

Complexity

Fault Identification, Diagnosis, and Prognostics Based on Complex Signal Analysis

Lead Guest Editor: Minvydas Ragulskis

Guest Editors: Chen Lu, Mao-Sen Cao, Gangbing Song, and Rafał Burdzik





Fault Identification, Diagnosis, and Prognostics Based on Complex Signal Analysis

Complexity

Fault Identification, Diagnosis, and Prognostics Based on Complex Signal Analysis

Lead Guest Editor: Minvydas Ragulskis

Guest Editors: Chen Lu, Mao-Sen Cao, Gangbing Song,
and Rafal Burdzik



Copyright © 2018 Hindawi. All rights reserved.

This is a special issue published in “Complexity.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

- José A. Acosta, Spain
Carlos F. Aguilar-Ibáñez, Mexico
Mojtaba Ahmadiéh Khanesar, UK
Tarek Ahmed-Ali, France
Alex Alexandridis, Greece
Basil M. Al-Hadithi, Spain
Juan A. Almendral, Spain
Diego R. Amancio, Brazil
David Arroyo, Spain
Mohamed Boutayeb, France
Átila Bueno, Brazil
Arturo Buscarino, Italy
Guido Caldarelli, Italy
Eric Campos-Canton, Mexico
Mohammed Chadli, France
Émile J. L. Chappin, Netherlands
Diyi Chen, China
Yu-Wang Chen, UK
Giulio Cimini, Italy
Danilo Comminiello, Italy
Sara Dadras, USA
Sergey Dashkovskiy, Germany
Manlio De Domenico, Italy
Pietro De Lellis, Italy
Albert Diaz-Guilera, Spain
Thach Ngoc Dinh, France
Jordi Duch, Spain
Marcio Eisencraft, Brazil
Joshua Epstein, USA
Mondher Farza, France
Thierry Floquet, France
Mattia Frasca, Italy
José Manuel Galán, Spain
Lucia Valentina Gambuzza, Italy
Bernhard C. Geiger, Austria
Carlos Gershenson, Mexico
Peter Giesl, UK
Sergio Gómez, Spain
Lingzhong Guo, UK
Xianggui Guo, China
Sigurdur F. Hafstein, Iceland
Chittaranjan Hens, Israel
Giacomo Innocenti, Italy
Sarangapani Jagannathan, USA
Mahdi Jalili, Australia
Jeffrey H. Johnson, UK
M. Hassan Khooban, Denmark
Abbas Khosravi, Australia
Toshikazu Kuniya, Japan
Vincent Labatut, France
Lucas Lacasa, UK
Guang Li, UK
Qingdu Li, Germany
Chongyang Liu, China
Xiaoping Liu, Canada
Xinzhi Liu, Canada
Rosa M. Lopez Gutierrez, Mexico
Vittorio Loreto, Italy
Noureddine Manamanni, France
Didier Maquin, France
Eulalia Martínez, Spain
Marcelo Messias, Brazil
Ana Meštrović, Croatia
Ludovico Minati, Japan
Ch. P. Monterola, Philippines
Marcin Mrugalski, Poland
Roberto Natella, Italy
Sing Kiong Nguang, New Zealand
Nam-Phong Nguyen, USA
B. M. Ombuki-Berman, Canada
Irene Otero-Muras, Spain
Yongping Pan, Singapore
Daniela Paolotti, Italy
Cornelio Posadas-Castillo, Mexico
Mahardhika Pratama, Singapore
Luis M. Rocha, USA
Miguel Romance, Spain
Avimanyu Sahoo, USA
Matilde Santos, Spain
Josep Sardanyés Cayuela, Spain
Ramaswamy Savitha, Singapore
Hiroki Sayama, USA
Michele Scarpiniti, Italy
Enzo Pasquale Scilingo, Italy
Dan Selişteanu, Romania
Dehua Shen, China
Dimitrios Stamovlasis, Greece
Samuel Stanton, USA
Roberto Tonelli, Italy
Shahadat Uddin, Australia
Gaetano Valenza, Italy
Dimitri Volchenkov, USA
Christos Volos, Greece
Zidong Wang, UK
Yan-Ling Wei, Singapore
Honglei Xu, Australia
Yong Xu, China
Xinggang Yan, UK
Baris Yuçe, UK
Massimiliano Zanin, Spain
Hassan Zargarzadeh, USA
Rongqing Zhang, USA
Xianming Zhang, Australia
Xiaopeng Zhao, USA
Quanmin Zhu, UK

Contents

Fault Identification, Diagnosis, and Prognostics Based on Complex Signal Analysis

Minvydas Ragulskis , Chen Lu , Maosen Cao , Gangbing Song , and Rafal Burdzik 
Editorial (2 pages), Article ID 4020729, Volume 2018 (2018)

Fault Diagnosis of Rolling Bearing Based on a Novel Adaptive High-Order Local Projection Denoising Method

Rui Yuan , Yong Lv , and Gangbing Song 
Research Article (15 pages), Article ID 3049318, Volume 2018 (2018)

Damage Diagnosis in 3D Structures Using a Novel Hybrid Multiobjective Optimization and FE Model Updating Framework

Nizar Faisal Alkayem , Maosen Cao , and Minvydas Ragulskis 
Research Article (13 pages), Article ID 3541676, Volume 2018 (2018)

Fault Diagnosis of Electromechanical Actuator Based on VMD Multifractal Detrended Fluctuation Analysis and PNN

Hongmei Liu , Jiayao Jing , and Jian Ma 
Research Article (11 pages), Article ID 9154682, Volume 2018 (2018)

Predicting the Remaining Useful Life of an Aircraft Engine Using a Stacked Sparse Autoencoder with Multilayer Self-Learning

Jian Ma , Hua Su , Wan-lin Zhao , and Bin Liu 
Research Article (13 pages), Article ID 3813029, Volume 2018 (2018)

Study on the Magnitude of Reservoir-Triggered Earthquake Based on Support Vector Machines

Hai Wei , Mingming Wang , Bingyue Song, Xin Wang, and Danlei Chen
Research Article (10 pages), Article ID 2830690, Volume 2018 (2018)

A Novel Fault Diagnosis Method for Rolling Bearing Based on Improved Sparse Regularization via Convex Optimization

Dongjie Zhong, Cancan Yi , Han Xiao , Houzhuang Zhang, and Anding Wu
Research Article (10 pages), Article ID 2169364, Volume 2018 (2018)

An Investigation of Stretched Exponential Function in Quantifying Long-Term Memory of Extreme Events Based on Artificial Data following Lévy Stable Distribution

HongGuang Sun , Lin Yuan, Yong Zhang , and Nicholas Privitera
Research Article (7 pages), Article ID 5913976, Volume 2018 (2018)

A Modified Time Reversal Method for Guided Wave Detection of Bolt Loosening in Simulated Thermal Protection System Panels

Guan-nan Wu, Chao Xu , Fei Du, and Wei-dong Zhu
Research Article (12 pages), Article ID 8210817, Volume 2018 (2018)

Multiplicative Fault Estimation-Based Adaptive Sliding Mode Fault-Tolerant Control Design for Nonlinear Systems

Ali Ben Brahim , Slim Dhahri, Fayçal Ben Hmida, and Anis Sellami
Research Article (15 pages), Article ID 1462594, Volume 2018 (2018)

Rigorous Solution of Slopes' Stability considering Hydrostatic Pressure

Chengchao Li , Pengming Jiang, and Aizhao Zhou

Research Article (10 pages), Article ID 2829873, Volume 2018 (2018)

Research Progress on Monitoring and Separating Suspension Particles for Lubricating Oil

Ziping Wang , Xian Xue, He Yin, Zhengxuan Jiang, and Yefei Li

Review Article (9 pages), Article ID 9356451, Volume 2018 (2018)

Contactless Modal Phenomena Based Approach to Detecting, Identifying, and Diagnosing of Electrical Connections

Pavel Orlov  and Talgat Gazizov 

Research Article (12 pages), Article ID 5081684, Volume 2018 (2018)

A General Purpose Adaptive Fault Detection and Diagnosis Scheme for Information Systems with Superheterodyne Receivers

Dengwei Song, Hongmei Liu , Le Qi , and Bo Zhou

Research Article (9 pages), Article ID 4763612, Volume 2018 (2018)

Statistical Identification of Parameters for Damaged FGM Structures with Material Uncertainties in Thermal Environment

Yalan Xu , Yu Qian, and Kongming Guo

Research Article (21 pages), Article ID 9034865, Volume 2018 (2018)

Gear Fault Diagnosis in Variable Speed Condition Based on Multiscale Chirplet Path Pursuit and Linear Canonical Transform

Xu Shuiqing , Zhang Ke , Chai Yi, He Yigang, and Feng Li 

Research Article (8 pages), Article ID 3904598, Volume 2018 (2018)

Fault Diagnosis for Hydraulic Servo System Using Compressed Random Subspace Based ReliefF

Yu Ding , Fei Wang , Zhen-ya Wang , and Wen-jin Zhang 

Research Article (14 pages), Article ID 8740989, Volume 2018 (2018)

Damage Detection of Refractory Based on Principle Component Analysis and Gaussian Mixture Model

Changming Liu, Di Zhou, Zhigang Wang, Dan Yang, and Gangbing Song 

Research Article (9 pages), Article ID 7356189, Volume 2018 (2018)

Lithium-Ion Battery Capacity Estimation: A Method Based on Visual Cognition

Yujie Cheng, Laifa Tao, and Chao Yang

Research Article (13 pages), Article ID 6342170, Volume 2017 (2018)

Editorial

Fault Identification, Diagnosis, and Prognostics Based on Complex Signal Analysis

Minvydas Ragulskis ¹, **Chen Lu** ², **Maosen Cao** ³,
Gangbing Song ⁴, and **Rafal Burdzik** ⁵

¹Center for Nonlinear Systems, Kaunas University of Technology, Kaunas, Lithuania

²School of Reliability and Systems Engineering, Beihang University, Beijing, China

³Institute of Vibration and Dynamics, Hohai University, Nanjing, China

⁴Department of Mechanical Engineering, University of Houston, Houston, USA

⁵Faculty of Transport, Silesian University of Technology, Katowice, Poland

Correspondence should be addressed to Minvydas Ragulskis; minvydas.ragulskis@ktu.lt

Received 21 October 2018; Accepted 6 December 2018; Published 17 December 2018

Copyright © 2018 Minvydas Ragulskis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This special issue titled “Fault Identification, Diagnosis, and Prognostics Based on Complex Signal Analysis” is a truly interdisciplinary issue. Manuscripts published in this special issue do represent such diverse fields as mechanical engineering, aviation engineering and technology, electric and electronic engineering, statistics, and so on. The geography of authors spans over three continents, Asia, Europe, and America. A total of 18 manuscripts were published in this issue.

This special issue gathered several studies that focus on fault detection and diagnosis of mechanical products, such as bolts, rolling bearings, gears, and hydraulic servo systems. For example, G. Wu et al. proposed a modified time reversal method and successfully realized bolt loosening detection and localization in simulated thermal protection system panels. Two studies aim to reduce the background noise in monitoring signal of rolling bearings, thus improving the effectiveness of fault diagnosis. R. Yuan et al. reported an adaptive high-order local projection denoising method and demonstrated the characteristic frequencies of simulated signals can be well extracted by the proposed method. D. Zhong et al. proposed a novel fault signal denoising scheme based on improved sparse regularization via convex optimization to extract the fault feature of rolling bearing. Experiments show that the proposed method has a better performance

than traditional methods. In an interesting study, a method based on the multiscale chirplet path pursuit and the linear canonical transform is proposed by X. Shuiqing et al. and applied to diagnose the gear fault in the variable speed condition for the first time. This method can diagnose the gear faults available. In another study, Y. Ding et al. presented a fault diagnosis scheme for hydraulic servo system using compressed random subspace based ReliefF (CRSR) method. The proposed CRSR method is able to enhance the robustness of the feature information against interference while selecting the feature combination with balanced information expressing ability. Finally, dealing with sliding mode Fault Tolerant Control (FTC) problem for nonlinear system, two adaptive sliding mode FTC schemes for an uncertain nonlinear system subject to multiplicative and process faults were presented by A. Ben Brahim et al. By solving a single-step multiobjective LMI optimization problem, the observer and controller gains are obtained, offering a solution to stabilize the closed-loop nonlinear system despite the occurrence of real fault effects.

This special issue also attracted several studies from the field of aviation engineering and technology. H. Liu et al., for example, proposed a fault diagnosis method for electromechanical actuators (EMAs) based on variational mode decomposition (VMD) multifractal detrended fluctuation analysis (MFDFA) and probabilistic neural network

(PNN). This method is demonstrated to achieve effective fault diagnosis for EMAs under different working conditions. In another elegant study, a deep learning approach is proposed by J. Ma et al. to predict the RUL of an aircraft engine based on a stacked sparse autoencoder and logistic regression. The proposed method also has significance for enhancing the safety of aircraft engines and prognosticating and managing the health of aircraft engines to reduce the cost of maintenance. In addition, as an interdisciplinary method, Y. Cheng et al. introduced visual cognition theory into the field of lithium-ion battery capacity estimation. Inspired by multi-channel and manifold sensing characteristics of human visual system, battery capacity is effectively estimated based on non-subsampled contourlet transform and Laplacian eigenmap manifold learning method.

Two studies in this special issue are related to electric and electronic engineering. D. Song et al. reported a general fault detection and diagnosis scheme based on observers and residual error analysis for superheterodyne receivers. The proposed method may provide a promising approach not only to the superheterodyne receiver but also to more complex signal receiving systems in which the transfer functions are difficult to obtain. Moreover, P. Orlov et al. presented a unified description of a new approach for contactless detection, identification, and diagnostics of electrical connections and described an idea and principles of using modal probing for these tasks.

Another two studies in this special issue investigated techniques in statistics. H. Sun reported an investigation of stretched exponential function in quantifying long-term memory of extreme events based on artificial data following Lévy stable distribution. It turns out that the stretched exponential distribution provides a reliable way to estimate the scaling behavior of extreme event intervals. In addition, considering the statistic numerical characteristics are often required in the probability-based damage identification and safety assessment of functionally graded material (FGM) structures, Y. Xu et al. presented a stochastic model updating-based inverse computational method to identify the second-order statistics (means and variances) of material properties as well as distribution of constituents for damaged FGM structures with material uncertainties.

This special issue also gathers several interdisciplinary studies from other fields. For example, an effective approach is introduced by H. Wei et al. to predict the magnitude of reservoir-triggered earthquake (RTE), based on support vector machines (SVM) and fuzzy support vector machines (FSVM) methods. Both the SVM and FSVM models are found to be effective in the prediction of the magnitude of RTE with high accuracy. Z. Wang et al. reviewed conventional methods for separating the suspended solids from lubricating oils and presented ultrasonic separation methods for particle separation, which is environment-friendly and has a high efficiency. N. F. Alkayem et al. presented a hybrid elitist-guided search combining a multiobjective particle swarm optimization, Lévy flights, and the technique for the order of preference by similarity to ideal solution investigated. The proposed method shows good performance even under noisy conditions and in the case of incomplete mode shapes. C. Liu

reported a damage detection method for refractory based on principle component analysis and Gaussian mixture model. Two types of damage were utilized to verify the effectiveness of the proposed method. Finally, C. Li proposed an algorithm which is built to obtain the rigorous solution approaching upper and lower bound values simultaneously, satisfying the static boundary and the kinematical boundary based on the slip line field, while stress discontinuity line and velocity discontinuity line are key points.

All in all, this special issue aims to aggregate the latest research efforts contributing to theoretical, methodological, and technological advances in detecting anomalies, forecasting potential degradation, and classifying faults from complex environments and signals. These methods are expected to address the existing challenges for a real-world PHM system.

Conflicts of Interest

The editors declare that they have no conflicts of interest regarding the publication of this special issue.

Minvydas Ragulskis
Chen Lu
Maosen Cao
Gangbing Song
Rafal Burdzik

Research Article

Fault Diagnosis of Rolling Bearing Based on a Novel Adaptive High-Order Local Projection Denoising Method

Rui Yuan ^{1,2}, Yong Lv ^{1,2} and Gangbing Song ³

¹Key Laboratory of Metallurgical Equipment and Control Technology, Wuhan University of Science and Technology, Ministry of Education, Wuhan 430081, China

²Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

³Smart Material and Structure Laboratory, Department of Mechanical Engineering, University of Houston, Houston, TX 77204, USA

Correspondence should be addressed to Yong Lv; lv Yong@wust.edu.cn and Gangbing Song; gsong@uh.edu

Received 20 January 2018; Revised 3 August 2018; Accepted 6 September 2018; Published 11 October 2018

Academic Editor: Marcin Mrugalski

Copyright © 2018 Rui Yuan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rolling bearings are vital components in rotary machinery, and their operating condition affects the entire mechanical systems. As one of the most important denoising methods for nonlinear systems, local projection (LP) denoising method can be used to reduce noise effectively. Afterwards, high-order polynomials are utilized to estimate the centroid of the neighborhood to better preserve complete geometry of attractors; thus, high-order local projection (HLP) can improve noise reduction performance. This paper proposed an adaptive high-order local projection (AHLP) denoising method in the field of fault diagnosis of rolling bearings to deal with different kinds of vibration signals of faulty rolling bearings. Optimal orders can be selected corresponding to vibration signals of outer ring fault (ORF) and inner ring fault (IRF) rolling bearings, because they have different nonlinear geometric structures. The vibration signal model of faulty rolling bearing is adopted in numerical simulations, and the characteristic frequencies of simulated signals can be well extracted by the proposed method. Furthermore, two kinds of experimental data have been processed in application researches, and fault frequencies of ORF and IRF rolling bearings can be both clearly extracted by the proposed method. The theoretical derivation, numerical simulations, and application research can indicate that the proposed novel approach is promising in the field of fault diagnosis of rolling bearing.

1. Introduction

Rolling bearings are of vital importance in rotary machinery systems, and they are prone to failures due to the complex running conditions [1]. The performance of rolling bearings can affect the reliability of operation of the entire system directly. Therefore, it is of great significance to conduct fault diagnosis of rolling bearing. At present, structural health monitoring of rolling bearing usually relies on analyzing its vibration signal [2–4]. During the operation of mechanical systems, the machine is usually accompanied by strong nonlinear and nonstationary vibrations, and the rolling bearing failure will further cause the unexpected vibration of other components, so the measured vibration signals are generally mixed with or submerged in vibration signal of other parts and background noise [5, 6]. The denoising of vibration

signal of faulty rolling bearing contributes to extracting the fault frequency of the faulty rolling bearing, to achieve the fault identification and diagnosis of rolling bearing, while the vibration signals collected from mechanical systems usually have nonlinear and nonstationary characteristics, which make denoising methods based on linear systems unsuitable [7, 8]. The rolling bearing vibration signals collected from operating mechanical systems are usually taken as chaotic signals [9, 10]. For nonlinear time series, useful information would be mistakenly filtered out if traditional linear denoising methods are conducted, which will subsequently result in distortion and deformation of the original signal [11]. There are some methods developed for nonlinear and nonstationary signal analysis, such as wavelet packet decomposition (WPD) [12–14], singular value decomposition (SVD) [15, 16], and ensemble empirical mode decomposition

(EEMD) [17]. While they rely on characteristic frequency components of the collected signal to some extent, however, it is difficult to extract the fault frequencies if the rolling bearing has incipient fault. On that occasion, the fault frequencies of rolling bearing would be submerged in all disturbing frequencies.

The denoising method based on phase space reconstruction has gradually become one of the most important tools to research nonlinear and nonstationary signals. A time series can be reconstructed into high-dimensional phase space through delay embedding, and the reconstructed phase space is diffeomorphic to original dynamic system, i.e., they have the same dynamic characteristics [18–21]. The dynamic characteristics contained in one-dimensional nonlinear time series can be revealed and extracted by studying the motion characteristics and distribution of attractor in high-dimensional phase space. The advantage of high-dimensional signal processing has also been verified in the fields biomedical science and communication engineering [22, 23]. The noise can be filtered out by adopting different projection ways of high-dimensional phase space. For instance, smooth orthogonal decomposition (SOD) [24–26] projects the points in a high-dimensional phase space into different subspaces to reduce noise, and characteristic frequencies of the original time series can be extracted. The main idea of the local projection algorithm was originally proposed to conduct noise reduction [27, 28]. It projects the points in the neighborhood onto the attractor, correcting the raw data by projecting into a few principle directions, which are preferred by locally smoothed data. The local projective filter method was improved to correct the location of points in phase space by determining locations of neighboring points [29]. The noise is reduced by requiring locally linear relations among the delay coordinates and moving delay vectors closer to smooth manifolds. Then, the local projective filters had been employed in spheres ranging from biology signal analysis [30] to space signal analysis [31]; the general framework of local projection was adopted to analyze signals from particular systems, such as being applied to remove interferential artifacts from an electroencephalography (EEG) signal among them. A local projection stabilized and characteristic decoupled scheme aiming at the fluid-fluid interaction problems was proposed. It can reflect the extensive applications of the local projective method [32]. More recently, the parameters of the local projection filtering method were optimized and applied to noisily observed variations in light intensity between the star and its observer [33]. Subspace was proposed to eliminate noise subsequently, and the sequences of temporally successive points in phase space are projected to the relative position of neighboring strands [24–26]. Among SOD, noises in the data are reduced by projecting to the tangent subspace. The vector directions of the projection can make sure the variance as maximum as possible, and the motions obtaining along these vector directions are the smoothest in terms of time.

The centroid selection of local neighborhoods is of vital importance in the local projection method to estimate the neighborhood correctly. The first-order statistics and the

TABLE 1: The algorithm of standard LP denoising method.

-
- (1) Select appropriate embedding dimension m and delay time τ to reconstruct noisy time series into m dimensional phase space.
 - (2) Determine neighborhood U_n for one phase point.
 - (3) Calculate centroid of its neighborhood by fixed neighborhood numbers.
 - (4) Calculate covariance matrix \mathbf{C} of its phase point neighborhood and then conduct eigenvalue decomposition to obtain eigenvector \mathbf{a}_q corresponding to $m - m_0$ smaller eigenvalues. ($q = 1, 2, \dots, m - m_0$).
 - (5) Subtract the projection of the phase point in noise subspace by

$$\mathbf{X}_n' = \mathbf{X}_n - \mathbf{R}^{-1} \sum_{q=1}^{m-m_0} \mathbf{a}_q [\mathbf{a}_q \cdot \mathbf{R}(\mathbf{X}_n - \bar{\mathbf{X}}_n)],$$
 where \mathbf{R} is a diagonal weight matrix, aiming at inhibiting the distortion of ending elements of phase points and reserving the middle stable elements (\mathbf{R}_{11} and \mathbf{R}_{mm} were set as 10^3 ; \mathbf{R}_{ii} values were set as 1 in our research). $\bar{\mathbf{X}}_n$ is the centroid of neighborhood; \mathbf{X}_n' denotes the denoised signal.
 - (6) Get back to step (2) and repeat subsequent steps until all phase points are processed.
-

second-order statistics have been widely used in the analysis of mechanical vibration signals, and these methods are theoretically applicable to the analysis of linear and Gaussian signal [29, 34, 35]. While for nonlinear and nonstationary signals, higher orders can be utilized to better reflect the signal characteristics, and high-order polynomials have been used to estimate the centroid of neighborhood among LP denoising method in the field of medical signal processing [36] and thermodynamics [37]. Compared with LP denoising method employing the first- and the second-order statistics, HLP denoising method employing higher-order statistics has the advantages in reducing noise and extracting characteristic frequencies, by adapting to phase space nonlinearity. The extension to HLP denoising method can reduce noise more effectively than previous LP filtering [34–37]. In this paper, by employing high-order polynomials to calculate the centroid of neighborhood during the LP procedures, the HLP was firstly adopted to denoise the vibration signals of faulty rolling bearing to extract its fault characteristic frequencies.

This paper proposes a novel approach called AHLP denoising method aiming at fault diagnosis of rolling bearings. In the proposed AHLP denoising method, except adopting high-order polynomials to calculate the centroid of neighborhood, optimal orders of dealing with different kinds of rolling bearing faults can be further estimated to achieve a better denoising effect. The AHLP denoising method can better denoise the vibration signal of faulty rolling bearing to extract its fault characteristic frequencies among all disturbing frequencies, which contributes to the field of fault diagnosis of rolling bearing. The organization of this paper is as follows: Section 2 introduces the methodologies of standard LP, HLP, and AHLP denoising methods and illustrates the scheme of AHLP denoising method in

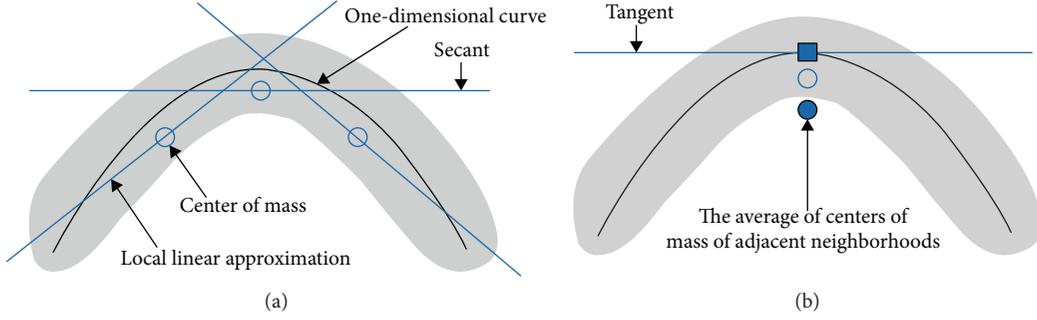


FIGURE 1: Consider a cloud of points around a curve. (a) The mean of phase points of neighborhood is estimated as centroid of neighborhood. (b) The second-order polynomial is used to estimate the centroid of neighborhood.

TABLE 2: The a_i value of different orders in HLP denoising method.

								Order	a_i		
1								1	[1]		
1	1							2	[2, -1]		
1	2	1						3	[3, -3, 1]		
1	3	3	1					4	[4, -6, 4, -1]		
1	4	6	4	1				5	[5, -10, 10, -5, 1]		
1	5	10	10	5	1			6	[6, -15, 20, -15, 6, -1]		
1	6	15	20	15	6	1			7	[7, -21, 35, -35, 21, -7, 1]	
1	7	21	35	35	21	7	1			8	[8, -28, 56, -70, 56, -28, -1]
1	8	28	56	70	56	27	8	1			

fault diagnosis of rolling bearing. Section 3 introduces the vibration signal model of faulty rolling bearings and presents numerical simulations including simulated vibration signals of ORF and IRF rolling bearings. By choosing different orders to denoise simulated signals and extract characteristic frequencies, the optimal orders are obtained. Section 4 presents the applications to bearing monitoring with two different practical experimental signals to validate the proposed method. The conclusions of the researches and necessary discussions are given in Section 5.

2. Methodology

2.1. Standard Local Projection Denoising Method. Assuming that $x(1), x(2), \dots, x(N) \in \mathbb{R}$ is a time series of length N extracted from a chaotic system, then it can be embedded in phase space with appropriate embedding dimension and delay time based on the embedding theorem [18–20]. The reconstructed phase space has the same diffeomorphism as the original time series, namely, they have the same dynamic characteristics. The reconstructed phase space can be expressed as follows:

$$\mathbf{X}_n = \left(x_{n-(m-1)\tau}, x_{n-(m-2)\tau}, \dots, x_n \right), \quad (1)$$

where \mathbf{X}_n denotes the n th phase point in reconstructed phase space, m is the embedding dimension, and τ is the delay time. The minimum embedding dimension m needs to meet the

criterion $m \geq 2d + 1$, when fractional dimension of system attractor is d . Under this circumstance, the reconstructed phase space is diffeomorphic to the original dynamic system, namely, they have the same dynamic characteristics. In m dimensional phase space, the attractor that reflects dynamic behavior of system is usually confined to a low-dimensional subspace of m_0 ($d < m_0 < m$). When there is no noise interference, zero subspace of $m - m_0$ exists. When noise is present, it is randomly distributed in phase space, and the components from subspace of $m - m_0$ are generated by noise. For each phase point, eigenvalue decomposition is conducted to deal with covariance matrix composed by all phase points within neighborhood. The subspaces corresponding to m_0 larger eigenvalues are taken as signal subspace, while other $m - m_0$ smaller eigenvalues correspond to the noise subspace. The LP denoising method is aimed at finding the abovementioned noise subspace and then subtract the projection of phase points in this subspace. The algorithm of the LP method is given in Table 1.

During step (1), embedding dimension m and delay time τ are critical parameters. The bigger embedding dimension m contributes to overembedding to fully extract the characteristics of the time series, but meanwhile, it decreases the computational efficiency. The smaller time delay τ can make neighborhood points closer to the attractor. Hence, in this paper, the mutual information method [38] is adopted to determine delay time τ , and then the proposed method by CAO [39] is used to determine embedding dimension m . As for step (3) of standard LP denoising method, the method

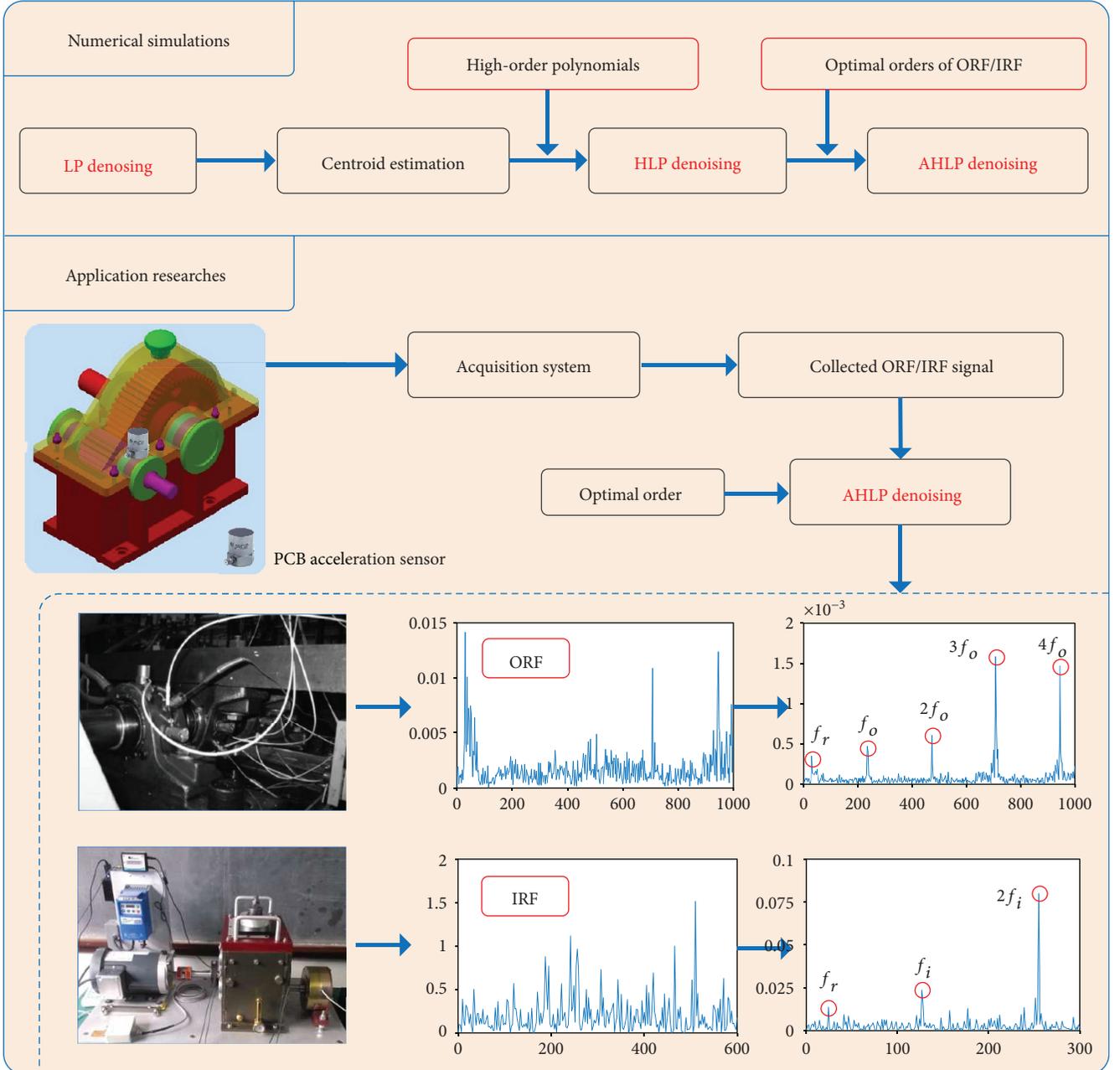


FIGURE 2: The scheme of AHLP denoising method in fault diagnosis of rolling bearing.

was proposed by using mean of phase points of neighborhood as centroid of neighborhood [34, 35].

$$\bar{\mathbf{X}}_n = \frac{1}{U_n} \sum_{k \in U_n} \mathbf{X}_k. \quad (2)$$

This method may generate certain errors because of local linearization; the centroids of each neighborhood are secants instead of tangents and all centroids (blue circles) are shifted inward regarding the curvature, as shown in Figure 1(a). Afterwards, a method of estimating the centroid by using

the second-order polynomial was developed to further suppress noise [29].

$$\bar{\mathbf{X}}'_n = 2\bar{\mathbf{X}}_n - \frac{1}{U_n} \sum_{k \in U_n} \bar{\mathbf{X}}_k, \quad (3)$$

where a tangent approximation is obtained by shifting the centroid of neighborhood outward regarding the curvature. The illustration is shown in Figure 1(b), where the filled circle denotes the centroid of phase points of adjacent neighborhoods, and the square is the modified centroid.

TABLE 3: The parameters of the vibration signal model of ORF rolling bearing.

A_0	f_m	f_o	τ_i	f_n	C_A	φ_A	φ_w	B
3	0	100	0.01	2000	1	0	0	800

Essentially, the method is performed by moving the centroid outward to make it approximately tangent to the hyperplane of projection. The above two methods are referred here as standard LP denoising method.

2.2. High-Order Local Projection Method. The method was proposed by applying high-order polynomials to estimate the centroid of the neighborhood more accurately. Significant achievements have been obtained by high-order algorithms in medical signal processing [36] and thermodynamics [37], because it can reflect mathematical characteristics more accurately. In this paper, this method is referred as the HLP denoising method. The basic principle is explained below.

As for any real number $\delta > 0$, define a continuous moving average operator as I_δ .

$$(I_\delta f)(x) = \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} f(t) dt, \quad (4)$$

where $f(t)$ is a time series, and δ is shift step length.

The effect of the LP denoising method is enhanced by considering using the linear combination of the first and the second orders to estimate the centroid of neighborhood, thus improving the denoising effect. Hence, it can be inferred that

$$f_n(0) = a_1(I_\delta f_n)(0) + a_2(I_\delta^2 f_n)(0), \quad (5)$$

where $f_n = (0)$ is applicable to all monomials, when $n = 0$, $f_n(x) = 1$, and when $n \neq 0$, $f_n(x) = x^n$, $0 \leq n \leq n_{\max}$, while n_{\max} is supposed to be as large as possible. As for any odd number, both sides of (5) are 0. Setting $n = 0$ and $n = 2$ results in a system combined with two linear equations, which have unique solution $a_1 = 2$ and $a_2 = -1$. In the m dimensional Euclidean space, the centroid of all phase points within a ball of radius δ is equivalent of moving average operator I_δ . In this way, the second-order LP denoising method was proposed. If (5) is applicable to all monomials, $f_n(x) = x^n$, $0 \leq n \leq 3$, which means

$$p_3(x) = a_1(I_\delta p_3)(x) + a_2(I_\delta^2 p_3)(x), \quad (6)$$

where $p_3(x)$ holds accurately for the third-order Taylor polynomial of any thrice differentiable function. As for $p_3(x)$ in (6), which is analogue of $f_n(0)$ in (5), it holds exactly for multivariate Taylor polynomials of higher order up to three in each variable.

Then for the HLP denoising method, the linear combination of I_δ^i needs to be identified to represent higher-order

Taylor polynomial more accurately. Through induction and deduction, it can be obtained that as for each $j, n \in \mathbb{N}$,

$$(I_\delta^j f_n)(x) = \frac{1}{\delta^{n+j}} \left(-\frac{1}{2}\right)^j \frac{n!}{(n+j)!} \times \sum_{r=0}^j \binom{j}{r} (-1)^r (x + (2r-j)\delta)^{n+j}. \quad (7)$$

This formula is used to prove that as for $k = 1, 2, \dots, 20$, the system constituted by k linear equations has the unique solution. The system and the solution are as follows:

$$f_n(0) = \sum_{i=1}^k a_i (I_\delta^i f_n)(0), \quad n = 0, 2, \dots, 2(k-1), \quad (8)$$

$$a_i = (-1)^{i-1} \binom{k}{i}, \quad i = 1, 2, \dots, k. \quad (9)$$

Obviously, all coefficients determined by (9) include two situations of estimating centroid of neighborhood when order is 1 and 2, namely, $i = 1$ [34] and $i = 2$ [29]. The equation also expands the estimation of centroid of neighborhood to higher orders ($i \geq 3$) such as $i = 1, 2, \dots, k$. The difference between algorithms of standard LP and HLP denoising methods lies in the 3rd step of all procedures as shown in Table 1. The 3rd step of the standard LP denoising method uses the mean and second-order polynomial to calculate the centroid of the local neighborhood. As for HLP denoising method, higher-order polynomials are utilized to estimate the centroid.

In this paper, the orders up to 8 are analyzed. In mathematics, binomial coefficients refer to a set of positive integers appear in binomial theorem as coefficients. They are indexed by two nonnegative integers, namely, k and i , which are the coefficients of x^i term in the polynomial expansion of binomial power $(1+x)^k$. Under appropriate situations, the value of the coefficient $\binom{k}{i}$ is set by $(k!)/i!(k-i)!$. Arranging binomial coefficients into rows for i changes from 1 to k . Thus, during the algorithm, the solutions a_i of different orders can be derived and calculated as given in Table 2. The left side shows the coefficients of deployed high polynomials.

2.3. Adaptive High-Order Local Projection Denoising Method. As mentioned above, as a denoising reduction method, LP method has been verified to be effective in reducing noises existing in vibration signals of rolling bearings after theoretical derivation and extensive tests. Estimation of the centroid of the neighborhood by using high-order polynomials is beneficial to achieve better denoising effect; hence, HLP denoising method is adopted in this paper to deal with vibration signals of faulty rolling bearings. Owing to the reason that vibration signals of ORF and IRF rolling bearings have different nonlinear geometric structures, denoising effects of different orders differ, so the optimal orders of different kinds

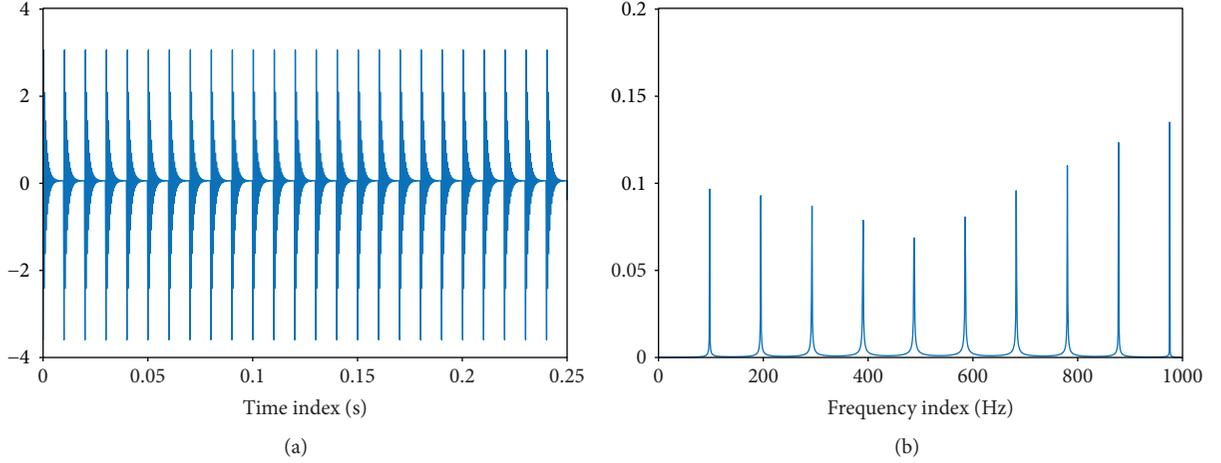


FIGURE 3: Time and frequency domain plots of simulated vibration signal of ORF rolling bearing.

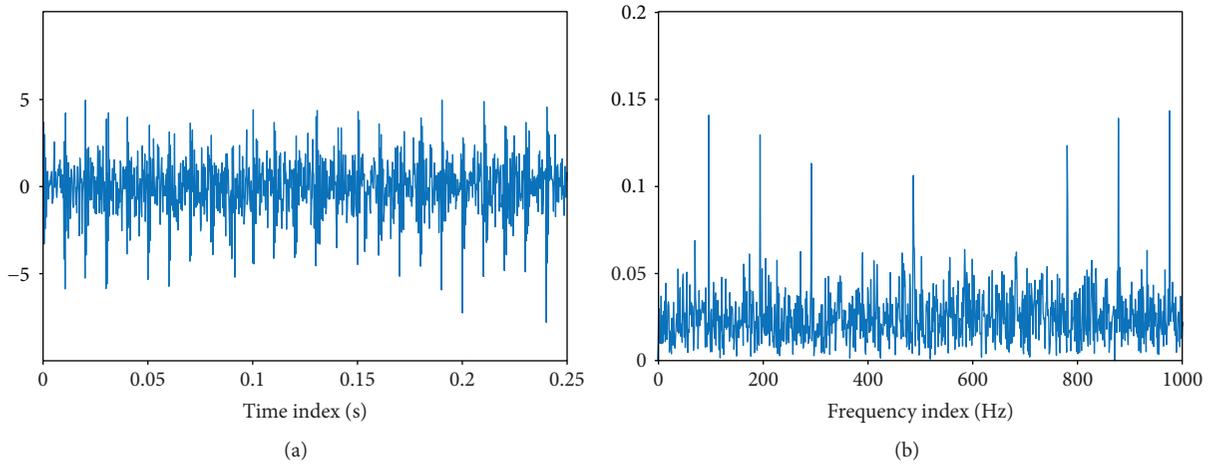


FIGURE 4: Time and frequency domain plots of simulated noisy signal of ORF rolling bearing.

TABLE 4: The SNRs of denoised signal with different orders.

Order	1	2	3	4	5	6	7	8
SNR	8.92	9.73	10.21	11.62	8.50	5.61	5.09	4.43

of faults are different. By choosing the optimal orders before conducting denoising process, a novel AHLP denoising is hereby proposed in this paper, and the scheme of the proposed method is illustrated in Figure 2.

3. Numerical Simulations

3.1. Vibration Signal Model of Faulty Rolling Bearing. The vibration signal model of faulty rolling bearing used hereinafter was proposed by Randall et al. [40–42], and it takes rolling bearing's construction (geometry), tolerance, amplitude modulation, ball sliding, surface wear, and other factors into consideration. This bearing model has been successfully

applied in related researches [42–44]. The main faults of rolling bearing include local pitting on the inner ring, outer ring, and rolling element, which are caused by shock effects on the rolling bearing and other parts during the operation of the entire system. The frequency of the periodic impact reflects the type of rolling bearing faults. A local pitting often encourages the system to produce vibration of the inherent frequency of the rolling bearing or the system. During the operation of the rolling bearing, the absolute position of pitting may change periodically versus sensor position. For instance, when a local pitting occurs on the inner ring, the absolute position of the local pitting would change with the rotation speed periodically, and when it occurs on the rolling element, the absolute position would change with the rotation period of bearing retainer periodically. Since the rotation period of the absolute position of a local pitting is greater than the period of shock, the shock is modulated distinctively. Compared to the derivative of resonant frequency, the shock time is rather short and usually decays rapidly. Under a steady-state condition, it can be considered that modulation amplitude does not change during the process

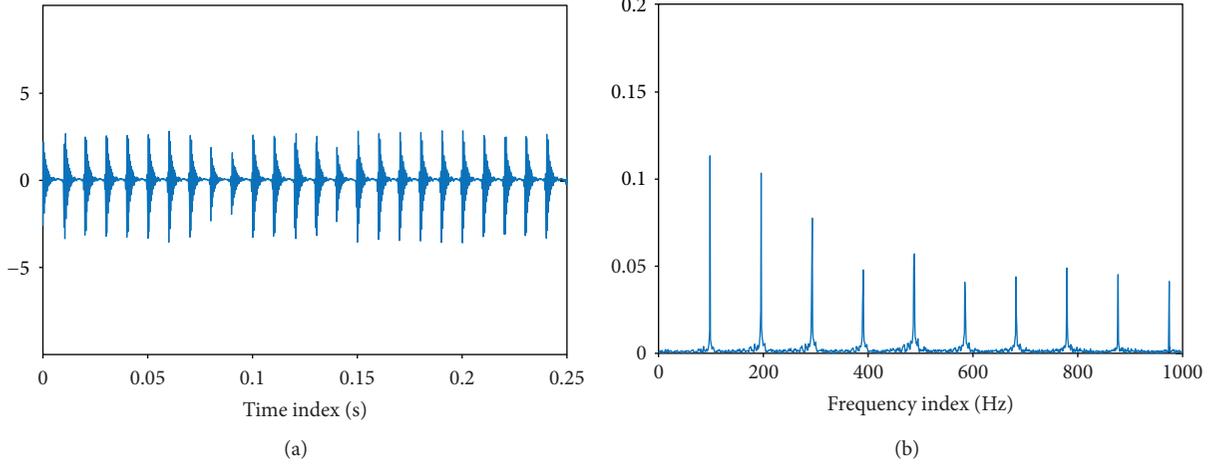


FIGURE 5: Time and frequency domain plots of denoised signal of ORF rolling bearing.

TABLE 5: The SNRs of denoised signal with different orders.

