

## Research Article

# Multilevel Feature Extraction Method for Adaptive Fault Diagnosis of Railway Axle Box Bearing

Zhigang Liu , Long Zhang , and Guoliang Xiong 

State Key Laboratory of Performance Monitoring Protecting of Rail Transit Infrastructure, East China Jiaotong University, Nanchang 330013, China

Correspondence should be addressed to Guoliang Xiong; [lgxcxx@ecjtu.edu.cn](mailto:lgxcxx@ecjtu.edu.cn)

Received 17 April 2023; Revised 8 July 2023; Accepted 13 July 2023; Published 7 August 2023

Academic Editor: Xingxing Jiang

Copyright © 2023 Zhigang Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Railway axle box bearing fault signal contains high  $Q$ -factor resonance and low  $Q$ -factor transient shock components with periodic transient shock features that can characterize bearing faults. However, extracting fault features is usually difficult due to noise, transmission paths, and high-amplitude accidental shocks. Therefore, to address the abovementioned problems, the multilevel feature extraction method for adaptive fault diagnosis of railway axle box bearing was proposed. This paper used the maximum second-order cyclostationary blind convolution (CYCBD) to weaken the influence of disturbances, enhancing weak transient shock components. Resonant sparse decomposition (RSSD) is used for the selection of quality factor  $Q$  due to its excellent fault feature separation property. Considering the strict fault frequency component periodicity in the envelope spectrum, the envelope spectrum multipoint kurtosis (ESMK) is proposed as a metric. The  $Q$  factor is optimized using the gray wolf optimization algorithm (GWO) to obtain an adaptive sparse decomposition method (GWO-RSSD) for extracting bearing transient fault shocks to eliminate the signal effects of high-amplitude disturbance shocks and background noise. The simulation results and measured signal analysis of railway axle box bearing show that ESMK can effectively measure periodic transient shocks under strong shock interference compared with the MED-RSSD method. Thus, GWO-RSSD can adaptively separate the optimal low- $Q$  resonance components, verifying the methods' effectiveness and superiority.

## 1. Introduction

Rolling bearings are widely used in machinery, transportation, aerospace, and other fields and are an important part of rotating machinery [1]. However, their harsh working environments can easily lead to fault, which can cause serious consequences if not detected on time. Therefore, an accurate determination of the rolling bearing health status is essential to improve the reliability and availability of mechanical equipment and ensure the safe operation of equipment [2, 3].

Periodic transient pulses capable of characterizing localized bearing faults are disturbed by transmission paths, strong background noise, and high-amplitude incidental shocks during signal acquisition, leading to difficulties in periodic transient-shock feature extraction [4]. Therefore, an effective feature extraction method is essential for bearing-

fault detection. Empirical mode decomposition (EMD), as a band decomposition method, can effectively decompose a signal into several approximate and detailed signals. Su used EMD to preprocess the original signal for noise reduction and a kurtogram for resonance filtering [5]. However, the kurtosis index used by the kurtogram cannot consider the periodic nature of the fault features, and EMD lacks a theoretical basis for the decomposition process, endpoint effects, and modal aliasing. Selesnick [17] proposed a signal resonance sparse decomposition (RSSD) method, which is different from linear methods based on band or scale decomposition, that decomposes the signal by the signal resonance properties (that is, quality factor size difference) of the fault signal into a high- $Q$  resonance component consisting of a continuous oscillation component and a low  $Q$  resonance component consisting of a transient shock component [6]. This method is being rapidly applied

to the field of rotating machinery fault diagnosis because of its unique pulse extraction advantages.

Wang first performed the overall empirical modal decomposition of the original signal with the maximum kurtosis as the index and further used RSSD to decompose the selected optimal Imf component to complete the fault diagnosis [7]; Li et al. [8] first used the kurtosis as the optimization index and intrinsic characteristic-scale decomposition (ICD) to preprocess the original signal. They further selected the best RSSD component for analysis based on the characteristic frequency domain ratio. Although the processing effect of the RSSD method is improved by some extent, the selection of its quality factor  $Q$  plays a decisive role in the final decomposition result; therefore, the selection of a suitable index to evaluate the  $Q$  value is a key to ensure the final decomposition effect. The abovementioned studies mostly use the kurtosis, which is susceptible to accidental disturbance shocks, as the optimization index. However, this index does not consider the characteristics of periodic rolling bearing fault shock characteristics, which affects diagnosis.

Vibration signals collected by sensors are usually the convolution between the fault transient impulses and background noise, and the impulse response function (transmission path) of the bearing system. To remove the interference of the transmission path on bearing fault characteristics, a series of deconvolution methods have emerged [9–11]. Barszcz and Sawalhi used minimum entropy deconvolution (MED) to eliminate the influence of the transmission path and achieve fault diagnosis by envelope spectrum analysis [12]. Ricci considered that it is difficult to effectively remove the influence of strong interference noise by a single signal processing method and proposed the use of EMD and MED to improve rolling bearing fault diagnosis [13]. Although the abovementioned methods can have an effect, MED takes the minimum entropy (the maximum kurtosis) as the measure, which has the same drawback as the kurtogram; that is, it does not consider the periodic characteristics of the fault impact characteristics and therefore prefers to extract a single transient pulse. To address these problems, maximum correlation kurtosis deconvolution (MCKD) has been widely used in recent years. Cheng et al. optimized the filter coefficients of the MCKD using the particle swarm algorithm and achieved better results [14]. However, the MCKD method sometimes fails because of strong background noise and variations in the fault cycle caused by speed fluctuations. McDonald et al. proposed the multipoint optimal minimum entropy deconvolution-adjusted (MOMEDA) method in recent years [11]. Shang et al. combined MOMEDA with IEWT for the feature extraction of vibration signals, further enhancing the effect of single MOMEDA feature extraction [15]. In recent years, BUZZONI proposed the maximum second-order cyclostationary blind deconvolution (CYCBD) algorithm, which uses the maximum second-order indicators of cyclostationary (ICS2) as the basis for finding the best inverse filter to extract the fault signal [16]. A comparison of MED, MCKD, and MOMEDA shows that CYCBD has a stronger fault feature extraction capability.

Based on the abovementioned analysis, this study proposes a new metric of periodic shock characteristics that is immune to accidental shocks and strong background noise—envelope spectral multipoint kurtosis (ESMK), based on the typical periodic characteristics of the fault characteristic frequency and its harmonic components in the bearing fault signal envelope spectrum. First, CYCBD is used to weaken the influence of transmission paths in the signal. Furthermore, to address the shortcomings of the existing RSSD method in which the quality factor  $Q$  relies on manual experience selection, the new metric—ESMK is used as the fitness function, and the gray wolf optimization algorithm (GWO) is used to optimize the quality factor  $Q$ . Then, an adaptive sparse decomposition method PSO-RSSD is obtained for transient shock feature extraction, which effectively eliminates noise interference such as chance shocks, and further realizes rolling bearing fault diagnosis using the envelope spectrum. The main innovations of the method are as follows: (i) a new index—ESMK that can effectively measure the periodic fault shock features despite the high amplitude of the chance shock interference is proposed; (ii) to reduce the RSSD quality factor reliance on manual experience, GWO is employed to optimize the  $Q$  factors; and (iii) both pre- and postprocessing steps take into account the periodic occurrence characteristic of fault shock features, which is expected to provide better diagnostic results. The rest of this paper is structured as follows: Section 2 is devoted to describe the basic principles of CYCBD. Section 3 introduces a brief description of RSSD and its drawbacks. Section 4 presents the principle of GWO and its advantages. Section 5 introduces the proposed new index—EASMK. Section 6 describes the detailed procedure of the proposed method and performs simulation experiments. In Section 7, the suggested approach is further examined for the detection of simulated signals, experimental signals, and railroad locomotive-bearing data. Finally, Section 8 presents the conclusions.

## 2. Maximum Second-Order Cyclostationary Blind Deconvolution

Similar to the other deconvolution algorithms, the main objective of the maximum second-order cyclostationary blind deconvolution (CYCBD) is to extract the fault signal from the complex observed signal. We assumed that the acquired vibration signals were as follows:

$$x = g * s_0 + n, \quad (1)$$

where  $n$  is the Gaussian white noise,  $s_0$  is the transient impulse of the fault,  $g$  is the response of the transmission path of the bearing system,  $*$  denotes convolution, and  $x$  is the original fault signal.

