

sensors

Special Issue Reprint

Intelligent Sensors for Structural Health Monitoring and Mechanical Fault Diagnosis

Edited by
Zijian Qiao, Ke Feng and Zhihui Lai

mdpi.com/journal/sensors



Intelligent Sensors for Structural Health Monitoring and Mechanical Fault Diagnosis

Intelligent Sensors for Structural Health Monitoring and Mechanical Fault Diagnosis

Guest Editors

Zijian Qiao

Ke Feng

Zhihui Lai



Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors

Zijian Qiao

Faculty of Mechanical

Engineering and Mechanics

Ningbo University

Ningbo

China

Ke Feng

School of Mechanical

Engineering

Xi'an Jiaotong University

Xi'an

China

Zhihui Lai

College of Mechatronics and

Control Engineering

Shenzhen University

Shenzhen

China

Editorial Office

MDPI AG

Grosspeteranlage 5

4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Sensors* (ISSN 1424-8220), freely accessible at: https://www.mdpi.com/journal/sensors/special_issues/192H3X3UC3.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. <i>Journal Name</i> Year , Volume Number, Page Range.
--

ISBN 978-3-7258-4965-9 (Hbk)

ISBN 978-3-7258-4966-6 (PDF)

<https://doi.org/10.3390/books978-3-7258-4966-6>

© 2025 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

Baoxiang Wang and Chuancang Ding

An Adaptive Signal Denoising Method Based on Reweighted SVD for the Fault Diagnosis of Rolling Bearings

Reprinted from: *Sensors* **2025**, 25, 2470, <https://doi.org/10.3390/s25082470> 1

Abraham Perez-Alonzo, Luis Alvarez-Icaza and Gabriel E. Sandoval-Romero

Static Calibration of a New Three-Axis Fiber Bragg Grating-Based Optical Accelerometer

Reprinted from: *Sensors* **2025**, 25, 835, <https://doi.org/10.3390/s25030835> 20

Sakthivel Gnanasekaran, Lakshmi Pathi Jakkamputi, Jegadeeshwaran Rakkiyannan, Mohanraj Thangamuthu and Yogesh Bhalerao

A Comprehensive Approach for Detecting Brake Pad Defects Using Histogram and Wavelet Features with Nested Dichotomy Family Classifiers

Reprinted from: *Sensors* **2023**, 23, 9093, <https://doi.org/10.3390/s23229093> 39

Tao Wen, Congxin Wei, Zhenyi Wang, Linzhu Wang, Zihan Yang and Tingting Gu

Regional Thermal Radiation Characteristics of FY Satellite Remote Sensing Based on Big Data Analysis

Reprinted from: *Sensors* **2023**, 23, 8446, <https://doi.org/10.3390/s23208446> 59

Shuzhen Han, Pingjuan Niu, Shijie Luo, Yitong Li, Dong Zhen, Guojin Feng and Shengke Sun

A Novel Deep Convolutional Neural Network Combining Global Feature Extraction and Detailed Feature Extraction for Bearing Compound Fault Diagnosis

Reprinted from: *Sensors* **2023**, 23, 8060, <https://doi.org/10.3390/s23198060> 83

Jian Liu, Jiaqi Guo, Bing Hu, Qiqing Zhai, Can Tang and Wanjia Zhang

Controlled Symmetry with Woods-Saxon Stochastic Resonance Enabled Weak Fault Detection

Reprinted from: *Sensors* **2023**, 23, 5062, <https://doi.org/10.3390/s23115062> 99

Zhihui Lai, Zhangjun Huang, Min Xu, Chen Wang, Junchen Xu, Cailiang Zhang, et al.

High-Performance Adaptive Weak Fault Diagnosis Based on the Global Parameter Optimization Model of a Cascaded Stochastic Resonance System

Reprinted from: *Sensors* **2023**, 23, 4429, <https://doi.org/10.3390/s23094429> 116

Min Xu, Chao Zheng, Kelei Sun, Li Xu, Zijian Qiao and Zhihui Lai

Stochastic Resonance with Parameter Estimation for Enhancing Unknown Compound Fault Detection of Bearings

Reprinted from: *Sensors* **2023**, 23, 3860, <https://doi.org/10.3390/s23083860> 135

Article

An Adaptive Signal Denoising Method Based on Reweighted SVD for the Fault Diagnosis of Rolling Bearings

Baoxiang Wang ¹ and Chuancang Ding ^{2,*}

¹ School of Mechanical Engineering, Suzhou University of Science and Technology, Suzhou 215009, China; wangbaoxiang@usts.edu.cn

² School of Rail Transportation, Intelligent Urban Rail Engineering Research Center of Jiangsu Province, Soochow University, Suzhou 215131, China

* Correspondence: ccding@suda.edu.cn

Highlights:

What are the main findings?

- This paper proposed an adaptive signal denoising method based on frequency domain multipoint kurtosis (FDMK) and singular value decomposition (SVD).
- FDMK is used to identify sensitive singular components that contain fault-related information, and an estimation process is proposed to calculate the fault characteristic frequency.

What is the implication of the main finding?

- The proposed method FDMK-SVD can effectively extract fault features from raw vibration signals, even when faced with significant background noise and other interferences, enabling accurate fault diagnosis of rolling bearings.

Abstract: Due to the harsh and complex operating conditions, rolling element bearings (REBs) are prone to failures, which can result in significant economic losses and catastrophic breakdowns. To efficiently extract weak fault features from raw signals, singular value decomposition (SVD)-based signal denoising methods have been widely adopted in the field of rolling bearing fault diagnosis. In traditional SVD-based methods, singular components (SCs) with significant singular values are selected to reconstruct the denoised signal. However, this approach often overlooks low-energy SCs that contain important fault information, leading to inaccurate diagnosis. To address this issue, we propose a new selection scheme based on frequency domain multipoint kurtosis (FDMK), along with a reweighting strategy based on FDMK to further emphasize weak fault features. In addition, the estimation process of fault characteristic frequency is introduced, allowing FDMK to be calculated without prior information. The proposed FDMK-SVD can adaptively extract periodic fault features and accurately identify the health condition of REBs. The effectiveness of FDMK-SVD is validated using both simulated and experimental data obtained from a locomotive bearing test rig. The results show that FDMK-SVD can effectively extract fault features from raw vibration signals, even in the presence of severe background noise and other types of interferences.

Keywords: singular value decomposition; frequency domain multipoint kurtosis; signal denoising; rolling element bearings; fault diagnosis

1. Introduction

Rolling element bearings (REBs) are essential components of rotating machinery and play a significant role in the modern industry. Due to their harsh operating conditions, REBs are prone to unexpected failures, which can lead to machine breakdowns and substantial economic losses [1,2]. Consequently, the condition monitoring of REBs has become crucial, attracting significant attention from both academia and industry [3,4].

Since the vibration signal contains abundant fault information that can effectively reflect the health conditions of REBs, vibration analysis-based diagnostic methods have been widely applied for fault diagnosis [5]. When a fault occurs in a bearing, a series of repetitive impulses with a unique period are produced in the vibration signal and excite the resonances of the bearing and mechanical system [6,7]. Therefore, fault diagnosis of REBs can be achieved by detecting the periodical fault impulses in the time domain waveform or fault characteristic frequency (FCF) in the envelope spectrum domain. However, fault-induced repetitive transients are often submerged by strong background noise and other interferences, making fault diagnosis in real-world applications challenging. Therefore, extracting the repetitive impulses from the raw vibration signal is a critical and difficult task in the fault detection of REBs.

In recent years, many researchers have worked to address this issue, proposing various diagnostic methods, including spectral kurtosis [8,9], wavelet transform (WT) [10], empirical mode decomposition (EMD) [11,12], minimum entropy deconvolution (MED) [13–15], and singular value decomposition (SVD). Among these methods, SVD has garnered significant attention due to its ability to perform adaptive decomposition and its high computational efficiency.

Due to the abovementioned merits, SVD-based signal denoising methods have been widely employed in condition-based maintenance (CBM). Zhao et al. [16] illustrated the similarities between SVD and WT when the Hankel matrix is selected to convert one-dimension discrete signal into two-dimension matrix. On the basis of the above findings, a singular value decomposition packet (SVDP) algorithm was developed for decomposing vibration signals and extracting weak fault features [17]. Moreover, SVD has been employed to enhance the signal-to-noise ratio (SNR) in both the time and frequency domains [18], which efficiently promotes the improvement of the reliability of fault detection. Cong et al. presented a hybrid method consisting of SVD, singular value ratio, and short-time matrix series for the vibration signal analysis [19], which achieves an accurate diagnosis of rolling bearings. Duan et al. [20] proposed an impulsive feature extraction with improved SVD and sparsity-closing morphological analysis, in which A length factor is designed to optimize the construction of the matrix. Jiang et al. [21] combined SVD with the whale optimization algorithm for the quantitative detection of multiple damages in turbine blades. In addition to the SVD-based signal denoising method, SVD can also be combined with other fault diagnosis methods, where SVD is utilized to extract the fault feature for further signal processing. Jiang et al. [22] proposed a novel method combining SVD with a continuous hidden Markov model, achieving high classification accuracy in rolling bearing diagnosis. In addition, Kang et al. [23] presented an SVD-based method for fault classification, in which the features extracted using SVD were adopted as the input of the multi-layer support vector machines. Muruganatham et al. [24] took the singular value or the energy of SCs as the input features for the neural network and successfully achieved the automated fault diagnosis of REBs.

The key to the SVD-based signal denoising method is to determine an appropriate criterion for selecting the SC for subsequent signal reconstruction. A difference spectrum [25] was proposed to capture the sudden change point, which reflects the boundary between the signal and noise. These change points are then used to select effective singular values. A similar approach based on the correlation coefficient was introduced in [26]. However,

periodicity and impulsiveness are not taken into account in the above two methods, which may cause effective singular values to be ignored and even incorrect diagnostic results. To address this issue, a novel information index, the periodic modulation intensity (PMI) [27], was used to evaluate whether the SCs contained diagnostic information. This index helps to accurately select SCs that carry fault-related information, thereby improving the reliability of fault diagnosis. However, one challenge of PMI is that the fault period of interest must be predetermined, which can hinder its application in real-world situations in which prior knowledge of the fault period is not always available.

To overcome the drawbacks of the abovementioned SVD-based signal denoising method, this paper proposes a new selection scheme based on frequency domain multi-point kurtosis (FDMK) to select the effective SCs. FDMK evaluates SCs based on their periodicity and impulsiveness, ensuring the accurate selection of SCs that contain fault-related information, unlike traditional SVD-based methods that retain only large-energy SCs. Additionally, a reweighting strategy based on FDMK is introduced to enhance weak fault features during the reconstruction of the denoised signal. In addition, to solve the problem of the fault characteristic frequency or fault period not being given in advance, the process of estimating the fault period was incorporated, which allowed the proposed FDMK-SVD to operate effectively without prior knowledge of the fault period or characteristic frequency. Therefore, the proposed FDMK-SVD can adaptively achieve fault diagnosis and health assessment of REBs without prior information.

The remainder of this paper is organized as follows. In Section 2, the principle of the SVD-based signal denoising method is briefly introduced. In Section 3, the proposed FDMK-SVD for fault diagnosis of REBs is presented. Specifically, Sections 3.1 and 3.2 introduce the definition of FDMK and the estimation process of fault characteristic frequency (FCF), respectively, and the detailed flowchart of the FDMK-SVD is illustrated in Section 3.3. Section 4 presents the preliminary validation of the FDMK-SVD using a simulated signal. Subsequently, in Section 5, the vibration signal acquired from the experiment is used to further confirm the effectiveness of FDMK-SVD. Finally, the conclusions are summarized in Section 6.

2. Principle of SVD-Based Signal Denoising Approach

For a real matrix $A \in R^{m \times n}$, the SVD of the matrix A can be defined as follows [16,25].

$$A = U \Sigma V^T \quad (1)$$

where the matrices $U = [u_1, u_2, \dots, u_m] \in R^{m \times m}$ and matrix $V = [v_1, v_2, \dots, v_n] \in R^{n \times n}$ are orthogonal to each other. The column vectors of matrices U and V are the eigenvectors of AA^T and $A^T A$, respectively. $\Sigma = [\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k), \mathbf{0}] \in R^{m \times n}$ is a diagonal matrix, where $\mathbf{0}$ is a zero matrix, $k = \min(m, n)$, σ is the singular value of matrix A , and $\sigma_1 \geq \sigma_2, \dots, \geq \sigma_k$.

As illustrated in [25,27,28], SVD is effectively used for vibration signal denoising, which can be achieved through the following steps:

Step 1: Construction of Hankel matrix

In practice, the vibration signal is usually a one-dimensional time series rather than a matrix; thus, converting the one-dimensional signal into a matrix A is indispensable. Some matrix forms have been presented, such as the Toeplitz matrix, the cycle matrix, and the Hankel matrix. Because the Hankel matrix has zero phase-shifting characteristics and wavelet-like properties, the Hankel matrix is employed in [16].

The Hankel matrix of the discrete signal $x = [x(1), x(2), \dots, x(N)]$ can be written as [27]

$$A = \begin{bmatrix} x(1) & x(2) & \cdots & x(n) \\ x(2) & x(3) & \cdots & x(n+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(m) & x(m+1) & \cdots & x(N) \end{bmatrix} \quad (2)$$

where $1 < n < N$ and $m + n - 1 = N$. Parameter m is usually selected less than n , and it is suggested to set as 3 times the number of sub-components of the vibration signal [29]. This choice of m also determines the number of SCs after performing SVD.

Step 2: Signal decomposition and reconstruction

Once matrix A is constructed, it can be expressed as the product of matrices U , Σ , and V^T according to Equation (1). This means matrix A can be decomposed into a sum of unit-rank elementary matrices that are mutually orthogonal, which can be expressed as

$$\begin{aligned} A &= [u_1, u_2, \dots, u_m] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \sigma_m & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} \\ &= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_m u_m v_m^T \\ &= A_1 + A_2 + \dots + A_m \end{aligned} \quad (3)$$

where $u_i \in \mathbb{R}^{m \times 1}$ and $v_i \in \mathbb{R}^{n \times 1}$ are the i th column vectors of matrices U and V , respectively.

As illustrated in Equation (3), matrix A can be decomposed into the sum of sub-matrix A_i , and each singular component x_i corresponding to sub-matrix A_i can be obtained using two methods, as demonstrated in Figure 1. In the direct method, the singular component x_i can be obtained by concatenating the first row and sub-column vector in the last column of A_i , which can be written as

$$x_i = [R_{i,1}, C_{i,n}^T] \quad (4)$$

where $R_{i,1} \in \mathbb{R}^{n \times 1}$ and $C_{i,n} \in \mathbb{R}^{(m-1) \times 1}$. Another way to obtain the singular component x_i is by diagonal averaging, where each element of x_i can be obtained by averaging along the anti-diagonals, as shown in Figure 1b. Interestingly, regardless of any of the methods mentioned above being used, the raw signal can be perfectly reconstructed by adding all decomposed SCs, as shown in Equation (5).

$$x = \sum_{i=1}^m x_i \quad (5)$$

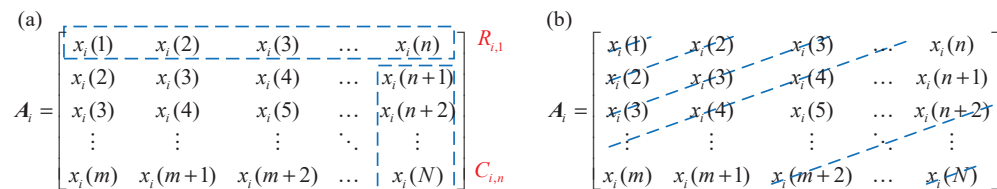


Figure 1. Obtaining x_i from the sub-matrix A_i . (a) direct method; and (b) anti-diagonals averaging method.

Step 3: Signal denoising

As illustrated in Equations (3) and (5), each singular value is associated with an SC. This means that the selection of effective SCs can be transformed into the selection of

the corresponding singular values. Therefore, the SVD-based signal denoising method can be achieved by selecting some effective SCs and simply adding their corresponding component signals together. In general, the SCs with large singular values are considered as the effective components and would be retained [25], which can be expressed as

$$\tilde{\mathbf{A}} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_k & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\tau \\ \mathbf{v}_2^\tau \\ \vdots \\ \mathbf{v}_n^\tau \end{bmatrix} = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^\tau = \sum_{i=1}^k \mathbf{A}_i \Rightarrow \tilde{\mathbf{x}} = \sum_{i=1}^k \mathbf{x}_i \quad (6)$$

where the first k largest singular values are added together; $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{x}}$ represent the denoized matrix and signal, respectively.

In the abovementioned SVD-based signal processing approach, it is important to determine the proper threshold k to guarantee denoising performance. The selection of k has been investigated in several studies, such as the difference spectrum [25], correlation coefficients [26], and the median value of singular values [28]. Among these methods, the difference spectrum is frequently adopted because of its easy implementation and good performance in applications. More details about the difference spectrum can be found in [28].

3. Reweighted Singular Value Decomposition Based on Frequency Domain Multipoint Kurtosis

3.1. Frequency Domain Multipoint Kurtosis

In the fault diagnosis of REBs, many indicators can be used to detect whether the signal contains fault information and quantify the extent of this information. Common indicators include Shannon entropy [30,31], Gini index [32], kurtosis [33], and l_2/l_1 norm [34]. However, most of these indicators mentioned above are introduced from the field of statistics, economics, and information theory, and the periodicity of the defect impulses caused by defective rotating machinery is not considered. In order to take into account the periodicity of the defect impulses, several other indicators, such as correlated kurtosis [35], envelope harmonic-to-noise ratio (EHNR) [36], and multipoint kurtosis (MK) [37], have been proposed in the literature. In this paper, the multipoint kurtosis that considers the periodicity of the defect impulses is utilized to select the sensitive SCs containing the information related to the health conditions of REBs.

The multipoint kurtosis of M -shift of zero-mean data $\mathbf{y}$ is expressed as [37]

$$MK(\vec{\mathbf{y}}, \vec{\mathbf{t}}) = \frac{\left(\sum_{n=1}^N t_n^2 \right)^2 \sum_{n=1}^N (t_n \mathbf{y}_n)^4}{\sum_{n=1}^N t_n^8 \left(\sum_{n=1}^N \mathbf{y}_n^2 \right)^2} \quad (7)$$

where N is the length of the original signal $\mathbf{y}$ and $\vec{\mathbf{t}}$ is the target vector. For example, $\vec{\mathbf{t}}$ can be defined as

$$\vec{\mathbf{t}} = [0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ \dots \ 0 \ 1]^T \quad (8)$$

The above $\vec{\mathbf{t}}$ controls the locations of the interested multiple points and the interval between adjacent peaks or impulses, where the interval is four points and the first impulse is located at the 4th point.

Although the MK can extract repetitive impulses, a major drawback still exists that hinders its application. As demonstrated in Equation (8), the locations of the impulses are assumed to be known in advance. However, it is impossible to determine the locations of fault impulses in actual situations, which makes it unrealistic to calculate the MK directly in the time domain. As illustrated in the previous literature, the position of the frequency lines associated with the FCF and its harmonics is constant in the envelope spectrum for periodic fault transients, regardless of their positions in the time waveform. Therefore, it is practical to calculate the MK in the frequency domain (envelope spectrum), and FDMK can be defined by

$$FDMK(\vec{y}, \vec{t}) = \frac{\left(\sum_{n=1}^N t_n^2\right)^2 \sum_{n=1}^N (t_n S_n)^4}{\sum_{n=1}^N t_n^8 \left(\sum_{n=1}^N S_n^2\right)^2} \quad (9)$$

where S represents the envelope spectrum of the signal and $\vec{t}$ represents the target vector. According to this definition, it can be observed that the FCF and sampling frequency f_s must be predetermined to construct the target vector $\vec{t}$. However, it is sometimes impossible to obtain the FCF in advance, or the theoretical FCF deviates from the real value, which leads to an inaccurate FDMK and even incorrect diagnostic results. To address this issue, a new strategy that can adaptively estimate the fault characteristic frequency is presented in the next section.

3.2. The Estimation of Fault Characteristic Frequency and Construction of Target Vector $\vec{t}$

As demonstrated in [38,39], the autocorrelation function can be utilized to estimate the period of the fault impulses. The definition of autocorrelation function of $x(t)$ is expressed as follows

$$r_x(\tau) = \int x(t)x(t+\tau)dt \quad (10)$$

As illustrated in Figure 2, τ is the time lag, and the autocorrelation function of $x(t)$ reaches a local maximum when $\tau = \tau_{\max}$. Therefore, $\tau_{\max}$ is considered the fault period.

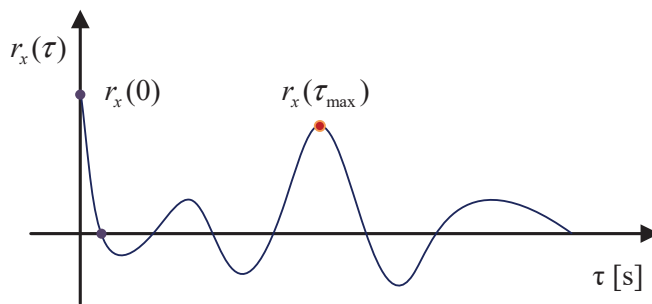


Figure 2. The autocorrelation function of $x(t)$.

However, the above method for estimating the period or FCF through the envelope autocorrelation function of the raw vibration signal may fail when the defect impulses are submerged by harmonics or severe background noise. To overcome these limitations, a new strategy based on SVD to estimate the period or fault characteristic frequency is proposed in this paper.

Step 1: Obtain the singular components using SVD.

In real applications, defect impulses are usually overwhelmed by harmonic components and background noise, leading to inaccurate estimates of the fault period or FCF from the raw signal. Therefore, SVD is adopted to separate the components in the raw signal, and SCs are obtained.

Step 2: Calculate the Gini index (GI) of each SC and then calculate the envelope harmonic-to-noise ratio (EHNR) to select the most sensitive SC.

In this part, GI and EHNR are introduced to select the most sensitive SC containing the diagnostic information. The definitions of GI and EHNR are provided below.

The Gini index (GI) was originally used to represent the wealth distribution of a nation's population. In recent years, the GI has been increasingly used as a measure of signal sparsity [32,40,41]. The GI values range from 0 to 1, and a large GI value indicates that the signal is sparser. For a real signal indexed in non-decreasing order, its GI can be given below

$$GI(x) = 1 - 2 \sum_{i=1}^N \frac{|x_i|}{\|x\|_1} \left(\frac{N - i + 0.5}{N} \right) \quad (11)$$

where $\|x\|_1$ represents the l_1 -norm of x and N is its length.

EHNR is proposed for detecting periodic impulses [39,42], and the formula is defined as follows:

$$EHNR = \frac{r_{E[x(t)]}(\tau_{\max})}{r_{E[x(t)]}(0) - r_{E[x(t)]}(\tau_{\max})} \quad (12)$$

where $E[x(t)]$ represents the envelope of signal x , and the autocorrelation function of $E[x(t)]$ reaches a local maximum when $\tau = \tau_{\max}$.

After separating the components contained in the raw signal, the GI value of each singular component is calculated, and the EHNR of the first several SCs with large GI values are further calculated. In this way, the periodicity and sparsity of the defect impulse are considered at the same time. The SC with the maximum ENHR is considered the most sensitive one, which is related to the fault feature.

Step 3: Estimate the fault period or fault characteristic frequency

The envelope autocorrelation function of the selected SC is calculated, and the time lag $\tau_{\max}$ that causes the envelope autocorrelation function to reach a local maximum is considered the fault period. Correspondingly, the fault characteristic frequency is $1/\tau_{\max}$.

3.3. The Proposed FDMK-SVD for Fault Detection of REBs

The framework of the proposed FDMK-SVD for bearing fault detection is shown in Figure 3, and the concrete steps are presented as follows:

Step 1: Signal decomposition

A one-dimensional signal must first be converted into a Hankel matrix, as illustrated in Section 2 for SVD. When a relatively large m is selected, the raw signal, consisting of fault impulse components, harmonic components, and other interferences, can be separated, and the computational cost increases as m increases. According to the analysis of the simulated signal and experimental signals, $m = 30$ is appropriate for separating the components contained in the vibration signal under different situations.

Step 2: Estimate the fault characteristic frequency

After obtaining SCs, GI and EHNR are set as the criteria to select the most sensitive SC, followed by the estimation of the fault period or FCF using the envelope autocorrelation function of the selected SC.

Step 3: Construct the target vector $\vec{t}$ After obtaining the characteristic frequency, the corresponding target vector $\vec{t}$ can be generated. Because the estimated fault characteristic has a slight deviation from the real value, the locations of '1' in the target vector $\vec{t}$ are determined by seeking the locations of the local maximum around the estimated value, which indicates that '1' is not equally distributed.

Step 4: Calculate the FDMK of each singular component

In this step, the FDMK of each singular component is calculated to evaluate whether the diagnostic information is included.

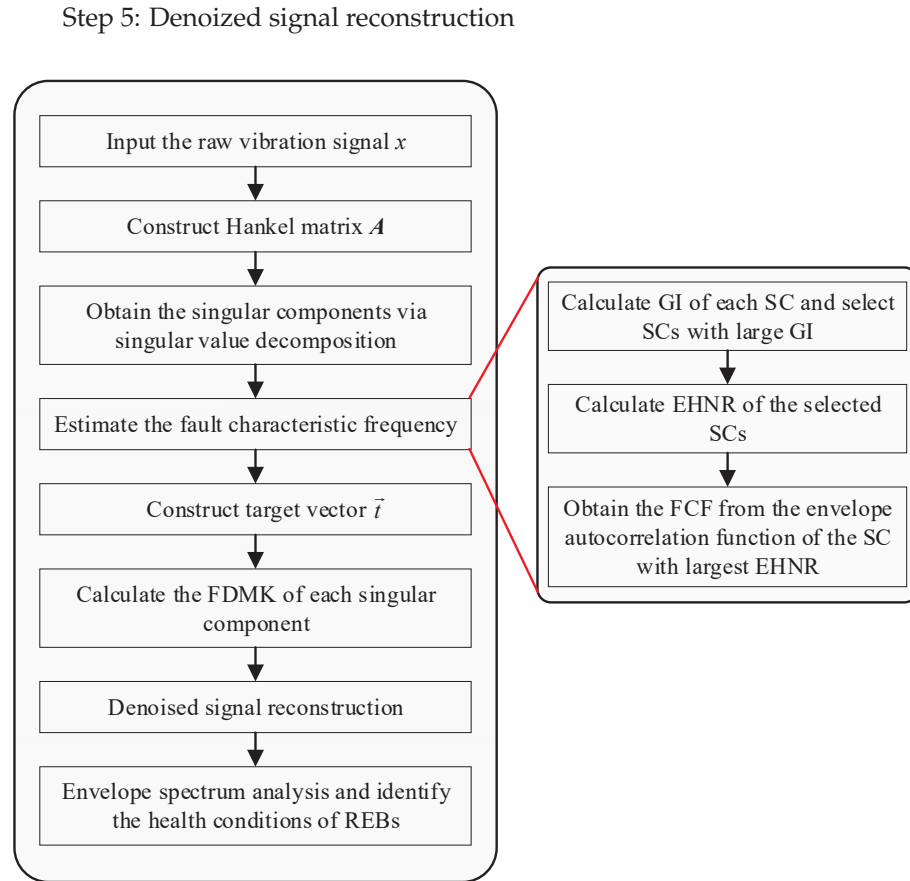


Figure 3. The flowchart of the proposed FDMK-SVD.

The denoised signal is obtained by assigning a new weighting coefficient to each SC and then adding the weighted SCs, as given in Equations (13) and (14).

$$\tilde{X} = w(1)X_1 + w(2)X_2 + w(3)X_3 + \cdots + w(m)X_m \quad (13)$$

$$w(i) = \begin{cases} 0 & mk_i \leq threshold \\ mk_i / \sum_i mk_i & mk_i > threshold \end{cases} \quad (14)$$

In Equation (14), the weighting function $w(i)$ determines whether each SC is highlighted or suppressed when reconstructing the denoised signal. The greater the weight value of each SC, the greater its contribution to reconstructing the denoised signal. For each SC, if its corresponding $mk_i \leq threshold$, then the SC is considered to contain no fault feature information, and it is culled by assigning it a weight of zero when reconstructing the denoised signal. Conversely, if the SCs with $mk_i > threshold$, then it is considered to contain fault information and should be preserved. Furthermore, since different SC contain different levels of information, SC is reweighted by the normalized FDMK, as shown in Equation (14). This ensures that SCs with larger FDMK can be highlighted more effectively in the reconstruction of the denoised signal. In this paper, the threshold is empirically set to 0.02.

Step 6: Envelope spectrum analysis and condition identification

Finally, the envelope spectrum of the reconstructed denoised signal is analyzed to detect the fault characteristic frequencies. By identifying the presence of these frequencies in the envelope spectrum, the health condition of the REBs can be accurately assessed.

4. Simulation Analysis

In this section, a simulated signal of a faulty rolling bearing is constructed to verify the effectiveness of the proposed FDMK-SVD. As mentioned above, the bearing vibration signal collected from an actual industrial environment contains multiple signal components while being submerged in heavy background noise. To create a more realistic signal model, we incorporate four common signal components into the simulated signal of a faulty rolling bearing, as described below [43].

$$y(t) = \underbrace{\sum_n A_n \cos(2\pi n f_0 t + a_n)}_{\text{Rotor vibration}} + \underbrace{\sum_n B_n \cos(2\pi n T f_0 t + \beta_n)}_{\text{Gear meshing}} + \underbrace{\sum_i D_i S_d(t - iT_d - \tau_i)}_{\text{Defect impulses}} + \underbrace{\sum_i R_i S_r(t - T_i)}_{\text{Random shocks}} + n(t) \quad (15)$$

The first part is the vibration from the rotor or shaft, where f_0 is the rotation frequency, A_n and a_n represent the amplitude and phase of the n th harmonic, respectively. The vibration signal originating from gear meshing is represented by the second part, where the number of gear teeth is denoted as T , and the amplitude and phase of the n th meshing frequency are represented by B_n and β_n . The third part represents the defect impulses caused by the failure of the rolling bearing, where the amplitude of the i th impulse is represented by D_i , and T_d represents the interval time between the two adjacent impulses. Moreover, τ_i is used to simulate the random slip of the bearing, which is usually set to 1~2% of the fault period.

In general, the defect impulses can be represented as an exponentially decaying sinusoid, as follows:

$$S_d(t) = e^{(-\alpha_r t)} \sin(2f_r t) \quad (16)$$

where f_r is the resonant frequency of the impulse response and α_r is the damping ratio of the impulses.

The fourth part is the external random shock, where R_i and T_i denote the amplitude and occurrence time of the i th random shock, respectively. Furthermore, $n(t)$ represents white Gaussian noise.

Table 1 lists the parameters of the simulated signal. In the simulation, the time length of the simulated signal and sampling frequency are set as 1 s and 10,000 Hz, respectively. Figure 4a–d display the time waveforms of the rotor vibration, gear meshing, defective impulses and random shocks, respectively. After adding white Gaussian noise with an SNR of −3 dB, the synthetic signal is shown in Figure 4e, and no obvious defect impulses can be recognized. The envelope spectrum of the simulated signal is presented in Figure 4f, where the frequency lines of the BPFO are ambiguous, and no signature indicates the existence of a bearing fault. Here, the BPFO represents the FCF of the outer race.

Table 1. The parameters of the simulated signal.

Rotor and Shaft				Gear Meshing				Defect Impulses				Random Shocks	
f_0	A_1	α_1	T	B_1	β_1	B_2	β_2	D_1	T_d	f_r	α_r	f_r	α_r
14	0.1	$\pi/2$	30	0.2	$\pi/2$	0.08	$\pi/2$	0.3	1/50	2400	200	4000	700

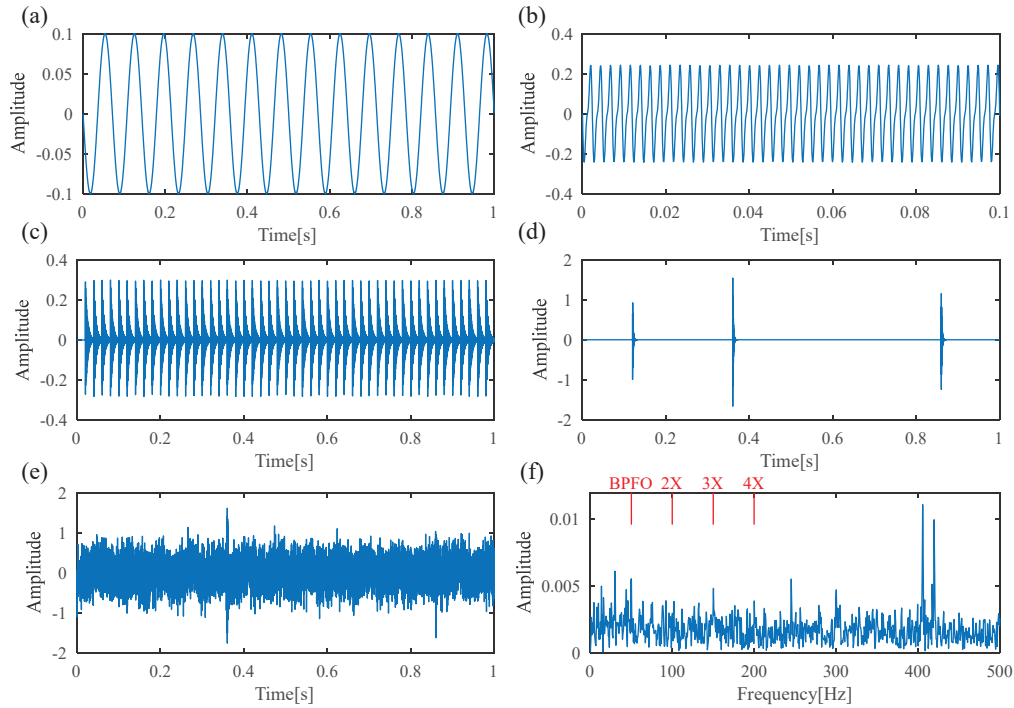


Figure 4. Simulated signal. (a) rotor vibration; (b) gear meshing; (c) defect impulses; (d) random shocks; (e) noise-added mixture signal; (f) envelope spectrum.

Next, the traditional DS-SVD [25] and the proposed FDMK-SVD are applied to analyze the simulated vibration signal. In traditional DS-SVD, the threshold k is determined by the difference spectrum (DS), and the fault information is considered to be contained in the first k SCs. Figures 5 and 6 show the results obtained using DS-SVD. According to the difference spectrum shown in Figure 5, the maximum peak is observed at $k = 2$, which means that only the first two singular components are selected to reconstruct the denoized signal. Figure 6a presents the denoized signal, and no obvious defect impulses are observed. The envelope spectrum of the reconstructed signal is displayed in Figure 6b, and it also fails to provide useful information for fault diagnosis. To further investigate the reason for this phenomenon, the decomposition results obtained using SVD are shown in Figure 7. As shown in Figure 7, the first two SCs selected for reconstructing the denoized signal based on DS-SVD are composed of the harmonic components originating from gear meshing, and the defect impulses are ignored in the DS-SVD. The above results indicate that the selection scheme of traditional DS-SVD is inappropriate when the diagnostic information is very weak and disturbed by other large-energy components.

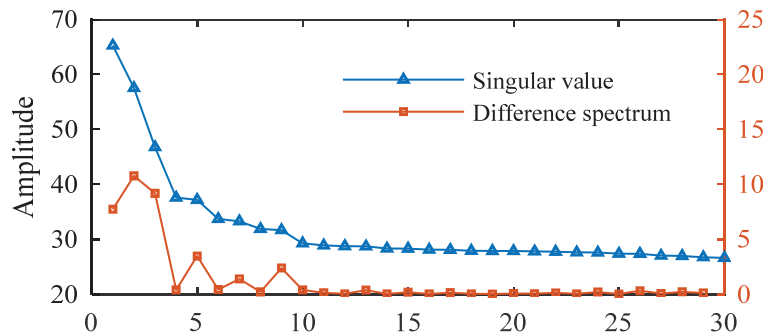


Figure 5. Singular values and difference spectra.

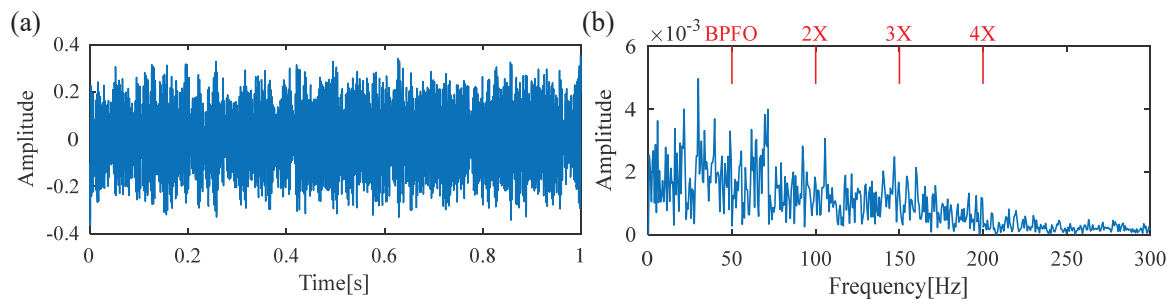


Figure 6. The results obtained by DS-SVD for simulated signal. (a) denoized signal; (b) envelope spectrum.

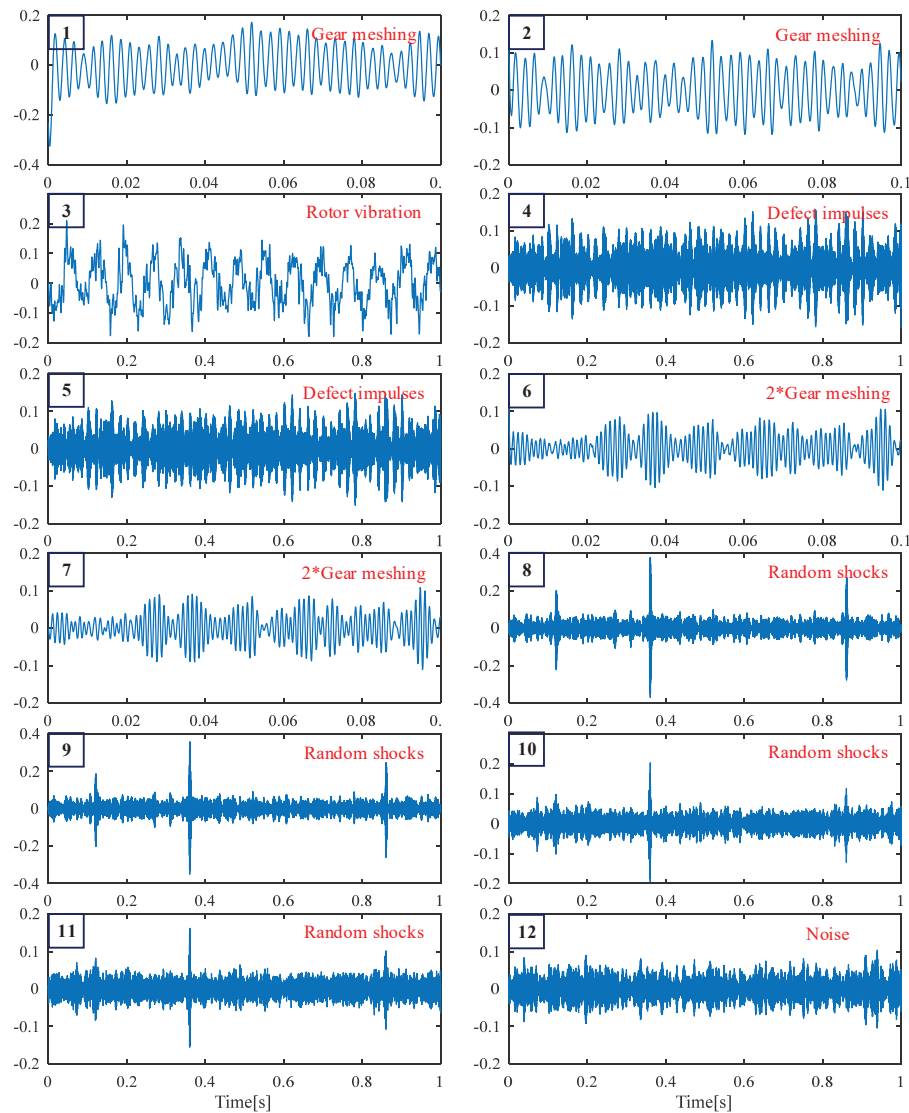


Figure 7. The first 12 singular components obtained using SVD.

For comparison, the same simulated signal is processed using the proposed FDMK-SVD. The envelope autocorrelation function of the selected most sensitive SC and raw signal is displayed in Figure 8. As shown in Figure 8a, the local maximum is located at $\tau_{\max} = 0.0202$ s, and its corresponding frequency is 49.505 Hz, which has a slight deviation from the real value. On the contrary, the fault period is masked by the harmonic period ($\tau_{\max} = 0.0024$ s) corresponding to gear meshing when the raw signal is used to estimate the period

or FCF. The reason for this phenomenon is that the energy of the defect impulses is less than that of the harmonic components. After obtaining the estimated fault characteristic frequency, the target vector $\vec{t}$ is constructed, and the FDMK value of SC is calculated and is displayed in Figure 9. As shown in Figure 9, although the first three SCs have large energies, their FDMKs are very small since they are not relevant to the bearing fault. On the contrary, the FDMK of the 4th and 5th components are significantly larger than those of the others. By checking the decomposition results obtained by SVD shown in Figure 7, the 4th and 5th SCs correspond to the defect impulses, which is significant information for identifying the health conditions of bearings. The time waveforms of the denoised signal and the corresponding envelope spectrum are displayed in Figure 10. According to Figure 10, the defect impulses are successfully extracted in the time domain, and BPFO and its harmonics are clearly recognized in the envelope spectrum. There is no doubt that the fault is successfully detected by the proposed FDMK-SVD.

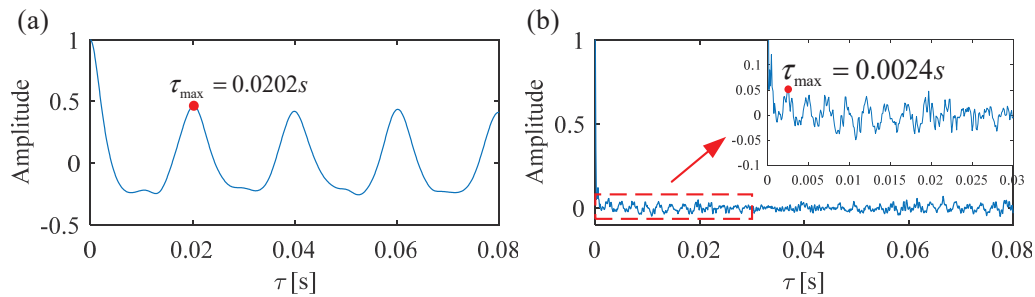


Figure 8. Envelope autocorrelation function of (a) selected singular component and (b) raw signal.

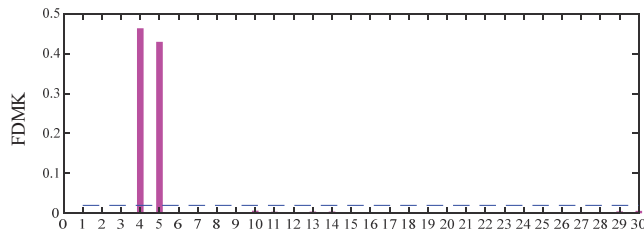


Figure 9. FDMK of each singular component for simulated signal.

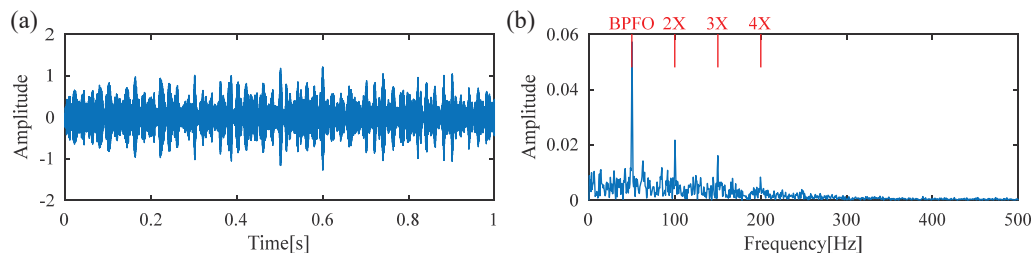


Figure 10. The results obtained by FDMK-SVD for simulated signal. (a) denoised signal; (b) envelope spectrum.

To evaluate the computational efficiency of the proposed method, the execution time is analyzed on a laptop equipped with a 12th Gen Intel® Core™ i7-12700H processor (2.7 GHz) and 16 GB DDR4 RAM, running MATLAB R2022a. The execution time for analyzing the simulated signal was 1.563 s. This relatively short processing time highlights the proposed method's suitability for real-time applications, including online monitoring. Therefore, the proposed method not only provides accurate fault diagnosis but also ensures computational efficiency, making it ideal for condition monitoring of bearings.

5. Experimental Verification

In this section, an experimental test is conducted on a locomotive bearing test rig to verify the validity of the proposed FDMK-SVD. The test rig consists of a driving wheel, a hydraulic motor, and a locomotive wheel, as shown in Figure 11. Notably, in this setup, the outer race of the bearing is driven by the hydraulic motor, while the inner race remains fixed. As a result, when a fault occurs in the outer race, amplitude modulation occurs, and sidebands are expected to appear in the envelope spectrum. In contrast, no amplitude modulation is observed for faults in the inner race. The vibration signal is collected by a three-axis PCB accelerometer with a sensitivity of 100 mV/g, which is placed at the end of the shaft. The sampling frequency is set to 76,800 Hz, and the data length is 1 s. Moreover, the speed information of the driving wheel is also acquired by a tachometer. In the experiment, the bearing under test is a tapered roller bearing, model 197,726. The geometric parameters of the locomotive bearings are listed in Table 2. To simulate real-world scenarios, the test rig was subjected to various noise sources, including mechanical vibrations from the surrounding machinery and environmental noise.

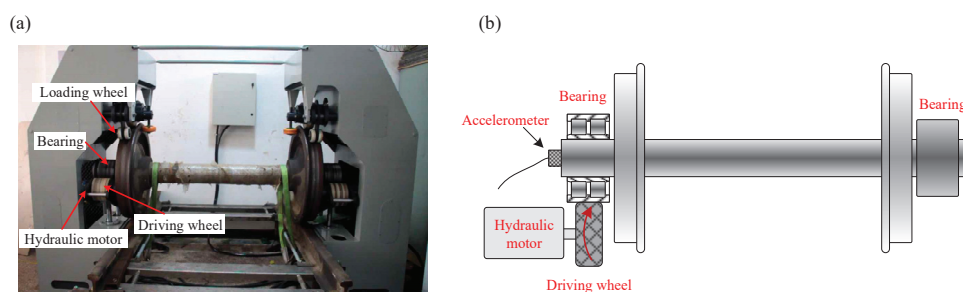


Figure 11. (a) Experiment setup of the test rig; (b) schematic view of the locomotive bearing test rig.

Table 2. Geometric parameters of the locomotive bearing.

Number of Rollers	Pitch Diameter (mm)	Roller Diameter (mm)	Contact Angle (Degree)
20	180	23.775	9

5.1. Experiment 1: The Bearing with Inner-Race-Fault

In this case, in order to further validate the validity of the proposed FDMK-SVD, the vibration signal from a bearing with an inner-race fault is processed. Table 3 lists the FCFs of the tested inner-race-fault bearing, where f_r specifies the rotating frequency of the outer race, and BPFO, BPFI, and BSF represent the FCF of the outer race, inner race, and rolling element, respectively.

Table 3. Bearing characteristic frequencies (Hz) in Experiment 1.

f_r	BPFO	BPFI	BSF
4.6183	40.1584	52.2083	17.1851

Figure 11 displays the raw vibration signal and its envelope spectrum. Owing to the heavy noise, the periodic defect impulses cannot be obviously observed from the time waveform, as shown in Figure 12a. Meanwhile, no significant BPFI or its harmonics are observed in the envelope spectrum, as shown in Figure 12b. To extract the weak fault feature from the raw vibration signal, DS-SVD is applied, and the corresponding denoised signal is illustrated in Figure 13. However, useful information related to the inner-race fault

cannot be observed from the waveform of the denoized signal or its envelope spectrum. Moreover, it can be seen that the rotating frequency f_r is dominant in the envelope spectrum, which illustrates that the DS-SVD method, which relies on an energy-oriented selection scheme, fails to effectively select the sensitive SCs containing diagnostic information when harmonic components dominate the signal.

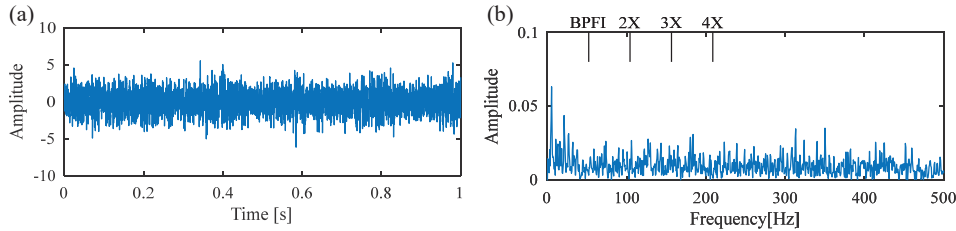


Figure 12. (a) Raw signal of the inner race fault bearing; (b) envelope spectrum.

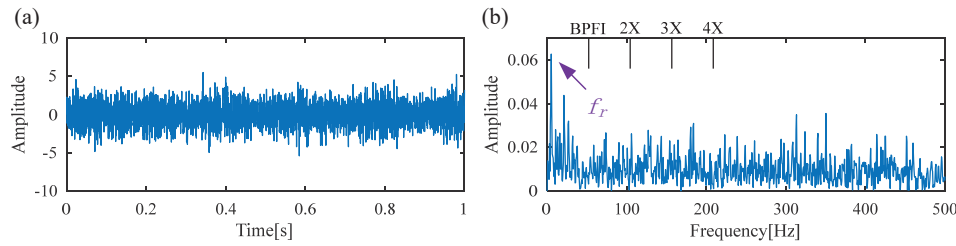


Figure 13. The results obtained by DS-SVD for the signal of the inner race fault bearing. (a) denoized signal; (b) envelope spectrum.

Spectral kurtosis (SK), a widely used and state-of-the-art method, is also adopted to extract the fault feature from the raw signal. Figure 14a presents the Kurtogram of the raw vibration signal, from which the resonance band with maximum spectral kurtosis (center frequency: 21,600 Hz; Bandwidth: 600 Hz) is easily determined. A band-pass filter with a center frequency of 21,600 Hz and a bandwidth of 600 Hz is designed to filter the vibration signal, and the filtered signal is displayed in Figure 14b. However, only several random impulses can be observed, and the periodic fault impulses are not detected. Similarly, the envelope spectrum in Figure 14c shows no obvious fault information. This failure results from the assumption that fault impulses occur in the frequency band with the largest kurtosis value. However, kurtosis is only a statistical indicator of signal sparsity, and the periodicity of the defect impulses is not taken into consideration. Therefore, it is susceptible to interference from non-Gaussian signals and random noise.

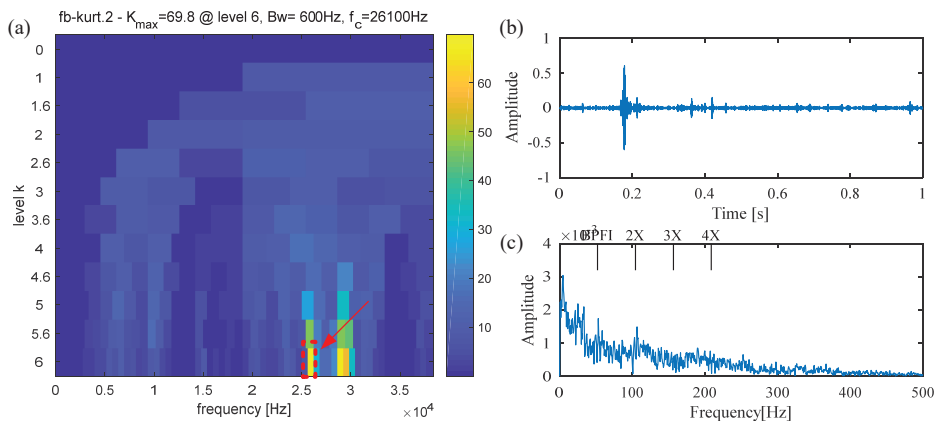


Figure 14. The results obtained by the SK method for the signal of the inner race fault bearing. (a) Kurtogram; (b) filtered signal; (c) envelope spectrum.

For comparison, the proposed FDMK-SVD is conducted on the same vibration signal. The FDMK value of each SC is calculated and is displayed in Figure 15. The denoized signal and its envelope spectrum are displayed in Figure 16. As expected, defect impulses are effectively detected, which illustrates the effectiveness of FDMK-SVD in extracting fault features and suppressing noise interference. Moreover, the BPFI and its harmonics are clearly observed, which provides evidence of the presence of an inner-race fault. To verify the results of the above analysis, the test bearing is disassembled, and the picture is illustrated in Figure 17, from which the inner-race fault of the bearing is clearly identified.

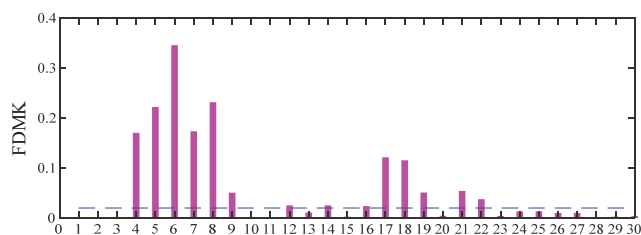


Figure 15. FDMK of each singular component for the signal of the inner race fault bearing.

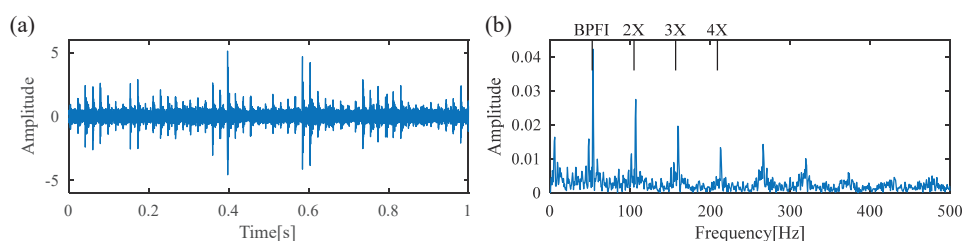


Figure 16. The results obtained by FDMK-SVD for the signal of the inner race fault bearing. (a) denoized signal; (b) envelope spectrum.



Figure 17. The fault on the inner race.

5.2. Experiment 2: The Bearing with Outer-Race Fault

In this case, the vibration signal from a bearing with an outer-race fault is used to further verify the effectiveness of FDMK-SVD in detecting bearing faults. Table 4 lists the fault characteristic frequencies for the outer-race fault.

Table 4. Bearing characteristic frequencies (Hz) in Experiment 2.

f_r	BPFO	BPFI	BSF
5.2747	45.8657	59.6281	19.6274

Figure 18 presents the time waveform of the raw vibration signal and its envelope spectrum, from which no periodic fault impulses or fault characteristic frequency can be observed. Then the DS-SVD is performed on the vibration signal, and the results are displayed in Figure 19. Similar to Experiment 1, it can be seen that the rotating frequency and

its multipliers are dominant in the envelope spectrum, while the BPFO and its harmonics are not prominent, which implies that the suitable SCs with abundant information are not identified. Figure 20a shows the Kurtogram of the signal, where the resonance band is determined to be at the center frequency of 282,000 Hz with a bandwidth of 1200 Hz. The filtered signal obtained through the corresponding band-pass filter is presented in Figure 20b, where only random shocks are extracted. Moreover, there is no obvious fault characteristic frequency or its multiplier in the envelope spectrum. Finally, the proposed FDMK-SVD is applied to the same signal, and the results are shown in Figures 21 and 22. According to Figure 21, it can be seen that the FDMK of several SCs exceeds the threshold line, which implies that these SCs may contain useful diagnostic information. To validate the above analysis, a reweighting scheme is adopted, and the reconstructed signal is shown in Figure 22. Periodic fault impulses are effectively extracted, which verifies the effectiveness of FDMK-SVD in extracting weak fault features and suppressing noise interference. Moreover, the BPFO and its harmonics are obvious in the envelope spectrum, which indicates that the outer race of the rolling bearing is faulty. Then the locomotive bearing is disassembled, from which the outer-race fault of the bearing is clearly recognized, and the picture is illustrated in Figure 23.

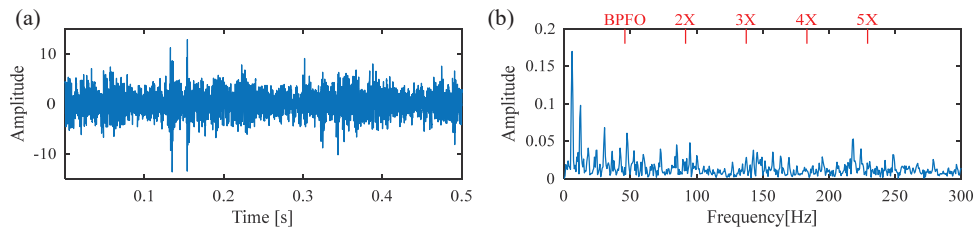


Figure 18. (a) Raw signal of the outer race fault bearing; (b) envelope spectrum.

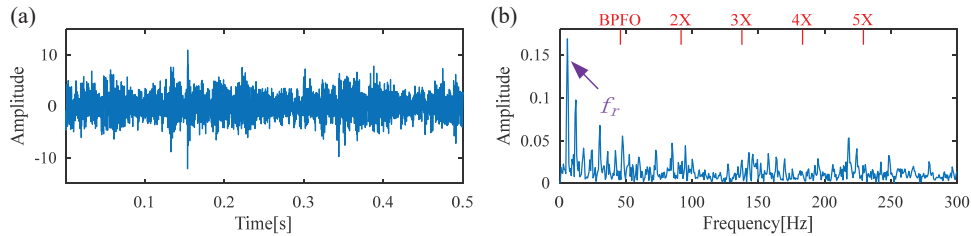


Figure 19. The results obtained by DS-SVD for the signal of the outer race fault bearing. (a) denoized signal; (b) envelope spectrum.

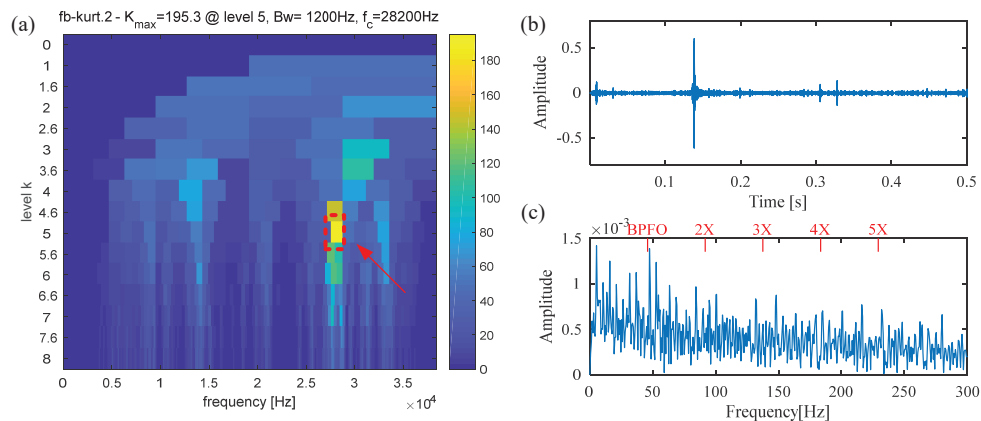


Figure 20. The results obtained by the SK method for the signal of the outer race fault bearing. (a) Kurtogram; (b) filtered signal; (c) envelope spectrum.

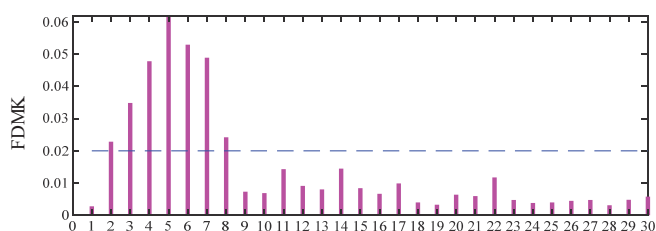


Figure 21. FDMK of each singular component for the signal of the outer race fault bearing.

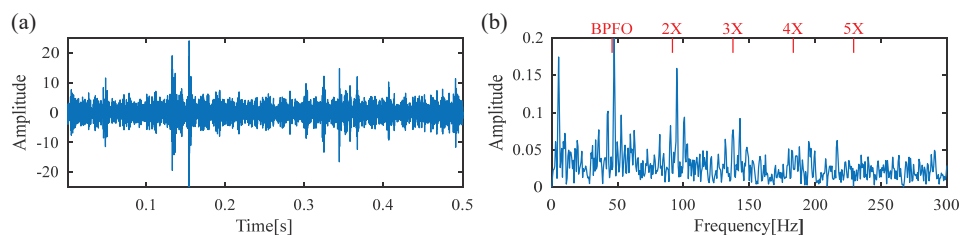


Figure 22. The results obtained by FDMK-SVD for the signal of the outer race fault bearing. (a) denoised signal; (b) envelope spectrum.

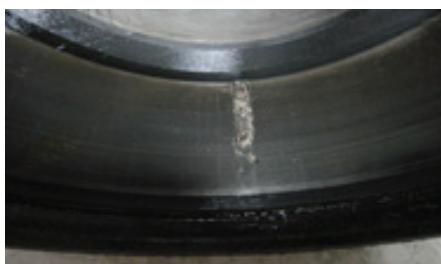


Figure 23. The fault on the outer race of the bearing.

6. Conclusions

This paper proposes a novel method, FDMK-SVD, for the diagnosis of rolling element bearings. In this work, the vibration signal is firstly decomposed via SVD, and then the sensitive SCs containing diagnostic information are selected and reweighted based on the FDMK index to reconstruct the denoised signal. Meanwhile, the estimation process of fault characteristic frequency is incorporated into the proposed method to address the issue that prior information about FCF is not predetermined in practice. Using the proposed method, weak fault features indicating the occurrence of bearings are successfully extracted. Moreover, the effectiveness and advantages of FDMK-SVD in detecting bearing faults are validated by comparing it with DS-SVD and Kurtogram. The results demonstrate that FDMK-SVD is a promising method for the condition monitoring of bearings, even in the presence of heavy noise and strong random interference.

The proposed FDMK-SVD has been validated for a constant-speed operation; however, real-world machinery often operates at varying speeds. Future work will focus on adapting SVD-based signal analysis to variable-speed conditions.

Author Contributions: Conceptualization, B.W.; methodology, B.W.; software, B.W.; writing—original draft preparation, B.W.; writing—review and editing, C.D.; and supervision, C.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China under Grant numbers. 52405124, 52205119, 52275157, the open project of the Intelligent Urban Rail Engineering Research Center of Jiangsu Province under Grant SDGC2414, and the Natural Science Foundation of Jiangsu Province under Grant BK20220497.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are unavailable due to privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Que, H.; Liu, X.; Jin, S.; Huo, Y.; Wu, C.; Ding, C.; Zhu, Z. Partial transfer learning method based on inter-class feature transfer for rolling bearing fault diagnosis. *Sensors* **2024**, *24*, 5165. [CrossRef] [PubMed]
2. Xu, Y.; Li, S.; Feng, K.; Sun, B.; Yang, X.; Kou, L.; Zhao, Z.; Huang, G.Q. Multiperspective temporal pooling convolutional neural networks for fault diagnosis of mechanical transmission systems. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 1–12. [CrossRef]
3. Shi, M.; Ding, C.; Chang, S.; Shen, C.; Huang, W.; Zhu, Z. Cross-domain class incremental broad network for continuous diagnosis of rotating machinery faults under variable operating conditions. *IEEE Trans. Ind. Inform.* **2024**, *20*, 6356–6368. [CrossRef]
4. Bhattacharjee, B.; Chakraborti, P.; Choudhuri, K. Selection of suitable lubricant for sliding contact bearing and the effect of different lubricants on bearing performance: A review and recommendations. *Tribol.-Finn. J. Tribol.* **2020**, *37*, 13–25. [CrossRef]
5. Ding, C.; Huang, W.; Shen, C.; Jiang, X.; Wang, J.; Zhu, Z. Synchroextracting frequency synchronous chirplet transform for fault diagnosis of rotating machinery under varying speed conditions. *Struct. Health Monit.* **2024**, *23*, 1403–1422. [CrossRef]
6. Ambrozkiwicz, B.; Litak, G.; Georgiadis, A.; Syta, A.; Meier, N.; Gassner, A. Effect of radial clearance on ball bearing's dynamics using a 2-DOF model. *Int. J. Simul. Model.* **2021**, *20*, 513–524. [CrossRef]
7. Dadouche, A.; Kerrouche, R. Bearing skidding detection for high speed and aero engine applications[C]//Turbo Expo: Power for Land, Sea, and Air. *Am. Soc. Mech. Eng.* **2021**, 85024, V09AT24A014.
8. Zhang, B.; Miao, Y.; Lin, J.; Li, H. Weighted envelope spectrum based on the spectral coherence for bearing diagnosis. *ISA Trans.* **2022**, *123*, 398–412. [CrossRef]
9. Mauricio, A.; Gryllias, K. Cyclostationary-based multiband envelope spectra extraction for bearing diagnostics: The combined improved envelope spectrum. *Mech. Syst. Signal Process.* **2021**, *149*, 107150. [CrossRef]
10. Zhang, Q.; Ding, J.; Zhao, W. An adaptive boundary determination method for empirical wavelet transform and its application in wheelset-bearing fault detection in high-speed trains. *Measurement* **2021**, *171*, 108746. [CrossRef]
11. Sun, Y.; Li, S.; Wang, Y.; Wang, X. Fault diagnosis of rolling bearing based on empirical mode decomposition and improved manhattan distance in symmetrized dot pattern image. *Mech. Syst. Signal Process.* **2021**, *159*, 107817. [CrossRef]
12. Zheng, J.; Su, M.; Ying, W.; Tong, J.; Pan, Z. Improved uniform phase empirical mode decomposition and its application in machinery fault diagnosis. *Measurement* **2021**, *179*, 109425. [CrossRef]
13. Miao, Y.; Li, C.; Zhang, B.; Lin, J. Application of a coarse-to-fine minimum entropy deconvolution method for rotating machines fault detection. *Mech. Syst. Signal Process.* **2023**, *198*, 110431. [CrossRef]
14. Xiong, M.; Lv, Y.; Dang, Z.; Yuan, R.; Song, H. Early fault diagnosis of rolling bearings based on parameter-adaptive multipoint optimal minimum entropy deconvolution adjusted and dynamic mode decomposition. *Meas. Sci. Technol.* **2022**, *33*, 125101. [CrossRef]
15. López, C.; Wang, D.; Naranjo, Á.; Moore, K.J. Box-cox-sparse-measures-based blind filtering: Understanding the difference between the maximum kurtosis deconvolution and the minimum entropy deconvolution. *Mech. Syst. Signal Process.* **2022**, *165*, 108376. [CrossRef]
16. Zhao, X.; Ye, B. Similarity of signal processing effect between Hankel matrix-based SVD and wavelet transform and its mechanism analysis. *Mech. Syst. Signal Process.* **2009**, *23*, 1062–1075. [CrossRef]
17. Zhao, X.; Ye, B. Singular value decomposition packet and its application to extraction of weak fault feature, Mechanical Systems and Signal Processing. *Mech. Syst. Signal Process.* **2016**, *70*, 73–86. [CrossRef]
18. Golafshan, R.; Sanliturk, K.Y. SVD and Hankel matrix based de-noising approach for ball bearing fault detection and its assessment using artificial faults. *Mech. Syst. Signal Process.* **2016**, *70*, 36–50. [CrossRef]
19. Cong, F.; Chen, J.; Dong, G.; Zhao, F. Short-time matrix series based singular value decomposition for rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2013**, *34*, 218–230. [CrossRef]
20. Duan, R.; Liao, Y. Impulsive feature extraction with improved singular spectrum decomposition and sparsity-closing morphological analysis. *Mech. Syst. Signal Process.* **2022**, *180*, 109436. [CrossRef]
21. Jiang, H.; Jiang, Y.; Xiang, J. Method using singular value decomposition and whale optimization algorithm to quantitatively detect multiple damages in turbine blades. *Struct. Health Monit.* **2024**, *23*, 1025–1036. [CrossRef]
22. Jiang, H.; Chen, J.; Dong, G.; Liu, T.; Chen, G. Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* **2015**, *52*, 338–359. [CrossRef]

23. Kang, J.; Kim, M. Singular value decomposition based feature extraction approaches for classifying faults of induction motors. *Mech. Syst. Signal Process.* **2013**, *41*, 348–356. [CrossRef]
24. Muruganatham, B.; Sanjith, M.A.; Krishnakumar, B.; Murty, S.A.V.S. Roller element bearing fault diagnosis using singular spectrum analysis. *Mech. Syst. Signal Process.* **2013**, *35*, 150–166. [CrossRef]
25. Zhao, X.; Ye, B. Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock. *Mech. Syst. Signal Process.* **2011**, *25*, 1617–1631. [CrossRef]
26. Qiao, Z.; Pan, Z. SVD principle analysis and fault diagnosis for bearings based on the correlation coefficient. *Mech. Syst. Signal Process.* **2015**, *26*, 085014. [CrossRef]
27. Zhao, M.; Jia, X. A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery. *Mech. Syst. Signal Process.* **2017**, *94*, 129–147. [CrossRef]
28. Zhao, X.; Ye, B.; Chen, T. Difference spectrum theory of singular value and its application to the fault diagnosis of headstock of lathe. *J. Mech. Eng.* **2010**, *46*, 100–108. [CrossRef]
29. Li, Q.; Weaver, B.; Chu, F.; Wang, W. A fast damping ratio identification method for rotating machinery based on the singular value of a directional power spectra density function matrix. *J. Sound Vib.* **2020**, *488*, 115630. [CrossRef]
30. Su, W.; Wang, F.; Zhu, H.; Zhang, Z.; Guo, Z. Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement. *Mech. Syst. Signal Process.* **2010**, *24*, 1458–1472. [CrossRef]
31. Kim, T.; Lee, G.; Youn, B.D. PHM experimental design for effective state separation using Jensen–Shannon divergence. *Reliab. Eng. Syst. Saf.* **2019**, *190*, 106503. [CrossRef]
32. Meng, Z.; Cui, Z.; Liu, J.; Li, J.; Fan, F. Maximum cyclic Gini index deconvolution for rolling bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [CrossRef]
33. He, W.; Jiang, Z.N.; Feng, K. Bearing fault detection based on optimal wavelet filter and sparse code shrinkage. *Measurement* **2009**, *42*, 1092–1102. [CrossRef]
34. Wang, D. Some further thoughts about spectral kurtosis, spectral L2/L1 norm, spectral smoothness index and spectral Gini index for characterizing repetitive transients. *Mech. Syst. Signal Process.* **2018**, *108*, 360–368. [CrossRef]
35. Zhang, C.; Liu, Y.; Wan, F.; Chen, B.; Liu, J.; Hu, B. Adaptive filtering enhanced windowed correlated kurtosis for multiple faults diagnosis of locomotive bearings. *ISA Trans.* **2020**, *101*, 421–429. [CrossRef]
36. Xu, X.; Zhao, M.; Lin, J.; Lei, Y. Envelope harmonic-to-noise ratio for periodic impulses detection and its application to bearing diagnosis. *Measurement* **2016**, *91*, 385–397. [CrossRef]
37. McDonald, G.L.; Zhao, Q. Multipoint optimal minimum entropy deconvolution and convolution fix: Application to vibration fault detection. *Mech. Syst. Signal Process.* **2017**, *82*, 461–477. [CrossRef]
38. Miao, Y.; Zhao, M.; Lin, J.; Lei, Y. Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2017**, *92*, 173–195. [CrossRef]
39. Xu, X.; Zhao, M.; Lin, J.; Lei, Y. Periodicity-based kurtogram for random impulse resistance. *Meas. Sci. Technol.* **2015**, *26*, 085011. [CrossRef]
40. Loza, C.A.; Principe, J.C. Transient model of EEG using Gini index-based matching pursuit. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 724–728.
41. Zonoobi, D.; Kassim, A.A.; Venkatesh, Y.V. Gini index as sparsity measure for signal reconstruction from compressive samples. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 927–932. [CrossRef]
42. Miao, Y.; Zhao, M.; Lin, J. Periodicity-impulsiveness spectrum based on singular value negentropy and its application for identification of optimal frequency band. *IEEE Trans. Ind. Electron.* **2018**, *66*, 3127–3138. [CrossRef]
43. Ding, C.; Zhao, M.; Lin, J. Sparse feature extraction based on periodical convolutional sparse representation for fault detection of rotating machinery. *Meas. Sci. Technol.* **2020**, *32*, 015008. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Static Calibration of a New Three-Axis Fiber Bragg Grating-Based Optical Accelerometer

Abraham Perez-Alonzo ^{1,*}, Luis Alvarez-Icaza ¹ and Gabriel E. Sandoval-Romero ²

¹ Instituto de Ingeniería, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; alvar@pumas.iingen.unam.mx

² Instituto de Ciencias Aplicadas y Tecnología, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; eduardo.sandoval@icat.unam.mx

* Correspondence: aperezal@iingen.unam.mx

Abstract: Optical sensors are a promising technology in structural and health monitoring due to their high sensitivity and immunity to electromagnetic interference. Because of their high sensitivity, they can register the responses of buildings to a wide range of motions, including those induced by ambient noise, or detect small structural changes caused by aging or environmental factors. In previous work, an FBG-based accelerometer was introduced that is suitable for use as an autonomous unit since it does not make use of any interrogator equipment. In this paper, we present the results of the characterization of this device, which yielded the best precision and accuracy. The results show the following: (i) improvements in the orthogonality of the sensor axes, which impact their cross-axis sensitivity; (ii) reductions in the electronic noise, which increase the signal-to-noise ratio. The results of our static characterization show that, in the worst case, we can obtain a correlation coefficient R^2 of 0.9999 when comparing the output voltage with the input acceleration for the X- and Y-axes of the sensor. We developed an analytical, non-iterative, 12-parameter matrix calibration approach based on the least-squares method, which allows compensation for different gains in its axes, offset, and cross-axis. To improve the accuracy of our sensor, we propose a table with correction terms that can be subtracted from the estimated acceleration. The mean error of each estimated acceleration component of the sensor is zero, with a maximum standard deviation of 0.018 m/s^2 . The maximum RMSE for all tested positions is $6.7 \times 10^{-3} \text{ m/s}^2$.