Order	1	2	3	4	5	6	7	8
1	8.92	9.73	10.21	11.62	8.50	5.61	5.09	4.43
2	10.17	12.42	15.65	15.63	12.46	10.30	8.89	7.06
SNR 3	11.65	12.45	14.90	16.85	14.70	13.97	10.45	7.90
4	12.76	14.34	18.25	19.79	16.80	13.46	12.38	10.66
5	12.83	13.78	14.67	17.01	14.38	13.78	11.79	11.51

of shock oscillation, and the oscillation signals caused by each shock are the same.

Define T as the period of shock and $s(t)$ as the shock oscillation caused by a pitting fault. Assume that the amplitude of i th shock is A_i . There are generally strong ambient noises since the working environment of rolling bearing is relatively poor. Thus, the model considers the interference of additive noises $n(t)$, which are assumed to be stationary random noises of zero mean. Hence, the vibration signal model of faulty rolling bearing $m(t)$ is expressed as follows:

$$\begin{aligned}
 m(t) &= \sum_{i=1}^M A_i \cdot s(t - iT - \tau_i) + n(t), \\
 A_i &= A_0 \cdot \cos(2\pi f_m t + \varphi_A) + C_A, \\
 s(t) &= \exp(-Bt) \cdot \cos(2\pi f_n t + \varphi_w),
 \end{aligned} \tag{10}$$

where A_0 is amplitude modulation factor function, f_m is modulation frequency, $s(t)$ is shock oscillation, τ_i is tiny slip-page between rolling element and raceway, f_n is system resonant frequency of rolling bearing, φ_A , φ_w and C_A are arbitrary normal numbers, and B is the intensity of resonance. The significant distinction between vibration signals of simulated ORF and IRF rolling bearings is when it comes to ORF rolling bearing that f_m equals to 0 Hz, while when it comes to IRF rolling bearing that f_m equals to f_r (f_r denotes rotation frequency).

3.2. Numerical Simulation of ORF Rolling Bearing. Based on the vibration signal model of faulty rolling bearing in Section

TABLE 6: The parameters of the vibration signal model of IRF rolling bearing.

A_0	f_m	f_i	τ_i	f_n	C_A	φ_A	φ_w	B
3	20	100	0.01	2000	1	0	0	800

3.1, as for the ORF rolling bearing, f_m was set as 0 Hz in our research; f_o denotes the fault frequency of ORF rolling bearing. The parameters of the simulated vibration signal of ORF rolling bearing are set as shown in Table 3 in this paper. The sampling point is $N = 4096$ and the sampling frequency is $f_s = 4096$ Hz. To verify the effectiveness of HLP denoising method in real situations, where the collected signal normally has noise, therefore, Gaussian white noise is added in numerical simulations. The power spectral density of Gaussian white noise is uniformly distributed, and its amplitudes obey the Gaussian distribution, making it is suitable for simulation analysis. Here, Gaussian white noise is added to the original signal, and the signal noise ratio (SNR) is 1. The time and frequency domain plots of the simulated vibration signal of ORF rolling bearing are shown in Figure 3. The time and frequency domain plots of the simulated noisy signal of ORF rolling bearing are shown in Figure 4. The time domain of the first 0.25 s of the signal is analyzed specifically to show the effectiveness of the HLP denoising method and clarify different characteristics between vibration signals of ORF and IRF rolling bearings.

It can be observed from the time and frequency domain plots in Figure 3 that the vibration signal of ORF rolling bearing has periodic characteristic. The frequency domain plot shows fault frequency $f_o = 100$ Hz and its harmonic frequencies containing second, third, and fourth harmonic frequency. By conducting HLP denoising method towards simulated noisy ORF signal with different orders, the denoised signal with different SNRs can be obtained as shown in Table 4.

It can be seen from Table 4 that when the original SNR is 1, the SNRs of the denoised signal with different orders increase along with the order until 4 and then decrease all

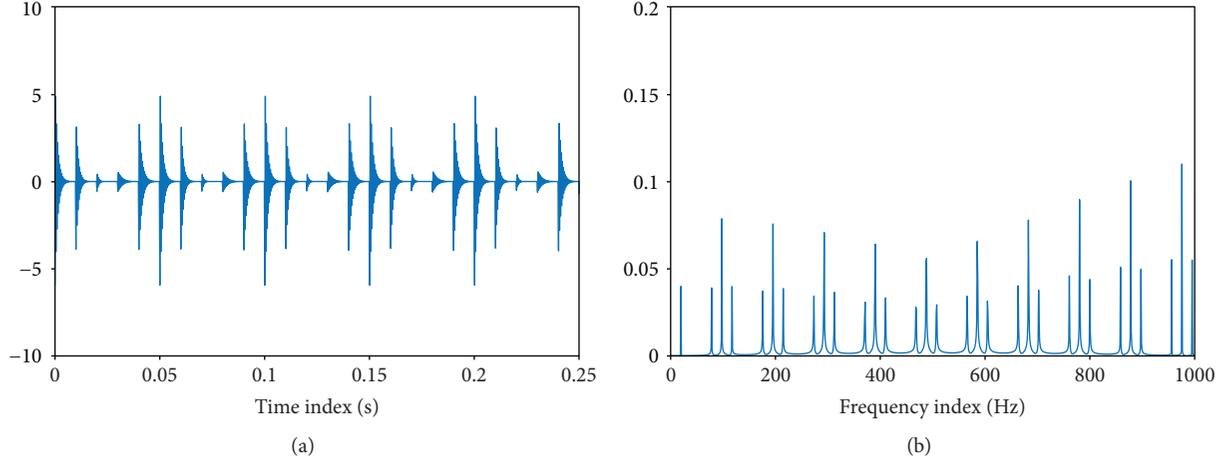


FIGURE 6: Time and frequency domain plots of simulated vibration signal of IRF rolling bearing.

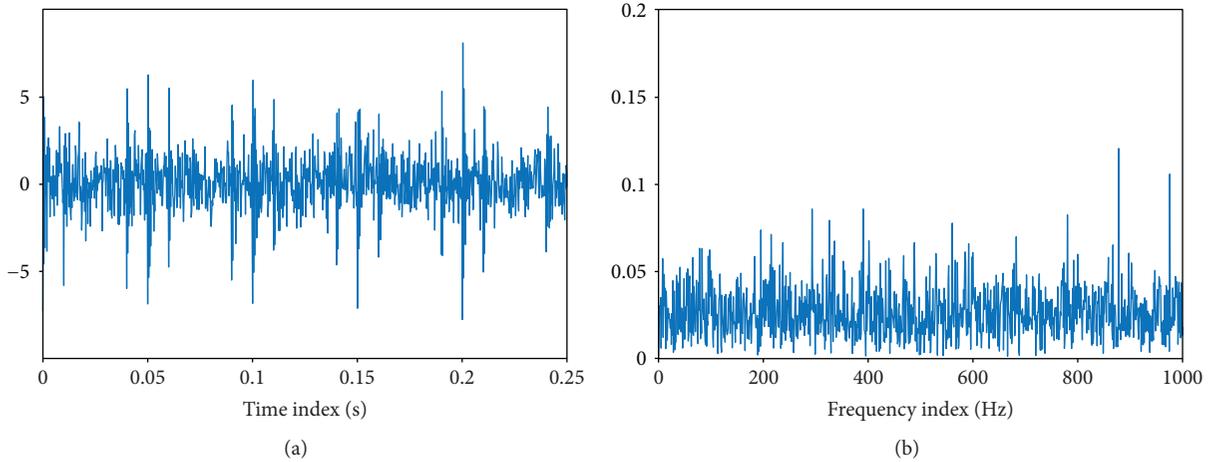


FIGURE 7: Time and frequency domain plots of simulated noisy signal of IRF rolling bearing.

TABLE 7: The SNRs of denoised signal with different orders.

Order	1	2	3	4	5	6	7	8
SNR	6.91	7.28	7.73	7.62	7.20	6.72	5.07	4.52

the way. The time and frequency domain plots of the denoised vibration signal of ORF rolling bearing, when the order of HLP denoising method is 4, are shown in Figure 5.

As shown in Figure 5, the characteristic frequencies f_o and its resonant frequencies can be clearly extracted under the influence of noise. The characteristic frequency amplitudes are obvious. The result verifies the effectiveness of HLP denoising method to extract the fault frequencies of vibration signal of ORF rolling bearing. To obtain the optimal orders for HLP denoising method aiming at vibration signal of ORF rolling bearing, Gaussian white noise with different SNRs, SNR = 1, 2, 3, 4, 5, respectively, is added to the original simulated ORF signal. Then, different SNRs of different denoised results are obtained, as shown in Table 5.

When the SNR is bigger than 5, which implies there are not too much additive noises in the signal, the HLP can

achieve proper denoising effect in most cases. Hence, the above simulations are conducted to find optimal orders aiming at vibration signal of ORF rolling bearing. As shown in Table 5, the primary optimal order of HLP denoising method is 4. In one case, the optimal order is 3, and SNR is only a little bit higher than the situation when the order is 4. Because of some uncertain parameters in vibration signal model of faulty rolling bearing, the slight fluctuation of optimal order can be accepted. Therefore, it can be concluded that the optimal order for vibration signal of ORF rolling bearing is 4.

3.3. Numerical Simulation of IRF Rolling Bearing. Based on the vibration signal model of faulty rolling bearing in Section 3.1, as for the IRF rolling bearing, f_m and f_r were set as 20 Hz in our research; f_i denotes the fault frequency of IRF rolling bearing. The parameters of the simulated vibration signal of IRF rolling bearing are set as shown in Table 6 in this paper. The sampling point is $N = 2048$ and the sampling frequency is $f_s = 4096$ Hz. Similar to that in Section 4.1, Gaussian white noise is added to the original signal, and the SNR is 1. The time and frequency domain plots of the simulated vibration signal of IRF rolling bearing are shown in Figure 6. The time

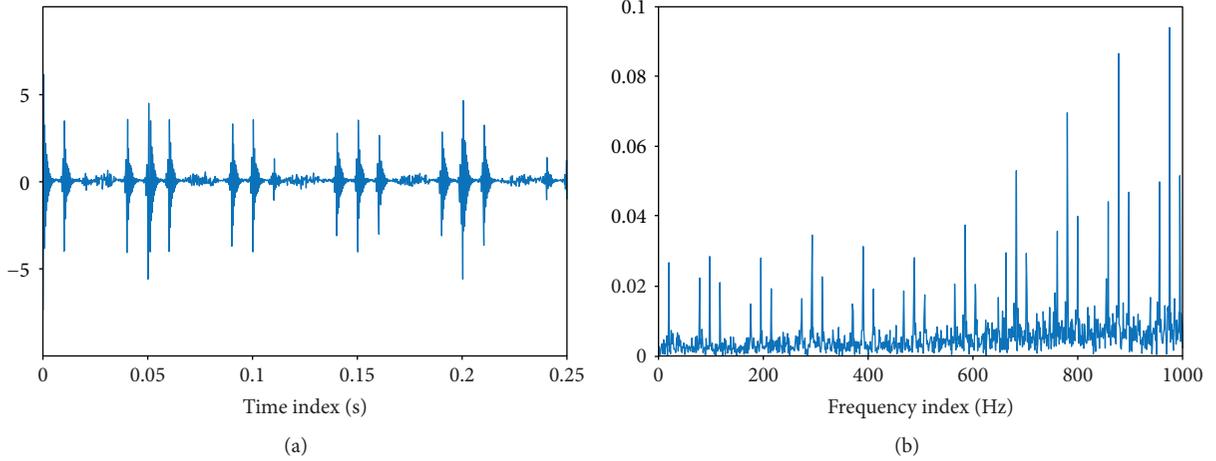


FIGURE 8: Time and frequency domain plots of the denoised signal of IRF rolling bearing.

TABLE 8: The SNRs of denoised signal with different orders.

Order	1	2	3	4	5	6	7	8
1	6.91	7.28	7.73	7.62	7.20	6.72	5.07	4.52
2	7.68	8.04	8.54	8.30	6.62	6.61	5.93	5.12
SNR 3	8.14	8.25	8.66	8.41	8.12	6.55	6.53	5.89
4	8.73	8.81	9.03	8.70	8.44	7.50	7.17	6.81
5	8.82	9.25	9.32	8.59	8.15	8.06	7.73	7.01

and frequency domain plots of the simulated IRF noisy signal are shown in Figure 7. The time domain of the first 0.25 s of the signal is analyzed specifically to show the effectiveness of HLP denoising method and clarify different characteristics between vibration signals of ORF and IRF rolling bearings.

It can be observed from the time and frequency domain plots in Figure 6 that the vibration signal of IRF rolling bearing has periodic characteristic. The frequency domain plot shows fault frequency $f_i = 100$ Hz and its harmonic frequencies containing the second, third, and fourth harmonic frequency. In addition to characteristic frequencies, there are sideband frequencies with modulation frequency $f_m = f_r = 20$ Hz and the resonant frequencies are $f_i \pm f_r, 2f_i \pm f_r, 3f_i \pm f_r, \dots$ and $f_r = 20$ Hz in the low-frequency range. Apply the HLP denoising method towards the noisy IRF signal with different orders that result in the denoised signals with different SNRs, as shown in Table 7.

It is clear from Table 7 that, when the original SNR is 1, the SNRs of the denoised signal with different orders increase with the order until 3 and then decrease all the way. The time and frequency domain plots of the denoised ORF noisy signal, when order of HLP denoising method is 3, are shown in Figure 8.

As shown in Figure 8, the characteristic frequencies f_o and its resonant frequencies $f_i \pm f_r, 2f_i \pm f_r, 3f_i \pm f_r, \dots$ and $f_r = 20$ Hz can be clearly extracted under the influence of noise. The characteristic frequency amplitudes are obvious. The result demonstrates the effectiveness of HLP denoising method in extracting fault frequencies of the vibration signal of IRF rolling bearing. To obtain the optimal orders

for HLP denoising method aiming at vibration signal of IRF rolling bearing, Gaussian white noise with different SNRs, SNR = 1, 2, 3, 4, 5, respectively, is added into original simulated IRF signal. Then, different SNRs of different denoised results are obtained, as shown in Table 8.

As shown in Table 8, the optimal order of the HLP denoising method is 3. Different from the situation when HLP denoising method is applied to ORF signal, the optimal order is a constant. Therefore, it can be obtained that the optimal order for HLP denoising method aiming at the vibration signal of IRF rolling bearing is 3.

4. Applications to Fault Diagnosis of ORF and IRF Rolling Bearings

Through numerical simulations, the proposed novel AHLP denoising method is applied to deal with vibration signals of ORF and IRF rolling bearings successfully, and there are certain optimal orders for different faults of rolling bearing. By choosing optimal order of HLP, the best denoising effect can be obtained. Here, two cases of test data are used to validate the effectiveness of the proposed AHLP denoising method in this paper.

4.1. Application to Processing of Intelligent Maintenance System Bearing Data. To verify the effectiveness of the proposed method in application to fault diagnosis of rolling bearing, the bearing data from Intelligent Maintenance Systems (IMS) Center of University of Cincinnati [45] is used to verify the proposed method. Three datasets are included in the downloaded data packet, and a test to failure experiment was conducted and dataset was collected by NI DAQ Card 6062E during no. 2 experiment. The 772nd data of no. 2 dataset was used to verify the effectiveness of the proposed method in this paper. The schematic diagram of the experimental apparatus and the position of sensors are shown in Figure 9. Four Rexnord ZA-2115 rolling bearings were installed on the shaft, and PCB 353B33 accelerometers were installed vertically on the bearing houses. The rotational speed was 2000 r/min, facilitated by an AC motor, namely,

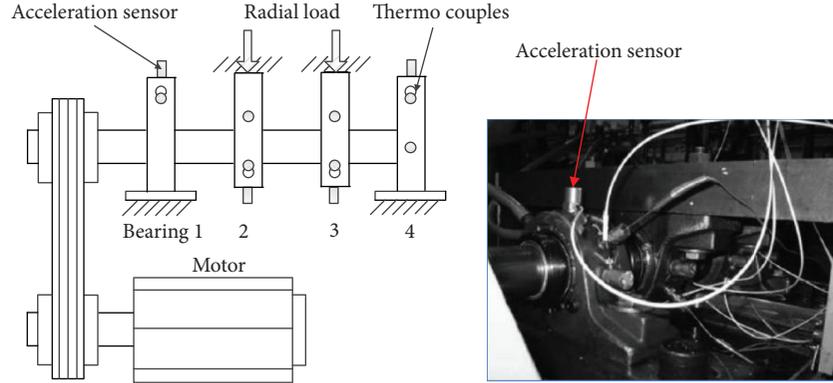


FIGURE 9: The schematic diagram and a picture of the apparatus [45].

TABLE 9: Rolling element bearing parameters.

Rolling element bearing parameters of ZA-2115 (diameter/cm)			
Ball number n	Contact angle α	Ball diameter d_r	Pitch diameter D_w
16	15.17	0.331	2.815

TABLE 10: Bearing characteristic frequencies.

Fault type	Failure frequency
Defect on inner race	$f_i = 0.5n(1 + d_r \cos \alpha/D_w)f_r$
Defect on outer race	$f_o = 0.5n(1 - d_r \cos \alpha/D_w)f_r$

TABLE 11: Characteristic frequencies of Rexnord ZA-2115 rolling bearing.

Fault type	Fault frequency (Hz)
Outer ring fault	$f_o = 236.4$
Inner ring fault	$f_i = 296.9$

the rotational frequency f_r is 33.33 Hz. The sampling frequency is 20 kHz. The specific parameters of fault rolling bearings are shown in Table 9. The calculating methods for different rolling bearing fault frequencies are shown in Table 10. The characteristic frequencies of rolling element bearing were computed as shown in Table 11.

At the end of the no. 2 experiment, ORF happened on the rolling bearing. The time and frequency domain plots of the collected vibration signal of the faulty rolling bearing (772nd data) are shown in Figure 10.

The optimal order is chosen as 4 among the proposed AHLP denoising method. To illustrate the effectiveness of the proposed method, the frequency domain plots of the original signal and signal processed by WPD are also presented. The wavelet function db15 with 11 layers is adopted in the WPD. To observe the fault characteristic frequency clearly and conduct a comparative analysis to verify the effectiveness of the proposed method, the partial enlarged frequency plots of frequency domain of collected signal and WPD denoised signal are shown in Figure 11. The frequency domain plots of signal processed by the LP denoising method and the proposed AHLP denoising method are presented in Figure 12.

It can be seen from Figure 11(a) that in the frequency domain plot of collected signal of ORF rolling bearing, the rotational frequency f_r can be seen along with harmonic frequencies $3f_o$ and $4f_o$, but the fault characteristic frequency itself f_o cannot be found. It indicates that the rolling bearing

had severe fault then and collected signal was interfered by background noises to a great extent, and there are many interfering frequencies. From Figure 11(b) and Figure 12(a), it can be seen after the processing of WPD and LP denoising method that all characteristic frequencies can be extracted but still along with some interfering frequencies. In addition, the effectiveness of LP denoising method is better than WPD. It can be seen from Figure 12(b) that after AHLP denoising method, the background noise is greatly reduced and eliminated, the characteristic frequencies are obvious, and there are not many interfering frequencies in the frequency domain plot. The frequency domain plot shows very clear the rotational frequency f_r , and fault characteristic frequencies f_o , $2f_o$, $3f_o$, $4f_o$, namely, fault characteristic frequencies can be well extracted. The denoising effect of AHLP is better than LP and WPD denoising methods, which indicates that the extension research from LP to AHLP is a successful exploration. Based on the displayed results and comparative analysis, it can be concluded that the vibration signal of ORF rolling bearing with severe fault can be well denoised by the proposed AHLP denoising method in practical application.

4.2. The Collected Signal Processing of Drivetrain Diagnostics Simulator. To further verify the effectiveness of the proposed method in the application of weak fault diagnosis, the experiment is conducted on Drivetrain Diagnostics Simulator. Here, ORF and IRF rolling bearings are used during the experiment, and fault signals are collected to be processed.

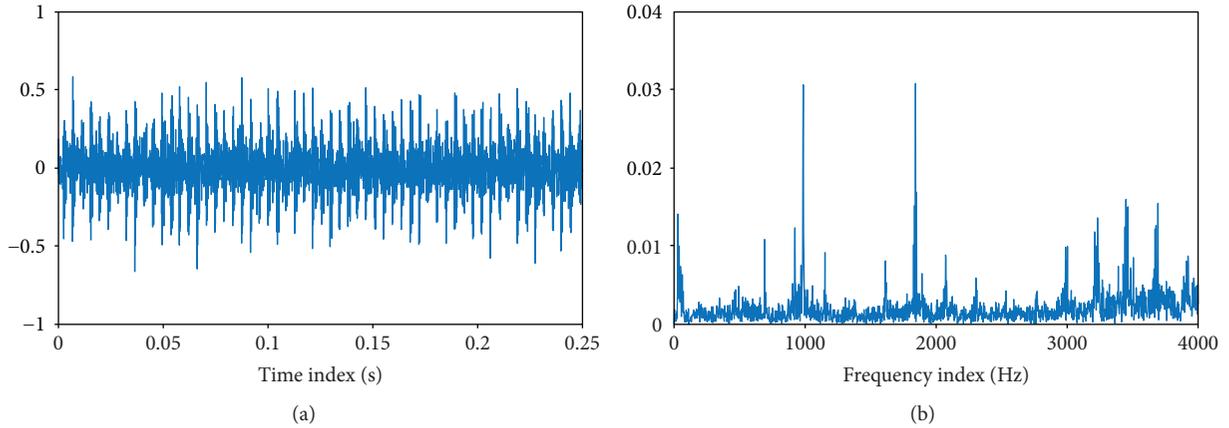


FIGURE 10: Time and frequency domain plots of collected vibration signal of ORF rolling bearing.

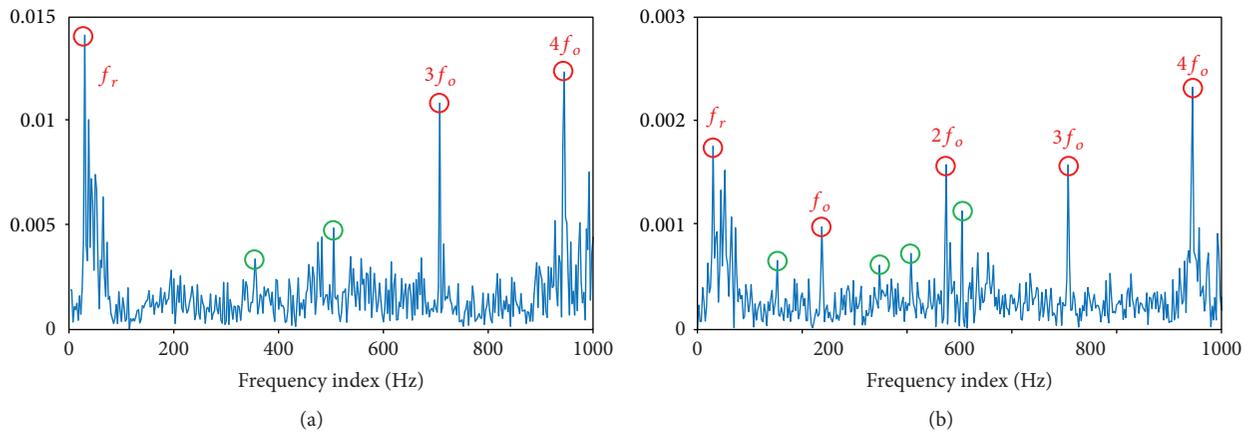


FIGURE 11: (a) Frequency domain plot of collected signal. (b) Frequency domain plot of denoised signal by WPD (red circles denote characteristic frequency, and green circles denote interfering frequency, as same hereinafter).

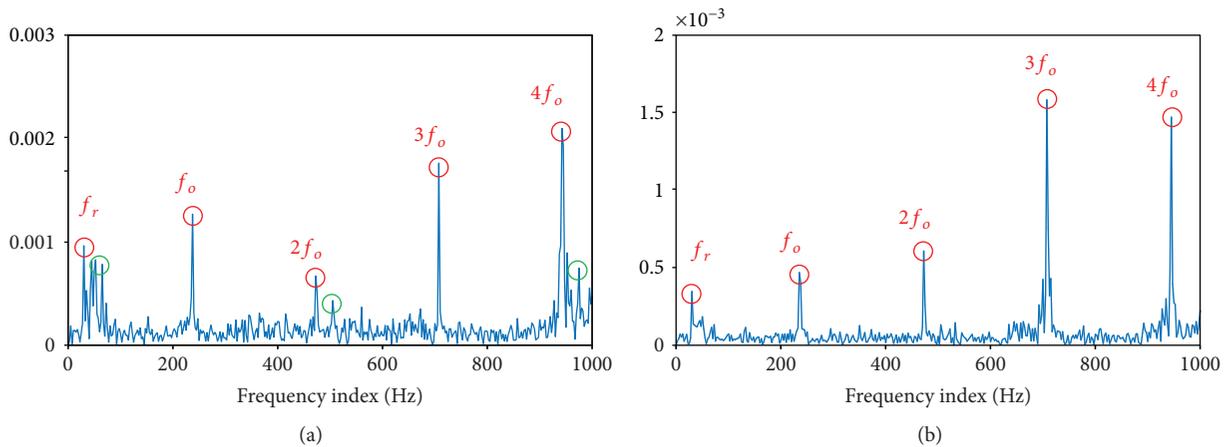


FIGURE 12: (a) Frequency domain plot of denoised signal by LP denoising method. (b) Frequency domain plot of denoised signal by AHLP denoising method.

The experimental apparatus is produced by SQI Company, United States. The experiment is aimed at collecting the signal of ORF and IRF rolling bearings. The sensor is placed on the rolling bearing end plate to collect its acceleration signal.

The experimental apparatus is composed of a variable speed drive, a torque transducer and encoder, a parallel shaft gear-box which includes two parallel shaft rolling bearing, and a programmable magnetic brake. The schematic diagram of

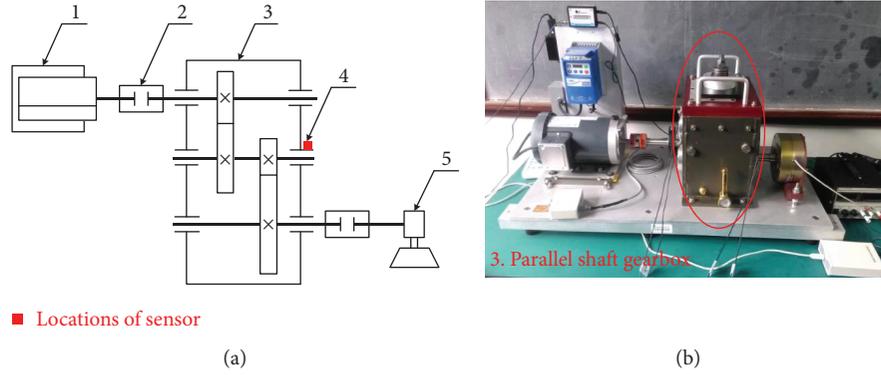


FIGURE 13: The schematic diagram and the picture of the experimental apparatus. 1: variable speed drive, 2: torque transducer and encoder, 3: parallel shaft gearbox, 4: test point, and 5: programmable magnetic brake.

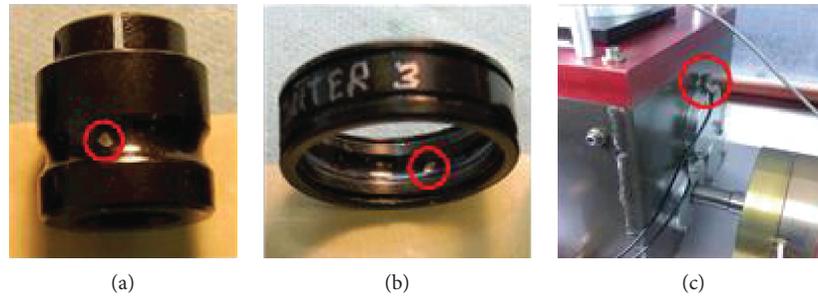


FIGURE 14: (a) IRF of rolling bearing, (b) ORF of rolling bearings, and (c) location of sensor.

TABLE 12: Characteristic frequencies of FAFNIR deep groove rolling bearing.

Fault type	Fault frequency (Hz)
Inner ring fault	$f_i = 130.32$
Outer ring fault	$f_o = 85.73$

the experimental apparatus and its photo are shown in Figure 13. The photos of IRF and ORF rolling bearings and location of sensor are shown in Figure 14.

The experiment was conducted to collect the vibration signal of weak fault rolling bearing. During the experiment, the sampling frequency is 8192 Hz, and the rotational frequency f_r is 25 Hz. The fault characteristic frequencies of experimental FAFNIR deep groove rolling bearing are shown in Table 12.

The time and frequency domain plots of the collected signal of fault rolling bearing are shown in Figure 15.

The vibration signal of IRF rolling bearing is used here to verify the effectiveness of proposed AHLPL denoising method; thus, the optimal order can be chosen as 3 among the proposed AHLPL denoising methods. To better analyze the denoising effect, the frequency range of 0–300 Hz is analyzed. To observe the fault characteristic frequency clearly and conduct a comparative analysis to verify the effectiveness of the proposed method, the partially enlarged frequency plots of frequency domain of collected signal and WPD denoised

signal are shown in Figure 16. The frequency domain plots of the vibration signal processed by the LP denoising method and the proposed AHLPL denoising method are presented in Figure 17.

It can be seen from Figure 16(a) that in the frequency domain plot of collected signal of IRF rolling bearing, the collected signal was interfered by background noises to a great extent, and there are many interfering frequencies. The rotational frequency and fault characteristic frequencies cannot be found, and it indicates there was an incipient fault happened on rolling bearing. From Figure 16(b), it can be seen after the processing of WPD that the harmonic frequency $2f_i$ can be extracted, along with many interfering frequencies. It is hard to identify among all interfering frequencies. From Figure 17(a), it can be seen after LP denoising that the interfering frequencies can be reduced or eliminated to a degree, but fault frequency f_i and rotational frequency f_r still cannot be extracted. While the effectiveness of LP denoising method is better than WPD from the comparative analysis. It can be seen from Figure 17(b) that after AHLPL denoising method, the characteristic frequencies are obvious, and the background noise is greatly reduced and eliminated. There are not many interfering frequencies in the frequency domain plot except the characteristic frequencies containing the rotational frequency f_r and fault characteristic frequencies f_i , $2f_i$. The denoising effect of AHLPL is better than LP and WPD denoising methods, which indicates the extension research of LP to AHLPL is a successful exploration. Based on the displayed results and comparative analysis, it draws

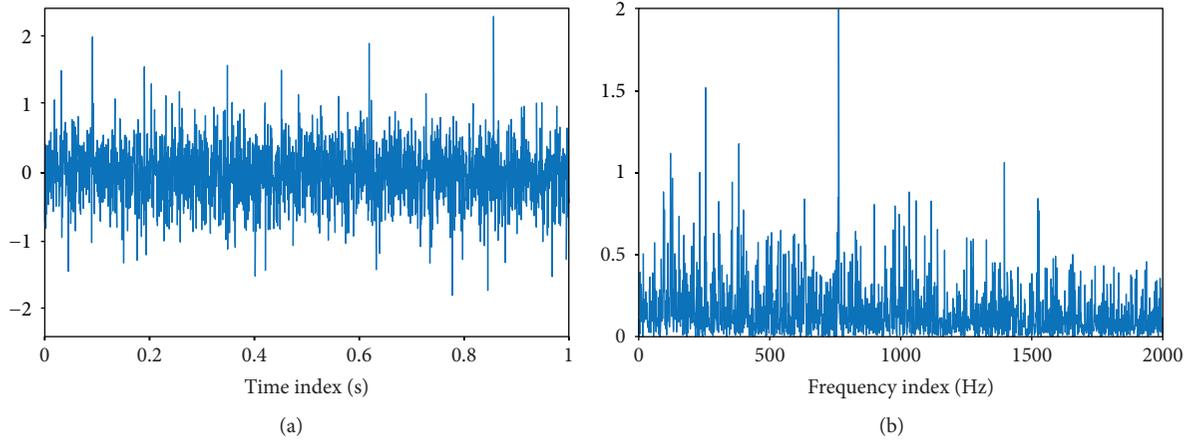


FIGURE 15: Time and frequency domain plots of collected vibration signal of IRF rolling bearing.

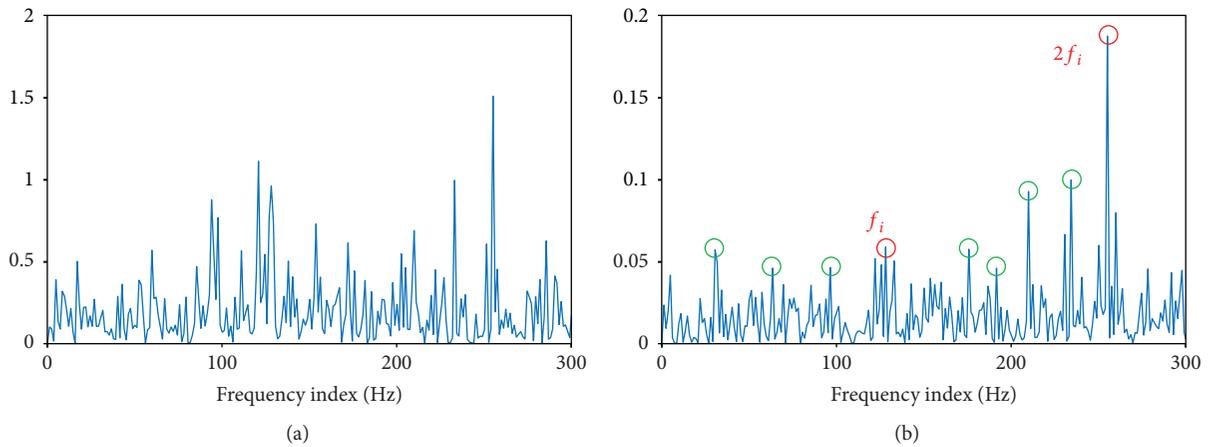


FIGURE 16: (a) Frequency domain plot of collected signal. (b) Frequency domain plot of denoised signal by WPD.

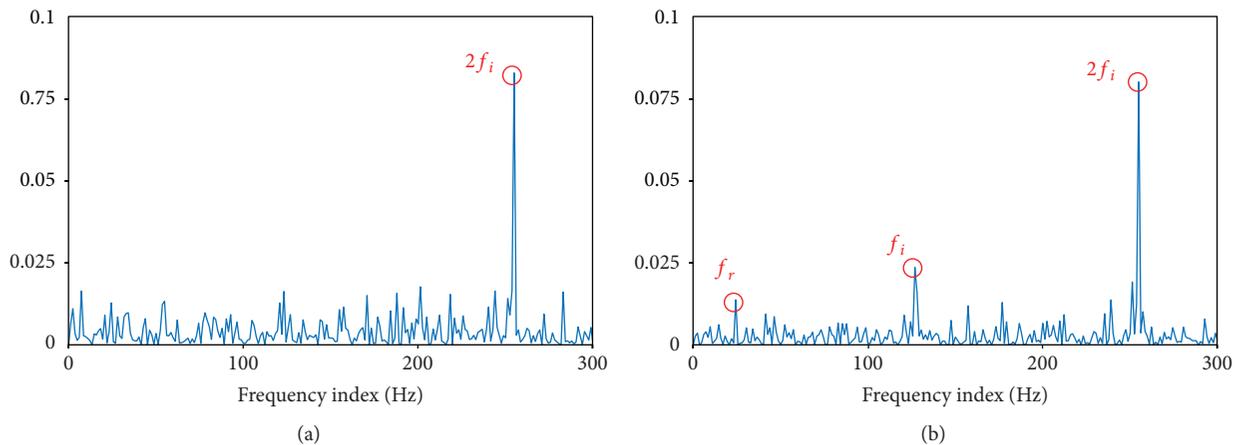


FIGURE 17: (a) Frequency domain plot of denoised signal by standard LP denoising method. (b) Frequency domain plot of denoised signal by AHLP denoising method.

a conclusion that the vibration signal of IRF rolling bearing with incipient fault can be well denoised by the proposed AHLP denoising method in practical application. It can further demonstrate the effectiveness of the proposed approach in this paper.

5. Conclusions

In this paper, the research work elaborates the validity and effectiveness of the proposed novel AHLP denoising approach in fault diagnosis of rolling bearing, through

theoretical deviation, numerical simulations, and practical applications. In the field of fault detection and isolation, it is of great significance to reduce noise in vibration signal of faulty rolling bearing, to extract fault characteristic frequencies correctly and effectively. By adopting high-order polynomials to the estimate centroid of neighborhood among LP denoising method, the proposed HLP method can achieve better noise reduction effect. By choosing the optimal order among HLP denoising method aiming at vibration signals of ORF and IRF rolling bearings, the optimal denoising effect can be obtained in the proposed AHLPL denoising method. The proposed method has exhibited good performance during the numerical simulations and application researches, containing simulated and practical vibration signal processing of ORF and IRF rolling bearings. In application researches, the fault characteristic frequencies can be both well extracted for the rolling bearing with severe fault corresponding to IMS bearing data processing and the rolling bearing with incipient fault corresponding to our own conducted experiment on Drivetrain Diagnostics Simulator. The displayed results and comparative analysis indicate that the proposed method can achieve good denoising results towards different degrees of faults. To sum up, the research work in this paper can demonstrate the significance and superiority of the proposed novel approach in fault diagnosis of rolling bearing.

Furthermore, this paper mainly deals with vibration signal of ORF and IRF rolling bearing, while in the field of rotary machinery, most of the denoising methods cater to various fault detections, such as gear faults and rotor faults. In our researches, gear faults have been dealt with and proper denoising effect can be achieved, so the proposed novel approach holds potential for wide employment in practical engineering applications. For the practical application in real industrial situation, the additional overhead costs concerning hardware, software, installation, and other factors should be also considered. This aspect of the proposed approach would be researched in our future work to explore its suitable application fields.

Data Availability

The authors also appreciate the free download of the original bearing failure data and one photo picture provided by Intelligent Maintenance Systems (IMS) Center of University of Cincinnati.

Conflicts of Interest

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

Acknowledgments

This research project is supported by the National Natural Science Foundation of China under Grant no. 51475339 and no. 51105284, the Natural Science Foundation of Hubei Province under Grant no. 2016CFA042, and the

Postgraduate Overseas Visiting Scholar Fund of Wuhan University of Science and Technology.

References

- [1] Y. Li, M. Xu, Y. Wei, and W. Huang, "Rotating machine fault diagnosis based on intrinsic characteristic-scale decomposition," *Mechanism and Machine Theory*, vol. 94, pp. 9–27, 2015.
- [2] Y. Lv, R. Yuan, T. Wang, H. Li, and G. Song, "Health degradation monitoring and early fault diagnosis of a rolling bearing based on CEEMDAN and improved MMSE," *Materials*, vol. 11, no. 6, p. 1009, 2018.
- [3] H. Liu, J. Zhang, Y. Cheng, and C. Lu, "Fault diagnosis of gearbox using empirical mode decomposition and multi-fractal detrended cross-correlation analysis," *Journal of Sound and Vibration*, vol. 385, pp. 350–371, 2016.
- [4] R. Burdzik, L. Konieczny, and P. Folega, "Structural health monitoring of rotating machines in manufacturing processes by vibration methods," *Advanced Materials Research*, vol. 1036, pp. 642–647, 2014.
- [5] Y. Tian, J. Ma, C. Lu, and Z. Wang, "Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine," *Mechanism and Machine Theory*, vol. 90, pp. 175–186, 2015.
- [6] Y. Tian, Z. Wang, and C. Lu, "Self-adaptive bearing fault diagnosis based on permutation entropy and manifold-based dynamic time warping," *Mechanical Systems and Signal Processing*, vol. 114, pp. 658–673, 2019.
- [7] X. Wang, Y. Zi, and Z. He, "Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 285–304, 2011.
- [8] H. Liu and M. Han, "A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings," *Mechanism and Machine Theory*, vol. 75, pp. 67–78, 2014.
- [9] Y. Lv, R. Yuan, and G. Song, "Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing," *Mechanical Systems and Signal Processing*, vol. 81, pp. 219–234, 2016.
- [10] Y. Li, M. Xu, H. Zhao, and W. Huang, "Hierarchical fuzzy entropy and improved support vector machine based binary tree approach for rolling bearing fault diagnosis," *Mechanism and Machine Theory*, vol. 98, pp. 114–132, 2016.
- [11] M. Cao and P. Qiao, "On the wavelet–fractal nonlinear damage diagnosis of mechanical systems," *Smart Materials and Structures*, vol. 18, no. 8, article 085022, 2009.
- [12] A. Tabrizi, L. Garibaldi, A. Fasana, and S. Marchesiello, "Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine," *Meccanica*, vol. 50, no. 3, pp. 865–874, 2015.
- [13] H. Ocak, K. A. Loparo, and F. M. Discenzo, "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics," *Journal of Sound and Vibration*, vol. 302, no. 4–5, pp. 951–961, 2007.
- [14] Y. Pan, J. Chen, and X. Li, "Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy c-means," *Mechanical Systems and Signal Processing*, vol. 24, no. 2, pp. 559–566, 2010.

- [15] Z. Qiao and Z. Pan, "SVD principle analysis and fault diagnosis for bearings based on the correlation coefficient," *Measurement Science and Technology*, vol. 26, no. 8, article 085014, 2015.
- [16] H. Jiang, J. Chen, G. Dong, T. Liu, and G. Chen, "Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 52-53, pp. 338–359, 2015.
- [17] W. Guo and P. W. Tse, "A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals," *Journal of Sound and Vibration*, vol. 332, no. 2, pp. 423–441, 2013.
- [18] F. Takens, *Detecting Strange Attractors in Turbulence// Dynamical Systems and Turbulence, Warwick 1980*, Springer, Berlin, Heidelberg, 1981.
- [19] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 6, pp. 3403–3411, 1992.
- [20] D. M. Walker and N. B. Tuffillaro, "Phase space reconstruction using input-output time series data," *Physical Review E*, vol. 60, no. 4, pp. 4008–4013, 1999.
- [21] C. A. L. Pires and A. Hannachi, "Independent subspace analysis of the sea surface temperature variability: non-Gaussian sources and sensitivity to sampling and dimensionality," *Complexity*, vol. 2017, Article ID 3076810, 23 pages, 2017.
- [22] I. Stojkovic and Z. Obradovic, "Sparse learning of the disease severity score for high-dimensional data," *Complexity*, vol. 2017, Article ID 7120691, 11 pages, 2017.
- [23] M. P. Mareca and B. Bordel, "Improving the complexity of the Lorenz dynamics," *Complexity*, vol. 2017, Article ID 3204073, 16 pages, 2017.
- [24] D. Chelidze and W. Zhou, "Smooth orthogonal decomposition-based vibration mode identification," *Journal of Sound and Vibration*, vol. 292, no. 3–5, pp. 461–473, 2006.
- [25] D. Chelidze and G. Chelidze, "Nonlinear model reduction based on smooth orthogonal decomposition," in *Proceedings of the Ninth IASTED International Conference on Control and Applications*, pp. 325–330, ACTA Press, Montreal, Quebec, 2007.
- [26] S. Ilbeigi and D. Chelidze, "Persistent model order reduction for complex dynamical systems using smooth orthogonal decomposition," *Mechanical Systems and Signal Processing*, vol. 96, pp. 125–138, 2017.
- [27] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, "On noise reduction methods for chaotic data," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 3, no. 2, pp. 127–141, 1993.
- [28] W. F. He, L. Cheng, X. G. Li, and J. Luo, "Nonlinear noise reduction and its application to the vibration signal analysis," *Journal of Vibration Engineering*, vol. 17, pp. 408–411, 2004.
- [29] T. Sauer, "A noise reduction method for signals from nonlinear systems," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1–4, pp. 193–201, 1992.
- [30] X. Wan, K. Iwata, J. Riera, T. Ozaki, M. Kitamura, and R. Kawashima, "Artifact reduction for EEG/fMRI recording: nonlinear reduction of ballistocardiogram artifacts," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 668–680, 2006.
- [31] N. Jevtić, P. Stine, and J. S. Schweitzer, "Nonlinear time series analysis of Kepler space telescope data: mutually beneficial progress," *Astronomische Nachrichten*, vol. 333, no. 10, pp. 983–986, 2012.
- [32] L. Qian, J. Chen, and X. Feng, "Local projection stabilized and characteristic decoupled scheme for the fluid–fluid interaction problems," *Numerical Methods for Partial Differential Equations*, vol. 33, no. 3, pp. 704–723, 2017.
- [33] N. Jevtic, J. S. Schweitzer, and P. Stine, "Optimizing nonlinear projective noise reduction for the detection of planets in mean-motion resonances in transit light curves," *Chaos Theory*, pp. 191–198, 2011.
- [34] R. Cawley and G. H. Hsu, "Local-geometric-projection method for noise reduction in chaotic maps and flows," *Physical Review A*, vol. 46, no. 6, pp. 3057–3082, 1992.
- [35] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: the TISEAN package," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 9, no. 2, pp. 413–435, 1999.
- [36] J. M. Moore, M. Small, and A. Karrech, "Improvements to local projective noise reduction through higher order and multiscale refinements," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 6, article 063114, 2015.
- [37] A. Cordero, E. Gómez, and J. R. Torregrosa, "Efficient high-order iterative methods for solving nonlinear systems and their application on heat conduction problems," *Complexity*, vol. 2017, Article ID 6457532, 11 pages, 2017.
- [38] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 2004.
- [39] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1–2, pp. 43–50, 1997.
- [40] J. Antoni and R. B. Randall, "A stochastic model for simulation and diagnostics of rolling element bearings with localized faults," *Journal of Vibration and Acoustics*, vol. 125, no. 3, pp. 282–289, 2003.
- [41] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.
- [42] R. B. Randall, J. Antoni, and S. Chobsaard, "The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals," *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 945–962, 2001.
- [43] J. Antoni and R. B. Randall, "Differential diagnosis of gear and bearing faults," *Journal of Vibration and Acoustics*, vol. 124, no. 2, pp. 165–171, 2002.
- [44] Y. N. Pan, J. Chen, and X. L. Li, "Spectral entropy: a complementary index for rolling element bearing performance degradation assessment," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 223, no. 5, pp. 1223–1231, 2009.
- [45] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, vol. 289, no. 4–5, pp. 1066–1090, 2006.