The essential part of the CYCBD is to establish an optimal FIR filter to restore the fault transient shock signal  $s$  by deconvoluting the collected signal  $x$ . The fault signal is extracted by determining the optimal inverse filter based on the maximum second-order cyclostationary index (ICS2). The deconvolution process can be expressed as

$$s = x * g \approx s_0, \quad (2)$$

where  $x$  is the measured signal,  $s$  is the deconvoluted signal, and  $*$  denotes the convolution operator.

The matrix form of equation (2) is as follows:

$$S = XG$$

$$\begin{bmatrix} s[N-1] \\ \vdots \\ s[L-1] \end{bmatrix} = \begin{bmatrix} x[N-1] \cdots x[0] \\ \vdots \ddots \vdots \\ x[L-1] \cdots s[L-N-2] \end{bmatrix} \begin{bmatrix} g[0] \\ \vdots \\ g[N-1] \end{bmatrix}, \quad (3)$$

where  $S$  is the discrete signal,  $L$  is the length of  $s$ ,  $g$  is the inverse filter, and  $N$  is the length of  $h$ . Thus, ICS2 can be defined as

$$ICS2 = \frac{g^H X^H W X g}{g^H X^H X g} = \frac{g^H R_{XWX} g}{g^H R_{XX} g}, \quad (4)$$

where  $R_{XWX}$  is the weighted correlation matrix and  $R_{XX}$  is the correlation matrix. The representation of the weighted matrix  $W$  is shown as follows:

$$W = \text{diag} \left( \frac{P[|s|^2]}{s^H s} \right) (L - N + 1) = \begin{pmatrix} \ddots & & 0 \\ & P[|s|^2] & \\ 0 & & \ddots \end{pmatrix} \frac{(L - N + 1)}{\sum_{l=N-1}^{L-1} |s|^2}$$

$$P[|s|^2] = \frac{1}{(L - N + 1)} \sum_k e_k (e_k^H |s|^2) = \frac{E E^H |s|^2}{L - N + 1}, \quad (5)$$

$$E = (e_1 \cdots e_k \cdots e_K),$$

$$e_k = \begin{pmatrix} e^{-j2\pi(k/T_s)(N-1)} \\ \vdots \\ e^{-j2\pi(k/T_s)(L-1)} \end{pmatrix},$$

where  $k$  is the sample index and  $T_s$  is the length of the fault cycle. Because the cycle frequency is considered to be the frequency associated with signal energy fluctuations which can be related to phenomena such as bearing or gear fault, the cycle frequency of a discrete signal can be defined as  $k/T_s$  [16], where  $k(n-1)/T_s$  in the superscript of any term of the vector  $e_k$  can be rewritten as  $k f_s t_{N-1}/T_s$ , where  $t_{N-1}$  is the  $N$ -th data corresponding to the time,  $f_s$  is the sampling frequency, and  $f_s/T_s$  is the rolling bearing fault.

The maximum ICS2 can be attained by resolving a generalized eigenvalue issue, where the maximum ICS2 is the maximum eigenvalue  $\lambda$  [16].

$$R_{XWX} h = R_{XX} h \lambda. \quad (6)$$

Because the weighting matrix is initialized by preconditioning the initial filter  $h$ , the largest ICS2 value must be obtained by the following iterative process:

- (1) Use the autoregressive (AR) model to initialize the whitening filter  $h$  for obtaining the filter coefficients
- (2) The weighting matrix  $W$  is calculated using the observed signals  $x$  and  $h$
- (3) Equation (6) is used to obtain the maximum eigenvalue  $\lambda$  and corresponding  $h$

- (4) Return to step (2) and use the newly obtained filter  $h$  to recalculate until convergence

### 3. Resonance Sparse Decomposition (RSSD)

In 2011 [17], Selenick proposed a sparse nonlinear decomposition method for signal resonance based on the tunable quality factor wavelet transform (TQWT) [18], which no longer uses the traditional method to divide the signal into different frequency bands but uses morphological component analysis (MCA) based on the quality factor  $Q$  difference corresponding to the harmonic and shock signals in the signal. Component analysis (MCA) [19] was used to separate the components possessing different oscillatory characteristics and obtain a low-quality resonant component containing the shock signal and a high quality resonant component containing the harmonic component. The quality factor  $Q$  is defined as the ratio of the center frequency to the bandwidth.

$$Q = \frac{f_w}{B_w}, \quad (7)$$

where  $f_w$  is the center frequency of the signal oscillation and  $B_w$  is the bandwidth.

In resonance sparse decomposition, the quality factor  $Q$  defines the resonance properties of a signal. If the transient shock is a single oscillation signal, the lower the quality factor  $Q$ , the lower the resonance property of the signal. If the harmonic part is a continuous oscillation signal, the higher the quality factor  $Q$ , the higher the resonance property of the signal. The resonance sparse decomposition method is based on the resonance properties of the signal and uses a double bandpass filter bank to implement the decomposition process, as shown in Figure 1. The basis function banks of high and low  $Q$  are obtained separately by the TQWT, and the corresponding transform coefficients are obtained by iterative calculations. The expressions of the low-pass and high-pass filter bank scale parameters are shown in the following equation (8):

$$\alpha = 1 - \frac{\beta}{r}, \beta = \frac{2}{Q+1}, \quad (8)$$

where  $r$  denotes the degree of redundancy. From equation (8), the low-pass and high-pass filters are determined by the scale factor, which in turn is determined by the quality factor and redundancy. Therefore, different quality factors and degrees of redundancy determine different wavelet filters.

In Figure 1,  $H_{i=0,1}(\omega)$  and  $H_i^* = 0, 1(\omega)$  are the frequency response functions of the filter decomposition and reconstruction, respectively;  $V_O(n)$  and  $V_1(n)$  are the filtered subband signals, and  $y(n)$  is the synthesized signal. The original fault signal  $x$  is expressed as

$$x = x_1 + x_2 + e, \quad (9)$$

where  $x_1$  is the harmonic signal,  $x_2$  is the fault shock signal, and  $e$  is the background noise. Because  $x_1$  and  $x_2$  have different quality factors, morphological component analysis is used to perform the nonlinear decomposition of the original signal  $x$ . The sparse representation process is a minimization problem, assuming that  $s_1$  and  $s_2$  denote the filter sets with high- and low-quality factors, respectively. The sparse decomposition objective function is constructed according to the morphological component analysis as follows:

$$\operatorname{argmin}_{w_1, w_2} \|x - S_1 W_1 - S_2 W_2\|_2^2 + \gamma_1 \|W_1\|_1 + \gamma_2 \|W_2\|_1, \quad (10)$$

where  $W_1$  and  $W_2$  denote the transformation coefficients of  $x_1$  and  $x_2$  under the basis functions  $S_1$  and  $S_2$ , respectively, and  $\gamma_1$  and  $\gamma_2$  are regularization parameter vectors.

Sparse decomposition can be viewed as the process of determining the optimal transformation coefficients  $W_1$  and  $W_2$  optimized using the split-increasing Lagrangian contraction algorithm to minimize the objective function. Therefore, the high- and low-resonance components obtained from the decomposition can be expressed as

$$\hat{x}_1 = S_1 W_1^* \quad \hat{x}_2 = S_2 W_2^*. \quad (11)$$

#### 4. Gray Wolf Optimization (GWO)

In 2014, Seyedai proposed the GWO algorithm, which was inspired by the division of labor in wolf foraging [20]. This new community optimization algorithm simulates the wolf ranking system and foraging behavior. The highest ranked wolf is species A, which is the top position of the pack and is able to lead and make decisions for other wolves. This is followed by species B, C, and E. Although wolves B and C are not at the top position, they can take over as new leaders when wolf A lost leadership. Wolf E is the lowest ranking wolf in the pack and is responsible for balancing the relationships in the pack.

Every wolf in the GWO method is taken as a possible solution, where wolf species A is the first optimal choice, while wolf species B and C are the second and second best choices, respectively. The GWO algorithm searches for prey by updating the distances and positions between them through an iterative optimization process in which the positions of wolf species A, B, and C are constantly updated, and the expressions are as follows:

$$D = |C \times X_p(t) - X(t)|, \quad (12)$$

$$X(t+1) = X_p(t) + D,$$

where  $D$  represents the distance between the gray wolf and the prey,  $X_p$  and  $X$  represent the prey as well as the location of the gray wolf, respectively, and the initial location coordinates are defined as  $(c, g)$ . The expressions for wolves A and C are  $A = 2a \times r_2 - a$  and  $C = 2r_1$ , respectively. The gray wolf will expand the search range in order to find the prey when  $|A| > 1$ . And the gray wolves will narrow the enclosure in order to search for nearby prey when  $|A| \leq 1$ . For  $a = 2 - (2t/t_{\max})$ , an increase in the number of iterations leads to a decrease in the convergence coefficient from 2 to 0.  $r_1$  and  $r_2$  are chosen randomly in the interval  $[0, 1]$ .