Keywords: sensor characterization; static calibration; FBG-based accelerometer; overlapping interrogation method

1. Introduction

In this work, we focus on the full characterization of the accelerometer developed in [1] for seismic applications, which is a fiber Bragg grating (FBG)-based triaxial accelerometer that uses the overlapping interrogation method for the demodulation of the Bragg wavelength of the gratings. The common approach is to use two FBGs for each sensing axis; however, we require just one FBG for each sensor axis and one reference FBG. Thus, this device only needs four FBGs in total.

The relevance of accelerometers in the field of structural health monitoring lies in the fact that they are part of a system that measures different vibration parameters, such as strain, where FBGs can be directly embedded in buildings, as studied in [2]. Although acceleration can be calculated from displacement or velocity measurements and vice versa, the derivation or integration process causes large errors; thus, the direct measurement of

the desired magnitude is preferred. Due to their good response, many researchers are working on the development of triaxial FBG-based accelerometers; for example, in [3], the authors developed a triaxial device with a measurement range between 0.1 Hz and 20 Hz and a high sensitivity of 606.2 pm/g; in [4], a triaxial sensor with an operating frequency between 0.1 Hz and 50 Hz was developed, achieving a sensitivity between 395 pm/g and 590 pm/g on its axes, and with a minimum detectable acceleration of 0.01 g; in both [3,4], the authors presented applications to seismic detection, similar to [5–7] and many other authors. These devices, and almost all FBG-based accelerometers, use an FBG demodulator (FBG interrogator) to measure the changes in the FBG. This FBG interrogator turns the sensor into a more complex and expensive device. We found that there is a lack of research into FBG sensors that do not use interrogators.

For the work presented here, we developed a new triaxial sensor assembled with new fiber Bragg gratings. Manufacturing helps to improve the orthogonality in its axes. This new sensor is based on the same operation principle, in which we do not use an interrogator to demodulate the FBG changes and use only one FBG for each axis and one additional reference FBG for all the axes. This approach was already employed, with satisfactory results, in the biaxial accelerometer developed in [8].

The characterization we perform in this work is of vital importance because it allows us to prove that our new device is suitable for the proposed application, and ensures that its measurements are reliable. Calibration methods are classified as static and dynamic characterizations. We selected a static characterization, which is performed by leaving the accelerometer fixed in different positions for a short period of time and recording the sensor output for each position; this procedure is the fundamental first step in the characterization of an accelerometer, since it allows us to determine the non-orthogonalities, the sensitivity in each axis, and the effects of temperature on the measurements.

Static calibration can be performed by iterative or analytical methods. Iterative methods have the advantage of not requiring one know the orientation of the accelerometer, as is the case with Newton's method [9], the Levenberg–Marquardt algorithm, or the Thin-Shell algorithm [10], which try to minimize the calibration error in each iteration until it falls to a value equal to or smaller than a previously established one. In analytical methods, there is no need for iterations; however, the tilt angles of the accelerometer for each position must be known.

The analytical method was chosen in this work, which allows the linearity of the sensor to be determined in each of its axes by considering that its measurement range is less than ± 0.5 g (± 4.89 m/s² in Mexico City), which can be provided by a tilt table [11]. The principle of this static calibration method is that the sum of three acceleration components in the accelerometer should be equal to the gravitational field vector [12].

On the other hand, since Bragg gratings are sensitive to both temperature and strain [13], we need to decouple the strain from temperature. In theory, the overlapping interrogation method of FBGs allows us to intrinsically compensate for temperature because the reference FBG does not undergo any change due to vibrations, only due to temperature changes; then, when each of the spectra of the sensing FBGs overlap with the spectrum of the reference FBG, the changes in the FBGs due to temperature are compensated. In practice, this compensation is not complete; in our prototype it is between 50% and 60%.

This incomplete compensation behavior has also been reported in other works, such as [14], where two FBGs that should have had the same sensitivity to temperature showed that they had different sensitivities in experiments: 23 pm/°C in one and 28 pm/°C in the other. Another example is [15], where the temperature coefficients of the two FBGs were

16.9 pm/°C and 18.2 pm/°C; thus, although the temperature dependence was reduced to 1.3 pm/°C, there was still some temperature dependence.

This imperfect temperature compensation approach has limited effects on the measurement of vibrations due to seismic motion, since the signal variation with respect to temperature is very small compared to the signal variations due to vibration. The longest period of the signal that our sensor is intended to measure is ten seconds; thus, the temperature variations are negligible, since they normally have a period of 24 h when the sensor is in a static position. Therefore, we can decouple the vibration signals from the temperature signals.

2. Materials and Methods

The static characterization was performed on the hand-made tilting platform shown in Figure 1, whose parameters can be seen in Table 1. This platform allows us to change the pitch and roll angle by rotating the corresponding bolt nuts to slide one end of the tilting table. The hinges are tightened in such a way that there is no gap between their parts, allowing only rotational movement.

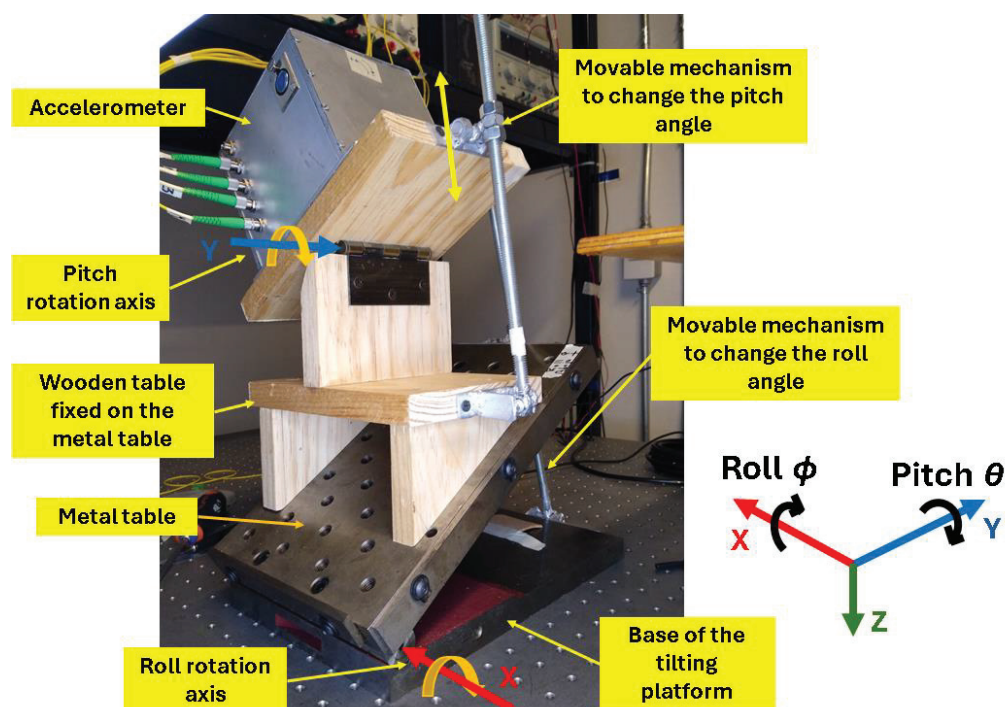


Figure 1. Tilting platform in pitch and roll rotations. The optical sensor is placed on top of the platform. The coordinate system is shown on the right of the figure.

Table 1. Characteristics of the tilting platform.

Parameter	Range	Resolution
Pitch rotation, θ (Rotation about the Y-axis)	± 60 deg	0.1 deg
Roll rotation, ϕ (Rotation about the X-axis)	± 30 deg	0.1 deg

The magnitude of the gravitational acceleration varies with latitude, altitude, gravity anomaly, and tidal effect. We use the International Gravity Formula derived in [9], which depends on latitude and altitude, as shown in Equation (1).

$$g(\gamma, h) = 9.780327[1 + 5.3024 \times 10^{-3} \sin^2(\gamma) - 5.08 \times 10^{-6} \sin^2(\gamma)] - 3.086 \times 10^{-6} h \text{ [m/s}^2\text{]}, \quad (1)$$

where γ is the latitude, and h is the height (in meters) with respect to sea level. Substituting $\gamma = +19.32078^\circ$, which corresponds to Mexico City's latitude, and $h = 2293$ m according to [16], results in a local acceleration value of 9.7789 m/s^2 ; this value will be used in the rest of the article as the magnitude of gravity.

As our accelerometer has a high sensitivity and a low range ($\pm 4.89 \text{ m/s}^2$), it is enough to rotate its horizontal axes between a range of ± 30 degrees. In the case of the vertical axis, to change the Z component of the gravity vector by 4.89 m/s^2 , it is necessary to tilt the platform within a ± 60 degrees range. Therefore, the characteristics of the tilting platform are sufficient to characterize our sensor. The tilt angle of the platform was measured with a digital protractor model Pro 360 950-315 from the Japanese manufacturer Mitutoyo, which has a tilt range of 360 degrees and a resolution of 0.1 degree.

When the accelerometer is at rest in the horizontal position, the acceleration on the vertical axis of the sensor is 9.78 m/s^2 downward, but since we want to measure the acceleration relative to the gravitational field, we take the acceleration as 0 m/s^2 in all three axes at rest position. Another thing to note in the vertical axis is that the changes in acceleration are symmetric regardless of whether the platform is tilted in the positive or negative sense. Thus, while the horizontal axes are tested over a range of $\pm 4.89 \text{ m/s}^2$, the vertical axis is only tested from 0 to -4.89 m/s^2 . The accelerometer cannot be turned over (rotated 180 degrees about the X- or Y-axis) because it is outside the measurement range of the sensor.

Static Calibration Procedure

The most general linear calibration of the accelerometer has a total of 12 parameters [17], where a recalibration is performed over the previous factory calibration. We use this same procedure to perform the first calibration of our sensor. This is a non-iterative method, which is based on the linear least-squares optimization.

At first, the general linear model that relates the dependent variable y to the N inputs x_0 to x_{N-1} , through N coefficients β_0 to β_{N-1} , is

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{N-1} x_{N-1}. \quad (2)$$

The N model parameters for each sensor axis β_0 to β_{N-1} are determined from the M measurements from $i = 0$ to $i = M - 1$, where $M \geq N$. Equation (2), written in vector form for all M measurements, is

$$\begin{pmatrix} y[0] \\ y[1] \\ \vdots \\ y[M-1] \end{pmatrix} = \begin{pmatrix} x_0[0] & x_1[0] & \cdots & x_{N-1}[0] \\ x_0[1] & x_1[1] & \cdots & x_{N-1}[1] \\ \vdots & \vdots & \ddots & \vdots \\ x_0[M-1] & x_1[M-1] & \cdots & x_{N-1}[M-1] \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{N-1} \end{pmatrix}. \quad (3)$$

The vector Y of the dependent variables for each sensor axis is defined as

$$Y = \begin{pmatrix} y[0] \\ y[1] \\ \vdots \\ y[M-1] \end{pmatrix}. \quad (4)$$

The matrix X of the measurements of the independent variables is defined as

$$X = \begin{pmatrix} x_0[0] & x_1[0] & \cdots & x_{N-1}[0] \\ x_0[1] & x_1[1] & \cdots & x_{N-1}[1] \\ \vdots & \vdots & \ddots & \vdots \\ x_0[M-1] & x_1[M-1] & \cdots & x_{N-1}[M-1] \end{pmatrix}. \quad (5)$$

The solution vector β of the N model parameters for each sensor axis is defined as

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{N-1} \end{pmatrix}. \quad (6)$$

Using Equations (4)–(6), Equation (3) can be rewritten as

$$Y = X\beta. \quad (7)$$

In the linear least-squares optimization, a residuals vector is defined as

$$r = Y - X\beta. \quad (8)$$

The optimal least-squares fit for β is the one that minimizes the performance function P defined as the squared modulus of the residuals vector [17]. Finally, the commonly known normal equations for least squares optimization result in:

$$\beta = (X^T X)^{-1} X^T Y. \quad (9)$$

The matrix X of measurements is reused in the calculation of the parameter of the three axes; the vectors Y and β are different for each axis. The calibrated accelerometer output G is defined by

$$G = \begin{pmatrix} G_x \\ G_y \\ G_z \end{pmatrix} = W V_0 + V = \begin{pmatrix} W_{xx} & W_{xy} & W_{xz} \\ W_{yx} & W_{yy} & W_{yz} \\ W_{zx} & W_{zy} & W_{zz} \end{pmatrix} \begin{pmatrix} V_{0x} \\ V_{0y} \\ V_{0z} \end{pmatrix} + \begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix}, \quad (10)$$

where W is the gain matrix, which has nine independent elements. It includes all possible cross-axis interactions and any internal rotation of the sensor structure with respect to its case. V_0 are the output voltages of the sensor, and V is the offset vector at 0 m/s². Using Equation (9), the optimum least squares solution for β_x of the X-axis is

$$\beta_x = \begin{pmatrix} W_{xx} \\ W_{xy} \\ W_{xz} \\ V_x \end{pmatrix} = (X^T X)^{-1} X^T Y_x = (X^T X)^{-1} X^T \begin{pmatrix} -\sin \theta[0] \\ -\sin \theta[1] \\ \vdots \\ -\sin \theta[M-1] \end{pmatrix}. \quad (11)$$

Similarly, the optimum least-squares solution for the remaining eight calibration parameters for the Y- and Z-axes is

$$\beta_y = \begin{pmatrix} W_{yx} \\ W_{yy} \\ W_{yz} \\ V_y \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_y = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \begin{pmatrix} \cos \theta[0] \sin \phi[0] \\ \cos \theta[1] \sin \phi[1] \\ \vdots \\ \cos \theta[M-1] \sin \phi[M-1] \end{pmatrix}, \quad (12)$$

$$\beta_z = \begin{pmatrix} W_{zx} \\ W_{zy} \\ W_{zz} \\ V_z \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_z = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \begin{pmatrix} \cos \theta[0] \cos \phi[0] \\ \cos \theta[1] \cos \phi[1] \\ \vdots \\ \cos \theta[M-1] \cos \phi[M-1] \end{pmatrix}. \quad (13)$$

The measurement matrix $\mathbf{X}$ is common to all three Equations (11)–(13). The angle θ is the pitch rotation about the Y-axis, the angle ϕ is the roll rotation about the X-axis, and M is the number of different sensor orientations used to find the calibration parameters.

3. Results

In this section, we show the dependence of the Bragg wavelength of the four FBGs used in the sensor on temperature changes, and how to compensate for them. We show that the reference FBG is not affected by vibrations, only by temperature. The response of our sensor to input acceleration, in wavelength and voltage, is shown, along with the 12-parameter model calibration and error analysis.

3.1. Bragg Wavelength Dependence of FBGs with Temperature

As mentioned in the introduction, the reference FBG acts as a temperature compensation device, which in theory must be 100%. In practice, all FBGs have different sensitivity coefficients to a greater or lesser degree. To characterize the dependence on temperature changes in all four FBGs used in our sensor, we kept the sensor stable at a fixed position and changed the temperature of the sensor. We placed an incandescent lamp over the sensor at a certain distance from it, which was periodically reduced to increase the sensor temperature. The internal temperature of the sensor, measured using a DHT22 sensor, was changed from 24 °C to 31 °C. The Bragg wavelength was recorded with a device, the Micron Optics—Optical Sensing Interrogator sm130, with a sampling rate of 100 Hz. To reduce noise, 1000 samples were averaged for each data point [18]. Although the external temperature was changed from 24 °C to about 50 °C, the internal temperature did not reach the same value, which means that the sensor case helps to reduce the variation of the output voltages due to temperature. The total time to change the temperature from 24 °C to 31 °C was 3 h and 25 min.

Figure 2a shows the Bragg wavelengths of all FBGs used in our sensor with respect to its internal temperature. To compare the responses of the gratings, they must have the same initial Bragg wavelength; therefore, an offset was applied to the Bragg wavelength of the sensing FBGs. Then, the Bragg wavelength of the sensing FBGs was adjusted by a factor using Equation (14), which changes the temperature sensitivity. The reference FBG has the lowest sensitivity to temperature; thus, the adjustment factors found for the sensing gratings were less than 1—0.51, 0.51, and 0.59 for the X-, Y-, and Z-axes, respectively. The temperature adjustment is then of the form

$$\lambda_{BS,adj,i} = \lambda_{BS,0} + (\lambda_{BS,i} - \lambda_{BS,0})f_a + O \quad (14)$$

where $\lambda_{BS_{adj,i}}$ is the i -th adjusted Bragg wavelength of the i -th measurement point, $\lambda_{BS,0}$ is the Bragg wavelength of the first measurement point, $\lambda_{BS,i}$ is the Bragg wavelength of the i -th measurement point, f_a is the adjustment factor, and O is the offset. Equation (14) is applied for each of the sensing axes.

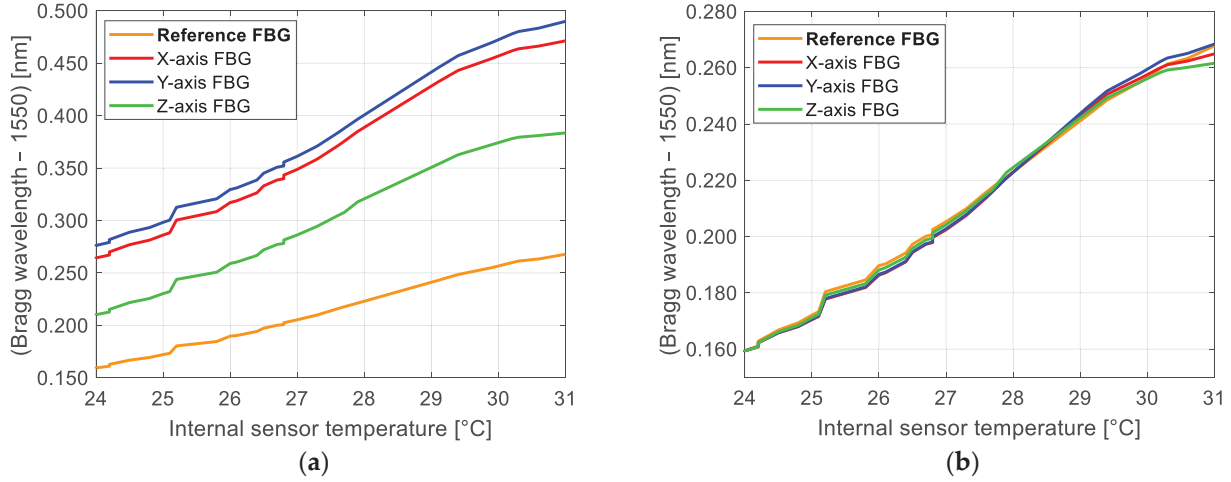


Figure 2. Dependence of the FBGs of our prototype on temperature: (a) apart from an offset, the sensitivity in all FBGs is different; (b) the Bragg wavelength of sensing FBGs is adjusted by a factor and an offset; now all FBGs show the same trend.

In Figure 3, we can see the responses of the FBGs of our sensor, which was left recording in the laboratory for 17 h. The accelerometer was left on an optical table with active pneumatic isolation; thus, it would not suffer from the ambient vibrations of the building. The changes in the Bragg wavelength of the FBGs are due to temperature, in the same way as in Figure 2. This figure was added since the accelerometer is not intended to work at a controlled ambient temperature, and the temperature surrounding it will change in a similar way as in Figure 3. The trend in this figure confirms once again that the Bragg wavelength of all FBGs shows the same tendency with time, as the ambient temperature changes. The initial Bragg wavelength of the FBGs may vary depending on the sensor calibration.

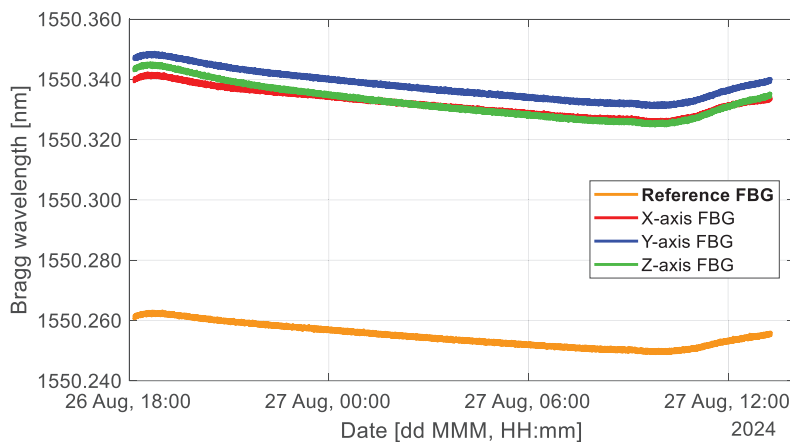


Figure 3. Changes in Bragg wavelength due to temperature over 17 h at laboratory room temperature. Measurements were made with a sampling period of 0.1 s, and a simple moving average with a window size of 5 was applied to the data.

Therefore, we can state that the reference FBG, when used in the overlapping interrogation method, helps to decrease the influence of temperature by reducing the drift of

the sensor output. By using the adjustment factors to change the sensitivity of sensing FBGs, we can almost eliminate the signal variation. In addition, the signal variation due to temperature is very small compared to the variations due to vibrations. This is because the temperature changes at a very slow rate compared to the period of vibrations.

3.2. Immunity of Reference FBG to Vibrations

The reference FBG has two purposes in our sensor: the first is to complete the demodulation scheme for the accelerometer, and the second is to compensate for temperature. Therefore, it is very important that the reference FBG is not affected by vibrations. In our prototype sensor, this fiber grating is fixed in an aluminum clamp with a set screw for pretension [1]. This grating is not connected to the inertial mass as the sensing gratings are; thus, this FBG is not affected by vibrations as demonstrated in Figure 4. In Figure 4a, we apply a directional vibration by hitting the tilting platform to which the accelerometer is attached. It can be seen that the reference FBG is not affected by these hits. In Figure 4b, we apply aleatory shocks, and we can observe that all three axes are affected, but, again, the reference FBG is not affected by these impulses.

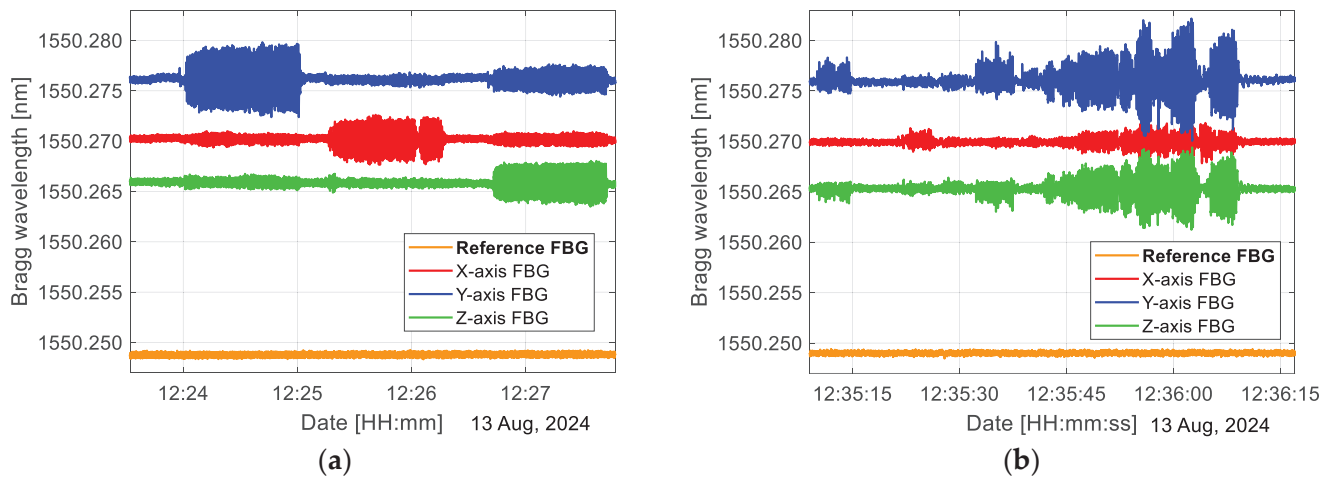


Figure 4. Vibration is applied to the tilting platform to which our sensor is attached; it can be noted that the reference FBG is not affected by such vibration: (a) vibration applied in a preferred direction; the X-, Y-, or Z-axis; (b) vibration applied in arbitrary directions.

We performed a simple moving average with a window size of 10 on data in Figure 4 that were sampled at 1 kHz.

The reference FBG is attached to the sensor structure by two points with a 25 mm clearance between the fixing points; this clearance is the same as that of the sensing grating; therefore, the immunity of the reference FBG to vibrations is proof that the possible bending of optical fibers under the influence of acceleration is decreased by the pre-strain applied to the fiber at the time of its fixing, as reported in [19]. Although the pre-strain limits the maximum measurable acceleration, since in our case the maximum acceleration to be measured is only $\pm 4.89 \text{ m/s}^2$, this limitation does not affect it. Due to the pre-strain, the fiber is never slack, but its tension is higher or lower depending on the direction of motion; therefore, the bending of the fiber is decreased, and its effect is not seen in the result of the experiment.

3.3. Static Characterization of the Accelerometer

3.3.1. Wavelength Characterization

As described in the materials and methods section, the X-axis was tilted in a ± 60 degrees range to obtain the response of the vertical axis to -0.5 g , while the Y-axis was only tilted in a ± 30 degrees range. The step in the measurements was 5 degrees.

In Figure 5, we can see a little offset between the forward and the backward sweep measurements due to temperature changes; this offset is clearer in Figure 5b. Before analyzing this data, we performed compensation based on the changes of the reference FBG using Equation (15).

$$\lambda_{BS_{adj},i} = \lambda_{BS,i} - (\lambda_{BR,i} - \lambda_{BR,0})f_{ar}, \quad (15)$$

where $\lambda_{BS_{adj},i}$ is the i -th adjusted Bragg wavelength of the i -th measurement point, $\lambda_{BS,i}$ is the Bragg wavelength of the i -th measurement point, $\lambda_{BR,i}$ is the Bragg wavelength of the i -th measurement point of reference FBG, $\lambda_{BR,0}$ is the Bragg wavelength of the initial measurement point of the reference FBG, and f_{ar} is the adjustment factor. Equation (15) applies to each of the sensing axes. The adjustment factors used in the compensation were 1.15, 0.90, and 1.00 for the X-, Y-, and Z-axes, respectively. The results are shown in Figure 6, where the response of reference FBG is shown without compensation.

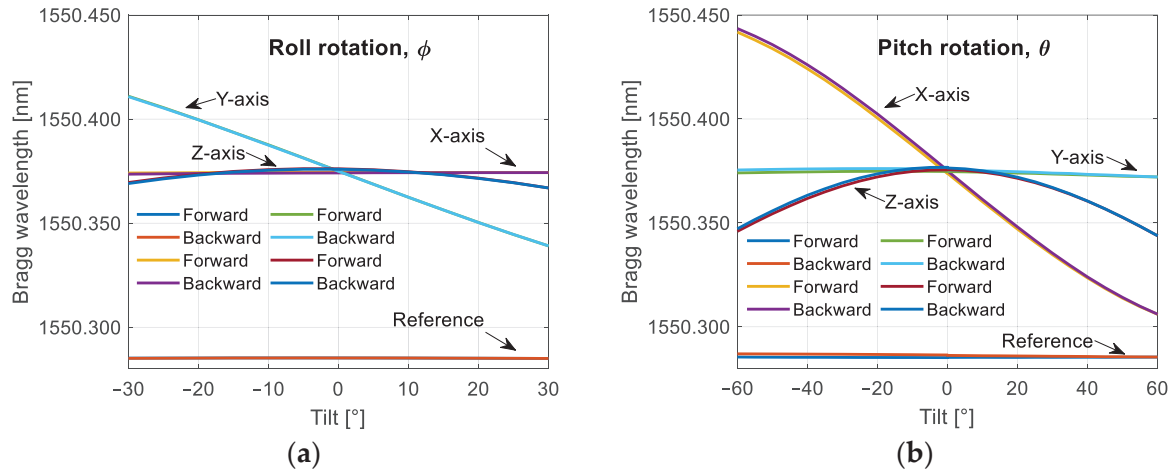


Figure 5. Response in wavelength of the FBGs used in our prototype with respect to tilt. The tilt is referenced to the horizontal plane; two sweeps were performed: one forward and one backward. (a) The sensor is rotated about its X-axis; (b) the sensor is rotated about its Y-axis.

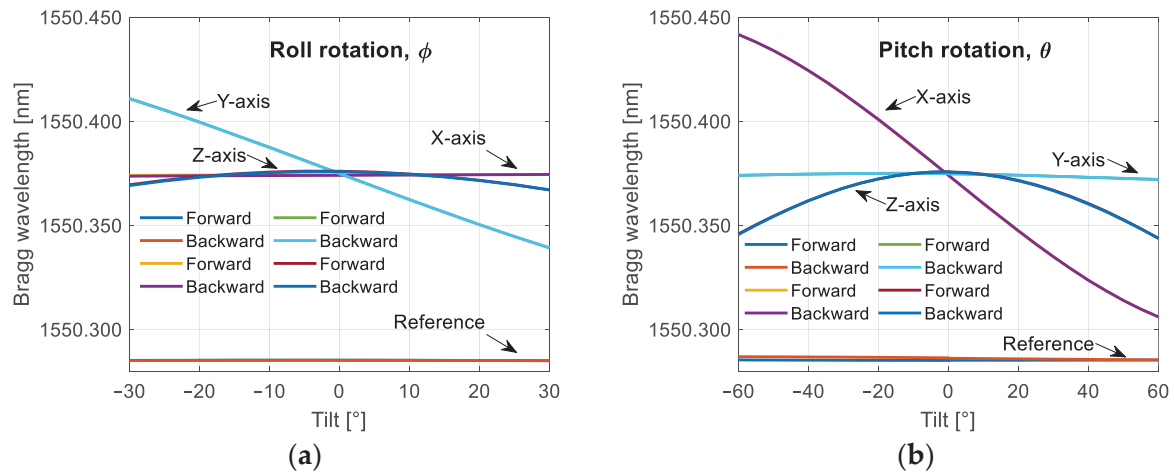


Figure 6. Response in wavelength of the FBGs used in our prototype with temperature compensation. The tilt is referenced to the horizontal plane; the temperature compensation was performed with respect to the reference FBG: (a) the sensor is rotated about its X-axis; (b) the sensor is rotated about its Y-axis.

After temperature compensation, the forward and backward sweep data are averaged for each sensor orientation and plotted with the horizontal axis in units of acceleration. The acceleration was calculated with Equation (16) by applying the sine function to the tilt angles in the X- and Y-axes and the cosine function for the vertical axis of the sensor. The results are shown in Figure 7.

$$\begin{aligned} a_x &= -\sin \theta \\ a_y &= \cos \theta \sin \phi \\ a_z &= \cos \theta \cos \phi \end{aligned} \quad (16)$$

where a_x , a_y , and a_z are the components of the gravity vector coupled to the X-, Y-, and Z-axes of the sensor, θ is the pitch rotation about the Y-axis, and ϕ is the roll rotation about the X-axis of the sensor.

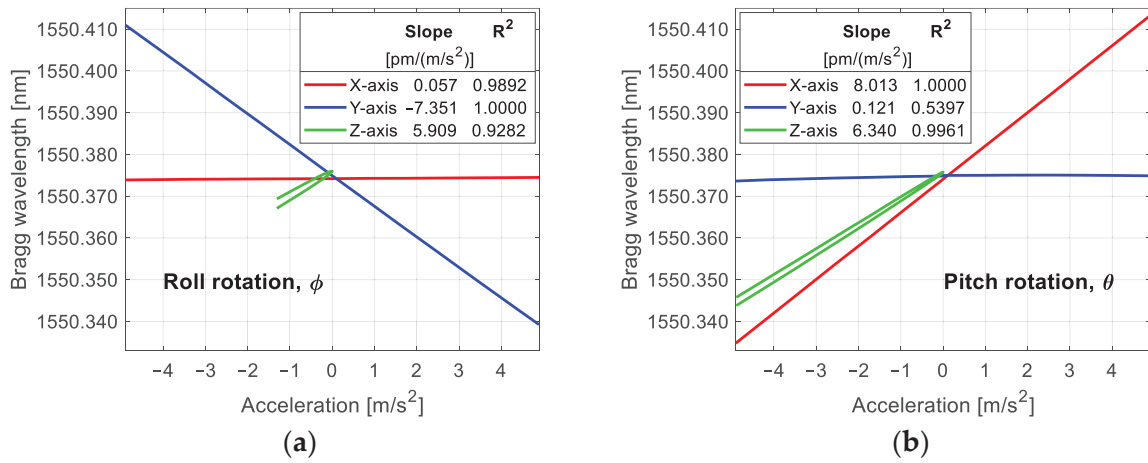


Figure 7. Response of the sensing FBG of our sensor, the horizontal axis of the plot was converted to acceleration by using the sine function of the tilt angle for the X- and Y-axes and the cosine function for the Z-axis. In the Z-axis, the magnitude of the gravity vector is subtracted from the measurements: (a) the sensor is rotated about its X-axis; (b) the sensor is rotated about its Y-axis.

For Figure 7a, where the sensor rotates about its X-axis, after a linear fit, we obtained a sensitivity of $-7.351 \text{ pm}/(\text{m}/\text{s}^2)$ for the Y-axis with an R-squared (correlation factor) of 1.0000, while we have a cross-sensitivity of $0.057 \text{ pm}/(\text{m}/\text{s}^2)$ on the X-axis. This means a cross-sensitivity 129 times lower than the corresponding axis sensitivity. In the second case, where the sensor is rotated about its Y-axis, we obtained a sensitivity of $8.013 \text{ pm}/(\text{m}/\text{s}^2)$ for the X-axis with an R-squared value of 1.0000, and $0.121 \text{ pm}/(\text{m}/\text{s}^2)$ on the Y-axis; thus, we have a cross-sensitivity 66 times lower than the corresponding axis.

For the Z-axis, where the acceleration is symmetrical with respect to both senses of rotation of the horizontal axes of the sensor, the ideal would be to obtain a line that goes from -5 to $0 \text{ m}/\text{s}^2$, then from 0 to $-5 \text{ m}/\text{s}^2$, arriving at the same place; however, in Figure 7a,b, we can see that its sensitivity is somewhat different depending on the sense of rotation of the sensor. If we analyze Figure 6, the difference in sensitivity can be attributed to a slight misalignment of the Z-axis with respect to the vertical. This is because the maximum wavelength of the Z-axis FBG is not obtained when the sensor is in a flat position, but when it is somewhat tilted. The average sensitivity of this axis, when the sensor is rotated about the X-axis, is $5.909 \text{ pm}/(\text{m}/\text{s}^2)$, with R-squared equal to 0.9282, and $6.340 \text{ pm}/(\text{m}/\text{s}^2)$ when the sensor is rotated about the Y-axis, with R-squared equal to 0.9961.

3.3.2. Voltage Characterization

The sensitivities and response curves we obtained earlier were at the wavelength scale with respect to the acceleration experienced in each axis. These values help to reveal the

response of our sensor to temperature and acceleration during characterization. However, one of the main advantages of our sensor is that we do not need interrogation equipment; instead, the reference FBG allows us to demodulate the Bragg wavelength of the sensing fiber gratings into optical power changes, which are measured with simple photodetectors. The output of the photodetectors is a current, which is proportional to the input optical power. Hence, the next step is to obtain the response of our sensor by measuring the voltage at the output of the electronic stage. The way of characterization is the same as in the wavelength characterization; the sweeps are in the same range, and the results are shown in Figure 8.

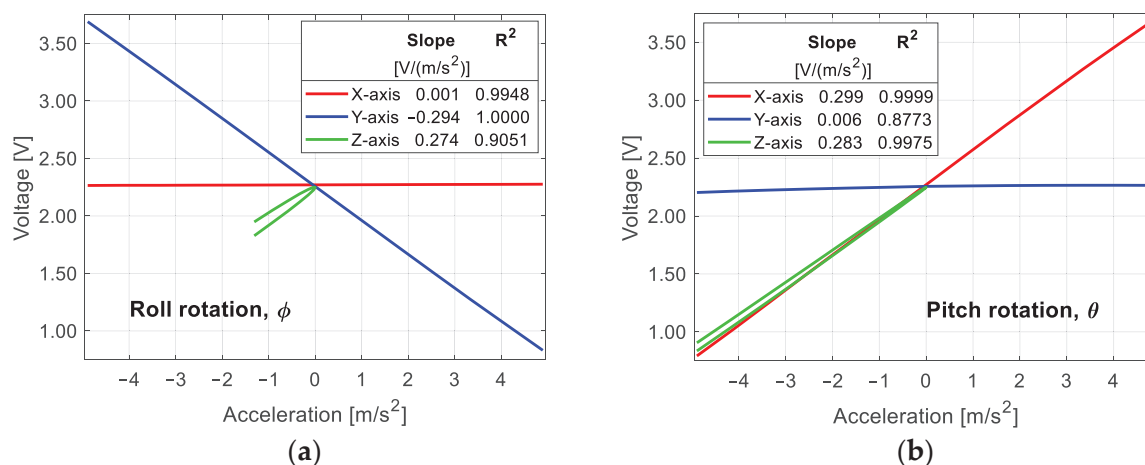


Figure 8. Sensor output in voltage at the final sensor stage with respect to the input acceleration: (a) rotation about the X-axis; (b) rotation about the Y-axis.

According to Figure 8a, the sensitivity of the Y-axis when the sensor rotates about the X-axis (roll rotation) is $-0.294 \text{ V}/(\text{m}/\text{s}^2)$ with a correlation coefficient of 1.0000, and $0.001 \text{ V}/(\text{m}/\text{s}^2)$ for the X-axis, which means a cross-sensitivity 267 times lower (0.34%) than that of the corresponding Y-axis. When the sensor rotates about the Y-axis (pitch rotation), the sensitivity for the X-axis is $0.299 \text{ V}/(\text{m}/\text{s}^2)$ with R-squared equal to 0.9999, and a cross-sensitivity of $0.006 \text{ V}/(\text{m}/\text{s}^2)$ for the Y-axis, i.e., a cross-sensitivity 48 times lower (2.08%) than that of the X-axis. For the Z-axis, the average sensitivity when the sensor rotates about the X-axis is $0.274 \text{ V}/(\text{m}/\text{s}^2)$ and $0.283 \text{ V}/(\text{m}/\text{s}^2)$ when it rotates about the Y-axis. The cross-axis interference for our sensor is comparable to that of some commercial accelerometers, such as the SF3000L calibrated in [12].

The overlapping interrogation method has a nonlinear response because the output power of the method is a convolution of the two FBG spectra, modeled as Gaussians [20]. There is a region where the response of this method is nearly linear; this region can be achieved with a suitable separation between the Bragg wavelength of the reference FBG and that of the sensing FBG. The correlation coefficient that we found in the sensor for the X- and Y-axes is very close to unity, which means that the FBGs are working in the linear region. The proper separation between the reference FBG and the sensing FBGs found experimentally in our sensor is shown in Table 2.

Table 2. Optimal separation between the Bragg wavelength of sensing FBGs and the reference FBG.

FBG ₁	FBG ₂	Bragg Wavelength Separation (pm)
Reference	X-axis FBG	102 ± 2
Reference	Y-axis FBG	98 ± 2
Reference	Z-axis FBG	95 ± 2

3.3.3. The 12-Parameters Calibration Model

Along with the voltage characterization, we performed the calibration model described in Equation (10). The sensor orientations used as input for parameter estimation are shown in Table 3, where θ is the angle of rotation about the Y-axis of the sensor from the horizontal plane, and ϕ is the angle of rotation about the X-axis of the sensor from the horizontal plane too. For the estimation of the X- and Y-axes parameters, the positions 0, 1, 2, 4, 5, and 6 were used. For the Z-axis, all eight positions were used.

Table 3. Orientations for sensor calibration.

Orientation Number	Pitch, θ [deg]	Roll, ϕ (deg)
0	0	−30
1	0	0
2	0	+30
3	−60	0
4	−30	0
5	0	0
6	+30	0
7	+60	0

The 12-parameter model includes three main gains, cross-axis gains, sensor axes misalignment, and offsets for zero input. The model parameters' values found for our sensor are shown in Equation (17).

$$\mathbf{G} = \begin{pmatrix} G_x \\ G_y \\ G_z \end{pmatrix} = \begin{pmatrix} 3.3610 & 0.0171 & -0.1107 \\ 0.0719 & -3.4242 & 0.0844 \\ -0.0608 & -0.1500 & 3.5454 \end{pmatrix} \begin{pmatrix} V_{ox} \\ V_{oy} \\ V_{oz} \end{pmatrix} + \begin{pmatrix} -7.4219 \\ 7.3733 \\ -7.5029 \end{pmatrix}. \quad (17)$$

In the estimated parameters, the vertical acceleration is considered relative to the magnitude of the gravity field; then, at the rest position of the sensor, all three axes have a zero value. G_x , G_y , and G_z are the acceleration components in m/s^2 for the X-, Y-, and Z-axes, respectively. V_{ox} , V_{oy} , and V_{oz} are the output voltages of the accelerometer electronic stage in volts.

3.3.4. Errors in the 12-Parameter Calibration Model

We verified the estimated calibration model parameters for our sensor by introducing the sensor output voltages from each of the sensor orientations obtained during the sweeps, as shown in Figure 5, and comparing the model output to each acceleration component exerted at each position. We calculate the full-scale error (FSE) as a percentage for all measurements with Equation (18).

$$\text{FSE [\%]} = \frac{\text{measured value} - \text{reference value}}{\text{full-scale acceleration}} \times 100\%. \quad (18)$$

The maximum full-scale errors are 0.58%, 0.48%, and 0.39% for the X-, Y-, and Z-axes, respectively. In absolute values, the maximum acceleration errors were 0.057 m/s^2 , 0.047 m/s^2 and 0.038 m/s^2 for the X-, Y-, and Z-axes, respectively. These errors found are comparable to the errors found after calibration in [12].

In the estimation of the calibration parameters, the 38 recorded positions were also considered to see how much the maximum error decreased, finding that there was no improvement. Therefore, we can conclude that using eight orientations to estimate the model parameters is a good compromise between the number of positions and the error obtained.

In Figure 9, we show the errors present after calibration for each sensor orientation in each axis. The largest errors are present at the extremes of the sensor range, i.e., near $\pm 4.89 \text{ m/s}^2$. The plot has the horizontal axis in units of angles because this unit is common to the horizontal and vertical axes of the sensor, but acceleration is not. The errors are not totally random, as they show a tendency to increase or decrease systematically.

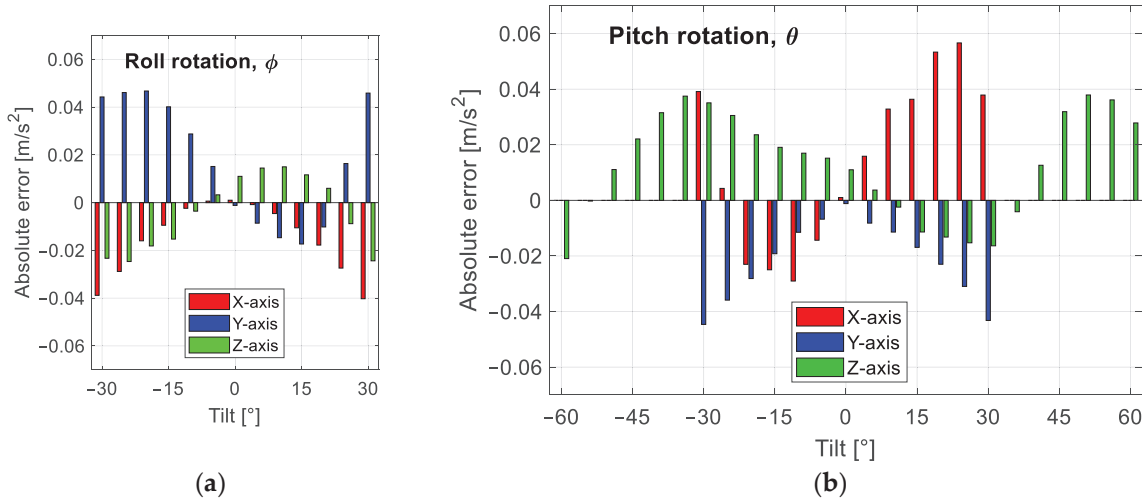


Figure 9. Absolute acceleration error comparing the theoretical acceleration due to each tilt angle with the estimated acceleration using the sensor output voltages: (a) the sensor is rotated about its X-axis; (b) the sensor is rotated about its Y-axis.

To further reduce errors, we propose to subtract one term from each estimated component of acceleration. To make sure our proposal is adequate and that the errors decrease, we performed the sweep for the 38 positions five times. We then subtracted the mean error of the five measurements at each position. In Table 4, we can see the terms to be subtracted from the estimated acceleration together with the standard deviation of the error at each orientation. The remaining errors are plotted in Figure 10, where we group the errors for each position and each axis. Now, the errors have a zero mean and show a random behavior.

Table 4. Correction terms to be subtracted from estimated accelerations and the respective standard deviation (SD) of the zero-mean error.

Ref. Angle $\theta; \phi$ [deg]	Correction Term C_{Gx}, C_{Gy}, C_{Gz} (m/s^2)	Error SD $\sigma_{Ex}, \sigma_{Ey}, \sigma_{Ez}$ (m/s^2)	Ref. Angle $\theta; \phi$ (deg)	Correction Term C_{Gx}, C_{Gy}, C_{Gz} (m/s^2)	Error SD $\sigma_{Ex}, \sigma_{Ey}, \sigma_{Ez}$ (m/s^2)
0; -30	0.0388; -0.0443; 0.0233	0.0047; 0.0106; 0.0027	-30; 0	-0.0391; 0.0446; -0.0351	0.0154; 0.0018; 0.0049
0; -25	0.0288; -0.0461; 0.0247	0.0025; 0.0058; 0.0027	-25; 0	-0.0043; 0.0359; -0.0305	0.0116; 0.0019; 0.0053
0; -20	0.0160; -0.0468; 0.0181	0.0054; 0.0055; 0.0039	-20; 0	0.0230; 0.0281; -0.0236	0.0144; 0.0028; 0.0040
0; -15	0.0095; -0.0401; 0.0153	0.0024; 0.0072; 0.0031	-15; 0	0.0250; 0.0192; -0.0191	0.0086; 0.0021; 0.0055
0; -10	0.0024; -0.0288; 0.0036	0.0019; 0.0041; 0.0012	-10; 0	0.0290; 0.0115; -0.0170	0.0105; 0.0014; 0.0034
0; -5	-0.0006; -0.0151; -0.0033	0.0028; 0.0066; 0.0020	-5; 0	0.0143; 0.0068; -0.0152	0.0107; 0.0024; 0.0052
0; 0	-0.0010; 0.0012; -0.0110	0.0000; 0.0000; 0.0000	0; 0	-0.0010; 0.0012; -0.0110	0.0000; 0.0000; 0.0000
0; 5	0.0008; 0.0086; -0.0145	0.0025; 0.0064; 0.0012	5; 0	-0.0159; 0.0082; -0.0037	0.0037; 0.0032; 0.0101
0; 10	0.0046; 0.0147; -0.0150	0.0025; 0.0057; 0.0008	10; 0	-0.0329; 0.0114; 0.0024	0.0072; 0.0031; 0.0100
0; 15	0.0105; 0.0173; -0.0116	0.0032; 0.0037; 0.0009	15; 0	-0.0364; 0.0169; 0.0114	0.0120; 0.0014; 0.0067
0; 20	0.0178; 0.0102; -0.0060	0.0048; 0.0041; 0.0017	20; 0	-0.0533; 0.0230; 0.0132	0.0102; 0.0015; 0.0062
0; 25	0.0274; -0.0164; 0.0088	0.0077; 0.0016; 0.0026	25; 0	-0.0567; 0.0301; 0.0153	0.0095; 0.0041; 0.0069
0; 30	0.0403; -0.0459; 0.0244	0.0073; 0.0065; 0.0029	30; 0	-0.0379; 0.0432; 0.0164	0.0112; 0.0025; 0.0064
-60; 0	-----; -----; 0.0209	-----; -----; 0.0106	35; 0	-----; -----; -0.0172	-----; -----; 0.0071
-55; 0	-----; -----; 0.0002	-----; -----; 0.0066	40; 0	-----; -----; -0.0339	-----; -----; 0.0076
-50; 0	-----; -----; -0.0111	-----; -----; 0.0076	45; 0	-----; -----; -0.0532	-----; -----; 0.0099
-45; 0	-----; -----; -0.0221	-----; -----; 0.0023	50; 0	-----; -----; -0.0592	-----; -----; 0.0072
-40; 0	-----; -----; -0.0315	-----; -----; 0.0025	55; 0	-----; -----; -0.0574	-----; -----; 0.0089
-35; 0	-----; -----; -0.0375	-----; -----; 0.0037	60; 0	-----; -----; -0.0491	-----; -----; 0.0180

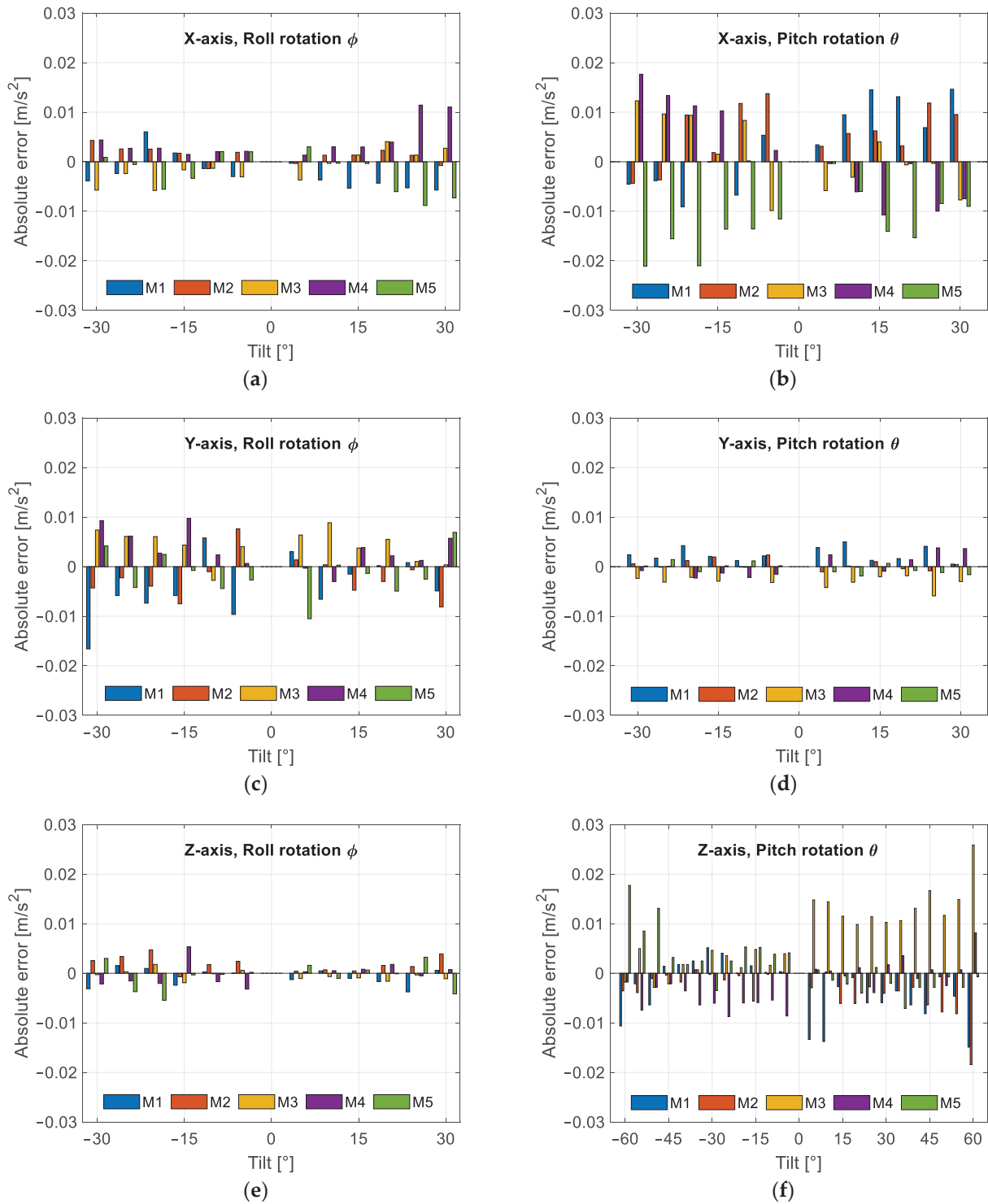


Figure 10. Absolute acceleration error after subtracting the mean error for each sensor orientation: (a) X-axis response to roll rotation; (b) X-axis response to pitch rotation; (c) Y-axis response to roll rotation; (d) Y-axis response to pitch rotation; (e) Z-axis response to roll rotation; (f) Z-axis response to pitch rotation.

From Figure 10a,d,e, we see that the errors are below $\pm 0.01 \text{ m/s}^2$, while in Figure 10b,c, the maximum errors are around $\pm 0.02 \text{ m/s}^2$, with the largest maximum error in Figure 10f being between $+0.03$ and -0.02 m/s^2 . With the correction term applied to the estimated acceleration components, the maximum errors for all sweeps are as follows: for the X-axis $\pm 0.02 \text{ m/s}^2$, for the Y-axis between $+0.01$ and -0.02 m/s^2 , and for the Z-axis between $+0.03$,

and -0.02 m/s^2 . In percentage, the maximum error was reduced by 58%, 58%, and 54% for the X-, Y-, and Z-axes, respectively.

Figure 11 shows the error between the gravity magnitude of the calibrated accelerometer output and the magnitude of the gravity field. Figure 11a shows the errors in each sensor orientation for each of the five sweeps and the root mean square error (RMSE), with the maximum value being 0.025 m/s^2 before error correction. From Figure 11b, we can see that the RMSE decreases to 0.007 m/s^2 ; therefore, the RMSE is reduced by 72%. In Table 5, we can see the pitch and roll angle for each of the positions in Figure 11.

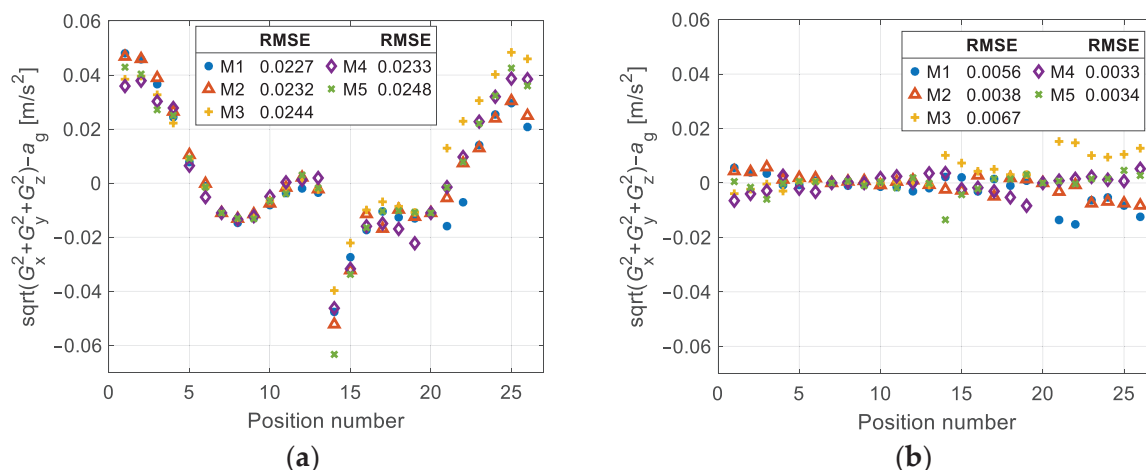


Figure 11. Absolute acceleration error, between the magnitude of the calibrated output of the accelerometer and the magnitude of the gravity field, and their respective RMSEs: (a) errors before subtracting the correction term; (b) errors after subtracting the correction term.

Table 5. Pitch and roll angle for each of the positions in Figure 11.

Orientation	Pitch, θ (deg)	Roll, ϕ (deg)	Orientation	Pitch, θ (deg)	Roll, ϕ (deg)	Orientation	Pitch, θ (deg)	Roll, ϕ (deg)
1	0	−30	10	0	15	19	−5	0
2	0	−25	11	0	20	20	0	0
3	0	−20	12	0	25	21	5	0
4	0	−15	13	0	30	22	10	0
5	0	−10	14	−30	0	23	15	0
6	0	−5	15	−25	0	24	20	0
7	0	0	16	−20	0	25	25	0
8	0	5	17	−15	0	26	30	0
9	0	10	18	−10	0	-	-	-

Now, we compare the errors obtained during calibration and the noise present in the circuit. The root mean square (RMS) values of the noise, which we can see in Figure 12, are $9.8 \times 10^{-4} \text{ V}$, $5.3 \times 10^{-4} \text{ V}$, and $4.9 \times 10^{-4} \text{ V}$ for the X-, Y-, and Z-axes, respectively. These values are converted to acceleration using the sensitivities shown in Figure 8, and result in 0.0033 m/s^2 , 0.0018 m/s^2 and 0.0017 m/s^2 , for the X-, Y- and Z-axes, respectively. The cutoff frequency is 20 Hz, and the window size is 58 s. The noise we see in Figure 12 is the sum of all noise sources present in the electronic circuit, which includes the thermal noise within the analog-to-digital converter and quantization noise due to the analog-to-digital conversion process [21]. The SNR (signal-to-noise ratio) is calculated using

$$\text{SNR} = 20 \log \left(\frac{\text{full-scale input}}{\text{noise}} \right). \quad (19)$$

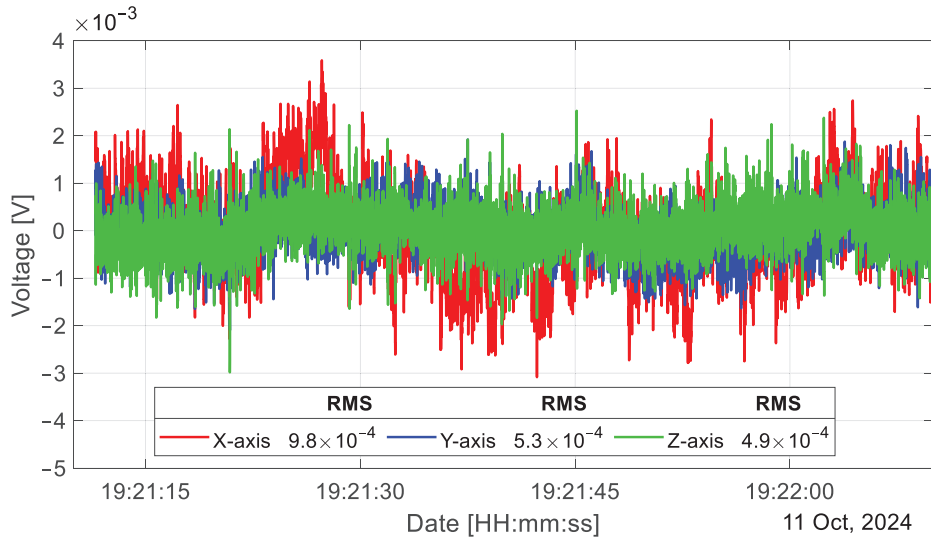


Figure 12. Electrical noise present at the output of the accelerometer electronic stage in a 60 s window.

The full-scale input of our sensor is 9.7789 m/s^2 , which, when converted to RMS, is 3.4574 m/s^2 . The full-scale input (or range) is converted to RMS for consistency with the RMS noise used in the calculation. Using Equation (19), the SNR values for the X-, Y-, and Z-axes are 60.40 dB, 65.7 dB, and 66.2 dB, respectively. On the other hand, the dynamic range (DR) is defined as

$$\text{DR} = 20 \log \left(\frac{\text{range}}{2 \times \text{noise}} \right). \quad (20)$$

DR is found to be 54.4 dB, 59.6 dB, and 60.1 dB for the X-, Y-, and Z-axes, respectively. We consider the minimum detectable acceleration as twice the noise value, i.e., the minimum detectable acceleration is $6.6 \times 10^{-3} \text{ m/s}^2$, $3.6 \times 10^{-3} \text{ m/s}^2$, and $3.4 \times 10^{-3} \text{ m/s}^2$.

4. Discussion and Conclusions

The characterization of a new triaxial FBG-based accelerometer has been performed. This triaxial accelerometer uses only a total of four fiber gratings. This has been achieved by reusing the reference FBG, which is not part of any sensor axis, with each of the sensing axes. We need to confirm that the Bragg wavelength of the reference FBG is not affected by vibrations, and that we can reuse it for demodulating the Bragg wavelength of the three sensing axes without loss of linearity. Therefore, the first experiment was to apply acceleration on the sensor axes, where the results show that the reference FBG is immune to vibrations. The reference FBG is oriented along the X-axis of the sensor, but because it is vibration-insensitive, its orientation has no effect on the responses of the sensing FBGs.

We have shown that all FBGs exhibit similar behaviors with respect to temperature, and that although the temperature compensation in the sensing FBGs is not at 100%, the reference FBG helps to compensate between 50% and 60%. The remaining temperature dependence can be neglected because the signal variation due to temperature is much slower than the signal variation due to vibration. Over a short period of time, the signal variation is also very small.

The overlapping interrogation method is nonlinear; then, the linearity of the output power from the overlapping interrogation method with respect to the input acceleration is another issue. A nearly linear operating range can be used with a suitable Bragg wavelength difference between the sensing FBG and the reference grating. The optimal spacing for the FBGs used in our sensor was found to be $102 \pm 2 \text{ pm}$, $98 \pm 2 \text{ pm}$, and $95 \pm 2 \text{ pm}$ for the X-, Y-, and Z-axes, respectively. Since the correlation factors for the X- and Y-axes are 1.0000

and 0.9999, respectively, we are operating within the nearly linear operating range, and the linear model of the sensor could be constructed.

We chose an analytical calibration method instead of an iterative one in order to determine the linearity of our sensor and to make the necessary adjustments to the Bragg wavelength of the gratings until we were operating in the center of the linear range of the demodulation scheme. We performed a 12-parameter calibration based on the linear least squares optimization by placing the sensor in seven different positions. The full-scale errors found were all under 1%—0.58%, 0.48%, and 0.39% for the X-, Y-, and Z-axes, respectively. We performed five sweeps in all sensor orientations, and the errors in the estimated acceleration components were repetitive; thus, we proposed to subtract a term, equal to the average of the five sweeps, from the estimated acceleration. After error compensation, the values obtained were 0.24%, 0.20%, and 0.18% for the X-, Y-, and Z-axes, which means reductions of 58%, 58%, and 54%, respectively.

The dynamic range found in the sensor was 54.4 dB, 59.6 dB, and 60.1 dB for the X-, Y-, and Z-axes, respectively. The minimum detectable acceleration was $6.6 \times 10^{-3} \text{ m/s}^2$, $3.6 \times 10^{-3} \text{ m/s}^2$, and $3.4 \times 10^{-3} \text{ m/s}^2$ for the X-, Y-, and Z-axes, respectively. The minimum detectable acceleration was small enough to detect a seismic event of magnitude less than 3.0, and the errors in the measurements were all low enough to estimate the amplitude of the motions accurately. The larger cross-sensitivity value for our sensor was found to be 1.51%.

The characteristics of our sensor are comparable to those of the high-figure-of-merit, low-cross-sensitivity accelerometer developed in [3], where the minimum detectable acceleration was $1.67 \times 10^{-3} \text{ m/s}^2$ with a transverse sensitivity of 2.3%. Although the accelerometer in [3] showed a higher sensitivity, 606.2 pm/g, compared to 78.4 pm/g, 71.9 pm/g, and 57.8 pm/g for the X-, Y-, and Z- axes of our sensor, we did not have the limitation of the interrogator resolution, and this has allowed us to achieve similar minimum detectable acceleration. Our sensor improved by 15-fold the minimum detectable acceleration when compared to that in [4], the value of which was $9.8 \times 10^{-2} \text{ m/s}^2$. The low-frequency accelerometer developed in [22] had a sensitivity of 681.7 pm/g and a transverse interference of less than 4.9%, and, considering the errors in the sensor, the minimum acceleration amplitude that it could measure was around $2.9 \times 10^{-1} \text{ m/s}^2$. In [23], a high sensitivity of 763.2 pm/g was achieved, but again, the minimum detectable acceleration was limited by the resolution of the interrogator used, which in this case was 1 pm, meaning that its minimum detectable acceleration was about $1.3 \times 10^{-2} \text{ m/s}^2$, and its cross-sensitivity was 5%. It is worth mentioning that the accelerometers mentioned above used an interrogator. Therefore, it is demonstrated that our sensor, which does not need an interrogator and uses only one FBG for each sensing axis, shows similar or better characteristics than those that do.

Our accelerometer can be used to measure vibration response in buildings, and help preserve structures and human lives. This device does not require an interrogator and can be operated as an autonomous unit. As future work, since the sensor sensitivity may vary with frequency, a dynamic characterization of the sensor is needed. The reduction in its size will improve internal temperature uniformity and further reduce the possibility of optical fiber bending with vibration. Further, the construction of the sensor frame in one piece will prevent possible resonances after prolonged use. This developed accelerometer prototype is suitable for field applications, and is not only a laboratory experiment.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s25030835/s1>, Table S1: F2-Temperature Dependence; Table S2: F3-Peaks_20240826180838; Table S3: F4a-Peaks_20240813122330; Table S4: F4b-Peaks_20240813123508; Table S5: F5to7-Caract_est_for_art_X; Table S6: F5to7-Caract_est_for_art_Y; Table S7: F8-Caract_est_for_art_X_Volt_avg; Table S8: F8-Caract_est_for_art_Y_Volt_avg; Table S9: F9-Positions_Voltaje15nov2024; Table S10: F10-Positios_Voltaje15nov2024_Sev_meas; Table S11: F12-Noise_data.