Research Article

Damage Diagnosis in 3D Structures Using a Novel Hybrid Multiobjective Optimization and FE Model Updating Framework

Nizar Faisal Alkayem ¹, Maosen Cao ¹ and Minvydas Ragulskis ^{1,2}

¹Department of Engineering Mechanics, Hohai University, Nanjing, Jiangsu 210098, China

²Department of Mathematical Modeling, Kaunas University of Technology, Kaunas, Lithuania

Correspondence should be addressed to Maosen Cao; cmszhy@hhu.edu.cn

Received 31 December 2017; Revised 30 June 2018; Accepted 9 July 2018; Published 19 September 2018

Academic Editor: Ana Meštrović

Copyright © 2018 Nizar Faisal Alkayem et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Structural damage detection is a well-known engineering inverse problem in which the extracting of damage information from the dynamic responses of the structure is considered a complex problem. Within that area, the damage tracking in 3D structures is evaluated as a more complex and difficult task. Swarm intelligence and evolutionary algorithms (EAs) can be well adapted for solving the problem. For this purpose, a hybrid elitist-guided search combining a multiobjective particle swarm optimization (MOPSO), Lévy flights (LFs), and the technique for the order of preference by similarity to ideal solution (TOPSIS) is evolved in this work. Modal characteristics are employed to develop the objective function by considering two subobjectives, namely, modal strain energy (MSTE) and mode shape (MS) subobjectives. The proposed framework is tested using a well-known benchmark model. The overall strong performance of the suggested method is maintained even under noisy conditions and in the case of incomplete mode shapes.

1. Introduction

Structural health monitoring (SHM) and failure diagnosis have recently witnessed increasing attention because of its importance for guaranteeing the integrity and service perfection of a structure. One of the main research foci of SHM is damage detection in structures under service, in which a preliminary assessment of damage is conducted to ensure the reliability and persistent performance of the structure. The evolution of structural damage detection techniques has been mainly inspired by the invention of accurate sensors and transducers in addition to improvements in signal processing techniques, which in combination were applied to nondestructive vibrational testing that reveals the dynamic characteristics of a structure. Those dynamic characteristics carry the changes in real structural parameters that in turn illustrate the existence and rate of structural damage [1–3].

Structural damage could be defined as variations between the FE model of a structure and the structure in-service that brings the concept of FE model updating into a picture [4]. FE model updating can provide an efficient

way of tracking damage in structures under service. For a structural damage identification problem, the structural parameters related to the FE model of a structure are slightly and gradually regulated and dynamic responses are registered. Thereafter, distortions between the model's and measured dynamic responses are minimized until a relative consistency is reached. During that process, such local modifications can indicate damage in the studied structure. The minimization task can be thought through by utilizing efficient optimization algorithms [5–7].

In order to integrate the FE model updating into a structural damage tracking procedure, the basic task is to establish the dynamic characteristic subobjectives. The developed subobjectives are milestones in formulating the objective function that addresses the optimization problem of damage characterization. Modal analysis can provide an effective manner for gathering dynamic features. Mode shapes (MSs) are fundamental characteristics that can give general information about the alterations that have occurred in a structure due to damage. For that reason, MS subobjective can be developed by using some correlation paradigms such as the

modal assurance criterion (MAC) [8, 9]. Another modal property with great damage sensitivity, modal strain energy (MSTE), has been used by various researchers [10–14]. It performs efficiently when utilized for tackling damage problems in complex structures experiencing damage. Furthermore, MSTE is applicable in three-dimensional structures [14, 15]. Hence, both mode shape-based and MSTE-based subobjectives are more than sufficient to be employed in this research.

Computational intelligence has been implemented for solving complex prediction, diagnosis, and detection problems in various fields such as mechanical engineering [16–18], computer science [19, 20], biomedical engineering [21, 22], and electrical engineering [23]. As major disciplines of computational intelligence, swarm intelligence and EAs are powerful stochastic optimization techniques that have been implemented for solving various engineering problems in recent years. Single-objective EAs have been widely and successfully utilized for damage localization in structures. Research has shown the efficacy of genetic algorithms (GAs), the most common form of EA [24–27]. Other representative research can be found in [28, 29]. Swarm intelligent algorithms have also been implemented for the same purpose. Ding et al. [30] developed a hybrid swarm for damage identification using modal properties with pleasing results. Qian et al. [31] implemented a hybrid PSO simplex method for delamination detection in laminated beams using a delamination parameter-based objective function with robustness and efficient performance. Kang et al. [32] as well as Gökdağ and Yildiz [33] proposed two different PSO versions to track damage with successful performance. Zhu et al. [34] developed a bird mating optimizer (BMO) in the time-frequency domain for damage detection in 2D structures. Other techniques have been used efficiently, as in the work of Seyedpoor et al. [35, 36]. From the above survey, it is seen that single-objective EAs have performed well in solving structural damage detection problems. Nevertheless, most applications were 1D or 2D problems.

Multiobjective EAs (MOEAs) have also been implemented by various researchers for damage localization in structures. Cha and Buyukozturk [14, 15] developed a MOEA for damage prognostic in three-dimensional structures. Their research outcomes showed the robustness of the proposed method even under noisy environments. Liu et al. [37] used natural frequency and MS subobjectives with the successful incorporation of the weighted sum method and a GA to perform a multiobjective optimization for damage identification. Jung et al. [38] utilized the NSGAI for damage identification in truss structures, concluding that further enhancements were needed for the proposed method to be recommended for damage deduction. NSGAI was also implemented successfully by Shabbir and Omenzetter [39]. Other representative research into the application of NSGAI for damage detection can be found in [40, 41]. Farokhzad et al. [42] compared multiobjective GA (MOGA) and modified multiobjective GA (MMOGA) for damage localization in Timoshenko beams, with remarkable results obtained from MMOGA. MOGA was also successfully employed for damage detection in truss by Jung et al. [43].

Other MOEAs have been implemented for structural damage identification, as in [44, 45].

A literature survey of the implementation of MOEAs showed that relevant research has rarely addressed the problem of structural damage in complex structures. Hence, further research is important to explore the effectiveness of MOEAs and swarm intelligence in structural damage localization. Moreover, it is essential to introduce more powerful algorithms that can efficiently and accurately solve the problem of damage identification in 3D structures. For those reasons, a novel multiobjective optimization algorithm combines the multiobjective particle swarm optimization (MOPSO), Lévy flights (LFs), and the technique for the order of preference by similarity to ideal solution (TOPSIS) with a FE model updating framework is proposed for solving the problem of damage estimation in complex 3D structures. The coherence of the proposed algorithm can be measured by the combination between the global and local search of MOPSO and LFs. Furthermore, the integration of TOPSIS inside the algorithm to iteratively select a leader solution can provide an efficient paradigm that improves the overall performance of the algorithm.

This work illustrates the application of the novel multiobjective optimization algorithm called MOLFPSO/TOPSIS applied on FE model updating for structural damage localization. Two dynamic characteristic-based subobjectives, namely, MS- and MSTE-based subobjectives, are embodied in an objective function. The proposed technique is applied on a 4-story benchmark building model. In order to examine the performance of the developed technique, two damage scenarios are studied. Furthermore, the algorithm is tested under noisy conditions in correspondence to the two damage scenarios. Also, the incomplete mode shape problem is evaluated. At last, a detailed discussion of the performance of the suggested method is presented.

2. The Objective Function

It has been observed that in structural damage tracking combined with the FE model updating, the most critical key point is the proper selection of dynamic characteristics that includes the changes in structural parameters due to the occurrence of structural damage. On that basis, the subobjectives of errors between the dynamic features of the FE model of the intact structure and those measured from the damaged structure are developed. By the formulation of those subobjectives, they can be merged in such a way that an objective function for the damage detection optimization problem can be established. In the current research, both MS- and MSTE-based subobjectives are employed and engrafted into the objective function to be solved for the damage localization.

To develop the MS subobjective, a correlation technique should be employed [8]. One of the most commonly used correlation techniques is the MAC [46]. The MAC measures the uniformity between two vectors, one of which being a reference MS and the other a measured MS. The MAC matrix can be stated as

$$\text{MAC}(\{\varphi_I\}, \{\varphi_D\}) = \frac{|\{\varphi_I^*\}^T \{\varphi_D\}|^2}{\{\varphi_I^*\}^T \{\varphi_I\} \{\varphi_D^*\}^T \{\varphi_D\}}, \quad (1)$$

where $\{\varphi\}$ is the MS, I and D represent the undamaged and the damaged structures, respectively, and T and $*$ are the transpose and the conjugate. It is defined that full correlation is observed when MAC is one and no correlation is deduced when MAC is zero.

The MS subobjective can be expressed as

$$F_1(\varphi^I, \varphi^D) = \sum_{i=1}^M (1 - \text{diag}(\text{MAC}_i(\{\varphi_i^I\}, \{\varphi_i^D\}))), \quad (2)$$

where $F_1(\varphi^I, \varphi^D)$ is the MS subobjective, i is the MS number, and M is the total number of MSs diag is the diagonal element of MAC.

MSTE can serve as a sensitive indicator of structural damage. In a large-scale complex structure, MSTE can hold better information about the existence and position of minor damage than other modal characteristics. One of the most commonly occurring types of structural damage can be thought of as a stiffness lessening [47]. In spite of the reality that stiffness alterations cannot reflect all sorts of structural damage, it can be adopted in this research to simulate linearly structural damage. Consequently, by handling the MSs of the structure and the stiffness matrix \mathbf{K} of the model, the MSTEs corresponding to healthy and damaged structures can be defined as

$$\begin{aligned} \text{MSTE}_i^I(\varphi_i^I, K) &= \frac{1}{2} \varphi_i^{I^T} K \varphi_i^I, \\ \text{MSTE}_i^D(\varphi_i^D, K) &= \frac{1}{2} \varphi_i^{D^T} K \varphi_i^D, \end{aligned} \quad (3)$$

where MSTE_i^I and MSTE_i^D are the MSTEs of the robust and deteriorated structures for the i th MS, respectively. The global MSTE subobjective can be written as

$$F_2(\varphi^I, \varphi^D, K) = \sum_{i=1}^M \left(1 - \frac{\varphi_i^{I^T} K \varphi_i^I}{\varphi_i^{D^T} K \varphi_i^D} \right)^2, \quad (4)$$

where F_2 is the MSTE subobjective. Based on the above description, the overall hybrid objective function is built as

$$\begin{aligned} \text{Min}(F_1, F_2) = \text{Min} & \left(\sum_{i=1}^M (1 - \text{diag}(\text{MAC}_i(\{\varphi_i^I\}, \{\varphi_i^D\}))), \right. \\ & \left. \sum_{i=1}^M \left(1 - \frac{\varphi_i^{I^T} K \varphi_i^I}{\varphi_i^{D^T} K \varphi_i^D} \right)^2 \right). \end{aligned} \quad (5)$$

Damage causes changes in the mechanical parameters of a structure [48]. The parametric changes can be directly used to track damage by performing a FE model updating framework [8, 14]. Commonly, to select the appropriate updating parameters that can be used as damage indicators, the possible damage patterns must be defined depending on the structure under consideration and loading conditions. Some

recommendations can be followed before choosing the updating parameters. First, greater importance should be assigned to the positions where structural damage is more likely to exist. Then, dynamic characteristics that are more damage-sensitive must be considered. Finally, efforts should be made to reduce the size of the updating parameter set so that the overall complexity of the problem can be tackled, by ignoring structural parameters that do not illustrate possible existing damage scenarios [49, 50].

One of the most widely used parameterization methods is to adopt the material and geometrical properties directly [51]. Material properties can be chosen as damage indexes for tracking structural damage, because the stiffness and mass matrices are proportional to material properties. A useful strategy is to develop a vector of damage indexes for all elements (el) as $\Theta = [\theta_1, \theta_2, \dots, \theta_{el}]$, where $\theta \in [0, 1]$. If we choose a damage conductor as v , Θ is utilized to reflect the variations between the FE model parameter v^I and the parameter v^U related to damage simulation. The damage parameter can be defined as

$$v_i^U = v_i^I (1 - \theta_i), \quad (6)$$

and the changes in model's element matrixes \mathbf{M} and \mathbf{K} can be written as

$$\begin{aligned} M_i^U &= M_i^I + \Delta M_i, \\ K_i^U &= K_i^I + \Delta K_i, \\ \Delta M_i &= v_i^U \cdot M_i^I, \\ \Delta K_i &= v_i^U \cdot K_i^I, \end{aligned} \quad (7)$$

where ΔM_i and ΔK_i are the deviations in model's matrixes.

To examine the performance of the proposed MOLFPSO/TOPSIS, the structural damage is presumed to be a reduction of some elements' moduli of elasticity (E). It is important to mention that reductions in E cannot mimic all sorts of structural damage, but it can be used to simulate damage to check the performance and robustness of the proposed technique.

3. Structural Damage Localization Framework

To tackle the structural damage estimation problem, it is essential to organize the optimization problem into two main divisions: the single-objective EAs and the MOEAs. The key points to understand the differences between the two divisions can be shown in [52, 53].

Structural damage tracking using MOEAs and FE model updating is shown in Figure 1. We summarize the framework as follows: initially, a FE model of the robust structure is evolved and the damage indicators are selected. Then, the framework is partitioned into two subprocesses. The former is the damage simulation process where the damage-simulated MSTE and MS subobjectives can be deduced. The second subprocess is to perform the minimization task applied on the objective function. The detailed overall framework can be well observed in [52].

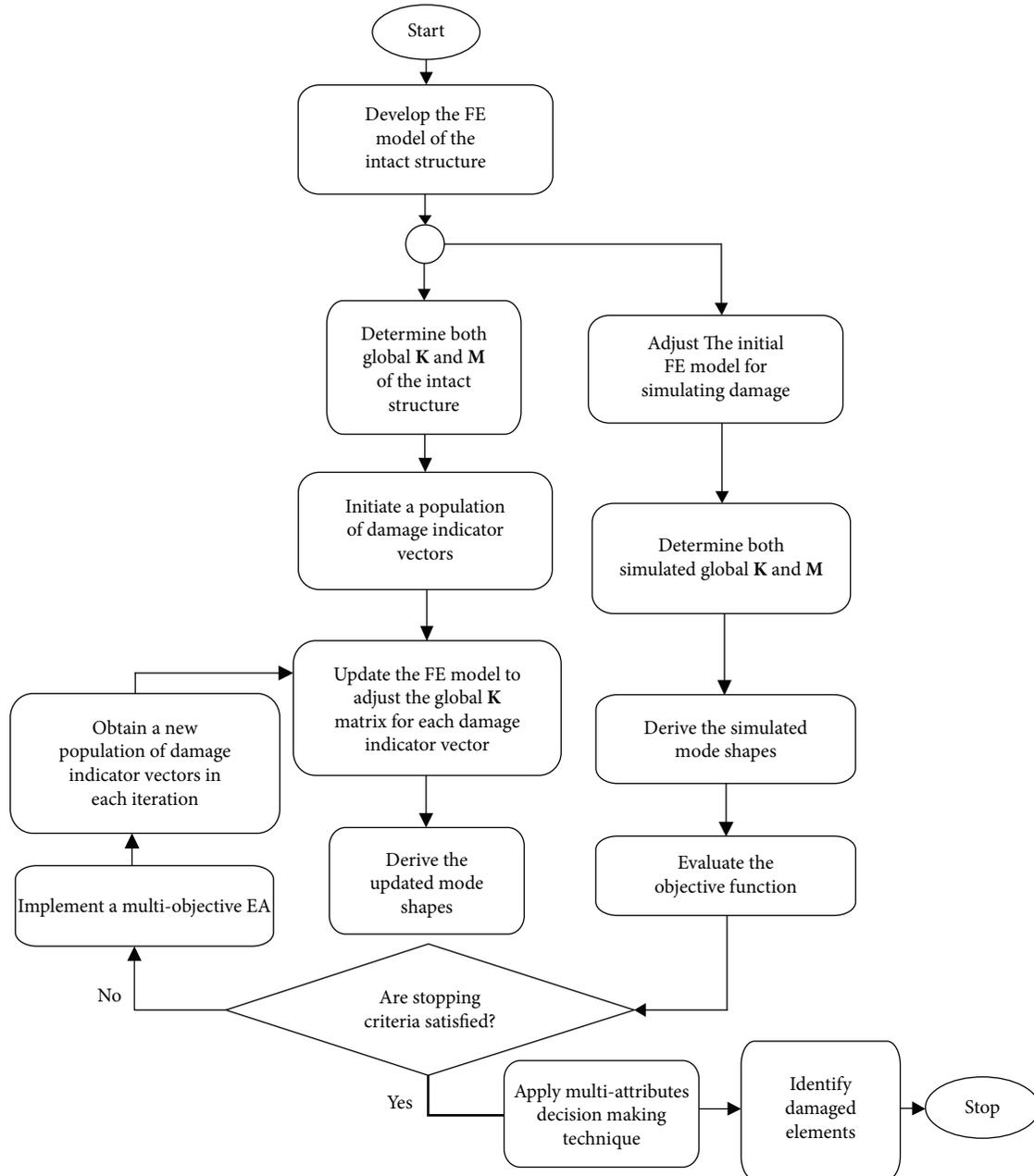


FIGURE 1: Damage prognostic paradigm using the MOEAs and FE model updating.

4. The MOLFPSO/TOPSIS

By following back to the excellent research conducted by Cha and Buyukozturk [14, 15], it is observed that one dynamic characteristic has been used inside the multiobjective optimization problem by dividing the MSTE-based objective function into two subobjectives. In the current research, an improvement is suggested by adding an extra subobjective that relies on the MSs. The use of independent mode shape subobjective can study the alterations in the structure independent from the stiffness matrix utilized in MSTE and without losing the benefit of damage detectability when only MSTE is utilized. This can help

to improve the overall ability of the damage identification procedure when noisy conditions exist as well as when the complete MSs cannot be obtained. Moreover, because MSTE can describe only the linear type of structural damage, adding the mode shape term can enhance the damage deducibility of other types.

The MOLFPSO/TOPSIS algorithm combines various paradigms and concepts to perform a strong and reliable multiobjective optimization framework to solve complex problems comprising the problem of damage tracking in structures with FE model updating. To provide a comprehensive elucidation of the developed method, some preliminary concepts and techniques must be explained. Firstly, Lévy

1. Create a population of random solutions Pop and corresponding speeds Vel .
2. Initialize an external archive Ac and define a memory of flight experience for each individual, i.e. $Pbest$.
3. Evaluate Pop using the multi-objective functions.
4. Find the non-dominated solutions and store them in Ac .
5. Develop hypercubes of the search space and distribute the individuals within the hypercubes using the multi-objective function values.
6. Apply TOPSIS on Ac to determine the leader solution or the global best individual $Gbest$.
7. **While** stopping criteria are not satisfied **do**:
 - a) Generate a new population A by using LFs using a modified version of Eq. (9) as:

$$s = a_0(Gbest - x_{G,i}) \oplus \text{Lévy}(\lambda) \sim 0.01(u/|v|^{1/\lambda})(Gbest - X_{G,i})$$

$$\overline{P}_{i+1} = P_i + s$$
 - b) Evaluate all particles in A , compare A with Pop using the non-domination criterion and update Pop .
 - c) Generate a new population B by using PSO as:

$$V_i^{t+1} = W \times V_i^t + r_1 \times (Pbest_i - P_i) + r_2 \times (Gbest - P_i),$$

$$P_{i+1} = P_i + v_i^{t+1},$$
 where p_i is the current particle; v_i is the particle's velocity; w is the inertia factor; r_1 and r_2 are acceleration coefficients; r_1 and r_2 imply random numbers $\in [0, 1]$; t is the current generation number.
 - d) Evaluate all particles in B , compare B with Pop using the non-domination criterion and update Pop .
 - e) Remove the non-dominated solutions and fill the empty positions with randomly created solutions as:

$$In_i = \text{randper}(P_{new,i}),$$

$$\overline{P}_{new,i} = \text{rnd}(0, 1) + P_{new}(In_i),$$
 where randper is a Gaussian random permutation operator; $P_{new,i}$ is the i^{th} removed dominated solution; rnd is a random number $\in [0, 1]$.
 - f) Evaluate all particles in Pop and insert the non-dominated solution into Ac .
 - g) Update the hypercubes in the current Ac .
 - h) Remove the extra individuals in Ac by eliminating the crowded individuals within the corresponding hypercubes.
 - i) Apply TOPSIS on Ac to determine the global best individual $Gbest$.
 - j) Update $Pbest$ for all individuals.
- 8 **End**.

ALGORITHM 1

flights (LFs) can be defined as random permutations that are capable of mimicking the random movements of creatures seeking food, reproduction, or other activities [54]. LFs can be implemented efficiently to expose unknown search space for optimization purposes. Moreover, LFs can perform both global and local searches around promising solutions that can be referred to as the exploration and exploitation of the global search space. To generate a new solution X_i^{t+1} by LF, (8) can be used:

$$X_i^{t+1} = X_i^t + s, \quad (8)$$

where s is the step size, proposed by Yang and Deb [55]:

$$s = a_0(x_{G,j} - x_{G,i}) \oplus \text{Lévy}(\lambda) \sim 0.01 \frac{u}{|v|^{1/\lambda}} (X_{G,j} - X_{G,i}), \quad (9)$$

where a_0 is a constant, $X_{G,j}$ and $X_{G,i}$ are two random solutions, and u and v can be defined as

$$\begin{aligned} u &= N(0, \sigma_u^2), \\ v &= N(0, \sigma_v^2), \end{aligned} \quad (10)$$

where $\sigma_u = [\Gamma(1 + \lambda) \sin(\pi\lambda/2) / \Gamma((1 + \lambda)/2)] \lambda 2^{(\lambda-1)/2}]^{1/\beta}$ with $1 \leq \lambda \leq 2$ is an index, Γ is a gamma function, and $\sigma_v = 1$.

TOPSIS is a MCDM technique that was proposed by Hwang and Yoon [56]. From any decision matrix with m number of alternatives and n number of attributes, it chooses two sets of solutions known as the positive (X^+) and negative (X^-) ideal solutions. The solution with the best data from all the alternatives is called the positive ideal solution; similarly, the negative ideal solution has the worst data from all the alternatives. TOPSIS calculates the Euclidean distance of all points (alternatives) from two positive and negative solutions and compares them. The point that has the least Euclidean

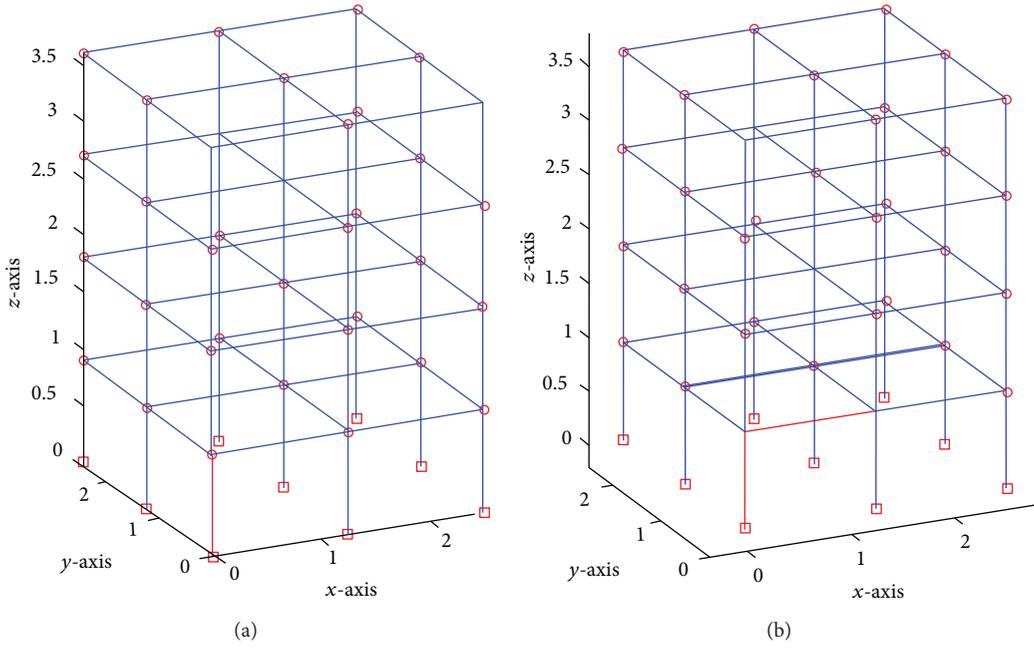


FIGURE 2: Damage scenarios: (a) scenario 1 and (b) scenario 2.

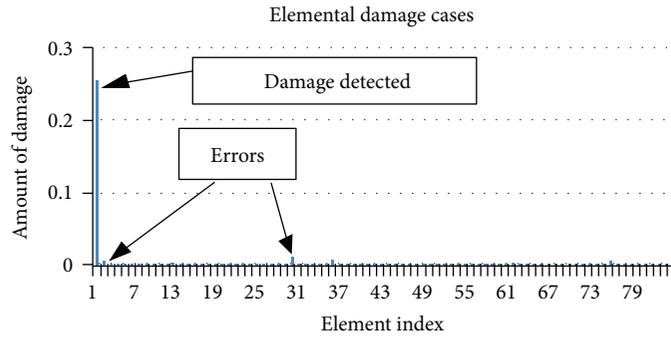


FIGURE 3: Damage locations in damage scenario 1.

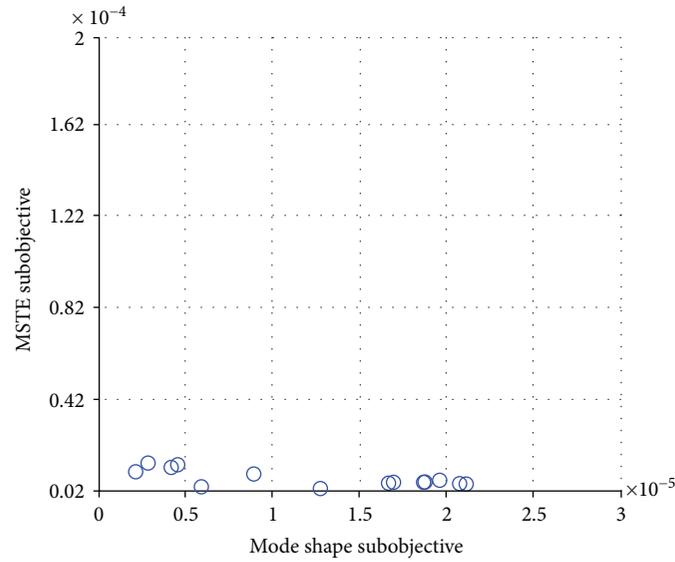


FIGURE 4: Optimal front obtained from MOLFPSO/TOPSIS in damage scenario 1.

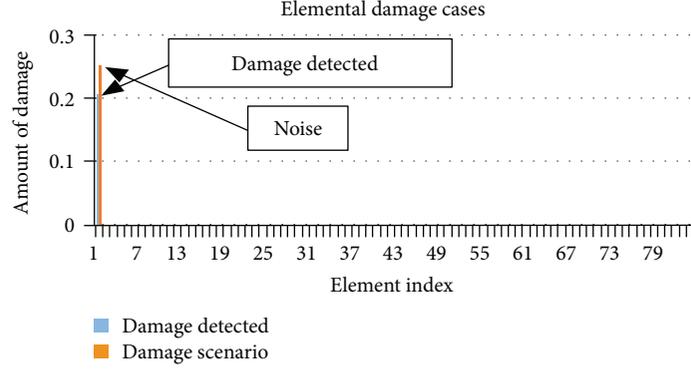


FIGURE 5: Damage positions in damage scenario 1 with noisy MSs.

distance from (X^+) and the greatest distance from (X^-) is considered the best. The steps of applying TOPSIS are as follows:

Step 1. The input \mathbf{x} matrix contains $m \times n$ elements (the elements of \mathbf{x} matrix are the Pareto solutions).

Step 2. To convert the dimensional problem into a nondimensional problem, a normalized decision matrix is performed from the \mathbf{x} matrix.

$$\bar{x}_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}}, \quad (11)$$

where $i = 1, \dots, m$ and $j = 1, \dots, n$, ($m = 100$ and $n = 5$).

Step 3. By assuming a set of weights for each attribute w_j for $j = 1, \dots, n$, we can find the weighted normalized decision matrix \mathbf{v} .

$$v_{ij} = w_j \bar{x}_{ij}, \quad (12)$$

where $\sum_{j=1}^n w_j = 1$.

Step 4. Determine the positive and negative ideal solutions.

$$\begin{aligned} X^+ &= \{v_1^+, \dots, v_n^+\}, \\ X^- &= \{v_1^-, \dots, v_n^-\}, \end{aligned} \quad (13)$$

where $v_j^+ = \{\max_i v_{ij} | j \in J, \min_i v_{ij} | j \in J'\}$ and $v_j^- = \{\min_i v_{ij} | j \in J, \max_i v_{ij} | j \in J'\}$, where J is the set of benefit attributes and J' is the set of cost attributes.

Step 5. Calculate the distances d_i^+ and d_i^- from the positive ideal and negative ideal solution, respectively.

$$\begin{aligned} d_i^+ &= \sqrt{\sum_{j=1}^n (v_j^+ - v_{ij})^2}, \\ d_i^- &= \sqrt{\sum_{j=1}^n (v_j^- - v_{ij})^2}, \end{aligned} \quad (14)$$

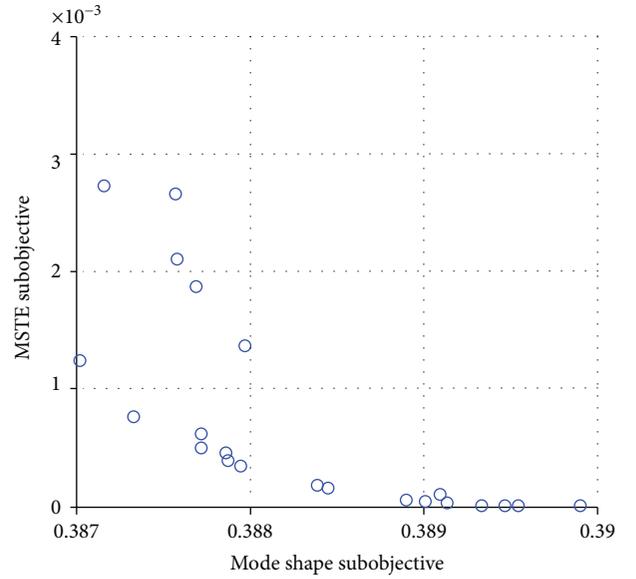


FIGURE 6: Optimal front obtained from MOLFPSO/TOPSIS in damage scenario 1 under noisy conditions.

Step 6. Determine the relative closeness of alternatives to the ideal solution.

$$C_i^+ = \frac{d_i^-}{d_i^- + d_i^+}, \quad (15)$$

where $0 \leq C_i^+ \leq 1$ and the best solutions have relative closeness C_i^+ closest to 1.

MOPSO was originally proposed by Coello and Lechuga [57] and further explained by Coello and Reyes-Sierra [58]. They extended the PSO algorithm into MOPSO by employing the Pareto ranking scheme as well as initializing an archive or repository to register the superior performance of any individual in each generation by means of sets of non-dominated solutions. The archive is exploited to choose a global best solution that leads the swarm to reach the Pareto front. In each generation, the archive should be updated by a spatially or geographically based process where the search space is partitioned into a set of hypercubes using multiobjective function rates.

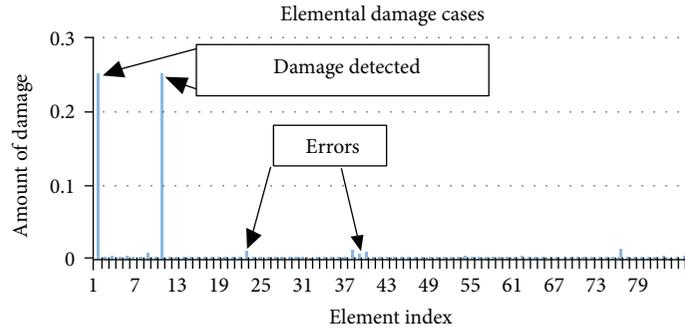


FIGURE 7: Damage locations in damage scenario 2.

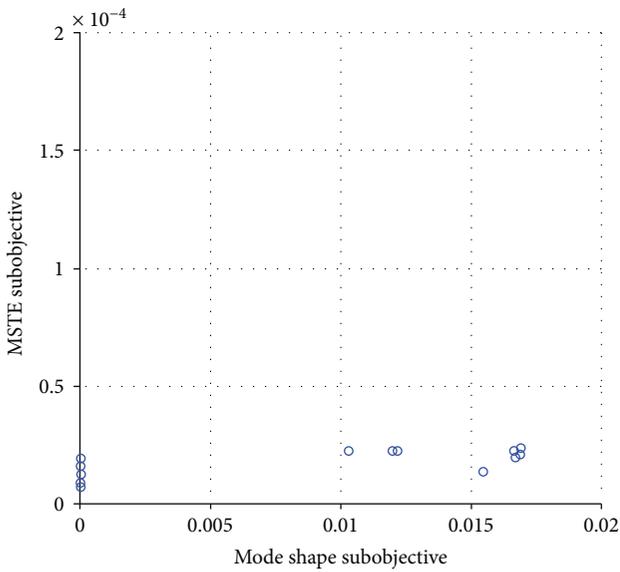


FIGURE 8: Optimal front obtained from MOLFPSO/TOPSIS in damage scenario 2.

The proposed MOLFPSO/TOPSIS is designed using the above-mentioned concepts. The key point in the proposed method is to use TOPSIS to rank all solutions in the archive and to sort them to choose the leading best solution in each generation. That helps to reduce the complexity of the problem because there is no need for a MCDM at the end of the structural damage detection process when the leader solution is chosen in each generation by TOPSIS. Another advantage is that both LFs and PSO are utilized within the algorithm in such a way that there is no need for a large population size and exploration and exploitation are well performed in each generation, reducing the computational time and improving the overall efficiency and reliability. The spatial representation in MOPSO is still employed in MOLFPSO/TOPSIS for removing crowded solutions among the archive. The MOLFPSO/TOPSIS algorithm can be summarized as follows:

Although multiobjective genetic algorithms have shown a prominent performance when applied by Cha and Buyukozturk [14] for structural damage tracking and also

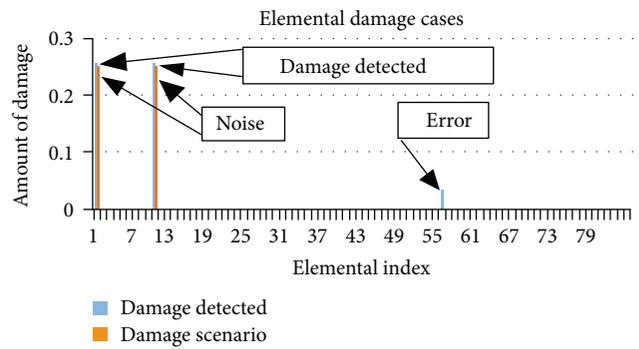


FIGURE 9: Damage positions in damage scenario 2 with noisy MSs.

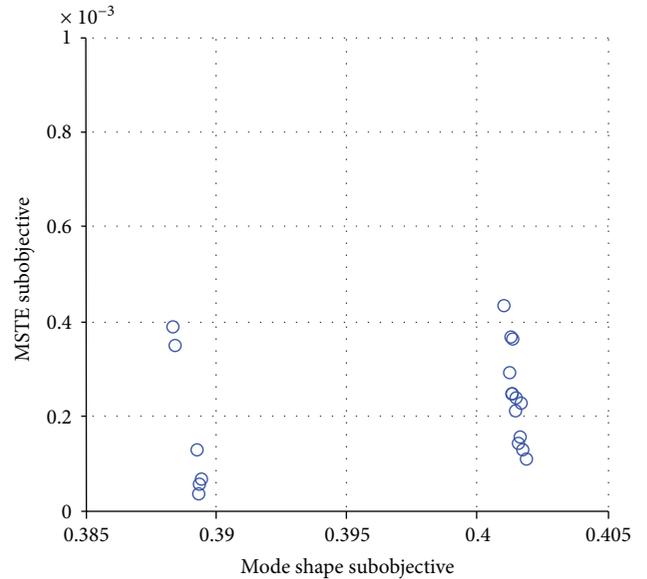


FIGURE 10: Optimal front obtained from MOLFPSO/TOPSIS in damage scenario 2 under noisy conditions.

when implemented for active control device optimal placement by Cha et al. [59], it is important to develop a stronger multiobjective optimization approach able to overcome more complexity in the structure and can not only detect but also

TABLE 1: Numerical evaluation of the MOLFPSO/TOPSIS with FE model updating.

		Min	Mean	Max
Case 1	MS subobjective	$2.18E-06$	$1.45E-05$	$2.11E-05$
	MSTE subobjective	$2.975E-06$	$6.99E-06$	$1.364E-05$
	Computational time	2551.97823 seconds		
Case 2	MS subobjective	$3.08E-06$	0.012812	0.017029
	MSTE subobjective	$7.33E-06$	$2.09E-05$	$2.56E-05$
	Computational time	4783.06526 seconds		
Case 1 with $\pm 5\%$ white noise	MS subobjective	0.387019	0.388337	0.390032
	MSTE subobjective	$1.74E-06$	0.00066	0.002721
	Computational time	4912.28630 seconds		
Case 2 with $\pm 5\%$ white noise	MS subobjective	0.388335	0.398434	0.401908
	MSTE subobjective	$7.33E-06$	$2.09E-05$	$2.56E-05$
	Computational time	7523.17741 seconds		

determine the accurate severity of structural damage. The MOLFPSO/TOPSIS algorithm can combine the well-known global optimization feature of PSO when used for FE model updating [60] with the local search ability of LFs. Such combination can serve better when damage detection accuracy is required and complex structures are under investigation. Furthermore, the guided elitist search feature provided by TOPSIS can help to reduce the computational cost and provide better convergence towards the optimal Pareto front. Also, it makes the MCDM determine the best compromise solution unnecessary.

5. Case Study: 3D Modular Structure

After benchmarking the algorithm and evaluating the performance of the proposed MOLFPSO/TOPSIS algorithm, a three-dimensional model is built similar to the benchmark model [61, 62] as shown in Figure 2. The model's dimensions are $2.5\text{ m} \times 2.5\text{ m} \times 3.6\text{ m}$. The model's beams and columns are Euler-Bernoulli beams built of hot rolled grade 300 W steel with 300 MPa nominal yield stress. The material properties of the structure are given in [62]. The model is composed of 84 elements and 45 nodes with 270 DOFs. To examine the efficiency of the developed MOLFPSO/TOPSIS technique, two damage scenarios are taken into account as in Figure 2. Damage is simulated by reducing 25% of the modulus of elasticity.

In testing the application of MOLFPSO/TOPSIS to identify damage in 3D structures, first, the previously mentioned objective function is utilized and the first 12 MSs are employed. Then, after several trials, MOLFPSO/TOPSIS is implemented using 50 particles and $W = 0.5$. The number of hypercubes in set to 7 and repository size is 25. In LFs, the parameters λ and β are set to 1.5 [54, 55]. The weighting factors in (12) are set to 0.5; that is, both subobjectives will have similar significance. The framework is executed 20 times, and results are recorded. In damage case 1, the algorithm is proved to be able to locate the damage with very minor errors, as shown in Figure 3. The optimal front

achieved by the MOLFPSO/TOPSIS is presented in Figure 4. Also, when the simulated MSs are contaminated with $\pm 5\%$ white noise, the MOLFPSO/TOPSIS successfully tracks the damage with very few significant errors, as observed in Figure 5. The optimal front in damage case 1 with noise is illustrated in Figure 6. It is obvious that the MOLFPSO/TOPSIS can detect and locate damage in the studied structure accurately and efficiently, even under noisy conditions.

In damage case 2, which is the more difficult case, the proposed technique shows excellent performance in tracking the structural damage with insignificant errors, as shown in Figure 7. The optimal front resulting from the MOLFPSO/TOPSIS is highlighted in Figure 8. Moreover, when $\pm 5\%$ Gaussian noise contaminates the simulated MSs, the developed methodology shows good ability to determine damage with acceptable errors, as shown in Figure 9. For damage case 2 with noise, the optimal front is exhibited in Figure 10. It is clear from the figures that the MOLFPSO/TOPSIS combined with FE model updating can serve as a powerful and reliable framework for identifying and diagnosing damage in the 3D structures, even within noise-polluted environments.

To fulfill the evaluation of the overall paradigm, all result outcomes are tabulated in Table 1. All minimum, mean, and maximum values of both MS-based and MSTE-based subobjectives (subobjectives) corresponding to all Pareto solutions as well as the computational time and number of successful runs are all given in Table 1. It can be seen that the multiobjective values almost match the target zero values of the dynamic subobjectives in damage cases 1 and 2. Moreover, even in noisy environments, the increase in multiobjective values due to noise has not affected the overall performance and ability to localize structural damage. Over and above these virtues, the computational time is convenient in all cases. Only in damage case 2 with noise is the number of successful runs lower, due to difficulty in inferring damage when two damaged beams are connected perpendicularly.

Another important factor to judge the overall performance is to evaluate the problem of incomplete MSs. Two

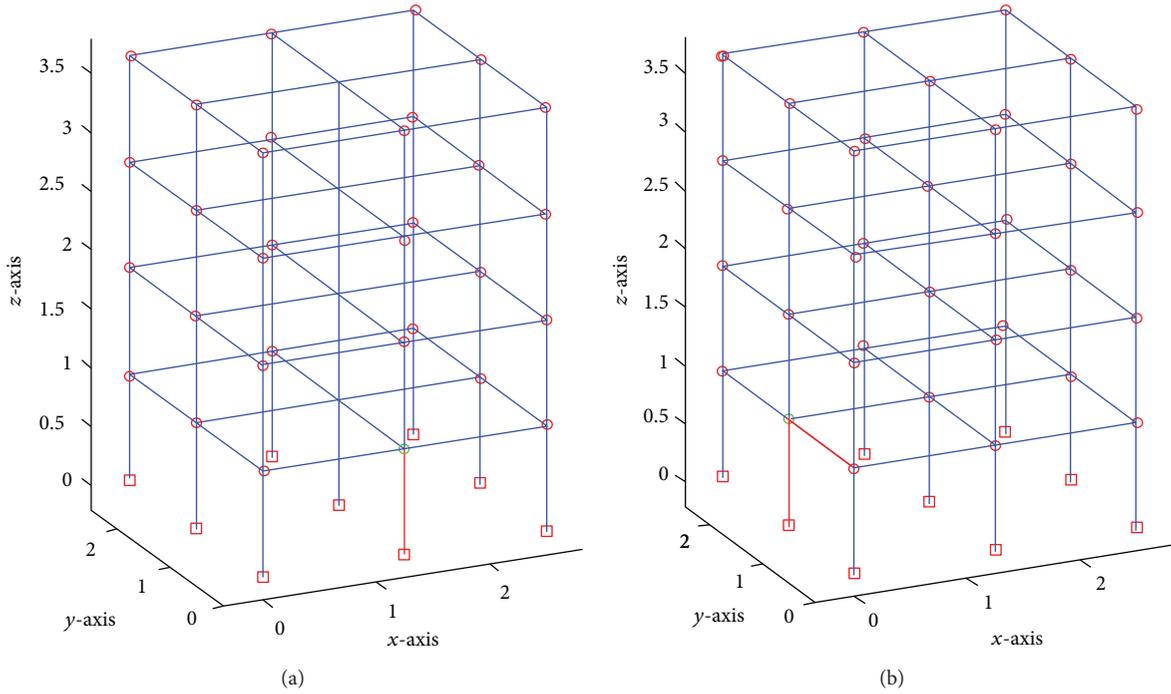


FIGURE 11: Incomplete MSs damage scenarios: (a) scenario 3 and (b) scenario 4.

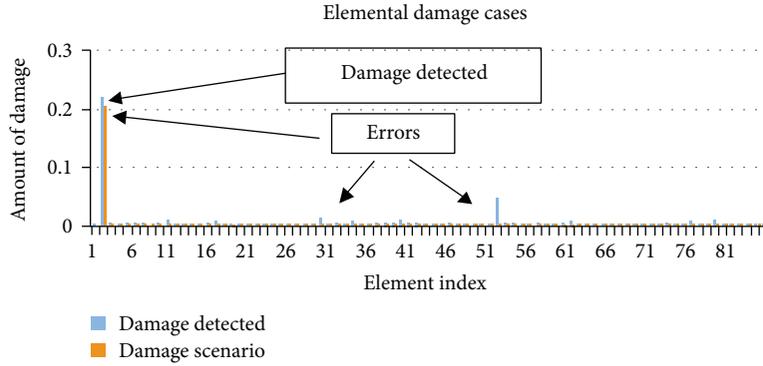


FIGURE 12: Damage locations in damage scenario 3.

damage scenarios are considered to study the incomplete MSs problem. Scenario 3 in which 20% damage is assigned to element 2 (red beam) and MS information are unknown at node 12 (green dot) as well as scenario 4 in which 20% damage is assigned to elements 4 and 16 (red beam) and MS information are missing at node 13 (green dot), as they are observed in Figure 11. Considering the use of the first three mode shapes, the results of damage tracking when complete MSs cannot be defined are shown in Figures 12 and 13. The results show that the proposed algorithm has proven good ability in detecting structural damage even when applied with unavailability of complete MSs. Nevertheless, it has shown inconsistent performance when executed for several times and needed almost the double computational time.

Finally, it is clear that the proposed framework has evidenced good performance when applied to damage identification in 3D structures. Moreover, its accuracy and relative

consistency, together with its the ability to track superior solutions in a single run, make the suggested MOLFPSO/TOPSIS with FE model updating paradigm suitable and reliable when damage localization in 3D structures is needed.