Wolf A leads wolves B and C to catch prey when wolves detect the location of prey; since wolves A, B, and C are nearest to the prey, they catch the prey according to

$$D_a = |C_1 \times X_a(t) - X(t)|,$$

$$D_b = |C_2 \times X_b(t) - X(t)|, \quad (13)$$

$$D_c = |C_3 \times X_c(t) - X(t)|,$$

where  $X_a$  represents the current position of wolf A,  $X_b$  that of wolf B, and  $X_c$  that of wolf C.  $C_1$ ,  $C_2$ , and  $C_3$  are random variables.  $X(t)$  denotes the present position of the wolf. The step length and direction of wolf E to wolves A, B, and C are determined by equations (14)–(16), while the final position of wolf E is defined by equation (17).

$$X_1 = |C_1 \times X_a - A_1 D_a|, \quad (14)$$

$$X_2 = |C_2 \times X_a - A_2 D_b|, \quad (15)$$

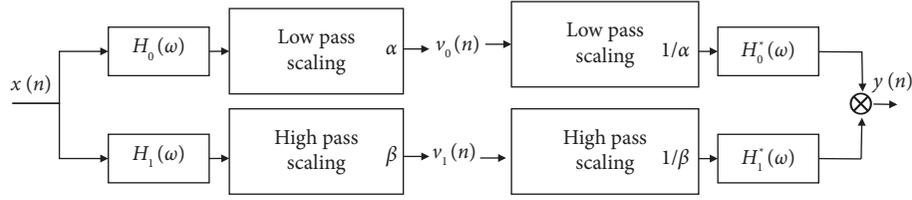


FIGURE 1: The two-channel filter bank.

$$X_3 = |C_3 \times X_a - A_3 D_c|, \quad (16)$$

$$X_{(t+1)} = \left| \frac{X_1 + X_2 + X_3}{3} \right|. \quad (17)$$

During the hunting process, wolves A, B, and C have different levels of adaptation to their prey. The different adaptation degrees are calculated to arrive at the first, second, and third optimal solutions and retain its current location information. Meanwhile, the wolf decides to move in the direction through this location information and finally approaches the prey and hunts successfully. Afterwards, the position of the wolf is updated. The location corresponding to the optimal solution is defined as  $(C_{best}, b_{best})$ .

## 5. Envelope Spectrum Multipoint Kurtosis

It is known from the cyclostationary theory that the transient shocks of rolling bearing fault signals have typical second-order cyclostationary characteristics; that is, their time-domain signals are not strictly periodic, but the transient energy components in their frequency spectra are periodic. Based on this, the fluctuation of the transient energy can be examined by calculating the multipoint kurtosis of the signal envelope spectrum.

Let the original signal be  $x(t)$  and the filtered signal be represented as

$$x_{an}(t) = x_{filtered}(t) + j\hat{x}_{filtered}(t), \quad (18)$$

where  $x_{filtered}(t)$  is the real part of the filtered signal and  $\hat{x}_{filtered}(t)$  is the Hilbert transform, so that the envelope of the filtered signal can be obtained according to equation (7).

$$S(t) = \sqrt{x_{filtered}^2(t) + \hat{x}_{filtered}^2(t)}. \quad (19)$$

The envelope spectrum of  $S(t)$  is represented as follows:

$$V(m) = \sum_{n=0}^{N-1} S(n) \exp\left(\frac{-i2\pi mn}{N}\right). \quad (20)$$

The envelope spectrum multipoint kurtosis ESMK is used as a metric for the periodic shock characteristics of localized bearing faults, and the standardized ESMK can be defined as

$$E_{kurt} = \frac{\left(\sum_{n=1}^{N-L} t_n^2\right) \sum_{n=1}^{N-L} (t_n V(m))^4}{\sum_{n=1}^{N-L} t_n^8 \left(\sum_{n=1}^{N-L} V(m)^2\right)^2}. \quad (21)$$

When the signal contains only one transient shock component, the signal has a large kurtosis value; when the signal contains multitransient shock sequences with a periodic distribution, the kurtosis value of the signal is smaller. The ESMK overcomes the disadvantage that the kurtosis index is easily disturbed by high-amplitude accidental shocks and can effectively identify faulty periodic transient shocks under strong disturbances.

## 6. The Proposed Bearing Fault Diagnosis Method

Periodic transient pulses that can characterize bearing faults are often disturbed by the transmission path, strong background noise, and high-amplitude incidental shocks during signal acquisition, making it difficult to extract the characteristics of periodic transient shocks, while the selection of the quality factor of the resonant sparse decomposition parameters seriously affects the decomposition results. The larger the quality factor  $Q$ , the higher the corresponding resonance, and the opposite corresponds to a lower resonance property.  $Q$  value that is too large or too small will affect the decomposition results. In the traditional RSSD method, the selection of the  $Q$  factor relies heavily on priori knowledge, and most of the current fault metrics are susceptible to disturbances, such as high-amplitude chance shocks. For example, it is difficult to effectively measure the periodic characteristics of cyclic transient shocks. Therefore, to address the abovementioned problems, the multilevel feature extraction method for adaptive fault diagnosis of rolling bearings is proposed. This method was developed considering the typical periodic characteristics of the fault feature frequencies and their harmonic components in the fault bearing envelope spectrum and the excellent performance of GWO global search. The implementation process is given below and the specific flow is shown in Figure 2:

- (1) Perform CYCBD deconvolution for the original signal to eliminate the transmission path effects and initially highlight fault shocks.
- (2) Set the initial conditions of the gray wolf optimization: the population size  $M = 10$  and the maximum number of iterations to 10. Furthermore, the variation ranges of the high- and low-quality factors  $Q1$  and  $Q2$  are  $[1.0, 3.0]$  and  $[4.0, 12.0]$ , respectively, and the redundancy factor  $r$  is 3.0.

- (3) Using the maximum ESMK of the low-resonance component as a metric, the optimal parameters of RSSD, high and low  $Q$  factors  $Q_1$  and  $Q_2$ , were optimized using GWO to acquire the optimal low-resonance components containing transient shock features.
- (4) Find the envelope spectrum of the best low-resonance transient component and compare it with the theoretical fault frequency of the bearing to complete fault diagnosis.

The simulation signal of the pure inner race fault transient shock is shown in Figure 3(a), where the characteristic frequency of the inner race fault is 90 Hz, and the signal sampling frequency is 20480 Hz. In order to make the simulated signal closer to the actual working condition, the Gaussian white noise component is further incorporated. The amplitude is 0.4, and the results are shown in Figure 3(b). The final signal obtained by convolving the impulse response function of the transmission path is shown in Figure 3(c), where the fault transient shock of inner race is no longer identifiable from the time domain under the interference of Gaussian noise. Figure 3(d) shows the envelope spectrum of Figure 3(c), from which the effective fault characteristic frequency components could not be found.

From the method shown in Figure 2, CYCBD is first used to preprocess the original signal to eliminate the influence of the transmission path, and the result is shown in Figure 4(a), indicating that the fault shock component is initially enhanced, but the shock component periodicity is still not sufficiently apparent to detect a bearing fault owing to the serious noise interference. Furthermore, the GWO-RSSD proposed in this study is used for transient shock feature extraction, and the parameters of the GWO algorithm are set with the variation range of high- and low-quality factors  $Q$ , as described in the previous section. With ESMK as the objective function, the optimal quality factors obtained were  $Q_1 = 11.3488$  and  $Q_2 = 1.0367$ , respectively. Then, the RSSD decomposition of the CYCBD deconvoluted signal is performed according to the optimal  $Q$  factor. The obtained optimal resonance components as well as the residual signals are shown in Figure 5. The final decomposition of the obtained  $Q_1$  signal is shown in Figure 5(b), and it includes mainly harmonic components. Figure 5(c) shows the signal corresponding to  $Q_2$ , where the shock component is relatively obvious and contains mainly transient shock information. It can be observed that the transient shock component is distinctly augmented. The envelope spectrum is presented as Figure 4(c). It reveals that the frequency component of 90 Hz, with obvious amplitude and side bands, and obvious harmonic components, such as 181 Hz and 269 Hz. Therefore, we can conclude that an inner race fault happened in the bearing. Therefore, the simulation signal analysis outcome validates the feasibility of the proposed bearing fault feature extraction method.