Author Contributions: Conceptualization, A.P.-A., L.A.-I. and G.E.S.-R.; methodology, A.P.-A. and G.E.S.-R.; software, A.P.-A.; validation, L.A.-I. and G.E.S.-R.; formal analysis, A.P.-A.; investigation, A.P.-A.; resources, L.A.-I. and G.E.S.-R.; data curation, A.P.-A.; writing—original draft preparation, A.P.-A.; writing—review and editing, A.P.-A., G.E.S.-R. and L.A.-I.; visualization, A.P.-A.; supervision, L.A.-I. and A.P.-A.; project administration, A.P.-A., G.E.S.-R. and L.A.-I.; funding acquisition, L.A.-I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by UNAM Posdoctoral Program, UNAM-PAPIIT grant IT100623, and UNAM-PAPIIT grant IT104123.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Perez-Alonzo, A.; Velazquez-Carreón, F.; Sandoval-Romero, G.E. High sensitivity three-axis optical vibration sensor for low-frequency measurement. In Proceedings of the 2023 World Congress on Advances in Structural Engineering and Mechanics (ASEM23), Seoul, Republic of Korea, 16–18 August 2023; GECE: Seoul, Republic of Korea, 2023; pp. 1–11.
2. Zabihollah, A.; Hajyalikhani, P.; Vuddandam, R. Evaluation of embedded FBG sensors for strain monitoring of residential timber buildings under various wind speeds. *Struct. Monit. Maint.* **2024**, *11*, 263–275.
3. Qiu, Z.; Zhang, Y. Three-Dimensional Low-Frequency Earthquake Monitoring Vibration Sensor Based on FBG. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–9. [CrossRef]
4. Zhang, B.; Ning, T.; Pei, L.; Liu, S.; Teng, Y. A Triaxial Dual FBG Accelerometer That Can Be Used for Seismic Wave Detection. *IEEE Sens. J.* **2024**, *24*, 10031–10038. [CrossRef]
5. Reghuprasad, A.E.; Colombero, C.; Godio, A. Serially Connected Cantilever Beam-Based FBG Accelerometers: Design, Optimization and Testing. *Sensors* **2023**, *23*, 3188. [CrossRef] [PubMed]
6. Hong, L.; Mai, M.; Zhang, R.; Luo, Q.; Wang, X. Rotating cantilever-based low-frequency double-fiber grating acceleration sensor. *Rev. Sci. Instrum.* **2023**, *94*, 065004. [CrossRef] [PubMed]
7. Hong, L.; Zhang, J.; Zhang, R.; Wu, P.; Duan, L. Research on low-frequency FBG accelerometer based on double curved beam reed. *Opt. Commun.* **2023**, *543*, 1–9. [CrossRef]
8. Perez-Alonzo, A.; Velazquez-Carreón, F.; Sandoval-Romero, G.E. Biaxial FBG Vibration Sensor With a Single Edge Filter and Matching Demodulation. *IEEE Access* **2022**, *10*, 127798–127805. [CrossRef]
9. Frosio, I.; Pedersini, F.; Borghese, N.A. Autocalibration of MEMS Accelerometers. In *Advanced Mechatronics and MEMS Devices*; Zhang, D., Ed.; Springer: New York, NY, USA, 2013; pp. 53–88. [CrossRef]
10. Sipos, M.; Paces, P.; Rohac, J.; Novacek, P. Analyses of Triaxial Accelerometer Calibration Algorithms. *IEEE Sens. J.* **2012**, *12*, 1157–1165. [CrossRef]
11. Wielandt, E. Seismic Sensors and their Calibration. In *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*; Deutsches GeoForschungsZentrum GFZ: Potsdam, Germany, 2012. [CrossRef]
12. Won, S.P.; Golnaraghi, F.A. Triaxial Accelerometer Calibration Method Using a Mathematical Model. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 2144–2153. [CrossRef]
13. Hisham, H. Sensor Mechanisms in Fiber Bragg Gratings. In *Fiber Bragg Grating Sensors: Development and Applications*; CRC Press: Boca Raton, FL, USA, 2020; pp. 49–54.
14. Sun, L.; Liang, T.; Sun, X.; Li, C.; Zhang, C. Temperature self-compensating and high-sensitivity FBG inclination sensor based on the sliding mass principle. *Opt. Fiber Technol.* **2023**, *81*, 103539. [CrossRef]

15. Fan, W.; Zhou, R.; Li, H.; Yu, D.; Qiao, X. Low-frequency temperature compensation fiber Bragg grating accelerometer based on double-arched beam. *Opt. Commun.* **2023**, *528*, 129058. [CrossRef]
16. Mapas Topográficos. Available online: <https://es-mx.topographic-map.com/> (accessed on 2 November 2024).
17. Pedley, M. *High Precision Calibration of a Three-Axis Accelerometer*; Freescale Semiconductor: Austin, TX, USA, 2013; pp. 1–41.
18. Forsberg, T.; Grip, N.; Sabourova, N. Non-iterative calibration for accelerometers with three non-orthogonal axes, reliable measurement setups and simple supplementary equipment. *Meas. Sci. Technol.* **2013**, *24*, 1–14. [CrossRef]
19. Basumallick, N.; Chatterjee, I.; Biswas, P.; Dasgupta, K.; Bandyopadhyay, S. Fiber Bragg grating accelerometer with enhanced sensitivity. *Sens. Actuators A Phys.* **2012**, *173*, 108–115. [CrossRef]
20. Lobo Ribeiro, A.B.; Ferreira, L.A.; Santos, J.L.; Jackson, D.A. Analysis of the reflective-matched fiber Bragg grating sensing interrogation scheme. *Appl. Opt.* **1997**, *36*, 934–939. [CrossRef] [PubMed]
21. McCarthy, M. *Peak-to-Peak Resolution Versus Effective Resolution*; Analog Devices: Wilmington, MA, USA, 2003; pp. 1–2.
22. Qiu, Z.; Su, K.; Zhang, J.; Teng, Y. A miniaturized low-frequency FBG accelerometer based on symmetrical cantilever beam. *IEEE. Sens. J.* **2023**, *23*, 5831–5840. [CrossRef]
23. Liu, Q.; Liu, W.; Wang, C.; Zhang, R.; Fan, W.; Yu, D.; Qiao, X. High figure of merit and low cross sensitivity fiber Bragg grating accelerometer based on double grid-diaphragms. *IEEE. Sens. J.* **2021**, *21*, 27503–27509. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

A Comprehensive Approach for Detecting Brake Pad Defects Using Histogram and Wavelet Features with Nested Dichotomy Family Classifiers

Sakthivel Gnanasekaran ^{1,2}, Lakshmi Pathi Jakkamputi ², Jegadeeshwaran Rakkiyannan ^{2,*}, Mohanraj Thangamuthu ^{3,*} and Yogesh Bhalerao ⁴

¹ School of Mechanical Engineering, Vellore Institute of Technology, Chennai 600127, India; sakthivel.g@vit.ac.in

² Centre for Automation, Vellore Institute of Technology, Chennai 600127, India; lakshmipati@vit.ac.in

³ Department of Mechanical Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India

⁴ Department of Mechanical Engineering and Design, School of Engineering, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK; y.bhalerao@uea.ac.uk

* Correspondence: jegadeeshwaran.r@vit.ac.in (J.R.); t_mohanraj@cb.amrita.edu (M.T.)

Abstract: The brake system requires careful attention for continuous monitoring as a vital module. This study specifically focuses on monitoring the hydraulic brake system using vibration signals through experimentation. Vibration signals from the brake pad assembly of commercial vehicles were captured under both good and defective conditions. Relevant histograms and wavelet features were extracted from these signals. The selected features were then categorized using Nested dichotomy family classifiers. The accuracy of all the algorithms during categorization was evaluated. Among the algorithms tested, the class-balanced nested dichotomy algorithm with a wavelet filter achieved a maximum accuracy of 99.45%. This indicates a highly effective method for accurately categorizing the brake system based on vibration signals. By implementing such a monitoring system, the reliability of the hydraulic brake system can be ensured, which is crucial for the safe and efficient operation of commercial vehicles in the market.

Keywords: histogram feature; wavelet feature; nested dichotomy; class-balanced nested dichotomy; data-near-balanced nested dichotomy

1. Introduction

The hydraulic brake system is a critical component of an automobile and is responsible for controlling the vehicle's speed and reducing the stopping distance. Any malfunction in the brake system can have a detrimental impact on the vehicle's stability and the safety of its passengers. According to the National Motor Vehicle Causation Survey, it has been revealed that 22% of accidents occur as a result of brake system failures [1]. In 2007, BMW recalled over 155,000 Sports Utility Vehicles (SUVs) due to brake fluid issues. This recall was initiated to address concerns regarding the performance and functionality of the brake system in these vehicles. The recall aimed to ensure the safety of the vehicles and their occupants by rectifying the brake-fluid-related issues. Similarly, Chrysler also faced a recall in 2007 involving approximately 60,000 vehicles. The recall was due to braking issues, specifically, a lack of success in the braking system. Chrysler undertook this recall to address the identified deficiencies and ensure the proper functioning of the brakes in the affected vehicles [2].

These recalls demonstrate the commitment of automotive manufacturers to prioritize safety and take prompt action to rectify any issues related to braking systems. This statistic underscores the significance of maintaining a reliable and properly functioning brake

system to prevent accidents and ensure the safety of both drivers and passengers. Regular maintenance, inspections, and testing are crucial for vehicle owners, manufacturers, and regulatory authorities to mitigate the risk of brake system failures. By implementing effective monitoring systems and taking preventive measures, potential issues can be identified early on, allowing for timely corrective actions and reducing the occurrence of accidents stemming from brake system failures. Examining such systems for defects will decrease the total crashes and enhance safety to a certain extent. Currently, a specific method for monitoring brake system failure is unavailable. Therefore, it is essential to develop a suitable method to monitor failures related to the brake system. The brake system may become flawed owing to pad wear, mechanical fade, inadequate oil pressure in the control cylinder, and oil leakage. If these flaws are not observed, accidents will take place. Hence, proactive maintenance is recommended to prevent accidents and can be attained through continuous monitoring [3,4].

Condition monitoring (CM) is continuously observing a system's operational attributes to initiate necessary steps before the occurrence of a breakdown. Many CM methods utilize devoted sensors and data assessment tools to examine a specific type of disparity in functional attributes [5]. Thomas et al. outlined a new communications module for identifying brake faults through sensors. The module detects the malfunction through the recorded signals from the brake system [6]. A method was patented for monitoring the vehicle condition and generating a maintenance plan according to the vehicle condition [7–9]. A new approach for monitoring the aircraft brake was proposed to determine the brake condition [10].

Nevertheless, to the best knowledge of the authors, there is no available solution for monitoring the brake pad conditions. Hence, a new method is required to diagnose brake pad-related failure. Fault diagnosis is a sub-area of CM in which specific system conditions are continuously studied. Fault identification can be performed in numerous ways, like visual testing, acoustic emission (AE) [11], thermal imagers, ultrasonic, and vibration signals [12]. Visual testing is the most commonly used technique in the industry. However, visual inspection has drawbacks since it requires skilled laborers who can identify the problems through experience. Hence, it is unsuitable for monitoring hidden components such as the brake system [13].

A passive infrared thermal imager was utilized to monitor the weld quality of the specimen, but sophisticated equipment was needed to measure the temperature variance [14]. Since it is an offline process, it is also unsuitable for brake fault diagnosis. AE can be used to monitor the machining process [15]. The bearing fault in a helicopter gearbox was identified by measuring the AE and vibration signals [16]. The spur gear condition was monitored using AE and vibration signals [17]. The study proved that vibration analysis is more appropriate for fault than AE signal analysis. However, the AE signal acquisition is complicated in machine components such as a brake.

Recently, most of the fault diagnostic studies focused on the vibration signal. The vibration signals are examined through frequency domain analysis, wavelet analysis, waveform examination, spectral analysis, statistical learning, histogram learning, etc. These examinations gave the data needed to decide on the maintenance plan. The consequences of such examinations are utilized for fault investigation to determine the fault's original cause. The brake system produces vibrations under various working circumstances. The measured vibration signals correlate precisely with the fault conditions, which are used to identify the defects in the system. Therefore, vibration signals can be considered appropriate for condition monitoring through feature-based analysis.

The feature is a specific, measurable characteristic of an anomaly being examined. Selecting data, separating, and categorizing independent features is essential in classification, pattern recognition, and regression. Numerous features, specifically statistical [18], histogram [19], and wavelet [20], can be obtained from the acquired signals. The application of a wavelet plays a crucial role in CM. The main advantage of the wavelet transform (WT) is to denoise or compress a signal without reducing the original signal and computational

time. WT could be successfully applied to the different fault diagnosis studies. Hence, WT has been used for diagnosing brake failure in this study.

After extracting the features, the most significant features have to be selected. The attribute evaluator, the influence of the number of features, and Decision Tree (DT) are the three most important methods for selecting the features. DT utilizes a graphical structure to select the most significant features [21]. The attribute evaluator calculates the value of a characteristic subset by studying each feature's analytical capacity and severance level. The influence of the number of features is one of the most helpful techniques in feature selection. Instead of taking all the features, the performance of the specific features is tested using this examination. The influence of several features study was applied to diagnose the brake faults and identify the best feature combination that produces the maximum classification accuracy [22].

Feature classification is the last phase in the classifier. The selected features should be classified for predicting performance. The commonly used classifiers are Decision Tree [23,24], Fuzzy [25], Artificial Neural Network [20], Bayes Net [26], Naive Bayes [20], Support Vector Machine (SVM) [27], Convolutional Neural Network (CNN) [28] and the Hidden Markov Model [29]. Different classifiers monitored the tool wear using vibration signals with the discrete wavelet feature [20]. So, there is considerable scope for the wavelet features with other machine learning algorithms in the brake monitoring system. The nested dichotomy (ND) classifier was used for fault diagnosis in the hydraulic brake system. They proved that the class-balanced nested dichotomy (CBND) prediction accuracy was more than that of the SVM and DT classifiers. The authors used the statistical feature alone for classification. The other features can also be considered to make a more detailed study.

To the authors' best knowledge, there is no literature on CM of a brake system using histogram and wavelet features. Hence, histogram and wavelet features are considered to monitor the hydraulic brake system. An attribute evaluator is used for feature selection. The feature selection through the attribute evaluator is validated using the influence of the number of features studied. The preferred wavelet and histogram features were categorized using the family of ND classifiers, and the results were compared and discussed.

2. Experimental Procedure

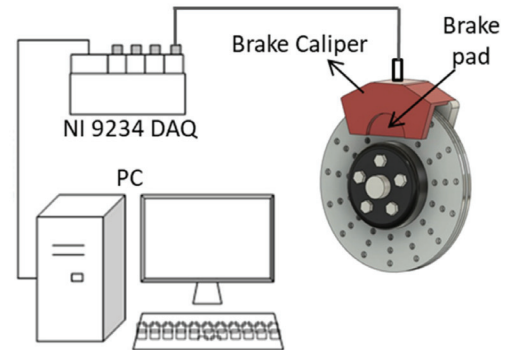
Experimentation was conducted to supervise the brake system using histogram and wavelet features extracted from the vibration signals. A brake system test rig was used to perform the experiments. The experimental test rig and procedure are briefed in the subsequent sections.

2.1. Experimental Test Rig

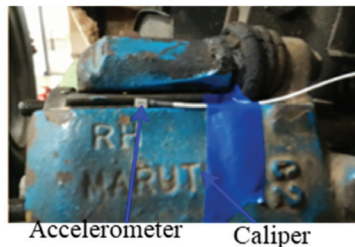
A commercial vehicle, as depicted in Figure 1, was utilized for the experiment. Its drive shaft consistently operated at 331 rpm on a test bed equipped with rollers. A piezo-electric uni-axial accelerometer (Make: Dytran, Sensitivity: 10 mV/g) was securely positioned at the center of the flat surface of the brake caliper, as illustrated in Figure 1c. As the drive shaft rotated, various dynamic forces and loads impacted the brake pad and caliper assembly. These forces resulted from the interaction between the brake pad and the brake rotor, leading to vibrations within the brake pad and caliper assembly. As the brake pad wears, it undergoes changes in both rigidity and mass distribution, which, in turn, cause fluctuations in the assembly's stiffness and mass. These variations in vibration within the brake pad assembly, associated with the brake pad's stiffness, are crucial parameters for understanding and identifying brake pad defects. The attached piezo-electric accelerometer captures these variations in vibration signals, which are then processed using the NI 9234 data acquisition (DAQ) system, converting them into digital data. The encoded data from the DAQ system is subsequently processed and presented in the LabVIEW 2021 software, providing results in the time domain as presented in Figure 1b. The experiments were performed with different brake pad conditions, as depicted in Figure 1d.



(a) Real time simulator test rig



(b) Block diagram of the experimental setup



(c) Exploded view of experiment setup



(d) Brake pad conditions

Figure 1. Experimental test rig.

2.2. Experimental Procedure

The frequently occurring failures, such as Brake Oil Spill (BOS), Drum Brake Mechanical Fade (DBM), Disc Pad Wear (Uneven) Inner (DBP(UE)I), Disc Pad Wear (Uneven) Inner and Outer (DBP(UE)IO), Disc Pad Wear (Even) Inner (DBPI), Disc Pad Wear (Even) Inner and Outer (DBPIO), Drum Brake Pad Wear (DBPW), Air in Brake Fluid (ABF), and Reservoir Leak (RL) were considered for simulating the fault conditions. The vibration signals were acquired from the test setup with good and various simulated faulty brake conditions. The LabVIEW program used for vibration signal acquisition is shown in Figure 2. The data acquisition was performed with the following specifications:

- Sample length and no. of samples: 2^{13} and 55 (arbitrarily chosen);
- Sampling frequency: 24 kHz (as per the Nyquist sampling theorem);
- Wheel speed and brake force: 331 rpm (~30 KMPH) and 68.7 N.

The wheel speed, which was 331 rpm, was measured and controlled using a tachometer. The brakes were applied gradually and evenly to prevent RPM fluctuations, and, importantly, as the brakes were applied, minor manual adjustments were made to the throttle to ensure a consistent RPM was maintained.

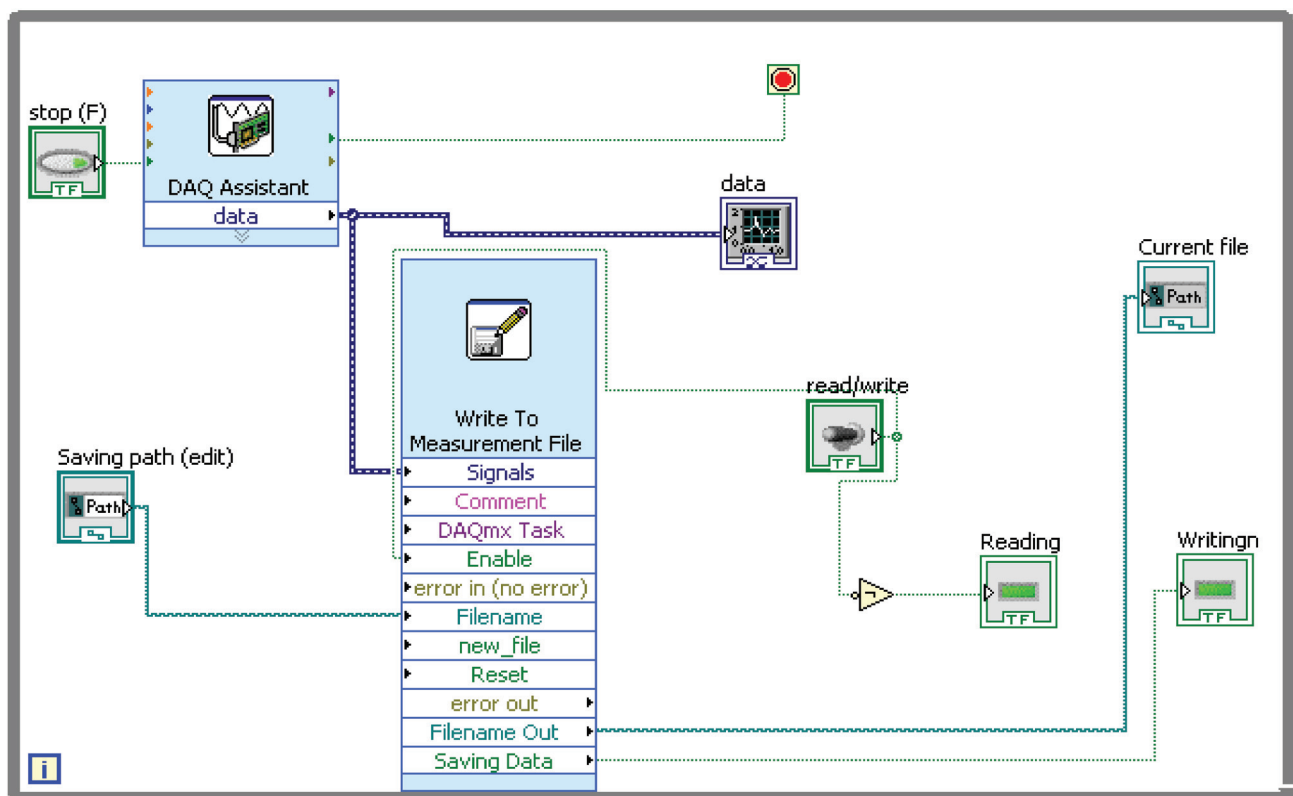


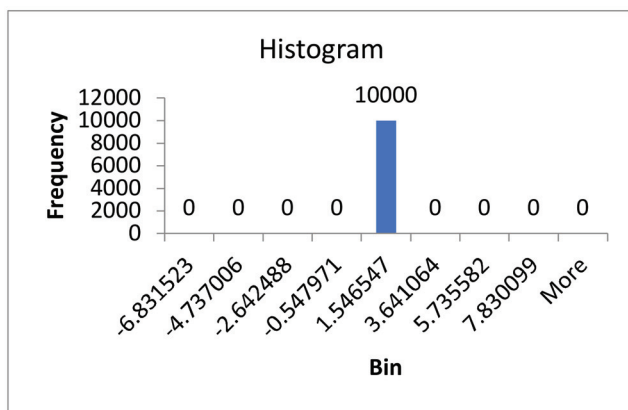
Figure 2. LabVIEW program for data acquisition.

3. Histogram Features

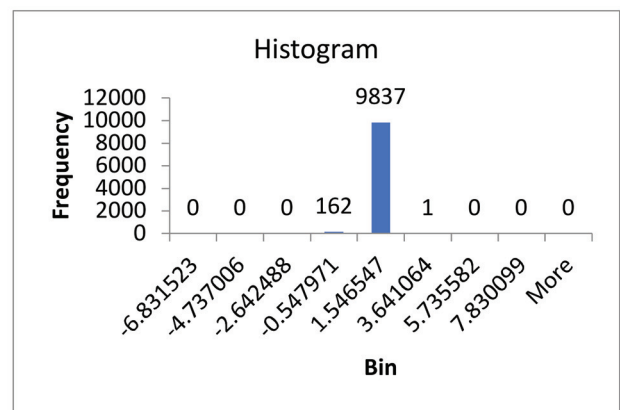
The histogram data is used as a feature in the brake CM. The magnitude range of the vibration is split into various sub-ranges (bins) from the smaller to the larger value of the vibration signal. The quantity of information focuses on whose magnitude of vibration value catarracts inside a specific bin are calculated. The bin and count frame the histogram's x and y axes. The histogram for each defect is plotted utilizing vibration motions as an independent graph. The goal is to discover the bins whose y-axis values are the same for a specific defect but not the same for another defect. The histogram for a specific brake defect might be small; however, it might be huge for other defects. The bin range chosen should suit all defects. The strategy for estimating the base of least magnitude is as follows:

- Compute the least magnitude of each signal in a specific condition;
- Repetition of step (i) for all conditions;
- Identify the least of those least magnitudes.

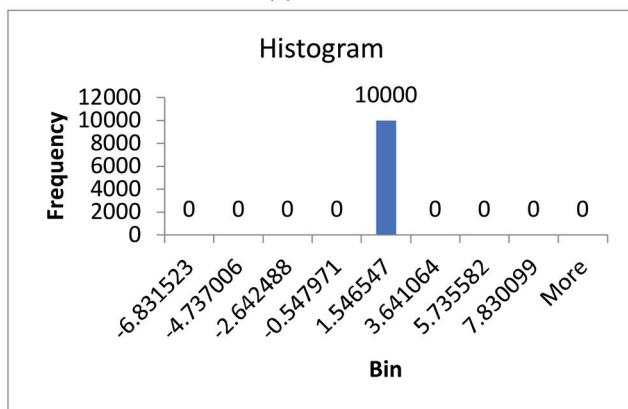
Similarly, the base of the largest magnitude is computed. The bin width should be fixed so that the bin height is diverse for various brake conditions. It requires not to be valid for all widths of the bin, but rather, at least a few of them should be taken after this measure is utilized as a feature for recognizing different conditions. This study used the bin width and range to plot the histogram, as shown in Figure 3a. The vibration value between the larger and smaller values was divided by a number of bins. The larger and smaller value range is divided into 69 bins (2 to 70). Figure 3a–j shows the histogram obtained from the vibration signals.



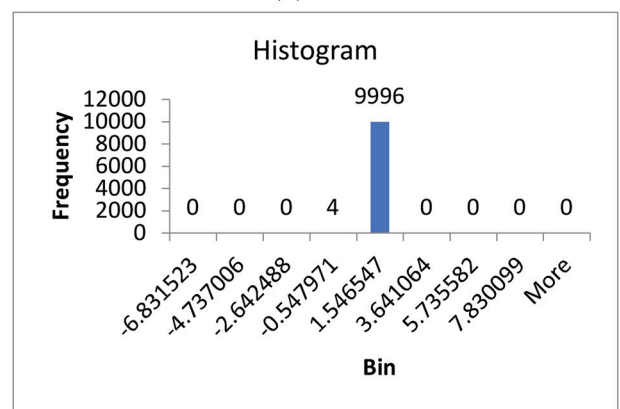
(a) GOOD



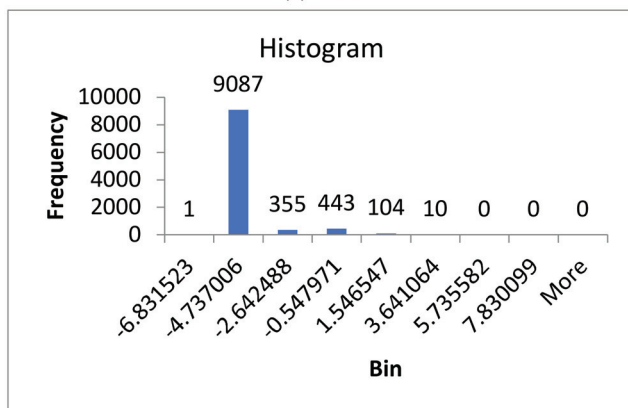
(b) ABF



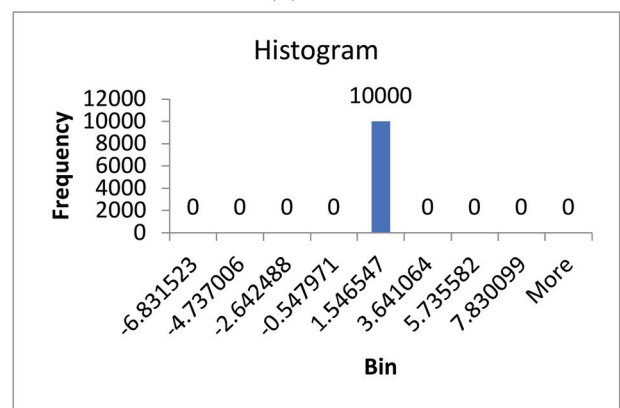
(c) BOS



(d) DBPI



(e) DBPIO



(f) DBP(UE) I

Figure 3. Cont.

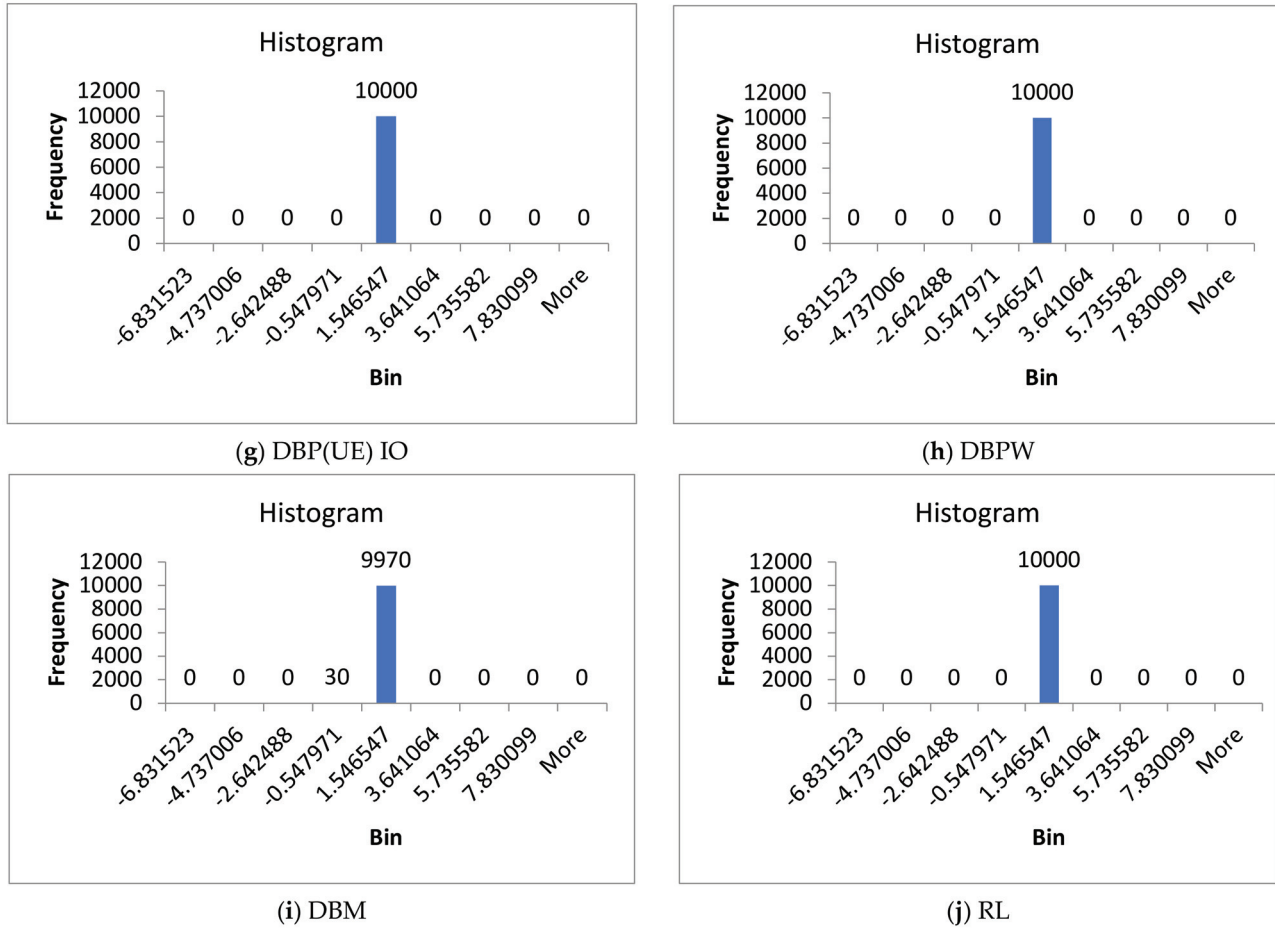


Figure 3. Histogram of different faulty conditions for 7 bin range.

4. Wavelet Filter

The wavelet transform (WT) is a prevailing time and frequency domain signal tool. WT can be continuous or discrete, permitting the investigation of fleeting, irregular, or local parts. The continuous WT can expose more insights about a signal but is computationally intensive. For fault analysis, discrete wavelet transform (DWT) aims to characterize a signal proficiently with fewer parameters and computation time. DWT is an appropriate method for this study [30].

4.1. Wavelet Transform

The WT is drawn as

$$WT_f(p, q) = f(t) \cdot \psi_{p,q}(t)$$

$$WT_f(p, q) = \frac{1}{\sqrt{p}} \int_0^{\infty} f(t) \psi\left(\frac{t-q}{p}\right) dt \quad (1)$$

where $\psi_{p,q}(t)$ is the alternative of the “Mother wavelet”, which is considered as

$$\psi_{p,q}(t) = \frac{1}{\sqrt{p}} \int_0^{\infty} \psi\left(\frac{t-q}{p}\right) \quad (2)$$

where $t, p, q \in \mathbb{R}$, and $p \neq 0$ are continuously varying quantities, p is the scale factor, q is a shift factor, and $\frac{1}{\sqrt{p}}$ is used for energy conservation. If p is small, extreme-frequency components are analyzed, and vice versa.

4.2. Discrete Wavelet Transform

DW performs scaling and translation procedures in distinct phases, restraining the selection of scales (a) and translation (τ) to specific numbers. However, the examination is satisfactorily precise and articulated in the following equation:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right) \quad (3)$$

where j and k are integers.

The time scale is tested at a distinct interval due to discretizing the wavelets, called decomposition, and succeeds a sequence of wavelet coefficients [31]. The wavelet features were obtained for every condition utilizing the Daubechies wavelets “db1” to “db10”. The wavelets studied for this examination are provided below:

- Haar—‘haar’;
- Daubechies—‘db’;
- Symlet—‘sym’;
- Coiflets—‘coif’;
- Biorthogonal—‘bior’;
- Reverse biorthogonal—‘rbio’;
- Discrete approximation of Meyer—‘dmey’;
- Fejer–korovkin—‘fk’.

The WT removed the noise from the raw vibration signal and reconstructed the signals. The successive approximation filters certain more high-frequency data from the raw signal. Using this method, substantial data loss is possible. Therefore, a new method is required for denoising. Optimal denoising demands a more sensitive approach known as thresholding, which includes discarding a small number of progressive elements that exceed a specific threshold. Thresholding is reformed from level to level. For the current study, level 5 approximation was applied. The eight types of wavelet features were extracted for every condition. Figure 4a–j shows the denoised sample signal attained through DWT for the Coiflets wavelet. From the denoised signals, relevant statistical features were collected.

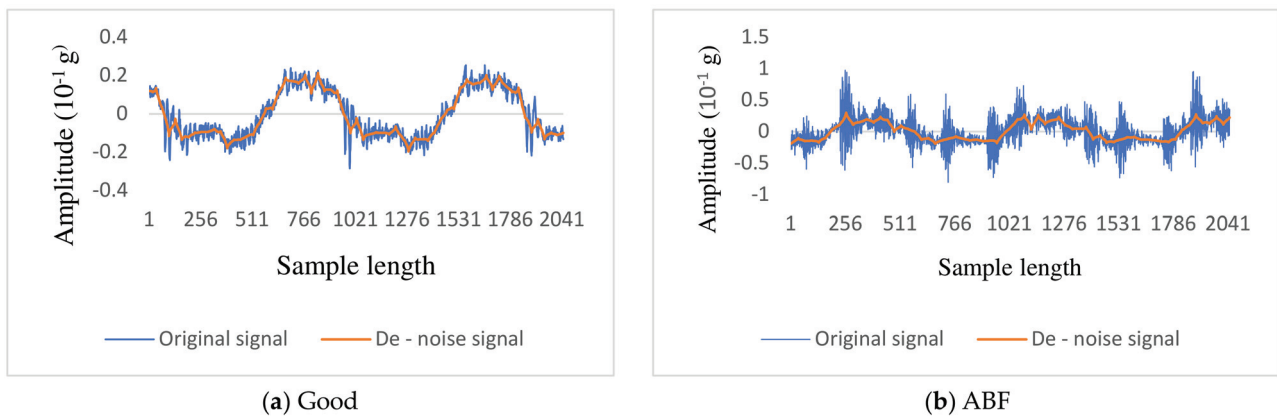


Figure 4. Cont.

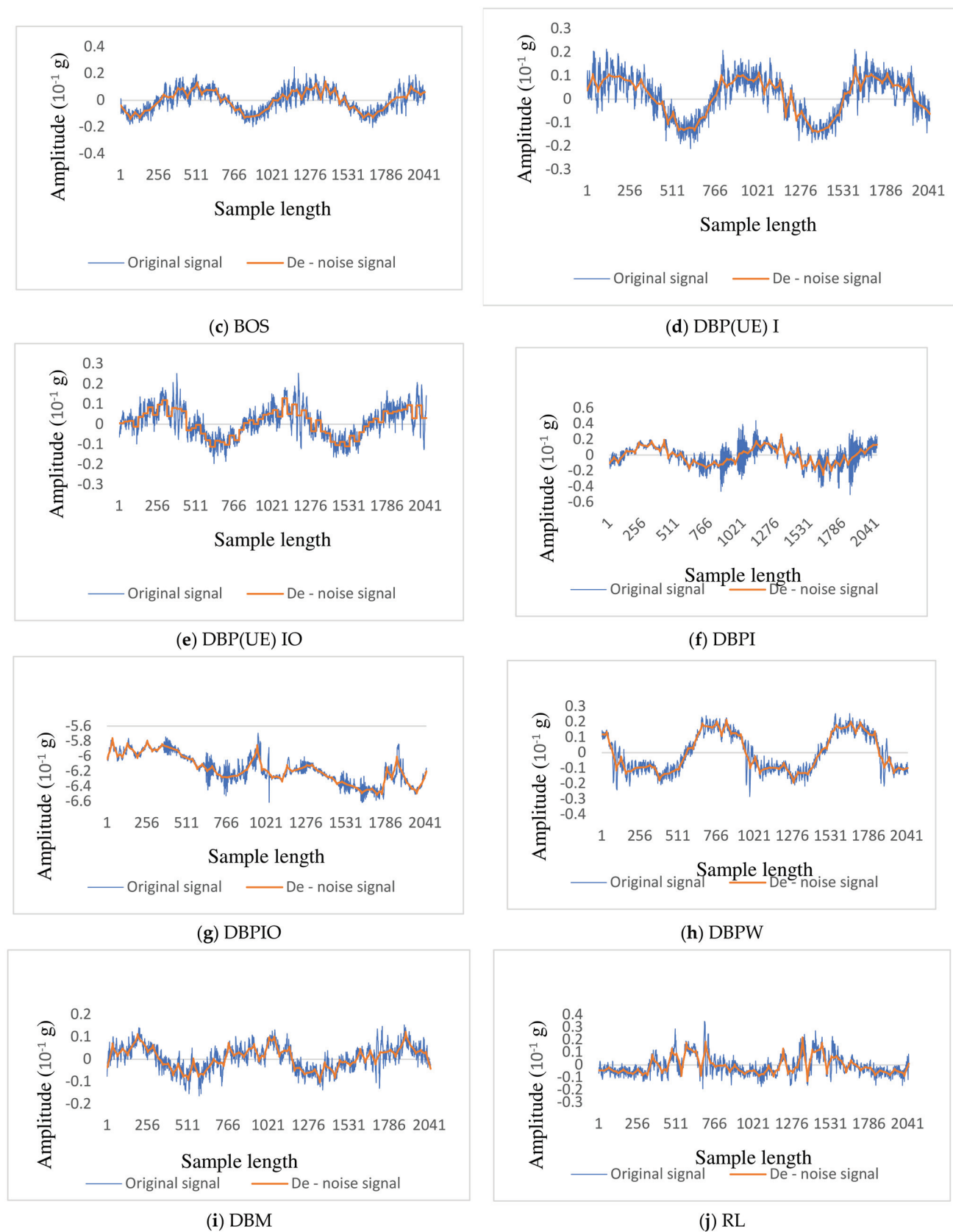


Figure 4. Original and denoised signals for various faulty conditions.

5. Feature Selection

The attribute evaluator, DT, and the influence of the number of features studied were used to select the most significant features. This process is called feature selection. The attribute evaluator was utilized to find the best feature set from the obtained features. In this study, the two different search methods, Correlation-based Feature Selection Subset Eval and Best First Search were applied to select the sequence of the significant features. The chosen feature was tested through the effect of the number of features studied. The first element suggested by the attribute evaluator was classified first, and the accuracy was recorded. Then, the second feature recommended by the evaluator was combined with the first feature and classified. The classification accuracy with two feature combinations was also recorded. The process was repeated until all the features were combined and classified. The best feature combination that only provides maximum accuracy was selected for further study. The DT was used for choosing the most significant features from the histogram feature.

6. Feature Classification

The selected features should be classified to obtain the classification result. A system of ND is employed to disintegrate a multinomial problem into multiple binary problems. The disintegration is described using a binary tree. A set of class labels is stored in each tree node, and a binary classifier from the training data. The training data is divided into two subsets conforming to the two meta-classes, and one subset of training data is considered a true example.

In contrast, the other subset of testing data is considered a false example. The two next-in-line root nodes receive two subsets of the true class labels with their conforming training data. A tree is constructed by continuously employing this process. Lastly, this process succeeds a leaf node if the node has only one label. The choice of tree structure significantly impacts the accuracy.

The CBND method depends on regulating the number of conditions at every node. As an alternative to examining each possible tree's space, CBND examined the space of each balanced tree. The quantity of possible trees in CBND is lesser than the entire tree of ND. The detailed algorithm for developing a system of CBND is given in [32]. The usual difficulty with the CBND method is that few multinomial issues are not balanced and more populated than others. A CBND does not have additional data balanced. This can undesirably affect the computational time of the algorithm. DNBND arbitrarily assigns the classes to two subsets until the training data size in one subset surpasses 50% of training data at the node. With the DNBND algorithm, assortment suffers when class dispersal is disturbed. The number of classes depends on the class distribution in the data.

7. Results and Discussion

The wavelet and histogram features were extracted, and the most significant features were chosen. The chosen features were classified using ND family classifiers.

7.1. Histogram Feature Classification Using Nested Dichotomy Family Classifiers

The DT algorithm chose the most significant features [23]. All the 69 bins ranging from 2 to 70 were classified using the DT algorithm one after another. The classification accuracy for different bin ranges is shown in Figure 5. The ND classifier delivers a maximum classification accuracy for the 59th bin of 97.64%. The algorithm uses decision rules for classification. These decision rules are graphically represented as a DT, as shown in Figure 6. Hence, the graphical DT shows the contributing bin ranges. Even though 59 bin ranges were used as input in the algorithm, it uses only 7 for classification. Figure 6 selected the contributing bin range (59th) alone which have the maximum classification accuracy (97.64%). The features were categorized using various ND-family algorithms, including ND, CBND, and DNBND. The parameters used for the classification are given below:

- Random no. seed: 1;

- No. of trees: 10;
- Base learner: best first tree;
- Tree depth: unlimited.

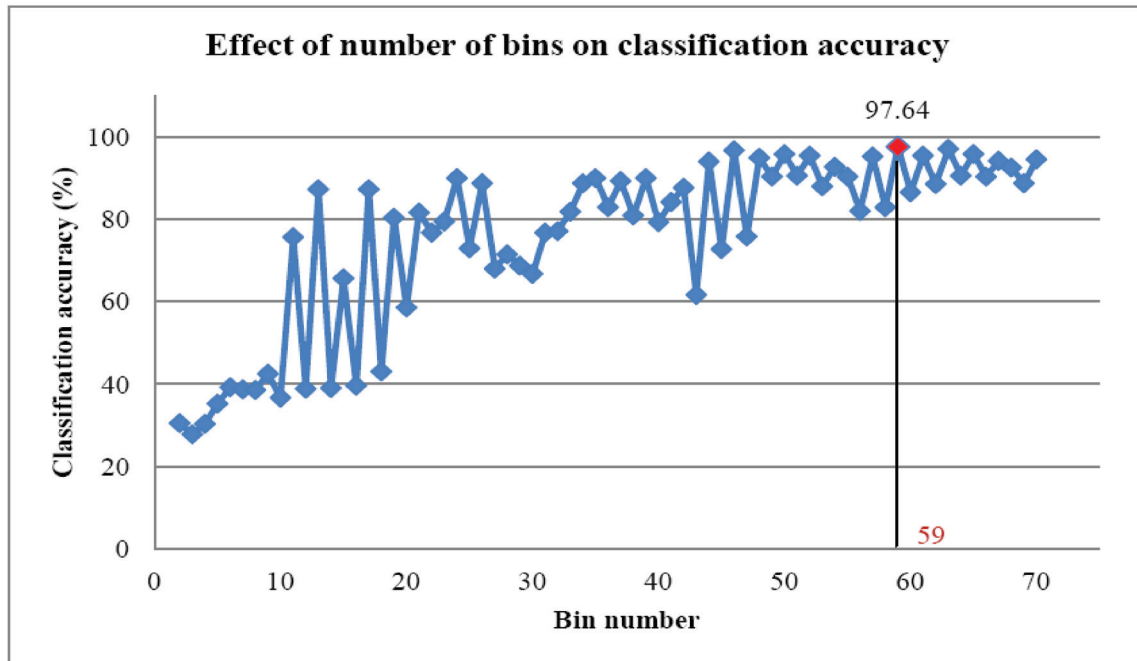


Figure 5. Classification accuracy vs. number of bins.

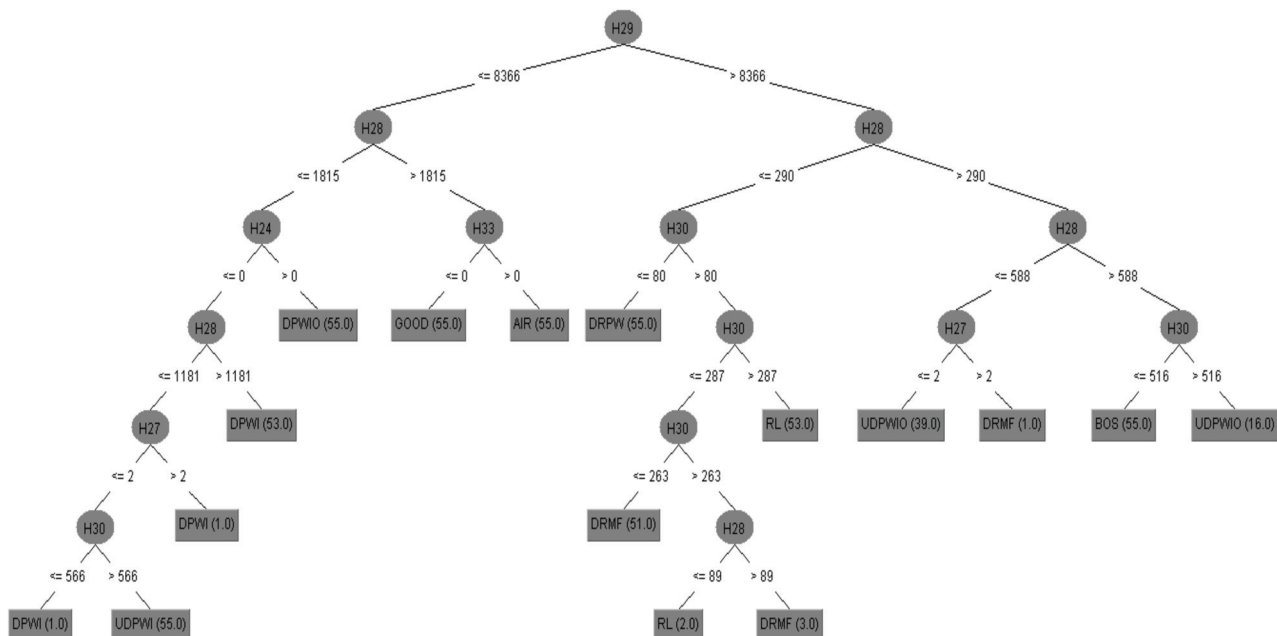


Figure 6. Graphical DT with histogram features.

The categorization accuracy is displayed in Figure 7. In Figure 7, the first cell signifies the number of data conforming to the “GOOD” class. The sum of data points in the first row is 55. All are correctly categorized and yield a classification accuracy of 100%. Likewise, the other element in the first row represents the false categorization. There are no data points on the first row beside the first column. This means that there is no false classification. Comparably, the sixth row sixth column signifies the DBPI condition. Among the 55 datapoints, 53 were

categorized accurately, and two were wrongly classified under the DBP(UE)I condition. The classification accuracy was estimated as 96.4%. Among the 550 datapoints, 538 were accurately classified, with an overall classification accuracy of 97.5%.

Category	GOOD	BOS	DBM	DBP(UE)I	DBP(UE)IO	DBPI	DBPIO	DBPW	ABF	RL	Precision	False discovery rate
GOOD	55	0	0	0	0	0	0	0	0	0	100	0.00
BOS	0	55	0	0	0	0	0	0	0	0	100	0.00
DBM	0	0	53	0	2	0	0	0	0	0	96	3.64
DBP(UE)I	0	0	0	54	0	1	0	0	0	0	98	1.82
DBP(UE)IO	0	0	0	0	54	0	0	1	0	0	98.2	1.82
DBPI	0	0	0	2	0	53	0	0	0	0	96	3.64
DBPIO	1	0	0	0	0	0	54	0	0	0	98	1.82
DBPW	0	0	0	0	1	0	0	54	0	0	98.2	1.82
ABF	0	0	0	0	0	0	0	3	52	0	95	5.45
RL	0	0	2	0	1	0	0	0	0	52	94.5	5.45
TPR	98.2	100.0	96.4	96.4	93.1	98.1	100.0	93.1	100.0	100.0	97.5	97%
FNR	1.8	0.0	3.6	3.6	6.9	1.9	0.0	6.9	0.0	0.0	2.4	

Figure 7. Confusion matrix for ND with histogram features.

The rows and columns represent the anticipated and actual conditions, respectively. The diagonal cells correspond to data points that are appropriately categorized. The off-diagonal cells represent wrongly categorized data. The no. of data is shown in each cell, and the exact prediction accuracy is shown in the last cell. The last two columns on the right show the % of all the data predicted to belong to each appropriate (precision) and wrongly classified (false discovery rate) condition. The last two rows at the bottom show the % of the data predicted appropriately (True Positive Rate (TPR)) and wrongly (False Negative Rate (FNR)) classified condition. The cell in the bottom right shows the overall accuracy.

Similarly, the selected features were classified using CBND and DNBND classifiers, and the corresponding classification accuracies were found. The overall accuracy for CBND and DNBND was 97% and 99%, respectively. Figures 8 and 9 represent the confusion matrix obtained for CBND and DNBND with histogram features.

Category	GOOD	BOS	DBM	DBP(UE)I	DBP(UE)IO	DBPI	DBPIO	DBPW	ABF	RL	Precision	False discovery rate
GOOD	55	0	0	0	0	0	0	0	0	0	100	0.00
BOS	0	54	0	0	0	0	0	0	1	0	98	1.82
DBM	0	0	52	0	1	0	0	0	0	2	95	5.45
DBP(UE)I	0	0	0	54	0	0	1	0	0	0	98	1.82
DBP(UE)IO	0	0	0	0	55	0	0	0	0	0	100.0	0.00
DBPI	0	0	0	1	0	54	0	0	0	0	98	1.82
DBPIO	1	0	0	0	0	0	54	0	0	0	98	1.82
DBPW	0	0	0	0	1	0	0	54	0	0	98.2	1.82
ABF	0	5	0	0	0	0	0	3	47	0	85	14.55
RL	0	0	1	0	0	0	0	0	0	54	98.2	1.82
TPR	98.2	91.5	98.1	98.2	96.5	100.0	98.2	94.7	97.9	96.4	97.5	97%
FNR	1.8	8.5	1.9	1.8	3.5	0.0	1.8	5.3	2.1	3.6	2.4	

Figure 8. Confusion matrix for CBND with histogram features.

7.2. Wavelet Feature Classification Using ND Family Classifiers

Eight families of wavelets were selected, and a denoised signal was obtained from each family. From the denoised signal, twelve statistical features were extracted. The attribute evaluator was used to select the most significant features, and nine statistical features were acknowledged and considered as input for the classifier. The nominated nine features were appraised using the influence of the no. of feature study. Table 1 indicates the influence of the no. of features studied performed on the Coiflets wavelet. The CBND produces a better accuracy (99.45%) with the top eight features.

The feature assortment process was applied for all wavelet families, and the resultant accuracy was recorded and presented in Table 2. In Table 2, the Coiflets wavelet (Coif) produced a maximum classification accuracy compared to other wavelets. Therefore, the Coiflets wavelet was selected. From the influence of the no. of feature study, the eight features that contributed to the classification were chosen for the study. They were as follows: (1) Mean, (2) Standard Error, (3) Median, (4) Standard Deviation, (5) Kurtosis, (6) Skewness, (7) Range, and (8) Minimum. The classification accuracy for the ND family (ND, CBND, and DNBND) algorithms was found. Figure 10 represents the confusion matrix for ND. Figure 10 reveals the overall classification accuracy of 99%.

Category	GOOD	BOS	DBM	DBP(UE)I	DBP(UE)IO	DBPI	DBPIO	DBPW	ABF	RL	Precision	False discovery rate
GOOD	55	0	0	0	0	0	0	0	0	0	100	0.00
BOS	0	54	0	0	1	0	0	0	0	0	98	1.82
DBM	0	0	52	0	1	0	0	0	0	2	95	5.45
DBP(UE)I	0	0	0	55	0	0	0	0	0	0	100	0.00
DBP(UE)IO	0	0	0	0	55	0	0	0	0	0	100.0	0.00
DBPI	0	0	0	1	0	54	0	0	0	0	98	1.82
DBPIO	0	0	0	0	0	0	55	0	0	0	100	0.00
DBPW	0	0	0	0	1	0	0	54	0	0	98.2	1.82
ABF	0	0	0	0	0	0	0	0	55	0	100	0.00
RL	0	0	1	0	0	0	0	0	0	54	98.2	1.82
TPR	100.0	100.0	98.1	98.2	94.8	100.0	100.0	100.0	100.0	96.4	97.5	99%
FNR	0.0	0.0	1.9	1.8	5.2	0.0	0.0	0.0	0.0	3.6	2.4	

Figure 9. Confusion matrix for DNBND with histogram features.

Table 1. Influence of no. of feature study with Coiflets wavelet.

No. of Features	Categorization Accuracy (%)		
	ND	CBND	DNBND
1	28.91	28	28.36
2	89.45	88.55	90
3	96	95.27	95.82
4	94.55	94.91	95.27
5	98	98.73	98.55
6	98	99.27	98.73
7	98.73	98.55	98.73
8	98.73	99.45	99.27
9	98.55	99.27	98.91

Table 2. Classification accuracy for various wavelet families.

Wavelet	Classification Accuracy (%)		
	ND	CBND	DNBND
Coif1L5	98.73	99.45	99.27
Dmey	99.27	99.09	99.09
Bior2.2L5	98.18	98.18	98.36
Db5L5	99.09	98.73	98.36
Haar	98.73	98.55	98.55
Sym3L5	98.73	98.55	98.55
Rbior2.8L5	86.73	85.82	86
Fk4L5	97.64	98.18	98.36

Category	GOOD	BOS	DBM	DBP(UE)I	DBP(UE)IO	DBPI	DBPIO	DBPW	ABF	RL	Precision	False discovery rate
GOOD	55	0	0	0	0	0	0	0	0	0	100	0.00
BOS	0	55	0	0	0	0	0	0	0	0	100	0.00
DBM	0	0	53	0	1	0	0	0	0	1	96	3.64
DBP(UE)I	0	0	0	55	0	0	0	0	0	0	100	0.00
DBP(UE)IO	0	1	0	0	53	1	0	0	0	0	96.4	3.64
DBPI	0	0	0	0	0	55	0	0	0	0	100	0.00
DBPIO	0	0	0	0	0	0	55	0	0	0	100	0.00
DBPW	0	0	1	0	0	0	0	54	0	0	98.2	1.82
ABF	0	0	0	0	0	0	0	0	55	0	100	0.00
RL	0	0	2	0	0	0	0	0	0	53	96.4	3.64
TPR	100.0	98.2	94.6	100.0	98.1	98.2	100.0	100.0	100.0	98.1	97.5	99%
FNR	0.0	1.8	5.4	0.0	1.9	1.8	0.0	0.0	0.0	1.9	2.4	

Figure 10. Confusion matrix for ND with Coiflets wavelet.

Comparably, the remaining data points in the first row signify the false classification details. All other values are zero in the first row other than the first element, revealing no false classification. Similarly, the last element in the last row represents RL. Among the 55 datapoints, 53 belong to RL and are accurately categorized, and 2 were wrongly

categorized as the DBM condition. Among the 550 datapoints, 543 were categorized accurately, with an accuracy of 99%.

Likewise, the classification accuracy was computed for CBND and DNBND. Both the classifier produced 99.45% and 99.27% as classification accuracy, respectively. Figures 11 and 12 show the graphical representation of the confusion matrix attained for the CBND and DNBND. The overall accuracy of the ND classifier is presented in Table 3. The CBND gives an enhanced accuracy of 99.45% among the three ND family classifiers. Among the 550 datapoints, 547 were accurately classified. Denoising using wavelet significantly reduced noise and increased classification accuracy.

Category	GOOD	BOS	DBM	DBP(UE)I	DBP(UE)IO	DBPI	DBPIO	DBPW	ABF	RL	Precision	False discovery rate
GOOD	55	0	0	0	0	0	0	0	0	0	100	0.00
BOS	0	55	0	0	0	0	0	0	0	0	100	0.00
DBM	0	0	55	0	0	0	0	0	0	1	98	1.79
DBP(UE)I	0	0	0	55	0	0	0	0	0	0	100	0.00
DBP(UE)IO	0	1	0	0	54	0	0	0	0	0	98.2	1.82
DBPI	0	0	0	0	0	55	0	0	0	0	100	0.00
DBPIO	0	0	0	0	0	0	55	0	0	0	100	0.00
DBPW	0	0	1	0	0	0	0	54	0	0	98.2	1.82
ABF	0	0	0	0	0	0	0	0	55	0	100	0.00
RL	0	0	1	0	0	0	0	0	0	54	98.2	1.82
TPR	100.0	98.2	96.5	100.0	100.0	100.0	100.0	100.0	100.0	98.2	97.5	99%
FNR	0.0	1.8	3.5	0.0	0.0	0.0	0.0	0.0	0.0	1.8	2.4	

Figure 11. Confusion matrix for CBND with wavelet.

Category	GOOD	BOS	DBM	DBP(UE)I	DBP(UE)IO	DBPI	DBPIO	DBPW	ABF	RL	Precision	False discovery rate
GOOD	55	0	0	0	0	0	0	0	0	0	100	0.00
BOS	0	55	0	0	0	0	0	0	0	0	100	0.00
DBM	0	0	55	0	0	0	0	0	0	1	98	1.79
DBP(UE)I	0	0	0	55	0	0	0	0	0	0	100	0.00
DBP(UE)IO	0	2	0	0	53	0	0	0	0	0	96.4	3.64
DBPI	0	0	0	0	0	55	0	0	0	0	100	0.00
DBPIO	0	0	0	0	0	0	55	0	0	0	100	0.00
DBPW	0	0	1	0	0	0	0	54	0	0	98.2	1.82
ABF	0	0	0	0	0	0	0	0	55	0	100	0.00
RL	0	0	1	0	0	0	0	0	0	54	98.2	1.82
TPR	100.0	96.5	96.5	100.0	100.0	100.0	100.0	100.0	100.0	98.2	97.5	99%
FNR	0.0	3.5	3.5	0.0	0.0	0.0	0.0	0.0	0.0	1.8	2.4	

Figure 12. Confusion matrix for DNBND with Coiflets wavelet.

Table 3. Overall classification accuracy of ND family classifiers.

Classifier	Features	Total No. of Data	Correctly Categorized Instances		Incorrectly Categorized Instances		Kappa Statistic
ND	Histogram	550	540	98.18%	10	1.82%	0.9758
CBND	Histogram	550	542	98.55%	8	1.45%	0.9818
DNBND	Histogram	550	542	98.55%	8	1.45%	0.9818
ND	Coiflets wavelet	550	543	98.73%	7	1.27%	0.9859
CBND	Coiflets wavelet	550	547	99.45%	3	0.55%	0.9939
DNBND	Coiflets wavelet	550	546	99.27%	4	0.73%	0.9919

7.3. Comprehensive Accuracy by Condition

The specific condition implementation of the ND can be studied using comprehensive accuracy by condition study, termed as TPR, FPR, Precision, and ROC. The TPR corresponds to the accurately identified conditional class, and the FPR is the rate at which the classes are wrongly classified. For a machine learning algorithm, TPR must be 1, and FPR should be 0. This study's TPR and FPR are closer to 1 and 0, respectively. Similarly, the other parameter values are also closer. Table 4 presents the classification accuracy for CBND for wavelet feature extraction.

Table 4. Comprehensive accuracy by conditions—CBND with Coiflets wavelets.

Category	TPR	FPR	Precision	Recall	F-Measure	ROC Area
ABF	1	0	1	1	1	1
BOS	1	0.002	0.982	1	0.991	1
DBM	1	0.004	0.965	1	0.982	1
DBP(UE) inner	1	0	1	1	1	1
DB(UE) IO	0.982	0	1	0.982	0.991	1
DBPI	1	0	1	1	1	1
DBPIO	1	0	1	1	1	1
DBPW	0.982	0	1	0.982	0.991	1
GOOD	1	0	1	1	1	1
RL	0.982	0	1	0.982	0.991	1
Weighted Average	0.995	0.001	0.995	0.995	0.995	1

7.4. Comparative Analysis

The captured raw vibration signals were denoised through WT. The carefully chosen features were categorized using ND, CBND, and DNBND, and the overall accuracy is shown in Table 5. The results of wavelet features were compared with histogram features. Table 5 shows the WT of the vibration signal with CBND, revealing a maximum overall accuracy of 99.45%. The comparative results show that the wavelet features and ND outperform the statistical features and all other classifiers considered in this study. Hence, the wavelets with CBND can be considered an appropriate classifier for diagnosing brake faults.

Table 5. Classification accuracy of ND classifiers.

S. No.	Name of the Classifier	Classification Accuracy (%)	
		Wavelet Features	Histogram Features
1	ND	99.20	97.50
2	CBND	99.45	97.00
3	DNBND	99.30	99.00

8. Conclusions

The paper discussed the wavelet and histogram features for supervising the brake system using vibration signals. The present study considered nine good simulated faults of the brake system. The vibration signals were recorded under all simulated defects. The wavelet and histogram features were extracted from the vibration signal, and the most significant features were selected. Then, the preferred features were classified using ND, CBND, and DNBND. The results reveal that the CBND classifier performed better than the ND and DNBND in classification accuracy. The competency of the wavelet features was evaluated with the histogram features, and the wavelet feature yielded an overall prediction accuracy of 99.45%. Henceforth, the CBND with wavelet features can be recommended for online brake fault diagnostics. This research is protracted with actual road conditions for monitoring the hydraulic brake system.

Author Contributions: Conceptualization, S.G. and J.R.; methodology, J.R.; software, L.P.J.; validation, J.R., M.T. and Y.B.; formal analysis, J.R.; investigation, L.P.J.; resources, J.R.; data curation, J.R.; writing—original draft preparation, J.R.; writing—review and editing, M.T.; visualization, S.G.; supervision, L.P.J.; project administration, J.R. and S.G.; funding acquisition, J.R. All authors have read and agreed to the published version of the manuscript.

Funding: The authors want to acknowledge the research funding from DST—Science and Engineering Research Board (SERB Grant) File No. ECR/2016/002068.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study will be provided upon request and institutional approval.

Acknowledgments: The authors thank the Management of VIT University, School of Mechanical Engineering, VIT Chennai and the Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India, and for their support to publish this work.

Conflicts of Interest: The authors declared no conflict of interest.

References

1. National Highway Traffic Safety Administration. *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*; US Department of Transportation: Washington, DC, USA, 2015; Volume 2, pp. 1–2.
2. NickJohnson. Brake Failure Accidents. 2007. Available online: www.articles3k.com/article/144/176520 (accessed on 1 September 2023).
3. Mohapatra, A.G.; Mohanty, A.; Pradhan, N.R.; Mohanty, S.N.; Gupta, D.; Alharbi, M.; Alkhayyat, A.; Khanna, A. An Industry 4.0 implementation of a condition monitoring system and IoT-enabled predictive maintenance scheme for diesel generators. *Alex. Eng. J.* **2023**, *76*, 525–541. [CrossRef]
4. Molla Salilew, W.; Karim, Z.A.A.; Lemma, T.A. Investigation of fault detection and isolation accuracy of different Machine learning techniques with different data processing methods for gas turbine. *Alex. Eng. J.* **2022**, *61*, 12635–12651. [CrossRef]
5. Goel, S.; Ghosh, R.; Kumar, S.; Akula, A. A methodical review of condition monitoring techniques for electrical equipment. In Proceedings of the National Seminar & Exhibition on Non-Destructive Evaluation (NDE 2014), Pune, India, 4–6 December 2014.
6. Wallace, T.E.; Rink, R.J.; Chandler, M.D. Brake Monitoring System and Method. US Patent No. 8,319,623, 27 November 2012.
7. Kanda, Y.; Terashima, N.; Katou, T. Vehicle Condition Monitoring System. US Patent No. 7,321,814, 22 January 2008.
8. Fiechter, C.-N.; Göker, M.H.; Grill, D.; Kaufmann, R.; Engelhardt, T.; Bertsche, A. Method and System for Condition Monitoring of Vehicles. US Patent No. 6,609,051, 19 August 2003.
9. Haugse, E.D.; Ikegami, R.; Trego, A. Vehicle Condition Monitoring System. US Patent No. 6,691,007, 10 February 2004.
10. Miller, R.J.; Marshall, R.J.; Bailey, D.A.; Griffin, N.C. Brake Condition Monitoring. US Patent No. 7,086,503, 3 August 2001.
11. Praveenkumar, T.; Saimurugan, M.; Ramachandran, K. Intelligent Fault Diagnosis of Synchromesh Gearbox Using Fusion of Vibration and Acoustic Emission Signals for Performance Enhancement. *Int. J. Progn. Health Manag.* **2019**, *10*, 2738. [CrossRef]
12. Yuvaraju, E.; Rudresh, L.; Saimurugan, M. Vibration signals based fault severity estimation of a shaft using machine learning techniques. *Mater. Today Proc.* **2020**, *24*, 241–250. [CrossRef]
13. Kim, Y.; Lee, S.; Seo, H.; Jung, J.; Yang, H. Development of dissolved gas analysis (DGA) expert system using new diagnostic algorithm for oil-immersed transformers. In Proceedings of the 2012 IEEE International Conference on Condition Monitoring and Diagnosis, Bali, Indonesia, 23–27 September 2012; pp. 365–369. [CrossRef]
14. Sreedhar, U.; Krishnamurthy, C.; Balasubramaniam, K.; Raghupathy, V.; Ravisankar, S. Automatic defect identification using thermal image analysis for online weld quality monitoring. *J. Mater. Process. Technol.* **2012**, *212*, 1557–1566. [CrossRef]
15. Sachin Krishnan, P.; Rameshkumar, K.; Krishnakumar, P. Hidden Markov modelling of high-speed milling (HSM) process using acoustic emission (AE) signature for predicting tool conditions. In *Advances in Materials and Manufacturing Engineering*; Springer: Singapore, 2020; pp. 573–580.
16. Elasha, F.; Greaves, M.; Mba, D.; Addali, A. Application of acoustic emission in diagnostic of bearing faults within a helicopter gearbox. *Procedia Cirp* **2015**, *38*, 30–36. [CrossRef]
17. Tan, C.K.; Irving, P.; Mba, D. A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears. *Mech. Syst. Signal Process.* **2007**, *21*, 208–233. [CrossRef]
18. Altaf, M.; Akram, T.; Khan, M.A.; Iqbal, M.; Ch, M.M.I.; Hsu, C.-H. A New Statistical Features Based Approach for Bearing Fault Diagnosis Using Vibration Signals. *Sensors* **2022**, *22*, 2012. [CrossRef] [PubMed]
19. Gnanasekaran, S.; Jakkamputi, L.; Thangamuthu, M.; Marikkannan, S.K.; Rakkiyannan, J.; Thangavelu, K.; Kotha, G. Condition Monitoring of an All-Terrain Vehicle Gear Train Assembly Using Deep Learning Algorithms with Vibration Signals. *Appl. Sci.* **2022**, *12*, 10917. [CrossRef]
20. Mohanraj, T.; Yerchuru, J.; Krishnan, H.; Aravind, R.N.; Yameni, R. Development of tool condition monitoring system in end milling process using wavelet features and Hoelder’s exponent with machine learning algorithms. *Measurement* **2021**, *173*, 108671. [CrossRef]
21. Steffan, J.J.; Jebadurai, I.J.; Asirvatham, L.G.; Manova, S.; Larkins, J.P. Prediction of Brake Pad Wear Using Various Machine Learning Algorithms. In *Recent Trends in Design, Materials and Manufacturing*; Springer: Singapore, 2022; pp. 529–543.
22. Liu, J.; Li, Y.-F.; Zio, E. A SVM framework for fault detection of the braking system in a high speed train. *Mech. Syst. Signal Process.* **2017**, *87*, 401–409. [CrossRef]
23. Joshuva, A.; Anaimuthu, S.; Selvaraju, N.; Muthiya, S.J.; Subramaniam, M. A Machine Learning Approach for Vibration Signal Based Fault Classification on Hydraulic Braking System through C4. 5 Decision Tree Classifier and Logistic Model Tree Classifier; SAE Technical Paper; SAE International: Warrendale, PA, USA, 2020. [CrossRef]

24. Satapathy, S.; Loganathan, D.; Kondaveeti, H.K.; Rath, R. Performance analysis of machine learning algorithms on automated sleep staging feature sets. *CAAI Trans. Intell. Technol.* **2021**, *6*, 155–174. [CrossRef]
25. Mazzoleni, M.; Sarda, K.; Acernese, A.; Russo, L.; Manfredi, L.; Glielmo, L.; Del Vecchio, C. A fuzzy logic-based approach for fault diagnosis and condition monitoring of industry 4.0 manufacturing processes. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105317. [CrossRef]
26. Bajaj, N.S.; Patange, A.D.; Jegadeeshwaran, R.; Kulkarni, K.A.; Ghatpande, R.S.; Kapadnis, A.M. A Bayesian optimized discriminant analysis model for condition monitoring of face milling cutter using vibration datasets. *J. Nondestruct. Eval. Diagn. Progn. Eng. Syst.* **2022**, *5*, 021002. [CrossRef]
27. Vos, K.; Peng, Z.; Jenkins, C.; Shahriar, M.R.; Borghesani, P.; Wang, W. Vibration-based anomaly detection using LSTM/SVM approaches. *Mech. Syst. Signal Process.* **2022**, *169*, 108752. [CrossRef]
28. Chao, Q.; Wei, X.; Tao, J.; Liu, C.; Wang, Y. Cavitation recognition of axial piston pumps in noisy environment based on Grad-CAM visualization technique. *CAAI Trans. Intell. Technol.* **2023**, *8*, 206–218. [CrossRef]
29. Jia, Y.; Xu, M.; Wang, R. Symbolic important point perceptually and hidden Markov model based hydraulic pump fault diagnosis method. *Sensors* **2018**, *18*, 4460. [CrossRef] [PubMed]
30. Wang, Y.; Yan, Y.; Wang, Q. Wavelet-based feature extraction in fault diagnosis for biquad high-pass filter circuit. *Math. Probl. Eng.* **2016**, *2016*, 5682847. [CrossRef]
31. Yao, Y.; Ma, J.; Ye, Y. Regularizing Autoencoders with Wavelet Transform for Sequence Anomaly Detection. *Pattern Recognit.* **2022**, *134*, 109084. [CrossRef]
32. Leathart, T.; Frank, E.; Pfahringer, B.; Holmes, G. Ensembles of nested dichotomies with multiple subset evaluation. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Macau, China, 14–17 April 2019.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Regional Thermal Radiation Characteristics of FY Satellite Remote Sensing Based on Big Data Analysis

Tao Wen ¹, Congxin Wei ^{2,*}, Zhenyi Wang ², Linzhu Wang ¹, Zihan Yang ¹ and Tingting Gu ³¹ Gansu Computing Center, Lanzhou 730030, China; went@cc.gs.cn (T.W.); wanglz@cc.gs.cn (L.W.)² School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China³ Gansu Zijin Cloud Big Data Development Co., Ltd., Jinchang 737100, China; gutt@gepic.cn

* Correspondence: weicx@lut.edu.cn

Abstract: It is of great significance to study the thermal radiation anomalies of earthquake swarms in the same area in terms of selecting abnormal characteristic determination parameters, optimizing and determining the processing model, and understanding the abnormal machine. In this paper, we investigated short-term and long-term thermal radiation anomalies induced by earthquake swarms in Iran and Pakistan between 2007 and 2016. The anomalies were extracted from infrared remote sensing black body temperature data from the China Geostationary Meteorological Satellites (FY-2C/2E/2F/2G) using the multiscale time-frequency relative power spectrum (MS T-FRPS) method. By analyzing and summarizing the thermal radiation anomalies of series earthquake groups with consistency law through a stable and reliable MS T-FRPS method, we first obtained the relationship between anomalies and ShakeMaps from USGS and proposed the anomaly regional indicator (ARI) to determine seismic anomalies and the magnitude decision factor (MDF) to determine seismic magnitude. In addition, we explored the following discussions: earthquake impact on regional thermal radiation background and the relationship between thermal anomalies and earthquake magnitude and the like. Future research directions using the MS T-FRPS method to characterize regional thermal radiation anomalies induced by strong earthquakes could help improve the accuracy of earthquake magnitude determination.