6. Conclusion

In this research, the problem of damage prognostic in three-dimensional structures implementing a novel MOEA incorporated with FE model updating was investigated. The novel algorithm called MOLFPSO/TOPSIS was designed to provide an efficient and reliable structural damage localization framework. The methodology included the use of TOPSIS as a MCDM technique to select a leading solution in each iteration within the multiobjective optimization, that is, the MOLFPSO. Proper selection of such a leader solution can highly and positively influence convergence to the Pareto

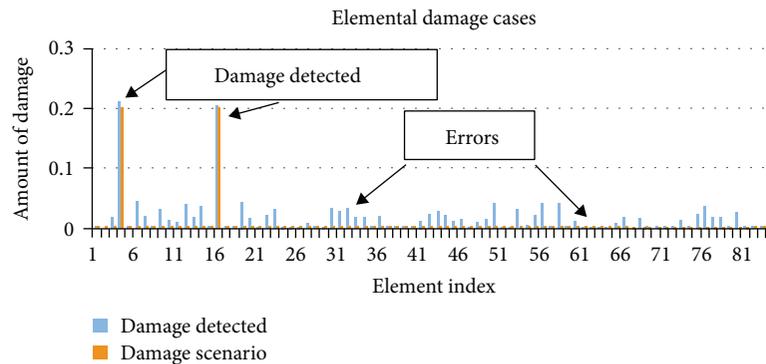


FIGURE 13: Damage locations in damage scenario 4.

front as well as saving postprocessing time when a best-compromised solution is required. To evaluate the performance of the proposed paradigm, a three-dimensional model was deemed a case study with two simulated damage scenarios. Furthermore, in each scenario, Gaussian noise was added to contaminate the simulated MSs to examine the reliability of the developed technique under polluted conditions. Finally, the incomplete MSs problem has been evaluated with relatively good performance. The proposed technique accomplished effective damage diagnosis in 3D structures with accuracy, reliability, and low computational time, even when tested under a noisy environment and with the absence of complete mode shapes.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors appreciate the financial help provided by the Key Program of the National Natural Science Foundation of China (Grant no. 11772115) and the Jiangsu Provincial Recruitment Program of Foreign Experts (Type B, Grant 172 no. JSB2017007).

References

- [1] W. Fan and P. Qiao, "Vibration-based damage identification methods: a review and comparative study," *Structural Health Monitoring: An International Journal*, vol. 10, no. 1, pp. 83–111, 2011.
- [2] Z. X. Yuan and K. P. Yu, "Finite element model updating of damped structures using vibration test data under base excitation," *Journal of Sound and Vibration*, vol. 340, pp. 303–316, 2015.
- [3] J. G. Chen and O. Büyüköztürk, "A symmetry measure for damage detection with mode shapes," *Journal of Sound and Vibration*, vol. 408, pp. 123–137, 2017.
- [4] F. Seguel and V. Meruane, "Damage assessment in a sandwich panel based on full-field vibration measurements," *Journal of Sound and Vibration*, vol. 417, pp. 1–18, 2018.
- [5] R. M. Lin, "Function-weighted frequency response function sensitivity method for analytical model updating," *Journal of Sound and Vibration*, vol. 403, pp. 59–74, 2017.
- [6] F. Adel, S. Shokrollahi, M. Jamal-Omidi, and H. Ahmadian, "A model updating method for hybrid composite/aluminum bolted joints using modal test data," *Journal of Sound and Vibration*, vol. 396, pp. 172–185, 2017.
- [7] M. Pedram, A. Esfandiari, and M. R. Khedmati, "Damage detection by a FE model updating method using power spectral density: numerical and experimental investigation," *Journal of Sound and Vibration*, vol. 397, pp. 51–76, 2017.
- [8] G.-H. Kim and Y.-S. Park, "An improved updating parameter selection method and finite element model update using multi-objective optimisation technique," *Mechanical Systems and Signal Processing*, vol. 18, no. 1, pp. 59–78, 2004.
- [9] R. Perera and A. Ruiz, "A multistage FE updating procedure for damage identification in large-scale structures based on multiobjective evolutionary optimization," *Mechanical Systems and Signal Processing*, vol. 22, no. 4, pp. 970–991, 2008.
- [10] C. J. Carrasco, R. Osegueda, C. Ferregut, and M. Grygier, "Damage localization in a space truss model using modal strain energy," in *SPIE The International Society for Optical Engineering*, pp. 1786–1792, SPIE International Society For Optical, Orlando, FL, USA, 1997.
- [11] Z. Y. Shi, S. S. Law, and L. M. Zhang, "Structural damage localization from modal strain energy change," *Journal of Sound and Vibration*, vol. 218, no. 5, pp. 825–844, 1998.
- [12] Z. Y. Shi, S. S. Law, and L. M. Zhang, "Structural damage detection from modal strain energy change," *Journal of Engineering Mechanics*, vol. 126, no. 12, pp. 1216–1223, 2000.
- [13] W.-J. Yan, T.-L. Huang, and W.-X. Ren, "Damage detection method based on element modal strain energy sensitivity," *Advances in Structural Engineering*, vol. 13, no. 6, pp. 1075–1088, 2010.
- [14] Y.-J. Cha and O. Buyukozturk, "Structural damage detection using modal strain energy and hybrid multiobjective optimization," *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, no. 5, pp. 347–358, 2015.
- [15] Y. Cha and O. Buyukozturk, "Modal strain energy based damage detection using multi-objective optimization," in *Structural Health Monitoring*, vol. 5, pp. 125–133, Springer, Cham, 2014.
- [16] G. Beruvides, F. Castaño, R. E. Haber, R. Quiza, and A. Villalonga, "Coping with complexity when predicting surface roughness in milling processes: hybrid incremental model with optimal parametrization," *Complexity*, vol. 2017, Article ID 7317254, 11 pages, 2017.
- [17] J. Ma, J. Wu, and X. Wang, "Fault diagnosis method of check valve based on multikernel cost-sensitive extreme learning

- machine,” *Complexity*, vol. 2017, Article ID 8395252, 19 pages, 2017.
- [18] I. Marović, I. Sušan, and N. Ožanić, “Development of ANN model for wind speed prediction as a support for early warning system,” *Complexity*, vol. 2017, Article ID 3418145, 10 pages, 2017.
- [19] H. Wang, Z. Zhao, Z. Guo, Z. Wang, and G. Xu, “An improved clustering method for detection system of public security events based on genetic algorithm and semisupervised learning,” *Complexity*, vol. 2017, Article ID 8130961, 10 pages, 2017.
- [20] C. S. Chin, J. Si, A. S. Clare, and M. Ma, “Intelligent image recognition system for marine fouling using Softmax transfer learning and deep convolutional neural networks,” *Complexity*, vol. 2017, Article ID 5730419, 9 pages, 2017.
- [21] L.-Y. Dai, C.-M. Feng, J.-X. Liu, C.-H. Zheng, J. Yu, and M.-X. Hou, “Robust nonnegative matrix factorization via joint graph Laplacian and discriminative information for identifying differentially expressed genes,” *Complexity*, vol. 2017, Article ID 4216797, 11 pages, 2017.
- [22] L. Yuan, C.-A. Yuan, and D.-S. Huang, “FAACOSE: a fast adaptive ant colony optimization algorithm for detecting SNP epistasis,” *Complexity*, vol. 2017, Article ID 5024867, 10 pages, 2017.
- [23] Y. Cheng, L. Tao, and C. Yang, “Lithium-ion battery capacity estimation: a method based on visual cognition,” *Complexity*, vol. 2017, Article ID 6342170, 13 pages, 2017.
- [24] X. Yong and H. Hong, “A genetic algorithm for structural damage detection based on vibration data,” in *IMAC XIX - 19th International Modal Analysis Conference*, pp. 1381–1387, Kissimmee, FL, USA, 2001.
- [25] J. H. Chou and J. Ghaboussi, “Genetic algorithm in structural damage detection,” *Computers & Structures*, vol. 79, no. 14, pp. 1335–1353, 2001.
- [26] F. T. K. Au, Y. S. Cheng, L. G. Tham, and Z. Z. Bai, “Structural damage detection based on a micro-genetic algorithm using incomplete and noisy modal test data,” *Journal of Sound and Vibration*, vol. 259, no. 5, pp. 1081–1094, 2003.
- [27] A. M. Raich and J. Ghaboussi, “Evolving structural design solutions using an implicit redundant genetic algorithm,” *Structural and Multidisciplinary Optimization*, vol. 20, no. 3, pp. 222–231, 2000.
- [28] R. S. He and S. F. Hwang, “Damage detection by an adaptive real-parameter simulated annealing genetic algorithm,” *Computers & Structures*, vol. 84, no. 31–32, pp. 2231–2243, 2006.
- [29] M. I. Friswell, J. E. T. Penny, and S. D. Garvey, “A combined genetic and eigensensitivity algorithm for the location of damage in structures,” *Computers & Structures*, vol. 69, no. 5, pp. 547–556, 1998.
- [30] Z. H. Ding, M. Huang, and Z. R. Lu, “Structural damage detection using artificial bee colony algorithm with hybrid search strategy,” *Swarm and Evolutionary Computation*, vol. 28, pp. 1–13, 2016.
- [31] X. Qian, M. Cao, Z. Su, and J. Chen, “A hybrid particle swarm optimization (PSO)-simplex algorithm for damage identification of delaminated beams,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 607418, 11 pages, 2012.
- [32] F. Kang, J.-j. Li, and Q. Xu, “Damage detection based on improved particle swarm optimization using vibration data,” *Applied Soft Computing*, vol. 12, no. 8, pp. 2329–2335, 2012.
- [33] H. Gökdag and A. R. Yildiz, “Structural damage detection using modal parameters and particle swarm optimization,” *Materials Testing*, vol. 54, no. 6, pp. 416–420, 2012.
- [34] J. J. Zhu, M. Huang, and Z. R. Lu, “Bird mating optimizer for structural damage detection using a hybrid objective function,” *Swarm and Evolutionary Computation*, vol. 35, pp. 41–52, 2017.
- [35] S. M. Seyedpoor, “A two stage method for structural damage detection using a modal strain energy based index and particle swarm optimization,” *International Journal of Non-Linear Mechanics*, vol. 47, no. 1, pp. 1–8, 2012.
- [36] S. M. Seyedpoor, S. Shahbandeh, and O. Yazdanpanah, “An efficient method for structural damage detection using a differential evolution algorithm-based optimisation approach,” *Civil Engineering and Environmental Systems*, vol. 32, no. 3, pp. 230–250, 2015.
- [37] H. Liu, K. Xin, and Q. Qi, “Study of structural damage detection with multi-objective function genetic algorithms,” *Procedia Engineering*, vol. 12, pp. 80–86, 2011.
- [38] S. Jung, S.-Y. Ok, and J. Song, “Robust structural damage identification based on multi-objective optimization,” *International Journal for Numerical Methods in Engineering*, vol. 81, no. 6, pp. 786–804, 2010.
- [39] F. Shabbir and P. Omenzetter, “Application of multi-objective optimization to structural damage estimation via model updating,” in *Proceedings of SPIE 8348, Health Monitoring of Structural and Biological Systems*, pp. 1–10, San Diego, CA, USA, 2012.
- [40] R. Perera, E. Sevillano, and A. Ruiz, “Multiobjective structural damage identification in uncertain environments,” in *International Conference on Structural Nonlinear Dynamics and Diagnosis*, pp. 1–5, Marrakech, Morocco, 2012.
- [41] P. da Silva Lopes Alexandrino, A. B. Jorge, and S. S. da Cunha Júnior, “Detection of holes in a plate using multiobjective optimization and multicriteria decision making,” in *22nd International Congress of Mechanical Engineering (COBEM)*, pp. 2890–2900, Ribeirão Preto, SP, Brazil, 2013.
- [42] R. Farokhzad, B. Mohebi, G. Ghodrati Amiri, and M. G. Ashtiany, “Detecting structural damage in Timoshenko beams based on optimization via simulation (OVS),” *Journal of Vibroengineering*, vol. 18, no. 8, pp. 5074–5095, 2016.
- [43] S. Jung, S.-Y. Ok, and J. Song, “Structural damage identification based on multi-objective optimization,” in *Structural Dynamics*, pp. 1239–1244, Springer, New York, NY, USA, 2010.
- [44] J. Zhou, A. Mita, and L. Rongshuai, “Multi-objective optimization strategies for damage detection using cloud model theory,” in *Proceedings of SPIE 8348, Health Monitoring of Structural and Biological Systems 2012*, pp. 1–7, San Diego, CA, USA, 2012.
- [45] M. Wang and J. C. Brigham, “Assessment of multi-objective optimization for nondestructive evaluation of damage in structural components,” *Journal of Intelligent Material Systems and Structures*, vol. 25, no. 9, pp. 1082–1096, 2014.
- [46] R. Allemang and D. Brown, “A correlation coefficient for modal vector analysis,” in *The First International Modal Analysis Conference*, pp. 110–116, Orlando, FL, USA, 1982.
- [47] L. R. Barroso and R. Rodriguez, “Damage detection utilizing the damage index method to a benchmark structure,” *Journal of Engineering Mechanics*, vol. 130, no. 2, pp. 142–151, 2004.

- [48] C. C. Ciang, J.-R. Lee, and H.-J. Bang, "Structural health monitoring for a wind turbine system: a review of damage detection methods," *Measurement Science and Technology*, vol. 19, no. 12, article 122001, 2008.
- [49] M. Ratcliffe and N. Lieven, "An improved method for parameter selection in finite element model updating," *The Aeronautical Journal*, vol. 102, no. 1016, pp. 321–329, 1998.
- [50] P. Avitabile, *Model Updating - Endless Possibilities*, Society for Experimental Mechanics, Bethel, CT, USA, 2000.
- [51] M. Friswell and J. Mottershead, *Finite Element Model Updating in Structural Dynamics*, Springer, Dordrecht, Netherlands, 1995.
- [52] N. F. Alkayem, M. Cao, Y. Zhang, M. Bayat, and Z. Su, "Structural damage detection using finite element model updating with evolutionary algorithms: a survey," *Neural Computing and Applications*, vol. 30, no. 2, pp. 389–411, 2018.
- [53] D. Savic, "Single-objective vs. multiobjective optimisation for integrated decision support," in *Proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society*, pp. 1–12, Lugano, Switzerland, 2002.
- [54] X.-S. Yang, "Firefly algorithm, Lévy flights and global optimization," in *Research and Development in Intelligent Systems XXVI*, pp. 209–218, Springer, London, UK, 2010.
- [55] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, pp. 210–214, Coimbatore, India, 2009.
- [56] C. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, New York, NY, USA, 1981.
- [57] C. Coello and M. Lechuga, "MOPSO: a proposal for multiple objective particle swarm optimization," in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, pp. 1051–1056, Honolulu, HI, USA, 2002.
- [58] C. A. Coello Coello and M. Reyes-Sierra, "Multi-objective particle swarm optimizers: a survey of the state-of-the-art," *International Journal of Computational Intelligence Research*, vol. 2, no. 3, 2006.
- [59] Y. J. Cha, A. Raich, L. Barroso, and A. Agrawal, "Optimal placement of active control devices and sensors in frame structures using multi-objective genetic algorithms," *Structural Control and Health Monitoring*, vol. 20, no. 1, pp. 16–44, 2013.
- [60] N. F. Alkayem and M. Cao, "Damage identification in three-dimensional structures using single-objective evolutionary algorithms and finite element model updating: evaluation and comparison," *Engineering Optimization*, pp. 1–20, 2018.
- [61] D. Bernal, S. J. Dyke, H.-F. Lam, and J. Beck, "Phase II of the ASCE Benchmark Study on SHM," in *Proceedings of the 15th ASCE Engineering Mechanics Conference*, pp. 1–5, New York, NY, USA, 2002.
- [62] E. A. Johnson, H. F. Lam, L. S. Katafygiotis, and J. L. Beck, "Phase I IASC-ASCE structural health monitoring benchmark problem using simulated data," *Journal of Engineering Mechanics*, vol. 130, no. 1, pp. 3–15, 2004.

Research Article

Fault Diagnosis of Electromechanical Actuator Based on VMD Multifractal Detrended Fluctuation Analysis and PNN

Hongmei Liu ¹, Jiayao Jing ¹ and Jian Ma ^{1,2}

¹School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

²Science & Technology on Reliability & Environmental Engineering Laboratory, Beijing 100191, China

Correspondence should be addressed to Jian Ma; 09977@buaa.edu.cn

Received 10 February 2018; Revised 10 June 2018; Accepted 13 June 2018; Published 1 August 2018

Academic Editor: Minvydas Ragulskis

Copyright © 2018 Hongmei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Electromechanical actuators (EMAs) are more and more widely used as actuation devices in flight control system of aircrafts and helicopters. The reliability of EMAs is vital because it will cause serious accidents if the malfunction of EMAs occurs, so it is significant to detect and diagnose the fault of EMAs timely. However, EMAs often run under variable conditions in realistic environment, and the vibration signals of EMAs are nonlinear and nonstationary, which make it difficult to effectively achieve fault diagnosis. This paper proposed a fault diagnosis method of electromechanical actuators based on variational mode decomposition (VMD) multifractal detrended fluctuation analysis (MFDFA) and probabilistic neural network (PNN). First, the vibration signals were decomposed by VMD into a number of intrinsic mode functions (IMFs). Second, the multifractal features hidden in IMFs were extracted by using MFDFA, and the generalized Hurst exponents were selected as the feature vectors. Then, the principal component analysis (PCA) was introduced to realize dimension reduction of the extracted feature vectors. Finally, the probabilistic neural network (PNN) was utilized to classify the fault modes. The experimental results show that this method can effectively achieve the fault diagnosis of EMAs even under different working conditions. Simultaneously, the diagnosis performance of the proposed method in this paper has an advantage over that of EMD-MFDFA method for feature extraction.

1. Introduction

Although most aircrafts and helicopters still adopt hydraulic actuation systems, electromechanical actuators have increasingly been applied as the key actuators for flight control systems of advanced aircrafts and helicopters in recent years. The main reason is that electromechanical actuator (EMA) has more superiorities in terms of reliability, economy, and other aspects than traditional hydraulic actuator. However, aircrafts and helicopters often perform mission under variable complex environments, and it will cause serious consequences when the faults of EMAs appear. Therefore, fault detection and diagnosis of EMAs in various working conditions play a vital role in the normal operation of aircrafts and helicopters. More and more researches have been done about the function of EMAs, but few are about fault. Consequently, it is very meaningful to carry out the research on fault diagnosis algorithms of EMAs under variable working conditions.

NASA Ames Research Center's researchers conducted failure mode and effect analysis of EMAs through extensive literature investigation, and the main fault modes of EMAs were obtained [1]. The researchers built the flyable electro-mechanical actuator (FLEA) test-bed, so that the normal data and fault data of EMA can be obtained through a large number of experiments [2]. A method based on neural network was proposed to realize the diagnosis for critical failure modes of EMAs [3]. Narasimhan et al. implemented the degeneration trend prognostics of EMAs by using the Gaussian process regression algorithm [4]. A method based on WPD-STFT time-frequency entropy and PNN was presented by Jing et al., which achieved the accurate diagnosis of EMAs [5]. At present, there are relatively few researches on fault diagnosis methods of EMAs under variable working conditions at home and abroad.

The vibration signal of rotating machinery contains abundant information about the running state of the equipment. And extracting the fault feature which represents the

fault information of the equipment is the most important step in fault diagnosis. However, vibration signal generally has the characteristics of nonlinear and nonstationary, and there are external disturbances such as noise, so that extracting features from vibration signal is the key problem for researchers. The commonly used methods for processing vibration signal to extract fault features include short-time Fourier transform (STFT), wavelet transform (WT), empirical mode decomposition (EMD), and local mean decomposition (LMD). STFT can depict signal in both time domain and frequency domain at the same time and can reflect the time-varying characteristics of the signal frequency spectrum. But the window function of STFT is fixed, so it is not suitable to analyze strong time-varying and nonstationary signal [6]. WT can realize the multiresolution analysis of signal, but its resolution in the frequency domain is not adjustable at the same scale, and it needs to preselect the basis function according to the characteristic of the signal [7]. EMD decomposes the signal into a finite number of single-component signals which are called intrinsic mode functions (IMFs). It has great potential for analyzing the nonlinear and nonstationary signal. However, EMD has a series of problems such as end effects, modal confusion, over-envelope and under-envelope, negative frequency, and lacking theoretical basis [8]. LMD is an adaptive time-frequency analysis method which is proposed on the basis of EMD. It can decompose the complex signal into several product functions (PFs). However, LMD also has the problem of end effects, modal confusion, and large amounts of calculation [9]. In addition, fault diagnosis methods based on various multidisciplinary algorithms have been studied in recent years. A rotating machinery fault diagnosis method combining bispectrum and image processing algorithm was proposed, and its validity was proved by experiments of hydraulic pump and centrifugal pump [10]. A method based on narrowband demodulation with frequency shift and spectrum edit was used to achieve the fault diagnosis of gears [11]. Variational mode decomposition (VMD) is a new signal processing method which has a different theoretical framework with EMD [12]. VMD transforms signal decomposition into non-recursive and variational mode decomposition problem which has theoretical foundation. It shows better noise robustness and can reduce the sampling effect and modal confusion.

Different from time-frequency analysis, fractal analysis can be used to reveal the fractal features of the signal, while fractal features can characterize the different operating states of a complex system. Therefore, fractal features can be utilized as fault features for fault diagnosis. Multifractal analysis can extract fractal features of different local scales, and researchers have applied classical multifractal theory to feature extraction of fault diagnosis in recent years. A method based on wavelet analysis and multifractal spectrum was applied to extract the fault features of hydropower unit [13]. And the multifractal spectrum was combined with PSO-SVM to achieve the fault diagnosis of gearbox [14]. However, the traditional multifractal theory can be easily disturbed by the trend of signal fluctuation and cannot reveal the multifractal characteristics hidden in nonstationary signal accurately. Thus, Liu et al. proposed a method

called multifractal detrended fluctuation analysis (MFDFA) combining multifractal (MF) with detrended fluctuation analysis (DFA), which can eliminate the influence of signal fluctuation and can further effectively extract the multifractal characteristics of nonstationary signal. MFDFA has been applied to the field of fault diagnosis for complex system. A method based on MFDFA and local characteristic-scale decomposition-Teager energy operator was proposed to realize the fault diagnosis of rolling bearing [15]. Tang et al. applied MFDFA into the fault diagnosis of nonlinear analog circuit [16].

A fault diagnosis method for EMA based on VMD-MFDFA and PNN is proposed in this paper. Firstly, the vibration signal of the accelerometer is collected. After pre-processing the vibration signal, a series of IMFs are obtained by using the VMD. Then, the multifractal features of IMFs are calculated by MFDFA, and the fault feature vectors are acquired by reducing the dimension with PCA. Finally, PNN model is trained to classify the fault modes.

2. Feature Extraction Method Based on VMD and MFDFA

The vibration signal of the EMA has the characteristics of nonlinear, nonstationary, and strong time-varying. In this paper, the vibration signal is decomposed by VMD, and the feature vectors are extracted by MFDFA to characterize the operating state of the EMA.

2.1. A Description of Variational Mode Decomposition (VMD). The VMD algorithm can obtain the optimal solution of the constrained variational problem and determine different central frequencies and bandwidths through iteration. The intrinsic mode functions (IMFs) of different frequencies are obtained by nonrecursive decomposition [17]. The implementation of VMD is divided into two parts: the construction of variational problem and the solution of variational problem [18].

The first part is the construction of variational problem. This time-frequency analysis method assumes that the multi-component signal f consisted of k intrinsic mode functions u_k with limited bandwidth, and the central frequency of each intrinsic mode function corresponds to ω_k .

The analytic signal of each intrinsic mode function is obtained by Hilbert demodulation as the following formula:

$$\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t). \quad (1)$$

A central frequency is estimated as $e^{-j\omega_k t}$ for each analytic signal, and the frequency spectrum of each IMF is modulated to the fundamental frequency band:

$$\left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t}. \quad (2)$$

The square norm L^2 of the above analytic signal gradient is calculated, and the bandwidth of each IMF is estimated.

Then, the constrained variational problem is obtained as the following formula:

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} & \sum_k u_k(t) = f, \end{aligned} \quad (3)$$

where $\{u_k\} = \{u_1, \dots, u_k\}$ represents one of the k intrinsic mode functions obtained by decomposition and $\{\omega_k\} = \{\omega_1, \dots, \omega_k\}$ represents the central frequency of each intrinsic mode function.

The second part is the solution of variational problem. In order to obtain the optimal solution of the variational model, Lagrange multiplication operator $\lambda(t)$ and quadratic penalty factor α need to be introduced to change the constrained variational problem into nonconstrained variational problem. The transformed Lagrange expression is

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_k u_k \right\|_2^2 + \left\langle \lambda(t) - \sum_k u_k(t) \right\rangle. \end{aligned} \quad (4)$$

The saddle point of formula (4) is obtained through iteratively updating u_k^{n+1} , ω_k^{n+1} , and λ^{n+1} by using the alternate direction method of multipliers (ADMM).

The update method of u_k^{n+1} is

$$u_k^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq k} u_i(\omega) + (\lambda(\omega)/2)}{1 + 2\alpha(\omega - \omega_k)^2}. \quad (5)$$

The update method of ω_k^{n+1} is

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k(\omega)|^2 d\omega}{\int_0^\infty |u_k(\omega)|^2 d\omega}. \quad (6)$$

The update method of λ^{n+1} is

$$\lambda^{n+1}(\omega) = \lambda^n(\omega) + \tau \left[f(\omega) - \sum_k u_k^{n+1}(\omega) \right]. \quad (7)$$

The real part after the Fourier transform of $\{u_k^{n+1}(\omega)\}$ combining formula (5) and formula (6) is the intrinsic mode functions $\{u_k(\omega)\}$.

The specific steps of VMD can be described as follows [19]:

- (1) Initialize $\{u_k^1\}$, $\{\omega_k^1\}$, $\{\lambda^1\}$, and n .
- (2) Set $n = n + 1$ and begin the circulation.
- (3) Update u_k and ω_k according to formula (5) and formula (6).
- (4) Set $K = K + 1$, and repeat step (3) until $K = k$.

(5) Update λ according to formula (7).

(6) Repeat step (3) to step (5), until iteration stop condition $\sum_k \|u_k^{n+1} - u_k^n\|_2^2 / \|u_k^n\|_2^2 < e$ is reached.

In the process of decomposition by VMD, the central frequency and bandwidth of each IMF are constantly updated to realize the adaptive decomposition of signal.

2.2. A Description of Multifractal Detrended Fluctuation Analysis (MFDFA). Multifractal detrended fluctuation analysis can effectively eliminate the effect of signal fluctuation trend and can accurately extract the implied multifractal features of nonlinear signal [20].

The steps of MFDFA can be described as follows [21]:

- (1) For time series x_k , construct cumulative deviation $Y(i)$ of the sequence to the mean:

$$\begin{aligned} Y(i) & \equiv \sum_{k=1}^i |x_k - \langle x \rangle|, \quad i = 1, \dots, N, \\ \langle x \rangle & = \frac{1}{N} \sum_{k=1}^N x_k. \end{aligned} \quad (8)$$

- (2) The new sequence $Y(i)$ is divided into nonoverlapping m subsequences with a fixed scale s :

$$m = \text{int} \left(\frac{N}{s} \right). \quad (9)$$

Then, the sequence is divided into m segments by the same scale from the reverse direction of the sequence, and $2m$ subsequences can be obtained.

- (3) Fit the polynomial trend of each subsequence by using the least square method, and calculate the variance as follows:

$$\begin{aligned} F^2(s, \nu) & \equiv \frac{1}{s} \sum_{i=1}^s \{ Y[(\nu-1)s + i] - y_\nu(i) \}^2, \quad \nu = 1, 2, \dots, m, \\ F^2(s, \nu) & \equiv \frac{1}{s} \sum_{i=1}^s \{ Y[N - (\nu-m)s + i] - y_\nu(i) \}^2, \\ & \nu = m+1, m+2, \dots, 2m, \end{aligned} \quad (10)$$

where $y_\nu(i)$ is the fitting polynomial of the ν subsequence.

- (4) Calculate the mean value of the q -order fluctuation function:

$$F_q(s) \equiv \left\{ \frac{1}{2m} \sum_{\nu=1}^{2m} [F^2(s, \nu)]^{q/2} \right\}^{1/q}, \quad (11)$$

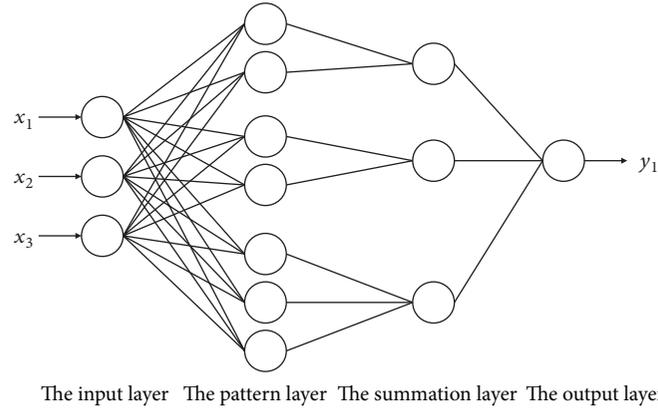


FIGURE 1: Basic structure of probabilistic neural network.

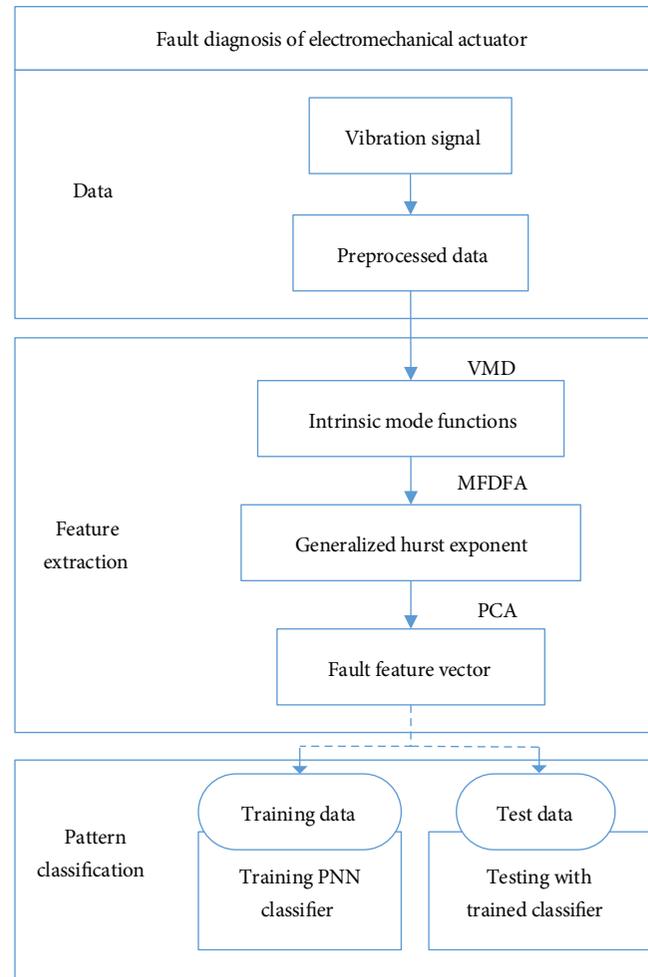


FIGURE 2: The fault diagnosis procedure of EMA.

where different values of q represent different degrees of fluctuation. And when $q=2$, MF DFA degenerates to DFA.

- (5) Change the length of the subsequence s and repeat steps (2) to (4).

- (6) $F_q(s)$ is the function of the length of the subsequence s and the fractal order q and has the following power-law relation with the scale s :

$$F_q(s) \sim s^{H(q)}, \quad (12)$$

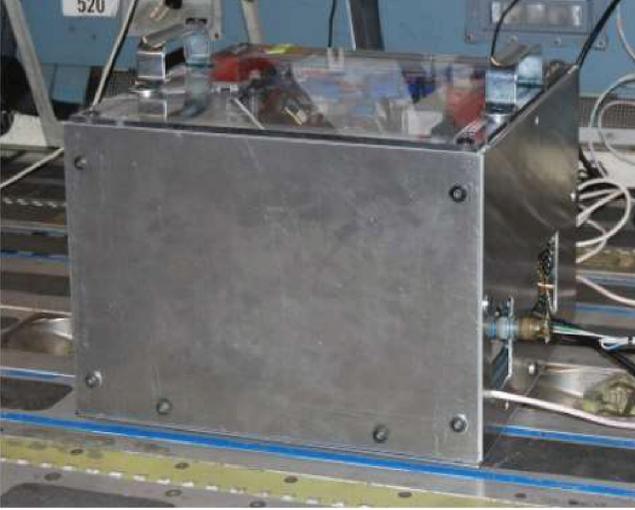


FIGURE 3: The FLEA test-bed.

where $F_q(s)$ is the mean value of q -order fluctuation function and H_q is the generalized Hurst exponent.

If x_k is a monofractal time series, H_q is a constant, and if x_k is a multifractal time series, H_q is the function of order q . The different order corresponds to the different generalized Hurst exponent.

The generalized Hurst exponent can describe the influence of the past time series on the present and the later time series, and the influences are different under different states of system.

Therefore, the generalized Hurst exponent can be used as the feature vector to describe the multifractal characteristics of the system and can characterize the different states of the system.

3. Fault Classification Based on PNN

The theoretical basis of probabilistic neural network (PNN) is Bayesian minimum risk criterion. PNN directly considers the probability characteristics of the sample space and takes the typical samples of the sample space as the nodes of the hidden layer. There is no need for training anymore once PNN is determined, and it is only necessary to append samples according to actual problems [22]. PNN has the advantages of short training time and global optimization and has great performance for classification.

The network structure of PNN is shown in Figure 1, which consists of the input layer, the pattern layer, the summation layer, and the output layer [23].

The input layer receives the values from the training data and transmits feature vectors to the network, and the number of neurons is equal to the dimension of the sample vectors.

The pattern layer calculates the matching regulation between the feature vectors and the different modes of the training data, and the number of neurons is equal to the

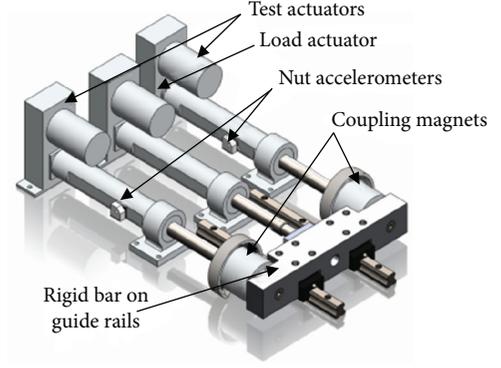


FIGURE 4: The model of FLEA test-bed.



FIGURE 5: The EMA in the FLEA test-bed.

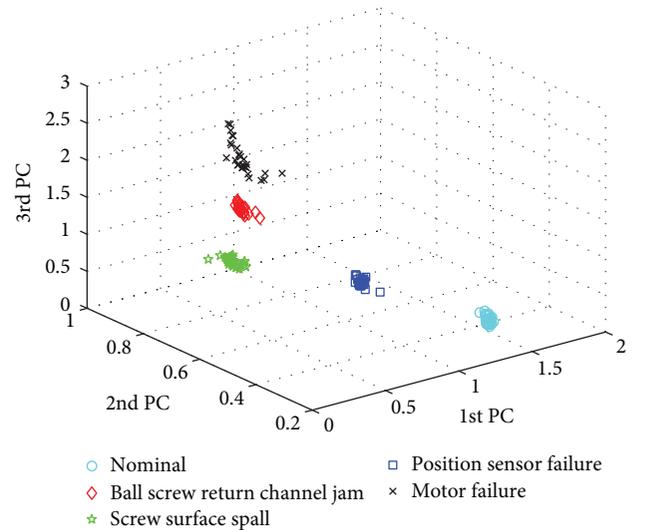


FIGURE 6: Clustering result of the fault features under condition 1.

sum of the training samples. The output of each unit of this layer is

$$f(X, W_i) = \exp \left[-\frac{(X - W_i)^T (X - W_i)}{2\delta^2} \right], \quad (13)$$

TABLE 1: The description of five working conditions.

Condition name	Position profile	Parameters	Load profile (load is in lbs)	Max. velocity (m/s)	Comments
Condition 1	Sine sweep	80 mm, 0.0625 Hz, 0.5 Hz	Constant at 0	0.08	Parameters represent amplitude and initial and final frequencies
Condition 2	Trapezoidal	40 mm, 22 s (1 + 1-second motion,	Constant at -10	0.04	Parameters represent amplitude and time period of trapezoidal wave
Condition 3	Trapezoidal	10 + 10-second hold)	Constant at 0	0.04	
Condition 4	Trapezoidal	40 mm, 21 s (0.5 + 0.5-second motion,	Constant at 0	0.08	
Condition 5	Trapezoidal	10 + 10-second hold)	Constant at 10	0.08	

where W_i is the weight of the input layer to the pattern layer and δ is the smoothing factor.

The summation layer adds the probability that each group of neurons belongs to a pattern and calculates the estimated probability density function of this pattern. A fault mode has only one summation layer neuron.

The output layer puts the mode with the greatest probability in the summation layer as the output. The output of a neuron with maximum probability is 1, and the output of other neurons is 0. The number of neurons in the output layer is equal to the number of the modes.

4. Fault Diagnosis Scheme of EMA Based on VMD-MFDFA and PNN

This paper presents a fault diagnosis method based on VMD-MFDFA and PNN for EMA, and the procedure of diagnosis scheme is shown in Figure 2.

- (1) Decompose the vibration signals of EMA into a series of IMFs by utilizing VMD.
- (2) Calculate the generalized Hurst exponents of IMFs as feature vectors by using MFDFA.
- (3) Reduce the dimension of the feature vectors to obtain the final fault features by utilizing PCA.
- (4) Classify the fault modes by using PNN.

5. Case Study

In order to verify the effectiveness of the fault diagnosis method proposed in this paper, we have conducted experiments by using the data from the FLEA test-bed of NASA database. The FLEA test-bed is shown in Figures 3 and 4, and the EMA in the test-bed is shown in Figure 5. The fault diagnosis experiments have been carried out in the following five states: the normal state, ball screw return channel jam, screw surface spall, motor failure, and position sensor failure. The vibration signals were acquired from the FLEA test-bed with a sampling frequency of 20 kHz for 30 seconds. And the data of each state has been divided into 29 groups (20,000 sampling points per group) to analyze conveniently. Moreover, for the sake of validating the adaptability to variable working conditions of the proposed method, the experiments have been carried out under five different conditions.

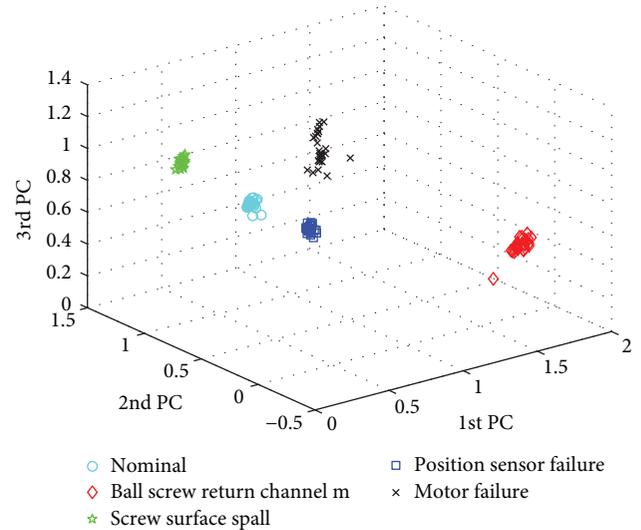


FIGURE 7: Clustering result of the fault features under condition 2.

5.1. Fault Feature Extraction Based on VMD and MFDFA. First, the collected raw vibration signals are preprocessed by normalizing.

Second, the preprocessed signals are decomposed by VMD. The VMD method needs to predetermine the number of modes k before decomposing signal. However, there will be modes with the same central frequency when the number of decomposition mode is more than 3. Therefore, the preprocessed signals are decomposed into three IMFs by using VMD in this paper.

Then, the MFDFA is applied to extract the multifractal features of decomposed IMFs. The generalized Hurst exponents are selected as fault features with the order $q = [-5, 0, 5]$. And in order to improve the accuracy of fault diagnosis, the 9-dimensional generalized Hurst exponents are reduced to 3-dimensional fault feature vectors by using PCA.

The clustering result of the final fault feature vectors is shown in Figure 6, which shows that the feature vectors acquired by the proposed method in this paper can characterize the state of EMA.

5.1.1. Fault Feature Extraction under Different Conditions. In practical application, EMA usually runs under variable working conditions, so it is of great significance to diagnose

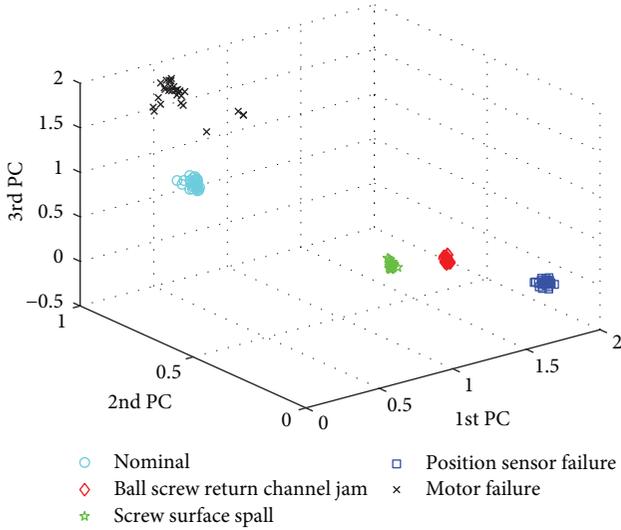


FIGURE 8: Clustering result of the fault features under condition 3.

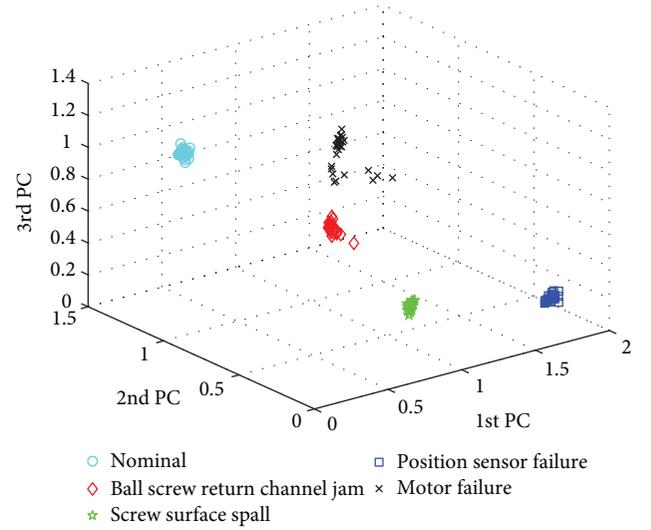


FIGURE 10: Clustering result of the fault features under condition 5.

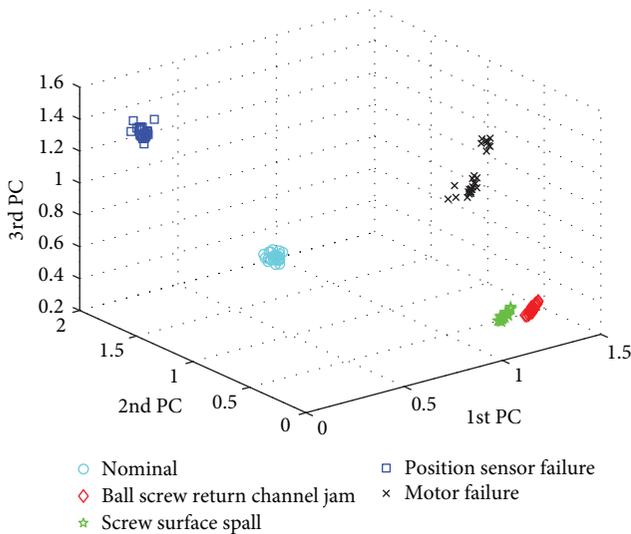


FIGURE 9: Clustering result of the fault features under condition 4.

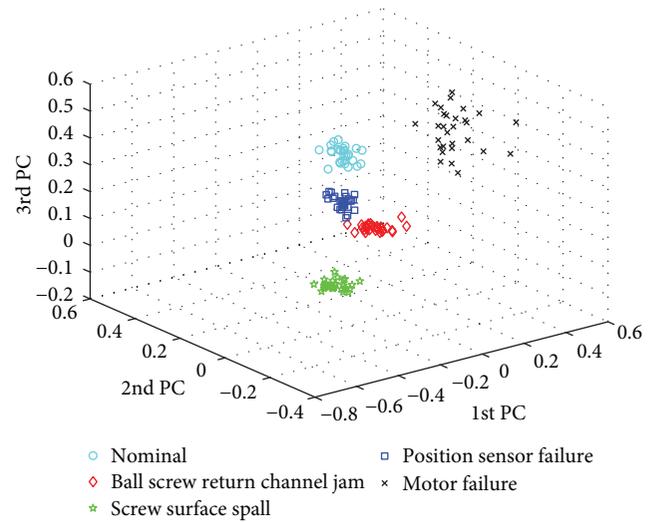


FIGURE 11: Clustering result of the fault features with EMD-MFDFA under condition 1.

accurately under different conditions. In order to prove the adaptability of the method to working conditions, the experiments are conducted under five different conditions as shown in Table 1.

The vibration signals of EMA are collected under the five working conditions, and the fault features are extracted by utilizing VMD-MFDFA-PCA method. Figures 7–10 show the clustering results of fault feature vectors under four conditions except condition 1.

It can be seen from the figures that the proposed method can accurately extract the fault features of EMA under different conditions and can adapt to the variable working conditions in the practical environment.

5.1.2. Comparison between the Proposed Method with EMD-MFDFA for Feature Extraction. In order to verify the excellent performance of the feature extraction method proposed

in this paper, the method based on EMD-MFDFA is applied to extract the fault feature vectors of EMA for comparison. This method combines the widely used empirical mode decomposition time-frequency analysis method with the MFDFA to extract the fault features of EMA.