Aiming to solve the problem that the early weak fault features of bearing are easily overwhelmed by strong background noise, MED was used to filter and denoise the acquired vibration signals, which improved the signal-to-noise ratio of vibration signals and underlines the early weak

fault features of bearing. To show the advantages of proposed method, MED is used as the preprocessing method, and the postprocessing is optimized using PSO with the RSSD quality factor as a metric to compare and analyze the methods. The results of the analysis are shown in Figure 6. In the best low-resonance-component time-domain waveform shown in Figure 6(b), no periodic fault transient shock is found, and the envelope spectrum in Figure 6(c) does not have any obvious fault characteristic frequency component, which is affected by the high-amplitude interference shock and cannot determine whether the rolling bearing has a fault, which confirms the effectiveness of the proposed method.

## 7. Experimental Results and Discussion

*7.1. Experimental Signal Analysis of Inner Race Fault.* To simulate the local bearing fault in actual railroads, machinery, and other large equipment, the experimental signal of the outer race fault generated by the homemade rotor-bearing fault simulation test bench shown in Figure 7 is first used for the analysis, which can simulate different rolling bearing and rotor faults. The experimental bench included a servo motor and controller, support bearing, disc, bearing seat, accelerometer, computer, and data acquisition card. The vibration signal was collected by the accelerometer and saved on a computer. The bearing type used for testing was the N205. To simulate the local fault of the bearing, a 0.5 mm wide groove was machined in the outer race of the bearing using wire-cutting technology. The motor speed during the test was 1000 rpm, and the acceleration sensor was mounted directly above the experimental bearing housing, as shown in Figure 7(a). The sampling frequency is 12 000 Hz, the inner race fault is shown in Figure 7(b), the bearing parameters and speed are not repeated, and the formula can be calculated by the inner race bearing fault characteristic frequency  $BPF_1 = 129.15$  Hz.

The time-domain waveform of the bearing outer-ring fault signal collected by the sensor is shown in Figure 8(a), which is the result of the convolution of the periodic transient shock caused by the fault with the transfer function of the bearing system. The fault shock component in Figure 8(a) is more obvious, which is owing to the more standard manual machining groove, resulting in a larger fault shock amplitude. To bring the collected vibration signal closer to the on-site situation, the signal is shown in Figure 8(b) after adding Gaussian random noise with an amplitude of 4. The bearing inner-ring fault shock characteristics, such as Gaussian noise, have not been clearly identified.

The results obtained using the method presented in this study are shown in Figure 9. First, CYCBD preprocessing was performed on the original signal, and the results are shown in Figure 9(a). Compared with Figure 8(c), the fault shocks were initially enhanced. Furthermore, GWO-RSSD is used for the secondary enhancement of transient shocks, and the optimal high- and low-quality factors are  $Q_1 = 4.76$  and  $Q_2 = 2.94$ , respectively, after setting the parameters of the GWO algorithm and the range of high- and low-quality factors  $Q$ .

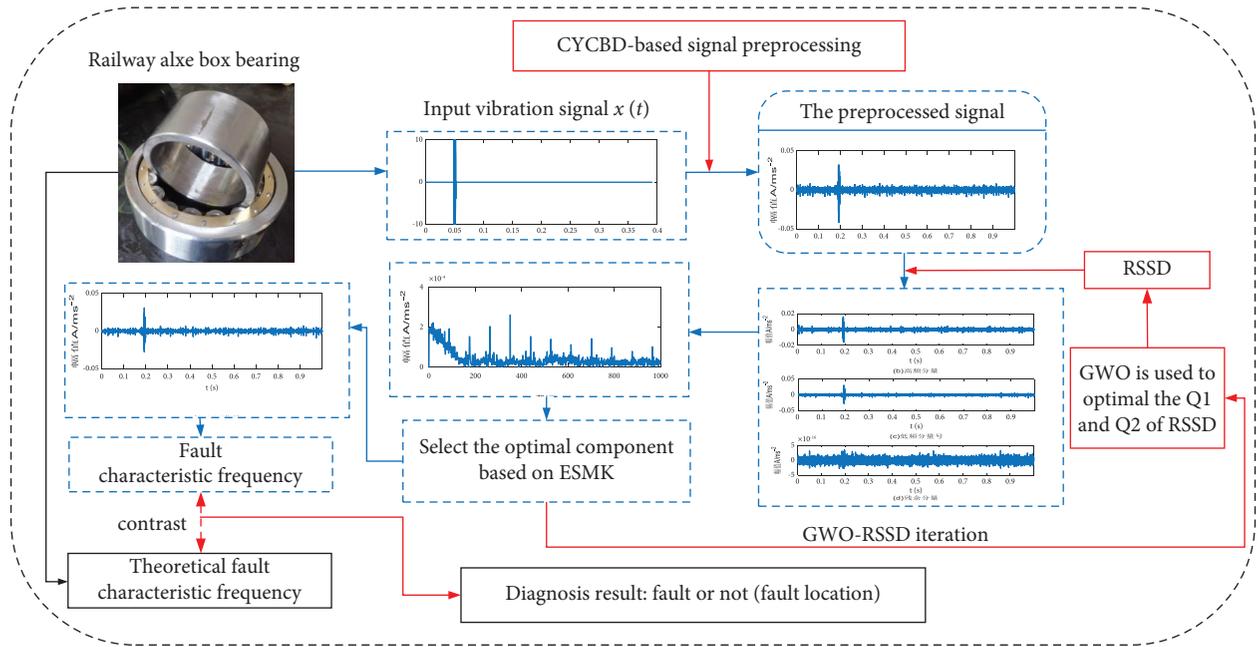


FIGURE 2: Flowchart of the proposed method.

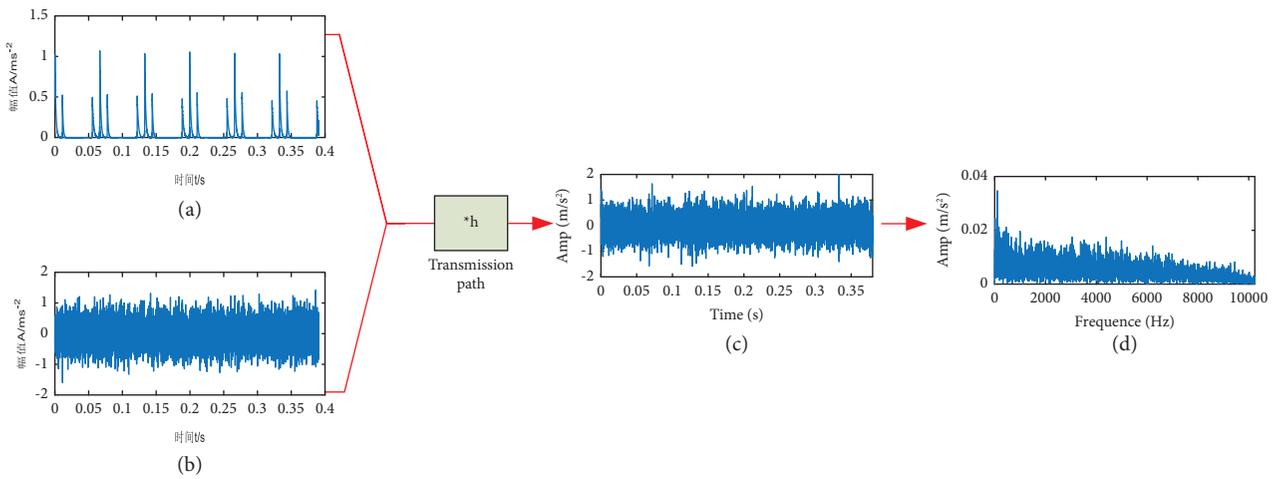


FIGURE 3: Simulated signals of inner race fault. (a) Fault transient impact. (b) Noise component. (c) Fault simulation signal. (d) Envelope spectrum.

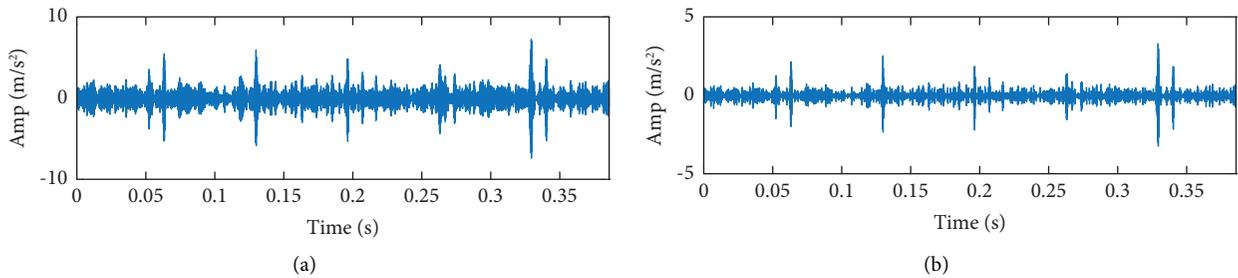


FIGURE 4: Continued.