Keywords: satellite thermal radiation remote sensing; multiscale time-frequency relative power spectrum analysis; thermal radiation background; thermal anomaly

1. Introduction

The application of satellite remote sensing thermal radiation data to study earthquake anomalies and mechanisms is one of the research directions of geophysics, which has obtained many research results. One of the earliest studies on the relationship between satellite radiation remote sensing and earthquakes dates back to the 1980s [1,2]. With the development of satellite remote sensing technology, many efforts have been devoted to studying the earthquake thermal radiation characteristics. It has been reported that significant thermal anomalies can be identified prior to earthquakes in seismically active areas [2–11]. Such examples include the following: Bhuj (Gujarat), Mw = 7.7, 26 January 2001, India [6,11]; Colima, Mw = 7.8, 21 January 2003, Mexico [11]; Kashmir, Mw = 7.6, 8 October 2005, [3]; Bhuj (Gujarat), Mw = 7.7, 26 January 2001, India [6,10]; Colima, Mw = 7.8, 21 January 2003, Mexico [11]; Yutian, Mw = 7.3, 21 March 2008, China [12]; Wenchuan, Mw = 8.0, 12 May 2008, China [13]; the Andaman Islands, Mw = 7.5, 11 August 2009 [14]; Honshu, Mw = 9.1, 11 March 2011, Japan [15]; and many others.

However, in seismic thermal radiation research, anomaly extraction is difficult because of the weak signals and complicated background noise. First of all, satellite remote sensing images were used in structural geological and geomorphologic research. Active faults and structures were mapped on the base of satellite images. This method is very limited in time

series analysis. There was no possibility to measure short-term processes before and after the earthquake. It was as simple as an extrapolation of air photo geological interpretation methods applied to space. The modern version of this method is active tectonic analysis with the application of alignment analysis. Case studies of various thermal remote sensing application methods for earthquakes have been recently reported. The use of MISR (Multi-angle Imaging Spectra Radiometer) data revealed a significant increase in surface moisture after the Bhuj (Gujarat) earthquake on 26 January 2001. This indicates that changes in surface moisture after an earthquake can be detected through thermal remote sensing technology [16]. The discovery that anomalous areas are consistent with seismic zones or active tectonics, with epicenters often located at the intersections of stress tropics, suggests a certain correlation between seismic activity and thermal remote sensing anomalies, possibly related to tectonic activity [17–21]. The analysis of satellite thermal infrared brightness temperature anomaly images reveals that short-term anomalies exhibit clearly identifiable radiation characteristics. This suggests that by analyzing radiation anomalies at short-term time scales, specific feature information can be obtained [22]. The study of the temporal-spatial variations of associated faulting inferred from satellite infrared information reveals a close correlation between the Mani earthquake radiation anomaly and the activity of the Altyn Tagh fault, illustrating the interactions between faults and deflection among the deformation exception, current fault activities, and earthquakes [23]. Secondly, time series of alignment distributions on the Earth's surface are investigated before and after an earthquake. Tramutoli and colleagues have developed a robust satellite data analysis technique for space–time thermal anomalies on the Earth's surface recorded by satellites months to weeks before the occurrence of earthquakes. Lastly, since 2000, Zhang et al. have researched anomaly extraction using case studies and have investigated underlying mechanisms [24]. They developed the time-frequency relative power spectrum (T-FRPS) method based on radiation brightness temperature remote sensing data from the China Geostationary Meteorological Satellites (FY-2C/2E/2F/2G) and proposed the 'union origin mechanism' to explain pre-earthquake thermal anomalies [25]. The thermal radiation anomaly characteristics, including characteristic period, amplitude, duration, and morphology, were clear and valuable for pre-earthquake thermal radiation research. The epicenters were located near the intersection of the anomaly migration processes, which was consistent with the remote sensing rock test mechanics [26,27].

Researchers have achieved certain results in the extraction methods of seismic anomaly information (from viewing pictures to time series and time-frequency analysis) and earthquake case studies. However, there are few studies in some directions, such as the characteristics of thermal radiation anomalies and the characteristics of swarms of earthquakes over many years; there are many studies on short-term and anomalous thermal radiation, but there are few studies on mid- and long-term background anomalies, and thermal radiation background characteristics, as well as the characteristics of thermal radiation backgrounds and the impacts of strong earthquakes on thermal radiation background field.

Therefore, this article details the multiscale time-frequency relative power spectrum (MS T-FRPS), it initially scrutinizes the results of Wavelet filtering results, and conducts an annual background anomaly analysis of Mw 7+ earthquakes. Additionally, it extracts short- and long-term thermal radiation anomalies of several strong earthquake swarms in Iran and Pakistan. Furthermore, it compares and analyzes the similarities and differences between short-term and long-term background abnormalities for the first time. The impacts of strong earthquakes on thermal radiation background field are also discussed for the first time.

2. Methodology

2.1. Multiscale Time-Frequency Relative Power Spectrum (MS T-FRPS) Analysis

Time-frequency analysis (TFA) is a powerful tool for analyzing time-varying non-stationary signals and has received increasing attention in recent years. TFA provides joint distribution information of the time domain and the frequency domain, which clearly

describes the signal frequency over time. The basic idea of TFA is to design a joint function of time and frequency that simultaneously describes the energy density or intensity of the signal at different times and frequencies. This joint function of time and frequency is referred to simply as the time-frequency distribution, by which we can draw the instantaneous frequency and amplitude at each moment and conduct time-frequency filtering and time-varying analysis of the signal.

Firstly, Daubechies wavelet transform (2nd-order scale and 7th-order scale filtering) is used to process the multi-year data of satellite thermal radiation daily values to separate the temperature changes caused by the basic temperature field, the annual temperature field, and other factors. In addition, the 6th- and 7th- order scales are respectively decomposed to obtain the annual and approximate annual background information. Secondly, the Welch's FFT power spectrum is applied to process the above wavelet filtering and decomposition results, in order to obtain the dominant frequency and amplitude in the time domain. Therefore, the method of successive wavelet processing and power spectrum processing is called time-frequency power spectrum analysis. Thirdly, in order to compare and analyze the power spectrum of the thermal radiation change before and after the earthquake, the power spectrum values of all the frequencies of each pixel are relatively processed to generate spatial-temporal data of the relative time-frequency power spectrum. The frequency analysis method through wavelet transform at different scales (6th-order scale, 7th-order scale, and 2nd- and 7th-order bandpass filtering) and relative Fourier power spectrum calculation for different window lengths (64 days, 128 days, 256 days, and 512 days) is called multiscale time-frequency relative power spectrum (MS T-FRPS).

2.1.1. Wavelet and Fourier Transform

Wavelet theory is a relatively new digital signal processing technology that originated a few decades ago. Due to its high resolution in both the time and frequency domains, it is called the 'mathematical microscope', which has been widely used and is developing rapidly in the field of digital signal processing. All wavelet transforms may be considered as forms of time-frequency representation for continuous-time signals and thus are related to harmonic analysis. Almost all practically useful discrete wavelet transforms use discrete-time filter banks, which are called wavelet and scaling coefficients in the wavelet nomenclature. The most commonly used set of discrete wavelet transforms was formulated by Ingrid Daubechies in 1988 [28]. The Daubechies wavelets are a family of orthogonal wavelets, with D2-D20 and db1-db10 commonly used. In this paper, the db8 wavelet basis function of the Daubechies (dbN) wavelet system is used as the basic transform window function to process daily value brightness temperature data.

Spectral analysis is an important method for studying signal characteristics. For deterministic signals, Fourier transform (FT) can be used to examine its spectral properties. Any signal that can be represented by amplitude that varies in time has a corresponding frequency spectrum, which contains all the information of the original signal. This means that the original function can be completely reconstructed by an inverse FT. In classical spectral estimation using a periodical diagram, the variance of spectral estimate at a specific frequency does not decrease with the increasing number of samples. Although it is successful in examining the frequency characteristics of noise-free functions, it cannot process noise-like signals, e.g., sinusoids with low signal-to-noise ratios. To mitigate this issue, Welch modified the periodic diagram method by taking average over time, which becomes popular for estimating the power of a signal at different frequencies. Due to its ease of understanding and calculation, it has been widely used and adopted in our analysis.

2.1.2. Relative Processing

In order to obtain the relative time-frequency power spectrum of each pixel, we calculate it by Equation (1),

$$R_{PS} = \frac{R_{i,j}}{R_{i,Mean}} \quad (1)$$

where R_{PS} is the calculated relative power spectrum value, $R_{i,j}$ is the power spectrum value of a certain pixel at any time, and $R_{i,Mean}$ is the ten-year average of the corresponding pixel.

2.2. Data Acquisition

Geostationary meteorological satellites are preferable to polar orbit satellites for earthquake research as they consistently observe the same location. We used radiation brightness temperature remote sensing data acquired from the China Geostationary Meteorological Satellites FY-2C/2E/2F/2G, which were launched during 2004–2014. The scan period was 30 min or 1 h, and the spatial resolution was 5 km. The data were collected from 1 January 2007 to 31 December 2016. Five midnight observations were selected from each day because of the lower solar radiation at night.

2.3. Data Processing

In this paper, the wavelet transform basis function and the relative FT are designed as a joint function of time and frequency, which describes the energy density or intensity of the signal at different times and frequencies. The TFA function was used to analyze the similarities and differences between time-frequency analyses at each scale.

Data processing and extraction were conducted in four steps, as shown in Figure 1:

1. The original database was formed after pre-processing the radiation remote sensing black body temperature data from the Chinese geostationary meteorological satellite. This process included data format conversion, data space range interception, removal of clouds, and selection of nighttime data for daily value calculations. The influence of partial cloud cover was removed by making up the window. Then, we calculated the mean value of five data points to obtain the daily value.
2. Various scale-factor transformation results were obtained by applying discrete Daubechies wavelet transformations. And the dominant frequency and the peak-to-peak amplitude values were obtained using the FFT power spectral method.
3. The decade (2007–2016) average value of each pixel was used as the denominator to obtain the relative time-frequency PS value of the corresponding pixel.
4. Time-frequency mapping was used to extract earthquake thermal radiation anomaly data. The non-seismic factors were effectively removed and earthquake thermal radiation anomalies were highlighted by scanning the time-frequency space data over the full spatial-temporal and band range.

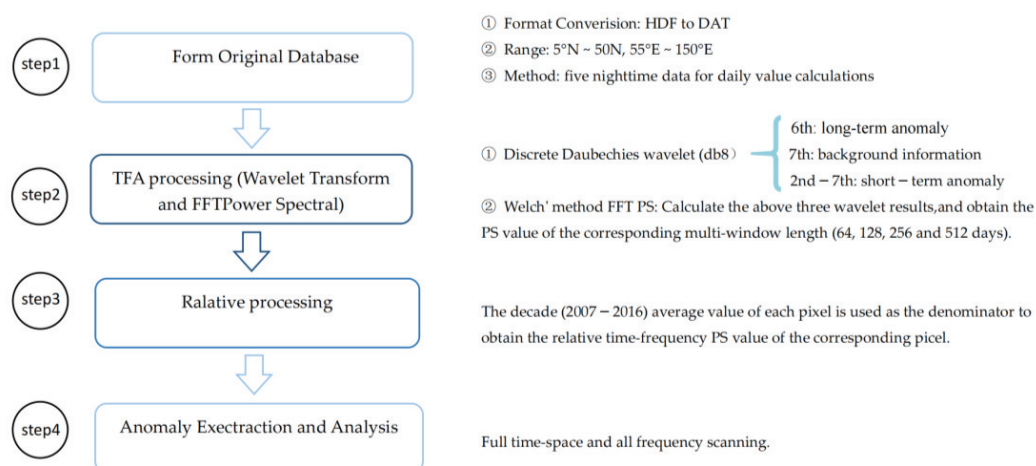


Figure 1. Data processing flowchart.

3. Study Area

3.1. Regional Tectonic Background

The tectonics of this investigated region is dominated by the collisions of the Arabian and Indian plates with Eurasia. The Arabian plate lithosphere subducts beneath the Eurasian plate at the Makran coast of Pakistan and Iran, and extends progressively deeper to the north [29]. Four major tectonic plates (Arabia, Eurasia, India, and Africa) and one smaller tectonic block (Anatolia) are responsible for seismicity in the Middle East and surrounding region. The geologic evolution of this region is a consequence of several first-order plates' tectonic processes including seductions, large-scale transform faulting, compression mountain building, and crustal extension. Continental thickening of the northern and western edge of the India subcontinent has produced the highest mountains in the world, including the Himalayan, Karakoram, Pamir, and Hindu Kush ranges. Earthquake activity and faulting found in this region as well as adjacent parts of Afghanistan and India are due to collisional plate tectonics. Shallower crustal earthquakes in the Pamir-Hindu Mountains occur primarily along the Main Pamir Thrust and other active Quaternary faults, which accommodate much of the region's crustal shortening. From 2007 to 2016, 4 earthquakes with Mw 6.0+ occurred in this region; the ShakeMaps of the main shocks are shown in Figure 2 and the source is the following website: (a) <https://earthquake.usgs.gov/earthquakes/eventpage/usp000gma2#shakemap>, (accessed on 1 July 2023); (b) <https://earthquake.usgs.gov/earthquakes/eventpage/usp000htb4#shakemap>, (accessed on 1 July 2023); (c) <https://earthquake.usgs.gov/earthquakes/eventpage/usb000g7x7#shakemap>, (accessed on 1 July 2023); (d) <https://earthquake.usgs.gov/earthquakes/eventpage/usb000jyiv#shakemap>, (accessed on 1 July 2023). In Figure 2, the star marks the epicenter, and the lines around the star represent the rupture direction of the earthquake.

3.2. Regional Earthquake Activity History

This region had strong seismic activity from 1900 to 2016, including 130 earthquakes (100 shallow earthquakes) of magnitude 6+, as shown in Figure 3, and the source is the following website: <https://earthquake.usgs.gov/earthquakes/search/>, accessed on 10 July 2023, Date: 1 January 1990–31 December 2016. The number of earthquakes decreased as magnitude increased. Previous studies have found that the thermal radiation anomalies of shallow earthquakes are more obvious and easier to extract. Therefore, we focused on analyzing shallow-source seismic activity.

Since 1900, the frequency of earthquakes per decade has gradually risen, peaking between 1997 and 2006, as shown in Figure 4, and the source is the following website: <https://earthquake.usgs.gov/earthquakes/search/>, accessed on 10 July 2023, Date: 1 January 1990–31 December 2016. Earthquakes occurred less frequently before 1947, which may be associated with poor seismic monitoring capacity and record keeping. Since 1947, the number of shallow earthquakes per decade has continuously increased. The last decade generally reflects the region's average seismic activity; therefore, its thermal radiation background information should be representative as well.

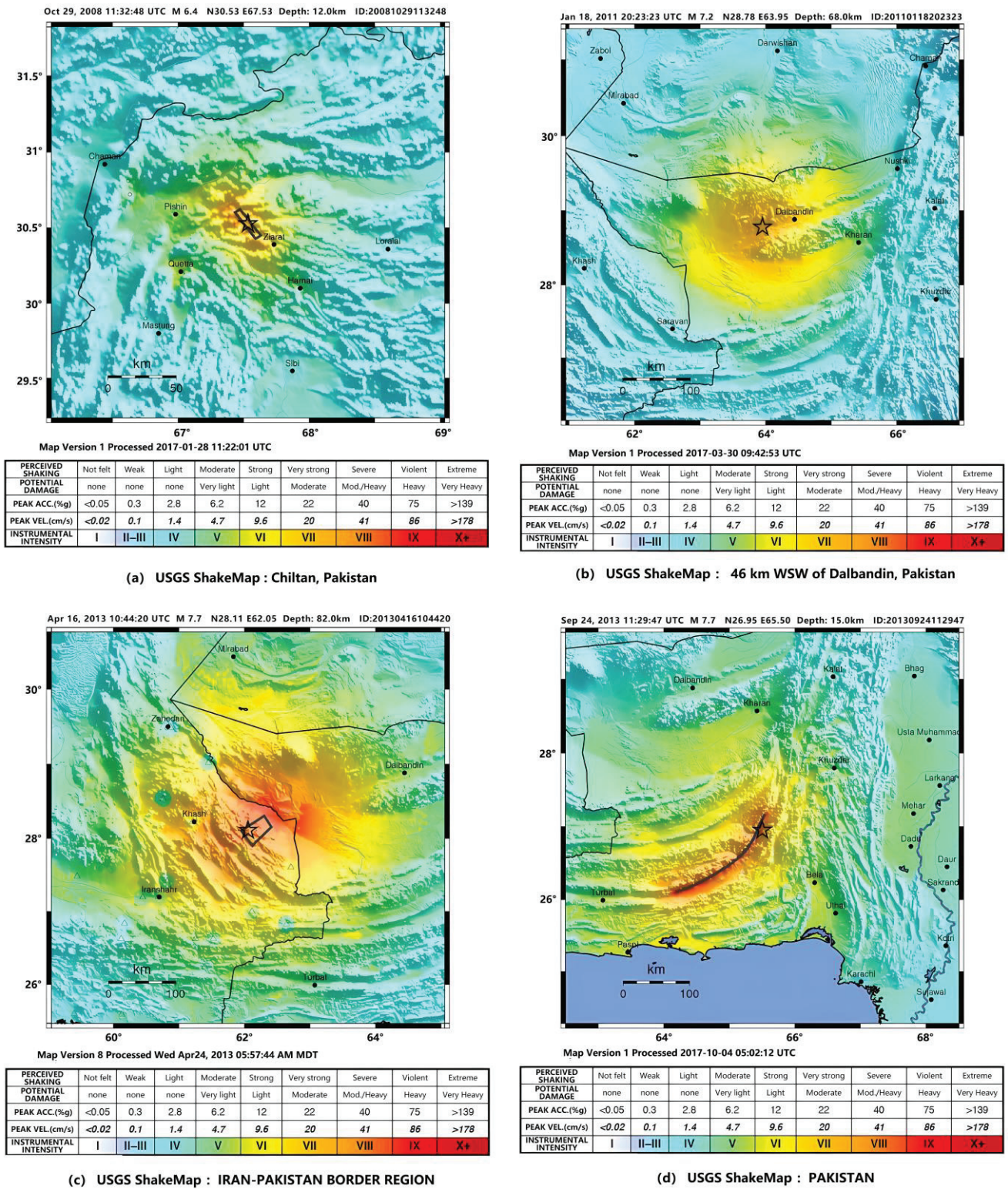


Figure 2. ShakeMap of four earthquakes. Source: United States Geological Survey.

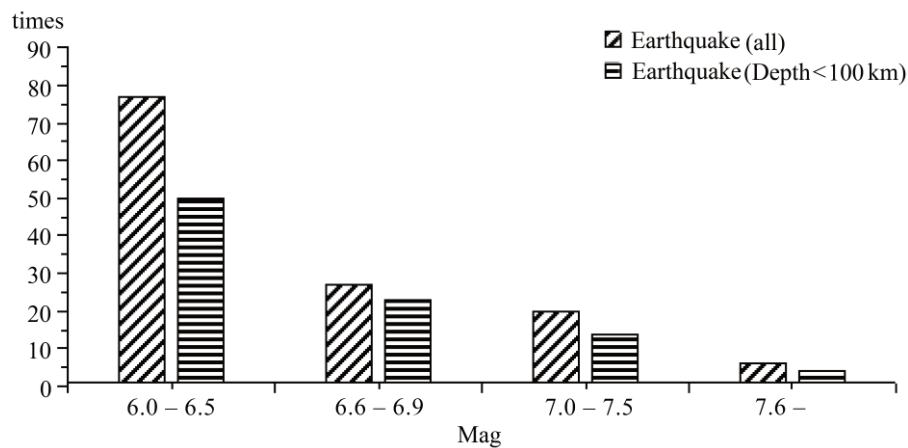


Figure 3. Regional MAG time map from 1900 to 2016. Region: 22–38° N, 55–71° E, earthquake magnitudes: Mw 6+; source: United States Geological Survey.

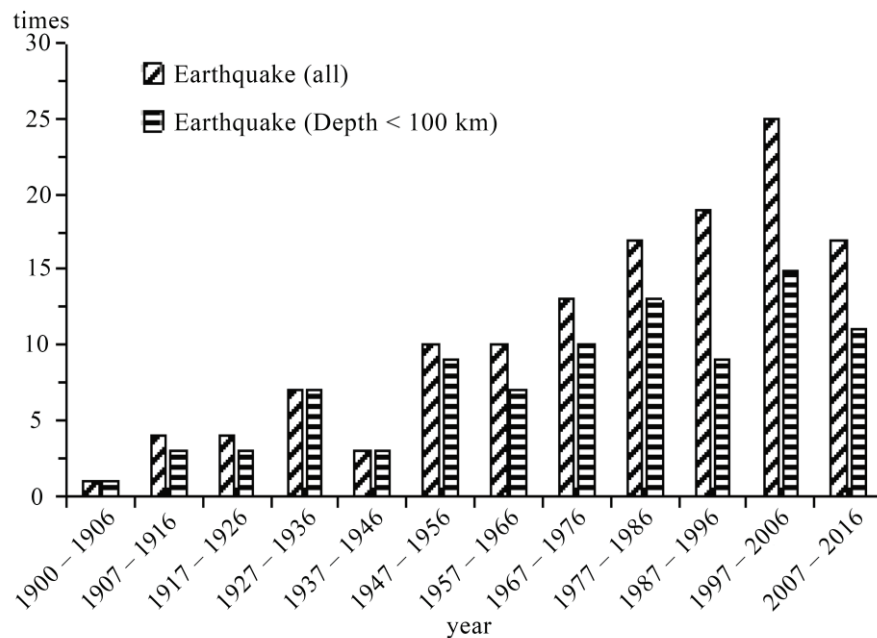


Figure 4. Regional decade interval earthquake frequency chart from 1900 to 2016. Region: 22–38° N, 55–71° E; earthquake magnitude: Mw 6+; source: United States Geological Survey.

3.3. Earthquake Catalogue and Grouping

Table 1 lists 10 shallow earthquakes that occurred over the last decade, which can be divided into four chronological swarms. The first earthquake group has three earthquakes in 2008, including an Mw 6.1 single earthquake and an Mw 6.4 double shock. The second group has three earthquakes in 2010 and 2011, including an Mw 7.2 single earthquake and an Mw 6.7 main earthquake in 2010 and its Mw 6.2 aftershock in 2011. The third group has two earthquakes of Mw 7.7 and Mw 6.1 in the first half of 2013. The last earthquake group has two earthquakes in the second half of 2013, including an Mw 7.7 main earthquake and its Mw 6.8 aftershock.

Table 1. Earthquake catalogue and grouping.

Date	Latitude (°)	Longitude (°)	Depth (km)	Magnitude and Type	Place	Group
10 September 2008	26.74	55.83	12	6.1 mwc	Southern Iran	G1
28 October 2008	30.64	67.35	15	6.4 mwc	Pakistan	
29 October 2008	30.60	67.46	14	6.4 mwc	Pakistan	
20 December 2010	28.41	59.18	12	6.7 mwc	Southeastern Iran	G2
18 January 2011	28.78	63.95	68	7.2 mww	Southwestern Pakistan	
27 January 2011	28.20	59.06	10	6.2 mww	Southeastern Iran	
16 April 2013	28.03	62.00	80	7.7 mww	83 km E of Khash, Iran	G3
11 May 2013	26.56	57.77	15	6.1 mww	94 km SE of Minab, Iran	
24 September 2013	26.95	65.50	15	7.7 mww	61 km NNE of Awaran, Pakistan	G4
28 September 2013	27.18	65.51	12	6.8 mww	85 km NNE of Awaran, Pakistan	

4. Results

In this section, the wavelet filtering results and wavelet annual background anomaly analysis of Mw 7+ earthquake are first analyzed. Secondly, the short-term seismic thermal radiation anomalies of the four earthquake swarms are analyzed. Then, the long-term background anomaly characteristics of regional thermal radiation are examined. Finally, the effects of strong earthquakes on the regional thermal radiation environment are investigated.

4.1. Wavelet Filtering Results and Wavelet Annual Background Anomaly Analysis of Mw 7+ Earthquake

The positive–negative time domain waveform processed by the second- to seventh-order band-pass filtering can be separated from the low-frequency annual temperature field and the high-frequency temperature field caused by other short time factors (cloud, short time heating up or cooling down), as shown in Figure 5.

In Figure 5a, the raw data exhibit distinguishable annual variations caused by season alternation. Due to the short-term weather changes, the fluctuations of the data are relatively large, which can be handled by wavelet analysis. By applying the sixth and seventh order wavelet transform, it is possible to separate the approximate annual period (sixth scale) and annual period (seventh scale) change information, respectively. The seventh scale has a longer period, which is closer to one year. Figure 5b,c show that Mw 7+ can cause changes in the annual period (seventh scale) and approximate annual period (6th scale) background curves. Also, three earthquakes occurred in the background transition.

In conclusion, the second- to seventh-order band-pass Daubechies wavelet filtering is used to reduce noise of the low-frequency annual temperature field and the high-frequency temperature field caused by other short time factors. The sixth- and seventh-order scales indicate the annual background anomaly and correspond to the Mw 7+ earthquake time.

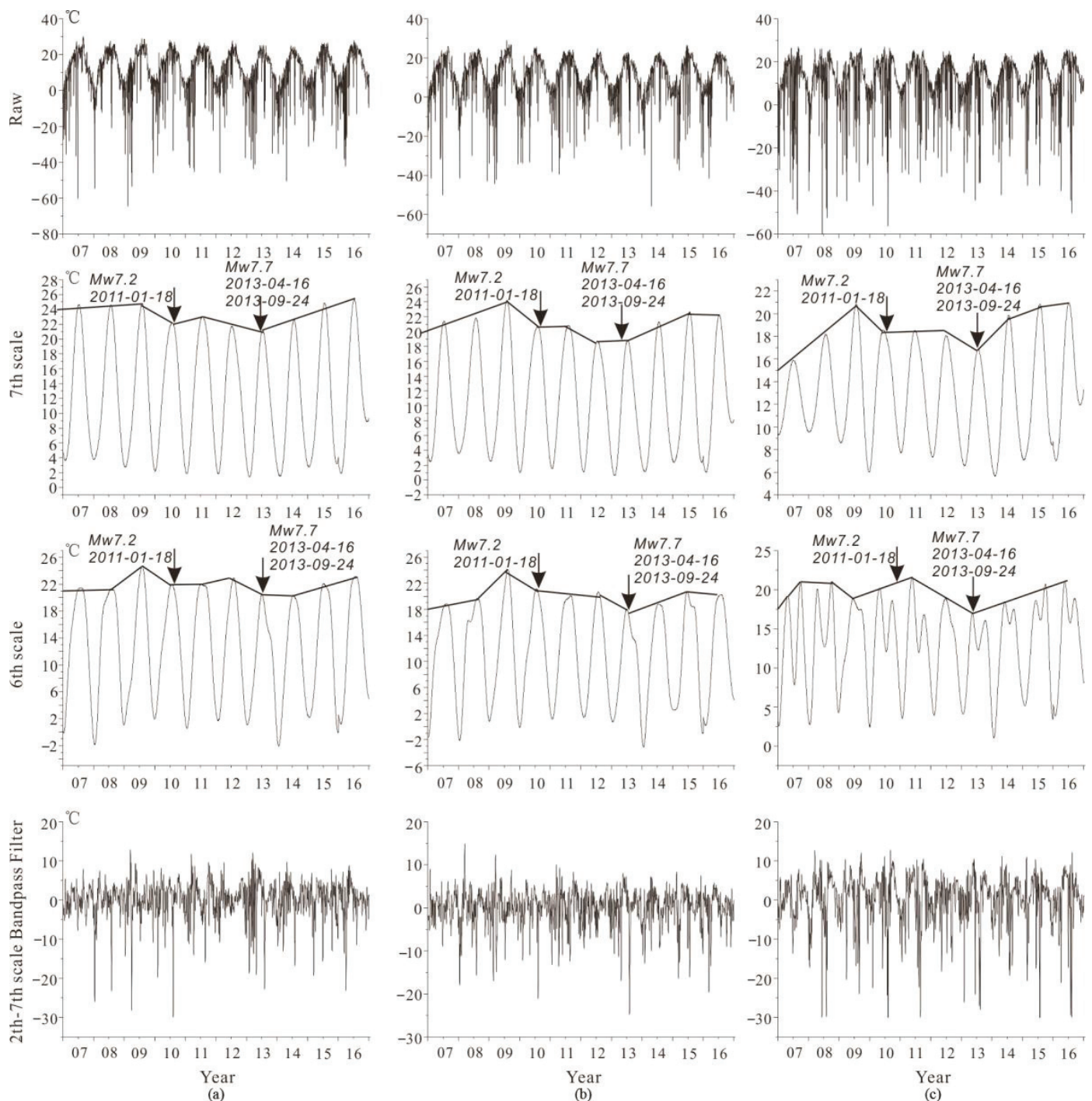


Figure 5. Different scale factor and band-pass filtering wavelet transform results for three Mw 7+ earthquakes. (a) Earthquake 1: 28.78° N, 63.95° E, Mw 7.2, 18 January 2011; (b) Earthquake 2: 28.03° N, 62.00° E, Mw 7.7, 16 April 2013; (c) Earthquake 3: 26.95° N, 65.50° E, Mw 7.7, 24 September 2013. Raw: raw data per day; 6th: sixth order; 7th: seventh order; 2nd–7th: second to seventh band-pass filter.

4.2. TFA of Earthquake Swarm Short-Term Thermal Radiation Anomalies

The TFRPS method was used to extract and identify the characteristic information of earthquake swarm thermal radiation anomalies. The analysis included the anomaly characteristic period, amplitude, duration, form, and the relationship among the anomaly range and earthquake epicenter and evolution characteristics. In the following section,

the TFAs of the thermal radiation anomalies of the four earthquake swarms (G1–G4) are described in chronological order.

1. Short-term thermal radiation anomalies of G1

As shown in Figure 6, the anomaly lasted about 70 days from 21 July 2008 to 29 September 2008. The maximum amplitude of the anomaly occurred on 5 August 2008, reaching more than 10 times the average value. The area of the anomaly was close to the double-shock epicenter, with the expansion direction nearly parallel to the line connection of these two epicenters (E(a), E(b)). The temporal and spatial evolution of the anomaly was clear and easy to extract and identify. The anomaly evolution presents a three-stage process of the abnormal occurrence–maximum–disappearance, in line with the earthquake propagation process.

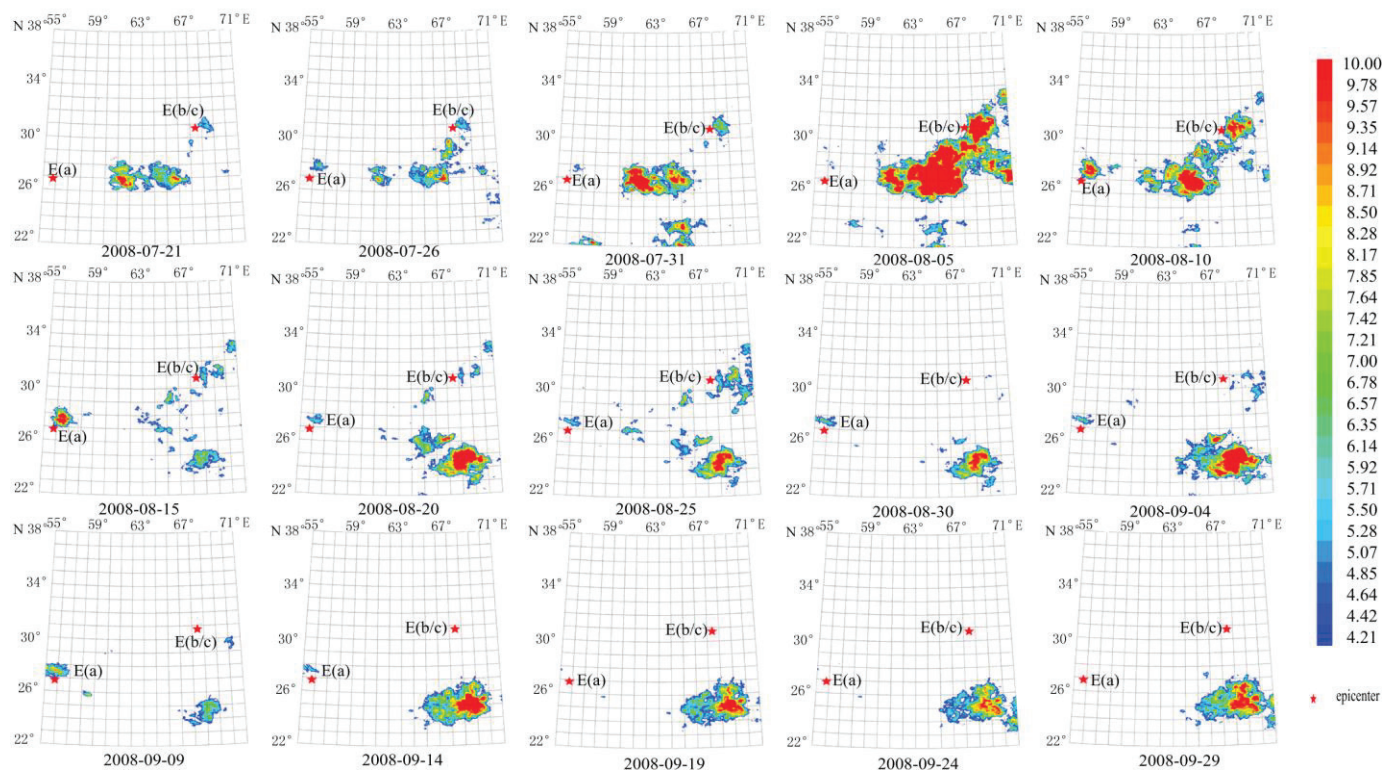


Figure 6. Temporal and spatial evolution of thermal radiation anomalies of earthquake group G1. Single earthquake: E(a) 26.74° N, 55.83° E, Mw 6.1, 10 September 2008. Double earthquake: E(b and c) 30.64° N, 67.35° E and 30.60° N, 67.40° E, Mw 6.4 and Mw 6.4 on 28 October 2008 and 29, respectively. Wavelet transform: second- to seventh-order band-pass filter using Daubechies wavelets. Window length of FT power spectrum: 64 days using Welch's method. Characteristic period: 13 days.

2. Short-term thermal radiation anomalies of G2

The anomaly duration of these earthquakes was also about 70 days from 20 August 2010 to 29 September 2010, as shown in Figure 7. The maximum amplitude of the anomaly reached more than 10 times the average value on 14 September 2010. The anomaly center was close to the Mw 7.2 epicenter. The anomalous shape is consistent with the ShakeMap of Figure 2b, which is mainly located in the southern epicenter.

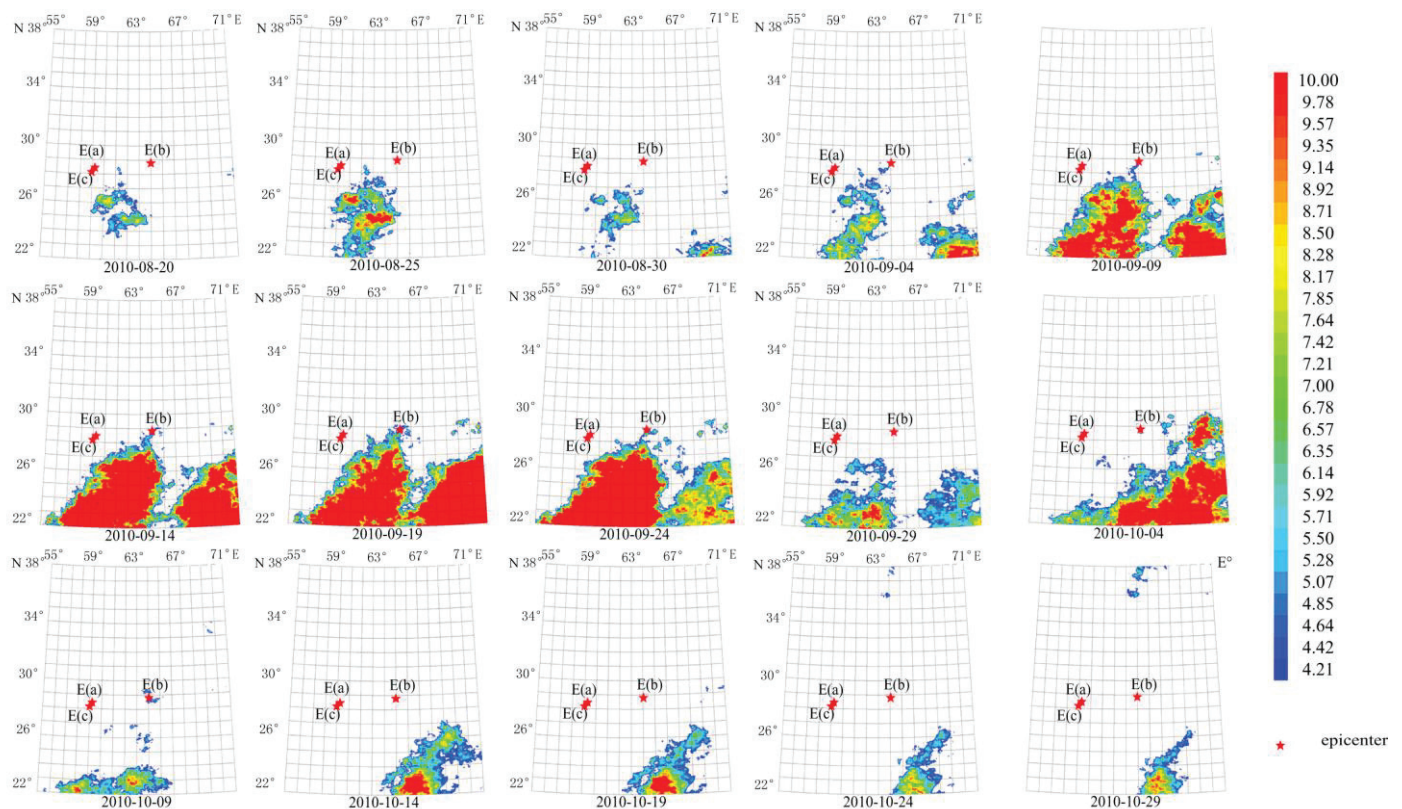


Figure 7. Temporal and spatial evolution of thermal radiation anomalies of earthquake group G2. Single earthquake: E(b) 28.78° N, 63.95° E, Mw 7.2, 18 January 2011. Main shock: E(a) 28.41° N, 59.18° E, Mw 6.7, 20 December 2010. Aftershock: E(c) 28.20° N, 59.02° E, Mw 6.2, 27 January 2011. Wavelet transform: second- to seventh-order scale band-pass filter using Daubechies wavelets. Window length of FT power spectrum: 64 days using Welch's method. Characteristic period: 13 days.

Using a longer characteristic period of 64 days, temporal and spatial evolution of thermal radiation anomalies of earthquakes in group G2 demonstrates a pattern that lasted for about 60 days, as shown in Figure 8. The maximum amplitude of the anomaly occurred on 14 September 2010. The circular anomaly appears to the southeast of the epicenter and agrees with the ShakeMap in Figure 2b.

3. Short-term thermal radiation anomalies of G3

The anomaly, lasting about 60 days from 7 March 2013 to 1 May 2013, is elliptical with the major axis extending diagonally. The anomaly center lies on top of the E(a) epicenter, and its maximum amplitude occurred from 27 March 2013 to 16 April 2013, rising to about 10 times the average value. The area of the anomaly overlapped with the Mw 7.7 epicenter.

4. Short-term thermal radiation anomalies of G4

The anomaly slightly extends from the epicenter to the southwest direction in an elliptical shape, which is consistent with Figure 2d. Its width along the south–north direction is shorter than the anomaly in Figure 9. The duration of this anomaly was about 55 days from 9 September 2013 to 3 October 2013. The maximum amplitude of the anomaly appeared during 14 August 2013 to 24 August 2013, reaching more than 10 times the average value.

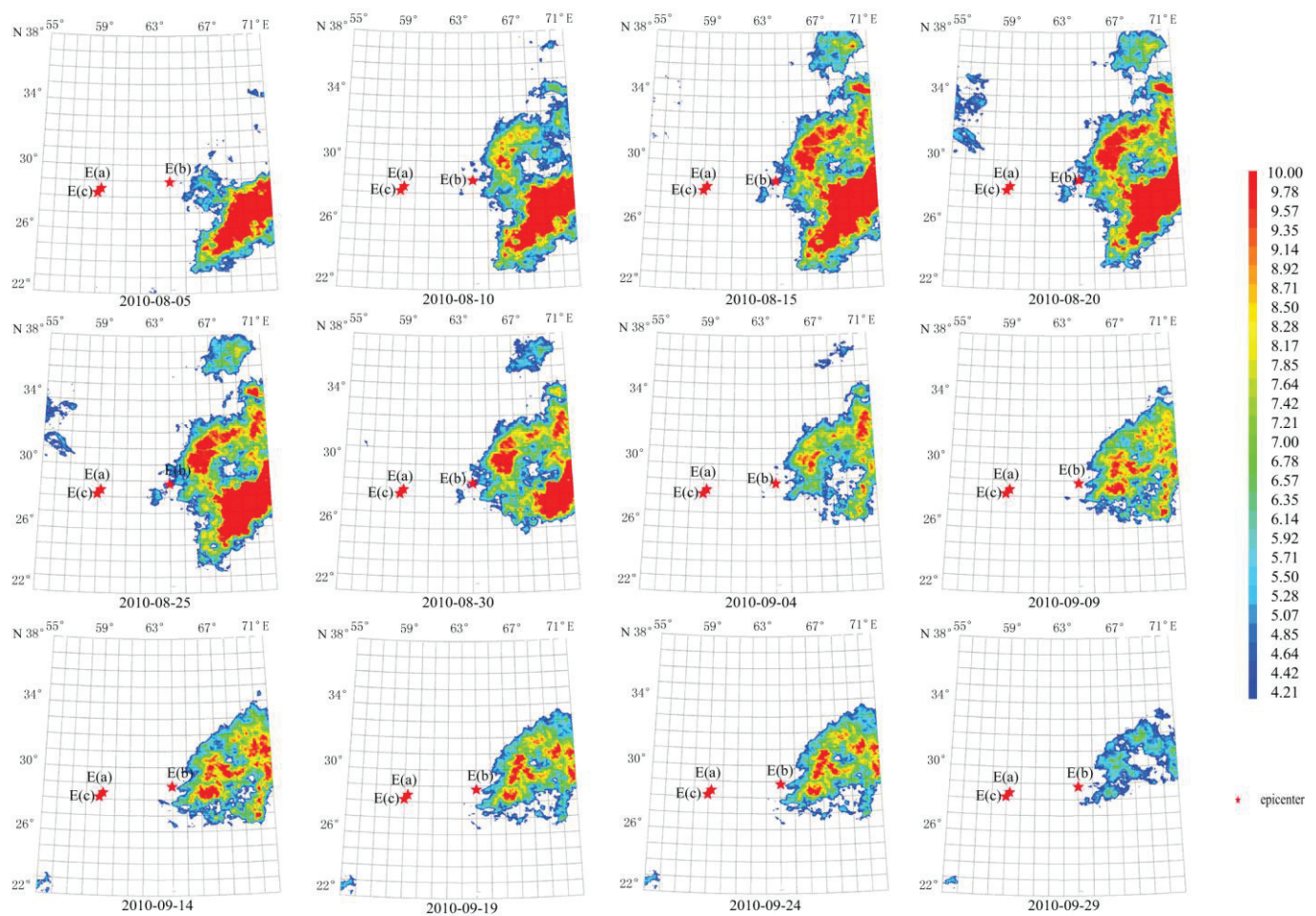


Figure 8. Temporal and spatial evolution of thermal radiation anomalies of earthquake group G2. Single earthquake: E(b) 28.78° N, 63.95° E, Mw 7.2, 18 January 2011. Main shock: E(a) 28.41° N, 59.18° E, Mw 6.7, 20 December 2010. Aftershock: E(c) 28.20° N, 59.02° E, Mw 6.2, 27 January 2011. Wavelet transform: second- to seventh-order scale band-pass filter using Daubechies wavelets. Window length of FT power spectrum: 64 days using Welch's method. Characteristic period: 64 days.

Compared to Figure 10, the anomaly in Figure 9, with a short characteristic period of 13 days, appeared 40 days pre-earthquake, reached the maximum when the earthquake occurred, and disappeared in 15 days after the earthquake. In addition, the anomaly area is much larger with the epicenter in the center, while in Figure 10, the anomaly appeared 45 days before the earthquake, and already started waning at the earthquake occurrence, and completely disappeared in the 10 days post-earthquake. The anomaly is more concentrated with the epicenter lying at its right edge.

As summarized in Table 2, short-term thermal radiation anomalies are irrelevant to the earthquake magnitude in terms of characteristic period, duration, and earthquake time and epicenter location. All anomalies lasted about 2 months and the earthquake occurred after they reached the maximum. Anomalies either overlap with or extend away from the epicenters. A positive relationship is identified between the anomaly morphology and the ShakeMap. That is, the higher the magnitude, the better they match each other.

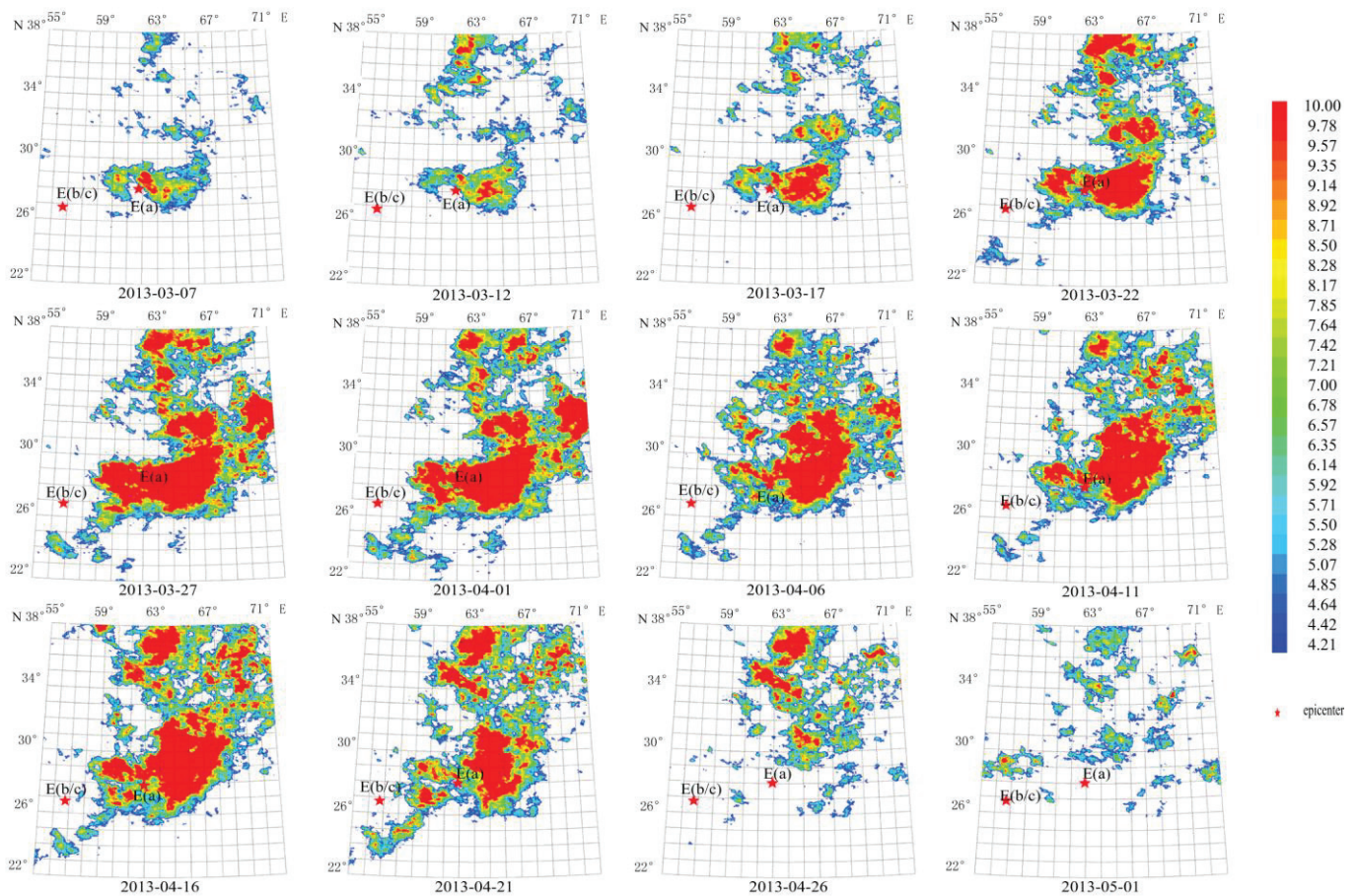


Figure 9. Temporal and spatial evolution of thermal radiation anomalies of earthquake group G3. Single earthquake: E(a) 28.03° N, 62.00° E, Mw 7.7, 16 April 2013; E(b) 26.56° N, 57.77° E, Mw 6.1, 11 May 2013. Wavelet transform: second- and seventh-order band-pass filter using Daubechies wavelets. Window length of FT power spectrum: 64 days using Welch's method. Characteristic period: 13 days.

In conclusion, the temporal and spatial evolution of the anomaly is clearly evident and can be easily extracted and identified. The anomaly evolution process presents a process of the appearance reaching the maximum disappearance, in line with the earthquake propagation process. The processes of the anomaly were not related to changes in weather but related to ShakeMap. Short-term seismic thermal radiation anomalies showed no obvious trends with earthquake magnitude, i.e., characteristic period, duration, earthquake time, and epicenter location are not positively related to earthquake magnitude. Strong earthquakes can affect regional thermal radiation background in that both the characteristic periods and the duration of impact became longer with increasing magnitude. By analyzing and summarizing the earthquake case, we found the following rules. The characteristic period of all earthquakes is 13 days if there is no earthquake of magnitude 7 within one year. If a magnitude 7 earthquake occurs within a year, the characteristic period of the earthquake in the region will increase from 13 days to 64 days. Therefore, we can use the short-term characteristic periods as an anomaly area indicator (AAI) to determine seismic anomalies.

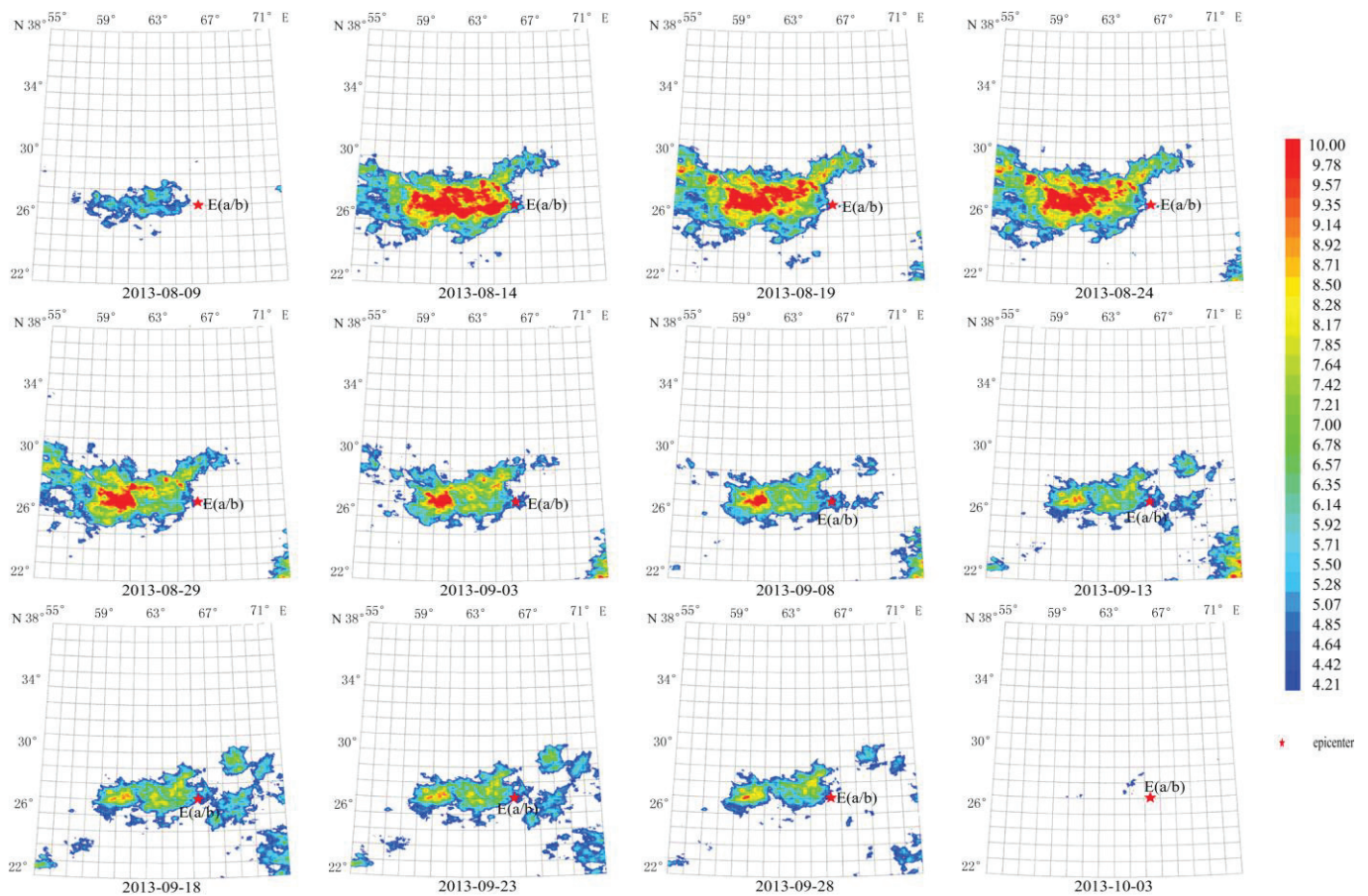


Figure 10. Temporal and spatial evolution of thermal radiation anomalies of earthquake group G4. Main shock: E(a) 26.95° N, 65.50° E, Mw 7.7, 24 September 2013. Aftershock: E(b) 27.18° N, 65.51° E, Mw 6.8, 28 September 2013. Wavelet transform: second- to seventh-order band-pass filter using Daubechies wavelets. Window length of FT power spectrum: 64 days using Welch's method. Characteristic period: 64 days.

Table 2. A summary of short-term thermal radiation anomalies of earthquakes.

Magnitude	Characteristic Period	Relationship of Abnormal Morphology and ShakeMap	Epicenter Location	Duration	Earthquake Time
Double Mw 6.4	13 Days	Inconsistency	Abnormal edge	about 2 months	2 months after the maximum anomaly
28&29 October 2008					
Mw 6.7	13 Days	Consistency	Abnormal edge	about 2 months	1 month after the maximum anomaly
20 December 2010					
Mw 7.2	64 Days	Consistency	Abnormal edge	about 2 months	2 months after the maximum anomaly
18 January 2011					
Mw 7.7	13 Days	Very consistent	Abnormal center	about 2 months	20 days after the maximum anomaly
16 April 2013					
Mw 7.7	64 Days	Very consistent	Abnormal edge	about 2 months	1 month after the maximum anomaly
24 September 2013					

4.3. Multiscale TFA of the Long-Term Thermal Radiation Anomalies of Mw 6+ Earthquake Swarms

In the following section, a multiscale TFA of the thermal radiation anomaly caused by large earthquakes is described in chronological order.

1. TF characteristics of long-term thermal radiation anomaly of Mw 6+ earthquakes in 2008

To extract the thermal radiation anomaly, the sixth-order transform using Daubechies wavelets and FT using Welch's method with the window length of 256 days and the characteristic period of 128 days were conducted. As shown in Figure 11, the anomaly that arose 3 months pre-earthquake lasted about 70 days from 10 July 2008 to 20 September 2008. The maximum amplitude of the anomaly occurred on 31 August 2008, reaching more than eight times the average value. The anomaly was far from the epicenter of the first shock but close to the double shock epicenter. The temporal and spatial evolution of the anomaly was clear and matches the ShakeMap in Figure 2a. The anomaly evolution process presents a process of the abnormal occurrence–maximum–disappearance, in line with the earthquake propagation process. The processes of the anomaly were not related to changes in weather, indicating that it was caused by the earthquakes.

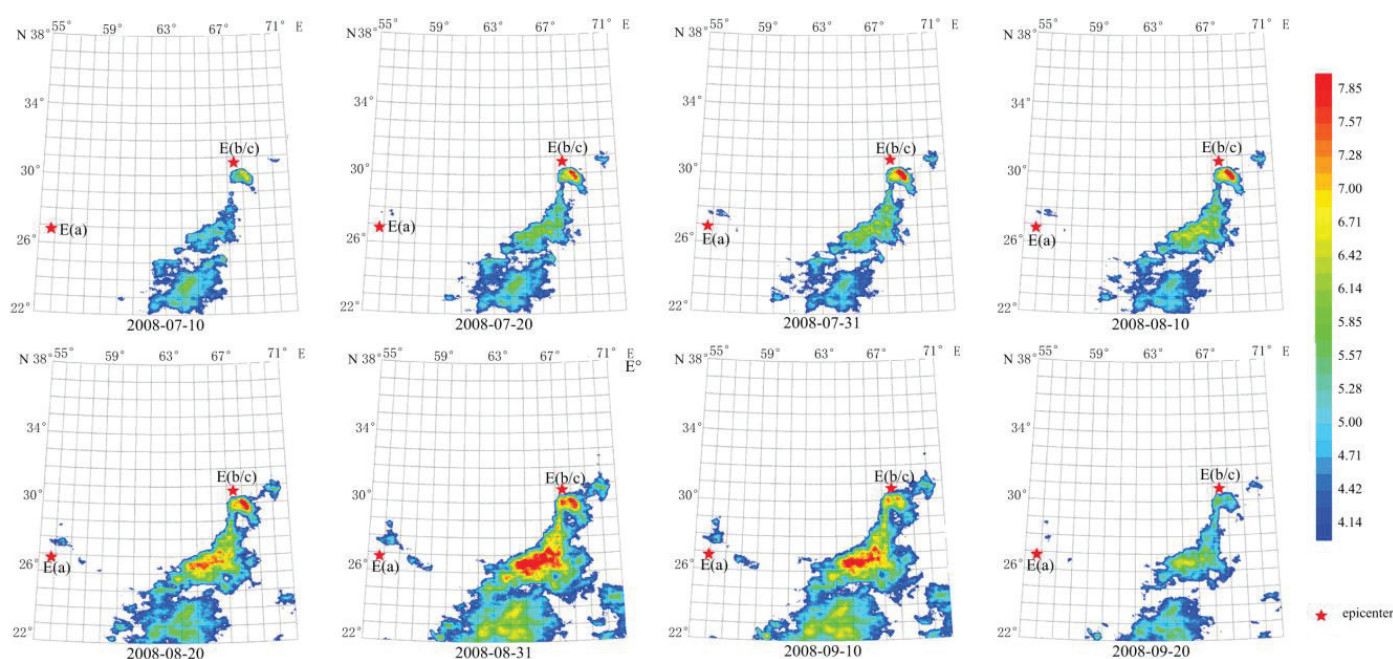


Figure 11. Temporal and spatial evolution of long-term thermal radiation anomaly of strong earthquakes. Single earthquake: E(a) 26.74° N, 55.83° E, Mw 6.1, 10 September 2008. Double earthquakes: E(b and c) 30.64° N, 67.35° E and 30.60° N, 67.40° E, Mw 6.4 and Mw 6.4, 28 October 2008 and 29, respectively. Wavelet transform: sixth order using Daubechies wavelets. Window length of FT power spectrum: 256 days using Welch's method. Characteristic period: 128 days.

2. TF characteristics of long-term thermal radiation anomalies of Mw7+ earthquakes in 2010 and 2011.

As shown in Figure 12, this anomaly showed up about 5 months pre-earthquake, lasting from 31 August 2010 to 31 January 2011. It reached the maximum amplitude from 31 October 2010 to 30 November 2011, which is more than eight times the average value. The area of the anomaly was close to the Mw 7.2 epicenter, and its shape is similar to that of the ShakeMap in Figure 1b. Different from the short-term anomaly, this long-term one disappeared with the earthquake occurrence. Thus, it can indicate the earthquake in advance. In this case, both the sixth- and seventh-order wavelet transforms showed anomalies, and FT used the window length of 512 days and the characteristic period of 256 days.

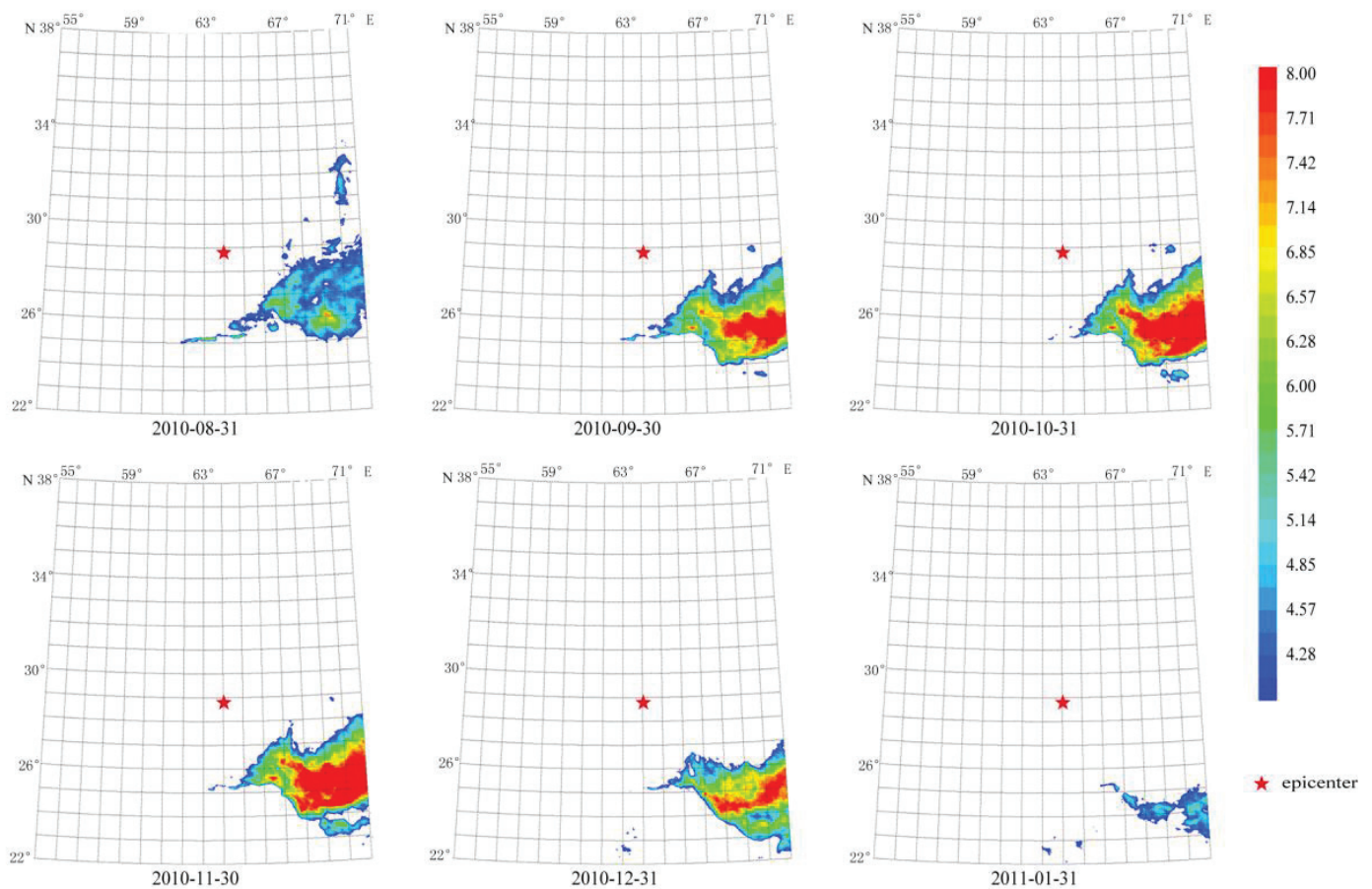


Figure 12. Temporal and spatial evolution of long-term thermal radiation anomaly of strong earthquakes. Single earthquake: E, 28.78° N, 63.95° E, Mw 7.2, 2011-01-18. Wavelet transform: sixth order using Daubechies wavelets. Window length of FT power spectrum using Welch's method: 512 days. Characteristic period: 256 days.

Figures 12 and 13 are similar in anomaly morphology, amplitude, characteristic period, and evolution processes. Nonetheless, the anomaly from the seventh-order transform appeared 3 months earlier and lasted 3 months longer than that from the sixth order transform. In addition, the sixth order anomaly disappeared with the earthquake occurrence, while the seventh order one faded in 2 months post-earthquake.

3. TF characteristics of long-term thermal radiation anomalies of Mw 7+ earthquakes in 2013.

Two Mw 7.7 earthquakes occurred separately in the first and second half of 2013 in this region. From Figures 9 and 10, the characteristic period of the short-term anomaly elongated from 13 days to 64 days, indicating that the first Mw 7.7 earthquake changed the regional thermal radiation. The second earthquake, along with the first one, increased the characteristic period of the long-term anomaly to 512 days (Figures 14 and 15).

The duration of the first anomaly was about 6 months from 31 July 2013 to 31 January 2014 (Figure 14). The maximum amplitude occurred from 30 September 2013 to 31 December 2013, which is eight times the average value. The second anomaly lasted about 9 months from 30 June 2013 to 31 March 2014. The maximum amplitude occurred from 30 September 2013 to 31 November 2013, which is also about eight times the average value. The shape of these two anomalies is similar to that of the ShakeMap provided by the USGS for the Mw 7.7 earthquakes (Figure 2d). These two anomalies showed similar characteristics, but their appearance and lasting time are obviously different. Specifically,

the seventh-order anomaly appeared 1 month earlier and lasted 3 months longer than the sixth order one.