Firstly, the original vibration signal is preprocessed and then decomposed by EMD into a series of IMFs with frequencies from high to low. Secondly, the first three IMFs containing the main fault information are selected. Then, the 9-dimensional generalized Hurst exponents are extracted by using MFDFA with the parameter $q = [-5, 0, 5]$. Finally, the PCA is applied to reduce the dimension of the 9-dimensional features to obtain a 3-dimensional generalized Hurst exponents as the ultimate fault feature vectors. The clustering result is shown in Figures 11–15.

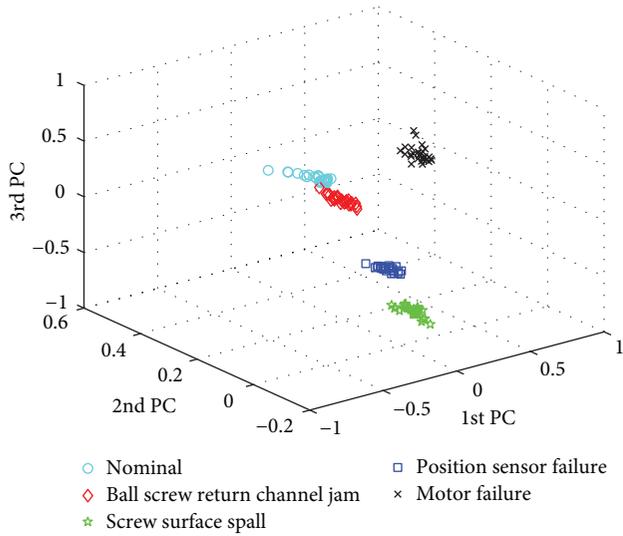


FIGURE 12: Clustering result of the fault features with EMD-MFDFA under condition 2.

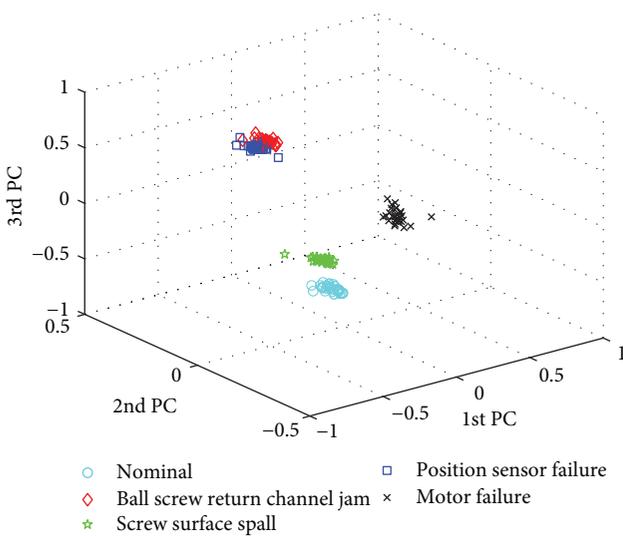


FIGURE 13: Clustering result of the fault features with EMD-MFDFA under condition 3.

It can be seen that scatter points of fault feature vectors by using feature extraction method based on EMD-MFDFA are relatively close, so that it is hard to clearly classify the fault modes. And it can also be proved that the method proposed in this paper has better performance for fault feature extraction.

5.2. Fault Classification Based on PNN. The PNN classifier model is trained to identify the fault modes of EMA under different working conditions. In each working condition, the 3-dimensional fault feature vectors of each fault mode are taken as the input of PNN, and the fault category labels are taken as the output of PNN. The first 15 sets of data of each fault mode are used as training data, and the training data is used to train the PNN classifier. Then, the fault labels

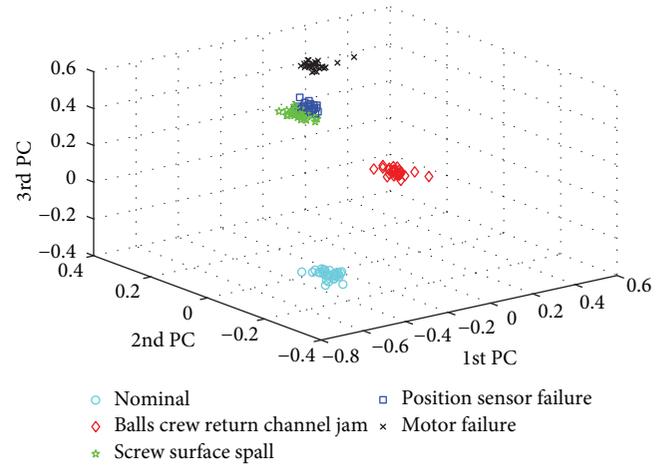


FIGURE 14: Clustering result of the fault features with EMD-MFDFA under condition 4.

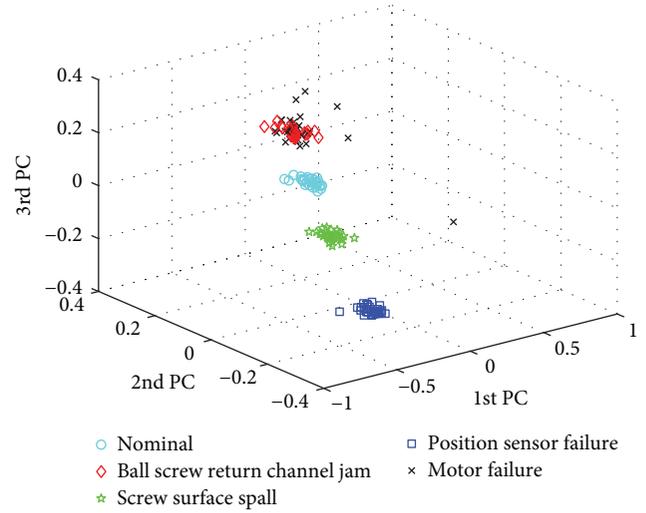


FIGURE 15: Clustering result of the fault features with EMD-MFDFA under condition 5.

of the last 14 sets of testing data are identified by the trained PNN classifier, and the final fault diagnosis result is obtained.

In the trained PNN classification model, 3 nodes in the input layer are determined according to the dimension of the feature vectors, 75 nodes in the pattern layer are determined according to the number of training samples, and 5 nodes in the summation layer and the output layer are determined according to the number of categories of the fault modes. And the smoothing factor of PNN is set to 1.0. The failure modes of the test samples correspond to the normal state, ball screw return channel jam, screw surface spall, position sensor failure, and motor failure, respectively, when the output of PNN is 1, 2, 3, 4, and 5.

The test results under five conditions are shown in Figures 16–20.

It can be seen from the figures that the diagnostic fault labels of the test samples are in high agreement with the

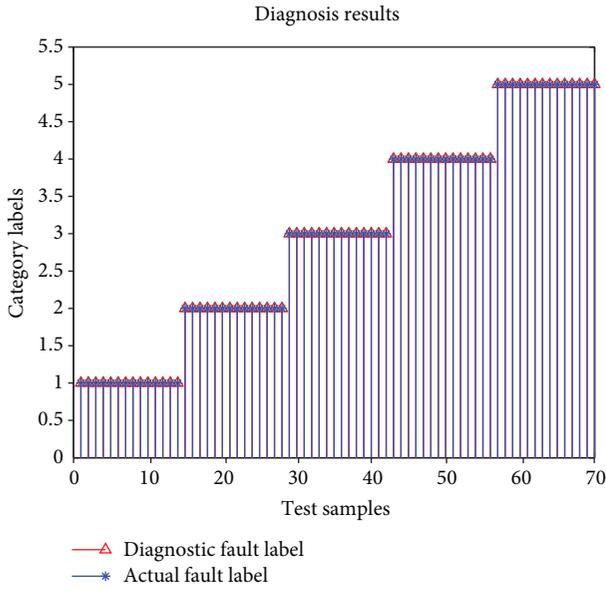


FIGURE 16: Classification result under condition 1.

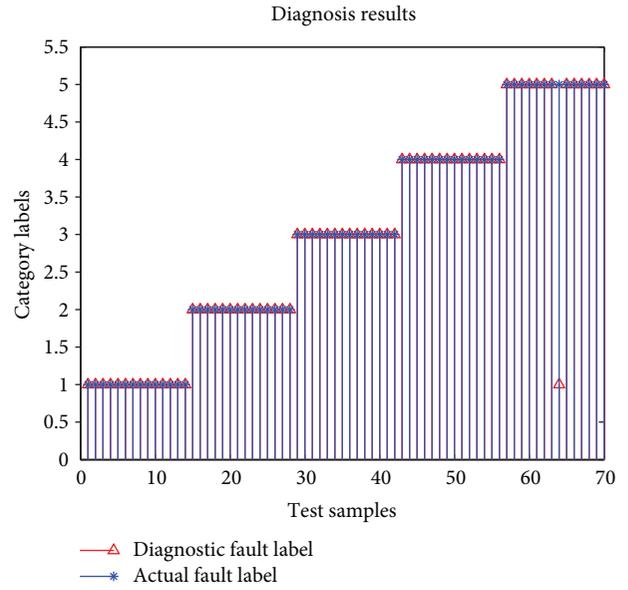


FIGURE 18: Classification result under condition 3.

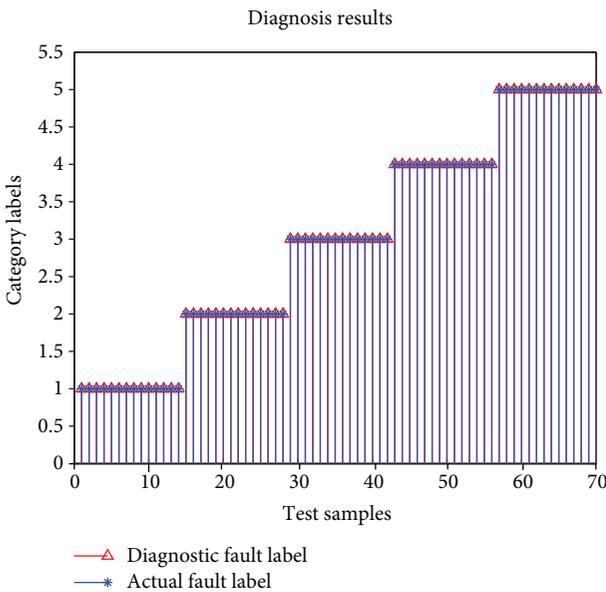


FIGURE 17: Classification result under condition 2.

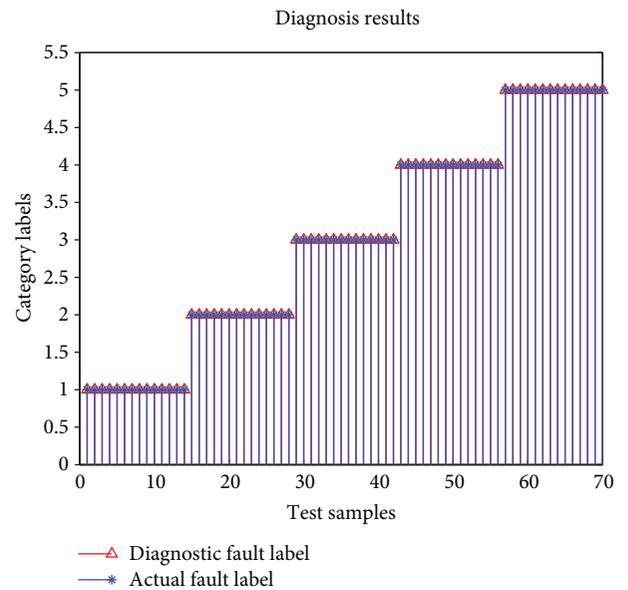


FIGURE 19: Classification result under condition 4.

actual fault labels. The diagnosis accuracy of condition 1 is 100%, the diagnosis accuracy of condition 2 is 100%, the diagnosis accuracy of condition 3 is 98.67%, the diagnosis accuracy of condition 4 is 100%, and the diagnosis accuracy of condition 5 is 100%. The diagnosis results indicate that the method proposed in this paper can accurately diagnose EMA and has great diagnostic performance.

6. Conclusion

EMA is more and more widely applied in the flight control system of aircrafts and helicopters, and it is of great importance to diagnose the fault of EMA in time for the safety of

aircrafts and helicopters. Thus, it is very meaningful to research the fault diagnosis of EMA. A fault diagnosis method based on VMD-MFDFA-PNN for EMA is presented in this study. Firstly, VMD is applied to decompose the vibration signal of EMA into the IMFs which contain the fault information. Secondly, the generalized Hurst exponents of IMFs are calculated as the fault features by MFDFA. Then, the PCA is utilized to reduce the dimension of the generalized Hurst exponents. Finally, the PNN model is trained to classify the fault modes. The results show that the method proposed in this paper can extract the features sensitive to the fault of EMA and can conduct accurate fault diagnosis under different working conditions. The great performance

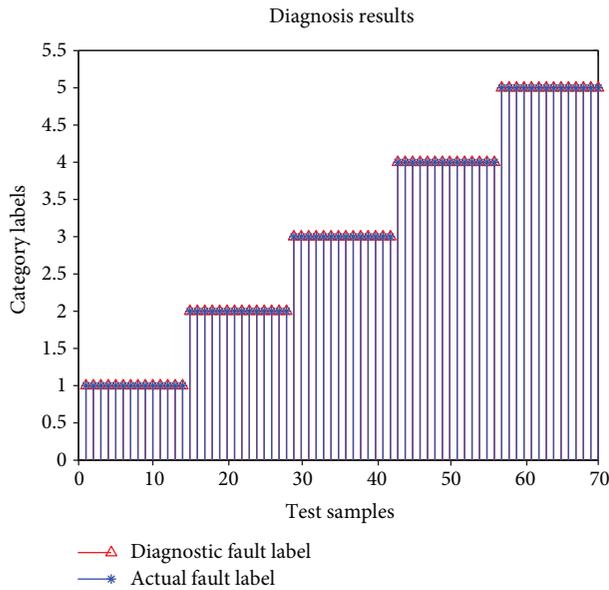


FIGURE 20: Classification result under condition 5.

of the proposed method is further validated by comparing with EMD-MFDFA-PCA.

However, the computational complexity of the proposed algorithm is relatively large. Therefore, in the case of limited computer resources, the calculation speed will be slightly slower. Future work will concentrate on two aspects. The first one is the study on improving the computational efficiency of the method proposed in this paper for the occasions with high real-time requirements. The second one is the study on health assessment and degeneration trend prognostics of EMA.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant nos. 51605014 and 51575021), the Technology Foundation Program of National Defense (Grant no. Z132013B002), the Fundamental Research Funds for the Central Universities (Grant no. YWF-16-BJ-J-18), and the Aviation Foundation (Grant nos. 20153351022, and 20163351018).

References

[1] E. Balaban, A. Saxena, S. Narasimhan et al., "Prognostic health-management system development for electromechanical

actuators," *Journal of Aerospace Information Systems*, vol. 12, no. 3, pp. 329–344, 2015.

[2] E. Balaban, A. Saxena, S. Narasimhan, I. Roychoudhury, K. F. Goebel, and M. T. Koopmans, "Airborne electro-mechanical actuator test stand for development of prognostic health management systems," *Journal of social welfare & Family Law*, vol. 32, no. 2, pp. 119–133, 2010.

[3] E. Balaban, P. Bansal, P. Stoelting, A. Saxena, K. F. Goebel, and S. Curran, "A diagnostic approach for electro-mechanical actuators in aerospace systems," in *2009 IEEE Aerospace conference*, pp. 1–13, Big Sky, MT, USA, March 2009.

[4] S. Narasimhan, I. Roychoudhury, E. Balaban, and A. Saxena, "Combining model-based and feature-driven diagnosis approaches—a case study on electromechanical actuators," in *21st International Workshop on Principles of Diagnosis, 2010*, pp. 1–8, Portland, OR, USA, October 2010.

[5] J. Jing, H. Liu, and C. Lu, "Fault diagnosis of electro-mechanical actuator based on WPD-STFT time-frequency entropy and PNN," *Vibroengineering PROCEDIA*, vol. 14, pp. 130–135, 2017.

[6] F. Jurado and J. R. Saenz, "Comparison between discrete STFT and wavelets for the analysis of power quality events," *Electric Power Systems Research*, vol. 62, no. 3, pp. 183–190, 2002.

[7] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 2015.

[8] J. Cheng, D. Yu, J. Tang, and Y. Yang, "Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery," *Shock and Vibration*, vol. 16, no. 1, 98 pages, 2013.

[9] H. Liu and M. Han, "A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings," *Mechanism and machine Theory*, vol. 75, no. 5, pp. 67–78, 2014.

[10] C. Lu, Y. Wang, M. Ragulskis, and Y. Cheng, "Fault diagnosis for rotating machinery: a method based on image processing," *PLoS One*, vol. 11, no. 10, article e0164111, 2016.

[11] Y. Guo, "Gear fault diagnosis based on narrowband demodulation with frequency shift and spectrum edit," *International Journal of Engineering and Technology Innovation*, vol. 6, no. 4, pp. 243–254, 2016.

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

[13] P. Guo, L. Sun, H. Li, and L. Xingqi, "Vibration fault diagnosis of hydropower unit based on multi-fractal spectrum and improved BP neural network," *Journal of Hydroelectric Engineering*, vol. 33, no. 3, pp. 299–305, 2014.

[14] S. Li, H. Pan, J. Zhang, and W. Zhao, "Fault diagnosis of gear-box based on multi-fractal and PSO-SVM," *Journal of Mechanical Transmission*, vol. 39, no. 2, pp. 132–136, 2015.

[15] H. Liu, X. Wang, and C. Lu, "Rolling bearing fault diagnosis based on LCD-TEO and multifractal detrended fluctuation analysis," *Mechanical Systems and Signal Processing*, vol. 60–61, pp. 273–288, 2015.

[16] J. Y. Tang, J. M. Chen, and C. Zhang, "Nonlinear analog circuit fault diagnosis based on MFDFA method," *Applied Mechanics and Materials*, vol. 263–266, no. 5, pp. 108–113, 2012.

[17] Y. Wang, R. Markert, J. Xiang, and W. Zheng, "Research on variational mode decomposition and its application in

- detecting rub-impact fault of the rotor system,” *Mechanical Systems and Signal Processing*, vol. 60-61, pp. 243–251, 2015.
- [18] Y. Liu, G. Yang, M. Li, and H. Yin, “Variational mode decomposition denoising combined the detrended fluctuation analysis,” *Signal Processing*, vol. 125, pp. 349–364, 2016.
- [19] C. Liu, W. U. Yingjie, and C. Zhen, “Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering,” *Proceedings of the Chinese Society of Electrical Engineering*, vol. 35, no. 13, pp. 3358–3365, 2015.
- [20] N. K. Vitanov, N. P. Hoffmann, and B. Wernitz, “Nonlinear time series analysis of vibration data from a friction brake: SSA, PCA, and MFDDFA,” *Chaos Solitons & Fractals*, vol. 69, no. 69, pp. 90–99, 2014.
- [21] S. Mukhopadhyay, N. K. Das, R. Kumar, D. Dash, A. Mitra, and P. K. Panigrahi, “Study of the dynamics of wind data fluctuations: a wavelet and MFDDFA based novel method,” in *Proceedings of 1st International Science & Technology Congress 2014*, pp. 541–545, India, October 2014.
- [22] S. Ding, X. H. Chang, and Q. H. Wu, “Application of probabilistic neural networks in fault diagnosis of three-phase induction motors,” *Applied Mechanics and Materials*, vol. 433–435, pp. 705–708, 2013.
- [23] K. Yang, G. Shan, and L. Zhao, “Application of wavelet packet analysis and probabilistic neural networks in fault diagnosis,” in *2006 6th World Congress on Intelligent Control and Automation*, vol. 1, pp. 4378–4381, Dalian, China, June 2006.

Research Article

Predicting the Remaining Useful Life of an Aircraft Engine Using a Stacked Sparse Autoencoder with Multilayer Self-Learning

Jian Ma ^{1,2} Hua Su ^{1,2} Wan-lin Zhao ^{1,2} and Bin Liu ^{1,2}

¹School of Reliability and Systems Engineering, Beihang University, Beijing, China

²Science & Technology on Reliability and Environmental Engineering Laboratory, Beijing, China

Correspondence should be addressed to Bin Liu; liubin6862@126.com

Received 6 January 2018; Revised 22 April 2018; Accepted 3 May 2018; Published 30 July 2018

Academic Editor: Minvydas Ragulskis

Copyright © 2018 Jian Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Because they are key components of aircraft, improving the safety, reliability and economy of engines is crucial. To ensure flight safety and reduce the cost of maintenance during aircraft engine operation, a prognostics and health management system that focuses on fault diagnosis, health assessment, and life prediction is introduced to solve the problems. Predicting the remaining useful life (RUL) is the most important information for making decisions about aircraft engine operation and maintenance, and it relies largely on the selection of performance degradation features. The choice of such features is highly significant, but there are some weaknesses in the current algorithm for RUL prediction, notably, the inability to obtain tendencies from the data. Especially with aircraft engines, extracting useful degradation features from multisensor data with complex correlations is a key technical problem that has hindered the implementation of degradation assessment. To solve these problems, deep learning has been proposed in recent years to exploit multiple layers of nonlinear information processing for unsupervised self-learning of features. This paper presents a deep learning approach to predict the RUL of an aircraft engine based on a stacked sparse autoencoder and logistic regression. The stacked sparse autoencoder is used to automatically extract performance degradation features from multiple sensors on the aircraft engine and to fuse multiple features through multilayer self-learning. Logistic regression is used to predict the remaining useful life. However, the hyperparameters of the deep learning, which significantly impact the feature extraction and prediction performance, are determined based on expert experience in most cases. The grid search method is introduced in this paper to optimize the hyperparameters of the proposed aircraft engine RUL prediction model. An application of this method of predicting the RUL of an aircraft engine with a benchmark dataset is employed to demonstrate the effectiveness of the proposed approach.

1. Introduction

Because they are core components of an aircraft, the failure of engines is often a major cause of major accidents and casualties [1]. Therefore, the safety and the reliability of engines are vital to the performance of aircraft. However, it is difficult to ensure their safety and reliability due to their complicated structures, and engine failure has arisen inevitably due to effects of aging, environment, and variable loading as the working time increases. For this reason, it is essential to detect underlying degradation, predict how soon an engine will fail effectively, implement maintenance promptly, and ultimately prevent catastrophic failure.

In the field of aircraft maintenance, traditional maintenance is either purely reactive (fixing or replacing an aircraft engine component after it fails) or blindly proactive (assuming a certain level of performance degradation with no input from the aircraft engine itself and maintaining the aircraft engine on a routine schedule whether maintenance is actually needed or not). Both scenarios are quite wasteful and inefficient, and neither is conducted in real time [2–5]. Given the scheduling of maintenance tasks based on fault diagnosis, performance degradation assessment and the predicted remaining useful life of the aircraft equipment and the need to prevent faults in advance, prognostics and health management (PHM) is gradually replacing these two maintenance

strategies. Prognostics, as the core of PHM, involves managing performance deterioration processes or faults in the aircraft engine and forecasts when components/systems of the engine will breakdown or when the performance will reach to an unacceptable level.

There are three main classes of RUL prediction methods: (1) data-driven methods, (2) physics model-based methods, and (3) methods that combine data-driven and physics model-based methods [6–9]. The data-driven methods use past condition monitoring data, the current health status of the system, and data on the degradation of similar systems. The methods based on physics models use system-specific mechanistic knowledge, failure regulation, and condition monitoring data to predict the RUL of a system or component. There are two main challenges in prognostics based on physics: (1) there is not enough physical knowledge to construct a physical degradation model and (2) the values of the physical model's parameters are difficult to determine exactly. Therefore, it is important to understand the failure mechanism of the system correctly, and experienced personnel are required for physics-based models [10, 11]. In addition, the peripheral environment during device operation (e.g., the temperature and humidity) and the operating conditions (e.g., the fan speed) may be used as inputs and constitute additional dimensions to be considered. Therefore, the requirements of data-driven methods to model the degradation and predict the RUL are easier to satisfy in reality. At present, data-driven methods are widely used in RUL prediction [12, 13].

The performance of many data-driven prognostics methods is heavily dependent on the choice of the performance degradation data to which they are applied [14]. However, engines have many sensor parameters. The sensitivity of the data from different sensors varies in terms of showing engine performance degradation; the data from some sensors is sensitive and the data from other sensors is not sensitive. Therefore, it is necessary to select suitable sensor parameters whose data are more sensitive to the engine's performance degradation trend as the training data for the RUL prediction model. By observing the characteristic variations of the data from all sensor parameters, quadratic fitting curve is used to fit the degradation data from different sensors and rank the engine's sensor parameters by sensitivity.

Three problems hinder the implementation of performance degradation feature extraction in practice. The first is to select the most sensitive performance degradation features for identifying performance degradation trends easily. The second is that the relevant performance degradation features are often not available and unknown a priori; a large number of candidate performance degradation features have been proposed to better represent the performance degradation state. The last is that most traditional methods of extracting performance degradation features for prognostics are unsupervised and cannot automatically adjust the feature extraction modal parameters based on feedback from the prediction [15–17]. Such feature extraction and choice is significant but represents a principal shortcoming of popular prognostics algorithms: the inability to extract and organize discriminative or trend information from data. Therefore, it

is important to develop an automatic feature extraction method that is capable of extracting the prominent feature to achieve better insight into the underlying performance degradation state.

Deep learning, a new method that has been put forward in the last few years, can be used to extract multilevel features from data, which means the method could express data at different levels of abstraction [18]. Deep learning is an end-to-end machine learning system. It can automatically process an original signal, identify discriminative and trend feature in the input data layer by layer, and then, directly output the classification/regression result. The whole process of feature learning and classifier/regression model training is based on optimizing an overall objective function. In contrast, traditional machine learning processes are divided into several discontinuous data preprocessing steps, such as manual feature extraction and classifier/regression model training, and each step is based on optimizing a separate objective function. Due to the advantage of feature self-learning, deep learning has had great success in applications in artificial intelligence, including computer vision (CS), natural language processing (NLP) [19, 20], object recognition [21], and image information retrieval [22, 23]. Deep learning is not only popular in the academic world but also favored in the industrial world. Companies such as Google, Microsoft, Apple, IBM, and Baidu [24], whose products are widely used, are researching deep learning and have made achievements, such as AlphaGo.

There are many deep learning methods: deep neural networks (DNNs), convolutional deep neural networks (CNNs), deep belief networks (DBNs), and so on [25], for instance, have been proposed. The stacked sparse autoencoder (SAE) [26] is one of the most commonly used deep neural network approaches. SAE consists of multilayer autoencoder such as sparse autoencoder, denoising autoencoder, and so on. Sparse autoencoder is on the basis of autoencoder and introduced sparse constraint condition to aid the expression code as sparse as possible. Denoising autoencoder can learn to remove the noise which is added to the initial input data and extract more robust expression of the input data [27]. For this reason, SAE can effectively capture the important factor of input data, extract more helpful and robust features of data, and then realize excellent performance in pattern recognition and machine learning.

In recent years, various researchers have demonstrated the success of DNN and SAE models in the application of machine health monitoring, such as fault classification of induction motor operated under six different conditions, vibration based fault diagnosis of rolling bearing and hydraulic pump, fault detection within tidal turbine's generator from vibration data acquired from an accelerometer sensor placed within the nacelle of the turbine, vibration based condition monitoring of air compressors, multi class fault classification of spacecraft using large variety of data generated during the spacecraft test, anomaly detection and fault disambiguation in large flight data, drill bit and steel plate health monitoring using vibration data, fault recognition of voltage transformer in electric power industry and so on [28–36]. Most of the research of SAE based

health monitoring mainly focus on anomaly detection and fault diagnosis at present. However, there are few applications on RUL prediction, especially for aircraft engine RUL prediction.

Consequently, a prognostics method based on a stacked sparse autoencoder is proposed to promote self-learning of multilayer features and to predict the RUL of an aircraft engine. The remainder of this paper is organized as follows: Section 2 presents the entire prediction method procedure and framework. Section 3 presents and discusses the prediction results. Finally, conclusions are drawn in Section 4.

2. Methodology

This section introduces the relevant algorithms used in this research. As depicted in Figure 1, the whole procedure for RUL prediction for an aircraft engine consists of two main steps: data preprocessing and RUL prediction using the SAE.

2.1. Data Preprocessing. Selection of sensors that are sensitive to performance degradation and standardization of sensor data with different dimensions are the primary tasks necessary to obtain a high RUL prediction accuracy. Three steps are needed to preprocess the data.

2.1.1. Sensor Selection. Different sensors in an aircraft engine have very different responses to the performance degradation process. Some sensors show unclear tendencies because of noise or insensitivity to degradation trends. Choosing insensitive parameter data may reduce the RUL prediction accuracy. To improve the performance of the prediction model, sensors that are more sensitive to the performance degradation process are chosen as inputs to the RUL prediction model. A method called slope analysis is proposed for sensitivity measurement. Its three main steps are as follows:

- Step 1: curve fitting is performed on the degradation data for each parameter of each engine. Then, the parameters of the best-fit curves, called slopes, are used to analyze the sensitivity of the degradation data.
- Step 2: the average values of all the engine parameters in the step 1 that belong to the same sensor are calculated. Then, the different average parameter values for the different sensors show the individual sensitivity of the degradation data.
- Step 3: the degradation data with larger slopes are selected for predicting the RUL of the engine.

2.1.2. Data Normalization. The linear function that best preserves the original performance degradation pattern of the aircraft engine is chosen to map the data for each selected sensor to $[0, 1]$.

2.1.3. RUL Normalization. The proposed prediction method outputs a result in the range from 0 to 1. In the training stage of the prediction model, the RUL of each cycle of aircraft engine should also be normalized to $[0, 1]$ using a linear

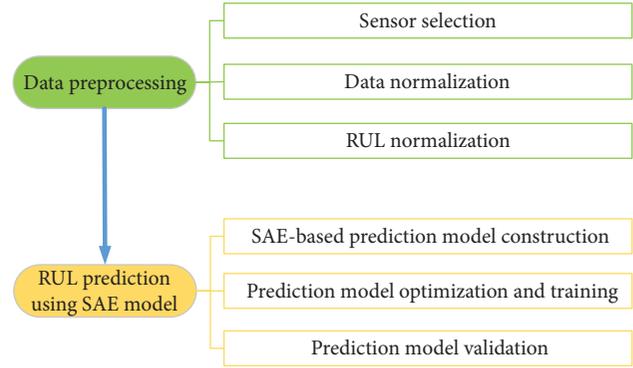


FIGURE 1: The procedure for predicting the RUL of an aircraft engine.

function. The test outputs of the prediction model need to be inversely mapped from $[0, 1]$ to the real RUL.

2.2. SAE Model Construction

2.2.1. Deep Architecture. Cortical computations in the brain have deep architecture and multiple layers of processing. For example, a visual image is processed in multiple stages by the brain, first by cortical area “V1,” then by cortical area “V2,” and so on [37]. Inspired by the information-processing scheme of the brain, deep neural networks have similar deep architectures and multiple hidden layers, which can support complex recognition tasks [6, 37]. As is typical of deep neural networks, the stacked sparse autoencoder (SAE) consists of multiple autoencoders. Compared with traditional neural networks with shallow architectures, it can learn features better and extract deeper discriminative representations [38].

However, it is difficult to train deep architectures [39]. This problem has been addressed by Hinton et al. [40–42], who showed that deep architectures can be trained by relying on two main procedures: (1) on the basis of an unsupervised autoencoder, the deep architecture layers are processed by pretraining, and the output of the top layer’s autoencoder is used as the input to a logistic regression and (2) fine-tuning based on backpropagation is used to adjust the model parameters to obtain accurate prediction results.

2.2.2. Sparse Autoencoder. An autoencoder, first introduced by Hinton et al. [40], is a general form of deep learning method [43] that has been extensively used in unsupervised feature learning. As shown in Figure 2, an autoencoder has three layers: an input layer, a hidden layer, and an output layer. The whole network is trained to realize the reconstruction from the input layer to the output layer, while the hidden layer is accepted as the key feature. However, the traditional autoencoder is not an efficient way to obtain significant representativeness due to its intrinsic limitations. The SAE, as an extension of an autoencoder, can be trained to obtain relatively sparse representatives by introducing a sparse penalty term into the autoencoder [44]. The sparse features learned by the SAE have meanings that are more practical in experiments and applications.

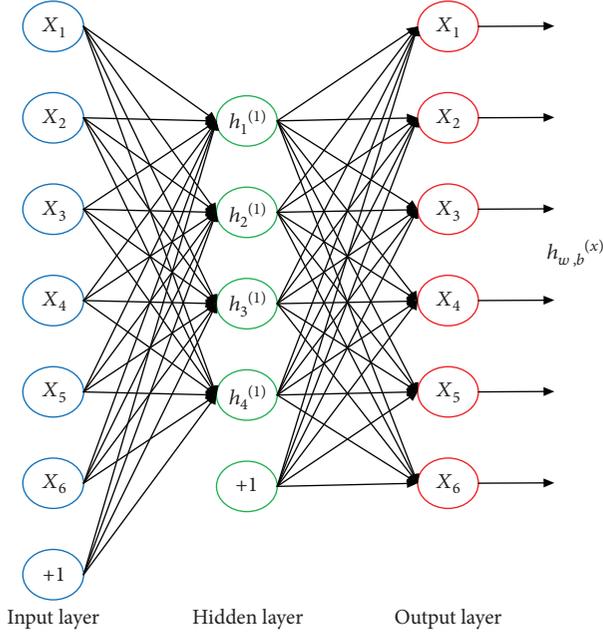


FIGURE 2: The structure of an autoencoder.

The SAE model contains two parts:

(i) An encoder map

The encoder maps an input vector $x^{(i)} \in \mathbb{R}^{d_0}$ (the i th training example) to the latent representation $a^{(i)}$ through deterministic mapping

$$a^{(i)} = f_1(x^{(i)}) = \text{sigmoid}(W_1 x^{(i)} + b_1), \quad (1)$$

where sigmoid is the activation function of the encoder with weight matrix W_1 and bias vector b_1 .

(ii) A decoder map

The decoder maps feature $a^{(i)}$ back to a reconstruction of the vector $h_{W,b}(a^{(i)}) \in \mathbb{R}^{d_0}$ in the output space [45] through a mapping function

$$\hat{x}^{(i)} = h_{W,b}(a^{(i)}) = \text{sigmoid}(W_{11} a^{(i)} + b_2). \quad (2)$$

The decoder map tries to learn a function $h_{W,b}(a^{(i)}) \approx x^{(i)}$, which means making the output $\hat{x}^{(i)}$ similar to the input $x^{(i)}$. Similarly, sigmoid is set as the activation function of the decoder map with weight matrix W_{11} and bias vector b_2 .

During the learning process, the parameters of the SAE are adjusted using backpropagation by minimizing the cost function within the sparsity constraint. The sparsity constraint works on the hidden layer to limit its units and makes it into a sparse vector in which most elements are zero or close to zero [44]. For the autoencoder's network structure, a neuron with a sigmoid activation function is in the active state if its output is close to 1 and the inactive state if its

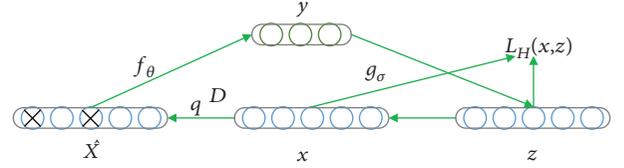


FIGURE 3: The concept of the denoising autoencoder.

output is close to 0. Therefore, the sparsity constraint is introduced to restrict most of the neurons to inactivity most of the time.

The activation of hidden unit j is denoted by $a_j(x)$, and the average activation of hidden unit j is as follows:

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [a_j^{(i)}(x^{(i)})]. \quad (3)$$

Then, we define the sparsity constraint as $\hat{\rho}_j = \rho$, where ρ denotes the sparsity criterion and has a value that is close to zero, that is, most of the neurons in the hidden layer are inactive.

To reach the goal of sparsity, a penalty term is introduced to the objective function that penalizes $\hat{\rho}_j$ if it deviates significantly from ρ . In our study, the KL divergence [45] is selected as the penalty term;

$$\text{KL}(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}. \quad (4)$$

The training set of m training examples is denoted by $\{(x^{(1)}, y^{(1)}) \dots, (x^{(m)}, y^{(m)})\}$, and the original cost function is defined as

$$\begin{aligned} J(W, b) &= \left[\frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}; y^{(i)}) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= \left[\frac{1}{m} \sum_{i=1}^m \frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2. \end{aligned} \quad (5)$$

The first term in (5) is an average sum-of-squares error term, and the second term is a regularization term or weight decay term, which tends to decrease the magnitude of the weights. Here, W and b are the same as in (1) and (2), and λ is the weight decay parameter.

By adding the sparse penalty term, the cost function is modified to

$$J(W, b)_{\text{sparse}} = J(W, b) + \beta \sum_{j=1}^{s_2} \text{KL}(\rho \parallel \hat{\rho}_j), \quad (6)$$

where β represents the weight of the sparsity penalty term.

2.2.3. Denoising Autoencoder. Despite the process described above, learning features well to improve the performance and generalization ability of the prediction model continues to face challenges because of the noise and outliers that commonly appear in real-world data. To force the hidden layer to

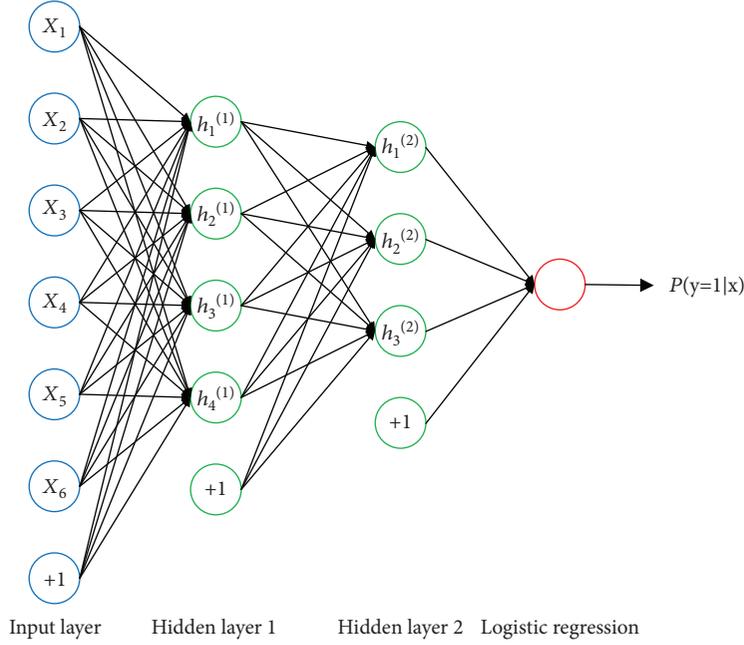


FIGURE 4: A two-layer stacked sparse autoencoder and logistic regression.

discover more robust features, the autoencoder can be trained by reconstructing the input from a corrupted version of it, which is the idea behind denoising autoencoders [37], as shown in Figure 3.

These data corruption is implemented by corrupting the initial input x to create a partially destroyed version \tilde{x} by means of a stochastic mapping,

$$\tilde{x} \sim q_D(\tilde{x}|x). \quad (7)$$

The standard approach is to apply masking noise to the original data by setting a random fraction of the elements of x to zero. Next, the corrupted data \tilde{x} pass through a basic autoencoder process and is mapped to a hidden representation,

$$y = f_\theta(\tilde{x}) = \text{sigmoid}(W \cdot \tilde{x} + b). \quad (8)$$

From this equation, we reconstruct

$$z = g_\theta'(y). \quad (9)$$

In the last stage, the parameters are trained to minimize the average reconstruction error

$$L_H(x, z) = H(B_x \| B_z) \quad (10)$$

to make z as close as possible to the uncorrupted input x .

2.2.4. Structure of the Stacked Sparse Autoencoder. As a typical neural network, the stacked autoencoder consists of multiple layers of sparse or denoising autoencoders (discarding the decoder) and a logistic regression. The outputs of each layer of the stacked autoencoder are wired to the inputs of the subsequent layer. The architecture of a two-layer stacked sparse autoencoder is shown in Figure 4. Each sparse or denoising autoencoder generates a representation of the inputs (data from the aircraft engine's sensors) that is more

abstract and high dimensional than the previous layer's because it is obtained by applying an additional nonlinear transformation. The output of the last layer of the sparse autoencoder are input to the logistic regression and then, the results (the predicted RUL) are obtained.

(1) *Prediction Using Logistic Regression.* The purpose of logistic regression is to find an optimal model for matching independent variables and class distinctions of dependent variables (probabilities of the occurrence of an event). The logistic function is expressed by

$$\text{prob}(\text{event}) = p(x) = \left(1 + e^{-e(x)}\right)^{-1}. \quad (11)$$

The logistic or logit model is

$$\begin{aligned} \text{Logit} = g(x) &= \log(p(x)(1-p(x))^{-1}) \\ &= \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k, \end{aligned} \quad (12)$$

where $g(x)$ is a linear combination of the independent variables x_1, x_2, \dots, x_k .

Parameters of models (such as $\alpha, \beta_1, \dots, \beta_k$) need to be determined beforehand, which is the major premise for determining $P(x)$. Because of the existence of dichotomous dependent variables, it is improper to estimate the values of the parameters using the least-squares method [46]. Therefore, compared with the method of minimizing the sum of the squared errors, the paper uses the maximum likelihood method to estimate the parameters (such as $\alpha, \beta_1, \dots, \beta_k$) of the logistic regression [47]. Then, the probability of the occurrence of the event can be obtained using (11) once the vector x has been determined.

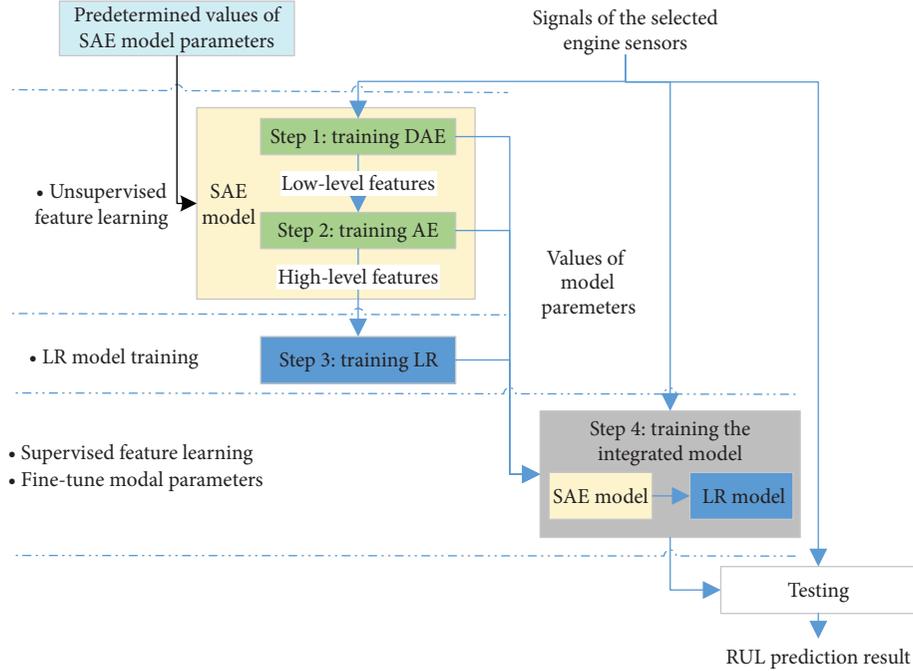


FIGURE 5: Procedure for training the SAE-based RUL prediction model.

(2) *Fine Tuning.* The process of fine-tuning mainly focuses on adjusting the weights in the SAE network, which leads to much better prediction performance.

First, feed-forward is used to compute the activations for all the autoencoder layers.

In the next step, we set $\delta^{(n_l)} = -(\nabla_{a^{n_l}} J) \bullet f'(z^{n_l})$ for the output layer, where $\nabla J = \theta(I - P)$, I is the input label, and P is the vector of conditional probabilities. Then, for layers $l = n_l - 1, n_l - 2, \dots, 3, 2$, we set $\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) \bullet f'(z^{n_l})$, and then, the desired partial derivatives are

$$\begin{aligned} \nabla W^l J(W, b; x; y) &= \delta^{(l+1)} \left(a^{(l)} \right)^T, \\ \nabla b^l J(W, b; x; y) &= \delta^{(l+1)}, \end{aligned} \quad (13)$$

where W , b , and a are as in (1) and (2).

Finally, the batch gradient descent algorithm is used to minimize the overall cost function.

2.3. Training and Optimization of SAE-Based RUL Prediction

2.3.1. Procedure for Training the SAE-Based RUL Prediction Model. A two-layer stacked sparse autoencoder and a logistic regression (LR) model were used as an example to illustrate the training procedures in the proposed deep learning-based RUL prediction methodology. The values of the SAE parameters are predetermined. A grid search is used to find a set of optimal SAE parameters. The four major steps of the procedure are as follows:

Step 1: a single-layer denoising autoencoder (DAE), the first layer of the SAE, is trained to extract robust performance degradation features using unsupervised learning [37]. The signals of the selected

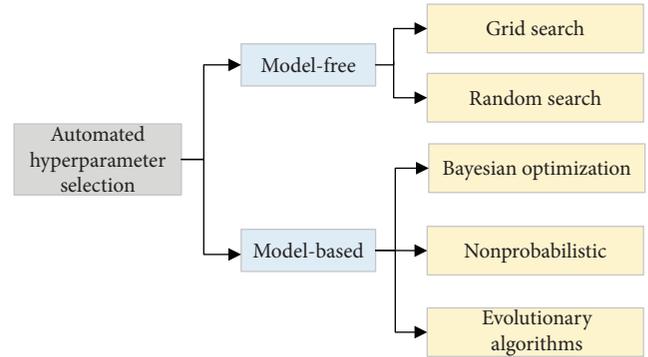


FIGURE 6: Approaches to automated hyperparameter selection.

sensors are input into the DAE, and then, low-level features are output by the hidden layer of the DAE.

Step 2: a single-layer sparse autoencoder (AE), the second layer of the SAE, is trained for unsupervised self-learning of features. The low-level features are input into the AE, and the high-level features are output by the hidden layer of the AE.

Step 3: high-level features are used as inputs to train the LR model for RUL prediction. The target output of the LR model is the normalized RUL of the aircraft engine.