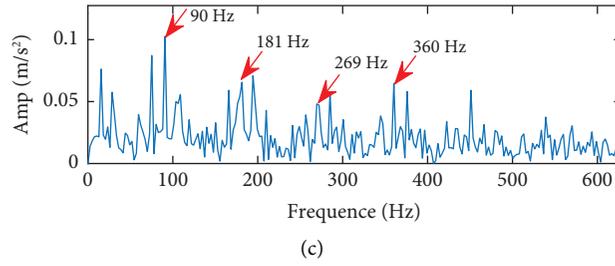


FIGURE 4: Simulated signal results using the proposed method.

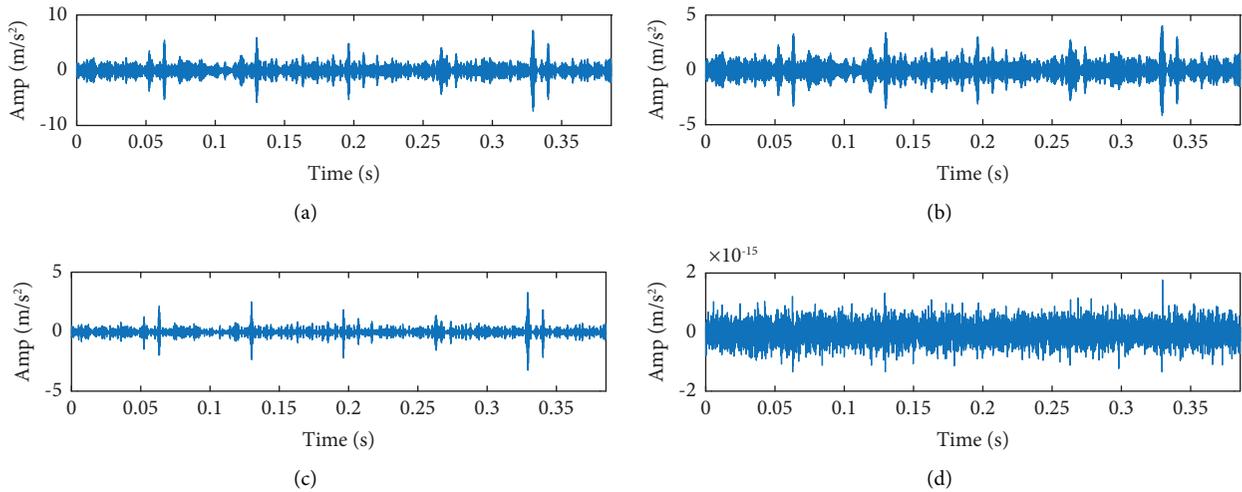


FIGURE 5: Results on sparse decomposition of simulation signal.

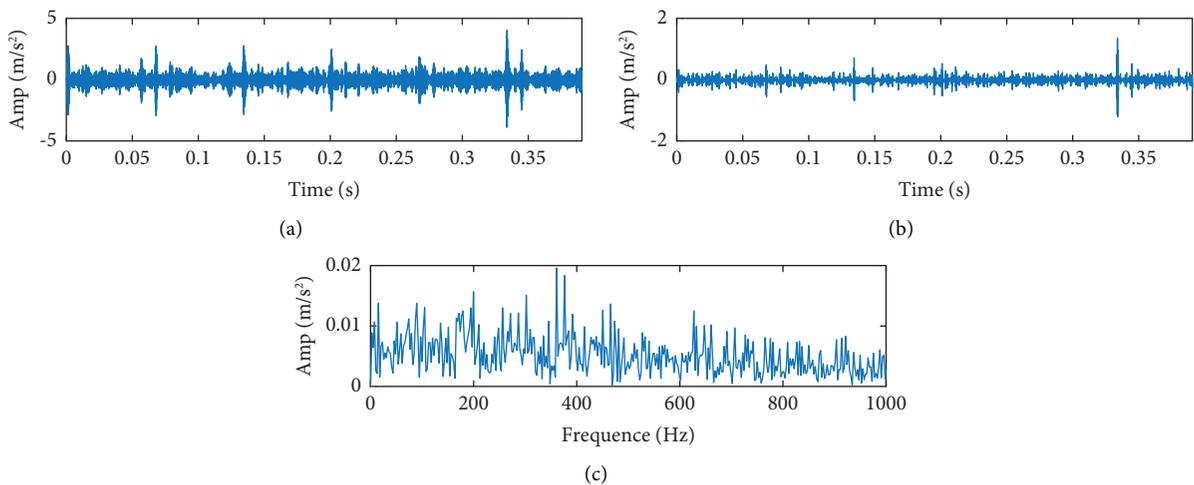


FIGURE 6: Results for the inner race fault simulated signal for comparison.

The RSSD decomposition of the preprocessed signal according to the optimal Q factors and the high- and low-resonance and residual components is shown in Figure 10. Figure 10(b) shows the best Q1 signal obtained, which consists mainly of harmonic components. Figure 10(c) shows the final obtained Q2 signal, looking to see the

apparent repetitive transient shock component. Figure 9(c) presents the envelope spectrum, in which a frequency component of 129 Hz with prominent amplitude can be observed. In addition, there are obvious harmonic components, such as 258 Hz and 387 Hz. At this point, it can be judged that an outer race fault happens in the bearing.

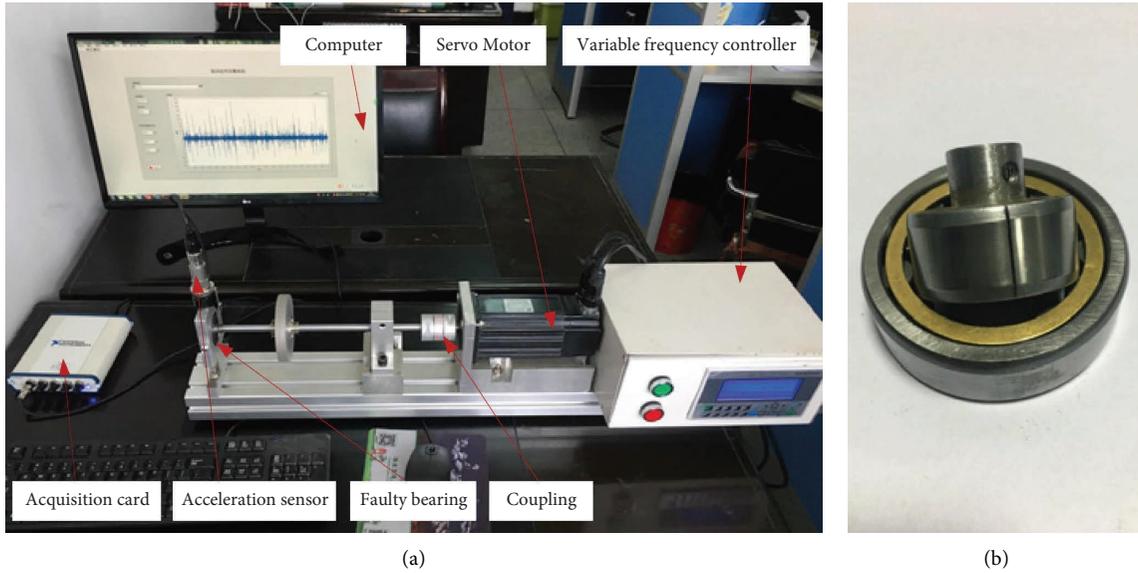


FIGURE 7: (a) Self-made rotor-bearing fault simulation test bench and (b) physical drawing of inner race fault bearing.

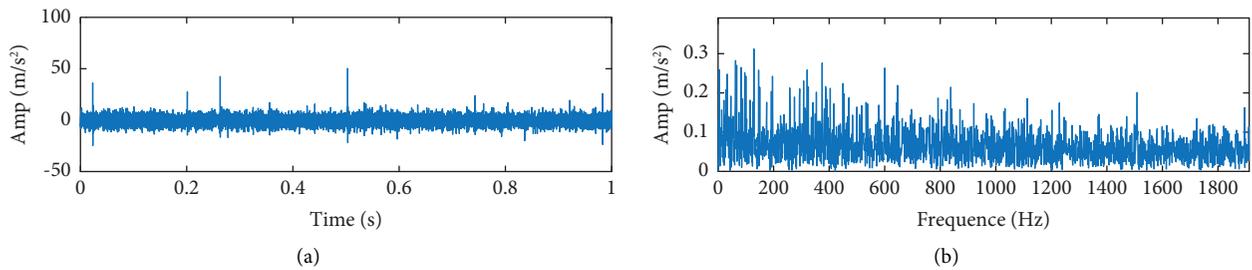


FIGURE 8: The signal and envelope spectrum of the inner race fault.