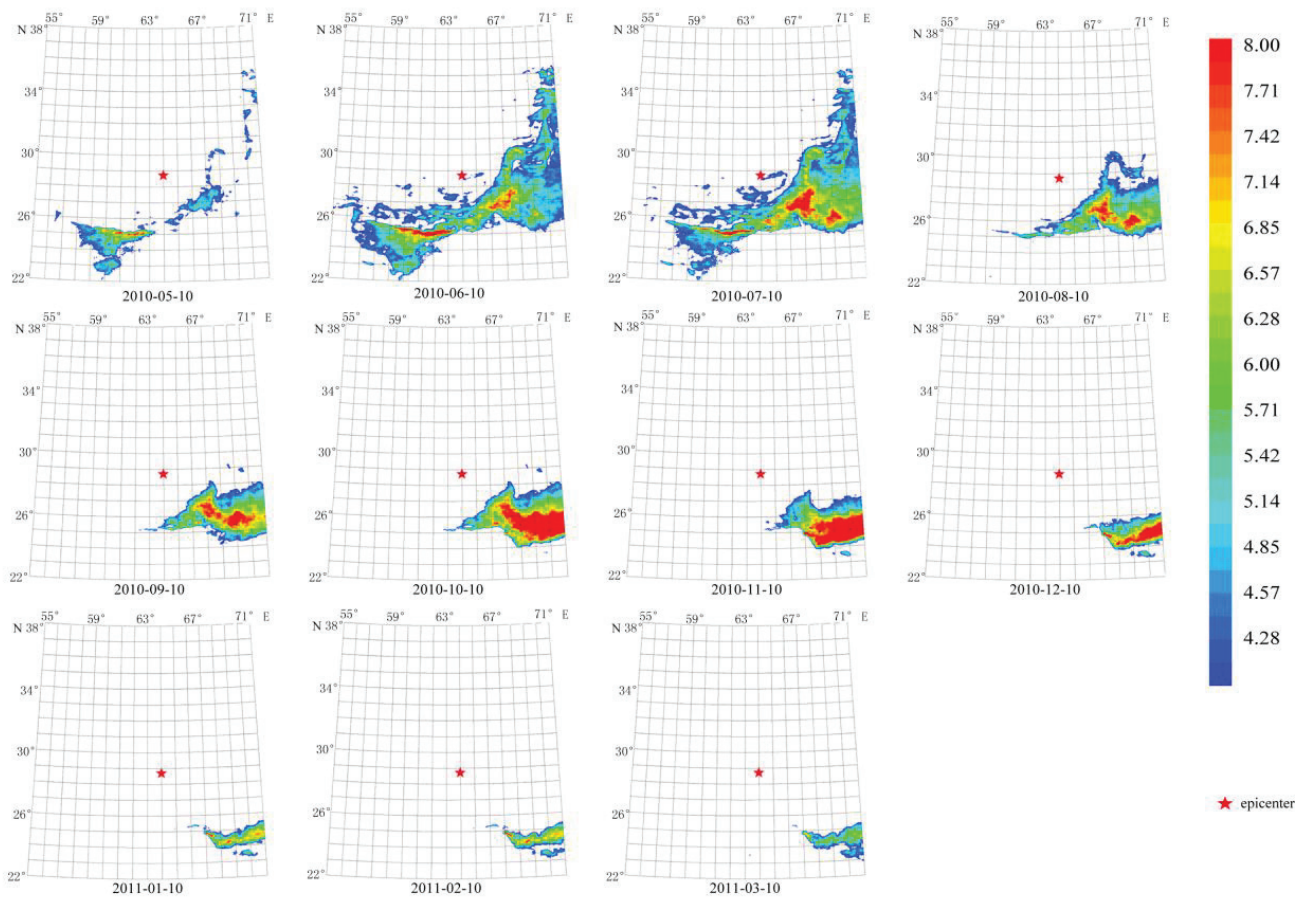


Figure 13. Temporal and spatial evolution of long-term thermal radiation anomaly of strong earthquakes. Single earthquake: E 28.78° N, 63.95° E, Mw 7.2, 18 January 2011. Wavelet transform: seventh order using Daubechies wavelets. Window length of FT power spectrum using Welch's method: 512 days. Characteristic period: 256 days.

Long-term anomalies are positively correlated to the earth magnitude in that the higher the magnitude the longer the window length and characteristic period (Table 3). For earthquakes < Mw 7, the window length of relative power spectrum is 256 days; for Mw 7.0+ earthquakes, the window length is 512 days. Moreover, only the sixth-order transform shows an anomaly when the earthquake Mw is smaller than 7.0; while both the sixth- and seventh-order transforms show anomalies for Mw 7.0+ earthquakes. In addition, the anomalies of Mw 7.0+ earthquakes lasted 6 months or more and the anomaly duration from seventh-order transforms is longer than that from sixth-order transforms, but those of earthquakes < Mw 7.0 only lasted 2 months. Last, all anomalies reached their maximum before or at the earthquake occurrence.

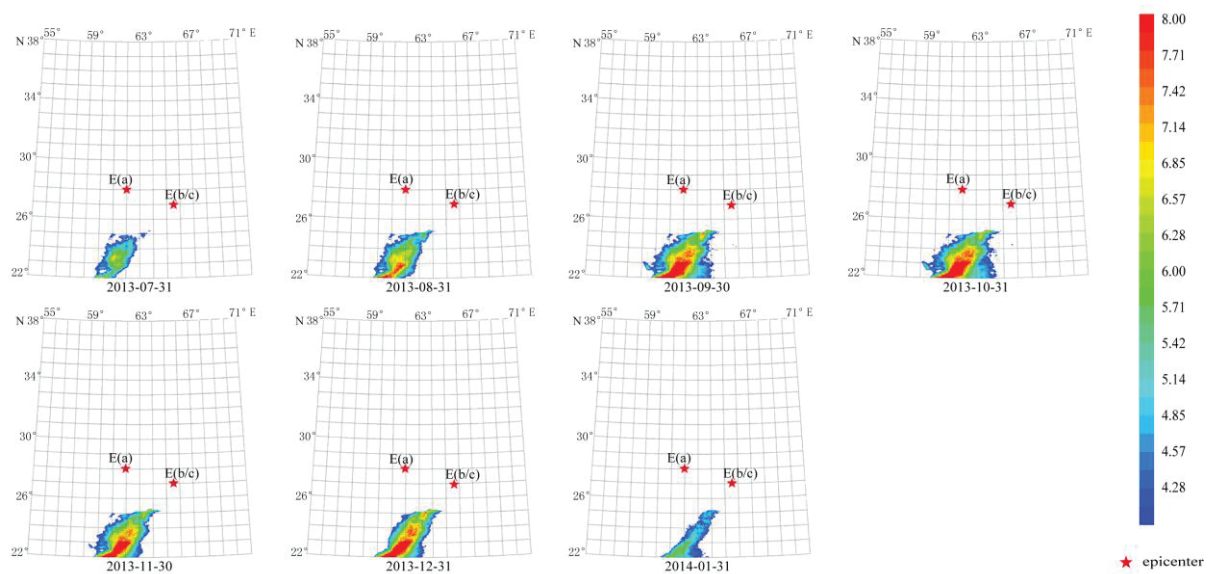


Figure 14. Temporal and spatial evolution of long-term thermal radiation anomaly of strong earthquakes. Single earthquake: E(a) 28.03° N, 62.00° E, Mw 7.7, 16 April 2013; E(b) 26.95° N, 65.50° E, Mw 7.7, 24 September 2013. Wavelet transform: sixth order using Daubechies wavelets. Window length of FT power spectrum using Welch's method: 512 days. Characteristic period: 512 days.

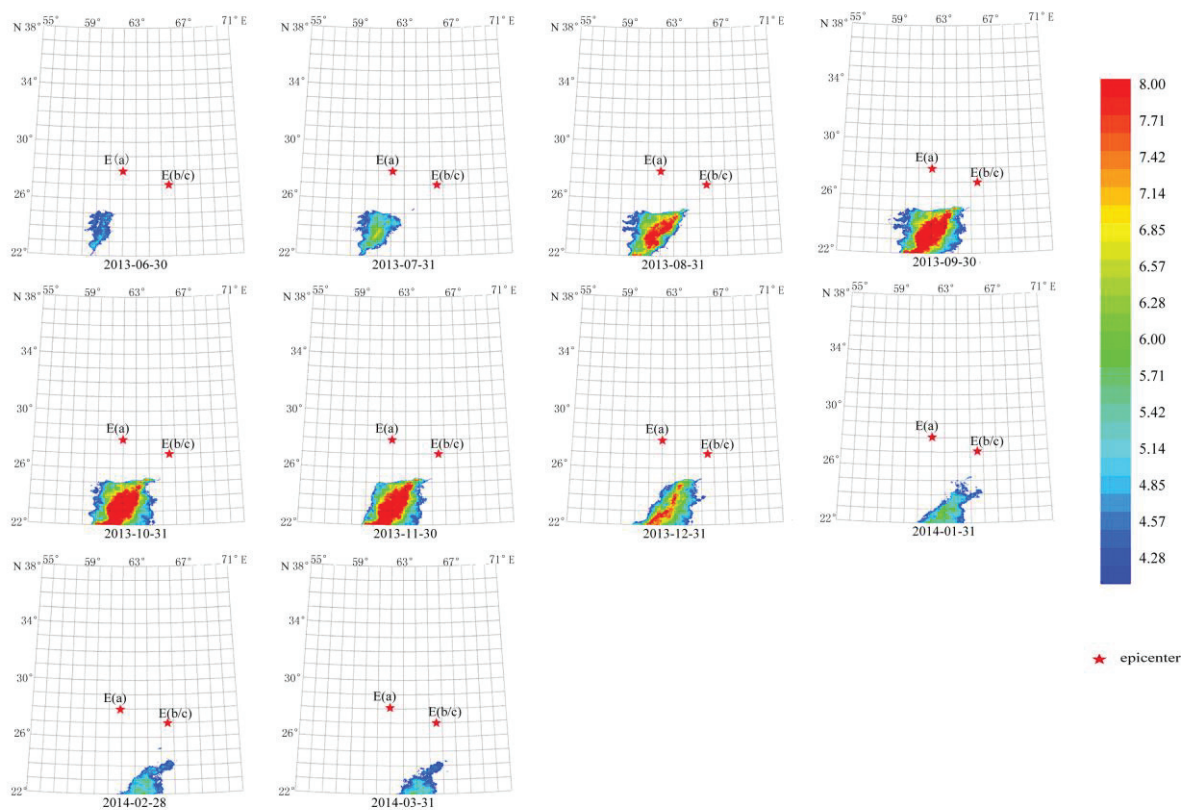


Figure 15. Temporal and spatial evolution of long-term thermal radiation anomalies of strong earthquakes. Single earthquake: E(a) 28.03° N, 62.00° E, Mw 7.7, 16 April 2013; E(b) 26.95° N, 65.50° E, Mw 7.7, 24 September 2013. Wavelet transform: seventh order using Daubechies wavelets. Window length of FT power spectrum using Welch's method: 512 days. Characteristic period: 512 days.

Table 3. A summary of the relationship between seismic anomaly and magnitude.

Magnitude	Wavelet Scale	Window Length of RPS	Characteristic Period	Duration	Earthquake Time
Mw 6.4	Only 6th	256 Days	128 Days	About 2 months	2 months after the maximum anomaly
Mw 6.7	6th and 7th	512 Days	256 Days	About 5 months (6th)	2 months after the maximum anomaly
Mw 7.2				About 10 months (7th)	3 months after the maximum anomaly
Two	6th and 7th	512 Days	512 Days	About 6 months (6th)	When the maximum anomaly occurred
Mw 7.7				About 9 months (7th)	When the maximum anomaly occurred

In conclusion, long-term anomalies are positively correlated to magnitude, i.e., both the window length of RPS and characteristic period increased with increasing magnitude. In addition, for Mw 7.0– earthquakes, only the sixth-order wavelet transform showed anomalies, while both the sixth- and seventh-order wavelet transforms showed anomalies. Therefore, we can use the long-term characteristic periods and window lengths as a magnitude decision factor (MDF) to determine seismic magnitude.

5. Discussion

5.1. Earthquake Impact on Regional Thermal Radiation Background

1. Characteristic period of short-term anomalies

The periods of the thermal radiation anomalies were changed by Mw 7+ earthquakes. Earthquake group G2 consists of foreshock, mainshock, and aftershock. From August to October 2010, short-term thermal radiation anomalies with characteristic periods of 13 and 64 days appeared (Figures 7 and 8). Later from September to November, another anomaly with the period of 64 appeared (Figure 16). Similarly, for G3 and G4, the characteristic period of the 16 April 2013 Mw 7.7 earthquake was 13 days but that of the 24 September 2013 Mw 7.7 earthquake increased to 64 days. However, Mw 6+ earthquakes from G1 do not show such an increasing trend in characteristic period. Therefore, the relative power spectrum of the 64-day characteristic period can be used to analyze the regional thermal radiation anomalies caused by strong earthquakes.

2. Duration of the impact and earthquake magnitude

Thermal radiation anomalies could be useful for studying strong earthquake processes and stress adjustment mechanisms. The 18 January 2011 Mw 7.2 earthquake affected the regional thermal radiation background for about 2 years (2010–2011), as shown in Figures 8 and 16, while the 16 April 2013 Mw 7.7 and 24 September 2013 Mw 7.7 earthquakes had a 4-year effect (2013–2016), as shown in Figures 10, 17 and 18. The duration of the impact increases with the magnitude of the earthquake, which may be further related to the regional stress adjustments and recovery post-earthquakes. Thus, longer duration of the impact could probably indicate slower regional stress adjustment and recovery.

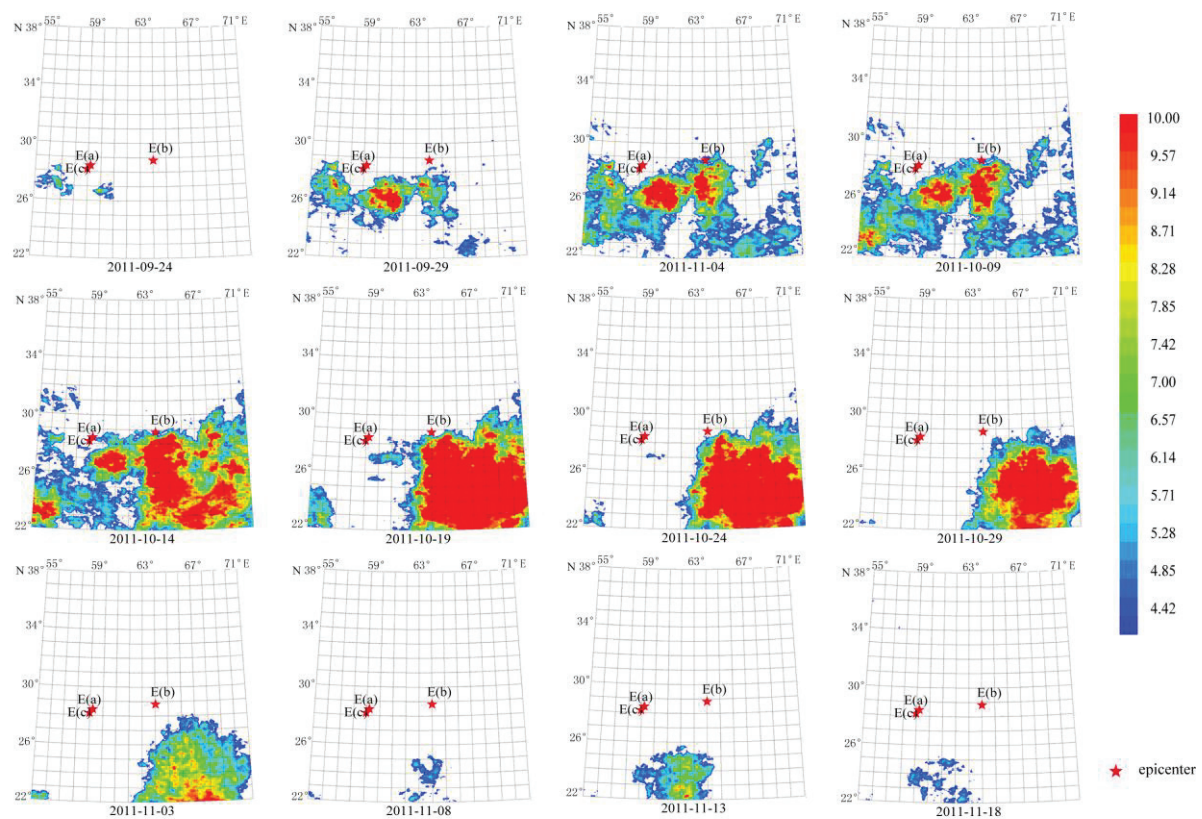


Figure 16. Spatial-temporal evolution of the thermal radiation background in 2011 caused by the Mw 7.2 earthquake. Single earthquake: E 28.78° N, 63.95° E, Mw 7.2, 18 January 2011. Wavelet transform: second- and seventh-order band-pass filter using Daubechies wavelets. Window length of FT power spectrum using Welch's method: 64 days. Characteristic period: 64 days.

By carrying out the TFA for the past 10 years' data with the characteristic period of 64 days, we found the thermal radiation anomaly caused by the Mw 7.2 earthquake impact lasted 2 years from 2010 to 2011. The duration of the anomaly was about 55 days from 24 September 2011 to 18 November 2011. The maximum amplitude of the anomaly occurred from 14 October 2011 to 29 October 2011, reaching more than 10 times the average value. The anomalous area was closer to the Mw 7.2 epicenter. The long-term and concentrated thermal radiation anomalies were similar to those of the Mw 7.2 earthquake group in terms of characteristic period, amplitude, form, and evolution.

From the data from 2013 to 2016, we found the thermal radiation anomaly caused by the double Mw 7.7 earthquakes. Their impact lasted for 4 years from 2013 to 2016. The anomaly continued for about 70 days from 17 March 2015 to 26 May 2015, which is consistent with the anomaly of the 16 April 2011 Mw 7.7 earthquake. The maximum amplitude of the anomaly that is more than 10 times the average value occurred from 26 April 2015 to 16 May 2015. The anomaly was close to the double Mw 7.7 epicenter. In addition, the long-term and concentrated thermal radiation anomalies are similar to those of the 16 April 2013 Mw 7.7 earthquake in terms of amplitude, morphology, and evolution (Figure 17).

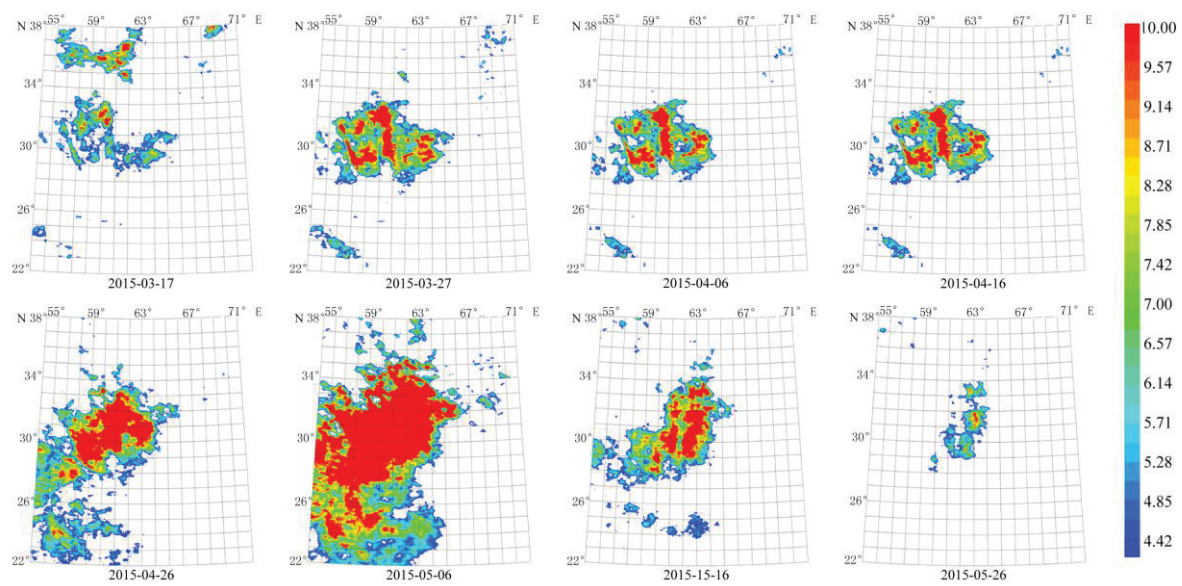


Figure 17. Spatial-temporal evolution of the thermal radiation background in 2015 caused by the Mw 7.7 earthquake. Earthquakes: E 28.03° N, 62.00° E, Mw 7.7, 16 April 2013; E 26.95° N, 65.50° E, Mw 7.7, 24 September 2013. Wavelet transform: second- and seventh-order band-pass filter using Daubechies wavelets. Window length of FT power spectrum: 64 days using Welch's method. Characteristic period: 64 days.

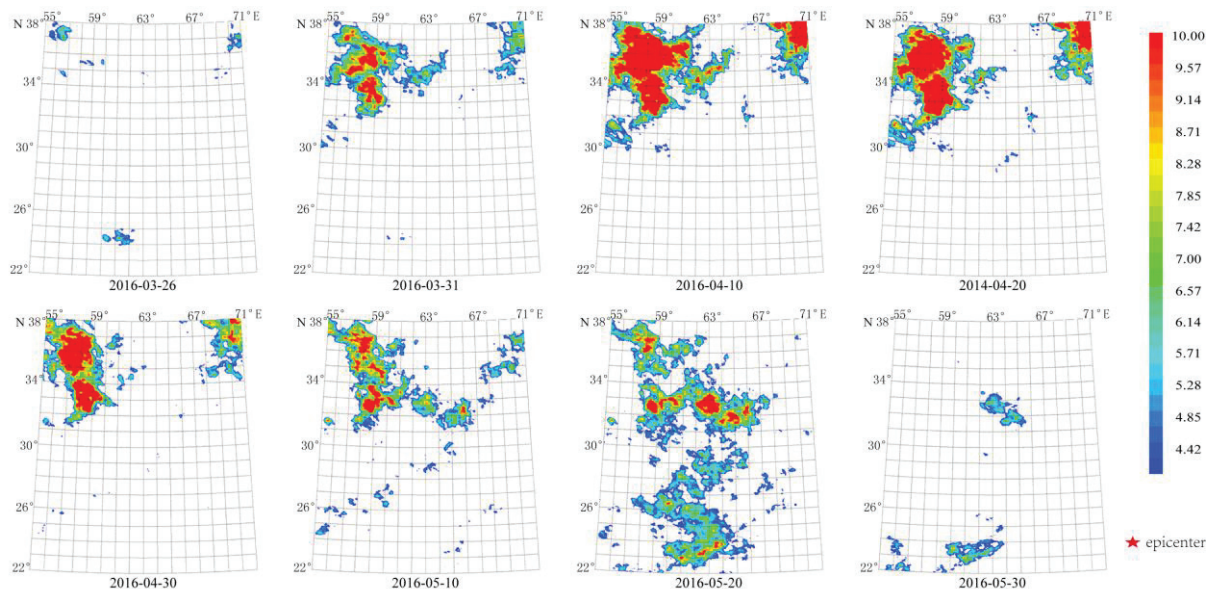


Figure 18. Spatial-temporal evolution of the thermal radiation background from 2016 caused by Mw 7.7 earthquakes. Earthquakes: E 28.03° N, 62.00° E, Mw 7.7, 16 April 2013; E 26.95° N, 65.50° E, Mw 7.7, 24 September 2013. Wavelet transform: second- and seventh-order band-pass filter using Daubechies wavelets. Window length of FT power spectrum: 64 days using Welch's method. Characteristic period: 64 days.

The anomaly from 26 March 2016 to 30 May 2016 with the maximum amplitude occurring from 10 April 2016 to 30 April 2016 was also closer to the double Mw 7.7 epicenter (Figure 18) and similar to that of the 16 April 2013 Mw 7.7 earthquake. The thermal radiation anomalies of the 24 September 2013 Mw 7.7 earthquake are different from that of the 2015–2016 in morphology. Specifically, the thermal radiation anomalies from the 24 September 2013 Mw 7.7 earthquake were concentrated at the epicenter and

nearby areas, and the area of the anomaly was smaller (Figure 10) than that of the anomaly from 2015 (Figure 17), and less dispersed than that of the anomaly from 2016 (Figure 18). The absence of thermal radiation in 2014 might be a consequence of the Mw 7.7 earthquake in September 2013.

3. Superposition of short-term anomalies induced by multiple earthquakes

If more than one Mw 6.0+ earthquake occurred over a 1-year time scale, the characteristics of the seismic thermal radiation anomalies became complex and were difficult to extract and identify, which is due to the superposition of the regional thermal radiation caused by the mutual coupling between strong earthquakes. The thermal radiation anomalies of the 2010–2011 earthquake swarms were complex, and the characteristic periods of 13 and 64 days appeared simultaneously, possibly because the earlier Mw 7.2 earthquake changed the latter regional thermal radiation. In addition, the characteristic period of the 16 April 2013 Mw 7.7 earthquake was 13 days, whereas that of the 24 September 2013 Mw 7.7 earthquake was 64 days, which may also have been affected by the earlier 16 April 2013 Mw 7.7 earthquake. This is also supported by the fact that there were no thermal radiation anomalies in March–May 2014.

5.2. Relationship between Thermal Anomalies and Earthquake Magnitude

From Table 2, short-term seismic thermal radiation anomalies showed no obvious change with earthquake magnitude, indicating that this method is suitable for extracting anomaly characteristics. In addition, characteristic period, duration, earthquake time, and epicenter location are not positively related to earthquake magnitude. All anomalies endured about 2 months, with the maximum amplitude appearing before the earthquake. The positive correlation identified is that the higher the magnitude, the better correspondence between the anomaly morphology and the ShakeMap.

Table 3 suggests strong relationships between long-term anomaly characteristics and earthquake magnitude. Since generally the longer the energy accumulates, the higher the earthquake magnitude, a longer characteristic period indicates longer energy accumulation time. This finding could serve as a criterion for determining the earthquake magnitude. In terms of wavelet scale, if only the sixth-order transform shows anomaly, the earthquake magnitude can be determined to be Mw 7.0–; if both the sixth- and seventh-order transforms show anomalies, the magnitude can be determined to be Mw 7.0+. Moreover, the longer the characteristic period, the higher the earthquake magnitude. Specifically, a period of 256 days corresponds to Mw 7.5–, and a period of 512 days indicates Mw 7.5+ earthquakes.

6. Conclusions

In this paper, we analyzed the characteristics of seismic thermal radiation anomalies in the same area. The results of this study have found the regular pattern and mechanism of anomaly seismic thermal radiation in the same region, and also improved the accuracy of the analytical processing model. The MS T-FRPS method was used to identify and determine the anomalous information of series earthquake groups with consistency law, which proves the stability and reliability of the method. By analyzing and summarizing the earthquake case, we propose the anomaly area indicator (AAI) to determine seismic anomalies using short-term characteristic periods and the magnitude decision factor (MDF) to determine seismic magnitude using the long-term characteristic periods and window lengths. In addition, we have explored the following: earthquake impact on regional thermal radiation background and relationship between thermal anomalies and earthquake magnitude and the like. Future research directions using the MS T-FRPS method to characterize regional thermal radiation anomalies induced by strong earthquakes could be (1) carrying out more case studies with accurate parameter quantification; (2) studying the relationship between thermal radiation anomalies and the earthquake magnitude; (3) employing methodologies where the correlation and causality between spatial points is examined in other phenomena; (4) utilizing machine learning algorithms to analyze earthquake precursors, crustal

movements, and other data to enhance the accuracy of earthquake predictions. Meaningful findings could help improve the accuracy of earthquake magnitude determination.

Author Contributions: Conceptualization, T.W. and C.W.; Validation, T.W.; Formal analysis, C.W.; Investigation, T.W. and Z.W.; Data curation, Z.W.; Visualization, L.W. and Z.Y.; Writing—review & editing, T.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the grant of National Natural Science Foundation of China, grant number 42161056, by Natural Science Foundation of Gansu Province, China, grant number 20JR10RA501, by National Special Projects of China, grant number Y612ZX1701, by the Central Funds Guiding the Local Science and Technology Development of Gansu Province, China, grant number 22ZY1QA006 and 22ZY2QC001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Acknowledgments: The authors gratefully acknowledge the anonymous reviewers for their helpful comments and suggestions. A portion of the computational resources was also provided by Gansu Computing Center.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gorny, V.I.; Salman, A.G.; Tronin, A.A. The earth outgoing IR radiation as an indicator of seismic activity. *Proc. Acad. Sci. USSR* **1988**, *301*, 67–69.
2. Tronin, A.A.; Hayakawa, M.; Molchanov, O.A. Thermal IR satellite data application for earthquake research in Japan and China. *J. Geodyn.* **2002**, *33*, 519–534. [CrossRef]
3. Panda, S.K.; Choudhury, S.; Saraf, A.K.; Das, J.D. MODIS land surface temperature data detects thermal anomaly preceding 08 October 2005 Kashmir earthquake. *Int. J. Remote Sens.* **2007**, *28*, 4587–4596. [CrossRef]
4. Qiang, Z.J.; Xu, X.D.; Dian, C.G. Thermal infrared anomaly precursor of short-term earthquakes. *Chin. Sci. Bull.* **1991**, *36*, 319–323.
5. Qiang, Z.J.; Xu, X.D.; Dian, C.G. Short-term-earthquake satellite thermal infrared and ground temperature increase anomalies. *Chin. Sci. Bull.* **1991**, *36*, 1894–1900.
6. Saraf, A.K.; Choudhury, S. Satellite detects surface thermal anomalies associated with the Algerian earthquakes of May, 2003. *Int. J. Remote Sens.* **2005**, *26*, 2705–2713. [CrossRef]
7. Singh, R.P.; Ouzounov, D. Earth processes in wake of Gujarat earthquake reviewed from space. *EOS Trans. Am. Geophys. Union* **2003**, *84*, 244. [CrossRef]
8. Ouzounov, D.; Freund, F. Mid-infrared emission prior to strong earthquakes analyzed by remote sensing data. *Adv. Space Res.* **2003**, *33*, 268–273. [CrossRef]
9. Tramutoli, V.; Bello, G.D.; Pergola, N.; Piscitelli, S. Robust satellite techniques for remote sensing of seismically active areas. *Ann. Geofis.* **2001**, *44*, 295–312. [CrossRef]
10. CORDIS RTD-PROJECTS. *Regular Update of Seismic Hazard Maps through Thermal Space Observations*; Project Reference: ENV4980741; European Communities: Madrid, Spain, 2000; Volume 10.
11. Ouzounov, D.; Bryant, N.; Logan, T.; Pulinet, S.; Taylor, P. Satellite thermal IR phenomena associated with some of the major earthquakes in 1999–2003. *Phys. Chem. Earth* **2006**, *31*, 154–163. [CrossRef]
12. Guo, X.; Zhang, Y.S.; Wei, C.X.; Zhong, M.J. Anomalies of middle infrared brightness before 2008 Yutian Ms7.3 and 2010 Yushu Ms7.1 earthquakes. *Acta Seismol. Sin.* **2014**, *36*, 175–183, (In Chinese with English Abstract).
13. Zhang, Y.S.; Guo, X.; Zhong, M.J.; Shen, W.; Li, W.; He, B. Wenchuan earthquake: Brightness temperature changes from satellite infrared information. *Chin. Sci. Bull.* **2010**, *55*, 1917–1924. [CrossRef]
14. Wei, C.; Zhang, Y.; Guo, X.; Hui, S.; Qin, M.; Zhang, Y. Thermal Infrared Anomalies of Several Strong Earthquakes. *Sci. World J.* **2013**, *2013*, 208407. [CrossRef] [PubMed]
15. Zhang, Y.S.; Guo, X.; Wei, C.; Shen, W.R.; Hui, S.X. The characteristics of seismic thermal radiation about Japan Ms9.0 and Myanmar Ms7.2 earthquake. *Chin. J. Geophys.* **2011**, *34*, 153–157, (In Chinese with English Abstract).
16. Pinty, B.; Gobron, N.; Verstraete, M.M.; Mélin, F.; Widlowski, J.-L.; Govaerts, Y.; Diner, D.J.; Fielding, E.; Nelson, D.L.; Madariaga, R.; et al. Observing earthquake-related dewatering using MISR/Terra satellite Data. *EOS Trans. Am. Geophys. Union* **2003**, *84*, 37–48. [CrossRef]
17. Qiang, Z.J.; Xu, X.D.; Dian, C.G. Thermal infrared anomalies under the non-temperature-increment background and the mechanism discussion. *Chin. Sci. Bull.* **1991**, *36*, 1901–1906.

18. Qiang, Z.J.; Dian, C.G.; Wang, X.J.; Hu, S.Y. Satellite thermal infrared anomalous temperature increase and short-term earthquake precursor. *China Sci. Bull.* **1992**, *37*, 1642–1646.
19. Qiang, Z.J.; Kong, L.C.; Guo, M.H.; Wang, G.P.; Zhao, Y. An experimental study on temperature increasing mechanism of satellitic thermal-infrared. *Acta Seismol. Sin.* **1997**, *10*, 247–252. [CrossRef]
20. Qiang, Z.J.; Dian, C.G. Satellite infrared brightness temperature anomaly earthquake imminent precursor. *Sci. China Ser. D* **1998**, *28*, 564–573, (In Chinese with English Abstract).
21. Qiang, Z.J.; Dian, C.G.; Li, L.Z.; Xu, M.; Ge, F.; Liu, T.; Zhao, Y.; Guo, M. Satellitic thermal infrared brightness temperature anomaly image—Short-term and short-term earthquake precursors. *Sci. China Ser. D-Earth* **1999**, *42*, 313–324. [CrossRef]
22. Xu, X.D.; Xu, X.M.; Ma, S.D.; Luo, Y. Preliminary inquiry into the causes of anomaly increase of air temperature by an short-term earthquake. *Acta Seismol. Sin.* **1995**, *8*, 149–154. [CrossRef]
23. Ma, J.; Chen, S.-Y.; Liu, P.-X.; Wang, Y.-P.; Liu, L.-Q. Temporal-spatial variations of associated faulting inferred from satellite infrared information: A case study of N-S seismo-tectonic zone in China. *Chin. J. Geophys.* **2006**, *49*, 707–716, (In Chinese with English Abstract). [CrossRef]
24. Zhang, L.F. The Applied Study in Predicting Earthquake Using Satellite Remote Sensing Data of Medium Wave Infrared. Master's Thesis, Lanzhou Institute of Seismology, Lanzhou, China, 2016.
25. Wei, C.X.; Zhang, Y.S.; Wang, Y. Time-frequency analysis of the influence of Japan Mw9.1 earthquake on regional thermal radiation background field. *Acta Seismol. Sin.* **2018**, *40*, 205–214, (In Chinese with English Abstract). [CrossRef]
26. Wang, Q.L.; Wang, W.P.; Liang, W.F.; Chen, B. Stress-dissipative heat geotemperature precursor mechanism of earthquakes. *Acta Seismol. Sin.* **1998**, *11*, 615–621, (In Chinese with English Abstract). [CrossRef]
27. Wu, L.X.; Liu, S.J.; Wu, Y.H.; Li, Y.Q. Remote-sensing rock mechanics (IV)—laws of thermal infrared radiation from compressively sheared fracturing of rock and its meanings for earthquake omens. *Chin. J. Rock Mech. Eng.* **2004**, *23*, 539–544, (In Chinese with English Abstract).
28. Akansu, A.N.; Haddad, R.A. *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*; Academic Press: Boston, MA, USA, 1992; ISBN 978-0-12-047141-6.
29. Hessami, K.; Jamali, F. Explanatory notes to the map of major active faults of Iran. *J. Seismol. Earthq. Eng.* **2006**, *8*, 1–11.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

A Novel Deep Convolutional Neural Network Combining Global Feature Extraction and Detailed Feature Extraction for Bearing Compound Fault Diagnosis

Shuzhen Han ^{1,2}, Pingjuan Niu ^{1,3,*}, Shijie Luo ², Yitong Li ², Dong Zhen ⁴, Guojin Feng ⁴ and Shengke Sun ⁵

¹ School of Mechanical Engineering, Tiangong University, Tianjin 300387, China; hanshuzhen@tiangong.edu.cn

² Office of the Cyberspace Affairs, Tiangong University, Tianjin 300387, China

³ School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China

⁴ School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China; G.Feng@hebut.edu.cn (G.F.)

⁵ School of Software, Tiangong University, Tianjin 300387, China; sunshengke@tiangong.edu.cn

* Correspondence: niupingjuan@tiangong.edu.cn; Tel.: +86-13821536307

Abstract: This study researched the application of a convolutional neural network (CNN) to a bearing compound fault diagnosis. The proposed idea lies in the ability of CNN to automatically extract fault features from complex raw signals. In our approach, to extract more effective features from a raw signal, a novel deep convolutional neural network combining global feature extraction with detailed feature extraction (GDDCNN) is proposed. First, wide and small kernel sizes are separately adopted in shallow and deep convolutional layers to extract global and detailed features. Then, the modified activation layer with a concatenated rectified linear unit (CReLU) is added following the shallow convolution layer to improve the utilization of shallow global features of the network. Finally, to acquire more robust features, another strategy involving the GMP layer is utilized, which replaces the traditional fully connected layer. The performance of the obtained diagnosis was validated on two bearing datasets. The results show that the accuracy of the compound fault diagnosis is over 98%. Compared with three other CNN-based methods, the proposed model demonstrates better stability.

Keywords: fault diagnosis; bearing compound fault; GDDCNN; CReLU; GMP

1. Introduction

With the development of modern industry, machine health monitoring becomes more and more important to maintain the safe operation of modern mechanical equipment [1,2]. Bearings are an indispensable part of modern machinery, especially in rotating machinery. Rolling bearing faults account for a large proportion of mechanical equipment faults, so it is necessary to ensure the normal operation of bearings [3–5]. In practical engineering, because of the complex and changeable operating conditions, once a fault occurs, it often affects other parts, leading to a compound fault [6–8]. Compared with single faults, the vibration signals are affected by mutual coupling and interference from various fault features, so the difficulty of feature extraction and fault diagnosis is greatly increased [9]. An effective and reliable compound fault diagnosis of rolling bearings is of great significance in guaranteeing safe operation.

In recent years, the compound fault diagnosis of rolling bearings has received increasing interest and attention from researchers. The research methods for compound fault diagnosis mainly include compound fault mechanism research [10,11], the blind source separation algorithm [12–15], the signal decomposition algorithm [16,17], and the artificial intelligence algorithm [18,19]. The method based on the compound fault mechanism, taking specific machineries as research objects, limits the application and portability of the model. Methods based on blind source separation have high requirements for the

number of channels of a raw signal. The number of sensors must meet the requirements of the algorithm, which increases the cost of a fault diagnosis. The modal decomposition algorithm is a typical signal decomposition method that is suitable for non-stationary signal processing.

However, there are some issues, such as mode aliasing and the end effect, that can directly affect the accuracy of a compound fault diagnosis. With the rapid development of artificial intelligence techniques, machine learning methods, including support vector machines (SVMs), Bayesian classifiers, artificial neural networks (ANNs), and convolutional neural networks (CNNs), have already been applied as potential tools to fault diagnosis, especially deep learning approaches. Ref. [20] proposed a deep inception net with atrous convolution and applied this model to a bearing fault diagnosis. This model overcomes the problem caused by the different feature distributions of characteristics between two data sets and achieves high accuracy. Ref. [21] proposed a model based on CNN, named WDCNN, for a bearing fault diagnosis that achieves high accuracy on raw vibration signals directly in the Case Western Reserve University (CWRU) bearing data set. Ref. [22] utilized an improved CNN for a fault diagnosis. In that network, a multiscale cascaded layer is added to CNN, which can enhance the classification information of the input. Ref. [23] constructed a CNN with feature alignment that addressed the finite-shift-invariance problem, which can extract a robust fault feature. Ref. [24] adopted a multi-task CNN through information fusion for a fault diagnosis on two bearing data sets. The experiment results showed that the proposed model improved the accuracy of the fault diagnosis. Ref. [25] implemented a lightweight CNN that combined transfer learning and self-attention, which can achieve higher diagnosis accuracy than traditional CNN models. Ref. [26] proposed a novel multiscale CNN model that incorporates multiscale learning during the feature extraction process to diagnose the fault of a wind turbine gearbox. A novel multiscale residual attention CNN model was proposed in [27], which utilizes multiscale features, an attention mechanism, and residual learning to enhance feature extraction ability. Experimental validation on two bearing datasets demonstrated that the algorithm achieved higher accuracy. Based on a CNN model, [28] designed a multi-task CNN model that utilizes a speed identification task and a load identification task as auxiliary tasks to improve the performance of a fault diagnosis task. The experimental results showed that multi-task learning can enhance the fault diagnosis performance of the model. In [29], an end-to-end fault diagnosis model combining CNN with LSTM is designed, which can realize a bearing fault diagnosis in a short time. Ref. [30] presented an improved one-dimensional multiscale model that combined different extended convolutional kernels with varying dilation rates. The superiority of this approach was validated on the CWRU and PU datasets. Ref. [31] integrated vibration signals and sound signals through a one-dimensional CNN for fusion and validated that it has a higher diagnostic accuracy. Ref. [32] used short-time Fourier transform to convert vibration signals into spectrograms and adopted a model based on CNN for feature extraction and health status classification. Ref. [33] proposed a lightweight CNN combined with data augmentation technology for a bearing fault diagnosis. A novel hybrid CNN-MLP model for diagnoses was proposed in [34], which combined mixed input to achieve a rolling bearing diagnosis. In [35], a lightweight CNN model with fixed feature graph dimensions is constructed with down-sampling vibration signals to construct spectral graphs, which can achieve high classification accuracy on low-dimensional input data. Ref. [36] put forward a model based on optimized-parameter maximum-correlated kurtosis deconvolution and CNN for a bearing compound fault diagnosis and verified its effectiveness.

The above models demonstrate that deep learning approaches have significantly improved fault diagnosis effectiveness. However, the relationship between compound fault and single fault is one-to-many or many-to-many, which is not a simple superposition of a single fault signal. Compared with signal faults, feature extraction and location of compound faults are more difficult and challenging, which brings great difficulties to compound fault diagnosis. Currently, intelligent diagnosis methods based on deep learning

mainly focus on single faults. Some models are suitable for single fault diagnosis; however, they are not effective for compound fault diagnosis. The research and application of deep learning approaches to bearing compound fault diagnosis are still in their infancy. Thus, in this study, a novel deep convolutional neural network combining global feature extraction with detailed feature extraction (GDDCNN) was proposed.

The proposed GDDCNN model is a deep convolutional neural network that incorporates both global feature extraction and detailed feature extraction, where G means global feature extraction, D means detailed feature extraction, and DCNN represents a deep convolutional neural network. DCNN is a feature-progressive learning algorithm in which the deep network can continue to learn more advanced fault features based on shallow features. By designing two feature extraction modules, G and D, better learning ability can be achieved for DCNN. Therefore, more abundant fault features can be extracted through GDCNN, thus improving the performance of fault diagnosis. The contributions of this study are summarized as follows:

1. A novel deep convolutional neural network combining global feature extraction with detailed feature extraction (GDDCNN) is proposed to extract features adaptively from a raw signal.
2. The modified activation concatenated ReLU (CReLU) is applied in the shallow layer of GDDCNN. It can improve the performance of global feature extraction.
3. The global max pooling (GMP) strategy is designed to replace the traditional fully connected layer, which can extract shift-invariance features and reduce model parameters. It overcomes the overfitting problem in the training process.

The rest of this study is described as follows. Section 2 reviews the basic theory of CNN. Section 3 expresses the proposed model in detail. Section 4 presents the experimental validation through two bearing datasets. Finally, Section 5 draws the conclusion.

2. Theoretical Background

CNN can be built with multiple layers, comprising a convolution layer, a pooling layer, an activation layer, and a classification layer. The convolution layer, pooling layer, and activation layer are used to extract features from input signals. One-dimensional CNN has been widely applied to 1-D vibration signal processing due to its powerful feature extraction ability. The classification layer applies the extracted features to classify.

The convolution layer is the core layer of the CNN structure, which convolves the input data with filter kernels. The network makes the filter learn to activate when it extracts certain features, then realizes feature extraction. The mathematical form can be described as follows:

$$y_j^l = K_i^l * x_i^l = \sum_{i \in M_j} x_i^l * w_{ij}^l + b_j^l \quad (1)$$

where the y_j^l denotes the output of the l -th layer; K_i^l is the i -th convolution kernel of the l -th layer; x_i^l denotes the input of the l -th layer; the notation $*$ represents the convolution operation; w_{ij}^l denotes the weights the convolution kernel; and b_j^l is the offset.

The activation layer is usually followed by the convolution layer, which is an essential layer. The activation function defines the output and input connections of a neuron, which is usually nonlinear. It makes the network learn nonlinear features from an input vibration signal to improve the capability of feature extraction. Rectified linear unit (ReLU) [37] is commonly used as an activation function in CNN, which is defined as follows:

$$\text{ReLU}(y_j^l) = \begin{cases} 0 & y_j^l < 0 \\ y_j^l & y_j^l \geq 0 \end{cases} \quad (2)$$

which denotes the activation value is 0 on the negative half-axis.

Batch normalization (BN) [38] is applied to deep neural networks, which can reduce the shift of internal covariance and improve the accuracy of the training model. In addition,

BN compels the learned options into a standard distribution with a mean value of 0 and a variance of 1, which can accelerate the training speed of the model. The transform of BN is described as follows:

$$\begin{cases} \mu_B = \frac{1}{m} \sum_{i=1}^m x_i \\ \delta_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \\ \hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\delta_B^2 + \varepsilon}} \\ y_i = \gamma \hat{x}_i + \beta \end{cases} \quad (3)$$

where m represents the mini-batch size, μ_B expresses the mini-batch mean, and δ_B^2 represents the mini-batch variance. γ and β are learnable parameters of the network.

The pooling layer performs a down-sampling operation, which removes redundant features and extracts deeper features. Max pooling and average pooling are the most common pooling operations. Max pooling generally outperforms average pooling for time series classification tasks, which is expressed as follows:

$$p_i^l = \max_{(j-1)s+1 \leq t \leq js} \{a_i^l(t)\} \quad (4)$$

where the p_i^l represents the output features of the l -th layer; max is max pooling, $a_i^l(t)$ means the output value of the t -th neuron in the i -th channel of the l -th layer, s denotes the stride of the pooling.

In the classification layer, the Softmax function is applied to normalize the probability of each category in the output. Softmax in the neural network is defined as

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}} \quad (5)$$

where K is the number of categories and x_i represents the logits of the j -th output neuron.

3. Proposed GDDCNN

To overcome the problems of compound fault feature extraction, a novel deep CNN is proposed. The structure of the proposed model, GDDCNN, is shown in Figure 1, composed of an input module, a feature extractor module, a GMP layer, and a Softmax classifier. Global feature extraction and detailed feature extraction constitute the feature extractor module. Besides, this study used two strategies: modified activation operation and GMP strategy during the feature extraction process to enhance the feature extraction ability of the proposed model, GDDCNN. Finally, a compound fault diagnosis is implemented. More details are illustrated in subsequent parts.

3.1. GDDCNN Architecture Design

Different kernel sizes for convolution: In CNN, convolutional kernel size plays an important role in the convolutional layer because kernels of different sizes can obtain different features. Generally, wider kernels pay more attention to global information during convolution operations, thereby extracting more global features, while smaller kernels can capture more detailed features. To obtain more robust features from a raw vibration signal, global feature extraction and detailed feature extraction are combined and applied to the feature extractor module. This study designed different kernel sizes for convolution; wide kernel sizes are applied in the shallow (first and second) convolution layers to extract global features, while the deep convolutional kernels are small, which help to obtain detailed features. Multi-convolution layers that adopt small convolutional kernels make the CNN networks deeper, which can improve the performance of compound fault feature extraction. Finally, the size of the convolution kernel for different layers is set to be [64, 16, 3].

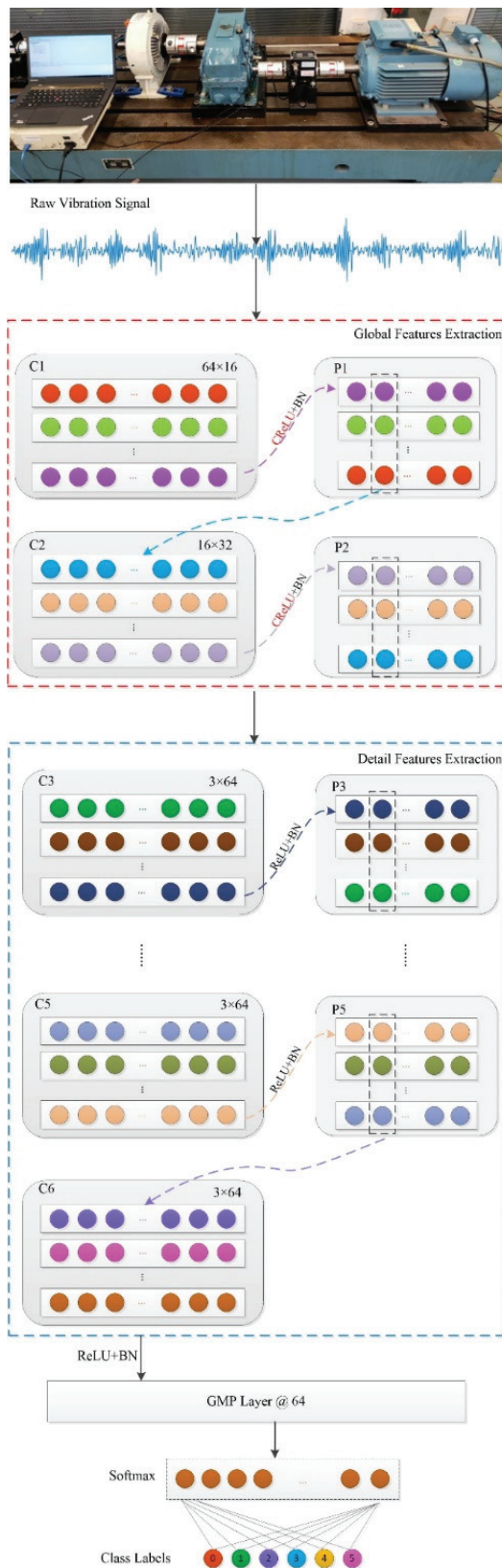


Figure 1. Structure of GDDCNN.

Modified activation operation: In addition to the convolution, the activation operation also affects the performance of CNN. Due to the advantages of simple calculation and no vanishing gradient problem, ReLU is widely applied in CNN as an activation unit.

However, when the inputs are negative, the neurons are always inactive. These dead neurons in a network may never activate, which stops learning and thus affects the learning ability of the network. In [39], it was found that in the shallow layer of CNNs, the parameter distribution of the network exhibits a stronger negative correlation, while with the deepening of the network, this negative correlation gradually becomes weak. CReLU is a concatenated ReLU, which, through inverting the feature map to activate the negative inputs, helps features transmit better backward. CReLU is defined as follows:

$$\text{CReLU}(x) = [\text{ReLU}(x), \text{ReLU}(-x)] \quad (6)$$

where ReLU represents the ReLU activation function. The modified activation CReLU is used in the shallow (first and second) convolution layers, which can improve the performance of global feature extraction.

GMP strategy: The fully connected layer is generally applied after the last convolutional or pooling layer, which can integrate the class-differentiated local features extracted by CNN. Each neuron in the fully connected layer is fully connected to all the neurons in the previous layer. Due to its fully connected characteristics, the parameters of the fully connected layer are numerous, and the calculation is extremely complicated. The right part of Figure 2 shows the GMP progress, that is, the max value of each channel is used as the new feature vector. The GMP strategy clearly reduces the dimension of the feature vector, which can avoid overfitting. Another advantage is that it is more native to the convolution structure by enforcing correspondences between feature maps and categories. Furthermore, this strategy can retain the spatial information that keeps shift-invariance, resulting in robust extracted features.

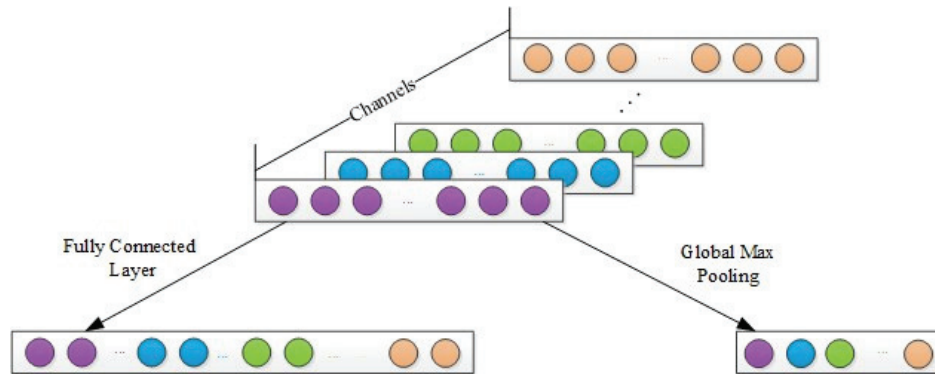


Figure 2. Global max pooling.

3.2. Training of GDDCNN

The architecture of GDDCNN is designed to take advantage of one-dimensional CNN. The cross-entropy loss function is used to estimate the consistency between the Softmax output probability distribution and the target class probability distribution. Suppose that $p(x)$ and $q(x)$ represent the target distribution and estimated distribution, respectively. The loss function can be expressed as follows:

$$\text{Loss}(L) = H(p(x), q(x)) = -\sum_x p(x) \log q(x) \quad (7)$$

The gradient of loss L and the gradient of parameters related to BN should be back-propagated in GDDCNN during training as per the following rules:

$$\frac{\partial L}{\partial \bar{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma \quad (8)$$

$$\frac{\partial L}{\partial \delta_B^2} = \sum_{i=1}^m \frac{\partial L}{\partial \bar{x}_i} \cdot (x_i - \mu_B) \cdot \frac{-1}{2} (\delta_B^2 + \epsilon)^{-3/2} \quad (9)$$

$$\frac{\partial L}{\partial \mu_B} = \left(\sum_{i=1}^m \frac{\partial L}{\partial \bar{x}_i} \cdot \frac{-1}{\sqrt{\delta_B^2 + \varepsilon}} \right) + \frac{\partial L}{\partial \delta_B^2} \cdot \frac{\sum_{i=1}^m -2(x_i - \mu_B)}{m} \quad (10)$$

$$\frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial \bar{x}_i} \cdot \frac{-1}{\sqrt{\delta_B^2 + \varepsilon}} + \frac{\partial L}{\partial \delta_B^2} \cdot \frac{2(x_i - \mu_B)}{m} + \frac{\partial L}{\partial \mu_B} \cdot \frac{1}{m} \quad (11)$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^m \frac{\partial L}{\partial y_i} \cdot \bar{x}_i \quad (12)$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^m \frac{\partial L}{\partial y_i} \quad (13)$$

Adam is a learning rate adaptive optimization algorithm that assembles the Adagrad algorithm and the RMSProp algorithm, which allows a model to allocate more updates to rarely occurring features, thereby aiding in the convergence of the optimization process. By bringing together the adaptive learning rates of Adagrad and the stability of RMSProp, the Adam algorithm provides an effective solution to optimize the proposed model, enabling it to converge quickly and efficiently while avoiding common optimization pitfalls. The choice of optimizer plays a crucial role in the success of the proposed model's training process.

In order to minimize the value of the loss function, the Adam optimization algorithm is applied to update the weights and obtain the optimal weights. The Adam optimizer is chosen for the following reasons: Bearing vibration signals typically contain a large number of data points due to their time-series nature. The efficient convergence property of the Adam algorithm allows the model to learn from a large volume of data more quickly. This, in turn, accelerates the assessment of the bearing's health status, saving training time. Features in bearing vibration signals may vary in importance over different time intervals. Adam's adaptive learning rate automatically adjusts the learning rate based on the gradient of each parameter. This capability helps the model better adapt to dynamic variations within a signal. When the Adam optimizer is initialed, it is necessary to set the learning rate ε , exponential decay rate of moment estimation ρ_1, ρ_2 , and constant δ . In this experiment, we set ε to 0.001, ρ_1, ρ_2 to default 0.9, 0.999, and the constant was set to prevent the numerical mutation from being set to 10^{-8} during the dividing operation. The detailed process of the Adam algorithm is shown in Table 1. More details on the Adam algorithm can be found in [40].

Table 1. Detailed flow of the Adam algorithm.

Adam Algorithm
Initialize variables $\varepsilon, \rho_1, \rho_2, \delta, \theta, s = 0, r = 0, t = 0$
While stopping criterion not met do
Sample a minibatch of m examples from the training set
$\{x^{(1)}, \dots, x^{(m)}\}$ with corresponding targets $y(i)$
Computing gradient:
$g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x^{(i)}; \theta), y^{(i)})$
$t \leftarrow t + 1$
Update biased first-order moment estimation: $s \leftarrow \rho_1 s + (1 - \rho_1)g$
Update biased second-order moment estimation: $r \leftarrow \rho_2 r + (1 - \rho_2)g \odot g$
Correcting the deviation of the first moment: $\hat{s} \leftarrow \frac{s}{1 - \rho_1^t}$
Correcting the deviation of the second moment: $\hat{r} \leftarrow \frac{r}{1 - \rho_2^t}$
Calculation update: $\Delta\theta = -\varepsilon \frac{\hat{s}}{\sqrt{\hat{r} + \delta}}$
Apply update: $\theta \leftarrow \theta + \Delta\theta$
end while

3.3. Diagnosis Procedure

1. The raw vibration signals of rolling bearings are collected by a data acquisition device.
2. The obtained vibration signals are sliced into samples of length 2048 for standardization processing and then used as network input.
3. Extracting fault features by combining global feature extraction with detailed feature extraction. Utilizing the GMP layer to integrate features. Then, Softmax is employed as a classifier to classify fault features.
4. Testing samples are entered into the network to realize fault diagnosis and validate the performance of the proposed model, GDDCNN.

4. Experiments and Results

4.1. Data Acquisition

In this study, the two-stage gear drive test bench shown in Figure 3, is composed of a drive motor, a two-stage helical gear box, a powder brake, a torque speed sensor, a coupling, tested bearings, and a data acquisition device. The experimental data was obtained by the intermediate shaft end bearing of the two-stage helical gearbox. In this task group, three rotating speeds (1200 rpm, 1500 rpm, and 1800 rpm) and three electrical machinery loads (0 A, 0.5 A, and 1.0 A) were set up in 9 working conditions in the experiment, respectively. In addition, the sampling frequency is 96 kHz, and the sampling time is 10 s. Collecting 9 raw data samples from each category, with a sampling length of 96,000 for each raw data sample. There are six categories of bearing faults: one normal, three single faults, and two compound faults. These six health conditions of bearings are designated as normal (NO), inner fault (IF), outer fault (OF), roll fault (RF), inner and roll compound fault (IRF), and outer and roll fault (ORF). All bearing faults simulated by electrical discharge machining (EDM) technology are shown in Figure 4.

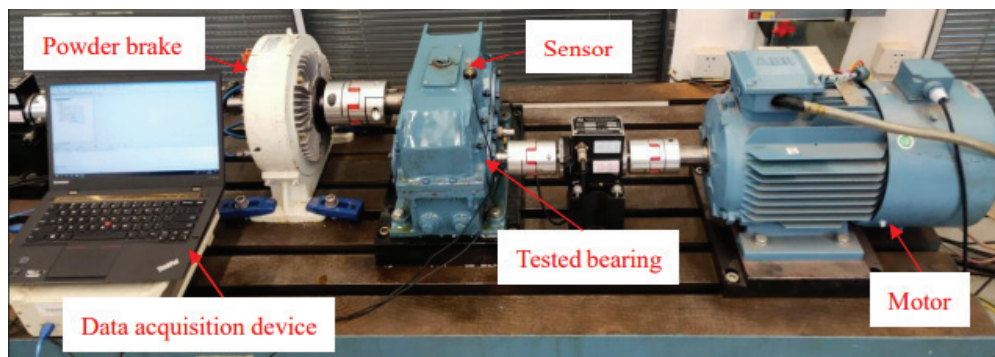


Figure 3. Two-stage gear drive test bench.

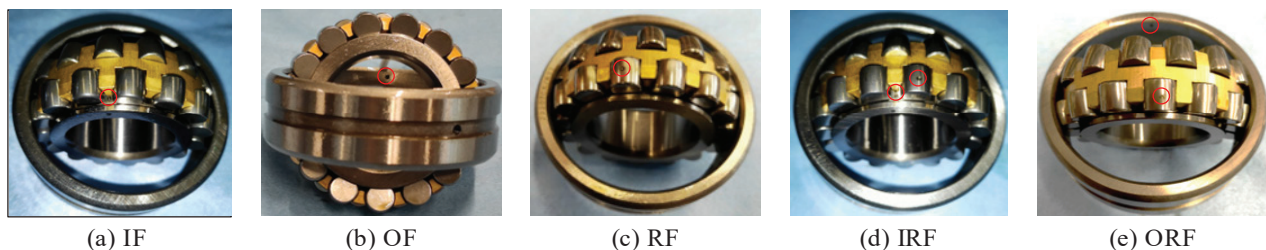


Figure 4. Bearing fault simulation: (a) inner fault; (b) outer fault; (c) roll fault; (d) inner and roll compound fault; and (e) outer and roll fault.

In addition to the data obtained from the above two-stage gear drive test bench, this study also utilized the public Case Western Reserve University (CWRU) dataset [41] to simulate compound fault signals. The raw vibration signals under a load of 1 hp with a rotating speed of 1772 r/min and a sampling frequency of 12kHz are selected. Four

compound fault categories are simulated by selected signals, which are IF + RF, OF + RF, IF + OF, and IF + OF + RF.

4.2. Data Preprocessing

(1) *Sample Segmentation*: After data acquisition, an easy data augmentation technique of slicing the raw vibration signals with overlap is applied to process training samples. In order to reflect the periodicity of raw signals, this study set the sampling size to 2048 and the overlapping window size to 512. The data segmentation process is shown in Figure 5. After data segmentation, 16,848 samples of each category were obtained, totaling 101,088 samples.

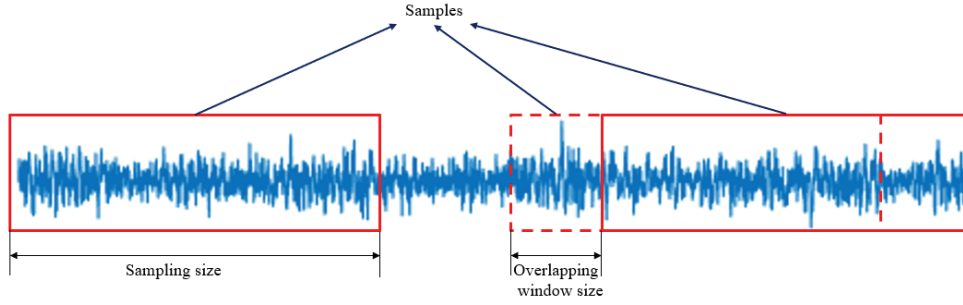


Figure 5. Data segmentation process.

(2) *Standardization*: After sample segmentation, in order to improve the convergence speed and accuracy of the model, the samples need to be standardized. The samples are standardized by the z-score method and follow the distribution of the mean value 0 and variance 1. The standardization process for the raw data $(x_1, x_2, x_3 \dots x_n)$ is as follows:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (14)$$

$$\delta = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (15)$$

$$\text{Standard}(x_i) = \frac{x_i - \bar{x}}{\delta} \quad (16)$$

4.3. Parameters of the Proposed Network

The experiments were performed using the above data to verify the performance of the proposed method. The architecture of the proposed method involves stacking 2 convolutional layers with a wide kernel size and pooling layers, 4 convolutional layers with a small kernel size and pooling layers, followed by GMP and a Softmax layer to implement fault diagnosis. The first convolutional kernel size and channel depth are, respectively, set to 64 and 16; the second convolutional kernel size and channel size are set to be 16 and 32, respectively; and the rest of the convolutional kernel size and channel are set to be 3 and 64, respectively. All five categories of pooling are max pooling, with a size of 2. The first two activation functions are the improved activation connection ReLU (CReLU), and the remaining activation functions are ReLU. To improve the performance of the proposed model, GDDCNN, BN is applied after each activation layer. Each convolution layer and pooling layer utilizes a zero-padding of causal and same type to prevent the loss of edge information. The detailed parameters of the proposed network are listed in Table 2.

Table 2. Detailed parameters of the proposed network.

Layer Type	Kernel Size/Stride	Kernel Channel Size	Padding	Output Size
Input	/	/	/	2048
C1	64/8	16	causal	256*16
P1	2/2	/	same	128*32
C2	16/2	32	causal	64*32
P2	2/2	/	same	32*64
C3	3/1	64	causal	32*64
P3	2/2	/	same	16*64
C4	3/1	64	causal	16*64
P4	2/2	/	same	8*64
C5	3/1	64	causal	8*64
P5	2/2	/	same	4*64
C6	3/1	64	same	4*64
GMP	/	/	/	64*1
Softmax	/	/	/	6*1

4.4. Comparative Experiment

In order to validate the feasibility of the proposed method, which is compared with several currently popular models, including classical CNN, CNN-SVM, and WDCNN [21]. This study selected 1500 samples from each category under all working conditions randomly for experiments: 1000 samples as training data, 300 samples as verification data, and 200 samples as testing data. The details of the experimental dataset are shown in Table 3. All experiments were run on a server equipped with an Intel Core i7-6800k CPU, an Nvidia GeForce GTX 1080 Ti GPU, and 32G of RAM.

Table 3. Description of samples distribution.

Fault Location	Sample Number	Training/Validation/Test	Label
Normal	1500	1000/300/200	0
IF	1500	1000/300/200	1
OF	1500	1000/300/200	2
RF	1500	1000/300/200	3
IRF	1500	1000/300/200	4
ORF	1500	1000/300/200	5

Accuracy is the most common metric used to evaluate the performance of a classifier. It calculates the proportion of correctly classified samples out of the total number of samples. Specifically, accuracy can be calculated using the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

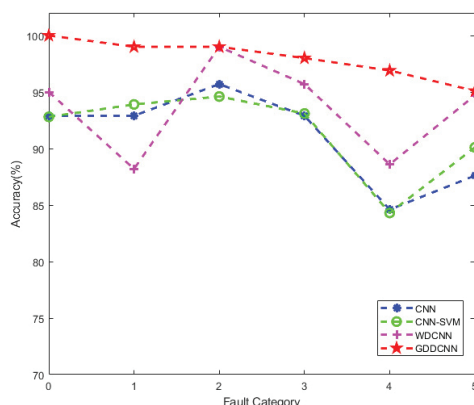
where TP represents the number of true positive samples, i.e., the number of samples correctly predicted as positive by the classifier; TN represents the number of true negative samples, i.e., the number of samples correctly predicted as negative by the classifier; FP represents the number of false positive samples, that is, the number of samples incorrectly predicted as positive by the classifier; and FN represents the number of false negative samples, that is, the number of samples incorrectly predicted as negative by the classifier.

The results of the diagnosis accuracy for each category using four different methods are summarized in Table 4. According to the statistical results, it can be observed that the single fault diagnosis accuracy of the proposed method can reach over 98% or even approach 100%, and the compound fault diagnosis accuracy is also over 95%. The accuracy of other methods in category 4 (IRF) of the compound fault diagnosis is lower than 90%.

Table 4. Classification accuracy results of the four methods.

Method	Class Label					
	0	1	2	3	4	5
CNN	92.9%	92.9%	95.7%	92.9%	84.6%	87.6%
CNN-SVM	92.8%	93.9%	94.6%	93.1%	84.3%	90.1%
WDCNN	95%	88.2%	99%	95.7%	88.6%	94.7%
GDDCNN	100%	99%	99%	98%	96.9%	95.1%

In order to demonstrate the advantages of the proposed method more visually and evidently, a line chart is used to compare the accuracy of the four methods, as shown in Figure 6. Where fault category 0–5 indicates a fault label, the details about fault labels correspond to the fault category, as shown in Table 3. It can be clearly observed that the proposed method has better performance than the other three methods. The diagnosis accuracy of each category is higher than that of other methods. Moreover, the diagnosis accuracy of each category is relatively balanced, with none of the six categories having particularly low accuracy. Although the WDCNN model has higher accuracy than CNN and CNN-SVM, its performance and stability are still not comparable to the proposed method.

**Figure 6.** Accuracy comparison of four methods.

Precision is similar to but different from accuracy, which is the proportion of correctly predicted positive samples among all samples identified as positive. Precision represents the prediction accuracy of the positive sample results. Recall refers to the proportion of correctly predicted positive samples among all true positive samples, which reflects the proportion of positive samples correctly predicted. F1-Score is the harmonic mean of precision and recall, which combines information from these two metrics. The calculation of precision, recall, and F1-score is as follows:

$$Precision = \frac{TP}{TP + FP} \quad (18)$$

$$Recall = \frac{TP}{TP + FN} \quad (19)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (20)$$

in which TP , FP , and FN have the same meanings as those in Formula 17. The precision, recall, and F1-score of the proposed method are shown in Figure 7. Where Class 0 represents normal, Class 1, Class 2, and Class 3 represent inner fault, outer fault, and roll fault. Both Class 4 and Class 5 are compound faults; Class 4 expresses inner and roll compound faults, while Class 5 represents outer and roll faults.

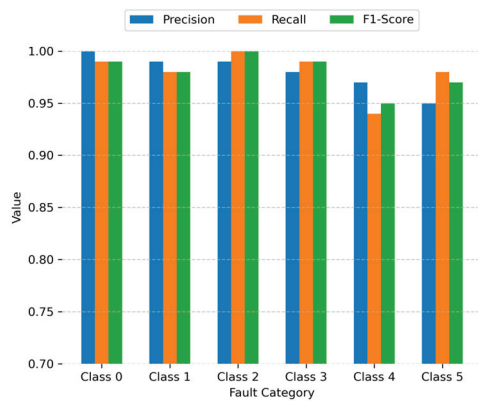


Figure 7. Precision, recall, and F1-score of the GDDCNN.

Precision, recall, and F1-score are also crucial performance metrics for a classifier; a higher precision, recall, and F1-score indicates a better performance of the classifier. It can be observed from Figure 7 that all the values of the proposed model for all conditions are high.