Step 4: the previously trained SAE and LR model are combined into an integrated feature learning and RUL prediction model. Then, the integrated model is trained using supervised learning for the final feature learning to obtain the RUL

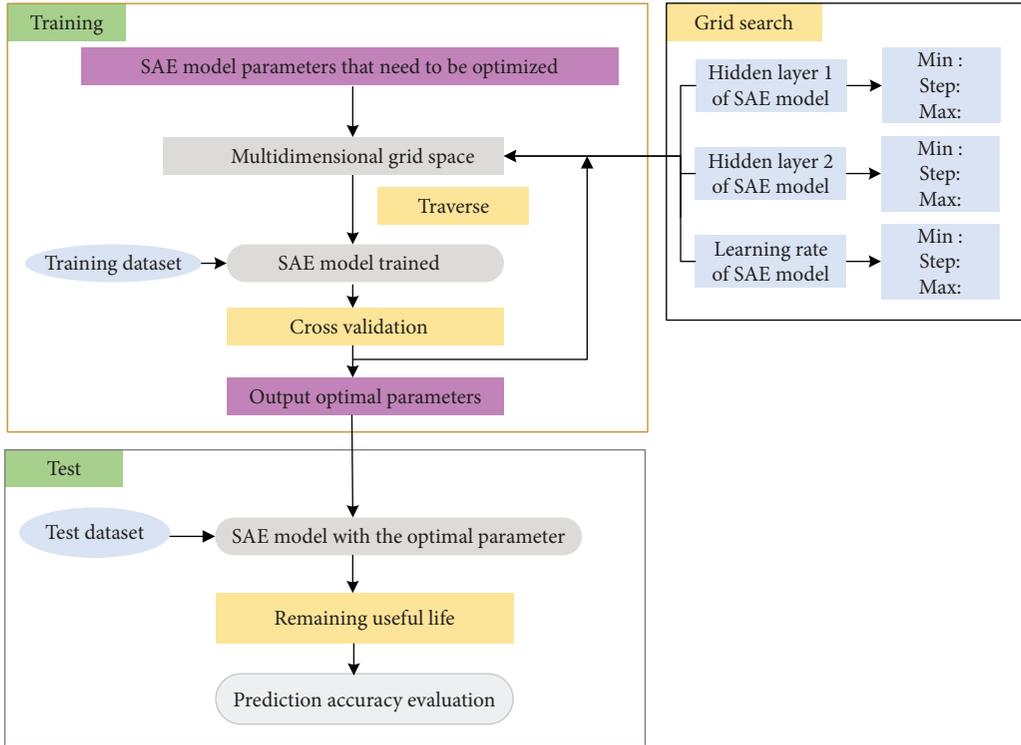


FIGURE 7: Grid search-based SAE RUL prediction model parameter optimization and validation.

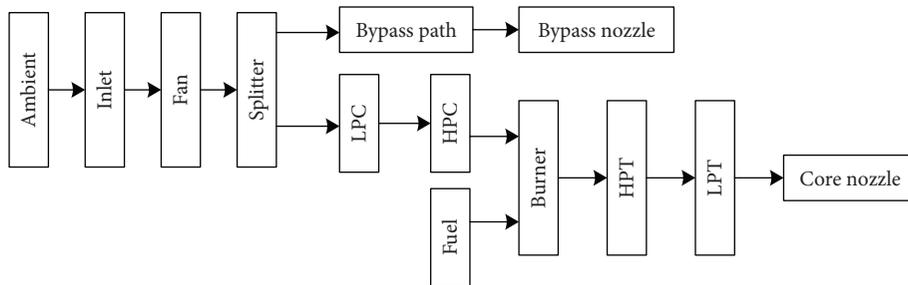


FIGURE 8: Connections of various modules in the simulation environment.

prediction model. The signals of the selected engine sensors are the inputs of the integrated model, and the normalized RUL of the engine is used as the target output during model training. Training the integrated model using supervised learning can fine-tune the modal parameters (the parameters of the DAE, the AE, and the LR models) based on the values obtained in the previous training, steps 1 to 3. The features obtained from the fine-tuned DAE and AE more clearly present the degradation trend of the engine performance. Based on these features, the LR model can provide a more accurate RUL prediction result.

The process of training the proposed RUL prediction method is summarized in Figure 5.

2.3.2. Grid Search-Based SAE RUL Prediction Model Parameter Optimization and Validation. The hyperparameters of the deep learning, which have significant impacts on the feature extraction performance, are adjusted in most cases based on expert experience. In view of the difficulty of adjusting hyperparameters by means of deep learning, a method of optimizing the hyperparameters is necessary.

There are currently two main types of method of automated hyperparameter selection for the SAE (which is shown in Figure 6). One includes model-free methods, which include the grid and random search methods; the other includes model-based methods, which mainly include three subcategories, the Bayesian optimization (e.g., spear-mint [48]), nonprobabilistic methods (such as RBF surrogate models [49]), and evolutionary algorithms (e.g., genetic algorithms [50] and particle swarm optimization [51]). Model-based methods efficiently explore the solution space

according to the algorithm selected and then, quickly obtain the accepted parameter value. However, the identified hyperparameter value may be a local optimum, and the method has several individual hyperparameters, which would increase its complexity.

Unlike model-based methods of hyperparameter selection, model-free hyperparameter selection methods search for the optimal parameters within the defined space; the main ones are grid and random searching [52, 53]. In the paper, the grid search method is chosen to search for the hyperparameters of the SAE.

There are a few reasons why grid search is chosen as the hyperparameter optimization algorithm used in the proposed SAE-based RUL prediction model.

- (1) Compared with the manual search method of optimizing the hyperparameters, a grid search is more likely to identify better model parameters than pure manual sequential optimization (in the same time).
- (2) Compared with model-based hyperparameter selection methods, a grid search is simple to implement, and parallel computing is easy to implement.
- (3) Compared with the random search method of hyperparameter optimization, mesh searches are recommended when few parameters need to be optimized.

Theoretically, when the space defined by the optimized parameters is large enough and the changes in the optimal parameters are small enough, the optimization method called mesh searching could be used to find the global optimal solution.

There are three main steps in the grid search-based hyperparameter optimization of the SAE RUL prediction model.

- Step 1: the hyperparameters to be optimized are defined in the space, and the space is divided into grids with a fixed step size. Each point on each grid is a combination of model parameters.
- Step 2: the training set is divided into several subsets of equal size. Then, the SAE is trained with one combination of model parameters. Details of the procedure for training the SAE-based RUL prediction model are in Section 2.3.1 of this paper.
- Step 3: step 2 is repeated until the grid search has been completed. The resulting optimal hyperparameters are output.

Figure 7 shows the process of the method. To obtain an SAE with the optimal parameters, predicted degradation data for the engines are input into the trained model, and the RUL of each engine is obtained.

3. Case Study

3.1. Engine Data Description. The challenge datasets used for the prognostics challenge competition at the 2008 PHM International Conference consist of multiple multivariate

TABLE 1: Description of the sensor signals for the aircraft gas turbine engine.

Index	Symbol	Description	Unit
1	T2	Total temperature at fan inlet	°R
2	T24	Total temperature at LPC outlet	°R
3	T30	Total temperature at HPC outlet	°R
4	T50	Total temperature at LPT outlet	°R
5	P2	Pressure at fan inlet	psia
6	P15	Total pressure in bypass-duct	psia
7	P30	Total pressure at HPC outlet	psia
8	Nf	Physical fan speed	rpm
9	Nc	Physical core speed	rpm
10	epr	Engine pressure ratio (P50/P2)	—
11	Ps30	Static pressure at HPC outlet	psia
12	phi	Ratio of fuel flow to Ps30	pps/psi
13	NRF	Corrected fan speed	rpm
14	NRC	Corrected core speed	rpm
15	BPR	Bypass ratio	—
16	farB	Burner fuel-air ratio	—
17	htBleed	Bleed enthalpy	—
18	Nf_dmd	Demanded fan speed	rpm
19	PCNFR_dmd	Demanded corrected fan speed	rpm
20	W31	HPT coolant bleed	lbm/s
21	W32	LPT coolant bleed	lbm/s

°R: Rankine temperature scale; psia: pounds per square inch absolute; rpm: revolutions per minute; pps: pulses per square inch; lbm/s: pound mass per second.

TABLE 2: Operating regime of the aircraft engine.

Operation regime	Altitude	Mach number	Throttle resolver angle (TRA)
R1	20	0.7	0

time series, which were collected via a dynamical simulation of an engine system. The model simulated various degradation scenarios in any of the five rotating components of the simulated engine (fan, LPC, HPC, HPT, and LPT), and the connections among the engine modules in the simulation are shown in Figure 8. The engine begins in normal operation, then, degradation appears in some cycle of the simulation. The degradation data for each engine are recorded until the engine fails. The simulation model results in 218 engine datasets defined as unit 1 through unit 218 with different failure times measured by the number of operating cycles for the same engine system.

The complete dataset for each cycle of each engine unit consists of the unit ID, the operating cycle index, the operational regime settings, and typical sensor measurements. A total of 21 sensors (shown in Table 1) are installed in different components of the aircraft engine. A total of 21 sensory signals are obtained under the R1 operation regime shown in Table 2. In this study, sensor data were collected from 200 aircraft engines injected with the HPC degradation

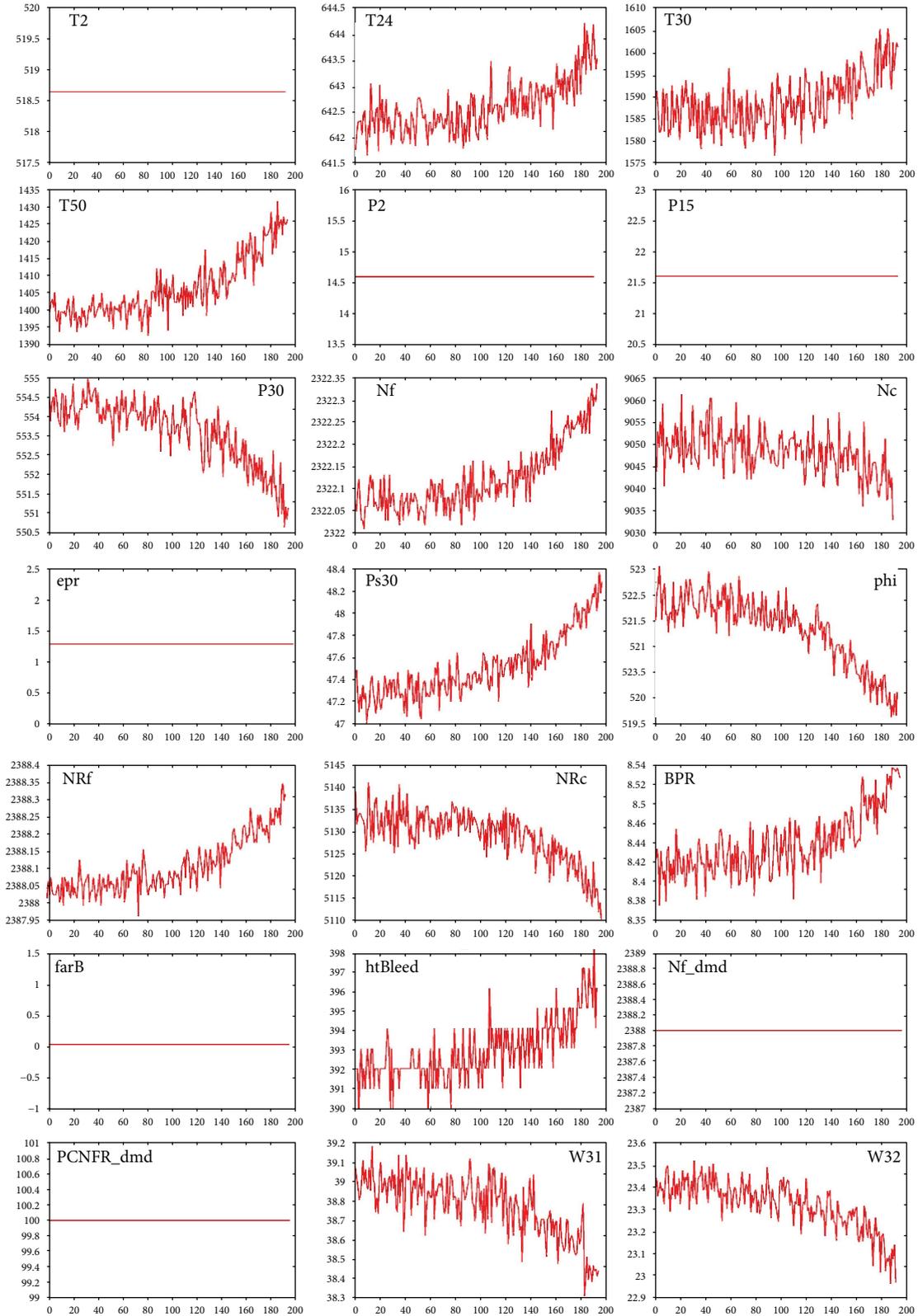


FIGURE 9: Engine degradation tendencies using the data from different sensors over time.

fault mode. The dataset considered in this study consists of three files, which include degradation data for 100 training and 100 testing units and the remaining useful life

of the 100 testing units. Each training unit runs to breakdown, and each testing unit stops running at some time before it breaks down. Through investigation and research,

TABLE 3: Operation regime of the aircraft engine.

Rank	Sensor	Sensitivity
1	Ps30	2.63785E-05
2	phi	2.58711E-05
3	P30	2.45237E-05
4	T50	2.43431E-05
5	BPR	2.35888E-05
6	Nf	2.30928E-05
7	NRf	2.25432E-05
8	W32	2.17608E-05
9	W31	2.1399E-05
10	htBleed	2.1267E-05
11	T24	2.1055E-05
12	T30	1.9856E-05
13	Nc	1.1581E-05
14	NRc	5.86587E-06

TABLE 4: Parameters of the SAE obtained using the grid search method.

Parameter name	Parameter value
Input layer	16
Learning rate of NN model	0.5
Number of training cycles	300

it is found that the dataset is highly authoritative and accurate [54–56].

3.2. Results and Discussion

3.2.1. Data Preprocessing. In the process of engine performance degradation, the performance data from the sensors gradually change over time, and the data indirectly reflect the degradation tendency of the engine’s performance. However, the sensitivity of different parameters to degradation varies over time. Figure 9 shows the degradation tendency of the 21 performance parameters. According to Figure 9, data from seven of the sensors (1, 5, 6, 10, 16, 18, and 19) exhibit no tendency, so the sensitivities of the remaining parameters to engine performance degradation are analyzed. The results of the 14 performance parameters for which the sensitivity analysis is conducted are shown in Table 3. To reduce the computational complexity, the data from the first six sensors (4, 7, 8, 11, 12, and 15) in the sensitivity ranking are selected. By surveying and analyzing relevant information about the RUL of engines, parameters T24 and T30 are also chosen as objects of study. Finally, eight performance parameters (2, 3, 4, 7, 8, 11, 12, and 15) are chosen for predicting the RUL of the aircraft engine [57–59].

3.2.2. SAE Parameters Optimized Using Grid Searching. The SAE used in this paper has eight hyperparameters: input layer, hidden layer 1, hidden layer 2, output layer, learning rate of SAE, learning rate of NN model, and number of

TABLE 5: Parameters of the SAE model obtained using the method.

Parameter name	Min	Step	Max
Hidden layer 1 of SAE	4	2	16
Hidden layer 2 of SAE	2	2	8
Learning rate of SAE	0.3	0.1	1

TABLE 6: Life prediction results for the first seven optimal parameters.

Rank	Accuracy rate	MAPE	Acceptable number	Acceptable rate
1	83.82%	16.18%	82	86.32%
2	83.67%	16.33%	81	85.26%
3	83.67%	16.33%	80	84.21%
4	83.66%	16.34%	80	84.21%
5	83.65%	16.35%	80	84.21%
6	83.62%	16.38%	81	85.26%
7	83.61%	16.39%	81	85.26%

TABLE 7: Engine life prediction accuracy of the 2008 PHM data challenge.

Rank	Score	MAPE	Accuracy rate
1	512.12	15.81%	84.19%
2	740.31	18.92%	81.08%
3	873.36	19.19%	80.81%
4	1218.43	20.15%	79.85%
5	1218.76	33.14%	66.86%
6	1232.27	32.90%	67.10%
7	1568.98	36.75%	63.25%
8	1645.77	30.00%	70.00%
9	1816.60	26.47%	73.53%
10	1839.06	27.72%	72.28%

training cycles. Based on the results, the parameters input layer, learning rate of NN model, and number of training cycles are shown in Table 4, which is a good parameter match. Then, there are three hyperparameters that need to be optimized using grid searching. The grid search method of automated hyperparameter selection for SAE is performed in a defined space with a fixed step size. The proposed parameters of the SAE obtained using the grid search method are shown in Table 5.

3.2.3. Results. Through automated selection of the DNN hyperparameters using a grid search, the experimental results show that the method is effective, with an accuracy rate of up to 83.82% and an acceptable rate of up to 86.32% of ranking first (shown in Table 6). Compared with the accuracy of the 2008 PHM data challenge engine life prediction, the first-rank prediction accuracy is 84.19% (shown in Table 7), and the RUL prediction accuracy is quite close. However, there are six types of working condition and 218 training and test

TABLE 8: The first seven optimal parameter arrays obtained by the grid search method.

Rank	Input layer	Hidden layer 1	Hidden layer 2	Output layer	Learning rate of SAE model
1	16	10	2	1	0.9
2	16	10	2	1	0.5
3	16	8	2	1	0.3
4	16	8	2	1	0.5
5	16	12	2	1	0.9
6	16	10	2	1	0.7
7	16	12	4	1	0.5

sets in the FD005T dataset, which were used in the data challenge. Then, comparing the method proposed in this paper and the 2008 PHM data challenge provides only a relative comparison. The accuracy of the RUL predictions obtained in this paper is acceptable in the field of engine prediction, and the results are satisfactory. Table 8 shows the first seven optimal parameter arrays obtained by the grid search method. Table 6 shows the life prediction results based on the first seven optimal parameters. Table 7 shows the seven most accurate engine life predictions in the 2008 PHM data challenge.

4. Conclusions

In this paper, a new data-driven approach to engine prognostics is developed based on deep learning that can capture effective nonlinear features by themselves and reduce manual intervention. The SAE, a type of deep learning model, is not only able to capture the tendency of the system to evolve but also sufficiently robust to noise. To automatically select the hyperparameters of the SAE, the grid search algorithm is used. The method of predicting an aircraft engine's remaining useful life is applied to the 2008 PHM data challenge dataset to demonstrate the effectiveness of the proposed approach. The experimental results, which show a satisfactory prediction accuracy and acceptance rate for all the samples, show that the method is effective at predicting the RUL of an aircraft engine. It also has significance for enhancing the safety of aircraft engines and prognosticating and managing the health of aircraft engines to reduce the cost of maintenance.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant nos. 51605014, 51105019,

and 51575021), the Aviation Science Fund (Grant no. 20163351018), the Technology Foundation Program of National Defense (Grant no. Z132013B002), and the Fundamental Research Funds for the Central Universities (Grant no. YWF-18-BJ-Y-159).

References

- [1] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 International Conference on Prognostics and Health Management*, pp. 1–9, Denver, CO, USA, October 2008.
- [2] S. H. Ding and S. Kamaruddin, "Maintenance policy optimization—literature review and directions," *International Journal of Advanced Manufacturing Technology*, vol. 76, no. 5–8, pp. 1263–1283, 2015.
- [3] A. Bousdekis, B. Magoutas, D. Apostolou, and G. Mentzas, "Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance," *Journal of Intelligent Manufacturing*, vol. 29, no. 6, 2018.
- [4] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [5] D. Djurdjanovic, J. Lee, and J. Ni, "Watchdog Agent—an infotonics-based prognostics approach for product performance degradation assessment and prediction," *Advanced Engineering Informatics*, vol. 17, no. 3–4, pp. 109–125, 2003.
- [6] F. Ahmadzadeh and J. Lundberg, "Remaining useful life estimation: review," *International Journal of System Assurance Engineering and Management*, vol. 5, no. 4, pp. 461–474, 2014.
- [7] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: state of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009.
- [8] F. Camci and R. B. Chinnam, "Health-state estimation and prognostics in machining processes," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 3, pp. 581–597, 2010.
- [9] J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto, and S. Chigusa, "Model-based prognostic techniques [maintenance applications]," in *Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference*, Anaheim, CA, USA, September 2003.
- [10] O. F. Eker, F. Camci, and I. K. Jennions, "Major challenges in prognostics: study on benchmarking prognostics datasets," in *European conference of the prognostics and Health Management*, Dresden, Germany, 2012.
- [11] H. Zhang, R. Kang, and M. Pecht, "A hybrid prognostics and health management approach for condition-based maintenance," in *2009 IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 1165–1169, Hong Kong, China, December 2009.
- [12] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation – a review on the statistical data driven approaches," *European Journal of Operational Research*, vol. 213, no. 1, pp. 1–14, 2011.
- [13] M. Schwabacher and G. Kai, "A survey of artificial intelligence for prognostics," Tech. Rep., AAAI Fall Symposium, 2007.

- [14] N. Gebraeel, A. Elwany, and J. Pan, "Residual life predictions in the absence of prior degradation knowledge," *IEEE Transactions on Reliability*, vol. 58, no. 1, pp. 106–117, 2009.
- [15] Y. Yang, Y. Liao, G. Meng, and J. Lee, "A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11311–11320, 2011.
- [16] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1-2, pp. 314–334, 2014.
- [17] R. F. Orsagh, J. Sheldon, and C. J. Klenke, "Prognostics/diagnostics for gas turbine engine bearings," in *2003 IEEE Aerospace Conference Proceedings (Cat. No.03TH8652)*, pp. 3095–3103, Big Sky, MT, USA, March 2003.
- [18] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [19] O. Abdel-Hamid, L. Deng, D. Yu, and H. Jiang, "Deep segmental neural networks for speech recognition," *Interspeech*, vol. 36, p. 70, 2013.
- [20] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, March 2013.
- [21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.
- [23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105, Curran Associates, Inc., 2012.
- [24] L. Deng and D. Yu, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3-4, pp. 197–387, 2014.
- [25] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [26] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," *Large-Scale Kernel Machines*, vol. 34, no. 5, pp. 1–41, 2007.
- [27] S. Hu, Y. Zuo, L. Wang, and P. Liu, "A review about building hidden layer methods of deep learning," *Journal of Advances in Information Technology*, vol. 7, no. 1, pp. 13–22, 2016.
- [28] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.
- [29] T. Junbo, L. Weining, A. Juneng, and W. Xueqian, "Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder," in *The 27th Chinese Control and Decision Conference (2015 CCDC)*, pp. 4608–4613, Qingdao, China, May 2015.
- [30] Z. Huijie, R. Ting, W. Xinqing, Z. You, and F. Husheng, "Fault diagnosis of hydraulic pump based on stacked autoencoders," in *2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, pp. 58–62, Qingdao, China, July 2015.
- [31] G. S. Galloway, V. M. Catterson, T. Fay, A. Robb, and C. Love, "Diagnosis of tidal turbine vibration data through deep neural networks," in *Proceedings of the Third European Conference of the Prognostics and Health Management Society*, pp. 172–180, 2016.
- [32] N. K. Verma, V. K. Gupta, M. Sharma, and R. K. Sevakula, "Intelligent condition based monitoring of rotating machines using sparse auto-encoders," in *2013 IEEE Conference on Prognostics and Health Management (PHM)*, pp. 1–7, Gaithersburg, MD, USA, June 2013.
- [33] R. Thirukovalluru, S. Dixit, R. K. Sevakula, N. K. Verma, and A. Salour, "Generating feature sets for fault diagnosis using denoising stacked auto-encoder," in *2016 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–7, Ottawa, ON, Canada, June 2016.
- [34] R. Kishore, K. Reddy, S. Sarkar, and M. Giering, "Anomaly detection and fault disambiguation in large flight data: a multi-modal deep auto-encoder approach," in *Annual Conference of the Prognostics and Health Management Society*, PHM Society, Denver, Colorado, 2016.
- [35] L. Wang, X. Zhao, J. Pei, and G. Tang, "Transformer fault diagnosis using continuous sparse autoencoder," *Springerplus*, vol. 5, no. 1, p. 448, 2016.
- [36] K. Li and Q. Wang, "Study on signal recognition and diagnosis for spacecraft based on deep learning method," in *2015 Prognostics and System Health Management Conference (PHM)*, Beijing, China, October 2015.
- [37] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning - ICML '08*, pp. 1096–1103, Helsinki, Finland, 2008.
- [38] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, vol. 7553, nature 521, 2015.
- [39] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, Italy, March 2010.
- [40] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [41] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, *Greedy Layer-Wise Training of Deep Networks*, Advances in neural information processing systems, 2007.
- [42] Y. L. Boureau and Y. L. Cun, *Sparse Feature Learning for Deep Belief Networks*, Advances in neural information processing systems, 2008.
- [43] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," in *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, June 2010.
- [44] N. Japkowicz, S. J. Hanson, and M. A. Gluck, "Nonlinear auto-association is not equivalent to PCA," *Neural Computation*, vol. 12, no. 3, pp. 531–545, 2000.
- [45] L. Chen, H. Qu, J. Zhao, B. Chen, and J. C. Principe, "Efficient and robust deep learning with correntropy-induced loss function," *Neural Computing and Applications*, vol. 27, no. 4, pp. 1019–1031, 2016.

- [46] J. Yan and J. Lee, "Degradation assessment and fault modes classification using logistic regression," *Journal of Manufacturing Science and Engineering*, vol. 127, no. 4, pp. 912–914, 2005.
- [47] W. M. Houston, D. J. Woodruff, and American College Testing Program, *Empirical Bayes Estimates of Parameters from the Logistic Regression Model*, ACT Research Report Series, ACT, Inc., 1997.
- [48] J. Snoek, H. Larochelle, and R. P. Adams, *Practical Bayesian Optimization of Machine Learning Algorithms*, Advances in neural information processing systems, 2012.
- [49] I. Ilievski, T. Akhtar, J. Feng, and C. A. Shoemaker, "Hyperparameter optimization of deep neural networks using non-probabilistic RBF surrogate model," *arXiv preprint arXiv*, vol. 1607, article 08316, 2016.
- [50] K. Liu, L. M. Zhang, and Y. W. Sun, "Deep Boltzmann machines aided design based on genetic algorithms," *Applied Mechanics and Materials*, vol. 568–570, pp. 848–851, 2014.
- [51] A. R. Syulistyo, D. M. J. Purnomo, M. F. Rachmadi, and A. Wibowo, "Particle swarm optimization (PSO) for training optimization on convolutional neural network (CNN)," *Jurnal Ilmu Komputer Dan Informasi*, vol. 9, no. 1, pp. 52–58, 2016.
- [52] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 281–305, 2012.
- [53] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, *Algorithms for hyper-parameter optimization*, Advances in neural information processing systems, 2011.
- [54] T. Wang, J. Yu, D. Siegel, and J. Lee, "A similarity-based prognostics approach for remaining useful life estimation of engineered systems," in *2008 International Conference on Prognostics and Health Management*, pp. 1–6, Denver, CO, USA, October 2008.
- [55] J. Xu, Y. Wang, and L. Xu, "PHM-oriented integrated fusion prognostics for aircraft engines based on sensor data," *IEEE Sensors Journal*, vol. 14, no. 4, pp. 1124–1132, 2014.
- [56] C. Hu, B. D. Youn, P. Wang, and J. Taek Yoon, "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life," *Reliability Engineering & System Safety*, vol. 103, pp. 120–135, 2012.
- [57] E. Ramasso and R. Gouriveau, "Remaining useful life estimation by classification of predictions based on a neuro-fuzzy system and theory of belief functions," *IEEE Transactions on Reliability*, vol. 63, no. 2, pp. 555–566, 2014.
- [58] M. El-Koujok, R. Gouriveau, and N. Zerhouni, "Reducing arbitrary choices in model building for prognostics: an approach by applying parsimony principle on an evolving neuro-fuzzy system," *Microelectronics Reliability*, vol. 51, no. 2, pp. 310–320, 2011.
- [59] K. Liu, N. Z. Gebrael, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 3, pp. 652–664, 2013.

Research Article

Study on the Magnitude of Reservoir-Triggered Earthquake Based on Support Vector Machines

Hai Wei , Mingming Wang , Bingyue Song, Xin Wang, and Danlei Chen

Faculty of Electric Power Engineering, Kunming University of Science and Technology, Kunming, Yunnan 650500, China

Correspondence should be addressed to Mingming Wang; wang.ming.ming@163.com

Received 31 January 2018; Revised 31 May 2018; Accepted 27 June 2018; Published 24 July 2018

Academic Editor: Minvydas Ragulskis

Copyright © 2018 Hai Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An effective approach is introduced to predict the magnitude of reservoir-triggered earthquake (RTE), based on support vector machines (SVM) and fuzzy support vector machines (FSVM) methods. The main influence factors on RTE, including lithology, rock mass integrity, fault features, tectonic stress state, and seismic activity background in reservoir area, are categorized into 11 parameters and quantified by using analytical hierarchy process (AHP). Dataset on 100 reservoirs in China, including the 48 well-documented cases of RTE, are collected and used to train and validate the prediction models established with SVM and FSVM, respectively. Through numerical tests, it is found that both the SVM and FSVM models are effective in the prediction of the magnitude of RTE with high accuracy, provided that sufficient samples are collected. While the results of FSVM which is extended from SVM by introducing a fuzzy membership to reduce the influence of noises or outliers are found to be slightly less accurate than those of SVM in the current analysis of RTE cases. The reason might be attributed to the high discreteness of the sample data in the current study.

1. Introduction

Reservoir-triggered earthquake (RTE), or reservoir earthquake (RE) in brief [1–3], also called as reservoir-induced earthquake (RIE), was caused by impounding water in reservoir created by the construction of dam across rivers. The first case of RTE was pointed out by Carder, which happened at Lake Mead in the USA [4]. Over the past decades, about 130 suspected cases of RTE have been reported in the world including the USA, India, Greece, Egypt, and New Zealand [3]. Among them, about 100 cases were universally acknowledged as RTE, since they were very hard to be distinguished from natural earthquakes. China is among the countries most prone to earthquakes, where about 40 RTE cases have been reported so far. They have brought significant threats to the safety of large dams and consequently are great hazards to the life and property of local residents [5], especially in the west of Sichuan province and the northwest of Yunnan province, where large dams have been built or are under construction in high density. It is necessary for dam engineers to know a reasonable prediction for the magnitude of RTE in order to assess the dam behavior subjected to this type of

seismic load. An effective prediction on the possibility of RTE and its magnitude in case of occurrence is thus essential in the design of large dams for the safety of life and property in related areas.

Like natural earthquakes, the mechanism of RTE is extremely complicated and remains almost completely unknown to seismologists. Since the occurrence time of RTE is very difficult to be predicted, previous studies on RTE have mainly focused on the prediction of its location and magnitude. Through extensive studies by seismologists, it has been widely acknowledged that RTE was closely influenced by combined factors including lithology, mechanical parameters of deep rock mass, the stress state of rock mass, and pore water distribution [6]. However, due to the difficulty in the measurement and quantification of those factors, great challenges have been posed on the analysis of RTE [7, 8].

In spite of this, various analytical approaches have been proposed on the study of RTE in the past decades, such as mechanical model method [7], geological analogy method [7], probabilistic and statistical method [8–10], artificial neural network method (ANN) [11, 12], and artificial intelligence technique [13]. Among them, the mechanical

model method is limited by the fact that the nonlinearity of rock, parameter variation, and pore water distribution as well as boundary conditions cannot be taken into full consideration. Therefore, the geological analogy method is proposed based on analogy with RTE cases, in terms of hydrology, engineering geology, and infiltration conditions. The possibility of RTE, as well as its magnitude and location in case of occurrence, is thus able to be analyzed. However, foundational mechanical principles like the force equilibrium condition cannot be included in this method. The probabilistic and statistical method [8], proposed based on Bayes' theorem, can also be used to evaluate the occurring probability of RTE. It is based on statistical analysis of previous RTE cases, and thus the accuracy of the result is strongly dependent upon sample numbers and prior probability of RTE. The artificial neural network method has strong ability in nonlinear mapping and thus shows great advantages in dealing with complex nonlinear relations in complicated problems, whereas the uncertainty of network structures and problems related to overfitting and underfitting have limited its application in the analysis of RTE.

The support vector machines (SVM) is a powerful machine learning method firstly proposed by Vapnik in 1995 [14]. It is derived from the statistical learning theory and has strong generalization capability. The best compromise between the complexity of model and learning ability could be achieved with the employment of structure risk minimization (SRM) principle, which is developed from the traditional empirical risk minimization (ERM) method [14–16]. SVM was applied initially to pattern classification and later to data regression problems. By introducing a kernel function, the original application of SVM in the optimal classification of linear separable data is extended to nonlinear problems. They could be solved by transferring to a quadratic programming (QP) problem with constraint condition [16]. With its excellent performance in dealing with nonlinear and high-dimensional problems, SVM has been widely applied in multiple fields, such as image recognition, classification, time series prediction, inverse analysis of geotechnical parameters, and hydrological forecasting [17–21]. However, conventional SVM is found very sensitive to noises or outliers. As a result, Lin and Wang [22] proposed fuzzy support vector machines (FSVM) by applying a fuzzy membership to each input point to reduce the effect of noises or outliers. In order to reduce the dependence of membership function on the geometric shape of sample data, Tang et al. [23] and Du et al. [24] introduced a new membership function based on the distance between a sample and its class hyperplane in FSVM. To further reduce the membership of nonsupport vectors and increase that of support vectors, Ding and Gu [25] proposed a new algorithm where a dual membership based on hypersphere was employed. For multiclass classification problems, Tsang et al. [26] defined a degree of membership of a sample to different classes by introducing a fuzzy membership to each training sample, while Abe [27] classified a sample into a multilabel class whose membership function is the largest.

In this study, we attempt to extend the application of SVM and FSVM to predict the magnitude of RTE, based on

dataset collected on reservoirs and RTE cases in China. Habibagahi [12] and Samui and Kim [13] have also used artificial intelligence techniques to predict the magnitude of induced earthquakes. In their models, only limited factors on reservoir parameters such as comprehensive parameter and maximum reservoir depth are taken into account. While geological factors, which are of great significance in the triggering of earthquakes, are not considered. Therefore, an effort has been made to carry out this research covering multiple geological factors of reservoir area and reservoir parameters for providing a step forward of research on RTE. Dataset on 100 large and medium-sized reservoirs, covering the 48 RTE cases, are collected from reputed enterprises and published literatures. Computational models are established based on SVM and FSVM. The analytical hierarchy process (AHP) [28] is employed to quantify geological factors based on geological analysis, providing the input data of the computational models. Three numerical tests are carried out to test the performance of the established models. Results obtained from the SVM and FSVM models are analyzed and compared in detail. Discussions are made on the applicability of SVM and FSVM, which offers an effective approach in the field of RTE prediction.

2. Mathematical Models and Numerical Procedure

2.1. The SVM and FSVM Models. The solution of FSVM classification can also be transferred into a quadratic programming (QP) problem like conventional SVM. By solving the dual problem of QP, we can easily obtain the solutions of FSVM. The calculation algorithm was proposed by Vapnik [15] and Deng and Tian [16], which can be summarized into the following steps:

- (1) Assume a training set $T\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in (X \times Y)$, where $x_i \in X = R^n$, $y_i \in Y = \{-1, 1\}$, $i = 1, 2, \dots, l$. Introduce a fuzzy membership s_i ($0 \leq s_i \leq 1$, $i = 1, 2, \dots, l$) [22], which can be regarded as the attitude of corresponding training point i toward one class. If $s_i = 0$, it indicates that training point i does not absolutely belong to one class; on the contrary, $s_i = 1$ indicates that training point i is completely within one class. The method of determining the value of s_i for training point i will be described later. For this binary classification (BC) problem, it can be transformed into QP problems with constraint condition including the fuzzy membership. The corresponding dual problem can be expressed as

$$\min_{\alpha} \left(\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^l \alpha_j \right), \quad (1)$$

$$\text{s.t.} \quad \begin{cases} \sum_{i=1}^l y_i \alpha_i = 0, \\ 0 \leq \alpha_i \leq s_i C, i = 1, 2, \dots, l, \end{cases} \quad (2)$$

where C is the regularization parameter, α is the Lagrange multiplier, K is a kernel function, and s_i is the fuzzy membership. When $s_i = 1$, this dual problem is equivalent to standard SVM.

Among several kernels, polynomial kernels and radial basis function (RBF) kernels are often used for pattern classification. While in most cases RBF kernels are found to perform better [27], it can be expressed as

$$K(x, x') = \exp\left(\frac{-\|x - x'\|^2}{\sigma^2}\right), \quad (3)$$

where x is the input vector and σ^2 is the parameter of the RBF kernel function.

By solving the above QP, the optimal solution of this dual problem could be expressed as $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*)^T$, in which the component α_i corresponds to sample T_i .

- (2) The construction of optimal hyperplane based on optimal solution α^* of above QP, the optimal hyperplane can be achieved as

$$w^* \cdot x + b^* = 0, \quad (4)$$

where $w^* = \sum_{i=1}^l y_i \alpha_i^*$ is the weight of sample i , and b^* can be calculated by choosing some positive α_j^* less than C , based on the following formula

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i \cdot x_j). \quad (5)$$

- (3) The determination of the decision function $f(x)$ for binary classification, the decision function $f(x)$ can be expressed as

$$f(x) = \text{sgn}\left(\sum_{i=1}^l y_i \cdot \alpha_i^* K(x_i, x) + b^*\right). \quad (6)$$

The determination of the fuzzy membership s_i is crucial to the performance of FSVM. At present, s_i is mostly determined based on the membership function, whose value is determined by the distance between the sample and its class center [22]. But this method is irrational for dataset with nonspherical-shape distribution. In the current study, we shall employ a new membership function proposed by Du et al. [24] and Ding and Gu [25], where the basic idea is to introduce the class hyperplane to substitute the cluster center. The class hyperplane is defined through the cluster center, and perpendicular to the line across the positive and negative cluster center, as shown in Figure 1. The fuzzy membership is then stated as the distance between the sample and the class hyperplane. In this way, the dependence of membership function on the geometric shape of sample data is greatly reduced, and thus the generalization of FSVM is significantly enhanced.

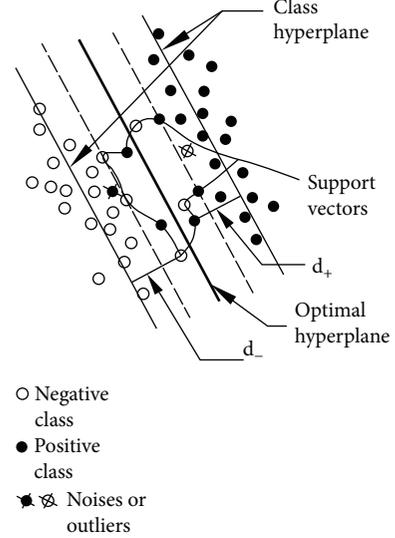


FIGURE 1: Scheme of SVM and FSVM (for 2-dimensional case).

Based on this new method, for nonlinear classification problems, the distance between the sample i and the class hyperplane, denoted by d_{i+} , d_{i-} , respectively, for samples in positive, negative class, as shown in Figure 1, can be obtained by using the kernel technique, as follows:

$$\begin{aligned} d_{i+} &= \left| \frac{1}{n_+} \sum_{j=1}^{n_+} K(x_j, x_i) - \frac{1}{n_-} \sum_{j=1}^{n_-} K(x_j, x_i) - \frac{1}{n_+^2} \sum_{i=1}^{n_+} \sum_{j=1}^{n_+} K(x_j, x_i) \right. \\ &\quad \left. + \frac{1}{n_+ n_-} \sum_{i=1}^{n_+} \sum_{j=1}^{n_-} K(x_j, x_i) \right|, \\ d_{i-} &= \left| \frac{1}{n_+} \sum_{j=1}^{n_+} K(x_j, x_i) - \frac{1}{n_-} \sum_{j=1}^{n_-} K(x_j, x_i) + \frac{1}{n_+^2} \sum_{i=1}^{n_+} \sum_{j=1}^{n_+} K(x_j, x_i) \right. \\ &\quad \left. - \frac{1}{n_+ n_-} \sum_{i=1}^{n_+} \sum_{j=1}^{n_-} K(x_j, x_i) \right|, \end{aligned} \quad (7)$$

where n_+ denotes the number of samples in positive class, n_- denotes the number of samples in negative class, and K is the kernel function.

The maximum distance between positive samples and the class hyperplane was denoted as $D_+ = \max(d_{i+})$, and that for negative samples is denoted as $D_- = \max(d_{i-})$.

Consequently, the membership degree of samples in each class can be defined as:

$$s_i = \begin{cases} 1 - \frac{d_{i+}}{D_+ + \delta} & (\text{for } x_i \text{ within positive class}), \\ 1 - \frac{d_{i-}}{D_- + \delta} & (\text{for } x_i \text{ within negative class}), \end{cases} \quad (8)$$

where δ is a very small positive number to avoid the case $s_i = 0$ and to guarantee $0 < s_i \leq 1$.

It can be seen from (8) that s_i gets smaller when sample x_i locates farther away from the class hyperplane and vice

versa. However, if a sample x_i is far away from its class hyperplane and not within its own class, it will most likely become a noise or outlier point and can severely affect the position of optimal hyperplane in standard SVM. While in FSVM, with the introduction of the fuzzy membership s_i , which becomes very small for noises or outliers, the effect of noises or outliers could then be eliminated on the position of the optimal hyperplane.

In addition, the position of the optimal hyperplane obtained with SVM is significantly influenced by the choice of the kernel function, the regularization parameter, and the relaxation coefficient, which need to be optimized first before applying [17, 18, 29].

2.2. Numerical Procedure

2.2.1. Quantitative Analysis on Parameters and Grading for the Magnitude of RTE.

The geological influence factors on RTE include the composition and structure of rocks at the base of the reservoir, the distribution, attitude, mechanical characteristics, growth level, and the present activity of the fault, as well as the hydrological conditions [30]. Nevertheless, those factors are very difficult to quantify, which brings great difficulty in the quantitative analysis of RTE. In the current study, the analytical hierarchy process (AHP) [28], which is an efficient way to deal with unquantifiable parameters, is employed to quantify these influence factors. Those geological factors are represented by natural number after employing AHP.

2.2.2. Data Normalization.

To prevent the models from domination by input variables with large value for different dimension, the original value of sample data is normalized by the following equation.

$$x_{ij}' = \frac{x_{ij} - \mu_{x_i}}{\sigma_{x_i}}, \quad (9)$$

where x_{ij} is the original value of sample, x_{ij}' is the scaled value of sample, μ_{x_i} is the expectation of sample, and σ_{x_i} is the standard deviation of sample.

2.2.3. The Establishment of the SVM and FSVM Models.

Since there are multiple classes in the prediction models of RTE, the SVM and FSVM models involve multiple classifications. For multiple classifications, SVM and FSVM models could be established by “one-against-all” and “one-against-one” ways [16].

Based on previous mathematical models and numerical procedure, the algorithm of SVM and FSVM is shown in the flow chart of Figure 2, for which computing codes can be programmed.

3. The Prediction Models of RTE Magnitude Based on SVM and FSVM

3.1. Data Construction.

Dataset on 100 large and medium-sized reservoirs in China, covering the 48 well-documented cases of RTE, are collected from reputed enterprises and published literatures for the current study. Based on them, 11

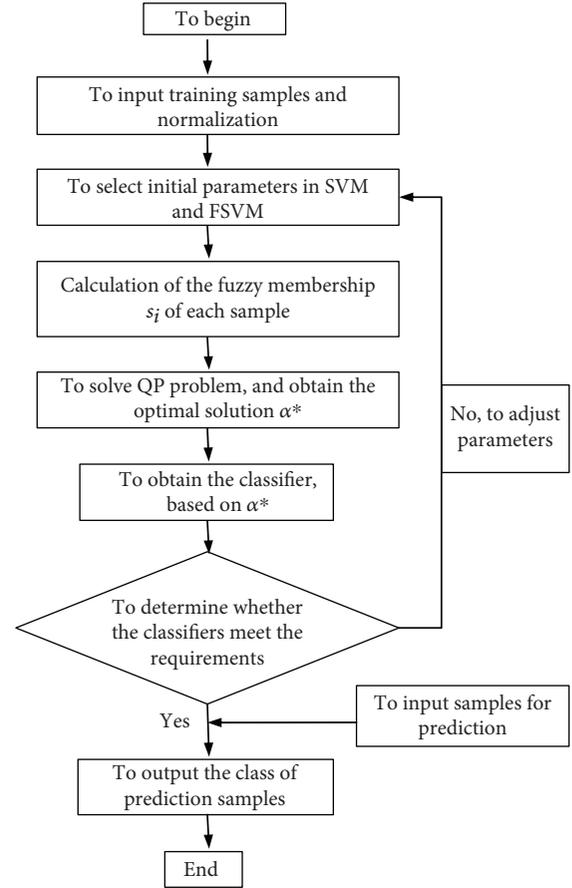


FIGURE 2: Flow chart of FSVM and SVM algorithm.

influence factors, including the geological factors and reservoir parameters, are taken into account in the analysis of RTE, as discussed below.

3.1.1. Dam Height and the Capacity of Reservoir.

Through statistical analysis of RTE cases, it has been found by Jiang et al. [30] that RTE has no obvious relation to dam height and the capacity of reservoir. This means that RTE may happen in any reservoir area regardless of its capacity. For example, the Dengjiaqiao reservoir, located in Hubei province in central China, has triggered an earthquake of Ms 2.2, while the capacity of reservoir is only $400 \times 10^3 \text{ m}^3$ [6].

3.1.2. Lithology at the Reservoir or Epicenter.

To study the influence of lithology at the reservoir area or the epicenter, we have performed statistical analysis on 48 RTE cases. It is found that the lithology in the epicenter is widely distributed, while a sufficient number of samples are essential to enhance the accuracy of SVM analysis. As a result, rocks with close lithology features are classified together. For example, limestone and dolomite are considered together as carbonatite, schist and slate are classified into one category, and pyroclastic rock and tuff are classified as volcanic rock. In this way, more than 10 types of rock are defined for lithology at the reservoir area or the epicenter based on those RTE cases. The representative

TABLE 1: Lithology classification.