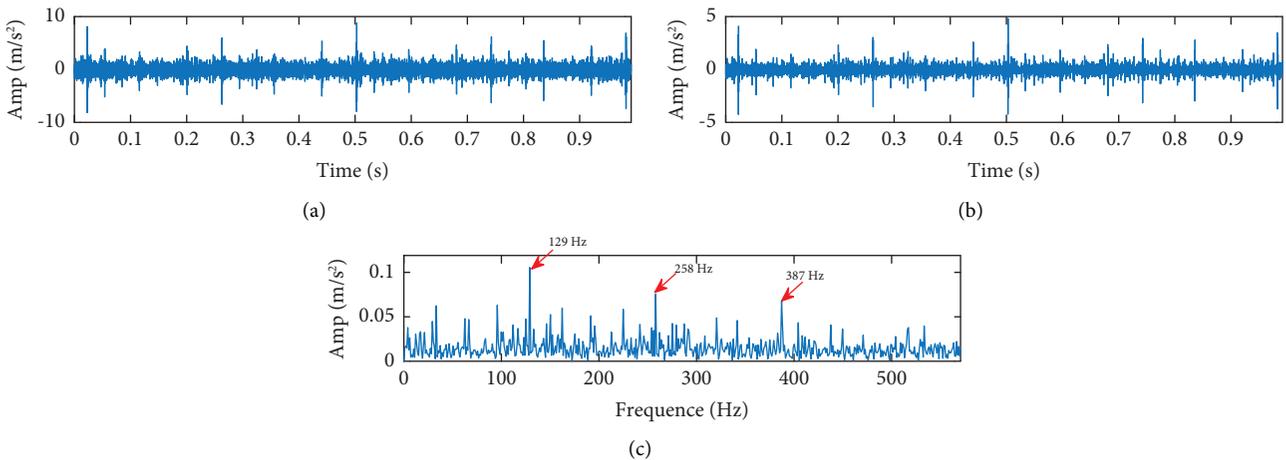


FIGURE 9: Results on the analysis of experimental signal using the proposed method.

Therefore, the results of the experimental data of the bearing outer race fault confirm the feasibility of the suggested approach to extract periodic defect characteristics under strong disturbances (such as accidental shocks with high amplitude).

For comparison, MED was used as the preprocessing method, and PSO was used as the optimization index for the RSSD quality factor in postprocessing, and the results are shown in Figure 11. The best low-Q resonance component time-domain waveform is shown in Figure 11(b), and there

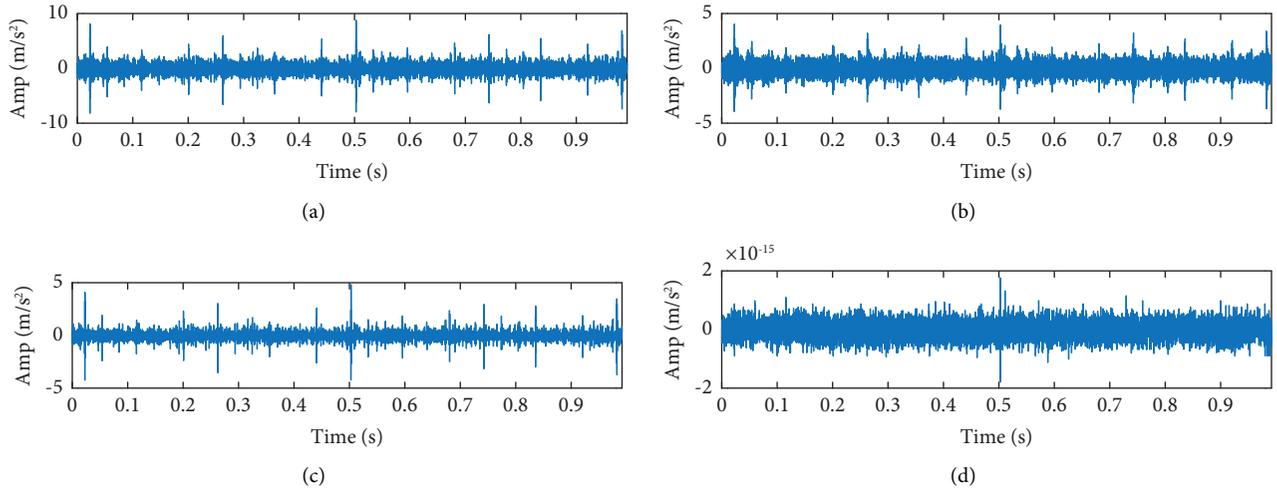


FIGURE 10: Results on sparse decomposition of experimental signal.

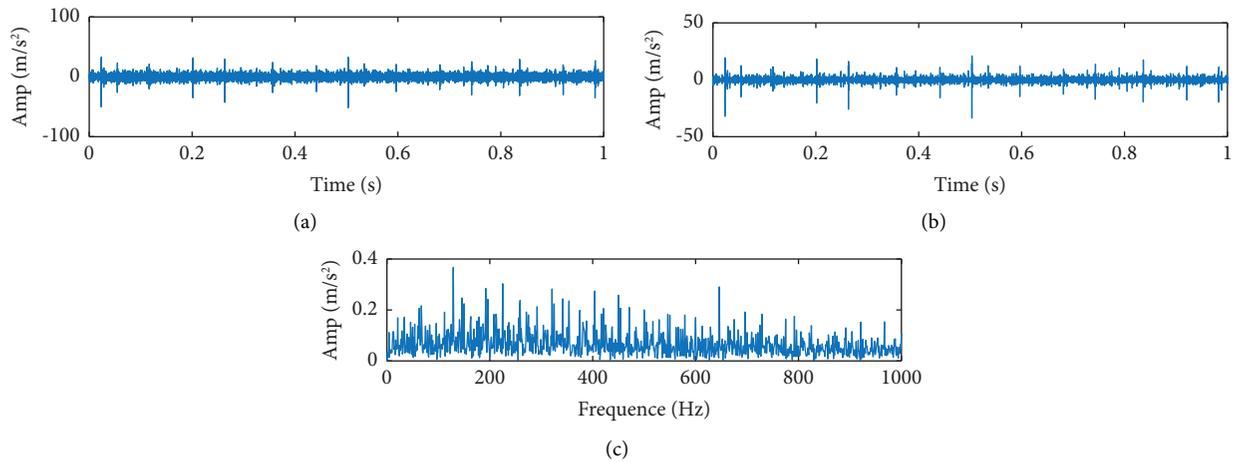


FIGURE 11: Results on inner race fault for comparison.

is a large amount of noise in its envelope spectrum in Figure 11(c). Furthermore, its feature extraction is poor compared with that of the proposed method.

**7.2. Engineering Practical Application of Railway Axle Box Bearing Fault Data Analysis.** The bearing used in this experiment was a faulty bearing of the DF4 internal combustion locomotive. DF4 locomotive is the most widely used diesel locomotive in China's railway history. Before the experiment, the outer ring failure bearings were divided into severe fault bearings with peeling outer ring, moderate fault bearings with chafing outer ring, and minor fault bearings with burns. Figure 12(a) shows the JL-501 locomotive wheel-to-wheel bearing fault test bench. The JL-501 locomotive bearing testing platform used in this manuscript comes from China Railway Group Limited. The test bench mainly includes four parts: headstock, hydraulic system, electrical system, and body. The platform uses a motor to drive the spindle, and the spindle speed range is 120 to 1200 r/min.

The maximum radial loading force of the tested bearing is 7500 N, and the radial force is applied by the hydraulic system. The minimum diameter of the inner ring is 120 mm and the maximum outer diameter of the outer ring is 340 mm, it can meet most engineering field practical axle box bearing models. The locomotive bearing model was NJ2232WB, and Figure 12(b) shows a faulty bearing with a fault size of 30 mm × 4 mm on the outer race. The local bearing fault diagram is shown in Figure 12(c). Three CA-YD-187T accelerometers were installed in the vertical and parallel directions on the outer race of the locomotive bearing, and vibration signals were collected using a national instrument data acquisition card. The bearing speed was 500 rpm and the sampling speed was 20,000 Hz. From the size of each component of the locomotive bearing and speed, the bearing outer race fault frequency was calculated as  $BPFO = 60.12$  Hz.