In order to further validate the effectiveness of the fault diagnosis model, three error metrics—maximum average error (MAE), mean squared error (MSE), and root mean squared error (RMSE)—are selected to evaluate the fitting effect of the fault diagnosis model. The formulas for each metric are defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_{i,true} - Y_{i,pred}| \quad (21)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_{i,true} - Y_{i,pred})^2 \quad (22)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_{i,true} - Y_{i,pred})^2} \quad (23)$$

where N is the number of testing samples, the value of $Y_{i,true}$ is 1, and $Y_{i,pred}$ is the probability value of the category to which it belongs. The value of $Y_{i,pred}$ calculated through the Softmax classifier is closer to 1, indicating a better fitting effect of the model. Therefore, smaller values of the three metrics represent better stability for the model. The metrics results of the four methods are listed in Table 5. From the statistical results, it can be seen that the three metrics of the GDDCNN are all smaller than those of other methods, indicating that the GDCNN proposed in this study has good stability.

Table 5. Metrics results of the four methods.

Method	MAE	MSE	RMSE
CNN	0.265	0.861	0.928
CNN-SVM	0.270	0.860	0.927
WDCNN	0.108	0.275	0.524
GDDCNN	0.032	0.065	0.255

Finally, the feature distribution in different layers of the proposed model, GDDCNN, can be visualized through the t-SNE technique. t-SNE is a nonlinear dimensionality reduction algorithm that calculates the positions of sample points based on the similarity between sample points in high-dimensional space and the distances between sample points in low-dimensional space. The experiment results, i.e., the position information after dimension reduction of the raw signal, six convolutional layers, and the GMP layer, are

shown in Figure 8. It can be obviously seen that the raw signal and the features in the shallow convolution layers are inseparable. With the deepening of the convolution layer, the feature distribution can be improved significantly. From the visualization of the GMP layer, we can observe the feature distribution, which suggests that the proposed model cannot discriminate between category 1 and category 4 very well. It may be due to the fact that category 4 is a compound fault that contains both category 1 and category 3. The samples of other categories can cluster in their own area, and the inter-category differences are good. The experiment results indicate that the proposed model, GDDCNN, can extract robust and effective features from a raw signal and achieve efficient fault diagnosis.

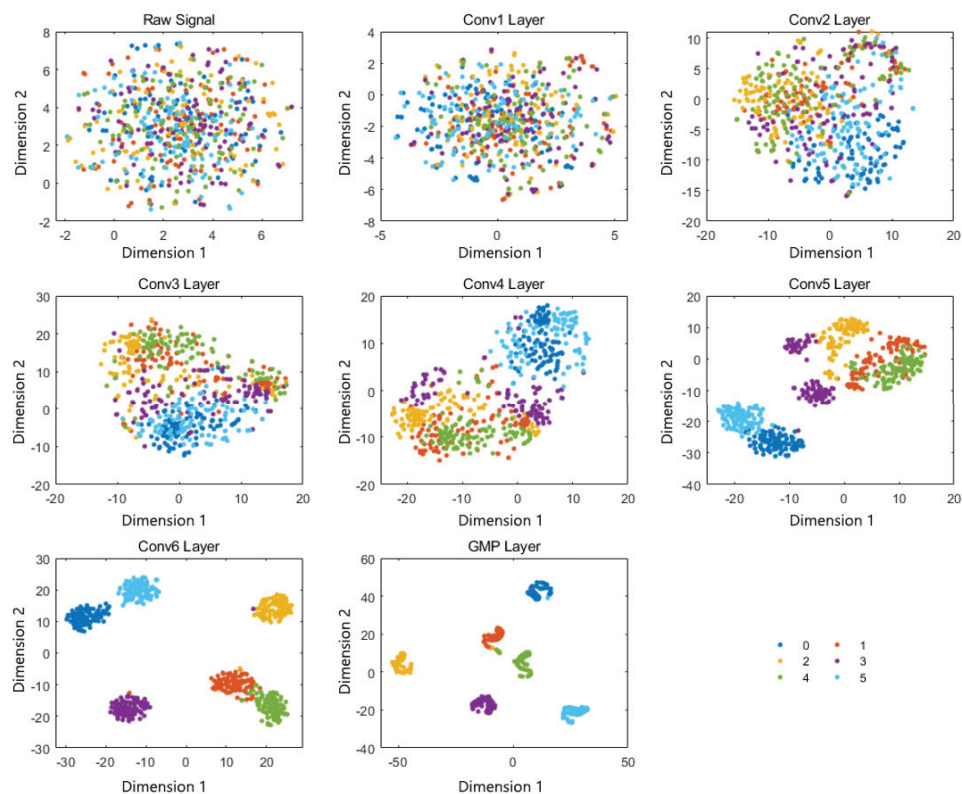


Figure 8. Feature visualization via t-SNE.

In order to further verify the effectiveness and performance of the proposed model, GDDCNN, fault diagnosis experiments are carried out with the simulated data from the CWRU dataset. Data is processed by overlapping sampling during simulation. The distribution of samples is listed in Table 6.

Table 6. Distribution of samples.

Fault Location	Sample Number	Training/Validation/Test	Label
Normal	1000	700/200/100	0
IF	1000	700/200/100	1
OF	1000	700/200/100	2
RF	1000	700/200/100	3
IF + RF	1000	700/200/100	4
OF + RF	1000	700/200/100	5
IF + OF	1000	700/200/100	6
IF + OF + RF	1000	700/200/100	7

The average accuracy of the four models is shown in Table 7. It is easy to see from the results that all models achieve high diagnostic accuracy. The accuracy of the proposed model, GDDCNN, reaches 100%, which indicates it has good performance.

Table 7. Average accuracy of the four models.

Model	CNN	CNN-SVM	WDCNN	GDDCNN
Accuracy (%)	98	98	98.75	100

5. Conclusions

This study proposed a novel deep CNN named GDDCNN for the diagnosis of practical engineering compound faults. GDDCNN extracted robust features adaptively from raw vibration signals and realized compound fault diagnosis without any manual processing. The datasets acquired by a two-stage gear drive test bench and simulated with the published dataset of CWRU were applied to experiments to verify the performance of the proposed model. The experimental results demonstrate that GDDCNN can achieve more than 98% accuracy on both datasets. Compared with three other existing CNN-based models, this model not only achieves satisfactory diagnostic accuracy but also exhibits strong stability. By combining different convolutional kernel sizes, improved activation CReLU, and GMP strategy, the performance of GDDCNN has been improved, and compound fault diagnosis has been achieved more effectively. In further work, we will continue to explore how to adaptively select hyperparameters to enhance the accuracy and stability of the diagnosis model. We also aim to research the application of transfer learning to achieve compound fault diagnosis across different operating conditions.

Author Contributions: Data curation, Y.L.; investigation, S.H., G.F. and S.S.; methodology, S.H.; resources, D.Z.; software, S.L.; validation, P.N.; writing—original draft, S.H.; writing—review and editing, P.N. and Y.L.; funding acquisition, P.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research is financially supported by the Tianjin Education Commission (2019ZD08).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wei, S.; Wang, D.; Wang, H.; Peng, Z. Time-varying envelope filtering for exhibiting space bearing cage fault features. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3504313. [CrossRef]
2. Wang, L.; Zhang, X.; Liu, Z.; Wang, J. Sparsity-based fractional spline wavelet denoising via overlapping group shrinkage with non-convex regularization and convex optimization for bearing fault diagnosis. *Meas. Sci. Technol.* **2020**, *31*, 055003. [CrossRef]
3. Gao, S.; Xu, L.; Zhang, Y.; Pei, Z. Rolling bearing fault diagnosis based on intelligent optimized self-adaptive deep belief network. *Meas. Sci. Technol.* **2020**, *31*, 055009. [CrossRef]
4. Rai, A.; Upadhyay, S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [CrossRef]
5. Qin, Y.; Jin, L.; Zhang, A.; He, B. Rolling bearing fault diagnosis with adaptive harmonic kurtosis and improved bat algorithm. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3508112. [CrossRef]
6. Tang, G.; Wang, Y.; Huang, Y.; Liu, N.; He, J. Compound bearing fault detection under varying speed conditions with virtual multichannel signals in angle domain. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 5535–5545. [CrossRef]
7. Zhang, J.; Zhang, Q.; He, X.; Sun, G.; Zhou, D. Compound-Fault diagnosis of rotating machinery: A fused imbalance learning method. *IEEE Trans. Control Syst. Technol.* **2020**, *29*, 1462–1474. [CrossRef]
8. Ma, X.; Hu, Y.; Wang, M.; Li, F.; Wang, Y. Degradation state partition and compound fault diagnosis of rolling bearing based on personalized multilabel learning. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3520711. [CrossRef]
9. Jin, Y.; Qin, C.; Zhang, Z.; Tao, J.; Liu, C. A multi-scale convolutional neural network for bearing compound fault diagnosis under various noise conditions. *Sci. China Tech. Sci.* **2022**, *65*, 2551–2563. [CrossRef]
10. Zhang, X.; Yan, C.; Liu, Y.; Yan, P.; Wang, Y.; Wu, L. Dynamic modeling and analysis of rolling bearing with compound fault on raceway and rolling element. *Shock Vib.* **2020**, *5*, 8861899. [CrossRef]
11. Huang, C.; Zhang, B.; Yi, C.; Jin, H. Vibration model for axle box bearings with multiple defective rolling elements for high-speed trains and the defects localization method. *Shock Vib.* **2020**, *39*, 34–43.

12. Tang, G.; Luo, G.; Zhang, W.; Yang, C.; Wang, H. Underdetermined blind source separation with variational mode decomposition for compound roller bearing fault signals. *Sensors* **2016**, *16*, 897. [CrossRef]
13. Huang, D.; Chen, D.; Ke, L.; Zhao, L.; Sun, G. A blind source separation method for multi-fault modal characteristic signals of rolling bearings with error influences. *Acta Armamentarii* **2018**, *39*, 1419–1428.
14. Xi, J.; Cui, J.; Jiang, L. JADE-ICA-based blind source separation of multi-fault signals of rolling bearings. *Shock Vib.* **2017**, *5*, 231–237.
15. Mika, D.; Budzik, G.; Jóźwik, J. Single Channel Source Separation with ICA-Based Time-Frequency Decomposition. *Sensors* **2020**, *20*, 2019. [CrossRef] [PubMed]
16. Pan, H.; Yang, Y.; Li, X.; Zheng, J.; Cheng, J. Symplectic geometry mode decomposition and its application to rotating machinery compound fault diagnosis. *Mech. Syst. Signal Process.* **2019**, *114*, 189–211. [CrossRef]
17. Miao, Y.; Zhao, M.; Lin, J. Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition. *ISA Trans.* **2018**, *84*, 82–95. [CrossRef] [PubMed]
18. Islam, M.M.; Manjurul, J.M. Kim Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vector Machines. *Reliab. Eng. Syst. Safty* **2019**, *184*, 55–56. [CrossRef]
19. Jin, Y.; Qin, C.; Huang, Y.; Liu, C. Actual bearing compound fault diagnosis based on active learning and decoupling attentional residual network. *Measurement* **2021**, *173*, 108500. [CrossRef]
20. Chen, Y.; Peng, G.; Xie, C.; Zhang, W.; Li, C.; Liu, S. ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis. *Neurocomputing* **2018**, *294*, 61–71. [CrossRef]
21. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **2017**, *17*, 425. [CrossRef] [PubMed]
22. Huang, W.; Cheng, J.; Yang, Y.; Guo, G. An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis. *Neurocomputing* **2019**, *359*, 77–92. [CrossRef]
23. Chen, J.; Huang, R.; Zhao, K.; Wang, W.; Liu, L.; Li, W. Multiscale convolutional neural network with feature alignment for bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3517010. [CrossRef]
24. Guo, S.; Zhang, B.; Yang, T.; Lyu, D.; Gao, W. Multitask convolutional neural network with information fusion for bearing fault diagnosis and localization. *IEEE Trans. Ind. Electron.* **2020**, *67*, 8005–8016. [CrossRef]
25. Zhong, H.; Lv, Y.; Yuan, R.; Yang, D. Bearing fault diagnosis using transfer learning and self-attention ensemble lightweight convolutional neural network. *Neurocomputing* **2022**, *501*, 765–777. [CrossRef]
26. Jiang, G.; He, H.; Yan, J.; Xie, P. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3196–3207. [CrossRef]
27. Jia, L.; Chow, T.W.S.; Wang, Y.; Yuan, Y. Multiscale Residual Attention Convolutional Neural Network for Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3519413. [CrossRef]
28. Liu, Z.; Wang, H.; Liu, J.; Qin, Y.; Peng, D. Multitask Learning Based on Lightweight 1DCNN for Fault Diagnosis of Wheelset Bearings. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3501711. [CrossRef]
29. Khorram, A.; Khalooei, M.; Rezghi, M. End-to-End CNN + LSTM Deep Learning Approach for Bearing Fault Diagnosis. *Appl. Intell.* **2021**, *51*, 736–751. [CrossRef]
30. He, J.; Wu, P.; Tong, Y.; Zhang, X.; Lei, M.; Gao, J. Bearing Fault Diagnosis via Improved One-Dimensional Multi-Scale Dilated CNN. *Sensors* **2021**, *21*, 7319. [CrossRef]
31. Wang, X.; Mao, D.; Li, X. Bearing Fault Diagnosis Based on Vibro-Acoustic Data Fusion and 1D-CNN Network. *Measurement* **2021**, *173*, 108518. [CrossRef]
32. Pham, M.T.; Kim, J.-M.; Kim, C.H. Accurate Bearing Fault Diagnosis under Variable Shaft Speed Using Convolutional Neural Networks and Vibration Spectrogram. *Appl. Sci.* **2020**, *10*, 6385. [CrossRef]
33. Oh, J.W.; Jeong, J. Data Augmentation for Bearing Fault Detection with a Light Weight CNN. *Procedia Comput. Sci.* **2020**, *175*, 72–79. [CrossRef]
34. Sinitsin, V.; Ibryaeva, O.; Sakovskaya, V.; Eremeeva, V. Intelligent Bearing Fault Diagnosis Method Combining Mixed Input and Hybrid CNN-MLP Model. *Mech. Syst. Signal Process.* **2022**, *180*, 109454. [CrossRef]
35. Yoo, Y.; Jo, H.; Ban, S.-W. Lite and Efficient Deep Learning Model for Bearing Fault Diagnosis Using the CWRU Dataset. *Sensors* **2023**, *23*, 3157. [CrossRef]
36. Gao, S.; Shi, S.; Zhang, Y. Rolling bearing compound fault diagnosis based on parameter optimization MCKD and convolutional neural network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3508108. [CrossRef]
37. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
38. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.
39. Shang, W.; Sohn, K.; Almeida, D.; Lee, H. Understanding and improving convolutional neural networks via concatenated rectified linear units. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 2217–2225.

40. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
41. Lou, X.; Loparo, K. A Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech. Syst. Signal Process.* **2004**, *18*, 1077–1095. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Controlled Symmetry with Woods-Saxon Stochastic Resonance Enabled Weak Fault Detection

Jian Liu ¹, Jiaqi Guo ¹, Bing Hu ^{2,*}, Qiqing Zhai ³, Can Tang ¹ and Wanjia Zhang ¹

¹ College of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China; liujian@nufe.edu.cn (J.L.); guojiaqi.orange@foxmail.com (J.G.)

² College of Modern Posts, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

³ College of Science, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

* Correspondence: hubing@njupt.edu.cn

Abstract: Weak fault detection with stochastic resonance (SR) is distinct from conventional approaches in that it is a nonlinear optimal signal processing to transfer noise into the signal, resulting in a higher output SNR. Owing to this special characteristic of SR, this study develops a controlled symmetry with Woods-Saxon stochastic resonance (CSwWSSR) model based on the Woods-Saxon stochastic resonance (WSSR), where each parameter of the model may be modified to vary the potential structure. Then, the potential structure of the model is investigated in this paper, along with the mathematical analysis and experimental comparison to clarify the effect of each parameter on it. The CSwWSSR is a tri-stable stochastic resonance, but differs from others in that each of its three potential wells is controlled by different parameters. Moreover, the particle swarm optimization (PSO), which can quickly find the ideal parameter matching, is introduced to attain the optimal parameters of the CSwWSSR model. Fault diagnosis of simulation signals and bearings was carried out to confirm the viability of the proposed CSwWSSR model, and the results revealed that the CSwWSSR model is superior to its constituent models.

Keywords: weak fault detection; controlled symmetry with Woods-Saxon stochastic resonance (CSwWSSR); particle swarm optimization (PSO); signal-to-noise ratio (SNR)

1. Introduction

As the industrialization has found its way into society, the safety and reliability of engineering systems and mechanical equipment play a critical role in the production process, which is an essential subject involving technology and people's livelihoods. It is common knowledge that mechanical failures occur from time to time with the increase in the age of the equipment or the imperfection of the technology in the use of mechanical equipment. Therefore, it is advantageous to find equipment problems as soon as possible because doing so can minimize financial damage and potentially save lives. Yet, in real-world applications, mechanical equipment operates in the presence of very strong background noise, which is manifested as a high frequency of noise overwhelming the presence of the effective signal in the time and frequency domain diagrams. Therefore, the development of novel methods that can identify and extract useful signals under the interference of substantial background noise has become an increasingly important topic in the field of weak fault detection [1,2].

In general, it is believed that the presence of noise will obstruct the detection of effective signals. The previous approaches to weak fault detection, such as filter technique, empirical mode decomposition, singular value decomposition, and wavelet transform, all use a noise suppression viewpoint [3–5]. These signal processing methods based on noise cancellation, however, are inevitably interfering with useful signals that may even lead to compromised detection performance. In their investigation of paleometeorological variability, Benzi et al. initially put forth the stochastic resonance (SR) theory [6–8]. Researchers began to notice the phenomenon of noise-enhanced weak periodic excitation in ring lasers in the following

years, and as a result, SR started to spark a research boom in a wide range of fields. For the first time, Hu et al. utilized SR to mechanical problem diagnostics in [9]. The traditional view of noise suppression has been challenged by the development of SR theory, and the notion of turning noise into valuable signal has captured the interest of many academics and researchers [10].

Over the past two decades, SR has been used in many different fields, including meteorology [11], pharmacology [12], nanomechanics [13–15], and medicine [16]. SR-aided signal processing is particularly useful for wireless communication [17], mechanical fault diagnosis [18], underwater target identification [19], image enhancement [20], and other applications. SR was found in [21] to improve the performance of the system, especially for the case of neural networks whose hidden layer consists of a threshold function. Nishiguchi and Fujiwara demonstrated that weak fault characteristic signals drowned in noise can be detected by a classical bistable stochastic resonance (CBSR) method performed by nano field effect transistors [22]. Li et al. proposed a noise-controlled second-order CBSR system method based on wavelet transform, and successfully implemented this method in wind turbine vibration signals [23]. Liu et al. applied SR to a lake eutrophic ecosystem and discussed in detail the effect of noise, time delay on the state transition of the lake in [24]. To overcome the restriction that CBSR can only be applied to periodic signals, Liu et al. concentrated on applying CBSR to binary pulse amplitude modulation (BPAM) signals [25]. He et al. proposed a detection method for non-collaborative SR for cognitive radio networks, which alleviates the signal-to-noise ratio (SNR) wall and the corresponding noise uncertainty problem of conventional energy detectors, especially in the case of low SNRs [26].

For the sake of practical engineering needs, time-delayed SR systems have attracted the attention of many scholars. In [27], Shi et al. proposed a time-delayed feedback tristable system and analyzed the effects of the time-delayed terms on the signal-to-noise ratio (SNR) and the probability density function (PDF), where they concluded that the time-delay can influence the SR phenomenon. Lu et al. [28] proposed a nonstationary weak signal detection strategy based on a time-delayed feedback SR model and demonstrated that it is applicable to the detection of strong nonlinear nonstationary signals. Mei et al. studied SR with time delay, correlated noise, and periodic signals using SNR theory to determine how time delay affected SR [29]. The SR problem in time-delayed bistable systems subject to Gaussian white noise was covered by He et al. in [30], along with an analysis of the impact of each parameter on the average first crossing time, the Shannon entropy, and the SNR.

The CBSR model has received the most attention among the several theoretical representations of SR. However, CBSR is only effective for small-amplitude signal identification due to the limitations of its own potential function. The frequent output saturation that CBSR encounters, which limits its ability to enhance performance in weak signal detection, is clearly described in [31–34]. In an effort to address this weakness of CBSR, Qiao et al. presented an adaptive unsaturated bistable stochastic resonance approach that, by adopting segmented potential walls with a fixed potential structure in [35] rather than the steep potential walls.

With the innovative development of SR, the model is not only limited to the CBSR, but also the multistable and monostable SR is gradually researched by scholars. Lu et al. investigated a tri-stable stochastic resonance (TSR) system and adaptively adjusted the parameters using particle swarm optimization (PSO) to obtain the best SNR and spectral output in [36]. Through simulation and actual fault signal detection, Lu et al. in [37] demonstrated the superiority of the proposed Woods-Saxon stochastic resonance (WSSR) method over the CBSR. Zhang et al. proposed a combined hybrid TSR system by combining WSSR and CBSR with simulations to demonstrate its effectiveness [38]. However, the TSR suggested in [38] is unable to controllably change the potential structure since the potential well depth and width of the CBSR change simultaneously with a single parameter. The two TSR models mentioned above have a significant impact on the potential structure when a

parameter is changed [36,38]. Hence, building a novel controlled potential structure with an optimal control strategy to catch weak signals remains a difficult problem.

In this paper, we propose a novel SR model called controlled symmetry with Woods-Saxon stochastic resonance (CSwWSSR). Since the parameters of the WSSR model control the potential wells singly and explicitly, we construct a controlled TSR based on the WSSR with three potential wells controlled by different parameters. In order to establish the optimal control strategy with the highest SNR, we then confirm the ideal parameters for the CSwWSSR model in various noise environments by utilizing PSO in the scenarios of controlled potential well depth and controlled potential well width. The feasibility and superiority of CSwWSSR for practical engineering applications are further verified by comparing the results with the original signal and its constituent model through the detection experiments of simulation signal and bearing.

The rest of this essay is structured as follows. The CSwWSSR model and its components are introduced in Section 2, which is then followed by a brief explanation of the procedure for optimizing the CSwWSSR parameters. Section 3 confirms the usefulness of the proposed CSwWSSR-based weak fault detection in simulation signals, and presents an analysis of its performance under various noise intensities and various fault characteristic frequencies. In Section 4, We demonstrate the efficiency of the CSwWSSR model by applying it to the bearing experiments. Finally, we conclude this paper in Section 5.

2. The Proposed CSwWSSR

2.1. Introduction of CSwWSSR

Most of the studies on SR primarily investigated the classical bistable SR model based on the four-time reflection-symmetric potentials. In these models, a change in a single parameter may induce a large change in the shape of the potential function. However, the shape of Woods-Saxon potential illustrated in Equation (1) can be adjusted separately or combinedly with the parameters H , W , and a . The potential well of WSSR is depicted in Figure 1 with steep potential well walls and a smooth bottom. Comparing the four curves in Figure 1, H is to control the depth of the potential well, W is to control the width of the potential well, and a is to determine the steepness of the potential well.

$$U_1(x) = -\frac{H}{1 + \exp(\frac{|x|-W}{a})} \quad (1)$$

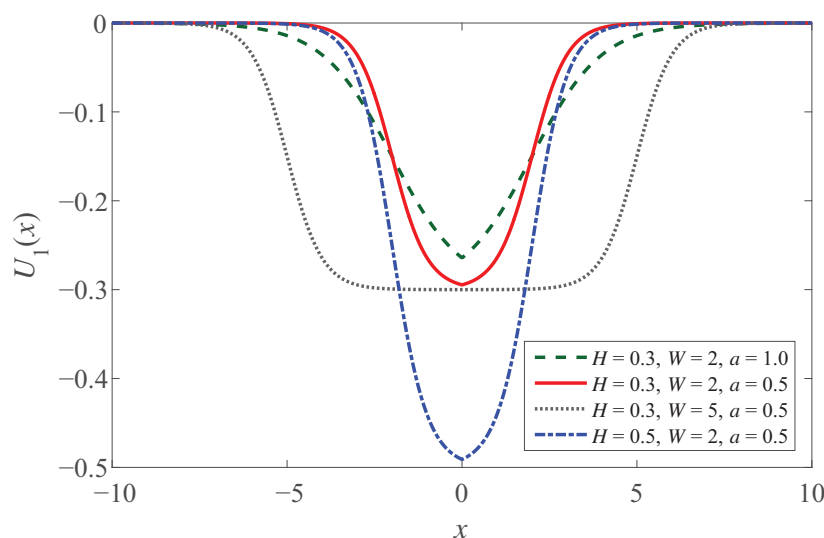


Figure 1. The potential function of WSSR.

Based on the previous analysis of WSSR, we propose a novel stochastic resonance model for weak fault detection, named CSwWSSR, whose potential function is expressed as

$$U(x) = U_1(x) + U_2(x) = -\frac{H}{1 + \exp\left(\frac{|x| - W}{a}\right)} - \frac{1}{2}A\omega_0^2x^2 + \frac{1}{4}B\beta x^4 \quad (2)$$

where $U_1(x)$ indicates the potential function of WSSR as Equation (1), and $U_2(x)$ is the potential function of controlled symmetric stochastic resonance (CSSR), which may be represented as Equation (3). The parameters of $U(x)$ are $(H, W, a, A, B, \omega_0^2, \beta)$, where (H, W, a) are the parameters of the WSSR and $(A, B, \omega_0^2, \beta)$ are the parameters of the CSSR, all of which have real values higher than 0.

$$U_2(x) = -\frac{1}{2}A\omega_0^2x^2 + \frac{1}{4}B\beta x^4 \quad (3)$$

Further, CSSR is defined as the controlled potential well depth and the controlled potential well width. In the case of the controlled well depth, let $A = B = k$. Then, through analyzing the potential function Equation (3), two minima, the potential wells, are located at $x_m = \pm\sqrt{\omega_0^2/\beta}$. The potential barrier is located at $x = 0$, with a height of $k\omega_0^4/4\beta$. So the parameter k is independent of the positions of the potential wells and the potential barrier, which are only related to ω_0^2 and β . For the ease of analysis, we assume that $\omega_0^2 = \beta = 1$, Figure 2 shows the effect of different k on the potential well depth, where k is taken as 0.5, 1, and 1.5, respectively. It is not tough to find that the depth of the two potential wells gradually increases and the walls of the potential wells gradually steepen as k increases [39–41].

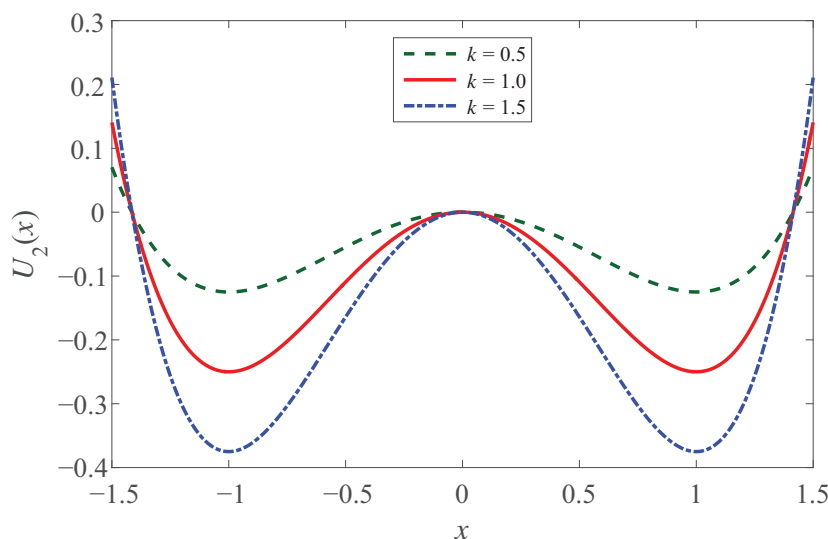


Figure 2. The potential function of CSSR in controlled depth case.

Under the circumstance of the controlled well width, let $A = 1/k^2$, and $B = 1/k^4$. Then, analyzing the potential function Equation (3), the potential barrier of Equation (3) exists at $x = 0$, which is similar to the controlled well depth. The difference is that the potential wells are at $x_m = \pm k\sqrt{\omega_0^2/\beta}$, where the depth of potential well is $\omega_0^4/4\beta$ and the width of the potential well is $k\sqrt{\omega_0^2/\beta}$. It is uncomplicated to note that the depth of the potential well is only related to ω_0^2 and β , and k controls the width of the potential well. Consistent with the above, we assume that $\omega_0^2 = \beta = 1$, Figure 3 shows the effect of different k on the potential well width, where k is taken respectively as 0.5, 1, and 1.5. With the increase of k , the depth of both potential wells remains constant, the width of the potential well increases, and the bottom of the potential well becomes smoother [42–45].

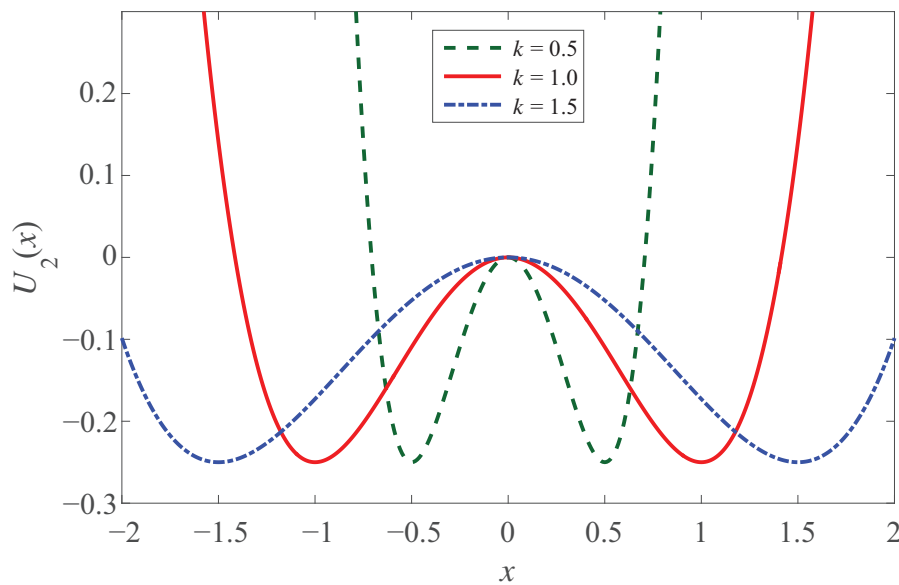


Figure 3. The potential function of CSSR in controlled width case.

Since the potential function of CSwWSSR is generated in the combination of WSSR and CSSR where CSSR has two cases, i.e., the controlled potential well depth and the controlled potential well width, CSwWSSR should also be discussed in two cases. On the basis of the aforementioned analysis of CSSR, we find that ω_0^2 and β do not affect the potential function either in the controlled well depth or the controlled well width, so we assume that $\omega_0^2 = \beta = 1$. Accordingly, the total four parameters controlling CSwWSSR are (H, W, a, k) in any case.

The parameters (H, W, a) of the WSSR control the intermediate potential well of the CSwWSSR, which have the same impact on the potential function as for the WSSR, as shown in Figures 4 and 5. In the case of the controlled potential well depth, the depth of the intermediate potential well gradually increases with increasing H , the width of the well gradually increases with increasing W , and the steepness of the potential well wall gradually becomes slower with increasing a in Figure 4a–c. The potential wells on the left and right sides depend on parameter k of the CSSR, which can be simply realized that the well depth increases with the increase of k in Figure 4d.

For the example of the controlled well width, Figure 5 shows the variation of the CSwWSSR potential function with each parameter. Similar to the controlled well depth, in Figure 5a–c, the depth and width of the intermediate potential well are proportional to the parameters H and W , as well as the steepness of the potential well walls is inversely proportional to the parameter a . Differently, k controls the width of the potential wells on both sides, which is positive to the width in Figure 5d. Comparing Figures 4a and 5a, the potential function is exactly the same for both cases when $k = 1$. Surprisingly, it is noteworthy that the CSwWSSR potential function will resemble that of the bistable state. After the analysis of the CSwWSSR, it contains both the characteristic of steep potential well walls in WSSR that allow for stable particle motion and the characteristic of controlled adjustment of the potential wells on both sides in CSSR. More importantly, the CSwWSSR is strongly adjustable, thus it has a significant impact on the SR efficacy and helps the system produce a strong output.

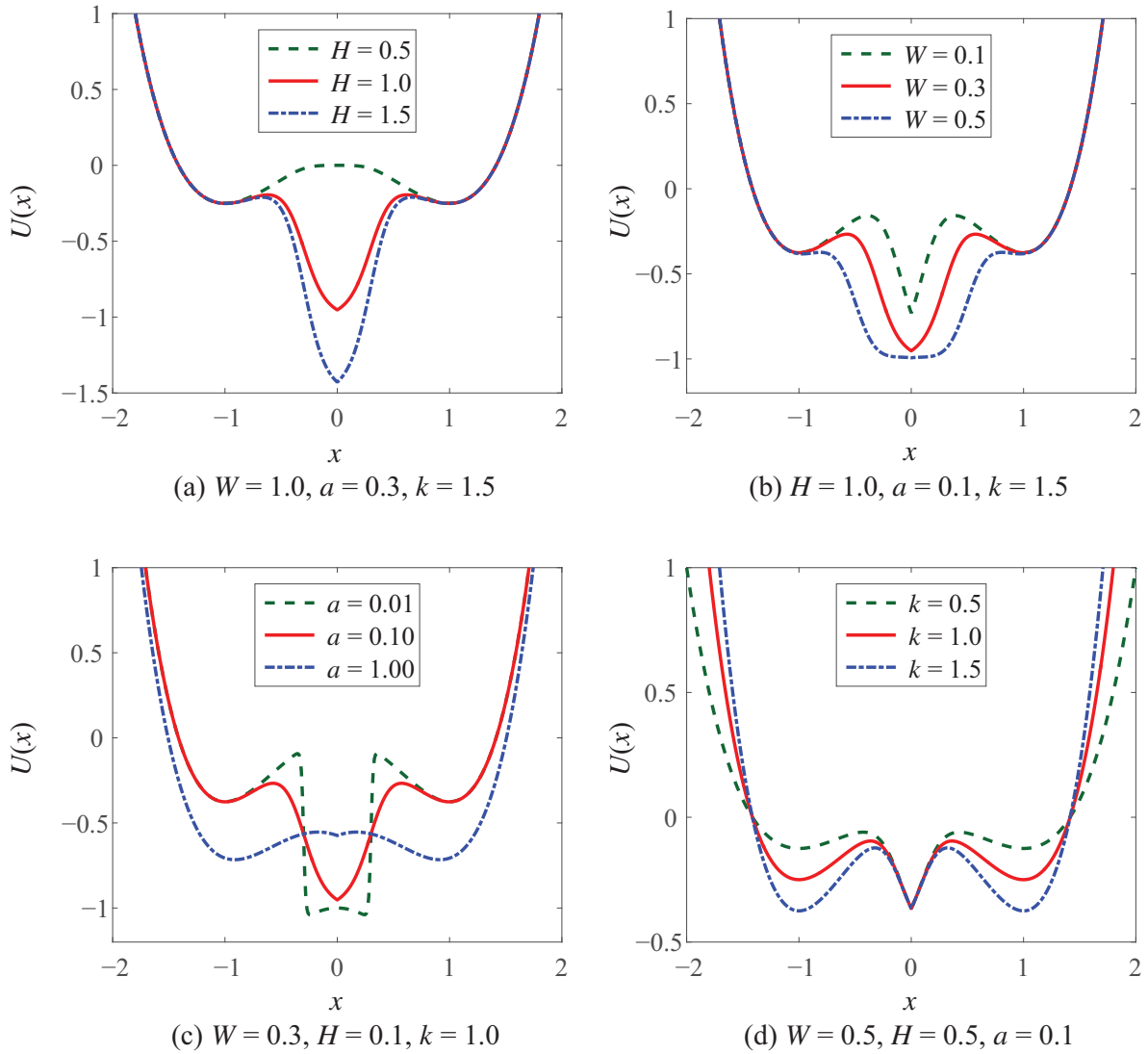


Figure 4. The potential function of CSwWSSR in controlled depth case.

2.2. Optimization of CSwWSSR Parameters

In light of the analysis of the CSwWSSR in Section 2.1, We draw the conclusion that the CSwWSSR depends on four parameters (H, W, a, k). In order to optimize the output, the particle swarm optimization (PSO) [46–48] is used to make the best match between the system parameters and the input signal. Figure 6 provides an illustration of the CSwWSSR model for weak fault detection using PSO. Here, the objective function of the PSO is output SNR from [35], which is defined as follows

$$SNR = 10\log_{10}\left(\frac{A_d}{\sum_{i=1}^{N/2} A_i - A_d}\right) \quad (4)$$

where A_d denotes the amplitude of the fault characteristic frequency in the output power spectrum, $\sum_{i=1}^{N/2} A_i - A_d$ represents the power amplitude sum of noise in the output power spectrum, and N stands for the length of the signal. A larger output SNR indicates superior performance in weak signal detection.

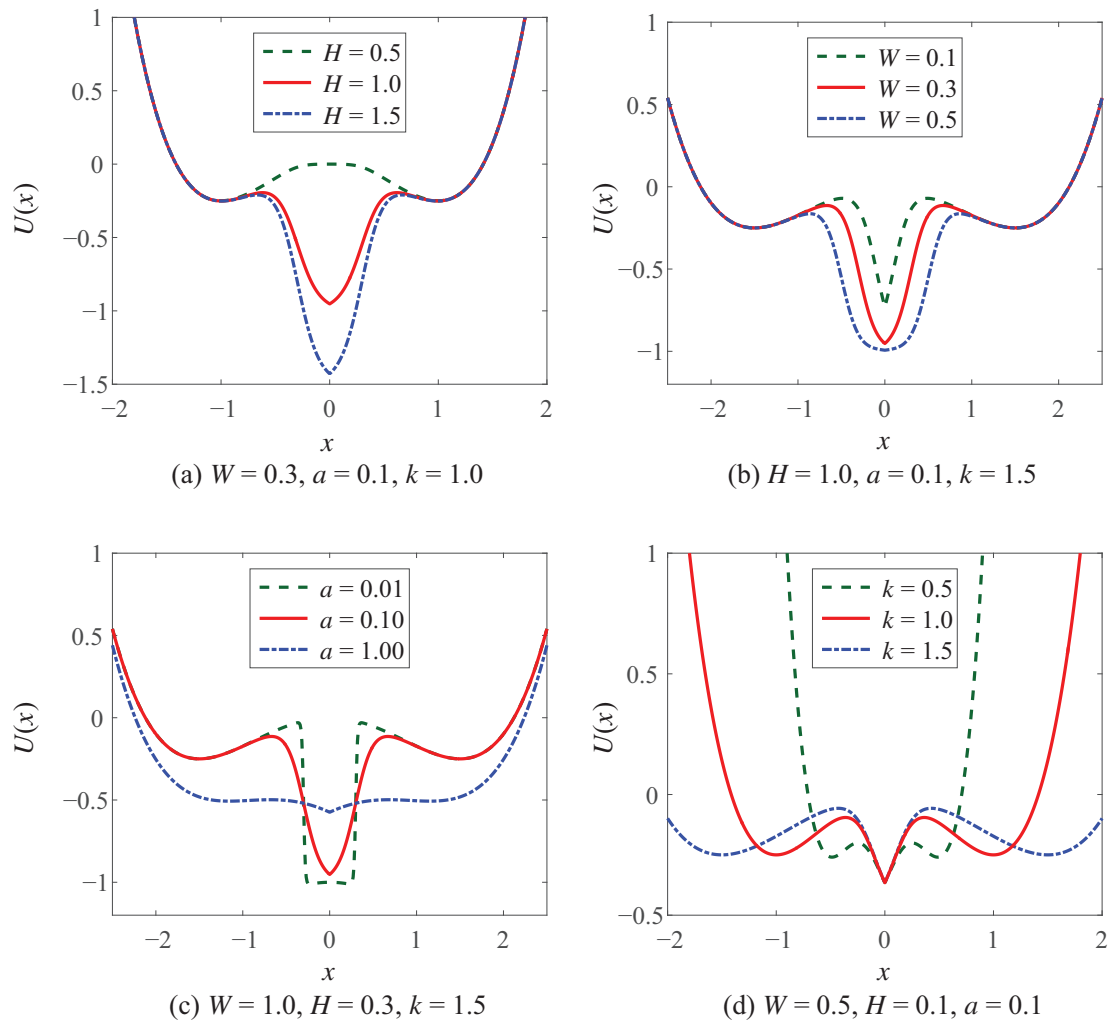


Figure 5. The potential function of CSwWSSR in controlled width case.

Step 1. Preprocess input signal. The adiabatic approximation theory, which stipulates that the driving frequency, signal amplitude, and noise intensity should all be considerably less than 1, was the basis for the initial proposal of SR. However, the majority of real signals are high frequency signals in the presence of severe noise, which does not agree with the theory of adiabatic approximation. A rescaling method was used in [27] to circumvent the restriction that SR is only applied to small parameter signals in an effort to more effectively apply it to numerous fields. The specific method first scales down the high frequency signal to get it into the adiabatic approximation theory's narrow parameter region, where SR is possible. The compressed signal is then transformed using SR, and finally the output signal is converted back to a high frequency signal with the same scale.

Step 2. Optimize parameters. The PSO process is shown in the dotted box in Figure 6. First, we need to initialize the PSO, including the population size, the maximum generations of iterations, the learning factor, the initial position of particles, the individual optimal position, the population optimal position, etc. Then, the fitness corresponding to the current position is calculated, i.e., the output SNR. In order to prevent the algorithm from entering a local optimum, the step to update the particle positions involves the use of a dynamic inertia factor whose value is correlated with the number of current iterations [49]. Afterwards, the fitness of the updated position is evaluated against the population and individual optimums. The updated position will take the place of the optimal position when the updated fitness is outperforming the individual optimal fitness and the population optimal fitness. The population optimal position is produced after multiple iterations.

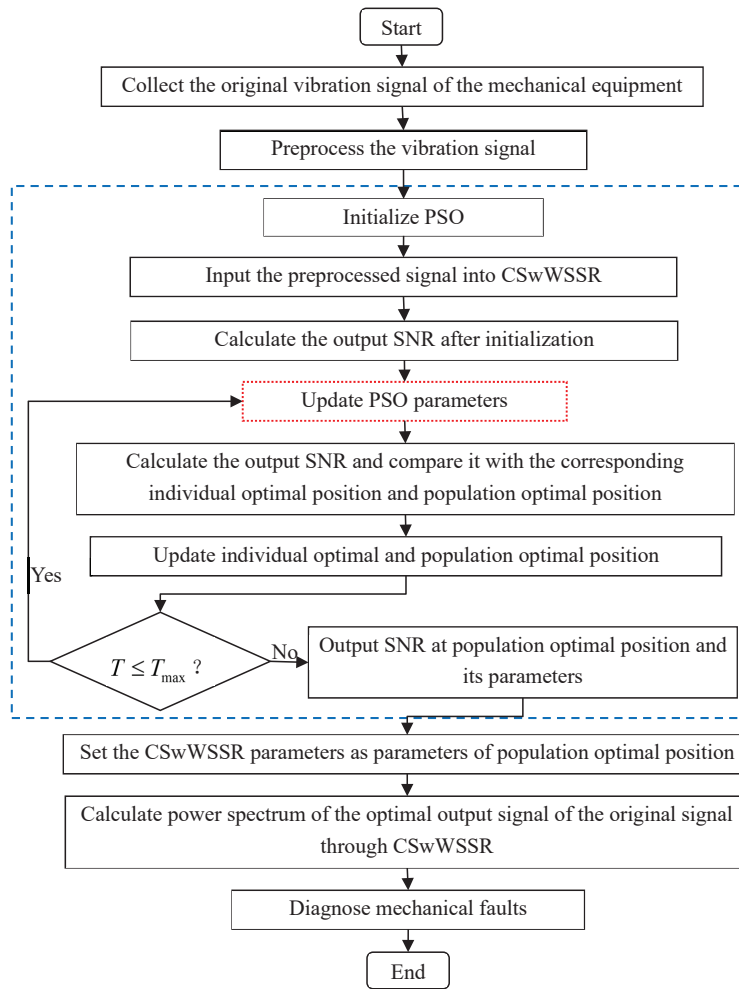


Figure 6. Process flow of CSwWSSR for weak fault detection through PSO.

Step 3. Calculate the CSwWSSR. The optimal combination of parameters output in the previous step is substituted into CSwWSSR to calculate the output signal. Due to the fact that the SR is based on a phenomenon generated by a nonlinear system, the analytical and approximate analytical methods are no longer applicable. As a consequence, a numerical method is usually used to solve the problem, i.e., fourth-order Runge-Kutta [36]. The solution method is as follows

$$\begin{cases} k_1 = \Delta * func_CSwWSSR(x_{in}(i), x_{out}(i), H, W, a, k) \\ k_2 = \Delta * func_CSwWSSR(x_{in}(i), x_{out}(i) + k_1/2, H, W, a, k) \\ k_3 = \Delta * func_CSwWSSR(x_{in}(i+1), x_{out}(i) + k_2/2, H, W, a, k) \\ k_4 = \Delta * func_CSwWSSR(x_{in}(i+1), x_{out}(i) + k_3, H, W, a, k) \\ x_{out}(i+1) = x_{in}(i) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \end{cases} \quad (5)$$

in which $x_{in}(i)$ is the i -th sample value of the input signal after summing the vibration signal and the noise signal, and $x_{out}(i)$ is the i -th sample value of the output signal. Δ is employed to represent the step size, which is scaled from $1/f_s$ to $1/f_{s\partial} = 1/(f_s/\partial)$ when using the rescaling method (∂ is scale compression ratio, f_s is the frequency of sample, $f_{s\partial}$ is two sampling frequency). $func_CSwWSSR()$ stands for the Langevin equation for CSwWSSR, expressed as follows

$$\begin{cases} func_CSwWSSR = \frac{dx}{dt} = -\frac{dU(x)}{dt} + s(t) + \zeta(t) \\ \frac{dU(x)}{dt} = \frac{H}{a} \operatorname{sgn}(x) \exp\left(\frac{|x|-W}{a}\right) \left(1 + \exp\left(\frac{|x|-W}{a}\right)\right)^{-2} - A\omega_0^2 x + B\beta x^3 \end{cases} \quad (6)$$

where $s(t)$ and $\xi(t)$ are vibration signal and noise respectively, $(H, W, a, A, B, \omega_0^2, \beta)$ are the parameters of CSwWSSR. We analyzed in Section 2.1 to know that ω_0^2 and β have no effect on the potential structure and set $\omega_0^2 = \beta = 1$. Both A and B can be represented by k in different cases, so the $func_CSwWSSR()$ only contains four parameters (H, W, a, k) .

Step 4. Diagnose faults. A power spectrum is created from the output signal via the Fourier transform to detect the weak fault and diagnose mechanical faults. In this work, the output SNR and the amplitude of the power spectrum at the fault characteristic frequency are observed and compared to assess the model's outstanding performance.

3. Simulation

3.1. Output Analysis of Simulation Signal

To illustrate the effectiveness of the proposed CSwWSSR model, the simulation signal is imported into the CSwWSSR model's weak signal detection for fault diagnosis. The simulation signal can be represented as

$$x(t) = A \sin(2\pi f_0 t) + \xi(t) \quad (7)$$

where $\xi(t)$ is additive Gaussian white noise with $\langle \xi(t)\xi(t+\tau) \rangle = 2D\xi(\tau)$ where noise intensity $D = 5$ and $\langle \bullet \rangle$ represent statistical average operator. A denotes the amplitude of the simulation signal with $A = 0.25$. To be specific, the fault characteristic frequency $f_0 = 140$ Hz, the sampling frequency $f_s = 10,000$ Hz, and the data length $N = 4000$. To satisfy the requirement of adiabatic approximation theory to achieve small parameters of SR, the scale compression ratio $\partial = 2000$, then the step size $\Delta = 1/(f_s/\partial) = 0.2$. Figure 7 displays the simulation signal's time domain waveform and power spectrum. It is obvious that the input signal without SR processing is completely drowned in the surrounding noise at the fault characteristic frequency $f_0 = 140$ Hz in the frequency domain diagram, so detecting the fault of simulation signal is challenging. Moreover, the waveform of the signal is not visible in Figure 7a.

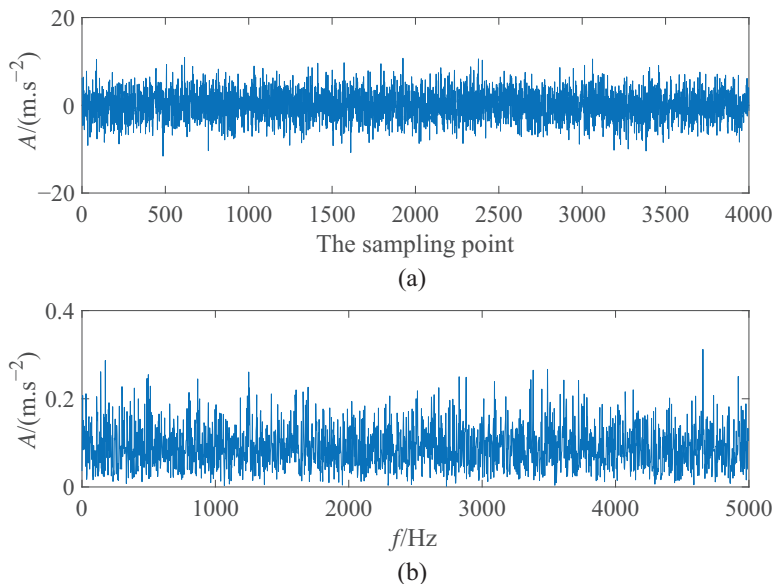


Figure 7. The input signal: (a) time domain diagram and (b) frequency domain diagram.

Aiming to demonstrate the fault characteristic detection capability of CSwWSSR, three different models are used to process the above simulation signal, that is WSSR, CSSR, and CSwWSSR. In addition, since CSSR and CSwWSSR are divided into controlled well depth and controlled well width, the input signal is subjected to five experiments. For the sake of fairness, the PSO is used for all five experiments to optimize parameters, where the population size and the number of iterations are both set to 50. After the PSO, Table 1

lists the optimized parameter combinations and output SNR, and Figure 8 provides their corresponding output signal time domain diagram and frequency domain diagram.

Table 1. Parameter combination and output SNR of each model after optimization of simulation signal.

Model	H	W	a	k	SNR/dB
WSSR	8.5624	6.2607	3.5634	-	-10.16
CSSR-depth	-	-	-	0.0417	-10.85
CSSR-width	-	-	-	1.6886	-11.63
CSwWSSR-depth	5.5987	5.1328	0.6898	0.0015	-9.863
CSwWSSR-width	1.6607	2.3355	0.1399	3.3592	-9.777

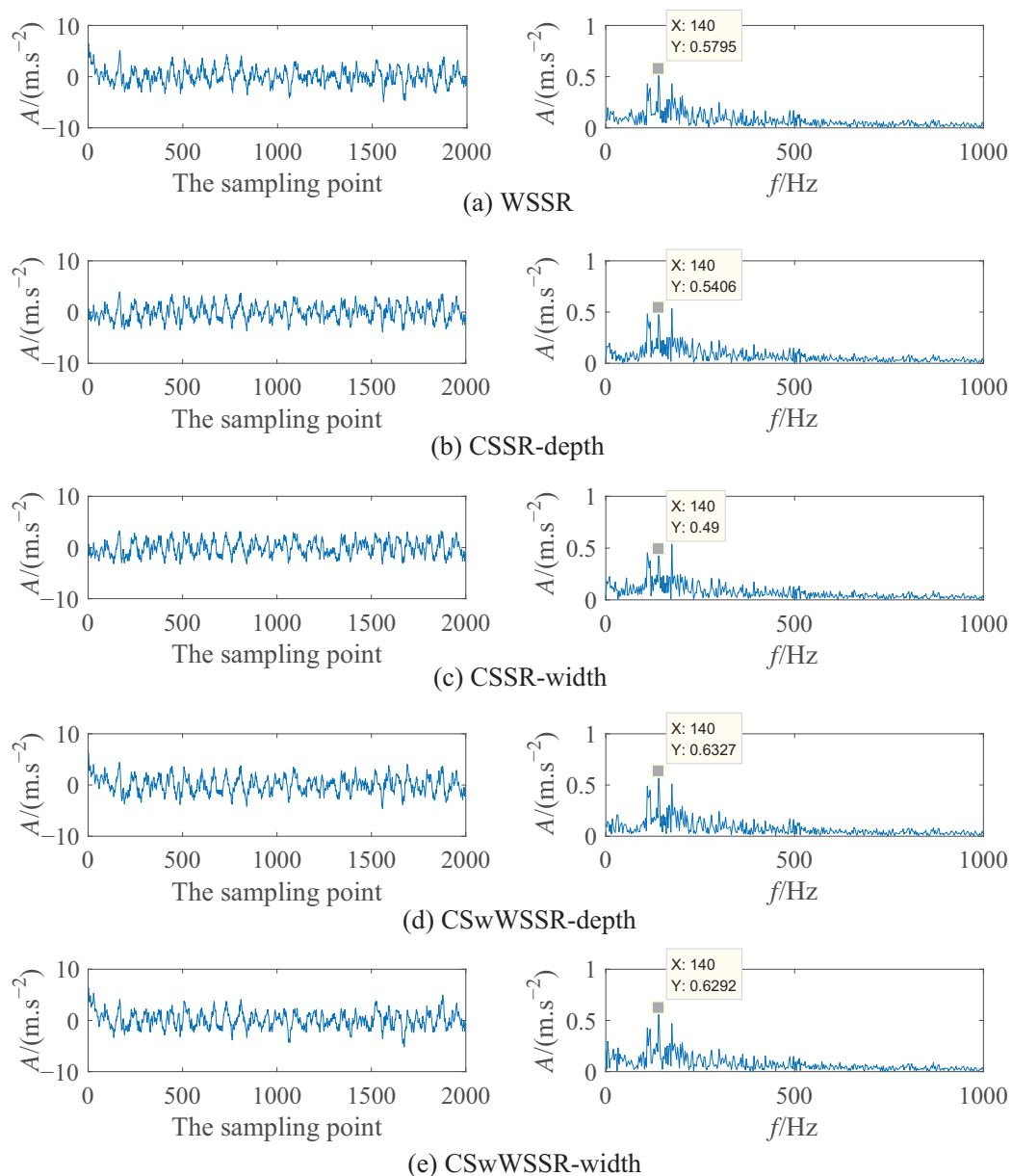


Figure 8. The output signal time-frequency diagram of each model in simulation signal. (**Left** column: Time domain diagram; **Right** column: Frequency domain diagram).

From Table 1, in view of the output SNR of the above simulation signal, it is seen clearly that the output SNR is: -9.777 dB (CSwWSSR-width) $>$ -9.863 dB (CSwWSSR-depth) $>$

-10.16 dB (WSSR) > -10.85 dB (CSSR-depth) > -11.63 dB (CSSR-width). In Figure 8a, the amplitude of the output signal after WSSR processing at the fault characteristic frequency of 140 Hz is 0.5795. From Figure 8b,c, the amplitude of the output signal after CSSR processing at the fault characteristic frequency is 0.5406 and 0.49. It can be viewed that the amplitude of the output signal at the fault characteristic frequency after CSwWSSR processing is 0.6327 and 0.6292 in Figure 8d,e. The amplitude of fault characteristic frequency is 0.6327 (CSwWSSR-depth) > 0.6292 (CSwWSSR-width) > 0.5795 (WSSR) > 0.5406 (CSSR-depth) > 0.49 (CSSR-width). So the fault characteristic extraction capability is CSwWSSR $>$ WSSR $>$ CSSR. Furthermore, Figure 8b,c show that the fault characteristic frequency is more significantly influenced by the neighboring frequencies, which is evident from the fact that the amplitude of the non-fault characteristic frequency is nearly similar to the fault characteristic frequency after CSSR processing. Based on the analysis hereinbefore, the method of using the CSwWSSR model to implement weak fault detection for simulation signal appeared more suitable.

3.2. Capability of Detecting Different Simulation Signals

To evaluate the capability of the proposed CSwWSSR model in detecting different fault signals, it is demonstrated by a series of simulation signals with different fault characteristic frequencies and noise intensities. Consistent with the above, Equation (7) is used as the input signal by simply changing its fault characteristic frequency f_0 and noise intensity D , in which the fault characteristic frequency f_0 is from 0.05 to 1500 Hz and the noise intensity D is from 1 to 6. In Figure 9a, let the sampling points $N = 4000$, the signal amplitude $A = 0.25$, the sampling frequency $f_s = 10,000$ Hz, the fault characteristic frequency $f_0 = 140$ Hz, and the rescaling ratio $\partial = 1000$. In Figure 9b, let the sampling points $N = 4000$, the signal amplitude $A = 0.2$, the noise intensity $D = 2$, and the sampling frequency $f_s = f_0 * 100$. To meet the adiabatic approximation theory, the rescaling ratio $\partial = f_0 * 8$. As before, PSO was used to find the best parameter match with a population size and iteration number of 50 and 10, respectively. The average of the five final output SNRs was taken as the output SNR of the current signal. Figure 9 shows the variation of SNR with D and f_0 under three different models.

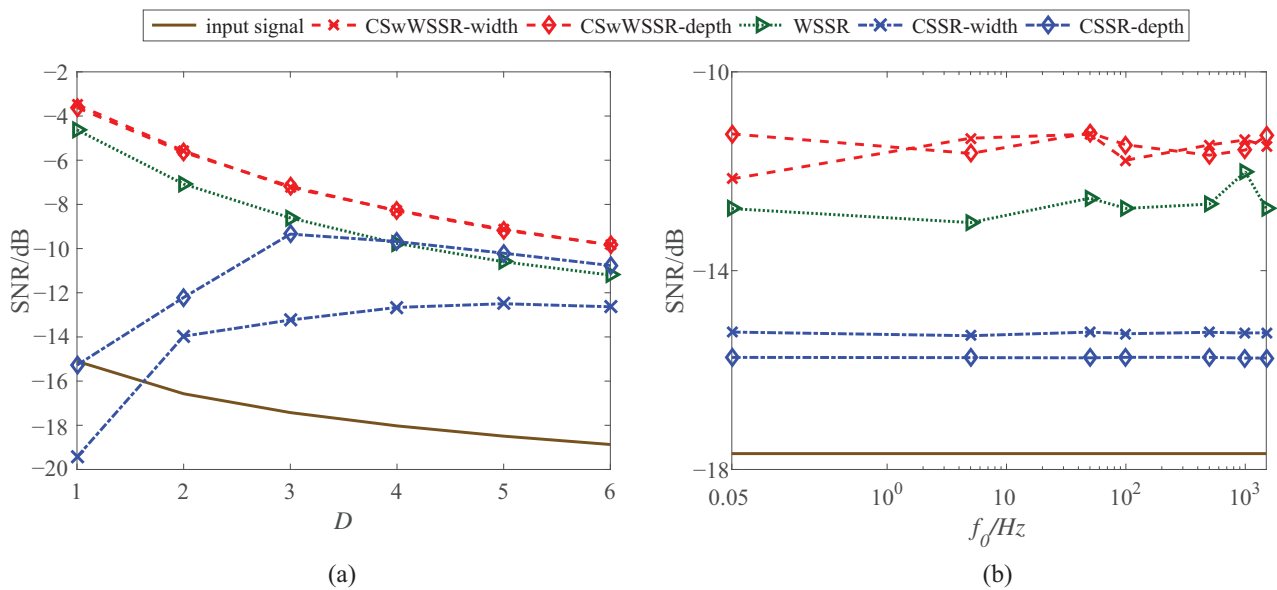


Figure 9. The variation of SNR with (a) D and (b) f_0 .

We can apparently conclude CSwWSSR is preferable to the other two models in terms of its ability to detect fault signals, whether anti-noise or enhancement of high frequency signals in Figure 9. On the one hand, the improvement of CSSR by CSwWSSR is particularly

significant, as can be seen that CSSR has no improvement effect on the input signal at low noise intensities, which is manifested by its output SNR being lower than the input SNR at low noise intensities. Although the enhanced performance of WSSR at high noise intensities is noticeably poorer and even worse than the output SNR of CSSR, it still outperforms input SNR in general. With the adequacies of CSSR and WSSR combined, CSwWSSR has an excellent ability to improve for both high and low noise intensities. On the other hand, the curves of the CSwWSSR are always higher than those of the other two models in Figure 9b for signals of various frequencies. Focusing on the above comparison, the suggested CSwWSSR shows more enhancement capability than CSSR and WSSR for the same input signal, different fault characteristic frequencies and noise intensities.

4. Experimental Demonstration of Bearing

In the previous section, we demonstrated the effectiveness of the CSwWSSR model for simulation signals. Furthermore, it is applied to practical engineering applications in this section to demonstrate effectiveness of the CSwWSSR model. Here, we adopt the data of rolling bearing provided by Case Western Reserve University, whose sampling frequency is 12 kHz, and speed is 1796 r/min. Taking the outer loop fault as an example, the theoretical value of the outer loop fault frequency is calculated by Equation (8) which is 107.30 Hz. The parameters needed to calculate the fault frequency are shown in Table 2.

$$f_i = \frac{n_1 N_1}{120} \left(1 - \frac{d_1 \cos \beta_1}{D_1} \right) \quad (8)$$

Table 2. The main structure parameters of the tested bearing 6205-2RSJEM.

Inner Diameter/mm	Outer Diameter/mm	D_1 /mm	n_1	d_1 /mm	Contact Angle/(°)
25.001	51.999	39.040	9.000	7.940	0

In Equation (8), n_1 represents the number of rolling elements, N_1 stands for the rotation rate, d_1 is employed to denote the rolling element diameter, β_1 is the bearing contact angle, and D_1 represents the pitch diameter. In this way, set the sampling point $N = 3000$ and the noise intensity $D = 4$. Since the fault characteristic frequency does not fulfill the adiabatic approximation theory, using a rescaling method is necessary, with two sampling frequency $f_{sd} = 6$ Hz.

Figure 10 plots the time domain waveform and spectrum of the input signal. It is shown that, with the absence of noise, the spectrum of the input signal is concentrated at 2000–4000 Hz, which resulted in the failure to detect the fault characteristic frequency. Furthermore, the time domain diagram contains waveforms of periodic shocks and vibrations. To accomplish weak fault detection, additive noise is added to the input signal with the purpose of generating the SR phenomenon. As in Section 3.1, we use PSO to achieve the best match between the signal and the parameters. The best matches for the five cases and their corresponding output SNRs are given in Table 3. Comparing the output SNR for the five cases in Table 3, it is clear that output SNR: -8.996 dB (CSwWSSR-width) > -9.702 dB (CSwWSSR-depth) > -9.813 dB (CSSR-depth) > -9.911 dB (CSSR-width) > -11.21 dB (WSSR).

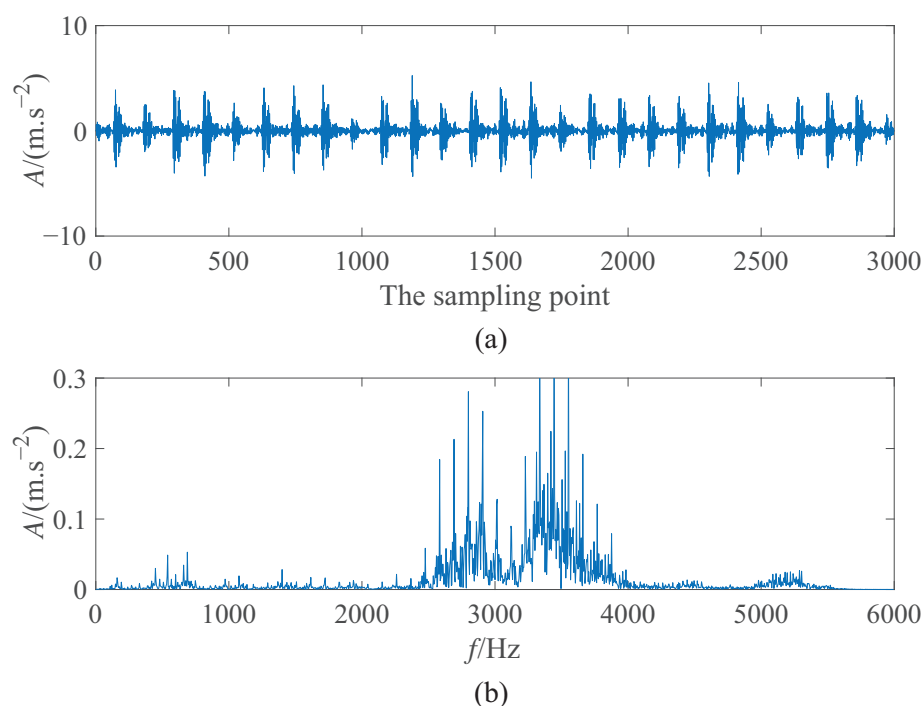


Figure 10. The input signal: (a) time domain diagram and (b) frequency domain diagram.

Table 3. Parameter combination and output SNR of each model after optimization of bearing.

Model	H	W	a	k	SNR/dB
WSSR	8.2734	3.8796	0.7568	-	−11.21
CSSR-depth	-	-	-	0.2897	−9.813
CSSR-width	-	-	-	1.1389	−9.911
CSwWSSR-depth	3.4026	3.4026	0.0367	0.1631	−9.702
CSwWSSR-width	3.2067	2.2868	0.0164	1.7698	−8.996

Figure 11 illustrates the time domain and frequency domain diagrams of the output signals for the five cases at the best match. 108 Hz is not hard to obtain from Figure 11, which is close to the fault characteristic frequency 107.30 Hz. In Figure 11a, the amplitude of the output signal after WSSR processing at the fault characteristic frequency of 108 Hz is 0.5423. From Figure 11b,c, the amplitude of the output signal after CSSR processing at the fault characteristic frequency is 0.4721 and 0.4431. It can be viewed that the amplitude of the output signal at the fault characteristic frequency after CSwWSSR processing is 0.5092 and 0.6009 in Figure 11d,e. The relationship between the amplitude at fault characteristic frequency in the five cases is clear: 0.6009 (CSwWSSR-width) > 0.5421 (WSSR) > 0.5092 (CSwWSSR-depth) > 0.4721 (CSSR-depth) > 0.4431 (CSSR-width). Although the fault characteristic frequency amplitude of WSSR is higher than this of CSwWSSR-depth, it is influenced by the low frequency, whose amplitude is almost equal to that of the fault characteristic frequency in Figure 11a. Taking into account, the fault characteristic extraction capability: CSwWSSR > CSSR > WSSR. Unlike the result in Section 3.1, CSSR outperforms WSSR in practical engineering applications. Consistently, CSwWSSR exhibits the excellent capability of weak fault detection. On account of the above analysis. CSwWSSR is also preferable to CSSR and WSSR in terms of effectiveness in practical engineering applications.

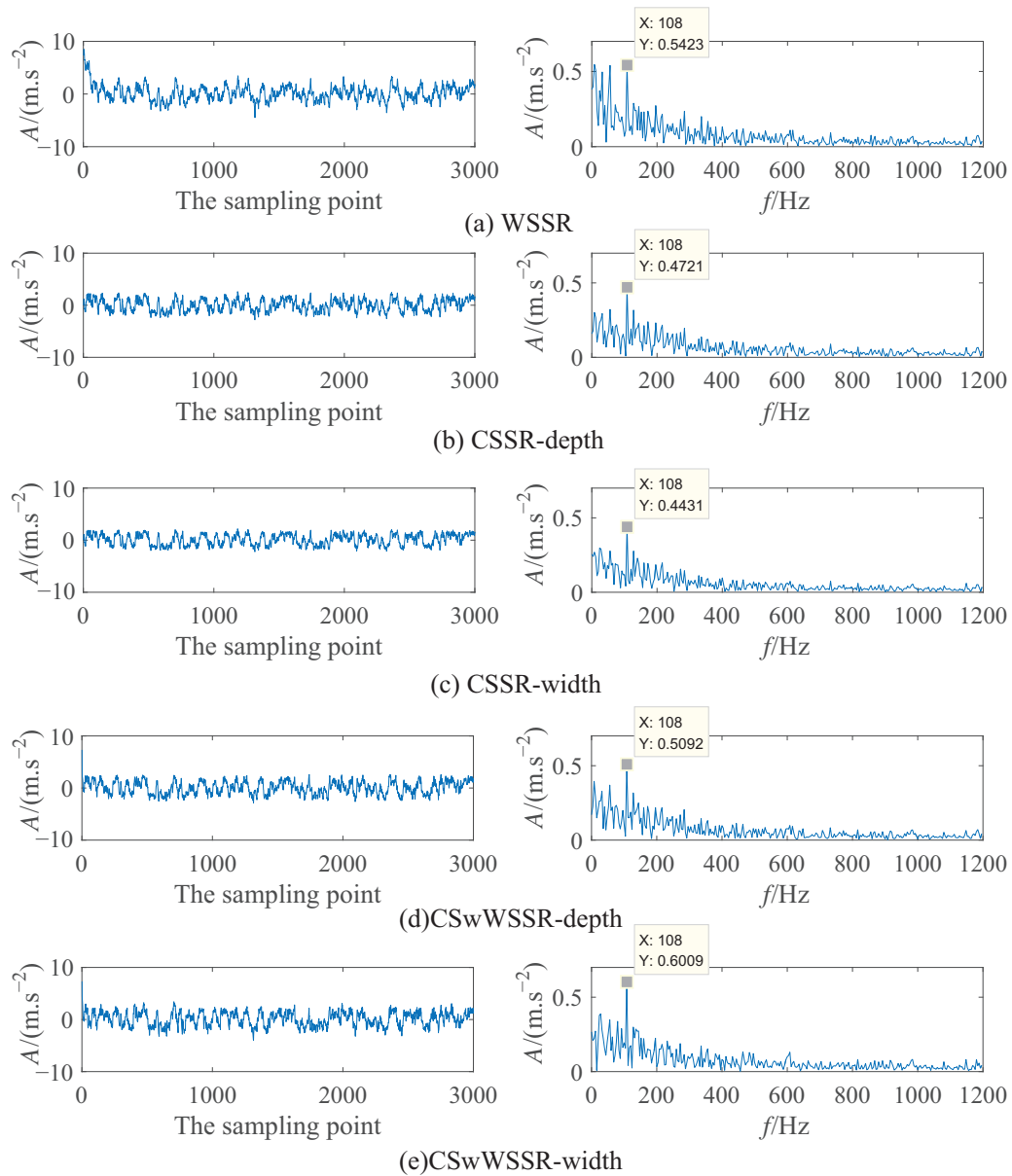


Figure 11. The output signal time-frequency diagram of each model in simulation signal. (**Left** column: Time domain diagram; **Right** column: Frequency domain diagram).