Lithology at reservoir area and epicenter	Representative value
Carbonatite	1
Sandstone	2
Clay rock	3
Gneiss	4
Schist and slate	5
Quartzite	6
Marble	7
Granite	8
Basalt	9
Oceanic rock	10
Others	11

TABLE 2: Rock mass integrity.

Rock mass grade	I	II	III	IV	V
Representative value	1	2	3	4	5

values of lithology are shown in Table 1, and those of rock integrity are shown in Table 2.

3.1.3. Type and Characteristics of Fault. The condition of joint and fault has strong influence on the integrity of the rock mass and thus the magnitude of RTE. It is known that surface rock stratum with developed joints and broken rocks has high permeability, for which microearthquake or ultra-microearthquake might be triggered. In relatively deep rock stratum, areas with active or tensional faults have high permeability, and water permeating from reservoir into the deep rock mass might change the pressure distribution of pore water. In this case, the earthquakes triggered are mostly weak and moderate ones. There are normally a large number of faults existing in the reservoir area. Thus, only some main faults are selected as the analyzing subjects considering their scale and influence degree. The representative values related to the type, the development degree, the activity level, and the permeability of the fault are shown in Tables 3–6, respectively.

3.1.4. Tectonic Stress State at the Reservoir Area. It is found that RTE has no obvious relation to tectonic stress state at the reservoir area [30]. This means that RTE could take place not only at areas with active tectonic activity but also at inactive areas. However, enormous elastic strain energy can be stored within rock mass in areas with high local stress. The state of rock mass may become close to rupture or slip, which could lead to relatively strong earthquakes. The representative values of different regional stress state are listed in Table 7.

3.1.5. The Background of Seismicity at Reservoir Area. Based on statistical analysis of RTE, it is found that RTE tends to occur at weak seismic regions, and very few RTE have been reported in relatively strong seismic regions [30]. This is owing to the fact that the strain energy could be easily

TABLE 3: Fault type.

Fault type	Normal fault	Reverse fault	Strike-slip fault
Representative value	1	2	3

TABLE 4: Fault development degree.

Fault development degree	Representative value
Very developed	1
Developed	2
Slightly developed	3
Undeveloped	4
Very undeveloped	5

TABLE 5: Fault activity.

Fault activity	Active	Medium	Inactive
Representative value	1	2	3

TABLE 6: Fault permeability.

Fault permeability	Well	Medium	Bad
Representative value	1	2	3

TABLE 7: Tectonic stress state.

Tectonic stress state	Representative value
Compression and twist	1
Transtension	2
Shear	3
Unobvious	4

TABLE 8: Seismic activity background.

Seismic activity background	Representative value
No seismicity	1
Weak seismicity	2
Moderate seismicity	3
Strong seismicity	4

released via common earthquakes at relatively strong seismic regions. However, RTE is still likely to happen in those areas, under the condition of certain geological structures. The representative value related to the background of seismicity is shown in Table 8.

3.1.6. Grading for the Magnitude of RTE. It is necessary to classify the magnitude of RTE in order to obtain sufficient samples in each class. From the monitoring data on RTE, it is found that the maximum magnitude of RTE is 6.1, denoted in surface wave magnitude (M_s). Combined with the research by Wang et al. [10], the magnitude of RTE is divided into 5 grades, which are strong seismicity ($M_s \geq 6.0$), moderate seismicity ($4.5 \leq M_s < 6.0$), weak

seismicity ($3.0 \leq M_s < 4.5$), microseismicity ($0.0 < M_s < 3.0$), and no seismicity ($M_s = 0.0$). The 5 grades are denoted by 5 growing natural numbers, from 1 to 5, respectively.

It should be noted that, since the number of RTE cases is very limited as 48 ones in China at present, the number of grades should thus be controlled to obtain sufficient number of samples in each class, so that the accuracy of the developed models could be guaranteed.

Some samples (out of the database on 100 reservoirs) on the quantified data of the geological factors and reservoir parameters are shown in Table 9, including earthquake grade classified.

3.2. Numerical Experiments and Results. Since the magnitude of RTE is classified into 5 grades as mentioned above, the SVM and FSVM models are involved in a 5-class classification problem. Considering the limited number of samples, the “one-against-one” method, which can reduce the effect of strongly uneven samples on classifier, is selected to construct the SVM and FSVM models. Consequently, 10 binary classifiers, that is, 1-2 BC, 1-3 BC, 1-4 BC, 1-5 BC, 2-3 BC, 2-4 BC, 2-5 BC, 3-4 BC, 3-5 BC, and 4-5 BC, need to be constructed.

To verify and validate the numerical models established with SVM and FSVM, three numerical tests are carried out. In each test, 90 reservoirs are randomly selected out of the total dataset to provide the data for the training set in the construction of the SVM and FSVM models. Data of the remaining 10 reservoirs are then used as the prediction set to test the accuracy of SVM and FSVM models.

The first training set is used to train the established SVM and FSVM models first. Through analysis and comparison of output results, it is found that the RBF works well as the kernel function. The corresponding values of parameter σ^2 in (3) and regularization parameter C in (2), which could effectively affect the performance of the SVM and FSVM models, are determined through a threefold cross validation method [15].

After the optimization of the parameters, the optimal parameters values of σ^2 and C for all BCs, obtained from the first training set for the SVM and FSVM models, are shown in Table 10. During the process of the threefold cross validation, the accuracy of SVM and FSVM models for the training set and the validation set is shown and compared in Table 11. The accuracy is defined as the matching percentage of the predicted set of labels and the true set of labels in each BC. Based on the optimal parameters, the output of the SVM model is completely identical with the sample data for the training set. While for the validation set, the accuracy remains above 85%, except for 3-4 BC with the accuracy of 61.97%. At the same time, the accuracy of FSVM with the optimal parameters remains above 85% for training set and above 87.59% for the validation set except for 3-4 BC with accuracy of 64.74%. Consequently, under this dataset, the performance of SVM model is slightly superior to that of FSVM model.

Based on the optimal parameters above, the SVM and FSVM models should be trained again by the training set; results show that no error happens out of the 10 BCs for training set, while for the FSVM model, only 2 BCs do not

exist error, the maximum error rate reaching 14.29% and the average error rate amounting to 4.67%.

We then use the trained SVM and FSVM models to predict the magnitude of RTE of the 10 reservoirs in the prediction set as described above. The results obtained from the FSVM and SVM models are shown in Table 12.

It can be seen from Table 12 that, except for Ankang reservoir and Liujiashan reservoir, the prediction results of the 10 reservoirs are identical for SVM and FSVM models in the current experiment. Among them, Chongbahu reservoir, Fengjiashan reservoir, and Bikou reservoir have no reported cases of RTE (denoted by 5 according to the quantification of RTE magnitude as described before). Both of the two models provide accurate results of RTE predictions for those reservoirs. In addition, the two models also provide accurate predictions on the grade of RTE for some of reservoirs with previous RTE cases, including Wujiangdu reservoir, Three Gorges reservoir, and Zipingpu reservoir, as shown in Table 12. While for Ankang reservoir and Liujiaxia reservoir, where no RTE has been monitored, the prediction grade is grade 3 ($3.0 \leq M_s < 4.5$) by FSVM model for the former and by SVM model for the latter, higher than the actual data, which suggests weak seismicity; whereas the prediction is grade 5 by SVM model for the former and by FSVM model for the latter, identical with monitored data. For Danjiangkou reservoir, previous RTE with the magnitude of $M_s 4.7$ has happened. The prediction grade is grade 3, corresponding to RTE with $3.0 \leq M_s < 4.5$. The prediction grade by the two models is also 3 for Wuxijiang reservoir, where actual RTE with the magnitude of $M_s 2.8$ was observed. To sum up, 7 samples have been predicted accurately by SVM and FSVM models among the 10 prediction samples in the current experiment. The errors for the left 3 samples are within small range.

In order to test the robustness of the models, another two training processes are undertaken with data selected in the same way as mentioned above, called group 2&3. The values of σ^2 and C remain the same as those obtained from the optimization processes with the first training set for the SVM and FSVM models, respectively.

After applying the second training set on the SVM model, the results of training process show that no BC exists error out of the total 10 BCs. The accuracy for the training set reaches to 100%. The trained SVM model is then employed to predict the magnitude of RTE of the 10 reservoirs in the prediction set, for which the results are shown in Table 13. It can be seen that 8 results have been predicted correctly, with the accuracy rate reaching 80%. The error of the results is kept in a small range of 1 grade.

While for the FSVM model, in the training process, only 2 BCs do not exist misjudgment among the 10 BCs. The maximum and average error rates are 14.29% and 4.74%, respectively. The prediction results with FSVM in the prediction set are also shown in Table 13, where it can be seen that only two of the results (on Shenwo reservoir and Dongfeng reservoir) are different from that of SVM by one grade, on which FSVM has provided accurate prediction with the accuracy of 80% as well, in spite of a little bit lower accuracy in the training set than that with the SVM model.

TABLE 9: Quantitative index on reservoirs.

Reservoir name	Height (m)	Storage capacity ($\times 10^8 \text{ m}^3$)	Water depth before dam (m)	Lithology at reservoir area and epicenter	Rock mass grade	Fault type	Fault development degree	Fault activity	Fault permeability	Tectonic stress state	Seismic activity background	Seismicity grade
Fouziling	74	4.7	65	1	3	2	1	1	1	2	3	2
Xinfengjiang	105	115	97	8	3	2	1	1	1	3	2	1
Nanshui	81.5	10.5	75	1	3	3	2	1	1	3	1	4
Nanchong	45	0.15	35	1	3	1	2	1	1	1	1	4
Huangshi	40.5	6.12	36	1	3	1	2	1	1	1	1	4
Qianjiang	50	0.168	44	1	3	1	2	1	1	1	1	3
Tuoling	63.5	71.7	47	1	3	2	1	1	1	2	1	3
Shenwo	50	5	36	1	3	3	2	1	1	3	2	2
Shiquan	65	4.7	53	5	3	1	2	1	1	1	2	3
Tongjiezi	82	2	70	9	3	2	2	1	1	2	2	4
Manwan	126	10.06	100	8	3	3	2	1	1	3	2	3
Shuikou	100	23.4	85	8	3	1	2	1	1	1	2	3
Geheyan	151	34	145	1	3	1	2	3	1	1	2	4
Dongfeng	168	10.3	160	1	3	1	2	3	1	1	2	4
Ertan	240	58	220	9	3	2	2	3	2	2	2	3
Tianshengqiao	178	102.6	155	1	3	1	2	3	3	1	2	3
Daqiao	92.2	6.58	85	8	3	2	2	1	1	2	4	2
Xiaolangdi	154	126.5	148	2	3	3	2	1	2	3	2	3
...					
Xudong	87	3.64	50	2	3	2	2	3	1	2	2	5

TABLE 10: Optimal parameters for respective BCs.

BCs	SVM		FSVM	
	σ^2	C	σ^2	C
1-2 BC	0.0039	2.0	0.0039	2.0
1-3 BC	0.0039	2.0	0.0039	2.0
1-4 BC	0.0039	2.0	0.0039	2.0
1-5 BC	0.0039	2.0	0.0039	2.0
2-3 BC	4.0000	1024	0.3536	128
2-4 BC	1.0000	4.0	1.4142	16.0
2-5 BC	0.5000	2.0	0.7071	2.0
3-4 BC	0.0156	2.0	0.0884	128.0
3-5 BC	0.7071	16.0	0.5000	8.0
4-5 BC	0.5000	16.0	0.5000	4.0

TABLE 11: Optimal output for respective BCs in cross validation test.

BCs	SVM		FSVM	
	Training set Accuracy (%)	Validation set Accuracy (%)	Training set Accuracy (%)	Validation set Accuracy (%)
1-2 BC	100	100	85	100
1-3 BC	100	100	95.69	100
1-4 BC	100	100	93.64	100
1-5 BC	100	100	97.70	100
2-3 BC	100	88.89	96.39	88.89
2-4 BC	100	85.71	97.62	90.48
2-5 BC	100	97.92	97.92	97.92
3-4 BC	100	61.97	100	64.74
3-5 BC	100	93.79	98.25	87.59
4-5 BC	100	94.74	98.25	94.74

TABLE 12: Prediction results of RTE with group 1.

Reservoir name	Observed seismicity grade	Predicted seismicity grade	
		SVM	FSVM
Chongbahu	5	5	5
Ankang	5	5	3
Fengjiashan	5	5	5
Liujiashan	5	3	5
Bikou	5	5	5
Danjiangkou	2	3	3
Wuxijiang	4	3	3
Wujiangdu	3	3	3
Three Gorges	3	3	3
Zipingpu	3	3	3

Based on the training set in group 3, no error is found for the 10 BCs in the trained SVM model. While in the FSVM model, no error is found in 2 BCs. The maximum and average error rates are 14.29% and 4.97%, respectively, for the left

TABLE 13: Prediction results of RTE with group 2.

Reservoir name	Observed seismicity grade	Predicted seismicity grade	
		SVM	FSVM
Xinanjiang	5	5	5
Jinshuitan	5	5	5
Gezhouba	5	5	5
Fengshuba	5	5	5
Tuoling	3	3	3
Shenwo	2	2	3
Dongfeng	4	3	4
Ertan	3	3	3
Panjiakou	4	5	5
Lijiaxia	3	3	3

TABLE 14: Prediction results of RTE with group 3.

Reservoir name	Observed seismicity grade	Predicted seismicity grade	
		SVM	FSVM
Gongzui	5	5	5
Jiangkou	5	5	5
Huanglongtan	5	5	5
Shizitan	5	5	5
Qianjin	3	3	4
Huangshi	4	4	4
Kezier	2	2	2
Yunpeng	3	3	3
Xiaowan	3	3	3
Nanshui	4	2	2

8 BCs. The results of the trained SVM and FSVM models applied in the prediction of the magnitude of 10 RTE in the prediction set are shown in Table 14. It can be seen that the accuracy is 90% for SVM and 80% for FSVM, with errors limited within 1~2 grades for the two models.

Through the results obtained with the SVM and FSVM models in the three tests, it can be seen that both the two models are proved to be effective in the prediction of the magnitude of RTE with high accuracy. Through comparison, it is further found that SVM model is slightly superior to FSVM in the current study on the prediction of the magnitude of RTE with dataset on RTE in China. This is in contradiction with the findings in [24]. In theory, FSVM model is superior to SVM in minimizing the effect of noises or outliers by introducing the fuzzy membership. However, it cannot be concluded that the FSVM model is superior to SVM in general. If the training samples are of high reliability and high discreteness, the introduction of the fuzzy membership might decrease the discreteness of the data and underestimate the effect of samples far away from the optimal hyperplane and thus reduce the accuracy of classifier. This might explain the reason of the findings in the current study in terms of the comparison between the two models. Therefore,

the choice between SVM and FSVM models in the application of classifying problems should be made according to the discreteness of data. For samples with high discreteness, SVM is likely to be superior to FSVM and vice versa.

4. Conclusions

In this study, we have applied the SVM and FSVM methods as an effective approach in the prediction of the magnitude of RTE. Dataset on 100 reservoirs in China, covering the 48 RTE cases, are collected and used as the database. The main 11 influence factors of RTE, including lithology, rock mass integrity, fault characteristic, tectonic stress state, and seismic activity background, are summarized and quantified with the analytical hierarchy process (AHP). Multiple numerical tests are carried out to test the performance of the established SVM and FSVM models in the prediction of the magnitude of RTE. The main conclusions obtained are summarized below.

Both the SVM and FSVM models are found to be effective in the prediction of the magnitude of RTE with high accuracy. Moreover, the results obtained by SVM are of slight higher accuracy than those by FSVM for the current analysis of RTE cases in concern. This might be caused by the introduction of the fuzzy membership in FSVM. It is effective in the minimization of the influence of the noises or outliers in some problems and yet is not favorable when dealing with samples with high discreteness. Therefore, the discreteness of samples needs to be considered when choosing between the SVM and FSVM models in practical problems. Based on the current study, the SVM model is found to be slightly superior to FSVM in the field of RTE prediction.

It should be noted that the magnitude of RTE is analyzed in terms of its maximum value in the present study. In general, besides the main shock, there could be multiple foreshocks and aftershocks in the process of RTE, whose magnitude and epicenter are quite different with those of the main shock. To predict the magnitude of foreshocks and aftershocks, the SVM and FSVM could also be employed; however, the data has to be established based on new data collected on the features of faults and locations of foreshocks and aftershocks. Furthermore, the SVM and FSVM models for the prediction of the magnitude of RTE are constructed based on the mass data of engineering geology and hydrology geology conditions. Thus, in order to obtain prediction results with high accuracy, it is essential to collect reliable data of sufficient samples on the main factors related to RTE, including the geological conditions at the reservoir area and the features of water permeation from reservoir to deep rock.

In addition, the current work is not on the prediction of the time history of RTE, which plays an important role in engineering practices. However, in theory, if we could collect and monitor sufficient data on the evolution of hydrology parameters and geological properties of rocks in a certain reservoir area, obtain their trends via means of numerical analysis, and make the analogy with the time series of RTE cases, the evolution and prediction of RTE in this area could be

studied intensively. Therefore, the current study also provide a creative idea for the prediction of the time history of RTE, while the main challenge lies in the access to reliable evolution trends of the multiple hydrological and geological factors involved.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are grateful to the reviewers for their very useful comments and suggestions. This research was financially supported by the National Natural Science Foundation of China (Grant no. 51669008).

References

- [1] L. X. Yi, G. Wang, and L. Li, "Hydrogeological structure and reservoir induced seismicity," *Hydrogeology & Engineering Geology*, vol. 31, no. 2, pp. 29–32, 2004.
- [2] T. Chang, "Forming conditions of karst collapsing induced earthquake," *Hydrogeology & Engineering Geology*, vol. 5, pp. 42–45, 2006.
- [3] H. K. Gupta, "A review of recent studies of triggered earthquakes by artificial water reservoirs with special emphasis on earthquakes in Koyna, India," *Earth-Science Reviews*, vol. 58, no. 3–4, pp. 279–310, 2002.
- [4] D. S. Carder, "Seismic investigations in the boulder dam area, 1940 – 1944, and the influence of reservoir loading on earthquake activity," *Bulletin of the Seismological Society of America*, vol. 35, pp. 175–192, 1945.
- [5] D. Chen, Y. Wang, and X. Zeng, "A study of reservoir-induced earthquake of three gorges project," *Chinese Journal of Rock Mechanics and Engineering*, vol. 27, no. 8, pp. 1513–1524, 2008.
- [6] Y. Ding, *The Reservoir Induced Earthquake*, Geological press, Beijing, 1989.
- [7] W. Ding, *Mechanical Problems in Earthquake Prediction*, Qinghua University Press, Beijing, 2012.
- [8] S. J. Kenner and P. Segall, "A mechanical model for intraplate earthquakes: application to the New Madrid Seismic Zone," *Science*, vol. 289, no. 5488, pp. 2329–2332, 2000.
- [9] J. Su, Q. Xia, Y. Wang, X. Wang, and Z. Ju, "Preliminary prediction on Datengxia reservoir induced seismicity," *The Chinese Journal of Geological Hazard and Control*, vol. 4, no. 4, pp. 42–51, 2009.
- [10] Q. Wang, Y. Yao, J. Xia, J. Li, and D. Wang, "Application of statistical forecasting models to the prediction of the three gorges reservoir induced seismicity," *Seismology and Geology*, vol. 31, no. 2, pp. 287–194, 2009.
- [11] J. Xia, C. Li, and X. Wang, "Application of neural network theory in prediction of earthquake induced by TGP reservoir," *Yangtze River*, vol. 38, no. 2, pp. 8–10, 2007.

- [12] G. Habibagahi, "Reservoir induced earthquakes analyzed via radial basis function networks," *Soil Dynamics and Earthquake Engineering*, vol. 17, no. 1, pp. 53–56, 1998.
- [13] P. Samui and D. Kim, "Applicability of artificial intelligence to reservoir induced earthquakes," *Acta Geophysica*, vol. 62, no. 3, pp. 608–619, 2014.
- [14] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [15] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998.
- [16] N. Deng and Y. Tian, *Support Vector Machine-Theory, Algorithm and Expansion*, Science Press, Beijing, 2009.
- [17] S. K. Das, P. Samui, and A. K. Sabat, "Prediction of field hydraulic conductivity of clay liners using an artificial neural network and support vector machine," *International Journal of Geomechanics*, vol. 12, no. 5, pp. 606–611, 2012.
- [18] M. Pal and S. Deswal, "Modeling pile capacity using support vector machines and generalized regression neural network," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 134, no. 7, pp. 1021–1024, 2008.
- [19] J. Lin and C. Cheng, "Application of support vector machine method to long-term runoff forecast," *Journal of Hydraulic Engineering*, vol. 37, no. 6, pp. 681–686, 2006.
- [20] H. Zhao, *Support Vector Machine Analysis in Rock and Soil Mechanics and Engineering*, Coal Industrial Press, Beijing, 2008.
- [21] K. Gopalakrishnan and S. Kim, "Support vector machines approach. to HMA stiffness prediction," *Journal of Engineering Mechanics*, vol. 137, no. 2, pp. 138–146, 2011.
- [22] C.-F. Lin and S. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002.
- [23] H. Tang, Y. Liao, F. Sun, and H. Xie, "Fuzzy support vector machine with a new fuzzy membership function," *Journal of Xi'an Jiao Tong University*, vol. 43, no. 7, pp. 40–43, 2009.
- [24] Z. Du, S. Liu, and X. Qi, "Fuzzy support vector machine with new membership function," *Journal of System Simulation*, vol. 21, no. 7, pp. 1901–1903, 2009.
- [25] S. Ding and Y. Gu, "A Fuzzy Support Vector machine algorithm with dual membership based on hypersphere," *Journal of Computational Information Systems*, vol. 7, no. 6, pp. 2028–2034, 2011.
- [26] E. C. C. Tsang, D. S. Yeung, and P. P. K. Chan, "Fuzzy support vector machines for solving two-class problems," in *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, pp. 1080–1083, Xi'an, China, November 2003.
- [27] S. Abe, "Fuzzy support vector machines for multilabel classification," *Pattern Recognition*, vol. 48, no. 6, pp. 2110–2117, 2015.
- [28] T. L. Satty, *The Analytic Hierarchy Process*, McGraw Hill, Inc., New York, NY, USA, 1980.
- [29] J. Valyon and G. Horvath, "Selection methods for extended least squares support vector machines," *International Journal of Intelligent Computing and Cybernetics*, vol. 1, no. 1, pp. 69–93, 2008.
- [30] H. Jiang, X. Zhang, and X. Shan, *Research on Prediction Method and Statistic Characteristics of Reservoir Induced Earthquake in China Mainland*, Seismic Press, Beijing, 2014.

Research Article

A Novel Fault Diagnosis Method for Rolling Bearing Based on Improved Sparse Regularization via Convex Optimization

Dongjie Zhong,^{1,2} Cancan Yi ,^{1,2,3} Han Xiao ,^{1,2} Houzhuang Zhang,^{1,2} and Anding Wu⁴

¹Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China

²Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

³Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China

⁴Wenzhou Special Equipment Inspection and Research Institute, Wenzhou 325007, China

Correspondence should be addressed to Cancan Yi; meycancan@wust.edu.cn

Received 12 January 2018; Revised 9 April 2018; Accepted 18 April 2018; Published 3 July 2018

Academic Editor: Minvydas Ragulskis

Copyright © 2018 Dongjie Zhong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Structural health monitoring and fault state identification of key components, such as rolling bearing, located in the mechanical main drive system, have a vital significance. The acquired fault signal of rolling bearing always presents the obvious nonlinear and nonstationary characteristics. Moreover, the concerned features are submerged in strong background noise. To handle this difficulty, a novel fault signal denoising scheme based on improved sparse regularization via convex optimization is proposed to extract the fault feature of rolling bearing. In this paper, the generalized minimax-concave (GMC) penalty is firstly researched to promote the sparsity of signal, which is based on traditional L_1 -norm and Huber function. It is designed to estimate the sparse solutions more accurately and maintain the convexity of the cost function. Then, the GMC penalty is extended to 1-D first-order total variation (TV) as nonseparability and nonconvex regularizer. Thus, a convex optimization problem, which involves a quadratic data fidelity term and a convex regularization term, is developed in this paper. To accelerate the convergence of the algorithm, it is solved by forward-backward (FB) iterative algorithm and thus the denoised signal can be obtained. In order to demonstrate its performance, the proposed method is illustrated for numerical simulation signal and applied in the feature extraction of the measured rolling bearing vibration signal.

1. Introduction

In the field of prognostics and health management (PHM) to mechanical equipment, the actual collected vibration signal contains wealthy information about operating status [1–3]. It is worthy to note that the faulty characteristic information can always be reflected from the measured signal. Rolling bearing is the key component in the main transmission system, and its operating performance is directly related to the status of the entire plant. When some faults occur in them, it will inevitably affect the work safety and production efficiency [4, 5]. Therefore, analyzing the possible fault characteristics of the rolling bearings has an important practical significance [6–8]. Modern signal processing

methods provide the main technical means of rolling bearing fault diagnosis.

Generally, the acquired signal is coupled by useful information and the strong noisy component, which has the typical nonlinear and nonstationary characteristics. Hence, the main task of mechanical equipment structural health monitoring is to effectively remove the noise component and improve the signal-to-noise ratio [9, 10]. Traditional signal processing method is based on inner product operator, which is built on the analyzed signal and basis function. The most representative ones are short-time Fourier transform (STFT) [11] and wavelet transform (WT) [12, 13]. Since the size of the analyzed window is fixed, STFT lacks sufficient capacity to deal with the complex nonstationary signals. The

performance of WT depends on the selection of wavelet basis function and the decomposition level. Then, local mean decomposition (LMD) algorithm is proposed as an adaptive time-frequency analysis method [14]. However, it still has the problem of mode aliasing. Variational mode decomposition (VMD) [15, 16] is proposed based on Wiener filtering, one-dimensional Hilbert transform, and heterodyne demodulation analysis. It is still affected by the selection of penalty parameter and the number of signal components. Presently, synchrosqueezing transform (ST) has attracted much attention due to its properties in time-frequency reassignment [17, 18]. Nevertheless, its performance is unsatisfied since the cross-term interference and poor scale separation.

Essentially, the signal denoising can be achieved by the calculation of a sparse approximate solution to the measured vibration signal. Thus, the novel denoising method based on convex optimization and sparsity has publicly employed in signal processing and image enhancement [19, 20]. It has been successfully applied in the field of mechanical fault diagnosis [21], spectral data processing, and baseline correction [22]. Total variation (TV) is the main content of convex optimization algorithm, which involves a quadratic data fidelity term and a convex regularization term [23]. It is developed on sparse signal models. Based on that, the first-order TV [24] and higher degree TV (HDTV) [25] regularization has been researched. However, the experimental analysis demonstrated that it may work when the signal is piecewise constant and it often produces undesirable staircase artifacts. Subsequently, wavelet total variation denoising [26] is proposed to improve the denoising performance, while the estimation of noise variance and penalty function selection is an inevitable question. Using either (1) L_1 -norm regularization and convex optimization or (2) nonconvex regularization and nonconvex optimization, the calculation of a sparse approximate solution to a linear system of equations is often performed. However, it tends to provide solutions that deviate the real values. Afterwards, a new penalty is suggested, which is a multivariate generalization of the minimax-concave (MC) penalty and is defined as generalized MC (GMC) penalty [27, 28]. GMC penalty simultaneously involves the generalized Huber function and the regular L_1 -norm regularization. In other words, it is an innovative nonseparability and nonconvex penalty function, which is using for achieving the sparse enhancement and signal smoothing. A nonseparable penalty has more superiority in meeting the requirement of preserving the convexity of the objective function [29]. The published result fully indicates that the GMC penalty is obviously superior to the common penalty function in noise artifacts removing for time-frequency analysis [27].

For early and weak fault detection, it has the feature of latent and dynamic response as the multifactor coupling and complex transmission path [30]. Since the fault is in its early state, the energy generated during the operation is low, and the signal that can be received by the sensor after attenuation is extremely weak [31]. Therefore, early failure signal is easily submerged by background noise, and effective extraction of these characteristics has been a difficult

problem. Inspired by the idea of jointing the nonconvex penalty and convex optimization algorithm, the traditional L_1 -norm regularization term is replaced by GMC penalty in the TV denoising scheme in this paper so as to effectively realize fault state identification. In order to verify the rationality and feasibility of the proposed method, it is used to analyze the numerical simulation and the actual fault vibration signal of rolling bearings in the bearing test rig. The result demonstrated that the proposed method has obvious advantages over traditional methods such as WT, TV, and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [32].

The rest of the paper is organized as follows. In Section 2, the basic ideas of GMC penalty and the researched convex optimization denoising algorithm are introduced. The simulation signal analysis is described in Section 3. The fault-bearing data from bearing test rig is analyzed in Section 4. Section 5 gives the final conclusions.

2. Theory Descriptions

2.1. Generalized Minimax-Concave (GMC) Penalty. Generally, the objective function is convex, and the optimization problem of constraint variable value in a convex set is called convex optimization problem. The penalty function is a measure of constraint violation, which makes the constraint zero when the constraint is satisfied. The one-dimension time series $x \in R^N$ can be simply expressed as $x = (x_1, x_2, \dots, x_N)$. Then, the L_1 -norm and L_2 -norm is denoted as $\|x\|_1 = \sum_n |x_n|$ and $\|x\|_2 = (\sum_n |x_n|^2)^{1/2}$ respectively.

The Huber function $s(x)$ can be defined as:

$$s(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq 1 \\ |x| - \frac{1}{2}, & |x| \geq 1. \end{cases} \quad (1)$$

The minimax-concave (MC) penalty $\phi(x)$ is defined as:

$$\phi(x) = \begin{cases} |x| - \frac{1}{2}x^2, & |x| \leq 1 \\ \frac{1}{2}, & |x| \geq 1. \end{cases} \quad (2)$$

Figures 1(a) and 1(c) have illustrated the Huber function and MC penalty. Hence, the relationship between Huber function and MC penalty can be described as

$$\phi(x) = |x| - s(x). \quad (3)$$

Defining the scaled versions of the Huber function and MC penalty is convenient. For $b \neq 0$, the scaled Huber function $s_b(x)$ is determined by

$$s_b(x) = \begin{cases} \frac{1}{2}b^2x^2, & |x| \leq \frac{1}{b} \\ |x| - \frac{1}{2b^2}, & |x| \geq \frac{1}{b}. \end{cases} \quad (4)$$

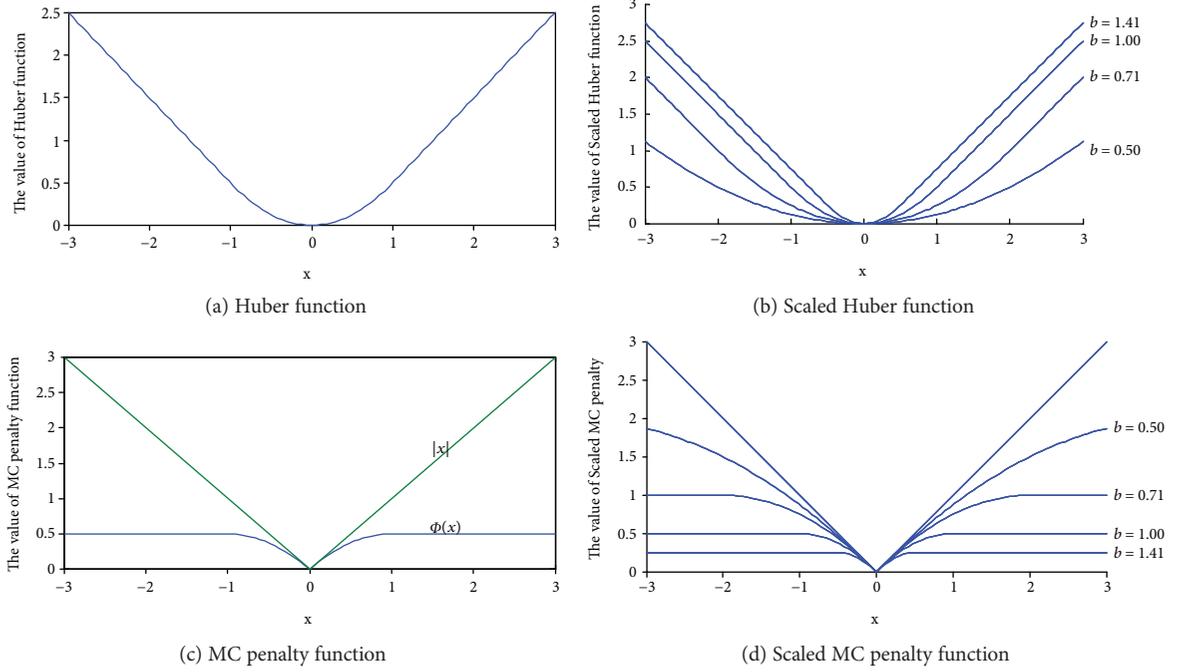


FIGURE 1: Different penalty functions.

Similarly, the scaled MC penalty function $\phi_b(x)$ is given by

$$\phi_b(x) = \|x\|_1 - s_b(x). \quad (5)$$

Figures 1(b) and 1(d) demonstrate the scaled Huber function and scaled MC penalty, respectively. Let $\mathbf{B} \in R^{M \times N}$ and we next define the generalized Huber function $S_B(x)$ as

$$S_B(x) = \inf_{v \in R^N} \left\{ \|v\|_1 + \frac{1}{2} \|\mathbf{B}(x - v)\|_2^2 \right\}. \quad (6)$$

Then, we propose a multivariate generalization of the MC penalty. The basic idea is to generalize (6) using the L_1 -norm and the generalized Huber function. Thus, we define the generalized MC (GMC) penalty function $\psi_B(x)$ as follows:

$$\psi_B(x) = \|x\|_1 - S_B(x). \quad (7)$$

2.2. The Denoising Algorithm Based on Convex Optimization.

Let $y \in R^M$ be the original observed signal, $\mathbf{A} \in R^{M \times N}$ and $\lambda > 0$ is the regularization parameter. A commonly used approach to obtain an optimal sparse approximate solution is to minimize the following objective function [24]:

$$Q(x) = \frac{1}{2} \|y - \mathbf{A}x\|_2^2 + \lambda \|x\|_1. \quad (8)$$

Undoubtedly, this is a typical convex optimization problem of first-order TV denoising, which is comprised by a quadratic fidelity term and an L_1 -norm regularization term. To achieve the signal denoising and early fault feature extraction, a sparse regularization denoising algorithm based on convex optimization is proposed in this paper. Specifically speaking, the traditional L_1 -norm regularization item is

replaced by the GMC penalty term and it is extended to the first-order TV denoising scheme. Thus, we redefine the objective function $F : R^N \rightarrow R$ as follows:

$$F(x) = \frac{1}{2} \|y - \mathbf{A}x\|_2^2 + \lambda \psi_B(\mathbf{D}x), \quad (9)$$

where \mathbf{A} is an oversampled inverse discrete Fourier transform, $\psi_B: R^N \rightarrow R$ is the generalized MC (GMC) penalty defined by (7), λ is the regularization parameter, and the symbol of $\mathbf{D}x$ stands for the TV operator. The bidiagonal matrix $\mathbf{D} \in R^{(n-1) \times n}$ is defined as

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & M & M & \\ & & & & -1 & 1 \end{bmatrix}. \quad (10)$$

The penalty ψ_B is parameterized by a matrix \mathbf{B} and the convexity of $F(x)$ depends on \mathbf{B} being suitably prescribed. It also should be pointed out that the choice of \mathbf{B} will depend on \mathbf{A} . If

$$\mathbf{B}^T \mathbf{B} \leq \frac{1}{\lambda} \mathbf{A}^T \mathbf{A}, \quad (11)$$

then $F(x)$ is a convex function. It is easy to satisfy the convexity condition. Given the matrix $\mathbf{A} \in R^{M \times N}$ ($N \geq M$), we may simply set

$$\mathbf{B} = \sqrt{\frac{\gamma}{\lambda}} \mathbf{A}, \quad (12)$$

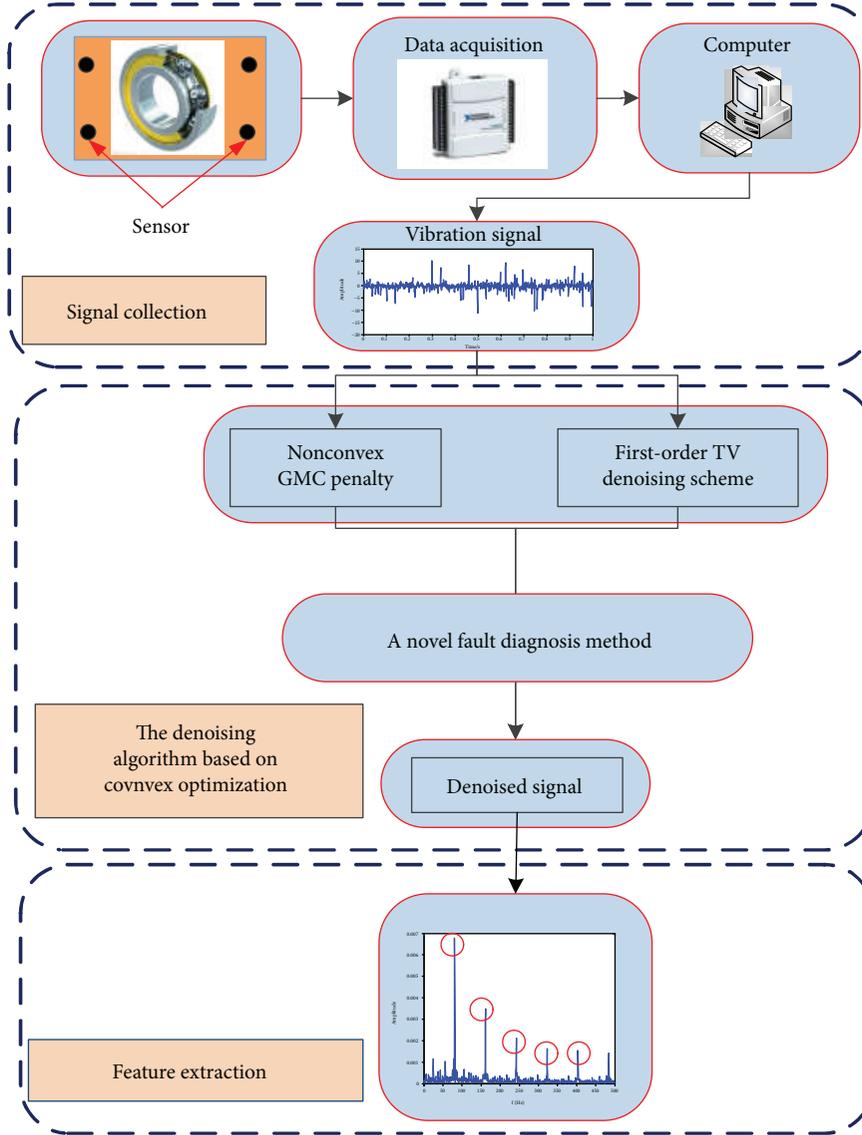


FIGURE 2: The flowchart of the method presented in this paper.

where M is the signal length and N is the transform length. It should be pointed out that $0 \leq \gamma \leq 1$ always meets the convexity condition. Generally, we use a nominal range of $0.5 \leq \gamma \leq 0.8$ so as to obtain better performance. In order to minimize the cost function by using the approximate algorithms, we rewrite it as a saddle-point problem:

$$(x^{\text{opt}}, v^{\text{opt}}) = \arg \min_{x \in \mathbb{R}^N} \max_{v \in \mathbb{R}^N} F(x, v), \quad (13)$$

where $F(x, v) = 1/2 \|y - Ax\|_2^2 + \lambda \|Dx\|_1 - \lambda \|v\|_1 - (\gamma/2) \|A(Dx - v)\|_2^2$ is the saddle function and x^{opt} presents the denoised signal.

According to abovementioned, it is obvious that the selection of regularization parameter λ also has a significant influence in denoising performance. Commonly, we chose $\lambda = 2$ to achieve the convergence of algorithm. Therefore,

the forward-backward (FB) iterative algorithm can be used to solve the problem $F(x, v)$ of this kind of saddle point [33]. The resulting iterative threshold algorithm uses the soft-threshold function, which is defined as

$$\text{soft}(y; \lambda) := \begin{cases} 0 & |y| \leq \lambda \\ (|y| - \lambda) \text{sign}(y) & |y| \geq \lambda. \end{cases} \quad (14)$$

The flowchart of the presented method in this paper is plotted in Figure 2.

3. Simulation Signal Analysis

Without loss of generality, the numerical signal analysis is used to simulate the rolling bearing fault feature identification. It is composed by frequency modulation signal,

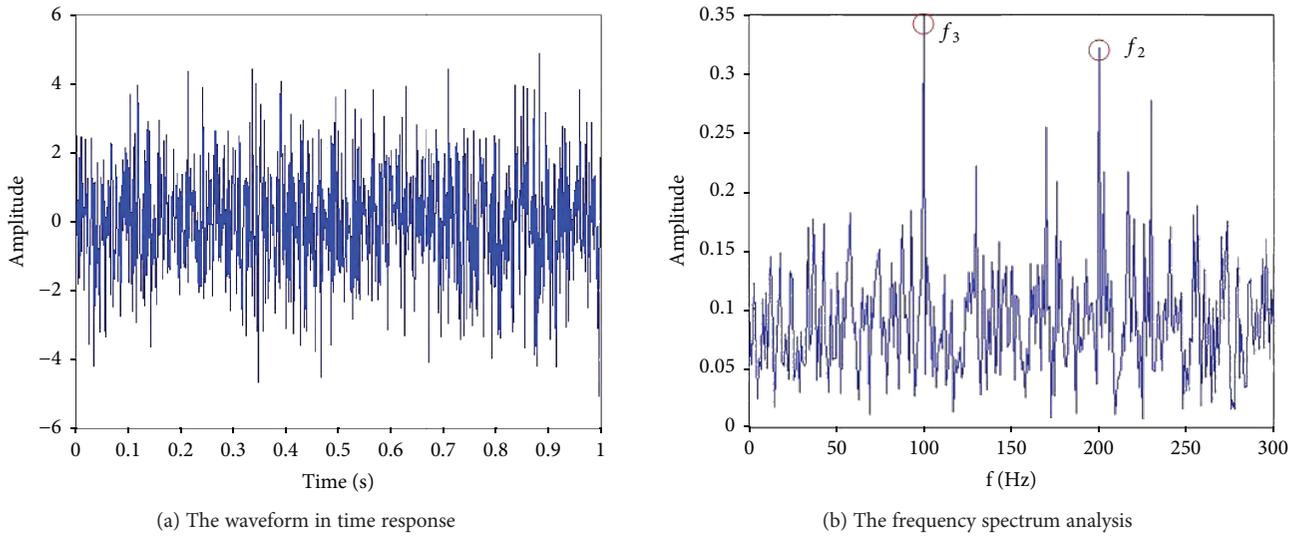


FIGURE 3: The synthetic signal of simulation.

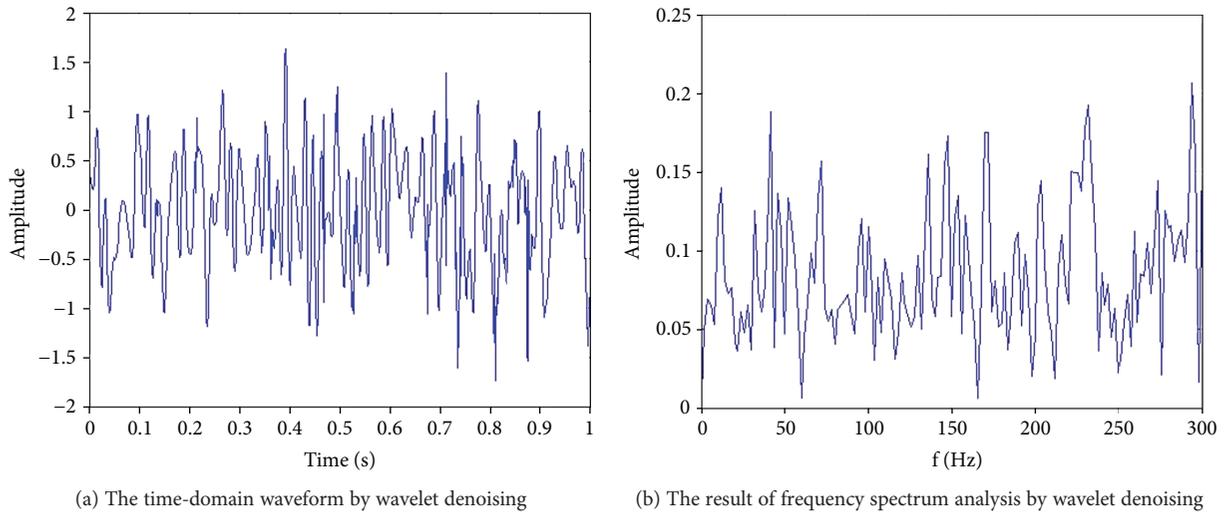


FIGURE 4: The result provided by wavelet denoising algorithm.