Figure 13(a) shows a segment containing 12,000 sampling points randomly intercepted from the original data, and the original signal is cluttered and the shock component

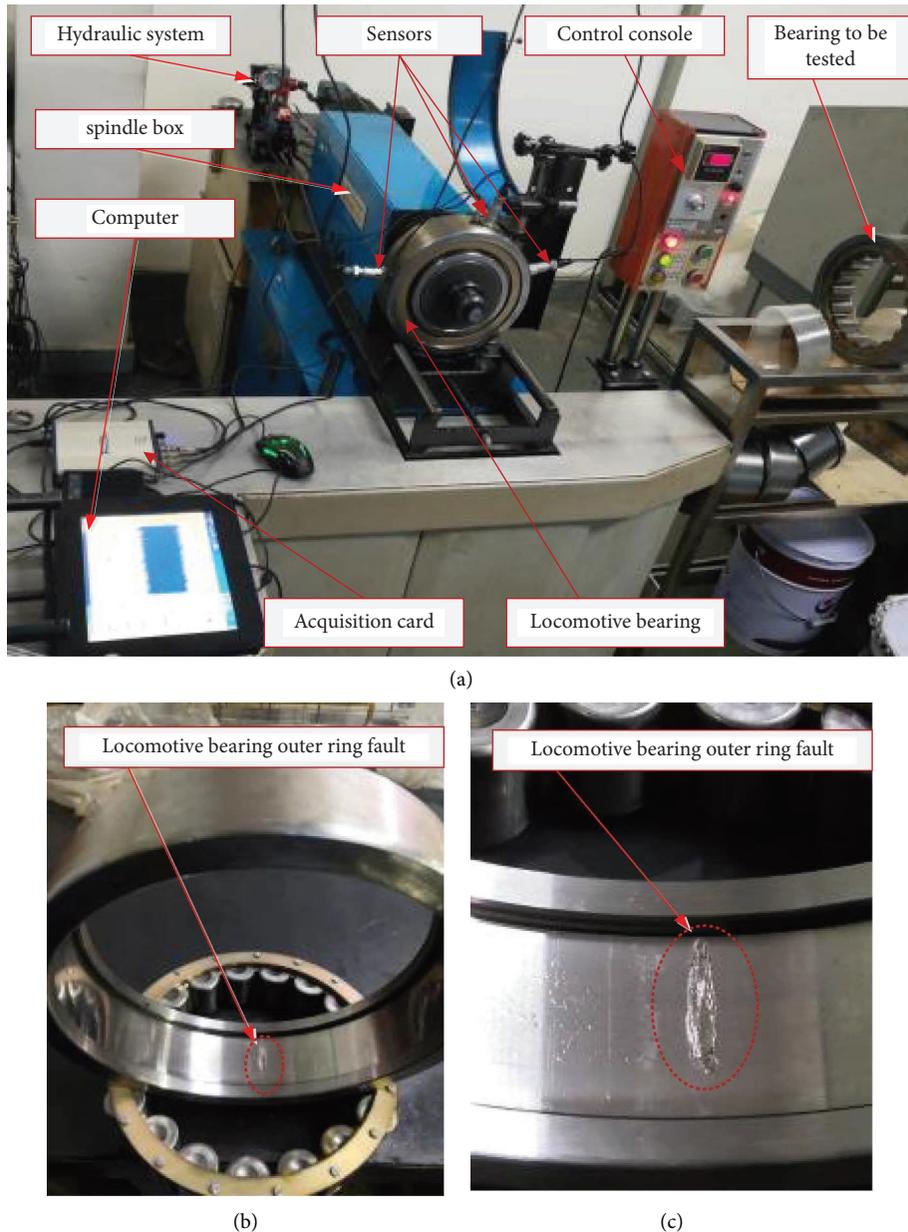


FIGURE 12: (a) JL-501 locomotive bearing testing platform; (b) bearing outer race fault; and (c) outer ring fault of partial enlargement.

is completely covered by noise. Figure 13(b) shows its envelope spectrum, and the inner race fault characteristic frequency cannot be clearly determined due to serious noise interference. The analysis results obtained using the proposed method are shown in Figure 14. First, CYCBD was used to preprocess the original signal, and the result is shown in Figure 14(a), indicating that the fault impact component was initially enhanced but not sufficiently for detecting bearing fault. Furthermore, GWO-RSSD was used for the secondary enhancement of transient shock characteristics, and the optimal high- and low-quality factors obtained were  $Q1 = 8.04$  and  $Q2 = 2.34$ , respectively, using the ESMK as the objective function. The time-domain waveform of the optimal low-resonance component is shown in Figure 14(b), showing that the shock component was significantly

enhanced, and its envelope spectrum is shown in Figure 14(c). The figure shows that the frequency component of 61 Hz coincides with the characteristic frequency of the inner race fault  $BPF1 = 61$  Hz with an obvious amplitude allowing with significant harmonic components such as 122 Hz and 183.9 Hz, indicating inner race bearing fault. The results of the experimental data analysis of the inner race fault verify the effectiveness of the method in the bearing vibration signal feature extraction in this paper.

For comparison, MED method was used as the preprocessing, and PSO was used as the optimization index for the RSSD quality factor in postprocessing, and the results are shown in Figure 15. The time-domain waveform of the optimal low  $Q$  resonance component is shown in Figure 15(b), and there is a large amount of noise in its

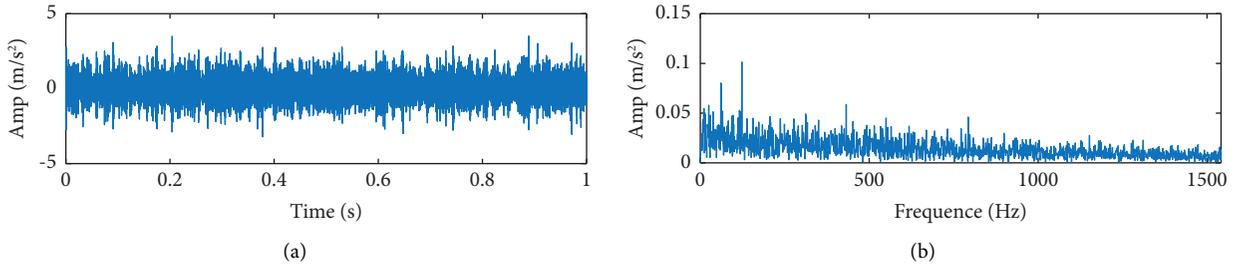


FIGURE 13: The signal and its envelope spectrum of outer race fault.