5. Conclusions

Linear optimal signal processing is typically less effective with a higher SNR in weak signal detection. However, because weak fault detection is regularly hindered by strong noise, the CSwWSSR model we proposed in this paper exploits noise to enhance weak signal detection. In addition to excellent noise utilization, it also exhibits notable features at the fault characteristic frequency, which is one of the benefits of combining WSSR and CSSR. What's more, by modifying the parameters, it may visibly alter the potential function's shape directly. The model parameters with the best SNR may be found for different input signals and noise thanks to an optimization algorithm. Then, comparative experiments are conducted for the controlled potential well depth and controlled potential well width to demonstrate the viability and superiority of the CSwWSSR model in fault characteristic detection. The following conclusions are drawn:

1. We analyze the potential functions of WSSR, CSSR, and CSwWSSR to investigate the effects of parameters on the potential structure and dynamical properties. The

intermediate potential well of the CSwWSSR is discovered to be controlled by the parameters (H , W , a), which have the same impact on the potential structure as the WSSR, and the potential wells on both sides depend on the parameter k . The CSwWSSR parameters are transparent to the control of the potential structure, which means that each parameter controls a different aspect of the potential structure.

2. The output SNR curves of WSSR, CSSR, and CSwWSSR were examined for various noise intensities and various fault characteristic frequencies, demonstrating how well CSwWSSR combines the benefits of WSSR and CSSR both in terms of anti-noise and enhancement of high frequency signals. It has both the features of stable particle motion in WSSR and the controlled adjustment of potential wells on both sides in CSSR.
3. We compare the output signal time domain and frequency domain diagrams of the WSSR, CSSR, and CSwWSSR after PSO optimized parameters in the simulation signal and bearing experiment. In terms of output SNR and amplitude of fault characteristic frequency, it is apparent that CSwWSSR outperforms CSSR and WSSR, which indicates that CSwWSSR has robust engineering applicability in the future.

In future work, we will consider the optimization problem of two-dimensional signals based on multistable stochastic resonance. In the low SNR environment, it will be a new challenge to make full use of SR techniques to overcome the spatial correlation of two-dimensional signals and recover two-dimensional signals disturbed by environmental noise. It is obvious that the results obtained in this study make a good foundation for the discussed future work.

Author Contributions: Methodology, J.L.; Validation, J.G.; Resources, B.H.; Data curation, C.T.; Writing—original draft, J.G.; Writing—review & editing, J.L.; Visualization, J.G. and W.Z.; Project administration, Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is financially supported National Natural Science Foundation of China (62001210, 62201284), Natural Science Foundation of Jiangsu (BK20190789).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gupta, P.; Pradhan, M.K. Fault detection analysis in rolling element bearing: A review. *Mater. Today Proc.* **2017**, *4*, 2085–2094. [CrossRef]
2. Wei, Y.; Li, Y.; Xu, M.; Huang, W. A review of early fault diagnosis approaches and their applications in rotating machinery. *Entropy* **2019**, *21*, 409. [CrossRef] [PubMed]
3. Duan, F.; Pan, Y.; Chapeau-Blondeau, F.; Abbott, D. Noise benefits in combined nonlinear bayesian estimators. *IEEE Trans. Signal Process.* **2019**, *67*, 4611–4623. [CrossRef]
4. Liu, J.; Hu, B.; Yang, F.; Zang, C.; Ding, X. Stochastic Resonance in a delay-controlled dissipative bistable potential for weak signal enhancement. *Commun. Nonlinear Sci. Numer. Simul.* **2020**, *85*, 105245. [CrossRef]
5. Qiao, Z.; Pan, Z. SVD principle analysis and fault diagnosis for bearings based on the correlation coefficient. *Meas. Sci. Technol.* **2015**, *26*, 085014. [CrossRef]
6. Xu, M.; Zheng, C.; Sun, K.; Xu, L.; Qiao, Z.; Lai, Z. Stochastic resonance with parameter estimation for enhancing unknown compound fault detection of bearings. *Sensors* **2023**, *23*, 3860. [CrossRef] [PubMed]
7. Benzi, R.; Sutera, A.; Vulpiani, A. The mechanism of stochastic resonance. *J. Phys. Math. Gen.* **1981**, *14*, L453–L457. [CrossRef]
8. Lu, S.; He, Q.; Wang, J. A review of stochastic resonance in rotating machine fault detection. *Sensors* **2019**, *116*, 230–260. [CrossRef]
9. Hu, N.; Chen, M.; Wen, X. The application of stochastic resonance theory for early detecting rub-impact fault of rotor system. *Mech. Syst. Signal Process.* **2003**, *17*, 883–895. [CrossRef]
10. Gammaitoni, L.; Hänggi, P.; Jung, P.; Marchesoni, F. Stochastic resonance: A remarkable idea that changed our perception of noise. *Eur. Phys. J. B* **2009**, *69*, 1–3. [CrossRef]
11. Caccamo, M.T.; Magazù, S. A physical-mathematical approach to climate change effects through stochastic resonance. *Climate* **2019**, *7*, 21. [CrossRef]

12. Wang, W.; Xiang, S.; Xie, S.; Xiang, B. An adaptive single-well stochastic resonance algorithm applied to trace analysis of Clenbuterol in human urine. *Molecules* **2012**, *17*, 1929–1938. [CrossRef] [PubMed]
13. Markina, A.; Muratov, A.; Petrovskyy, V.; Avetisov, V. Detection of single molecules using stochastic resonance of bistable oligomers. *Nanomaterials* **2020**, *10*, 2519. [CrossRef] [PubMed]
14. Zhang, Y.; Wang, Z.; Tazeddinova, D.; Ebrahimi, F.; Habibi, M.; Safarpour, H. Enhancing active vibration control performances in a smart rotary sandwich thick nanostructure conveying viscous fluid flow by a PD controller. *Waves Random Complex Media* **2021**, *33*, 1–24. [CrossRef]
15. Zare, R.; Najaafi, N.; Habibi, M.; Ebrahimi, F.; Safarpour, H. Influence of imperfection on the smart control frequency characteristics of a cylindrical sensor-actuator GPLRC cylindrical shell using a proportional-derivative smart controller. *Smart Struct. Syst.* **2020**, *26*, 469–480. [CrossRef]
16. Ko, L.-W.; Chikara, R.K.; Chen, P.-Y.; Jheng, Y.-C.; Wang, C.-C.; Yang, Y.-C.; Li, L.P.-H.; Liao, K.-K.; Chou, L.-W.; Kao, C.-L. Noisy galvanic vestibular stimulation (stochastic resonance) changes electroencephalography activities and postural control in patients with bilateral vestibular hypofunction. *Brain Sci.* **2020**, *10*, 740. [CrossRef] [PubMed]
17. Li, Q.; Li, Z. A novel sequential spectrum sensing method in cognitive radio using suprathreshold stochastic resonance. *IEEE Trans. Veh. Technol.* **2013**, *63*, 1717–1725. [CrossRef]
18. Qiao, Z.; Lei, Y.; Li, N. Applications of stochastic resonance to machinery fault detection: A review and tutorial. *Mech. Syst. Signal Process.* **2019**, *122*, 502–536. [CrossRef]
19. Dong, H.; He, K.; Shen, X.; Ma, S.; Wang, H.; Qiao, C. Adaptive intrawell matched stochastic resonance with a potential constraint aided line enhancer for passive sonars. *Sensors* **2020**, *20*, 3269. [CrossRef]
20. Liu, J.; Hu, B.; Wang, Y. Optimum adaptive array stochastic resonance in noisy grayscale image restoration. *Phys. Lett. A* **2019**, *383*, 1457–1465. [CrossRef]
21. Ikemoto, S.; DallaLibera, F.; Hosoda, K. Noise-modulated neural networks as an application of stochastic resonance. *Neurocomputing* **2018**, *277*, 29–37. [CrossRef]
22. Nishiguchi, K.; Fujiwara, A. Detecting signals buried in noise via nanowire transistors using stochastic resonance. *Appl. Phys. Lett.* **2012**, *101*, 193108. [CrossRef]
23. Li, J.; Chen, X.; Du, Z.; Fang, Z.; He, Z. A new noise-controlled second-order enhanced stochastic resonance method with its application in wind turbine drivetrain fault diagnosis. *Renew. Energy* **2013**, *60*, 7–19. [CrossRef]
24. Liu, J.; Guo, J.; Ding, X.; Qiao, Z.; Zang, C. Colored correlated noises induced regime shifts in a time-delayed lake eutrophication ecosystem. *Front. Sustain. Dev.* **2022**, *2*, 24–37. [CrossRef]
25. Liu, J.; Li, Z.; Guan, L.; Pan, L. A novel parameter-tuned stochastic resonator for binary PAM signal processing at low SNR. *IEEE Commun. Lett.* **2014**, *18*, 427–430. [CrossRef]
26. He, D.; Chen, X.; Pei, L.; Jiang, L.; Yu, W. Improvement of noise uncertainty and signal-to-noise ratio wall in spectrum sensing based on optimal stochastic resonance. *Sensors* **2019**, *19*, 841. [CrossRef]
27. Shi, P.; Yuan, D.; Han, D.; Zhang, Y.; Fu, R. Stochastic resonance in a time-delayed feedback tristable system and its application in fault diagnosis. *J. Sound Vib.* **2018**, *424*, 1–14. [CrossRef]
28. Lu, S.; He, Q.; Zhang, H.; Kong, F. Enhanced rotating machine fault diagnosis based on time-delayed feedback stochastic resonance. *J. Vib. Acoust.* **2015**, *137*, 051008. [CrossRef]
29. Mei, D.C.; Du, L.C.; Wang, C.J. The effects of time delay on stochastic resonance in a bistable system with correlated noises. *J. Stat. Phys.* **2009**, *137*, 625–638. [CrossRef]
30. He, M.; Xu, W.; Sun, Z. Dynamical complexity and stochastic resonance in a bistable system with time delay. *Nonlinear Dyn.* **2015**, *79*, 1787–1795. [CrossRef]
31. Liao, Z.; Wang, Z.; Yamahara, H.; Tabata, H. Echo state network activation function based on bistable stochastic resonance. *Chaos Solitons Fractals* **2021**, *153*, 111503. [CrossRef]
32. Lai, Z.H.; Wang, S.B.; Zhang, G.Q.; Zhang, C.L.; Zhang, J.W. Rolling bearing fault diagnosis based on adaptive multiparameter-adjusting bistable stochastic resonance. *Shock Vib.* **2020**, *2020*, 1–15. [CrossRef]
33. Liao, Z.; Wang, Z.; Yamahara, H.; Tabata, H. Low-power-consumption physical reservoir computing model based on overdamped bistable stochastic resonance system. *Neurocomputing* **2022**, *468*, 137–147. [CrossRef]
34. Zhang, X.; Wang, J.; Liu, Z.; Wang, J. Weak feature enhancement in machinery fault diagnosis using empirical wavelet transform and an improved adaptive bistable stochastic resonance. *ISA Trans.* **2019**, *84*, 283–295. [CrossRef] [PubMed]
35. Qiao, Z.; Lei, Y.; Lin, J.; Jia, F. An adaptive unsaturated bistable stochastic resonance method and its application in mechanical fault diagnosis. *Mech. Syst. Signal Process.* **2017**, *84*, 731–746. [CrossRef]
36. Lu, J.; Huang, M.; Yang, J.-J. A novel spectrum sensing method based on tri-stable stochastic resonance and quantum particle swarm optimization. *Wirel. Pers. Commun.* **2017**, *95*, 2635–2647. [CrossRef]
37. Lu, S.; He, Q.; Kong, F. Stochastic resonance with Woods-Saxon potential for rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* **2014**, *45*, 488–503. [CrossRef]
38. Zhang, G.; Li, H. Hybrid tri-stable stochastic resonance system used for fault signal detection. *J. Vib. Shock* **2019**, *38*, 9–17.
39. Khan, M.B.; Santos-García, G.; Noor, M.A.; Soliman, M.S. Some new concepts related to fuzzy fractional calculus for up and down convex fuzzy-number valued functions and inequalities. *Chaos Solitons Fractals* **2022**, *164*, 112692. [CrossRef]

40. Khan, M.B.; Santos-García, G.; Noor, M.A.; Soliman, M.S. New Hermite–Hadamard inequalities for convex fuzzy-number-valued mappings via fuzzy riemann integrals. *Mathematics* **2022**, *10*, 3251. [CrossRef]
41. Khan, M.B.; Othman, H.A.; Santos-García, G.; Saeed, T.; Soliman, M.S. On fuzzy fractional integral operators having exponential kernels and related certain inequalities for exponential trigonometric convex fuzzy-number valued mappings. *Chaos Solitons Fractals* **2023**, *169*, 113274. [CrossRef]
42. Macías-Díaz, J.E.; Khan, M.B.; Noor, M.A.; Abd Allah, A.M.; Alghamdi, S.M. Hermite-Hadamard inequalities for generalized convex functions in interval-valued calculus. *AIMS Math.* **2022**, *7*, 4266–4292. [CrossRef]
43. Khan, M.B.; Treanta, S.; Soliman, M.S. Generalized preinvex interval-valued functions and related Hermite–Hadamard type inequalities. *Symmetry* **2022**, *14*, 1901. [CrossRef]
44. Khan, M.B.; Noor, M.A.; Noor, K.I.; Chu, Y.-M. Higher-order strongly preinvex fuzzy mappings and fuzzy mixed variational-like inequalities. *Int. J. Comput. Intell. Syst.* **2021**, *14*, 1856. [CrossRef]
45. Khan, M.B.; Noor, M.A.; Noor, K.I.; Chu, Y.-M. New Hermite–Hadamard-type inequalities for -convex fuzzy-interval-valued functions. *Adv. Differ. Equ.* **2021**, *149*, 1–20. [CrossRef]
46. Wang, D.; Tan, D.; Liu, L. Particle swarm optimization algorithm: An overview. *Soft Comput.* **2018**, *22*, 387–408. [CrossRef]
47. Xia, X.; Gui, L.; Yu, F.; Wu, H.; Wei, B.; Zhang, Y.; Zhan, Z. Triple archives particle swarm optimization. *IEEE Trans. Cybern.* **2019**, *50*, 4862–4875. [CrossRef]
48. Shami, T.M.; El-Saleh, A.A.; Alswaitti, M.; Al-Tashi, Q.; Summakieh, M.A.; Mirjalili, S. Particle swarm optimization: A comprehensive survey. *IEEE Access.* **2022**, *10*, 10031–10061. [CrossRef]
49. Tian, D.; Shi, Z. MPSO: Modified particle swarm optimization and its applications. *Swarm Evol. Comput.* **2018**, *41*, 49–68. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

High-Performance Adaptive Weak Fault Diagnosis Based on the Global Parameter Optimization Model of a Cascaded Stochastic Resonance System

Zhihui Lai ^{1,2}, Zhangjun Huang ^{1,2}, Min Xu ³, Chen Wang ^{1,2}, Junchen Xu ^{1,2}, Cailiang Zhang ⁴, Ronghua Zhu ^{4,*} and Zijian Qiao ^{5,*}

- ¹ Shenzhen Key Laboratory of High Performance Nontraditional Manufacturing, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China; laizh@szu.edu.cn (Z.L.)
² Guangdong Key Laboratory of Electromagnetic Control and Intelligent Robots, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China
³ Ningbo Cigarette Factory, China Tobacco Zhejiang Industry Co., Ltd., Ningbo 315040, China
⁴ Ocean College, Zhejiang University, Zhoushan 316021, China
⁵ School of Mechanical Engineering and Mechanics, Ningbo University, Ningbo 315211, China
* Correspondence: zhu.richard@zju.edu.cn (R.Z.); qiaozijian@nbu.edu.cn (Z.Q.)

Abstract: Stochastic resonance (SR), as a type of noise-assisted signal processing method, has been widely applied in weak signal detection and mechanical weak fault diagnosis. In order to further improve the weak signal detection performance of SR-based approaches and realize high-performance weak fault diagnosis, a global parameter optimization (GPO) model of a cascaded SR system is proposed in this work. The cascaded SR systems, which involve multiple multi-parameter-adjusting SR systems with both bistable and tri-stable potential functions, are first introduced. The fixed-parameter optimization (FPO) model and the GPO models of the cascaded systems to achieve optimal SR outputs are proposed based on the particle swarm optimization (PSO) algorithm. Simulated results show that the GPO model is capable of achieving a better SR output compared to the FPO model with rather good robustness and stability in detecting low signal-to-noise ratio (SNR) weak signals, and the tri-stable cascaded SR system has a better weak signal detection performance compared to the bistable cascaded SR system. Furthermore, the weak fault diagnosis approach based on the GPO model of the tri-stable cascaded system is proposed, and two rolling bearing weak fault diagnosis experiments are performed, thus verifying the effectiveness of the proposed approach in high-performance adaptive weak fault diagnosis.

Keywords: weak fault diagnosis; stochastic resonance; cascaded system; global parameter optimization; particle swarm optimization algorithm

1. Introduction

Condition monitoring and fault diagnosis of rotating mechanical components, such as bearings, gears, and rotors, are important to ensure the safe operation of equipment [1]. Vibration analysis and magnetic flux leakage detection [2] are common and effective technologies in the condition monitoring and fault diagnosis of machinery. In many cases, especially when the machine operates in poor working conditions or the sensors have to be installed far away from the fault source, the fault characteristics are always submerged in strong background noise and other interference components, making them difficult to be identified. Therefore, many weak signal detection methods, such as spectral kurtosis (SK) [3], empirical mode decomposition (EMD) [4,5], digital filter [6], wavelet transform (WT) [7], etc., have been widely investigated to extract weak fault signals from noisy backgrounds. However, these weak signal detection methods inevitably weaken the fault characteristic when filtering out the noise; hence, the weak signal detection performance

is limited. Different from the conventional weak signal detection methods, stochastic resonance (SR)-based methods can utilize the noise (rather than eliminate the noise) to achieve weak signal detection, thus showing potential in weak fault diagnosis.

The concept of SR was first proposed by Benzi et al. [8,9] in 1981 when they studied the paleo-meteorological problem of alternating glacial and warm climate cycles. When SR occurs, a part of the noise energy is transferred to the low-frequency signal through a nonlinear system. Thus, the low-frequency weak signal can be significantly enhanced while filtering out the high-frequency noise, and the output signal-to-noise ratio (SNR) is significantly improved [10,11]. Due to the advantages of SR in weak signal detection, SR has attracted continuous attention in the field of weak fault diagnosis in the past 40 years [12–17].

In order to make SR-based methods suitable for practical engineering applications of mechanical fault diagnosis, many efforts have been made by researchers. To overcome the limitation of conventional SR that it can only process a signal with small parameters (e.g., a signal frequency that is much less than 1 Hz) due to the limitation of the adiabatic approximation theory [18], researchers have proposed many solutions to make SR capable of processing large-parameter signals, such as frequency re-scaling SR (FRSR) [19], frequency-shifted and re-scaling SR (FSRSR) [20], parameter-normalized SR (PNSR) [21], modulated SR (MSR) [22], and scale-transformation SR [23], among others. To appropriately select the system parameters to generate the optimal SR output for given signals, many adaptive parameter-tuning SR methods have been investigated [10,18,24], which are further applied in the fault diagnosis of planetary gearboxes [24], rolling bearing fault diagnosis [25,26], and rotor misalignment fault diagnosis [27].

The most important objective of SR-based fault diagnosis methods is to achieve a high-performance weak signal detection ability. The weak signal detection ability of an SR-based method is related to the potential function of the system [12]. It was reported that SR can be produced in nonlinear systems with monostable, bistable, tri-stable, and multi-stable potential functions [23,28–31]. Moreover, many novel forms of potential functions, such as the Woods–Saxon potential function proposed by Gao et al. [32], the pinning potential function employed by Zhang et al. [33], the piecewise linear potential function employed by Li et al. [34], and the compound tri-stable potential function proposed by Tang et al. [1], have been proposed recently for the realization of SR and the successful application of weak signal detection. However, how to quantitatively and reasonably evaluate the detection performance of an SR system with different potential functions is still an issue that needs to be solved.

Moreover, the weak signal detection ability of a single SR system is still limited [35], especially for weak signals with extremely low SNRs. Hence, some enhancement models based on the SR systems, such as the cascaded SR systems [36–38], coupled SR systems [29,39], and parallel SR systems [40,41], have been constructed to further enhance the weak signal detection ability of SR-based methods. Among these enhancement models, the cascaded SR models, which consist of several single SR systems connected in series, are of most interest. It is known that a single SR system acts as a low-pass filter that can transfer the energy from a high-frequency region to a low-frequency region, thus amplifying the low-frequency weak signal [42]. Hence, multiple filtering processes can be realized through a cascaded SR system, thus making the output signal smoother and making the characteristic signal more obvious. Therefore, the cascaded SR system may have a better weak signal detection performance compared to a single SR system.

Compared to a single SR system, the number of adjustable parameters of a cascaded system is significantly enhanced, and different parameter-adjusting strategies have been proposed for a cascaded system to realize SR output. Up to now, the so-called “fixed-parameter adjusting strategy” and “stage-by-stage parameter adjusting strategy” have been the most adopted. The fixed-parameter strategy means that the corresponding parameters in each subsystem are kept the same. Using this parameter-adjusting strategy, Hairan et al. [43] proposed a signal detection method under a dual-mode non-Gaussian noise

background based on a cascaded SR. This parameter-adjusting strategy can be further combined with a multi-parameter adaptive optimization algorithm to find out the optimal SR output adaptively [44]. The stage-by-stage parameter-adjusting strategy means the optimal SR output of each subsystem is obtained by adjusting the parameters of this subsystem, and its SR output is input into the next subsystem (if necessary). Therefore, the parameters of each subsystem may be different. Using this strategy, Xiao et al. [45] proposed an adaptive cascaded vibration resonance method and applied it to rotating machinery fault diagnosis; Liu et al. [46] proposed an adaptive cascaded piecewise linear SR system for bearing fault diagnosis; Xu et al. [37] proposed an adaptive cascaded SR method using the quantum particle swarm optimization (QPSO) algorithm; and Zhang et al. [47] proposed a cascaded three-steady-state SR and applied it to rolling bearing fault diagnosis. However, it is noted that in most of the literature, this strategy is adopted for a cascaded system with only two subsystems, as the output SNR cannot continue to increase with more subsystems. Hence, the fixed-parameter-adjusting strategy can produce a better SR output compared to the stage-by-stage parameter-adjusting strategy. Furthermore, it can be easily understood that as the cascaded system is not treated as a system as a whole, the optimal SR output of the cascaded system cannot be found by either the “fixed-parameter adjusting strategy” or the “stage-by-stage parameter adjusting strategy”. In contrast, a novel “global-parameter adjusting strategy”, which means all parameters of all subsystems are adjusted jointly, has the potential to realize the optimal SR output for the whole cascaded system. This parameter-adjusting strategy, which has not been fully studied yet, is worthy to be further investigated.

Therefore, in this paper, the global parameter optimization model of a cascaded SR system is proposed for high-performance adaptive weak signal detection and weak fault diagnosis. As it can achieve the theoretically optimal SR output of a cascaded system, the output SNR of this model can be also used to quantitatively evaluate the detection performance of SR systems with different potential functions. In this paper, the classical bistable and tri-stable potential functions are considered for demonstration. The rest of the paper is organized as follows. In Section 2, the background knowledge of the multi-parameter-adjusting SR is introduced, and both the fixed-parameter optimization (FPO) model and the global parameter optimization (GPO) model of cascaded SR systems are constructed with the help of the particle swarm optimization (PSO) algorithm. The weak signal detection performance of the single SR system and the FPO model and the GPO model of the cascaded systems with either bistable or tri-stable potential functions is fully studied and compared through numerical simulations in Section 3. Two rolling bearing weak fault diagnosis experiments are performed in Section 4 to demonstrate the effectiveness of the proposed GPO model of a cascaded system in weak signal detection and weak fault diagnosis. Conclusions are drawn in Section 5.

2. Weak Signal Detection Approaches Based on Adaptive Cascaded SR Systems

2.1. Multi-Parameter-Adjusting SR

SR is a nonlinear physical phenomenon reflecting the synergistic effect of the nonlinear system, the weak periodic signal, and the noise. When SR occurs, the weak periodic signal is enhanced by gaining energy from high-frequency noise. The mathematical model of an underdamped SR system, which describes the motion of a unit-mass Brownian particle in a potential field in the presence of periodic force and noise, can be written as:

$$\frac{dx^2(t)}{dt^2} + k\frac{dx(t)}{dt} = -\frac{dU(x)}{dx} + s(t) + n(t) \quad (1)$$

Here, k denotes the damping ratio; $U(x)$ denotes the potential field of the system, and $-dU(x)/dx$ is the potential field force applied to the particle; $s(t) = A\cos(2\pi f_0 t)$ represents a harmonic signal with amplitude A and frequency f_0 ; and $n(t) = \sqrt{2D}\xi(t)$ represents noise with an intensity of D , where $\xi(t)$ is zero-mean and unit-variance Gaussian

white noise. In this system, $sn(t) = s(t) + n(t)$ is defined as the input signal, and $x(t)$ denotes the output signal, which can be obtained by numerically solving the equation.

The nonlinear system shown in Equation (1) can generate SR with many different potential fields. Among these, the bistable potential field and the tri-stable potential field are two classical potential fields, whose generalized potential functions can be written as follows [28,48]:

$$U_B(x) = -\frac{1}{2}a_1x^2 + \frac{1}{4}b_1x^4 \quad (2)$$

$$U_T(x) = \frac{1}{2}a_2x^2 - \frac{1}{4}b_2x^4 + \frac{1}{6}c_2x^6 \quad (3)$$

where the potential field parameters a_1, b_1, a_2, b_2 , and c_2 all have positive values.

Examples of these two potential functions are plotted in Figure 1. It can be seen that the bistable potential field is symmetrical with two potential wells, and the tri-stable one is symmetrical with three potential wells.

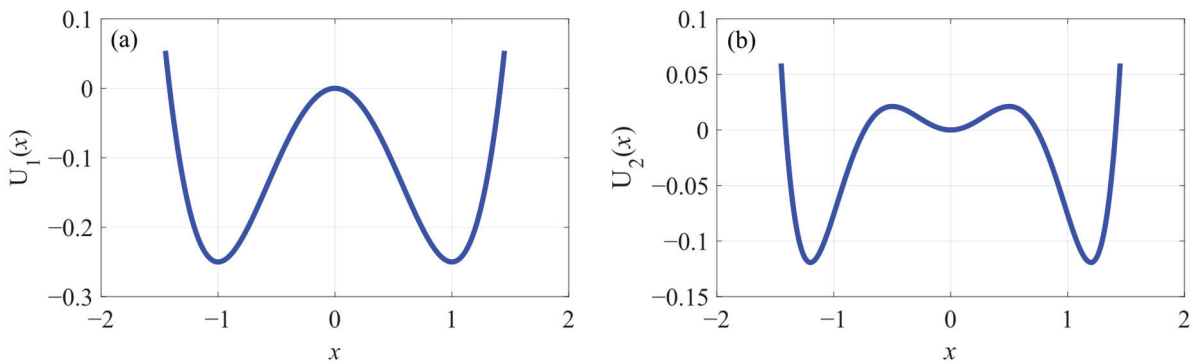


Figure 1. Potential functions of (a) the bistable potential field with $a_1 = 1$ and $b_1 = 1$ and (b) the tri-stable potential field with $a_2 = 0.36$, $b_2 = 1.69$, and $c_2 = 1.69$.

The output signal $x(t)$ of the SR system can be understood as the displacement of the Brownian particle moving in the potential field driven by the periodic signal and the random noise. When the signal is weak and noise is absent, i.e., $A < A_C$ and $D = 0$ (A_C indicates the critical amplitude for the particle to pass over the potential barrier to another potential well [48]), the Brownian particle can only oscillate within one of the potential wells. When noise is present ($D > 0$), the particle may pass over the barrier even though the signal is weak, as the particle can accumulate energy from high-frequency noise with the help of the nonlinear system. When the transition rate of the particle between potential wells accords with the signal period, the particle can realize regular inter-well oscillation, and SR occurs. Hence, the weak features of the input signal can be enhanced through SR and can be identified from the output signal $x(t)$. This advantage makes SR an advanced weak signal detection model, which has been applied in many fields including weak fault diagnosis.

Due to the limitation of the adiabatic approximation theory [23], the classical SR system shown in Equation (1) is only applicable to small-parameter conditions ($A < A_C$, $f_0 \ll 1$, and $D \ll 1$), which limits its applications in practical engineering. To solve this issue, a multi-parameter-adjusting SR system is further proposed by introducing an amplitude-transformation coefficient ε and a scale-transformation coefficient R to Equation (1), which can be written as [23]:

$$\frac{dx^2(t')}{dt'^2} + k \frac{dx(t')}{dt'} = -\frac{dU(x)}{dx} + \varepsilon \cdot sn(t') \quad (4)$$

Here, ε is used to linearly transform the amplitude of the input signal to be in an appropriate small-parameter range, and $t' = Rt$ is the times series after scale transformation,

where R is used to transform the frequency/time scale of the input signal to meet the small-parameter conditions. The scale transformation can be simply realized by adopting a time step of $h' = R/F_s$ instead of $h = 1/F_s$ in the calculation when numerically solving the mathematic models of the SR systems [23], where F_s is the sampling frequency of the input signal. Thus, the frequency of the characteristic signal can be regarded as $f_0' = f_0/R$ in the calculations, and the small-frequency condition of $f_0' \ll 1$ can be realized by appropriately adjusting the value of R . Hence, SR can be achieved under large-parameter conditions using the multi-parameter-adjusting SR system.

2.2. Models of Cascaded SR Systems

The SR system has shown its potential in weak signal detection, but its detection ability for a weak signal with an extremely low SNR is still limited, which should be further improved. One possible solution is to construct a cascaded SR system, which represents multiple SR systems connected in series, as shown in Figure 2. Note that the number of connected subsystems is defined as N in this work. Previous research shows that as the low-frequency weak signal can continuously accumulate energy from high-frequency signal and noise through multiple SR systems, the cascaded SR system has better output performance compared to the single SR system, such as a higher output SNR, a more obvious signal spectral peak, and a smoother output signal resulting from multiple high-frequency filterings [35].

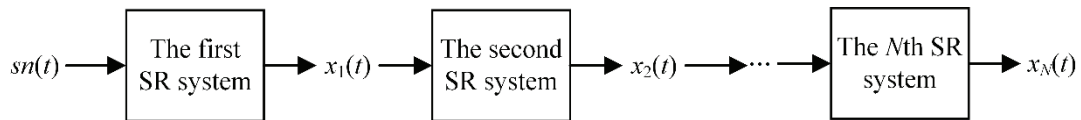


Figure 2. Framework of a cascaded SR system.

In order to utilize the cascade SR system to detect a large-parameter weak signal, the multi-parameters (k , ε , R , and the potential field parameters) of each subsystem should be optimized to achieve an optimal SR output. For this purpose, the multi-parameters can be optimized through different optimization models of the cascaded SR system. The first one is termed the fixed-parameter optimization (FPO) model of the cascade SR system, as shown in Equation (5), which has been studied by many scholars previously [44,49]. One can see that the multi-parameters in all subsystems stay the same. Thus, the cascaded bistable SR system only has five parameters (k , a_1 , b_1 , ε , and R) to be optimized, and the cascaded tri-stable SR system only has six parameters (k , a_2 , b_2 , c_2 , ε , and R) to be optimized. In the optimization process, the objective of the FPO model is to realize the optimal SR output of the last SR system by optimizing these limited parameters. Therefore, the optimization process based on the FPO model is relatively simple due to the small quantity of the parameters to be optimized.

$$\begin{cases} \frac{dx_1^2(t')}{dt'^2} + k \frac{dx_1(t')}{dt'} = -\frac{dU(x_1)}{dx_1} + \varepsilon \cdot sn(t') \\ \frac{dx_2^2(t')}{dt'^2} + k \frac{dx_2(t')}{dt'} = -\frac{dU(x_2)}{dx_2} + x_1(t') \\ \dots \\ \frac{dx_N^2(t')}{dt'^2} + k \frac{dx_N(t')}{dt'} = -\frac{dU(x_N)}{dx_N} + x_{N-1}(t') \end{cases} \quad (5)$$

In order to further improve the SR output of the cascaded SR system, a novel optimization model of the cascaded SR system is proposed in this work, which is termed the global parameter optimization (GPO) model, as shown in Equation (6). It can be seen that in this model, the multi-parameters of each subsystem are independent. In the optimization process, the objective is to realize the optimal SR output of the last subsystem by optimizing all multi-parameters jointly, and the number of parameters to be optimized is decided by the number of connected subsystems (N). It is easy to understand that the optimized result

obtained from the FPO model is one of the special cases of the results obtained from the GPO model. Therefore, the proposed GPO model has the potential to achieve a better SR output compared to the FPO model.

$$\begin{cases} \frac{dx_1^2(t')}{dt'^2} + k_1 \frac{dx_1(t')}{dt'} = -\frac{dU_1(x_1)}{dx_1} + \varepsilon \cdot sn(t') \\ \frac{dx_2^2(t')}{dt'^2} + k_2 \frac{dx_2(t')}{dt'} = -\frac{dU_2(x_2)}{dx_2} + x_1(t') \\ \dots \\ \frac{dx_N^2(t')}{dt'^2} + k_N \frac{dx_N(t')}{dt'} = -\frac{dU_N(x_N)}{dx_N} + x_{N-1}(t') \end{cases} \quad (6)$$

2.3. Optimization of the Cascaded SR System Based on the PSO Algorithm

To achieve the optimal SR output of both the FPO model and the GPO model of the cascaded system containing many adjustable parameters, all these parameters should be optimized through a multi-parameter optimization algorithm, which should be carefully selected to match the optimization objective. The particle swarm optimization (PSO) algorithm is a global algorithm based on the foraging behavior of birds. This method requires fewer parameters to adjust compared with other algorithms such as the genetic algorithm and the mosquito swarm, making it easier to implement and simplify the algorithm [28]. Therefore, the PSO algorithm is adopted in this work to find out the optimal SR output of both the FPO model and the GPO model of the cascaded system. The optimization process is briefly introduced in this subsection.

In order to evaluate the output performance of the cascaded SR system appropriately, the output SNR (SNR_{out}) of the system is first defined as the evaluation index [28]:

$$SNR_{out} = 10 \log_{10} \left(\frac{P_0^2}{\sum [X(f)]^2 - P_0^2} \right) \quad (7)$$

where $X(f)$ represents the single-side spectrum of the output signal of the single or cascade SR system and P_0 is the amplitude value of the output signal at the characteristic frequency. It is assumed that the optimal SR output of the single or cascaded system is achieved when SNR_{out} reaches its maximum value. Hence, SNR_{out} can be also used as the objective function of the PSO algorithm.

The PSO algorithm can be described as follows: assuming a particle swarm is a population of n particles in a Z -dimensional search space, The position and velocity of the i th particle are defined as $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{iZ})^T$ and $V_i = (V_{i1}, V_{i2}, V_{i3}, \dots, V_{iZ})^T$, respectively. Initially, random position and velocity populations are created. In each iteration, the particles update their own by tracking two optimal solutions: one is the individual optimal solution of each particle $p_i = (p_{i1}, p_{i2}, p_{i3}, \dots, p_{iZ})$, and the other one is the population optimal solution of all particles $p_g = (p_{g1}, p_{g2}, p_{g3}, \dots, p_{gZ})$ [10]. Then, the positions and velocities are updated according to the iterative formulas; thus, the individual and population optimal solution can be updated accordingly at each iteration. The optimization process is finished when the number of iterations reaches the pre-set value I_{max} , and the final population extreme value and the corresponding position of the particle can be obtained. In optimizing the cascaded system, the final population extreme value indicates the optimal SNR_{out} of the system, and the optimal parameter indicates the population optimal solution. Moreover, the iterative formulas of the velocity and position are given as:

$$V_{i,j}(t+1) = \omega V_{i,j}(t) + c_1 r_1 [p_{i,j} - x_{i,j}(t)] + c_2 r_2 [p_{g,j} - x_{i,j}(t)] \quad (8)$$

$$x_{i,j}(t+1) = x_{i,j}(t) + V_{i,j}(t+1), j = 1, \dots, Z \quad (9)$$

where c_1 and c_2 are positive learning factors or acceleration constants and r_1 and r_2 are random numbers uniformly distributed within $0 \sim 1$. In order to effectively control the flight velocity of the particle, an inertia weight ω is also introduced. In addition, by setting the

velocity range $[V_{\min}, V_{\max}]$ and position range $[x_{\min}, x_{\max}]$ of the particle, the movement of the particle can be appropriately limited. In this paper, c_1 and c_2 are set to 1.718, and ω is set to 0.8.

In this paper, the optimal SNR_{out} of the signal SR system and the FPO model and the GPO model of a cascaded SR system are obtained by using the PSO algorithm. The only difference in the optimization of these three systems is their optimized parameters.

3. Numerical Simulations

3.1. Weak Signal Detection Based on the FPO Model and the GPO Model

In this subsection, the weak signal detection results of the FPO model and GPO model of the cascaded bistable system are studied and compared by processing a simulated weak signal. The cascaded tri-stable SR system has similar results, which are not presented in this paper. It is noted that the fourth-order Runge–Kutta algorithm is adopted in this work to solve the equations of the SR systems.

A weak signal $sn(t)$ with $A = 0.1$, $f_0 = 0.01$ Hz, and $D = 0.4$ is first generated. The sampling frequency of the signal is set to $F_s = 5$ Hz, and the signal contains 4096 points. The waveform $sn(t)$ and its spectrum $sn(f)$ are shown in Figure 3a,b. One can see that the characteristic signal is almost submerged in the strong background noise due to a relatively low SNR_{in} of -22.90 dB, which can be calculated similarly to Equation (7). This weak signal is further processed by the single SR system Equation (4), the FPO model Equation (5), and the GPO model Equation (6) of a cascaded SR system ($N = 4$) with the bistable potential function Equation (2). It is noted that in the optimization processes, the values of ε and R are both set to 1, and the optimization ranges of other parameters are all set to $[0.01, 6]$. Moreover, the number of iterations I_{max} is set to 200, and the population n is set to 30 in the PSO algorithm.

First, the weak signal $sn(t)$ is processed by the single bistable SR system. In the optimization process, the values of k , a_1 , and b_1 are optimized to achieve an optimal SR output. The optimization result shows that with $sn(t)$ input, the single bistable system achieves optimal SR when $[k, a_1, b_1] = [1.58, 0.08, 0.20]$, and the optimal SNR_{out} is -3.43 dB. The output waveform $x(t)$ and its spectrum $x(f)$ are shown in Figure 3c,d. It can be seen that the characteristic of the weak signal is significantly enhanced after being processed by the single bistable system.

Next, the FPO model of the cascaded bistable system ($N = 4$) is used to process the weak signal $sn(t)$. The optimized parameters contain k , a_1 , and b_1 . An optimal SNR_{out} of -1.16 dB can be obtained from the FPO model when $[k, a_1, b_1] = [5.24, 0.01, 5.67]$. The output waveform $x_F(t)$ of the FPO model and its spectrum $x_F(f)$ are shown in Figure 3e,f. One can see that the characteristic of the weak signal is further enhanced through the FPO model by filtering other interference signals.

Lastly, the weak signal $sn(t)$ is processed by the GPO model of the cascaded bistable system ($N = 4$). In this case, there exist 12 optimized parameters including $k_1, a_{11}, b_{11}, k_2, a_{12}, b_{12}, \dots, k_4, a_{14},$ and b_{14} . When their values are 0.54, 0.01, 0.01, 4.74, 0.01, 0.03, 5.91, 0.95, 3.06, 2.25, 0.03, and 0.01, respectively, the GPO model achieves its optimal SNR_{out} of -0.72 dB. The output waveform $x_G(t)$ of the GPO model and its spectrum $x_G(f)$ are shown in Figure 3g,h. It can be seen that the spectral peak at the frequency of $f_0 = 0.01$ Hz is more obvious compared to previous results.

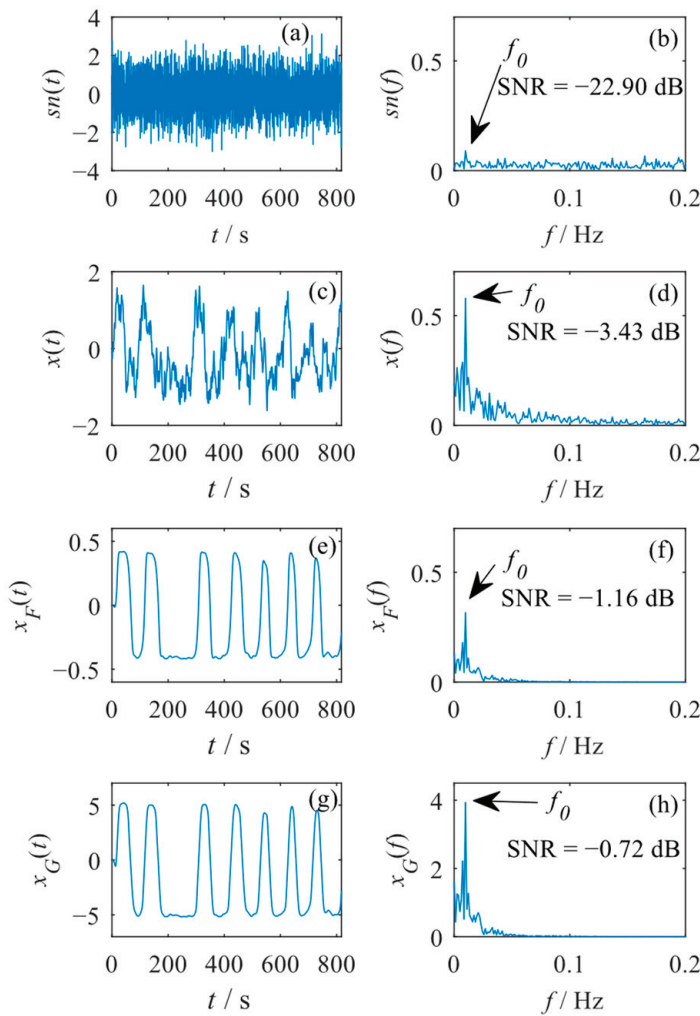


Figure 3. The waveform and spectrum of (a,b) the weak signal, (c,d) the output signal of the single bistable system, (e,f) the output signal of the FPO model of the cascaded bistable system ($N = 4$), and (g,h) the output signal of the GPO model of the cascaded bistable system ($N = 4$).

The simulated results show that by improving the signal spectral peak and filtering the interference signals/noise, the characteristic of the weak signal can be significantly enhanced through the single bistable system and the FPO model and the GPO model of the cascaded bistable system. The FPO model and the GPO model can achieve better SR outputs compared to the single bistable system, demonstrating the priority of the cascaded SR system in weak signal detection. Moreover, the GPO model is capable of achieving a better SR output compared to the FPO model, which accords with our previous analysis in Section 3.1. However, these conclusions are obtained from a special simulation case, which should be further demonstrated through statistical analysis. Moreover, it should be noted that the single bistable system and the FPO model and the GPO model of the cascaded bistable system all can detect the frequencies of multiple frequency components by setting the corresponding detected frequencies, and they will not misidentify non-existent frequency components. This can be demonstrated through further simulations, which are not presented in this paper.

3.2. Statistical Analysis of the Weak Signal Detection Performances of Different Models

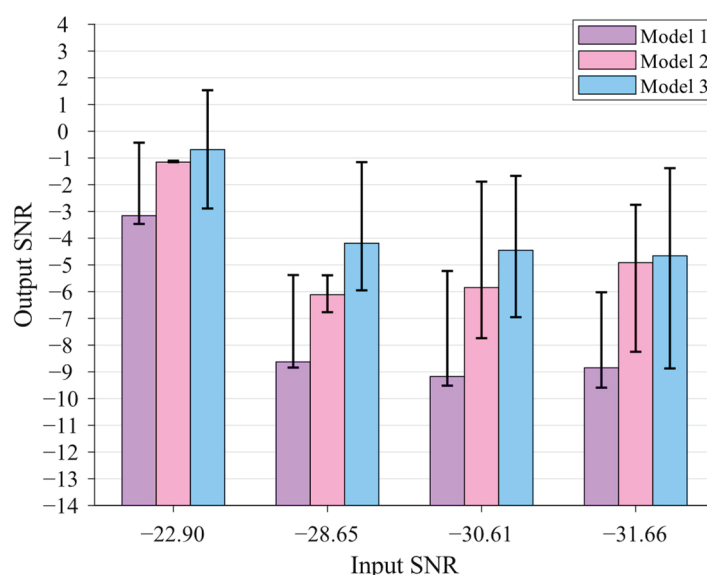
In this subsection, the weak signal detection performances of different models are fully investigated through statistical analysis. In this paper, six models are studied, including the single SR model, the FPO model, and the GPO model of the cascaded SR system with both bistable and tri-stable potential functions, as shown in Table 1.

Table 1. The SR models investigated in this work.

	Bistable Potential Field	Tri-Stable Potential Field
Single SR system	Model 1	Model 4
FPO model of the cascaded SR system	Model 2	Model 5
GPO model of the cascaded SR system	Model 3	Model 6

In the simulations, a weak signal with $A = 0.1$ and $f_0 = 0.01$ Hz is generated, and noises with intensities of 0.4, 1.6, 2.8, and 4.0 are added. Thus, four weak signals with an SNR_{in} of -22.90 dB, -28.65 dB, -30.61 dB, and -31.66 dB are obtained. The noisy signals are processed by the presented 6 models to obtain their optimal SR results, and the searching ranges of the optimization parameters are all set to $[0.01, 6]$ except $\varepsilon = 1$ and $R = 1$. Every signal is processed by each model 100 times, and the results are analyzed and compared statistically.

The statistical graph of the optimization results based on Models 1–3 is shown in Figure 4, in which the mean value of the output SNRs of each model for each input signal is plotted using a histogram and the maximal and minimum SNR_{out} for each case are plotted as well. One can see from Figure 4 that all the SNR_{out} are much larger than their corresponding SNR_{in} , indicating that all presented SR systems can significantly enhance the characteristic of the weak signal using the PSO algorithm. Moreover, for each input signal, the highest average SNR_{out} can be achieved from Model 3, and the second highest average SNR_{out} can be achieved from Model 2, indicating that in general, the GPO model of the cascaded bistable system has the best weak signal detection performance compared to the FPO model of the cascaded bistable system (second best) and the single bistable system. This conclusion can be also drawn from the fact that Model 3 can produce the highest maximum SNR_{out} for each input signal and also the highest minimum SNR_{out} in most cases. Another interesting finding is that as the SNR_{in} decreases from -28.65 dB to -31.66 dB, the average SNR_{out} for each case maintains a relatively stable value, demonstrating the robustness and stability of the SR systems in detecting low-SNR weak signals.

**Figure 4.** The statistical graph of the optimization results based on Models 1–3.

The statistical graph of the optimization results based on Models 4–6 is further shown in Figure 5. It can be seen that the statistical graph shows a similar pattern to Figure 4, indicating that the conclusions of the weak signal detection performance based on the bistable systems are applicable to those based on the tri-stable systems.

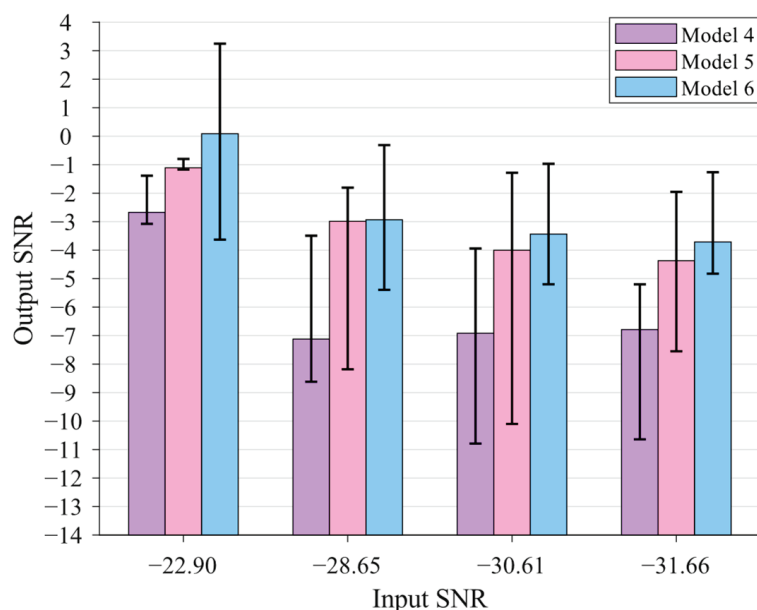


Figure 5. The statistical graph of the optimization results based on Models 4–6.

Previous analyses show the superiority of the proposed GPO model of the cascaded SR system compared to the conventional FPO model of the cascaded SR system and the single SR system. Furthermore, the weak signal detection performance of the cascaded systems with a bistable potential function and tri-stable potential function is compared through statistical analysis. In Figure 6, the mean values, maximum values, and minimum values of the output SNRs in 100 simulations based on the GPO models of the cascaded bistable systems and tri-stable systems are plotted for comparison. It can be seen that in most cases with different weak signals, especially the low- SNR_{in} cases, the GPO model of the cascaded tri-stable system has a higher average SNR_{out} , higher maximum SNR_{out} , and higher minimum SNR_{out} (except for one case) compared to those of the cascaded bistable system. This result indicates that the tri-stable SR system has a better weak signal detection performance compared to the bistable SR system, which has been reported in Refs. [28,50] for single SR systems.

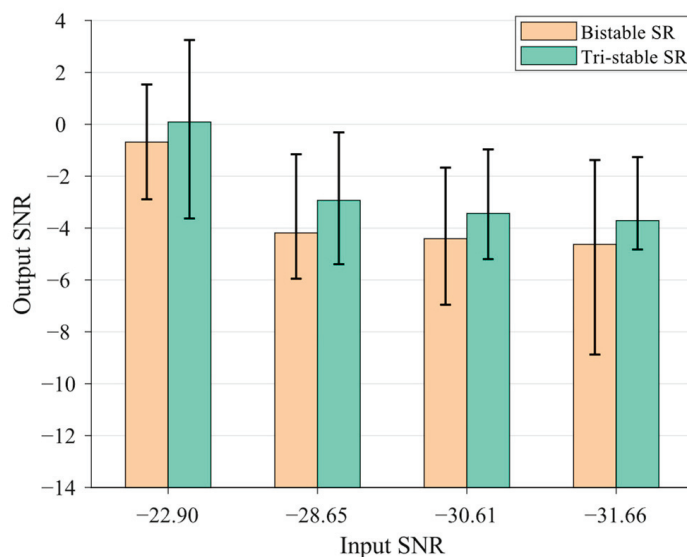


Figure 6. The statistical graph of the optimization results based on Model 3 and Model 6.

3.3. Influence of the Number of Subsystems

In previous simulations, the cascaded SR systems were pre-set as four-stage cascaded systems. Actually, the number of subsystems (N) has an influence on the weak signal detection performance and optimization time of the cascaded system, which are fully investigated in this subsection.

The GPO model of the cascaded tri-stable system is used for investigation. A total of 5 cascaded tri-stable systems with $N = 2, 3, 4, 5, 6, 7$, are adopted to analyze the same noisy signal with $A = 0.1$, $f_0 = 0.01$ Hz, and $D = 0.4$. The searching ranges of all adjustable parameters are set to $[0.01, 6]$ except $\varepsilon = 1$ and $R = 1$; the number of iterations $I_{\max}$ is set to 200, and the population n is set to 30 in the PSO algorithm. The input signal is processed by each cascaded system 100 times, and the average SNR_{out} and running times of each case are presented in Figure 7.

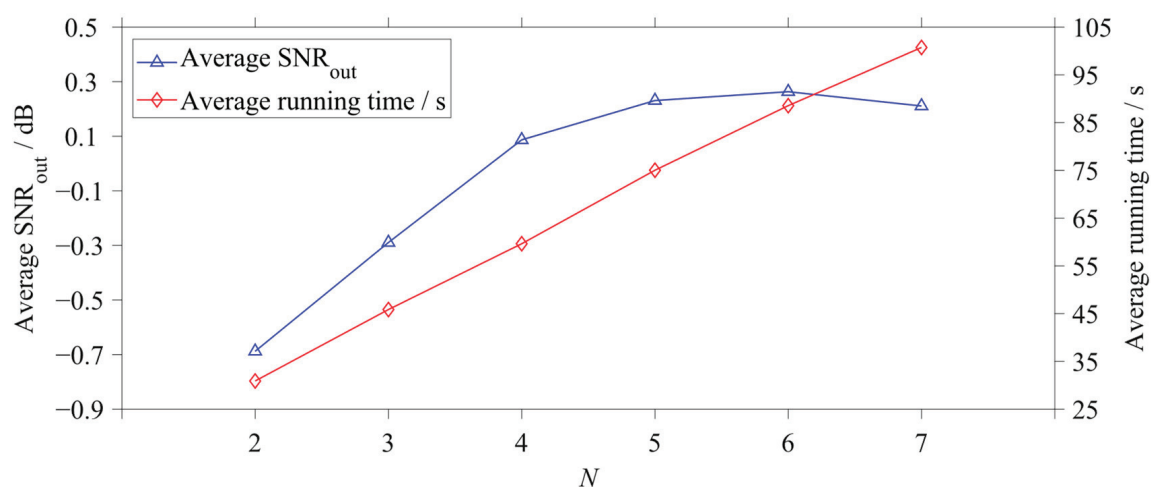


Figure 7. The statistical graph of the optimization results based on Model 6 with different numbers of subsystems.

It can be seen from Figure 7 that as N increases from 2 to 4, the average SNR_{out} gradually increases, which stays almost constant when the number of subsystems continues to increase from 4 to 6. The average SNR_{out} value starts to decrease when N increases to 7. The reason is that when the number of subsystems is large, too many subsystem parameters make it easy for the whole system to achieve a local optimum instead of the global optimum. On the other hand, with N increasing, the number of optimized parameters increases accordingly, and the average running time increases linearly, which can be seen in Figure 7. Therefore, the four-stage cascaded system is a high-cost-performance cascaded system that has a relatively high output performance and an acceptable optimization time. In the following sections of this paper, four-stage cascaded systems are used for weak signal detection and fault diagnosis.

4. Experimental Verifications

The simulation results of Section 3 show that the GPO model of the cascaded SR system has a rather good weak signal detection performance. Hence, it has great potential in mechanical weak fault diagnosis for faulty signals with extremely low SNR. In this section, two examples of practical mechanical weak fault diagnosis are presented to verify the effectiveness of the proposed GPO model of the cascaded SR system. It is noted that the SR system with tri-stable potential functions is used due to its better weak signal detection performance compared to that with a bistable potential function. The fault diagnosis process is shown in Figure 8.

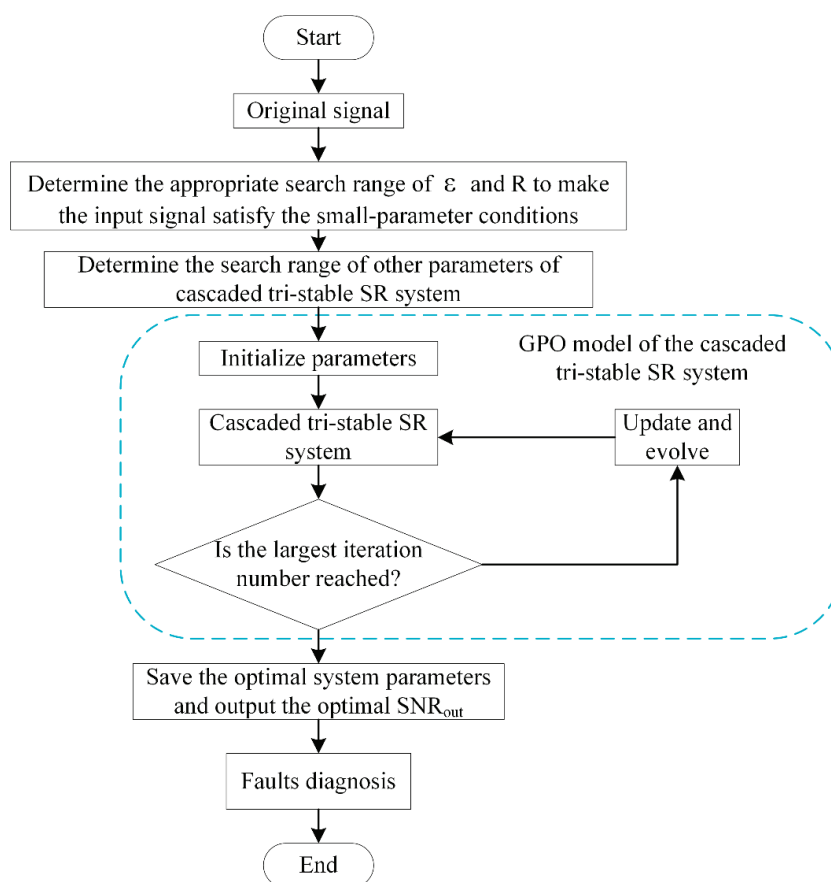


Figure 8. Flowchart of the fault diagnosis based on the GPO model of the cascaded tri-stable system.

4.1. Case 1: Diagnosis of Rolling Bearing Inner Raceway Fault

The data of the rolling bearing inner ring fault are obtained from the bearing data center website of the Case Western Reserve University (CWRU). The associated test rig consists of a 2-horsepower reliance electric motor, a torque transducer/encoder, a dynamometer, and control electronics. In the tests, a groove of width of 0.021 inches was cut in the face of the inner raceway of the drive-end bearing by using electrical discharge machining (EDM), thus introducing a single-point inner ring fault. The faulty bearing was a deep groove ball bearing (6205-2RS JEM SKF), whose parameters are shown in Table 2. According to the fault mechanism of the inner raceway fault, a characteristic frequency of f_i is contained in the vibration signal of the vibrating system. The value of f_i can be calculated according to [51]:

$$f_i = \frac{z}{2} \left(1 + \frac{d_r}{D_p} \cos \beta \right) f_R \quad (10)$$

where z is the number of rolling elements; d_r and D_p are the rolling elements' diameter and the pitch diameter, respectively; β is the contact angle; and f_R is the rotation frequency of the bearing. During the test, the shaft speed was 1774 rpm, i.e., $f_R = 29.57$ Hz; thus, the characteristic frequency of the inner raceway fault can be calculated as $f_i = 160.1$ Hz.

Table 2. The parameters of the tested rolling bearing.

Model of Bearing	Inside Diameter d_r (Inches)	Outside Diameter (Inches)	Thickness (Inches)	Ball Diameter (Inches)	No. of Rolling Elements Z	Contact Angle β
SKF 6205-2RS JEM	0.9843	2.0472	0.5906	0.3126	9	0°

The bearing vibration signals were acquired using an accelerometer placed at the motor housing, and the sampling frequency was 48,000 Hz. The length of the original signal used in this paper was $M = 4096$, and its input SNR was calculated as $\text{SNR}_{\text{in}} = -39.45$ dB. The waveform, global spectrum, and low-frequency spectrum of the original signal are shown in Figure 9. It can be seen that the fault characteristic frequency of $f_i = 160.1$ Hz cannot be identified from either the time domain or the frequency domain due to the strong noise interference. Therefore, the existing inner raceway fault of the bearing cannot be diagnosed directly from the original signal.

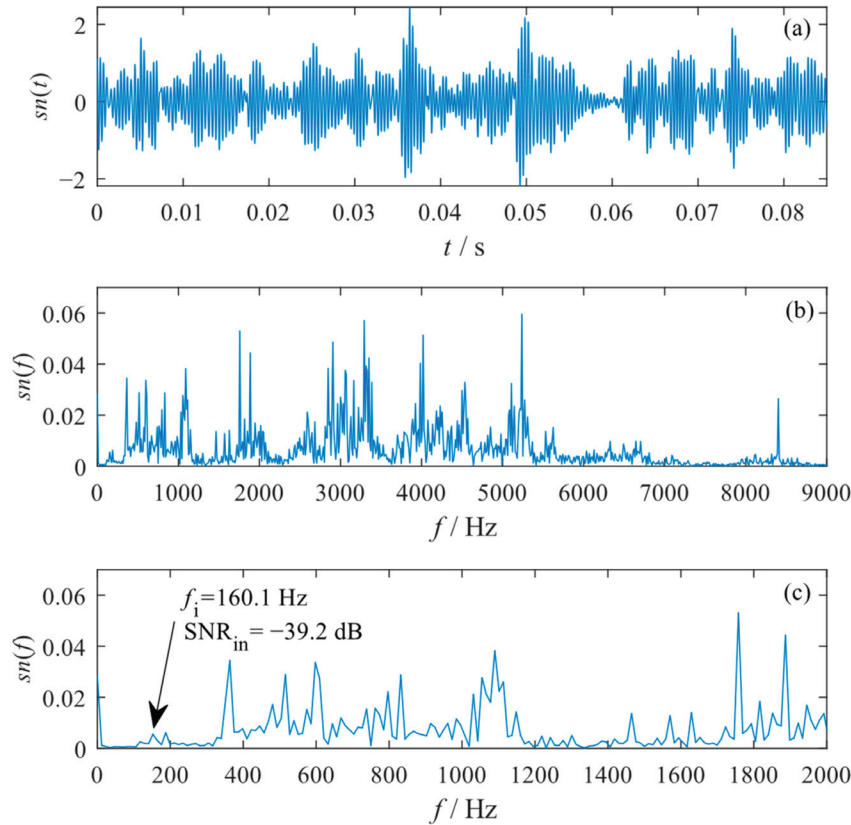


Figure 9. Original vibration signal with bearing inner raceway fault. (a) Waveform; (b) global spectrum; (c) low-frequency spectrum.

In order to diagnose the potential inner raceway fault of the rolling bearing, the proposed GPO model of a four-stage cascaded tri-stable system was adopted to further analyze the original signal. In the optimization of the PSO algorithm, the SNR_{out} at the frequency closest to the characteristic frequency f_i was selected as the objective function. The searching ranges of the amplitude-transformation coefficient and the scale-transformation coefficient were set to $\varepsilon \in [0, 30]$ and $R \in [2000, 3000]$, respectively, and the searching ranges of other optimization parameters ($a_1, b_1, c_1, k_1, a_2, b_2, c_2, k_2, \dots, a_4, b_4, c_4, k_4$) were all set to $[0.01, 6]$. The values of the optimal parameters of $\varepsilon, R, a_1, b_1, c_1, k_1, a_2, b_2, c_2, k_2, \dots, a_4, b_4, c_4, k_4$ obtained from the GPO model were 6.13, 2000.60, 5.95, 1.10, 1.07, 4.48, 5.39, 4.63, 2.77, 3.72, 3.79, 5.08, 0.93, 4.77, 0.27, 2.38, 1.62, and 0.01, respectively, which enabled the cascaded tri-stable system to produce an optimal SR output with an optimal SNR_{out} of +3.31 dB. The corresponding waveform, global spectrum, and low-frequency spectrum of the output signal were presented in Figure 10. One can see from the waveform that the system output presents a regular inter-well oscillation, and a spectral peak is prominent at $f = 164.1$ Hz with other interference components greatly reduced and nearly eliminated. It can be calculated that the frequency resolution of the spectrum is $\Delta f = F_s/N = 11.7$ Hz; thus, the frequency of the prominent spectral peak ($f = 164.1$ Hz) indicates the characteris-

tic frequency of an inner raceway fault ($f_i = 160.1$ Hz). Hence, an inner raceway fault in the experimental bearing can be identified from the system SR output, which accords with the practical truth. Thus, the effectiveness of the proposed GPO model of the cascaded SR system in mechanical fault diagnosis can be verified.

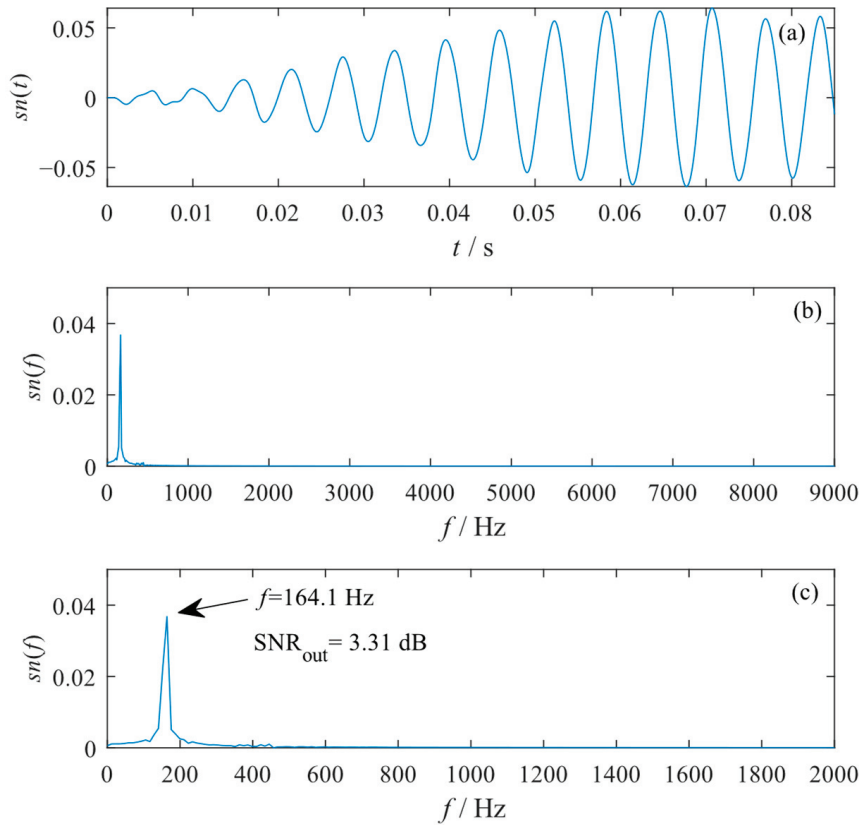


Figure 10. The output signal obtained from the 4-stage cascaded tri-stable system optimized by the GPO model with the rolling bearing inner raceway faulty signal input. (a) Waveform; (b) global spectrum; (c) low-frequency spectrum.

4.2. Case 2: Diagnosis of Rolling Bearing Outer Raceway Fault

The bearing outer raceway failure test rig is shown in Figure 11, which mainly contains a rotor shaft and two rolling bearings driven by an electric motor. In the tests, the outer raceway of the left bearing (cylindrical roller bearing, NU205) had a groove 0.2 mm in width and 0.1 mm in depth, which simulated a bearing with an outer raceway fault. According to the fault mechanism of the outer raceway fault, a characteristic frequency of f_o is contained in the vibration signal of the vibrating system, whose value can be calculated according to [51]:

$$f_o = \frac{z}{2} \left(1 - \frac{d_r}{D_p} \cos \beta \right) f_R \quad (11)$$

In the tests, the motor rotated at a speed of 950 rpm, i.e., $f_R = 15.83$ Hz. Hence, $f_o = 82.95$ Hz can be obtained with basic parameters of the faulty bearing of $z = 13$, $\beta = 0^\circ$, $d_r = 7.5$ mm, and $D_p = 39$ mm.

An accelerometer was installed on the test rig to acquire the vibration signal of the system using a signal-acquiring device (NI PXI-1033). In the tests, the sampling frequency was $F_s = 15,000$ Hz. An original signal with length $M = 4096$ was used for analysis. The waveform, global spectrum, and low-frequency spectrum of the original signal are shown in Figure 12. It can be seen that as the SNR of the original signal was as low as $SNR_{in} = -39.45$ dB, the waveform is disordered, and the fault characteristic is also

submerged in the spectrum. Therefore, the outer raceway fault of the bearing cannot be identified from the original signal.

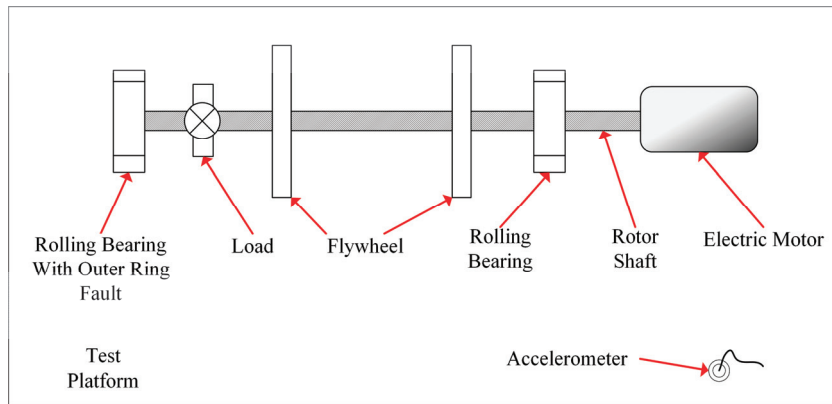


Figure 11. Bearing outer raceway fault test rig.