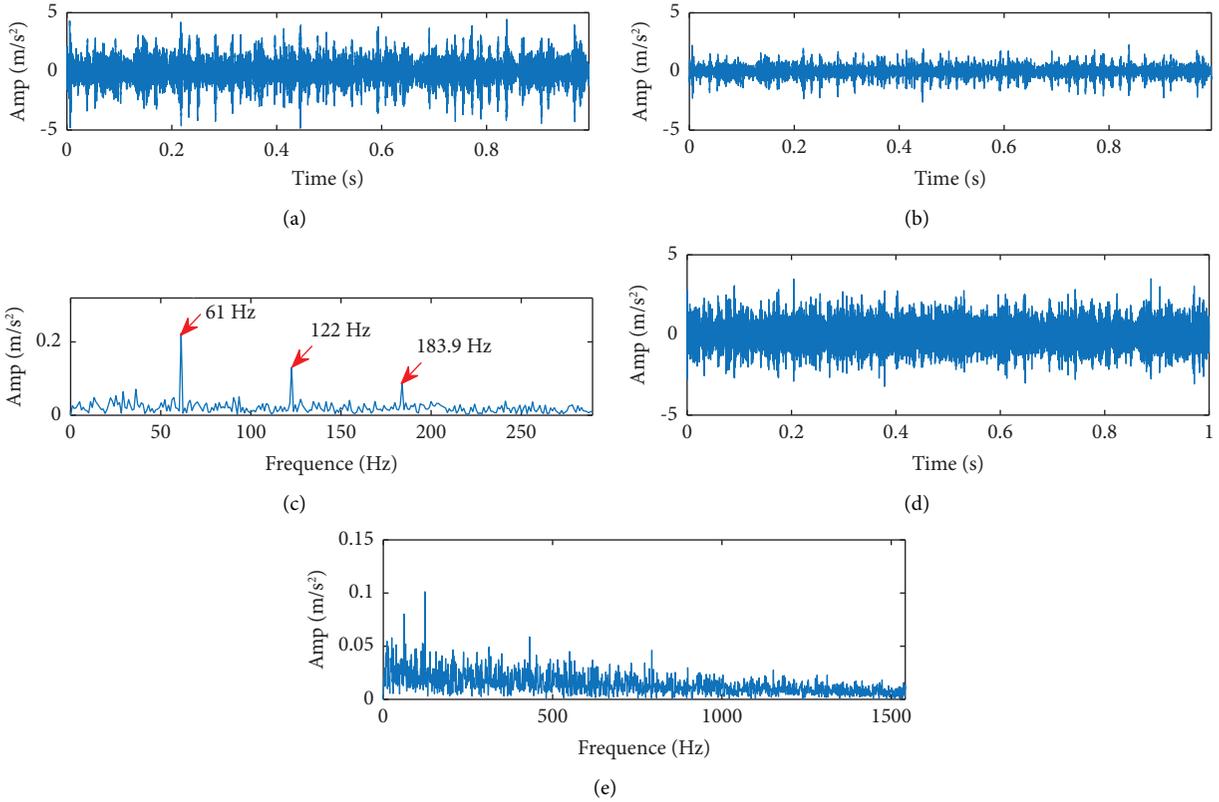


FIGURE 14: Results on the analysis of engineering actual signal using the proposed method.

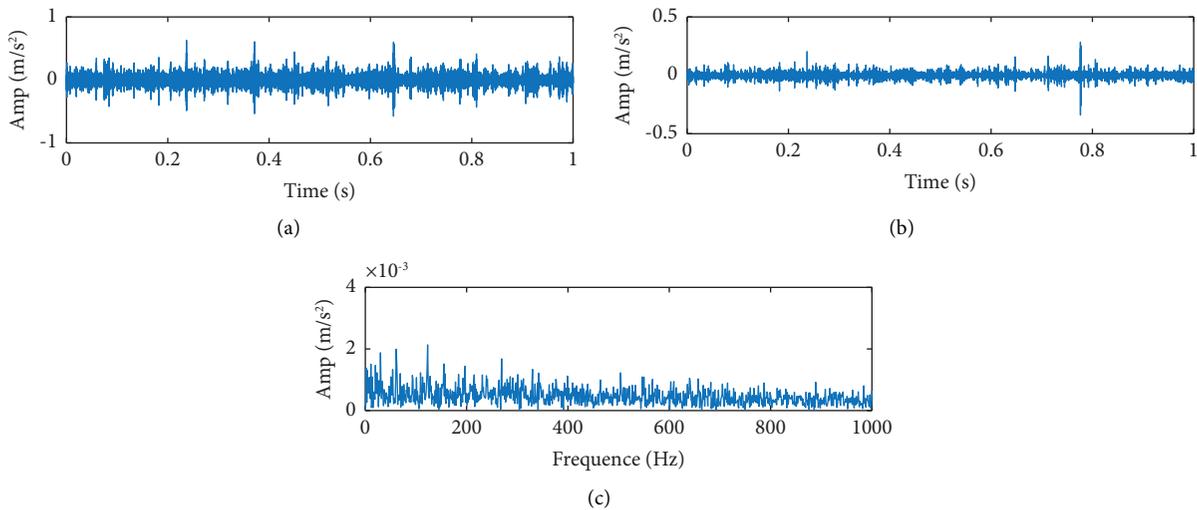


FIGURE 15: Results on outer race fault for comparison.

envelope spectrum in Figure 15(c), while its feature extraction effect is poor compared with the proposed method.

## 8. Conclusions

To improve current rolling bearing transient shock feature extraction methods mostly not considering the bearing transient shock cycle characteristics and the transmission path, accidental shock, interference noise, and other serious impacts on the fault diagnosis of the actual problem, a multilevel feature extraction method for adaptive fault diagnosis of rolling bearings was proposed. The following conclusions were drawn:

- (1) Signal resonance sparse decomposition (RSSD) can effectively separate high resonance and transient shock induced by fault. Considering the excellent performance of the GWO global search, using GWO for optimizing high and low Q-factors can effectively overcome the subjectivity of the RSSD quality factor relying on manual selection.
- (2) Considering the periodic characteristics of the frequency components in the ideal fault-bearing envelope spectrum, a new index of ESMK is proposed as the GWO optimization objective function, which can accurately eliminate the effects of strong disturbances, such as high-amplitude chance shocks.
- (3) The multilevel feature extraction method for adaptive fault diagnosis of rolling bearings can efficiently eliminate the impact of external accidental shocks and reduce the signal transmission path and noise interference. The proposed method can guarantee the effectiveness of fault diagnosis effectively and has greater advantages in bearing fault diagnosis compared with MED-RSSD method.

## Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

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

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 51865010) and the Natural Science Foundation of Jiangxi Province (nos. 20212BAB204007 and 20224ACB204017).

## References

- [1] S. Qiuyu, J. Xingxing, and D. Guifu, "Smart multichannel mode extraction for enhanced bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 189, 2023.
- [2] X. Jiang, J. Wang, C. Shen et al., "An adaptive and efficient variational mode decomposition and its application for bearing fault diagnosis," *Structural Health Monitoring*, vol. 20, no. 5, pp. 2708–2725, 2021.
- [3] C. Peeters, P. Guillaume, and J. Helsen, "A comparison of cepstral editing methods as signal pre-processing techniques for vibration-based bearing fault detection," *Mechanical Systems and Signal Processing*, vol. 91, pp. 354–381, 2017.
- [4] Q. Ni, J. C. Ji, K. Feng, and B. Halkon, "A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis," *Mechanical Systems and Signal Processing*, vol. 164, Article ID 108216, 2022.
- [5] W. Su, F. Wang, Z. Zhang, Z. Guo, and H. Li, "Application of EMD denoising and spectral kurtosis in early fault diagnosis of rolling element bearings," *Zhen Dong Yu Chong Ji*, vol. 29, no. 3, pp. 18–21, 2010.
- [6] I. W. Selesnick, "Resonance-based signal decomposition: a new sparsity-enabled signal analysis method," *Signal Processing*, vol. 91, no. 12, pp. 2793–2809, 2011.
- [7] H. Wang, J. Chen, and G. Dong, "Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform," *Mechanical Systems and Signal Processing*, vol. 48, pp. 103–119, 2014.
- [8] Y. Li, X. Liang, M. Xu, and W. Huang, "Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform," *Mechanical Systems and Signal Processing*, vol. 86, pp. 204–223, 2017.
- [9] R. A. Wiggins, "Minimum entropy deconvolution," *Geophysical Exploration*, vol. 16, no. 1-2, pp. 21–35, 1978.
- [10] G. L. McDonald, Q. Zhao, and M. J. Zuo, "Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection," *Mechanical Systems and Signal Processing*, vol. 33, pp. 237–255, 2012.
- [11] G. L. McDonald and Q. Zhao, "Multipoint optimal minimum entropy deconvolution and convolution fix: application to vibration fault detection," *Mechanical Systems and Signal Processing*, vol. 82, pp. 461–477, 2017.
- [12] T. Barszcz and N. Sawalhi, "Fault detection enhancement in rolling element bearings using the minimum entropy deconvolution," *Archives of Acoustics*, vol. 37, no. 2, pp. 131–141, 2012.
- [13] R. Ricci, P. Borghesani, S. Chatterton, and P. E. L. Pennacchi, "The combination of empirical mode decomposition and minimum entropy deconvolution for roller bearing diagnostics in non-stationary operation," in *Proceedings of the ASME 2012-International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 723–730, New York, NY, USA, June 2012.
- [14] Y. Cheng, Z. Wang, W. Zhang, and G. Huang, "Particle swarm optimization algorithm to solve the deconvolution problem for rolling element bearing fault diagnosis," *ISA Transactions*, vol. 90, pp. 244–267, 2019.
- [15] Z. Shang, G. Rui, L. Cheng, L. Xia, M. Gao, and J. Yun, "Rolling bearing fault diagnosis method based on MOMEDA and IEWT," *International Journal on Information and Management Sciences*, vol. 29, no. 4, pp. 345–363, 2018.
- [16] M. Buzzoni, J. Antoni, and G. D'Elia, "Blind deconvolution based on cyclostationarity maximization and its application to fault identification," *Journal of Sound and Vibration*, vol. 432, pp. 569–601, 2018.
- [17] I. W. Selesnick, "Wavelet transform with tunable Q-factor," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3560–3575, 2011.

- [18] J. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.
- [19] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.
- [20] H. Endo and R. B. Randall, "Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 906–919, 2007.