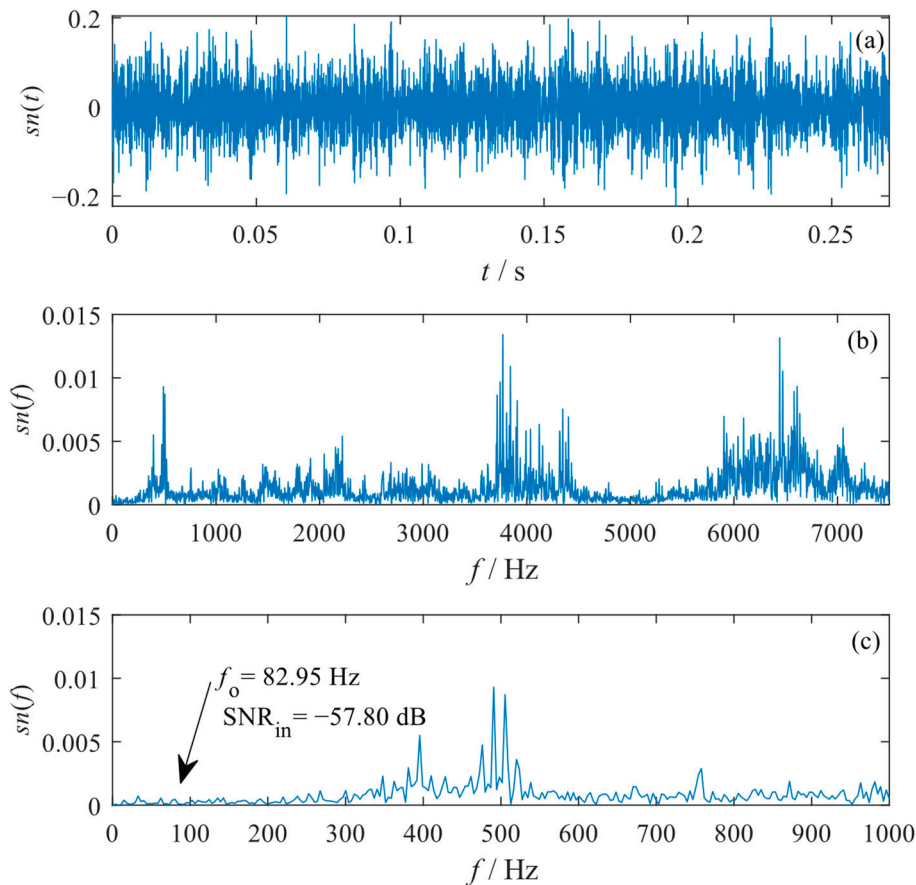


Figure 12. Original vibration signal with bearing outer raceway fault. (a) Waveform; (b) global spectrum; (c) low-frequency spectrum.

Again, the proposed GPO model of the four-stage cascaded tri-stable system was used to further process the original signal. The searching ranges of the amplitude-transformation coefficient and the scale-transformation coefficient were set to $\varepsilon \in [0, 100]$ and $R \in [1000, 2000]$, respectively, and the searching ranges of $a_1, b_1, c_1, k_1, a_2, b_2, c_2, k_2, \dots, a_4, b_4, c_4, k_4$ were all set to $[0.01, 6]$. The optimal SR output ($\text{SNR}_{\text{out}} = -11.33$ dB) was achieved when the values of $\varepsilon, R, a_1, b_1, c_1, k_1, a_2, b_2, c_2, k_2, \dots, a_4, b_4, c_4$, and k_4 were 74.06, 1002.90, 3.40, 2.81, 3.48, 3.88, 0.27, 0.97, 2.41, 0.01, 4.19, 3.44, 2.59, 3.49, 3.00, 5.16,

1.56, and 3.90, respectively. The corresponding waveform and spectrum of the output signal are presented in Figure 13. It can be seen from the waveform that the output signal presents a regular inter-well oscillation. The most prominent component has a frequency of 80.57 Hz, which can be clearly observed from the spectrum. Considering the frequency resolution of $\Delta f = F_s/N = 3.66$ Hz, this characteristic frequency indicates an outer raceway fault frequency of 82.95 Hz. Thus, an outer raceway fault in the rolling bearing can be identified, and the effectiveness of the proposed methods in mechanical fault diagnosis is verified again.

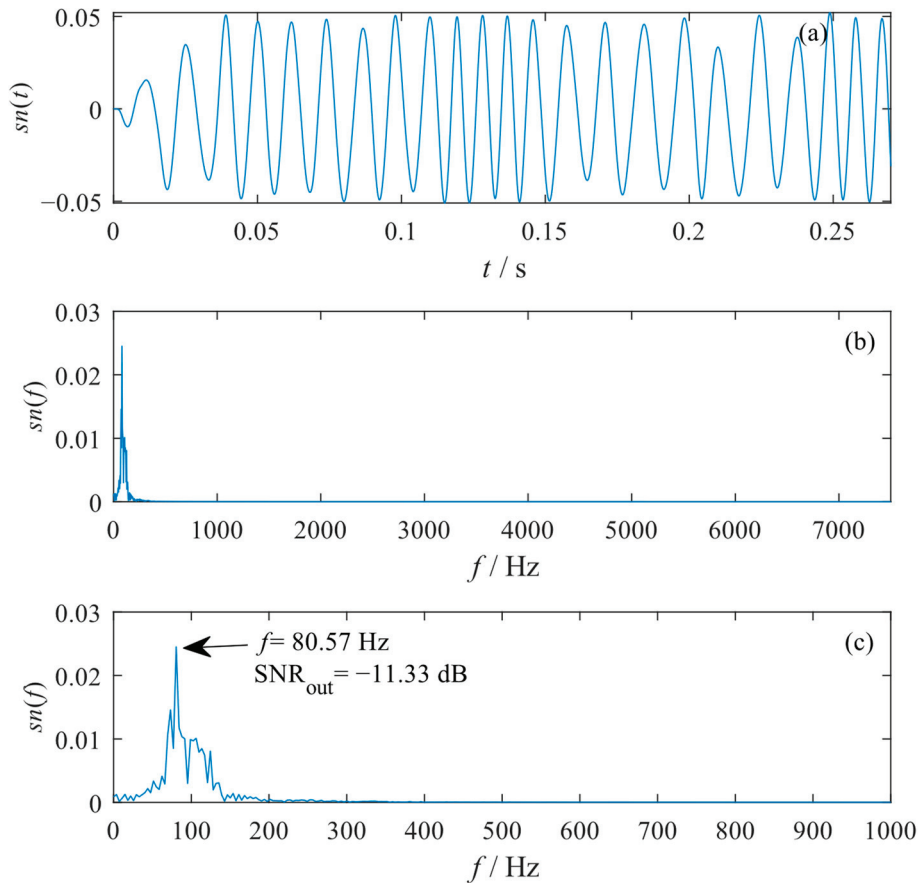


Figure 13. The output signal obtained from the 4-stage cascaded tri-stable system optimized by the GPO model with the rolling bearing outer raceway faulty signal input. (a) Waveform; (b) global spectrum; (c) low-frequency spectrum.

5. Conclusions

In this paper, by combining the multi-parameter-adjusting cascaded stochastic resonance (SR) system and the particle swarm optimization (PSO) algorithm, a global parameter optimization (GPO) model of a cascaded SR system is proposed for high-performance adaptive weak fault diagnosis. The cascaded SR system may involve different potential functions and contain multiple SR systems connected in series. Simulated results show that the GPO model of a cascaded SR system can realize adaptive weak signal detection even under large-parameter conditions with a better performance compared to the single SR system and the pre-proposed fixed-parameter optimization (FPO) model of the cascaded SR system, and the robustness and stability of the system in detecting low-SNR weak signals are rather good. Moreover, the cascaded SR system with a tri-stable potential function has a higher probability of producing a better SR output compared to that with a bistable potential function. The influence of the number of subsystems on the system output is also investigated. It is found that a four-stage cascaded system is a high-cost-performance

cascaded system that has a relatively high output performance and also an acceptable optimization time. The weak signal detection and weak fault diagnosis approach is further proposed based on the GPO model of a tri-stable cascaded SR system, which is applied to diagnose the inner raceway fault and the outer raceway fault of a rolling bearing. Experimental work shows that the proposed approach can be utilized to identify the weak characteristic frequency component submerged in a noisy background by significantly enhancing the signal-to-noise ratio (SNR) of the characteristic signal, thus demonstrating its effectiveness in high-performance adaptive mechanical weak fault diagnosis.

Author Contributions: Conceptualization, Z.L. and Z.Q.; Data curation, Z.H., C.W. and J.X.; Formal analysis, Z.H.; Funding acquisition, Z.L. and Z.Q.; Methodology, Z.L. and Z.H.; Project administration, R.Z. and Z.Q.; Resources, M.X.; Software, Z.H.; Supervision, Z.L. and R.Z.; Validation, C.Z.; Visualization, M.X.; Writing—original draft, Z.H.; Writing—review & editing, Z.L. and Z.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (grant nos. 51905349, U2013603, and 52205569), Guangdong Basic and Applied Basic Research Foundation (grant nos. 2022A1515010126 and 2020A1515011509), Zhejiang Provincial Natural Science Foundation of China (grant no. LQ22E050003), Ningbo Science and Technology Major Project (grant nos. 2020Z110, 2022Z057, and 2022Z002), Chuying Planning Project of Zhejiang Provincial Administration for Market Regulation (grant no. CY2023328), and Shandong Provincial Innovation Ability Improvement Project of Middle and Small-sized High-tech Enterprises (grant no. 2022TSGC2364).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available on request from the authors.

Acknowledgments: We sincerely thank the support by the National Natural Science Foundation of China (grant nos. 51905349, U2013603, and 52205569), Guangdong Basic and Applied Basic Research Foundation (grant nos. 2022A1515010126 and 2020A1515011509), Zhejiang Provincial Natural Science Foundation of China (grant no. LQ22E050003), Ningbo Science and Technology Major Project (grant nos. 2020Z110, 2022Z057, and 2022Z002), Chuying Planning Project of Zhejiang Provincial Administration for Market Regulation (grant no. CY2023328), and Shandong Provincial Innovation Ability Improvement Project of Middle and Small-sized High-tech Enterprises (grant no. 2022TSGC2364).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CWRU	Case Western Reserve University
EDM	Electrical Discharge Machining
EMD	Empirical Mode Decomposition
FPO	Fixed-parameter Optimization
FRSR	Frequency Re-scaling Stochastic Resonance
FSRSR	Frequency-Shifted and Re-scaling Stochastic Resonance
GPO	Global parameter Optimization
MSR	Modulated Stochastic Resonance
PNSR	Parameter-Normalized Stochastic Resonance
PSO	Particle Swarm Optimization
QPSO	Quantum Particle Swarm Optimization
SK	Spectral Kurtosis
SNR	Signal-to-noise Ration
SR	Stochastic Resonance
WT	Wavelet Transform

References

1. Tang, J.; Shi, B.; Li, Z.; Li, Y. Weak fault feature extraction method based on compound tri-stable stochastic resonance. *Chin. J. Phys.* **2020**, *66*, 50–59. [CrossRef]
2. Liu, J.; Shen, X.; Wang, J.; Jiang, L.; Zhang, H. An Intelligent Defect Detection Approach Based on Cascade Attention Network Under Complex Magnetic Flux Leakage Signals. *IEEE Trans. Ind. Electron.* **2023**, *70*, 7417–7427. [CrossRef]
3. Kolbe, S.; Gelman, L.; Ball, A. Novel Prediction of Diagnosis Effectiveness for Adaptation of the Spectral Kurtosis Technology to Varying Operating Conditions. *Sensors* **2021**, *21*, 6913. [CrossRef] [PubMed]
4. Yang, J.; Han, S.; Zhang, S.; Liu, H.; Tang, C. Empirical mode decomposition of weak fault characteristic signal of rolling bearing under strong noise background. *J. Vib. Eng.* **2020**, *33*, 582–589.
5. Zheng, J.D.; Su, M.X.; Ying, W.M.; Tong, J.Y.; Pan, Z.W. Improved uniform phase empirical mode decomposition and its application in machinery fault diagnosis. *Measurement* **2021**, *179*, 10. [CrossRef]
6. Wang, L.M.; He, J.F.; Xiao, H.F.; Zeng, Q.; Ding, X.X.; Shao, Y.M. Joint suppression of normal gear mesh component and background noise for early local fault detection based on dynamic evolutionary digital filter. *Measurement* **2022**, *201*, 16. [CrossRef]
7. Han, B.; Li, C.S.; Zhou, Y.Q.; Yu, G.; Wei, C.L. Second-Order Multisynchrosqueezing Wavelet Transform for Bearing Fault Detection. *J. Vib. Eng. Technol.* **2022**, *10*, 1541–1559. [CrossRef]
8. Benzi, R.; Parisi, G.; Sutera, A.; Vulpiani, A. Stochastic resonance in climatic change. *Tellus* **1982**, *34*, 10–16. [CrossRef]
9. Benzi, R.; Sutera, A.; Vulpiani, A. The mechanism of stochastic resonance. *J. Phys. A Math. Gen.* **1981**, *14*, L453. [CrossRef]
10. Shi, P.M.; Su, C.J.; Han, D.Y. Fault diagnosis of rotating machinery based on adaptive stochastic resonance and AMD-EEMD. *Shock. Vib.* **2016**, *2016*, 9278581. [CrossRef]
11. Tang, J.C.; Shi, B.Q.; Bao, H.R.; Li, Z.X. A new method for weak fault feature extraction based on piecewise mixed stochastic resonance. *Chin. J. Phys.* **2020**, *68*, 87–99. [CrossRef]
12. Lu, S.; He, Q.; Wang, J. A review of stochastic resonance in rotating machine fault detection. *Mech. Syst. Signal Process.* **2019**, *116*, 230–260. [CrossRef]
13. Qiao, Z.J.; Elhatab, A.; Shu, X.D.; He, C.B. A second-order stochastic resonance method enhanced by fractional-order derivative for mechanical fault detection. *Nonlinear Dyn.* **2021**, *106*, 707–723. [CrossRef]
14. Qiao, Z.; Shu, X. Coupled neurons with multi-objective optimization benefit incipient fault identification of machinery. *Chaos Solitons Fractals* **2021**, *145*, 110813. [CrossRef]
15. Zhang, J.J.; Zhang, T. Parameter-induced stochastic resonance based on spectral entropy and its application to weak signal detection. *Rev. Sci. Instrum.* **2015**, *86*, 6. [CrossRef]
16. Zhang, G.; Liu, Y.L.; He, L.F. Research on fault detection of asymmetric piecewise well-posed stochastic resonance system. *Rev. Sci. Instrum.* **2021**, *92*, 11. [CrossRef]
17. Qiao, Z.; Lei, Y.; Li, N. Applications of stochastic resonance to machinery fault detection: A review and tutorial. *Mech. Syst. Signal Process.* **2019**, *122*, 502–536. [CrossRef]
18. Yang, Z.; Li, Z.Q.; Zhou, F.X.; Ma, Y.J.; Yan, B.K. Weak fault feature extraction method based on improved stochastic resonance. *Sensors* **2022**, *22*, 6644. [CrossRef]
19. Leng, Y.G.; Leng, Y.S.; Wang, T.Y.; Guo, Y. Numerical analysis and engineering application of large parameter stochastic resonance. *J. Sound Vib.* **2006**, *292*, 788–801. [CrossRef]
20. Tan, J.; Chen, X.; Wang, J.; Chen, H.; Cao, H.; Zi, Y.; He, Z. Study of frequency-shifted and re-scaling stochastic resonance and its application to fault diagnosis. *Mech. Syst. Signal Process.* **2009**, *23*, 811–822. [CrossRef]
21. Yang, D.X.; Hu, Z.; Yang, Y.M. The analysis of stochastic resonance of periodic signal with large parameters. *Acta Phys. Sin.* **2012**, *61*, 10. (In Chinese)
22. Lin, M.; Huang, Y.M. Modulation and demodulation for detecting weak periodic signal of stochastic resonance. *Acta Phys. Sin.* **2006**, *55*, 3277–3282. (In Chinese) [CrossRef]
23. Lai, Z.-H.; Leng, Y.-G. Weak-signal detection based on the stochastic resonance of bistable Duffing oscillator and its application in incipient fault diagnosis. *Mech. Syst. Signal Process.* **2016**, *81*, 60–74. [CrossRef]
24. Xu, B.M.; Shi, J.C.; Zhong, M.; Zhang, J. Incipient fault diagnosis of planetary gearboxes based on an adaptive parameter-induced stochastic resonance method. *Appl. Acoust.* **2022**, *188*, 108587. [CrossRef]
25. Hu, B.B.; Guo, C.; Wu, J.M.; Tang, J.H.; Zhang, J.L.; Wang, Y. An adaptive periodical stochastic resonance method based on the grey wolf optimizer algorithm and its application in rolling bearing fault diagnosis. *J. Vib. Acoust. Trans. Asme* **2019**, *141*, 9. [CrossRef]
26. Cheng, W.; Xu, X.M.; Ding, Y.P.; Sun, K.H. Stochastic resonance in a single-well potential and its application in rolling bearing fault diagnosis. *Rev. Sci. Instrum.* **2020**, *91*, 16. [CrossRef]
27. Tong, L.; Li, X.G.; Hu, J.H.; Ren, L.T. A PSO optimization scale-transformation stochastic-resonance algorithm with stability mutation operator. *IEEE Access* **2018**, *6*, 1167–1176. [CrossRef]
28. Lai, Z.H.; Liu, J.S.; Zhang, H.T.; Zhang, C.L.; Zhang, J.W.; Duan, D.Z. Multi-parameter-adjusting stochastic resonance in a standard tri-stable system and its application in incipient fault diagnosis. *Nonlinear Dyn.* **2019**, *96*, 2069–2085. [CrossRef]
29. Cui, H.J.; Guan, Y.; Chen, H.Y.; Deng, W. A novel advancing signal processing method based on coupled multi-stable stochastic resonance for fault detection. *Appl. Sci.* **2021**, *11*, 5385. [CrossRef]

30. Qiao, Z.; Liu, J.; Ma, X.; Liu, J. Double stochastic resonance induced by varying potential-well depth and width. *J. Frankl. Inst.* **2021**, *358*, 2194–2211. [CrossRef]
31. Li, J.; Wang, X.; Li, Z.; Zhang, Y. Stochastic resonance in cascaded monostable systems with double feedback and its application in rolling bearing fault feature extraction. *Nonlinear Dyn.* **2021**, *104*, 971–988. [CrossRef]
32. Gao, R.; Jiao, S.; Wang, Y.; Li, Y. Weak signal detection of composite multistable stochastic resonance with Woods-Saxon potential. *Z. Fur Nat. Sect. A J. Phys. Sci.* **2022**, *77*, 963–976. [CrossRef]
33. Zhang, H.; He, Q.; Kong, F. Stochastic resonance in an underdamped system with pinning potential for weak signal detection. *Sensors* **2015**, *15*, 21169–21195. [CrossRef]
34. Li, Z.X.; Liu, X.D.; Han, S.J.; Wang, J.G.; Ren, X.P. Fault diagnosis method and application based on unsaturated piecewise linear stochastic resonance. *Rev. Sci. Instrum.* **2019**, *90*, 10. [CrossRef]
35. Li, J.M.; Zhang, Y.G.; Xie, P. A new adaptive cascaded stochastic resonance method for impact features extraction in gear fault diagnosis. *Measurement* **2016**, *91*, 499–508. [CrossRef]
36. Guo, W.; Zhou, Z.; Chen, C.; Li, X. Multi-frequency weak signal detection based on multi-segment cascaded stochastic resonance for rolling bearings. *Microelectron. Reliab.* **2017**, *75*, 239–252. [CrossRef]
37. Xu, D.; Ge, J.; Wang, Y.; Shao, J. Multi-Frequency Weak Signal Decomposition and Reconstruction of Rolling Bearing Based on Adaptive Cascaded Stochastic Resonance. *Machines* **2021**, *9*, 275. [CrossRef]
38. Zhang, G.; Wang, H.; Zhang, T.Q. Cascaded Nonlinear Mass Fluctuation Stochastic Resonance System and Its Application in Bearing Fault Diagnosis. *Fluct. Noise Lett.* **2021**, *20*, 2150045. [CrossRef]
39. Li, J.M.; Zhang, J.F.; Li, M.; Zhang, Y.G. A novel adaptive stochastic resonance method based on coupled bistable systems and its application in rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2019**, *114*, 128–145. [CrossRef]
40. Kim, J.; Harne, R.L.; Wang, K.W. Online signal denoising using adaptive stochastic resonance in parallel array and its application to acoustic emission signals. *J. Vib. Acoust. Trans. Asme* **2022**, *144*, 12. [CrossRef]
41. Liu, W.; Liu, Z.; Zhang, Q.; Xu, Y.; Liu, S.; Chen, Z.; Zhu, C.; Wang, Z.; Pan, M.; Hu, J.; et al. Magnetic Anomaly Signal Detection Using Parallel Monostable Stochastic Resonance System. *IEEE Access* **2020**, *8*, 162230–162237. [CrossRef]
42. He, H.L.; Wang, T.Y.; Leng, Y.G.; Zhang, Y.; Li, Q. Study on non-linear filter characteristic and engineering application of cascaded bistable stochastic resonance system. *Mech. Syst. Signal Process.* **2007**, *21*, 2740–2749. [CrossRef]
43. Hairan, R.; Jalimhan, E.; Dalabaev, S. Cascaded Bistable Stochastic Resonance Signal Detection under Dual-mode Non-Gaussian Noise Background. *Telecommun. Eng.* **2014**, *54*, 1260–1264. (In Chinese)
44. Li, B.; Li, J.; He, Z. Fault feature enhancement of gearbox in combined machining center by using adaptive cascade stochastic resonance. *Sci. China Technol. Sci.* **2011**, *54*, 3203–3210. [CrossRef]
45. Xiao, L.; Bajric, R.; Zhao, J.; Tang, J.; Zhang, X. An adaptive vibrational resonance method based on cascaded varying stable-state nonlinear systems and its application in rotating machine fault detection. *Nonlinear Dyn.* **2021**, *103*, 715–739. [CrossRef]
46. Liu, H.; Han, S.; Yang, J.; Liu, S. Improving the weak feature extraction by adaptive stochastic resonance in cascaded piecewise-linear system and its application in bearing fault detection. *J. Vibroeng.* **2017**, *19*, 2506–2520. [CrossRef]
47. Zhang, G.; Gao, J.P.; Hong-Wei, L.I. Research on stochastic resonance characteristics of cascaded three-steady-state and its application. *Comput. Sci.* **2018**, *45*, 146–151.
48. Lai, Z.H.; Leng, Y.G. Generalized parameter-adjusted stochastic resonance of Duffing oscillator and its application to weak-signal detection. *Sensors* **2015**, *15*, 21327–21349. [CrossRef]
49. Sun, H.; Wang, Z. Weak signal detection based on cascaded piecewise-linear stochastic resonance. *China Mech. Eng.* **2014**, *25*, 3343. (In Chinese)
50. Lai, Z.H.; Leng, Y.G. Dynamic response and stochastic resonance of a tri-stable system. *Acta Phys. Sin.* **2015**, *64*, 12. (In Chinese)
51. Zhang, G.; Yi, T.; Zhang, T.; Cao, L. A multiscale noise tuning stochastic resonance for fault diagnosis in rolling element bearings. *Chin. J. Phys.* **2018**, *56*, 145–157. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Stochastic Resonance with Parameter Estimation for Enhancing Unknown Compound Fault Detection of Bearings

Min Xu ¹, Chao Zheng ¹, Kelei Sun ¹, Li Xu ¹, Zijian Qiao ^{2,3,4,5,*} and Zhihui Lai ⁶

¹ Ningbo Cigarette Factory, China Tobacco Zhejiang Industry Co., Ltd., Ningbo 315040, China

² School of Mechanical Engineering and Mechanics, Ningbo University, Ningbo 315211, China

³ Yangjiang Offshore Wind Power Laboratory, Yangjiang 529500, China

⁴ State Key Laboratory of Performance Monitoring and Protecting of Rail Transit Infrastructure, East China Jiaotong University, Nanchang 330013, China

⁵ Zhejiang Provincial Key Laboratory of Part Rolling Technology, Ningbo 315211, China

⁶ Shenzhen Key Laboratory of High Performance Nontraditional Manufacturing, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China

* Correspondence: qiaozijian@nbu.edu.cn

Abstract: Although stochastic resonance (SR) has been widely used to enhance weak fault signatures in machinery and has obtained remarkable achievements in engineering application, the parameter optimization of the existing SR-based methods requires the quantification indicators dependent on prior knowledge of the defects to be detected; for example, the widely used signal-to-noise ratio easily results in a false SR and decreases the detection performance of SR further. These indicators dependent on prior knowledge would not be suitable for real-world fault diagnosis of machinery where their structure parameters are unknown or are not able to be obtained. Therefore, it is necessary for us to design a type of SR method with parameter estimation, and such a method can estimate these parameters of SR adaptively by virtue of the signals to be processed or detected in place of the prior knowledge of the machinery. In this method, the triggered SR condition in second-order nonlinear systems and the synergic relationship among weak periodic signals, background noise and nonlinear systems can be considered to decide parameter estimation for enhancing unknown weak fault characteristics of machinery. Bearing fault experiments were performed to demonstrate the feasibility of the proposed method. The experimental results indicate that the proposed method is able to enhance weak fault characteristics and diagnose weak compound faults of bearings at an early stage without prior knowledge and any quantification indicators, and it presents the same detection performance as the SR methods based on prior knowledge. Furthermore, the proposed method is more simple and less time-consuming than other SR methods based on prior knowledge where a large number of parameters need to be optimized. Moreover, the proposed method is superior to the fast kurtogram method for early fault detection of bearings.

Keywords: stochastic resonance; weak fault detection; mechanical fault diagnosis

1. Introduction

Rolling bearings are key components of rotating machinery and can provide reliable and stable support. However, during the operating process, rolling bearings are inevitable; their defects include wear, cracking and pitting due to insufficient lubrication, contact fatigue, etc. Therefore, it is very important for us to achieve early fault detection and fault diagnosis of bearings to avoid a serious accident. Up until now, there were two well-known categories of bearing fault diagnosis, including model updating [1,2] and direct feature extraction and analysis using machine learning [3,4] or signal processing [5,6]. In addition, some scholars focus on the prediction of remaining useful life [7,8].

Among them, because direct feature extraction and analysis based on signal processing does not require building complex mathematical models and complete fault samples, they

have attracted sustained attention in fault detection. However, most signal processing methods attempt to cancel the noise embedded in signals to detect weak fault features, while stochastic resonance is able to harvest the energy of noise to enhance weak useful signals and therefore has been widely used in weak fault detection and fault diagnosis of machinery. In 2019, Lu et al. [9] and Qiao et al. [10] reviewed these articles on SR-based fault diagnosis of machinery and pointed out the potential development directions, respectively. Two review articles are inspiring researchers to explore the potential of SR as well as to develop advanced research in this field.

Up until now, there were many scholars who investigated the application of SR in mechanical fault diagnosis. Among them, some paid attention to designing novel indicators to quantify SR. For example, López et al. [11] proposed hidden Markov models and a Box–Cox sparse measures-based SR method for bearing fault diagnosis in which the designed indicator does not depend on prior knowledge of the fault signature to be detected, but the parameter selection of hidden Markov models and Box–Cox sparse measures would complicate the proposed method. Lin et al. [12] designed a novel indicator to quantify the SR for bearing fault diagnosis, where the cross-correlation coefficient, impulse index and signal-to-noise ratio (SNR) are fused. Li et al. [13] presented a multi-parameter constrained potential underdamped SR method and applied it to weak fault diagnosis, where SNR is used to evaluate SR. Lai et al. [14] defined SNR_{input} and SNR_{output} as the signal-to-noise ratios at SR system input and output to quantify SR for mechanical fault diagnosis. Zhang et al. [15] used the opposite of SNR as a fitness function of salp swarm algorithms to optimize the parameters of the proposed SR method. Wang et al. [16] applied order tracking to address the time-varying signals and then designed a tristable SR method to enhance weak fault characteristics with the fitness function of SNR. Zhou et al. [17] used generative adversarial networks to fuse 10 statistical parameters as the fitness function for optimizing the proposed SR method. Shi et al. [18] proposed a novel adaptive multi-parameter unsaturation bistable SR method in which the output SNR is selected as the objective function. Qiao et al. [19] proposed a second-order SR method enhanced by a fractional-order derivative for mechanical fault detection where SNR is used to optimize its parameters, even in coupled neurons [20] or the proposed nonlinear resonant decomposition [21]. Li et al. [22] presented a frequency-shift multiscale noise tuning SR method for fault diagnosis of generator bearings in wind turbines and even presented the coupled bistable SR method [23] in which the parameters of the proposed method are optimized through modified SNR and genetic algorithms. Fu et al. [24] studied the SR in Duffing oscillators and proposed a moment-method-based bearing fault diagnosis algorithm where the output SNR is seen as the objective function of optimization algorithms. Xu et al. [25] studied the SR behaviors in a high-order-degradation bistable system, and the ratio of the amplitudes around the target orders to those of the interference orders was calculated as the objective function of optimization algorithms. In summary, most of the proposed SR methods' objective functions depend on prior knowledge of the fault characteristics to be detected, which would result in false SR and weaken the detection performance of SR further. It is improper to enhance and detect unknown weak fault characteristics, especially for real-world equipment. In addition, a large number of tuning parameters in the existing SR methods need to be optimized by using artificial intelligence, which is very time-consuming and complex and is not suitable to engineering application.

Triggering SR has clear mathematical conditions, but among these variables in the conditions, the intensity of noise is indefinite. If we could estimate the intensity of noise in a signal, a parameter-matching equation could be built. Motivated by such an idea, this paper will attempt to design a parameter-matched second-order SR method for enhancing unknown weak fault characteristics of bearings due to second-order SR that has a bandpass filtering property. The remainder of this paper is organized as follows. Section 2 gives clear mathematical conditions of parameter-matched second-order SR and then proposes the corresponding SR method for bearing fault enhancement. In Section 3, a bearing fault

experiment was performed to demonstrate the effectiveness of the proposed method, and a comparison is made. Finally, conclusions are drawn in Section 4.

2. A Parameter-Matched Second-Order SR Method

A second-order SR can be described as below [26],

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} = -\frac{dV(x)}{dx} + s(t) + n(t) \quad (1)$$

where γ is the damping factor and $0 < \gamma \leq 2\sqrt{2a}$ [27], $s(t)$ is a periodic signal to be detected and its expression can be written as $A \cos(2\pi f_0 t + \varphi)$ with the amplitude A , driving frequency f_0 and initial phase φ . $n(t)$ is noise with $\langle n(t), n(t + \tau) \rangle = 2D\delta(t)$ and D as the noise intensity. $V(x)$ is the bistable potential as below,

$$V(x) = -\frac{a}{2}x^2 + \frac{b}{4}x^4, a, b > 0 \quad (2)$$

where two stable states and one unstable state are located at $\pm x_m = \pm\sqrt{a/b}$ and $x_0 = 0$, respectively. Moreover, the barrier height is $\Delta V = \frac{a^2}{4b}$. According to two-state theory, the matched conditions for triggering the SR induced by a periodic signal is [28]

$$r_K = 2f_0 \quad (3)$$

where r_K is the Kramers' rate given by [29]

$$r_K = \frac{\omega_b \omega_0}{2\pi\gamma} \exp\left(-\frac{\Delta V}{D}\right) \quad (4)$$

in which $\omega_b = \sqrt{2a}$ and $\omega_0 = \sqrt{a}$.

For convenience, we rewrite the matched condition in Equation (3) as [30]

$$F(a, b, D, \gamma, f_0) = \frac{a}{2\sqrt{2}\pi\gamma f_0} \exp\left(-\frac{a^2}{4bD}\right) \quad (5)$$

It can be noticed from Equation (5) that SR can be induced when $F = 1$. In general, the input and output SNR in Equation (1) for a periodic signal plus additive noise can be written as [31]

$$SNR_{\text{input}} = \frac{A^2}{4D} \quad (6)$$

$$SNR_{\text{output}} \approx \frac{aA^2\sqrt{2a}}{4bD^2} \exp\left(-\frac{a^2}{4bD}\right) \quad (7)$$

Then, the SNR gain (Signal-to-noise ratio gain, SNRG) can be calculated as

$$SNRG = \frac{SNR_{\text{output}}}{SNR_{\text{input}}} \approx \frac{a\sqrt{2a}}{bD} \exp\left(-\frac{a^2}{4bD}\right) \quad (8)$$

It can be seen from Equation (8) that the SNRG would be controlled by bistable potential parameters and the noise intensity. Thus, the objective function of the signal to be detected is given as the following optimization problem:

$$D_{\text{opt}} = \arg\max SNRG \quad (9)$$

By solving Equation (8) with respect to D , we can obtain the optimal condition as below:

$$D_{\text{opt}} = \frac{a^2}{4b} = \Delta V \quad (10)$$

Therefore, the optimal parameter-matched condition for weak fault enhancement can be given as below:

$$\begin{cases} F = 1 \\ D = \frac{a^2}{4b} \end{cases} \quad (11)$$

However, Equation (1) is suitable to process weak signals subject to small-parameter limitation. In the real world, the mechanical signals are generally large parameter. For solving this issue, a normalized scale transformation is performed to Equation (1). Let $z = x\sqrt{b/a}$ and $\tau = at$, Equation (1) could be rewritten as

$$\frac{d^2z}{d\tau^2} + \gamma_h \frac{dz}{d\tau} = a_h z - b_h z^3 + A_h \cos(2\pi f_{0h}\tau + \varphi) + D_h \xi\left(\frac{\tau}{a}\right) \quad (12)$$

where $\gamma_h = \gamma/a$, $a_h = b_h = 1/a$, $f_{0h} = f_0/a$, $A_h = \sqrt{b/a^5}A$, and $D_h = b/a^5D$. By comparison, it is found that large-parameter signals can be transferred into small-parameter ones in which the amplitude is reduced by the time of $\sqrt{b/a^5}$ and the frequency is reduced by the time of $1/a$. By substituting the above parameters in Equation (12) into Equation (11), the parameter-matched conditions for triggering SR for weak large-parameter signal detection can be achieved as below:

$$\begin{cases} F(a_h, b_h, D_h, f_{0h}) = \frac{a_h}{2\sqrt{2\pi}\gamma_h f_{0h}} \exp\left(-\frac{a_h^2/4b_h}{D_h}\right) = 1 \\ D_h = a_h^2/4b_h \end{cases} \quad (13)$$

which is calculated further as

$$\begin{cases} a = 2\sqrt{2\pi}f_0\gamma e \\ b = a^4/4D \end{cases} \quad (14)$$

It can be seen from Equation (14) that the damping factor γ can be adjusted to enhance the weak fault characteristics of bearings. According to the condition $0 < \gamma_h \leq 2\sqrt{2a_h}$, we can obtain the limitation.

$$0 < \gamma < 16\sqrt{2\pi}f_0e \quad (15)$$

Therefore, the parameter-matched conditions can be given by using Equations (14) and (15) mathematically for weak large-parameter signal detection. According to the above parameter-matched conditions, we can design an SR method with parameter estimation to enhance the weak large-parameter fault signature of bearings. The proposed method is shown in Figure 1, and its detailed steps are given as below.

1. Signal pre-processing. In general, the bearing vibration signal $v(t)$ is a large-parameter signal where the bearing fault characteristics are modulated by natural vibration from the machinery itself. Hence, some signal demodulations are performed to the raw signals. Here, Hilbert envelope demodulation is used to pre-process the raw vibration signal of the tested bearing, and the corresponding envelope signal is denoted as $\tilde{v}(t)$.
2. Noise intensity estimation. In Equation (14), an important parameter D , namely, noise intensity, needs to be estimated from the envelope signal $\tilde{v}(t)$ of the raw signal of the tested bearing. Here, we use the principle of maximum likelihood estimation (MLE) [32] to achieve this, and we can obtain $D = MLE(\tilde{v}(t))$. The MLE algorithm can be downloaded by using the following website: http://www.biomecardio.com/matlab/evan_doc.html (accessed on 5 December 2022).
3. Damping factor initialization. The damping factor needs to be tuned to obtain the optimal detected result. Here, according to its range in Equation (15), we initialize the damping factor γ .

4. Output signal calculation and evaluation. By substituting $a_{\text{match}} = 2\sqrt{2}\pi f_0 \gamma e$, $b_{\text{match}} = a^4/4D$ and the corresponding damping factor into Equation (16), we can obtain the output signal $x(t)$. Then, we can calculate the corresponding $SNRG$ as the objective function $\gamma_{\text{opt}} = \text{argmax } SNRG$ for optimizing the damping factor. Finally, the optimal γ_{opt} is substituted into Equation (17) to solve the optimal $x(t)$.

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} = a_{\text{match}}x - b_{\text{match}}x^3 + \tilde{v}(t) \quad (16)$$

$$\frac{d^2x}{dt^2} + \gamma_{\text{opt}} \frac{dx}{dt} = a_{\text{match}}x - b_{\text{match}}x^3 + \tilde{v}(t) \quad (17)$$

5. Signal post-processing. Output the optimal $x(t)$ corresponding to the maximum of $SNRG$ as the detected signal. Here, the frequency spectral analysis is used to process the optimal $x(t)$ for observing the spectral peaks at the fault characteristic frequencies of bearings.

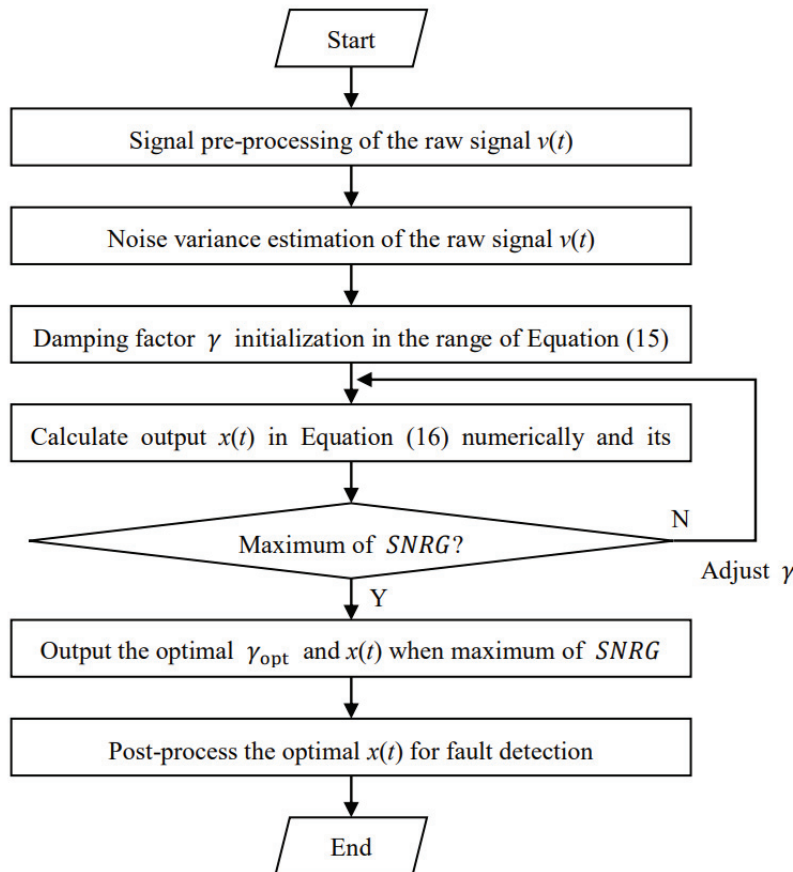


Figure 1. The flowchart of the proposed method.

3. Bearing Fault Experimental Verification

The bearing fault experimental setup is shown in Figure 2 where a bearing accelerated degradation test was performed throughout the whole operating life of the bearing, and the corresponding vibration signals were acquired by using two sensors [33]. The two sensors were placed on the housing of the tested bearings and positioned at 90 degrees to each other, i.e., one was placed on the vertical axis and the other one was placed on the horizontal axis. The tested bearing's parameters are given in Table 1. The sampling frequency was 25.6 kHz and 32,768 samples (i.e., 1.28 s) were recorded every 1 min. In addition, the rotating speed was 2100 rpm, and the radial force was 12 kN.

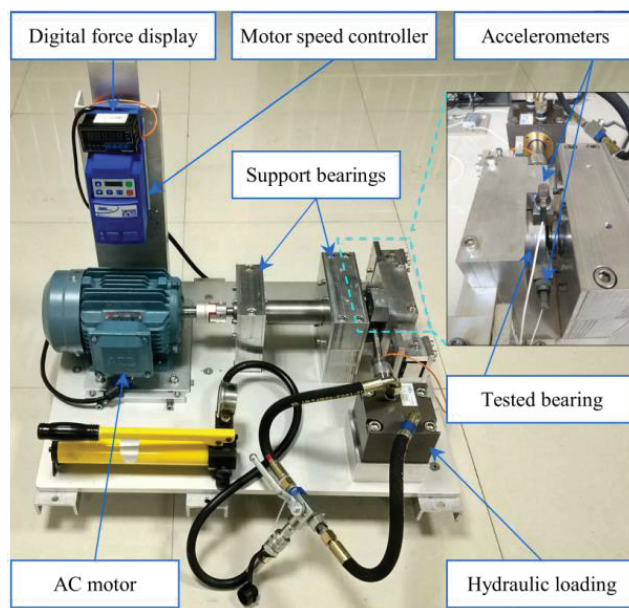


Figure 2. Bearing fault experimental setup.

Table 1. The tested bearing's parameters.

Parameters	Values	Parameters	Values
Outer race diameter	39.80 mm	Inner race diameter	29.30 mm
Mean diameter	34.55 mm	Ball diameter	7.92 mm
Number of balls	8	Contact angle	0°

In the process of the bearing accelerated degradation test, inner race and outer race wear occurred in the tested bearing, as shown in Figure 3. Figure 4 depicts the horizontal and vertical root mean square (RMS) indicators of the tested bearing throughout its whole operating life. It can be seen from Figure 3 that horizontal and vertical RMS indicators were kept unchanged initially and then rose rapidly after 30 min, suggesting that early wear may have occurred in the tested bearing. According to the horizontal and vertical RMS indicators, we can deduce that early compound wear may happen in the inner race and outer race of the tested bearing at the 24th min. Therefore, the horizontal and vertical raw vibration signals and their spectrum are shown in Figures 5 and 6, respectively. In the figures, we mark the inner race, outer race, roller and cage fault characteristic frequencies in the envelope spectrum by using different colors. According to the tested bearing's parameters, we can calculate the theoretical inner race, outer race, roller and cage fault characteristic frequencies as 172.09, 107.91, 72.33 and 13.49 Hz, respectively, by using the following equations:

$$f_{inner} = \frac{N}{2} f_r \left(1 + \frac{d}{D} \cos \theta \right) \quad (18)$$

$$f_{outer} = \frac{N}{2} f_r \left(1 - \frac{d}{D} \cos \theta \right) \quad (19)$$

$$f_{cage} = \frac{1}{2} f_r \left(1 - \frac{d}{D} \cos \theta \right) \quad (20)$$

$$f_{roller} = \frac{D}{d} f_r \left(1 - \left(\frac{d}{D} \cos \theta \right)^2 \right) \quad (21)$$

where N is the number of balls, f_r is the rotating frequency, d is the ball diameter, D is the mean diameter and θ is the contact angle.

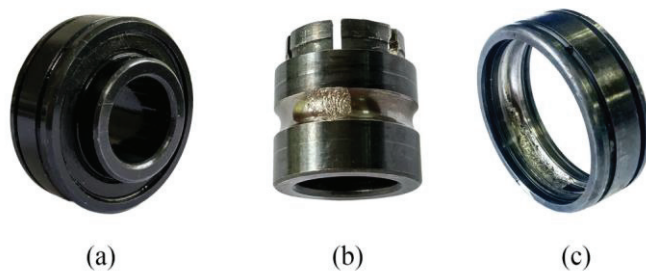


Figure 3. Photos of tested bearings: (a) Normal bearing, (b) Inner race wear and (c) Outer race wear.

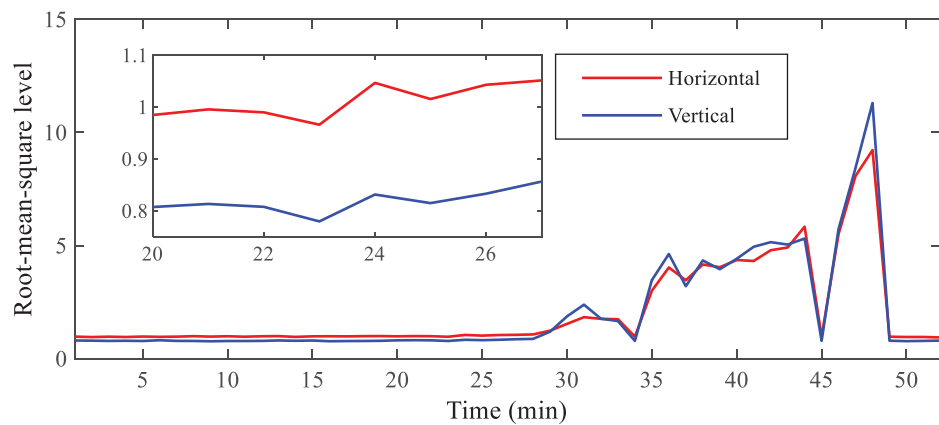


Figure 4. Horizontal and vertical RMS indicators of the tested bearing throughout its whole operating life.

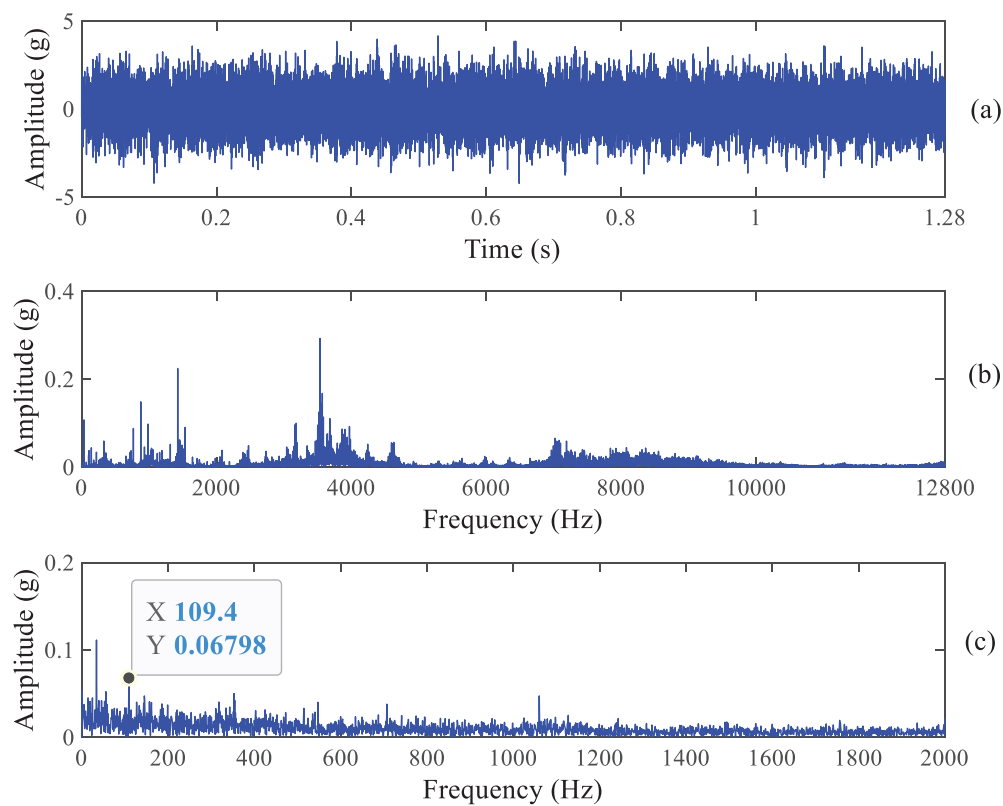


Figure 5. Horizontal raw vibration signal and its spectrum. (a) Raw signal, (b) Frequency spectrum and (c) Zoomed envelope spectrum.

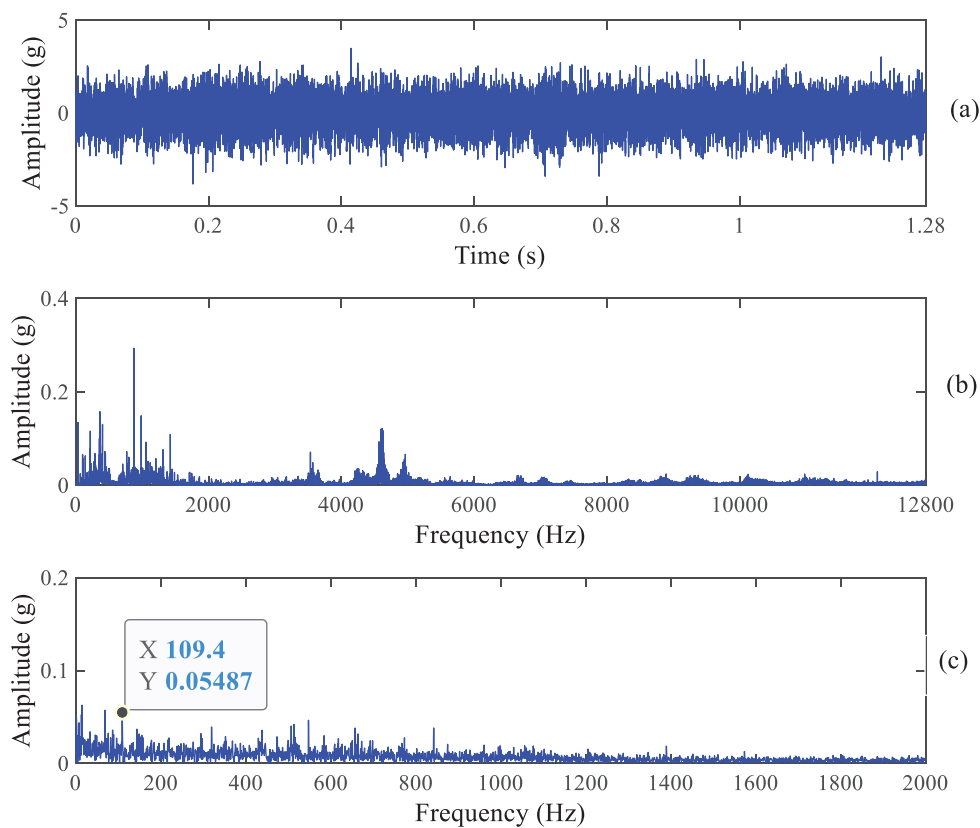


Figure 6. Vertical raw vibration signal and its spectrum: (a) Raw signal, (b) Frequency spectrum and (c) Zoomed envelope spectrum.

Compared with the information in the envelope spectrum in Figures 5 and 6, it is found that there is a clear spectrum peak at 109.4 Hz which is close to the theoretical value of the outer race fault characteristic frequency 107.91 Hz, but the spectrum peaks of its harmonics cannot be recognized. Obviously, the above signature demonstrates that early outer race wear occurs, which is consistent with the real experimental results. In addition, we can see from Figure 5 that there is a clear spectrum peak at 68.75 Hz, which is not a roller fault characteristic frequency. Other information cannot be observed from Figures 5 and 6. As a result, we can deduce that early outer race wear occurs in the tested bearing from the raw signal and its spectrum.

Although we can observe the weak spectrum peak at the outer race fault characteristic frequency, it cannot recognize the fault information excited by inner race wear. Therefore, the proposed method is used to process the raw vibration signal of the tested bearing. Inner race and outer race fault information can be enhanced by using the proposed method, as shown in Figures 7 and 8. It is found from Figure 7b that there is a clear spectrum peak at 171.9 Hz which is close to the theoretical value of the inner race fault characteristic frequency 172.09 Hz, suggesting that early inner race wear happens in the tested bearing. In fact, an inner race wear has been formed as shown in Figure 3c. Meanwhile, we can also observe the clear spectrum peak at 109.4 Hz from Figure 8b which is close to the theoretical value of the outer race fault characteristic frequency 107.91 Hz, indicating that early outer race wear happens. The period of enhanced signal in Figures 7a and 8a is equal to an inner race or outer race fault period, respectively.

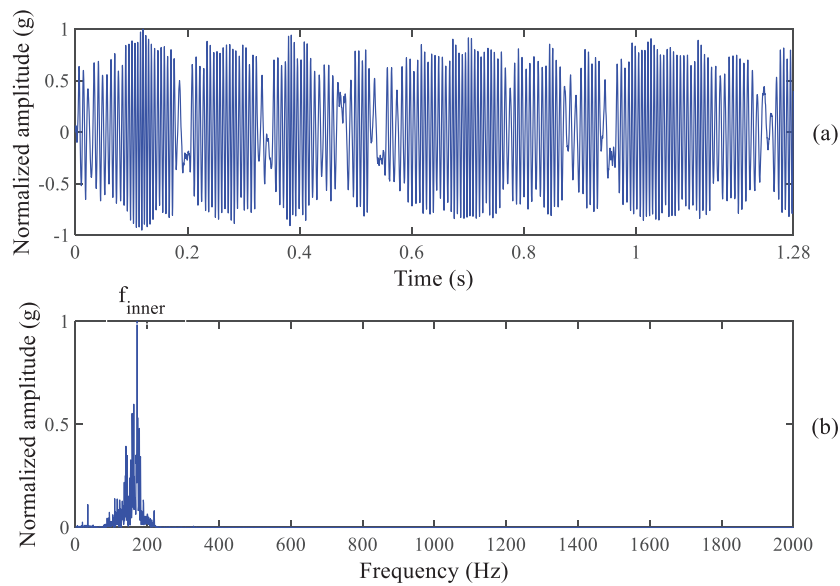


Figure 7. Inner race wear diagnosis result: (a) Enhanced signal (b) Zoomed frequency spectrum.

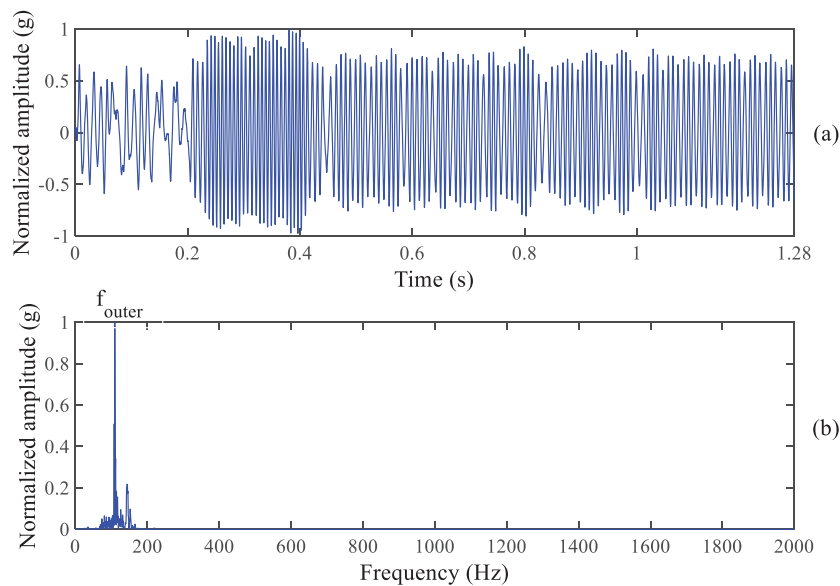


Figure 8. Outer race wear diagnosis result: (a) Enhanced signal (b) Zoomed frequency spectrum.

For comparison, the fast kurtogram [34,35], as a widely used method in mechanical fault diagnosis, is applied to process the raw vibration signal of the tested bearing. The kurtogram as shown in Figure 9 selects the optimal filtering frequency band as the optimal carrier frequency 3900 Hz with bandwidth 200 Hz to filter out the weak fault signature excited by the outer race and inner race wear in the tested bearing. The corresponding filtered result is shown in Figure 10. It can be seen from the amplitude spectrum of the filtered signal squared envelope in Figure 10 that there are clear spectrum peaks at 34.68 Hz and even at 75.81 Hz. Obviously, these are not fault signatures excited by inner race and outer race wear of the tested bearing. Therefore, we deduce that the optimal filtering frequency band may be the optimal carrier frequency 7866 Hz, as shown in Figure 9. The corresponding filtered signal is depicted in Figure 11. We cannot also observe clear spectrum peaks at the inner race and outer race fault characteristic frequencies from the amplitude spectrum of the filtered signal squared envelope in Figure 11. As a result, the fast kurtogram fails to detect the weak fault characteristics of the tested bearing

in this experiment. The comparison demonstrates the feasibility and superiority of the proposed method.

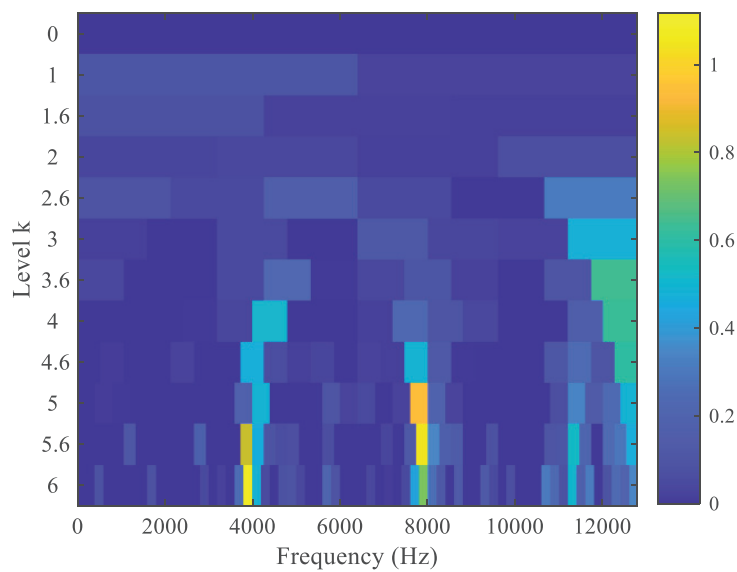


Figure 9. The kurtogram using the fast kurtogram method.

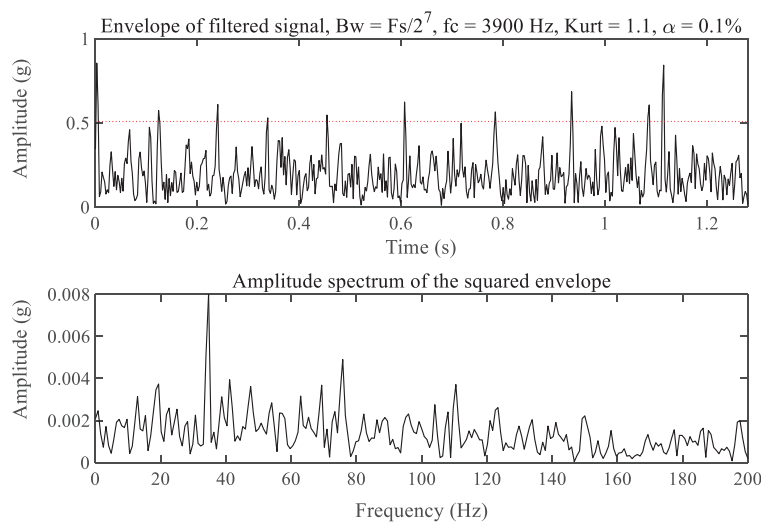


Figure 10. The filtered signal with the optimal carrier frequency 3900 Hz.

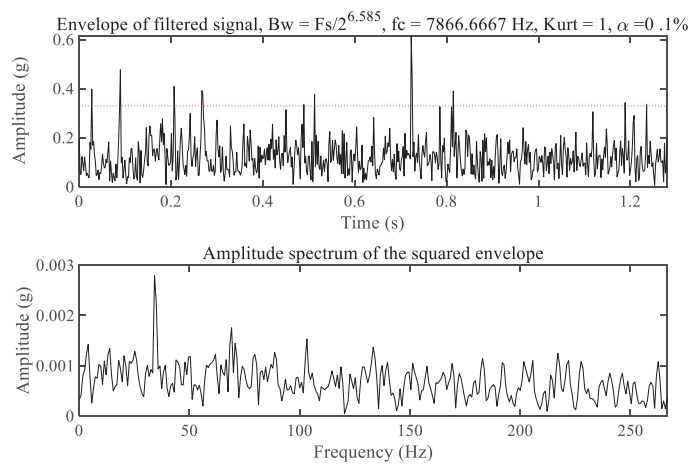


Figure 11. The filtered signal with the optimal carrier frequency 7866 Hz.

4. Conclusions

Stochastic resonances (SR) have been widely used in mechanical fault diagnosis, but most of them require designing the quantification indicators by virtue of prior knowledge of the fault signature to be detected, resulting in false SR and further impaired detection performance of SR. Therefore, it is very necessary for us to study the parameter-matched SR method based on the mathematical condition of triggering SR. For this purpose, we propose an SR method with parameter estimation for bearing fault diagnosis. Compared with other SR methods, the tuning parameters of the proposed method can be estimated by virtue of the raw signal of bearings and do not depend on prior knowledge of the defects to be detected. Moreover, the proposed method considers the conditions of triggering SR to avoid false SR and improve the detection performance of SR further. Such an advantage makes the proposed method more suitable in engineering application. In addition, the proposed method has a tuning parameter and a damping factor with clear mathematical limitation, making it more simple and less time-consuming than other SR methods with a large amount of parameter optimization. A bearing fault experiment was performed to verify this. The experimental results indicate that the proposed method can enhance the weak compound fault diagnosis of bearings at an early stage and is superior to the widely used fast kurtogram method.

Author Contributions: Conceptualization, M.X. and C.Z.; methodology, Z.Q. and K.S.; software, Z.L. and K.S.; validation, K.S., L.X. and Z.Q.; formal analysis, Z.L.; investigation, M.X.; resources, M.X.; data curation, M.X. and L.X.; writing—original draft preparation, M.X. and L.X.; writing—review and editing, K.S. and Z.Q.; visualization, C.Z. and L.X.; supervision, Z.Q.; project administration, Z.Q. and Z.L.; funding acquisition, Z.Q. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Zhejiang Provincial Natural Science Foundation of China (LQ22E050003), Foundation of the State Key Laboratory of Performance Monitoring and Protecting of Rail Transit Infrastructure of East China Jiaotong University (HJGZ2021114), Laboratory of Yangjiang Offshore Wind Power (YJOFWD-OF-2022A08), National Natural Science Foundation of China (52205569, 51905349, U2013603), Chuying Planning Project of Zhejiang Provincial Administration for Market Regulation (CY2023328), Shandong Provincial Innovation Ability Improvement Project of Middle and Small-sized High-tech Enterprises (2022TSGC2364), Ningbo Science and Technology Major Project (2020Z110, 2022Z057, 2022Z002), Ningbo Natural Science Foundation (2022J132), Guangdong Basic and Applied Basic Research Foundation (Grant Nos. 2022A1515010126, 2020A1515011509) and was also sponsored by K.C. Wong Magna Fund in Ningbo University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kim, W.; Kim, S.; Jeong, J.; Kim, H.; Lee, H.Y.; Youn, B.D. Digital twin approach for on-load tap changers using data-driven dynamic model updating and optimization-based operating condition estimation. *Mech. Syst. Signal Process.* **2022**, *181*, 109471. [CrossRef]
2. Jerez, D.J.; Jensen, H.A.; Beer, M. An effective implementation of reliability methods for Bayesian model updating of structural dynamic models with multiple uncertain parameters. *Reliab. Eng. Syst. Saf.* **2022**, *225*, 108634.
3. Liu, Y.; Jiang, H.; Liu, C.; Yang, W.; Sun, W. Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis. *Knowledge-Based Syst.* **2022**, *252*, 109439. [CrossRef]
4. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]
5. Wang, Z.; Zhou, J.; Du, W.; Lei, Y.; Wang, J. Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution. *Mech. Syst. Signal Process.* **2022**, *162*, 108018. [CrossRef]
6. Liang, M.; Zhou, K. Probabilistic bearing fault diagnosis using Gaussian process with tailored feature extraction. *Int. J. Adv. Manuf. Technol.* **2021**, *119*, 2059–2076. [CrossRef]

7. Song, W.; Liu, H.; Zio, E. Long-range dependence and heavy tail characteristics for remaining useful life prediction in rolling bearing degradation. *Appl. Math. Model.* **2022**, *102*, 268–284. [CrossRef]
8. Liu, H.; Song, W.; Zhang, Y.; Kudreyko, A. Generalized Cauchy Degradation Model With Long-Range Dependence and Maximum Lyapunov Exponent for Remaining Useful Life. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
9. Lu, S.; He, Q.; Wang, J. A review of stochastic resonance in rotating machine fault detection. *Mech. Syst. Signal Process.* **2018**, *116*, 230–260. [CrossRef]
10. Qiao, Z.; Lei, Y.; Li, N. Applications of stochastic resonance to machinery fault detection: A review and tutorial. *Mech. Syst. Signal Process.* **2019**, *122*, 502–536. [CrossRef]
11. López, C.; Naranjo, A.; Lu, S.; Moore, K.J. Hidden Markov Model based Stochastic Resonance and its Application to Bearing Fault Diagnosis. *J. Sound Vib.* **2022**, *528*, 116890. [CrossRef]
12. Lin, Y.; Xu, X.; Ye, C. Adaptive stochastic resonance quantified by a novel evaluation index for rotating machinery fault diagnosis. *Measurement* **2021**, *184*, 109920. [CrossRef]
13. Li, Z.; Liu, X.; Wang, X.; He, T.; Shan, Y. A multi-parameter constrained potential underdamped stochastic resonance method and its application for weak fault diagnosis. *J. Sound Vib.* **2019**, *459*, 114862. [CrossRef]
14. Lai, Z.H.; Liu, J.S.; Zhang, H.T.; Zhang, C.L.; Zhang, J.W.; Duan, D.Z. Multi-parameter-adjusting stochastic resonance in a standard tri-stable system and its application in incipient fault diagnosis. *Nonlinear Dyn.* **2019**, *96*, 2069–2085. [CrossRef]
15. Zhang, X.; Wang, J.; Liu, Z.; Wang, J. Weak feature enhancement in machinery fault diagnosis using empirical wavelet transform and an improved adaptive bistable stochastic resonance. *ISA Trans.* **2018**, *84*, 283–295. [CrossRef]
16. Wang, Z.; Yang, J.; Guo, Y.; Gong, T.; Shan, Z. Positive role of bifurcation on stochastic resonance and its application in fault diagnosis under time-varying rotational speed. *J. Sound Vib.* **2022**, *537*, 117210. [CrossRef]
17. Zhou, D.; Huang, D.; Hao, J.; Ren, Y.; Jiang, P.; Jia, X. Vibration-based fault diagnosis of the natural gas compressor using adaptive stochastic resonance realized by Generative Adversarial Networks. *Eng. Fail. Anal.* **2020**, *116*, 104759. [CrossRef]
18. Shi, P.; Li, M.; Zhang, W.; Han, D. Weak signal enhancement for machinery fault diagnosis based on a novel adaptive multi-parameter unsaturated stochastic resonance. *Appl. Acoust.* **2022**, *189*, 108609. [CrossRef]
19. Qiao, Z.; Elhatab, A.; Shu, X.; He, C. A second-order stochastic resonance method enhanced by fractional-order derivative for mechanical fault detection. *Nonlinear Dyn.* **2021**, *106*, 707–723. [CrossRef]
20. Qiao, Z.; Liu, J.; Xu, X.; Yin, A.; Shu, X. Nonlinear resonance decomposition for weak signal detection. *Rev. Sci. Instruments* **2021**, *92*, 105102. [CrossRef]
21. Qiao, Z.; Shu, X. Coupled neurons with multi-objective optimization benefit incipient fault identification of machinery. *Chaos Solitons Fractals* **2021**, *145*, 110813. [CrossRef]
22. Li, J.; Li, M.; Zhang, J.; Jiang, G. Frequency-shift multiscale noise tuning stochastic resonance method for fault diagnosis of generator bearing in wind turbine. *Measurement* **2018**, *133*, 421–432. [CrossRef]
23. Li, J.; Zhang, J.; Li, M.; Zhang, Y. A novel adaptive stochastic resonance method based on coupled bistable systems and its application in rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2019**, *114*, 128–145. [CrossRef]
24. Fu, Y.; Kang, Y.; Liu, R. Novel Bearing Fault Diagnosis Algorithm Based on the Method of Moments for Stochastic Resonant Systems. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–10. [CrossRef]
25. Xu, H.; Zhou, S.; Yang, T. Stochastic resonance of a high-order-degradation bistable system and its application in fault diagnosis with variable speed condition. *Mech. Syst. Signal Process.* **2023**, *186*. [CrossRef]
26. Qiao, Z.; Chen, S.; Lai, Z.; Zhou, S.; Sanjuán, M.A.F. Harmonic-Gaussian double-well potential stochastic resonance with its application to enhance weak fault characteristics of machinery. *Nonlinear Dyn.* **2023**, 1–15. [CrossRef]
27. Dong, H.; Wang, H.; Shen, X.; Jiang, Z. Effects of Second-Order Matched Stochastic Resonance for Weak Signal Detection. *IEEE Access* **2018**, *6*, 46505–46515. [CrossRef]
28. Gammaitoni, L.; Hänggi, P.; Jung, P.; Marchesoni, F. Stochastic resonance. *Rev. Mod. Phys.* **1998**, *70*, 223. [CrossRef]
29. Li, F.; Duan, F.; Chapeau-Blondeau, F.; Abbott, D. Signal estimation and filtering from quantized observations via adaptive stochastic resonance. *Phys. Rev. E* **2021**, *103*(5), 052108. [CrossRef]
30. Dong, H.; Wang, H.; Shen, X.; He, K. Parameter matched stochastic resonance with damping for passive sonar detection. *J. Sound Vib.* **2019**, *458*, 479–496. [CrossRef]
31. Lu, S.; He, Q.; Kong, F. Effects of underdamped step-varying second-order stochastic resonance for weak signal detection. *Digit. Signal Process.* **2015**, *36*, 93–103. [CrossRef]
32. Kay, S.M. *Fundamentals of statistical signal processing: Estimation theory*; Prentice-Hall, Inc.: New York, NY, USA, 1993; pp. 65–75.
33. Wang, B.; Lei, Y.; Li, N.; Li, N. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Reliab.* **2018**, *69*, 401–412. [CrossRef]
34. Antoni, J. Fast computation of the kurtogram for the detection of transient faults. *Mech. Syst. Signal Process.* **2007**, *21*, 108–124. [CrossRef]
35. Lei, Y.; Lin, J.; He, Z.; Zi, Y. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2011**, *25*, 1738–1749. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Sensors Editorial Office
E-mail: sensors@mdpi.com
www.mdpi.com/journal/sensors



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-4966-6