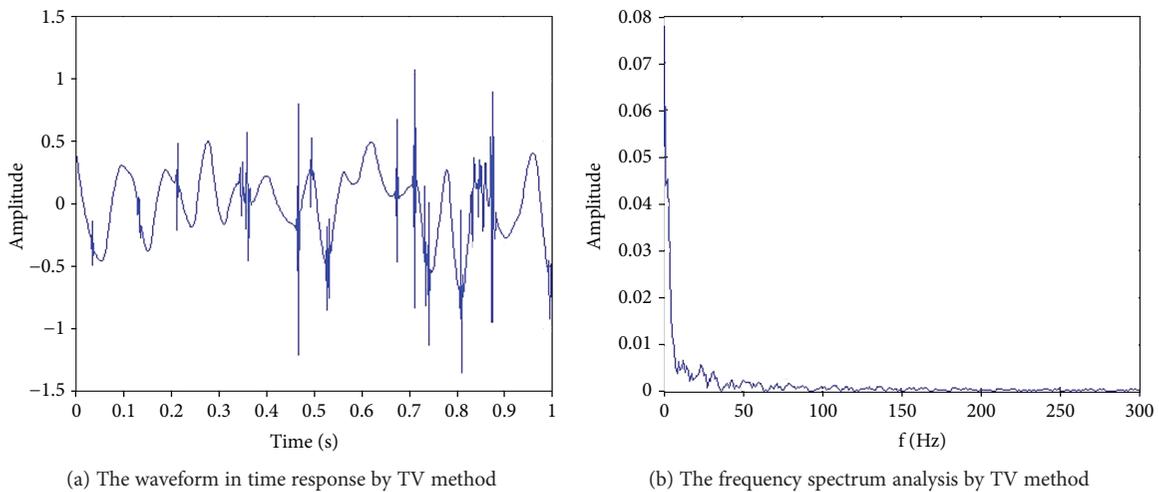


FIGURE 5: The result provided by the typical total variation (TV) denoising.

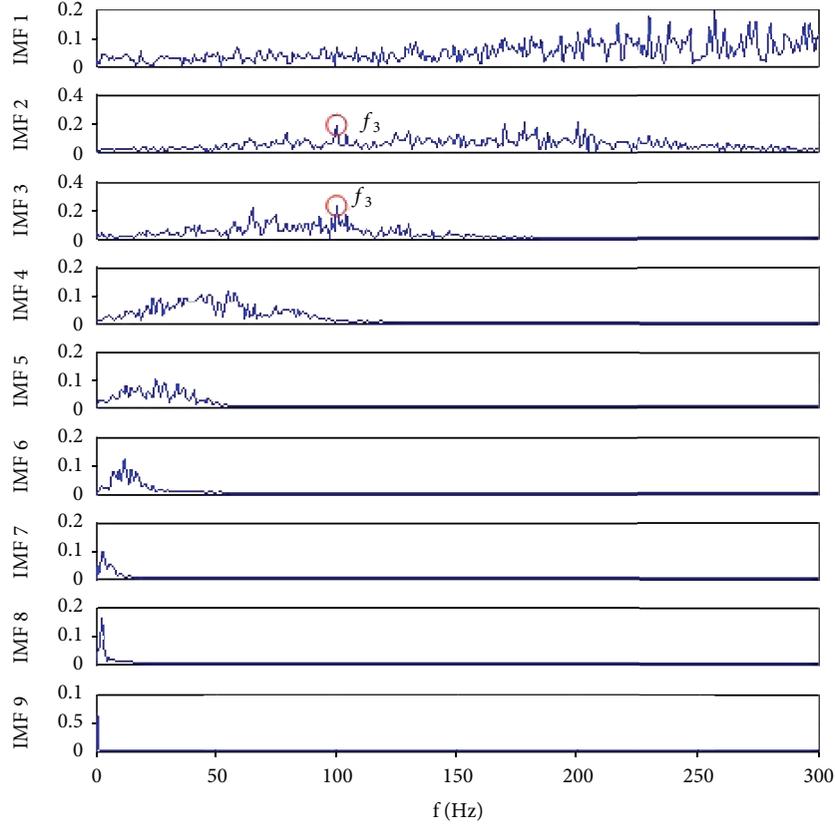


FIGURE 6: The result provided by CEEMDAN for simulation signal analysis.

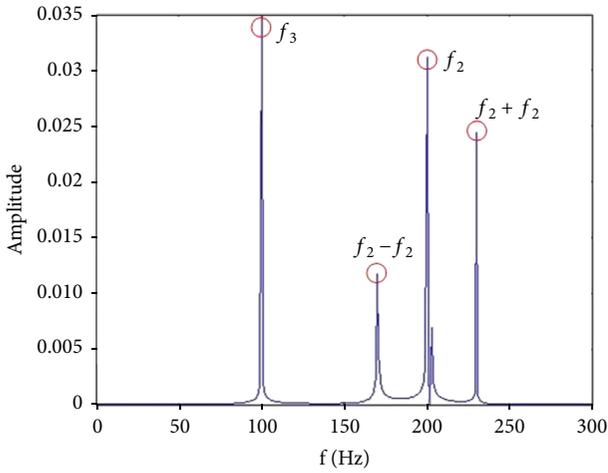


FIGURE 7: The result provided by the proposed method.

harmonic signal, and strong background noise components. The numerical simulation signal can be expressed as follows:

$$\begin{aligned}
 s_1 &= 0.5 \cos(2\pi f_2 t + \sin(2\pi f_1 t)), \\
 s_2 &= 0.3 \sin(2\pi f_3 t), \\
 s_3 &= 1.5 \cdot \text{randn}(1, 1024), \\
 y &= s_1 + s_2 + s_3,
 \end{aligned} \tag{15}$$

TABLE 1: The proposed algorithm in signal denoising.

Noise variance	0.5	0.9	1.3	1.5
RMSE	0.1522	0.1713	0.1854	0.2026

where s_3 is expressed a Gaussian white noise with a variance of 1.5, the composite signal y is a typical noisy multicomponent signal and the feature frequency is, respectively, set as $f_1 = 30$ Hz, $f_2 = 200$ Hz, and $f_3 = 100$ Hz.

Figure 3 shows the time and frequency responses of the synthetic signal y with strong noisy signal components. Figure 3(b) suggests that only the feature frequency f_2 and f_3 can be inspected, and the baseline of frequency spectrum has generated a drift. Unfortunately, the phenomenon of frequency modulation, namely, $f_2 \pm f_1$, is hardly identified since the interference of noise components. Nevertheless, frequency modulation is an important tool in fault state identification. Thus, we can make a conclusion that the traditional frequency spectrum analysis method should be improved. The immediate idea is that an effective denoising algorithm is performed before frequency spectrum analysis.

Subsequently, the commonly used method wavelet denoising and total variation (TV) denoising algorithm have been employed to analyze the complex simulation signal. The wavelet denoising is performed by wavelet packet scheme. Moreover, the wavelet base function is determined as “db5,” and the decomposition level is selected as 5. Then, the signal reconstruction is achieved using the

TABLE 2: Computational costs of four signal processing methods for simulated signal.

Method	Wavelet denoising	TV denoising	CEEMDAN	Proposed method
Computational costs	0.52	0.46	52	0.49

coefficient of the node (3, 0). Figure 4 has plotted the result generated by wavelet analysis in time domain and frequency domain. Judging from Figure 4, we can make a conclusion that the interesting components have been removed, and WT denoising method has failed to feature extraction. Traditional TV denoising algorithm has also been regarded as comparative analysis method in this paper, and the result is shown in Figure 5. According to Figure 5, we observe that most concerned signal components have been deleted, and its performance is still unsatisfied.

Subsequently, CEEMDAN is employed to simulation signal analysis, and the result is drawn in Figure 6. Original signal is decomposed into nine intrinsic mode functions (IMF). For the parameter setting for CEEMDAN algorithm, the noise standard deviation is 0.2, and the number of realizations is 200. It can be seen from Figure 6 that only the feature frequency $f_1 = 30$ Hz can be found in IMF2 and IMF3. Motivated by the classical first-order total variation (TV) denoising scheme, a sparse regularization method based on convex optimization is presented in this paper. In order to verify the effectiveness of this method, the researched method based on optimization method has also been applied to it. The parameters of the proposed method are chosen as $\gamma = 0.8$ and $\lambda = 2$. Figure 7 shows the result obtained by the proposed method. It is obvious that the feature frequency f_2 and f_3 can be clearly inspected. Most importantly, the frequency modulation $f_2 \pm f_1$ can also be determined. Comparing Figure 3(b) with Figure 7, it is obvious that the baseline in frequency domain has been corrected. The experimental results completely demonstrate that this proposed algorithm outperforms the classical wavelet denoising and TV method.

The actual noise reduction effect is directly related to the intensity of noise components. Theoretically, the smaller noise intensity will lead to the better actual denoising performance. When the intensity of Gaussian noise varies from 0.5 to 1.5, the root mean square error (RMSE) between the original noisy signal and the denoised signal is listed in Table 1. The result illustrated that the proposed method has advantage against the strong noise, and its robustness has been proved. The algorithms are run on a computer with an Intel Core i3-4160 CPU and 8.0 GB RAM. The computational costs of four signal processing methods for the simulated signal are listed in Table 2. Since the iterated operation with 200 times, the computational efficiency of CEEMDAN is low. It demonstrates that the computational complexity of the proposed method is acceptable.

4. Experimental Analyses

Rolling bearing is an important component of rotating machinery. Its main function is to support the mechanical rotary body, reduce the friction coefficient of its motion,



FIGURE 8: A photo of outer race fault.

TABLE 3: Rolling element bearing parameters.

Roller diameter/mm	Pitch diameter/mm	Number of elements	Contact angle
0.235	1.245	8	0

and guarantee its accuracy. Practically, the measured bearing failure signal is more complicated than the numerical simulation signal. Hence, the proposed method is also used for experimental signal analysis. The test rig was equipped with a NICE bearing with the outer race fault shown in Figure 8. The red circle indicates the location of the outer ring fault. The following parameters about the rolling bearing are listed in Table 3. The outer race fault conditions is expressed as follows: 300 lbs of load, input shaft rate $f_r = 25$ Hz, and sample rate of $f_s = 48828$ Hz. According to the theoretical calculation, the characteristic frequency of the outer ring fault is determined as $f_o = 80$ Hz.

The time domain waveform and frequency spectrum analysis results of the collected vibration signals are plotted in Figure 9. From the Figure 9(a) of the time response about measured bearing fault signal, the impact characteristics are obvious. However, we cannot detect the characteristic frequency of the outer ring in Figure 9(b). The frequency spectrogram analysis result indicates that the components of measured faulty signal are complicated. As the interference of the strong noisy components, the fault feature frequency can hardly been identified. The ideal result is to remove unnecessary signal components by effective denoising algorithm, and to retain or amplify the related signal components.

Then, the common signal processing methods, such as wavelet analysis and TV denoising, have been applied to the real inspected signal. For the method of wavelet denoising, we perform the wavelet packet decomposition denoising scheme. It should be noted that the wavelet base function is determined as “db5,” and the decomposition level is selected as 3. The coefficient of the node (2, 0) is used to signal reconstruction. Figures 10(a) and 10(b) show the results obtained by wavelet denoising and TV denoising, respectively. There is no obvious feature frequency corresponding to outer ring

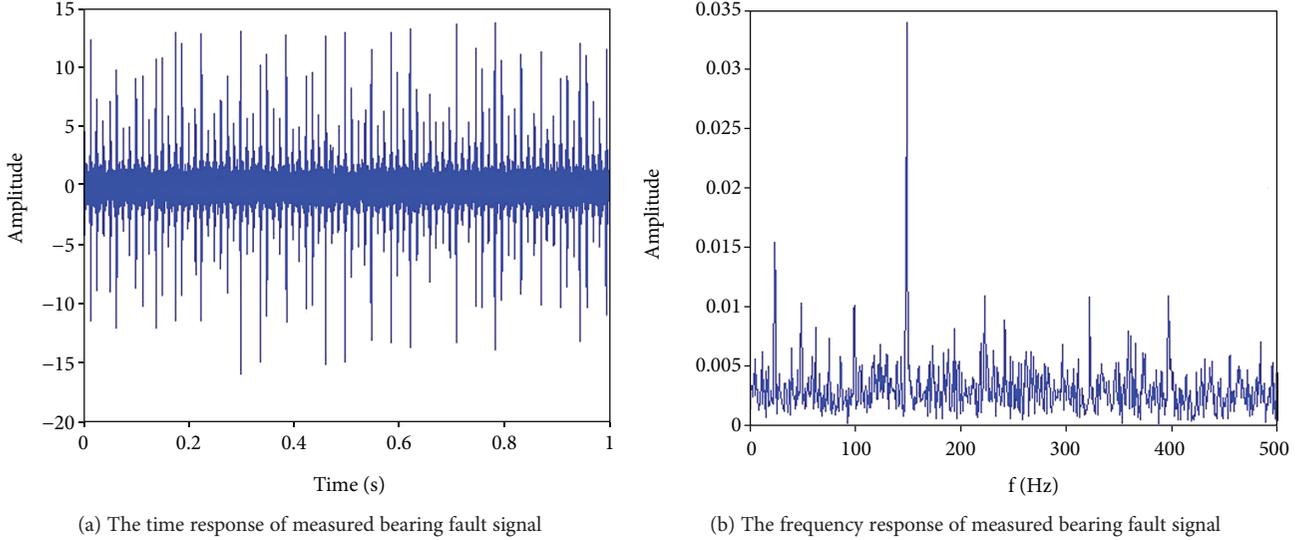


FIGURE 9: The time-frequency analysis of measured fault signal.

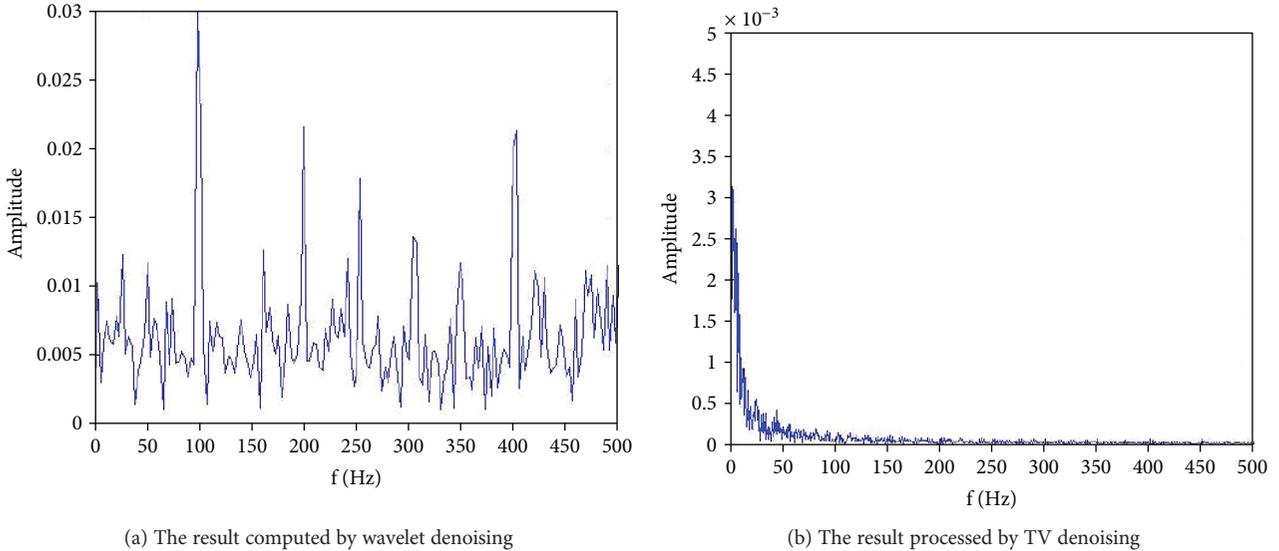


FIGURE 10: The result provided by wavelet denoising and TV denoising.

fault in Figure 10(a). Figure 10(b) suggested that the method of TV denoising has removed most of the useful signals, which is not suitable for vibration signal processing.

Similarly, CEEMDAN is also used to analyze the experimental data, and the corresponding result is plotted in Figure 11. Unfortunately, only the characteristic frequency of the outer ring fault f_o and its double frequency $2f_o$ can be inspected in IMF2, while the unwanted signal components have not been removed and it interferes with the identification of fault features. Since CEEMDAN still lacks theoretical support and exits the problem of mode aliasing, it fails to analyze the complex nonstationary vibration signal.

Finally, the proposed method based on GMC penalty function is performed to validate the method and illustrate its superiority. The concerned parameter of the proposed algorithm is chosen as $\gamma = 0.5$ and $\lambda = 2$, respectively. The computed result is drawn in Figure 12. Obviously, feature

frequency of outer ring fault f_o and its multiple frequencies ($2f_o$, $3f_o$, $4f_o$, $5f_o$, and $6f_o$) had been both identified in Figure 12. Meanwhile, the rotational frequency f_r can also be determined. The abovementioned characteristics fully indicate that the fault occurs in the outer ring, which is in accordance with the actual situation. Through the comparative analysis between Figures 10 and 12, the results show that the proposed method has obvious advantages in the fault state recognition of rolling bearings under the environmental of strong noise.

Theoretically, traditional signal processing methods such as fast Fourier transform and wavelet transform are based on the idea of matching the analyzed signal to the base function. Its performance largely depends on the signal structure and the base function selection. CEEMDAN is a typical signal adaptive decomposition algorithm, while it is restricted by endpoint effect and mode mixing. Distinguished it, convex

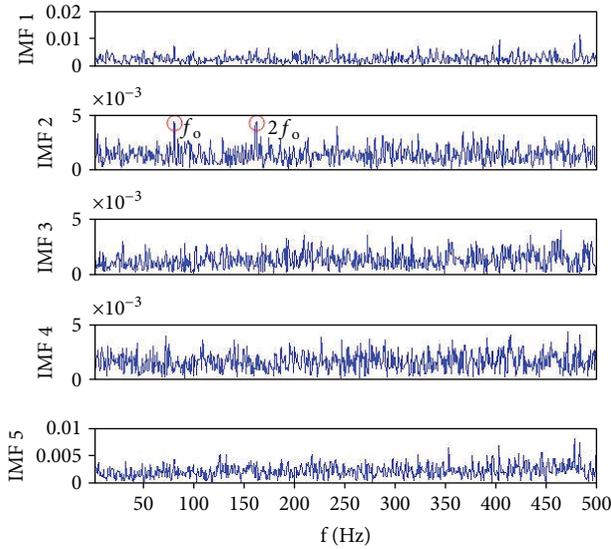


FIGURE 11: The result provided by CEEMDAN for experimental signal analysis.

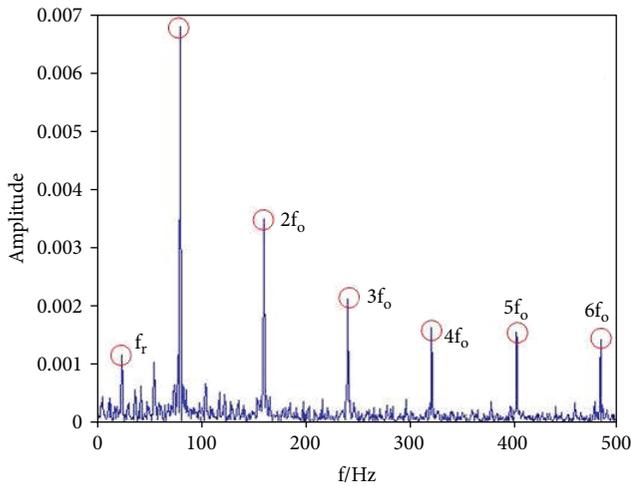


FIGURE 12: The result provided by the proposed method.

optimization denoising algorithm such as the proposed method is a typical iterative optimization process for objective functions. Using the nonconvex penalty function or cost function to realize the high efficiency, it is more suitable for analyzing the complex vibration signals.

5. Conclusions

For rolling bearing fault diagnosis, a novel method based on the improved sparse regularization via convex optimization is proposed in this paper. The main findings of this paper include (1) based on L_1 -norm and the Huber function, a novel nonconvex penalty function has been researched, namely, GMC penalty. It is employed to achieve the sparse representation of the signal and estimate the sparse solutions more accurately. (2) The proposed method is firstly put forward by combining the GMC penalty and TV denoising scheme. It is designed to enhance the performance of noise

reduction for the complex vibration signals. (3) Compared with the traditional methods such as WT and TV denoising, the proposed method has better performance, as demonstrated by both numerical and experimental studies.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this article.

Authors' Contributions

Dongjie Zhong and Cancan Yi conceived and designed the experiments. Dongjie Zhong and Han Xiao performed the experiments. Dongjie Zhong and Cancan Yi analyzed the data. Houzhuang Zhang and Anding Wu contributed to materials/analysis tools development. Dongjie Zhong, Cancan Yi, and Han Xiao wrote the paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant nos. 51475339, 51605344, and 51375354, the Natural Science Foundation of Hubei Province under Grant no. 2016CFA042, the Applied Basic Research Programs of Wuhan Science and Technology Bureau under Grant no. 2017010201010115, the Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education (Wuhan University of Science and Technology) under Grant no. MARC201701, Guangxi Key Laboratory of Optoelectronic Information Processing Open Foundation of China under Grant no. KFJJ2017-01, and Wenzhou Municipal Administration of Quality and Technology Supervision Research Project (201709).

References

- [1] M. Ragulskis, L. Chen, G. Song, and A. el Sinawi, "Intelligent fault diagnosis based on vibration signal analysis," *Shock & Vibration*, vol. 2017, Article ID 9186989, 1 pages, 2017.
- [2] J. Ma, J. Wu, and X. Wang, "Fault diagnosis method of check valve based on multikernel cost-sensitive extreme learning machine," *Complexity*, vol. 2017, Article ID 8395252, 19 pages, 2017.
- [3] Q. Feng, Q. Kong, and G. Song, "Damage detection of concrete piles subject to typical damage types based on stress wave measurement using embedded smart aggregates transducers," *Measurement*, vol. 88, pp. 345–352, 2016.
- [4] Q. Feng, H. Xiao, Q. Kong, Y. Liang, and G. Song, "Damage detection of concrete piles subject to typical damages using piezoceramic based passive sensing approach," *Journal of Vibroengineering*, vol. 18, no. 2, pp. 801–812, 2016.

- [5] I. Attoui, N. Fergani, N. Boutasseta, B. Oudjani, and A. Deliou, "A new time–frequency method for identification and classification of ball bearing faults," *Journal of Sound and Vibration*, vol. 397, pp. 241–265, 2017.
- [6] C. Yi, Y. Lv, Z. Dang, H. Xiao, and X. Yu, "Quaternion singular spectrum analysis using convex optimization and its application to fault diagnosis of rolling bearing," *Measurement*, vol. 103, pp. 321–332, 2017.
- [7] P. Yao, Q. Kong, K. Xu, T. Jiang, L.-s. Huo, and G. Song, "Structural health monitoring of multi-spot welded joints using a lead zirconate titanate based active sensing approach," *Smart Materials and Structures*, vol. 25, no. 1, article 015031, 2015.
- [8] H. Xiao, J. Zheng, and G. Song, "Severity evaluation of the transverse crack in a cylindrical part using a PZT wafer based on an interval energy approach," *Smart Materials and Structures*, vol. 25, no. 3, article 035021, 2016.
- [9] C. Peeters, P. Guillaume, and J. Helsen, "A comparison of cepstral editing methods as signal pre-processing techniques for vibration-based bearing fault detection," *Mechanical Systems and Signal Processing*, vol. 91, pp. 354–381, 2017.
- [10] X. Hong, J. Ruan, G. Liu, T. Wang, Y. Li, and G. Song, "Synergetics based damage detection of frame structures using piezoceramic patches," *Smart Structures and Systems*, vol. 17, no. 2, pp. 167–194, 2016.
- [11] G. Zhang, "Time-phase amplitude spectra based on a modified short-time Fourier transform," *Geophysical Prospecting*, vol. 66, no. 1, pp. 34–46, 2018.
- [12] E. Alickovic, J. Kevric, and A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction," *Biomedical Signal Processing and Control*, vol. 39, pp. 94–102, 2018.
- [13] B. Xu, T. Zhang, G. Song, and H. Gu, "Active interface debonding detection of a concrete-filled steel tube with piezoelectric technologies using wavelet packet analysis," *Mechanical Systems and Signal Processing*, vol. 36, no. 1, pp. 7–17, 2013.
- [14] L. Wang, Z. Liu, Q. Miao, and X. Zhang, "Time–frequency analysis based on ensemble local mean decomposition and fast kurtogram for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 103, pp. 60–75, 2018.
- [15] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [16] C. Yi, Y. Lv, and Z. Dang, "A fault diagnosis scheme for rolling bearing based on particle swarm optimization in variational mode decomposition," *Shock and Vibration*, vol. 2016, no. 2, Article ID 9372691, 10 pages, 2016.
- [17] I. Daubechies, J. Lu, and H. T. Wu, "Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243–261, 2011.
- [18] C. Yi, Y. Lv, H. Xiao, T. Huang, and G. You, "Multisensor signal denoising based on matching synchrosqueezing wavelet transform for mechanical fault condition assessment," *Measurement Science and Technology*, vol. 29, no. 4, article 045104, 2018.
- [19] H. H. Bauschke and L. Combettes, "Convex analysis and monotone operator theory in Hilbert space," in *CMS Books in Mathematics*, Springer, New York, NY, USA, 2011.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [21] H. Zhang, X. Chen, Z. Du, and R. Yan, "Kurtosis based weighted sparse model with convex optimization technique for bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 80, pp. 349–376, 2016.
- [22] C. Yi, Y. Lv, H. Xiao, K. Ke, and X. Yu, "A novel baseline correction method using convex optimization framework in laser-induced breakdown spectroscopy quantitative analysis," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 138, pp. 72–80, 2017.
- [23] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [24] I. W. Selesnick, A. Parekh, and I. Bayram, "Convex 1-D total variation denoising with non-convex regularization," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 141–144, 2014.
- [25] Y. Hu, G. Ongie, S. Ramani, and M. Jacob, "Generalized higher degree total variation (HDTV) regularization," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2423–2435, 2014.
- [26] Y. Ding and I. W. Selesnick, "Artifact-free wavelet denoising: non-convex sparse regularization, convex optimization," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1364–1368, 2015.
- [27] I. Selesnick and M. Farshchian, "Sparse signal approximation via nonseparable regularization," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2561–2575, 2017.
- [28] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4481–4494, 2017.
- [29] I. Selesnick, "Total variation denoising via the Moreau envelope," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 216–220, 2017.
- [30] J. Q. Wang, H. Xiao, Y. Lv, T. Wang, and Z. Xu, "Detrended fluctuation analysis and Hough transform based self-adaptation double-scale feature extraction of gear vibration signals," *Shock and Vibration*, vol. 2016, Article ID 3409897, 9 pages, 2016.
- [31] J. Li, X. Chen, Z. Du, Z. Fang, and Z. He, "A new noise-controlled second-order enhanced stochastic resonance method with its application in wind turbine drivetrain fault diagnosis," *Renewable Energy*, vol. 60, pp. 7–19, 2013.
- [32] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4144–4147, Prague, Czech Republic, May 2011.
- [33] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2006.

Research Article

An Investigation of Stretched Exponential Function in Quantifying Long-Term Memory of Extreme Events Based on Artificial Data following Lévy Stable Distribution

HongGuang Sun ¹, Lin Yuan,¹ Yong Zhang ^{1,2} and Nicholas Privitera²

¹State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, College of Mechanics and Materials, Hohai University, Nanjing, China

²Department of Geological Sciences, University of Alabama, Tuscaloosa, AL, USA

Correspondence should be addressed to Yong Zhang; y Zhang264@ua.edu

Received 29 January 2018; Revised 17 April 2018; Accepted 13 May 2018; Published 3 July 2018

Academic Editor: Gangbing Song

Copyright © 2018 HongGuang Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Extreme events, which are usually characterized by generalized extreme value (GEV) models, can exhibit long-term memory, whose impact needs to be quantified. It was known that extreme recurrence intervals can better characterize the significant influence of long-term memory than using the GEV model. Our statistical analyses based on time series datasets following the Lévy stable distribution confirm that the stretched exponential distribution can describe a wide spectrum of memory behavior transition from exponentially distributed intervals (without memory) to power-law distributed ones (with strong memory or fractal scaling property), extending the previous evaluation of the stretched exponential function using Gaussian/exponentially distributed random data. Further deviation and discussion of a historical paradox (i.e., the residual waiting time tends to increase with an increasing elapsed time under long-term memory) are also provided, based on the theoretical analysis of the Bayesian law and the stretched exponential distribution.

1. Introduction

Extreme events in complex systems have been widely explored for decades, such as natural hazards including extreme climate events [1], megalandslides [2], and earthquakes [3, 4] that cause severe challenges in economy, society, and environment. The clustering phenomena of extreme events imply the existence of long-term memory [5, 6]. Those phenomena were widely observed in river water levels [7], ocean temperature fluctuations [8], large-scale climate temperature [9], and so on. The generalized extreme value (GEV) distribution model (or the interval model) is designed to analyze the maximum within the interval R (see Figure 1(a), where the artificial random data is generated using the Lévy stable distribution, which will be further discussed in Section 2). According to the traditional extreme value theorem, these extremes will converge to the three

generalized extreme value distributions: Fréchet, Gumbel, and Weibull [10]. Though the GEV model has achieved many successful and empirical results [10, 11, 12], it is a statistical model based on independently and identically distributed (i.i.d.) data to investigate the probability density distribution characteristics without the impact of the temporal memory [13]. Prediction of the tail of the distribution, which has low probability but high impact, cannot be obtained accurately using the traditional extreme statistics, since it is impossible to obtain the effective description from the spatial probability density distribution or the i.i.d.

Previous studies have confirmed that the recurrence time analysis (see Figure 1(b)) is a powerful tool to characterize the temporal scaling properties and derive quantitative risk estimation of hazardous events [14]. This method can more efficiently use experimental data and characterize the physical correlations of time scales. Meanwhile, previous studies

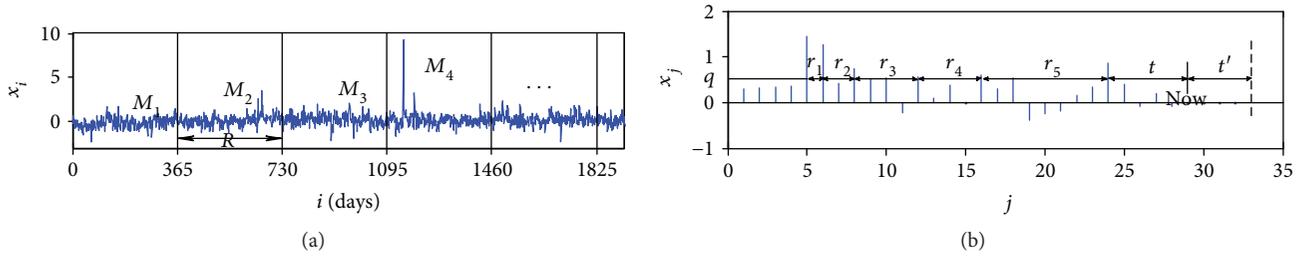


FIGURE 1: (a) Illustration of maxima series M_k in the GEV model, where R represents the interval size (here $R = 365$ days) and M_i represents the maximum value of the i th interval. (b) Definition of the return interval r_k in time series x_j for the threshold value q . The return interval r_k is the distance between two adjacent events that exceed the threshold q . The black dotted line in (b) represents the next unknown events that exceed the threshold q at the moment of “NOW”; t is the elapsed time since the last event; and t' is the residual waiting time to the next extreme event.

show that the recurrence of extreme events is not necessary to follow a pure memoryless Poisson distribution [15, 16].

This study aims at investigating extreme events with memory using two major steps. First, we will identify the memory effect embedded in extreme events. The memory effect can be characterized by the autocorrelation function [17, 18], given, for example, a normalized time series $\{x_i\}$ (where $i = 1, \dots, N$):

$$C_x(s) = \langle x_i x_{i+s} \rangle \equiv \frac{1}{N-s} \sum_{i=1}^{N-s} x_i x_{i+s}, \quad (1)$$

with a power law decay $C_x(s) \sim s^{-\gamma}$ and the correlation exponent $\gamma \in (0, 1)$. In this study, we use the detrended fluctuation analysis (DFA) [21] to detect this long-correlated behavior with Hurst exponent H . For long-term correlated data, the Hurst exponent equals to [22–25]

$$H = 1 - \frac{\gamma}{2}, \quad (2)$$

where $0.5 < H < 1$. The stretched exponential distribution with the correlation exponent γ proposed in [17–20] is then adopted to characterize the recurrence time.

Second, we will explore the influence of temporal memory on the forecast of extreme events based on the artificial data following the Lévy stable distribution. Under the memory behavior of previous events, there may be an improved estimate of the probability of a future event occurrence. To directly quantify this influence, we refer to Davis et al. [26]: “The longer it has been since the last earthquake, the longer the expected time till the next?” Therefore, we will apply the stretched exponential distribution, which is a widely used statistical model describing temporal memory features, to explore the possible “paradox” between the residual waiting and elapsed times. By extending the numerical analysis in literature [26–28], Sornette and Knopoff proposed a rigorous statistical framework for a quantitative conditional probability response and found that this framework is very sensitive to the assumed distribution [29]. Hereby, we will make an attempt to offer a derivation to this paradox (residual waiting time increases with the elapsed time under long-term memory), based on the theoretical analysis of the Bayesian law and the stretched exponential distribution.

It is also noteworthy that we select the Lévy stable distribution to quantify the heavy-tailed distribution of time series when analyzing extreme events. Cautions are needed when generating the time series data with long-term memory, because the non-Gaussian distribution feature of power-law processes cannot be well analyzed using traditional statistical models, such as the Gaussian distribution and lognormal distribution [17, 18, 30]. Based on extensive successful investigations of the Lévy stable distribution in real-world applications [31, 32], here we characterize heavy-tail behavior of time series using the Lévy stable distribution with a stability index α ($0 < \alpha < 2$). We apply the Lévy stable distribution and Hurst exponent in linear fractional stable noise (LFSN) to simulate heavy-tail and long-term memory processes and then investigate the property of extremal behavior using the methods proposed above.

The rest of this work is organized as follows. In Section 2, we introduce the LFSN model and explain the simulation parameters used to test the extreme value statistical behavior. In Section 3, we show the defects of traditional extreme value statistical models in describing the temporal behavior. The influence of temporal memory of the recurrence interval is then described using the stretched exponential distribution. In addition, based on the Bayesian theory and the stretched exponential statistics model, the “paradox” mentioned above is deduced in principle. Conclusions are drawn in Section 4.

2. Methods

2.1. Random Number Generation. The Lévy stable distribution includes four parameters: stability index α ($0 < \alpha \leq 2$), skewness parameter β ($-1 < \beta < 1$), scale parameter γ ($\gamma > 0$), and location parameter δ ($\delta \in \mathfrak{R}$). We employ the Lévy stable distribution to provide insight on the heavy-tail probability distribution, and this heavy-tail simulated fluctuation process is controlled by stability index $\alpha < 2$ in this study [33]. In the following, we use the random number generation method of the Lévy distribution proposed by Chambers et al. for analysis [34]. More details of the algorithm can be found in [35].

Linear fractional stable motion (LFSM) is a generalization of fractional Brownian motion (fBm) [36]. LFSN, which is an increment process of LFSM, displays both abnormal

fluctuations and long-term memory through Hurst exponent H and stability index α . The LFSM stochastic process is given as follows [37]:

$$M_{H,\alpha}(t) = C_{H,\alpha}^{-1} \int_{\mathfrak{R}} (t-s)^{H-1/\alpha} - (-s)^{H-1/\alpha} dL_\alpha(s), \quad (3)$$

where

$$C_{H,\alpha} = \left(\int_{\mathfrak{R}} |((1-s)^{H-1/\alpha} - (-s)^{H-1/\alpha})|^\alpha ds \right)^{1/\alpha}, \quad (4)$$

in which $0 < H < 1$, $0 < \alpha < 2$, and $L_\alpha\{L_\alpha, s \in \mathfrak{R}\}$ are a standard symmetric α -stable Lévy random measure on \mathfrak{R} . Linear fractional stable noise as the LFSM increment process is stated as

$$\begin{aligned} X_{H,\alpha}(t) &= M_{H,\alpha}(k) - M_{H,\alpha}(k-1) \\ &= -C_{H,\alpha}^{-1} \int_{\mathfrak{R}} (s)^{H-1/\alpha}_+ - (s-1)^{H-1/\alpha}_+ dL_\alpha(k-s), \end{aligned} \quad (5)$$

where $k \in \mathbb{Z}$ and $X_{H,\alpha}(t)$ presents the long-term memory when $H > 1/\alpha$, and it reduces to the fractional Gaussian noise when $\alpha = 2$. In our artificial data generation, the stability index is $\alpha = 1.8$, the Hurst parameter is $H = 0.8$ which corresponds to correlation exponent $\gamma = 0.4$ (see 2), and the number of generated data is $N = 2^{18}$.

2.2. Influence of Long-Term Memory and Non-Gaussian Processes on GEV Statistics. In the classical GEV model, one assumes that X_1, X_2, \dots, X_n are independent and identically distributed data described by the cumulative distribution function $F(x)$. The maximum value $M_n = \max\{X_1, X_2, \dots, X_n\}$ is also an element of the original data. Hence, the distribution $P_R(M_n)$ of the maxima M_n satisfies

$$P_R(M_n \leq x) = P_R(X_1 \leq x, \dots, X_n \leq x) = F^n(x). \quad (6)$$

According to the Fisher-Tippett extreme value theorem, if there are constant columns $\{a_n > 0\}$ and $\{b_n\}$, $P_R((M_n - b_n)/a_n \leq x) = F^n(a_n x + b_n) = H(x)$ is a nondegenerate distribution function. Hence, $H(x)$ must converge to one of the three types of extreme value distributions according to the distribution of the original data, when the number of data $n \rightarrow \infty$ [9]. For the original data following a power-law distribution, $H(x)$ converges to Fréchet distribution, or type II distribution, which is defined as

$$H(x) = \exp \left\{ - \left[1 + \xi \frac{(x - \mu)}{\sigma} \right]^{-1/\xi} \right\}, \quad \xi > 0, \quad (7)$$

where μ is the location parameter, σ represents the scale parameter, and ξ is the shape parameter.

Figure 2 shows a comparison of the probability density distribution $p_R(M_n)$ for both the correlated ($\gamma = 0.4$) and uncorrelated shuffled data for $R = 365$ (corresponding to the annual maxima). The distribution of long-term memory data (blue color) shifts to the left compared with the uncorrelated sequence, and the left tail exhibits an obvious broadening trend, which is consistent with the one offered in [18]. It

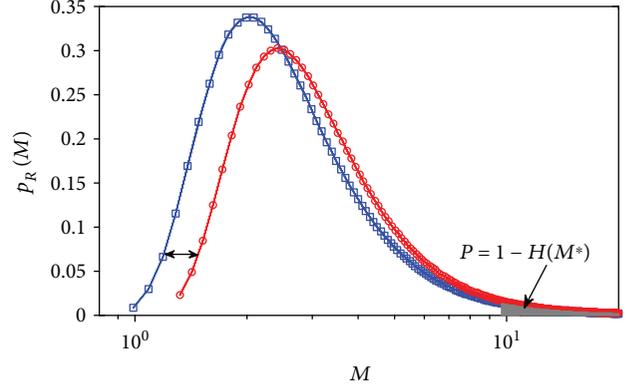


FIGURE 2: Comparison of the probability density distribution $p_R(M)$ of the maxima for two different data, where $R = 365$, representing the annual interval. The blue line represents the long-range correlation data ($\gamma = 0.4$), and the red line is shuffled uncorrelated data.

is noteworthy that the skewness of the probability density distribution of the correlated data following the Lévy stable distribution focused by this study is apparently larger than that following the Gaussian and exponential distributions considered in [18]. This discrepancy is mainly because the Lévy stable distribution is a middle part-dominated distribution with a heavy tail, compared with the Gaussian and exponential distributions. It is also clear that the distribution of long-term memory data (represented by the blue line in Figure 2) is more divergent compared with the uncorrelated sequence, especially on the left-hand side. This result indicates the influence of the memory or the role of correlation, which can make many large value events be clustered in certain time intervals while the maximum values in other periods are generally small. Because the large values will be still identified as annual maxima, the right tail of the extreme distribution is almost unaffected by the correlations. The GEV model cannot make a clear response to the temporal behavior.

When we investigate the extreme value problem, the calculation of the extreme value of the reoccurrence period is a very important part. For the T -year maximum value, the corresponding probability $P = 1/T$. Therefore, the maximum value of the return period T is estimated as the quantile M^* of the probability $P = 1 - \int_{-\infty}^{M^*} p_R(M) dM = 1 - H(M^*)$, and we can get

$$M^* = \mu - \frac{\sigma}{\xi} \left\{ 1 - [-\log(1-P)]^{-\xi} \right\}. \quad (8)$$

Here, we let $T = 100$ and then estimate a hundred-year maximum of two different memory behavior data (analyzed in Figure 2). Through (8), $M_{\gamma=0.4}^* = 28.0728$, $M_{\text{uncorr}}^* = 28.9301$, and we can find $M_{\gamma=0.4}^* \approx M_{\text{uncorr}}^*$. It is noteworthy that the obtained reoccurrence value has strong one-sidedness in practical sense, and the traditional extreme value model does not describe the factors of time-related behavior.

3. Results and Discussion

3.1. Statistics of Extreme Recurrence Times. We analyzed the return interval r_q over threshold values q for the return time statistics of long-term correlated time series. For uncorrelated data, such as “white noise,” the return intervals are also uncorrelated and follow the exponential distribution according to Poisson statistics [12]. When the return interval is affected by the long-term correlation, $P_q(r)$ exhibits a significant slower decay than the Poisson exponential distribution. This slower decay can be captured by the stretched exponential distribution [38, 39]:

$$P_q(r) \cong \frac{\alpha_\gamma}{R_q} \exp \left[-b_\gamma \left(\frac{r}{R_q} \right)^\gamma \right], \quad (9)$$

where the exponent γ is the correlation exponent to characterize the memory of the data, the parameters α_γ and b_γ are independent of q , and R_q is the average of the return interval at the given threshold q . In the study of the universality of (9), the return interval functions of four different original distribution data (Gaussian, exponential, power-law, and lognormal) were fitted in [14], where the results show that the stretched exponential agrees well with the Gaussian data and also good for the other distributions. It is also worth of note that the stretched exponential distribution of the recurrence time can be derived exactly from a deeper process, namely, the Hawkes process of interevent triggering [40].

Figure 3 shows the distribution of $P_q(r)$ of the return intervals r for both the original data (red symbols) and the shuffled data (rescaled by 10^{-1} , shown by the black symbols). In both cases, since $R_q P_q(r)$ reflects only the variation of the ratio r/R_q , the application of the recurrence time analysis is no longer limited to the actual threshold q . Figure 3 also confirms that the distribution function is exponential for the shuffled data.

Compared with the results of the uncorrelated data, the influence of the exponent γ in (9) makes the return intervals to exhibit an obvious two-stage differentiation. More intuitively speaking, the return intervals for both $r/R_q \ll 1$ and $r/R_q \gg 1$ are considerably more frequent for memory records than for the uncorrelated data. It means that the mean R_q is a poor description, because the analysis object has no typical scale or the “characteristic” scale is missing, or broadly referred to as “scale-free” phenomenon [41]. It also implies that the distribution changes from the exponential ($\gamma = 1$) to the power-law distribution ($\gamma \rightarrow 0$) when the index γ decreases (the degree of correlations increases) [42]. It means that the stretched exponential distribution is a subslow decay distribution between the exponential distribution and the power-law distribution with $0 < \gamma < 1$, in which the power-law relation is a statistical form of fractal which emphasizes the similarity of all scales [19, 43]. Therefore, the subslow decay of the stretched exponential distribution is in fact a scale-free statistical form or the result of transition from a nonsimilarity structure to a fully statistical fractal structure.

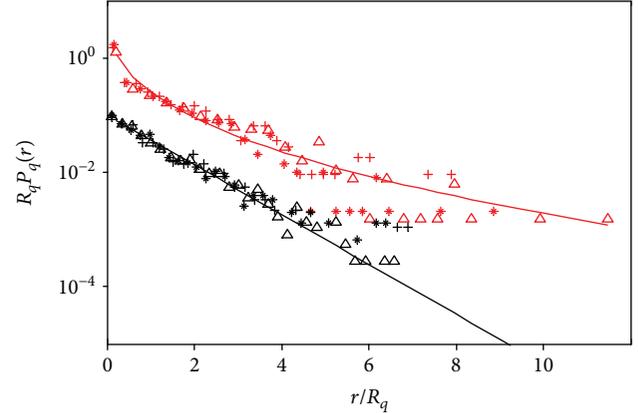


FIGURE 3: Normalized distribution density function $R_q P_q(r)$ as a function of r/R_q . The return interval r for the thresholds $q = 2.0$ ($\Delta R_q \approx 159$), 2.5 ($*R_q \approx 363$), 3.0 ($+R_q \approx 632$) for the simulated long-term correlated records with $\gamma = 0.4$ (red symbols). Black symbols represent the rescaled distribution ($\times 10^{-1}$) for the corresponding shuffled data with same thresholds.

3.2. The “Paradox” Phenomenon of the Residual Waiting Time under Long-Term Memory Effect. According to the discussion in the above section, the “cluster” phenomenon means that the data has a long memory effect, and the occurrence of the event is no longer a simple memoryless Poisson exponential distribution. The stretched exponential distribution of slow decay embodies the occurrence of events as a “scale-free” process. This temporal behavior can be quantified by the prior events, so there will be a corresponding predictable effect on the occurrence of the next event. In general, it reflects the dependence of the last return interval on the previous interval. The waiting time t' to the next event in the time interval also follows the stretched exponential distribution.

Bunde et al. [17] discussed the existence of memory effects according to the simulation data and obtained the “paradox” phenomenon of the residual waiting time that $t'_q(t | r_0)$ increases with increasing t and r_0 . Here, we try to derive a specific demonstration and quantification to the abnormal results, based on the fitting parameters for the stretched exponential distribution shown in Figure 3. In the following derivation, we adopt the numerical analysis method introduced by Sornette and Knopoff [29].

First, we assume $P(r)$ to be the return interval distribution, and the unknown residual waiting time t' satisfies the prior hypothesis $t' + t = r$. According to the Bayesian conditional probability theorem, the distribution function of t' satisfies

$$P(t' | t) = \frac{P(t' + t)}{\int_t^\infty P(r) dr}. \quad (10)$$

Here, we let the expected waiting time $\langle t' \rangle$ to be a function of time t and analyze the variation rule of $d\langle t' \rangle/dt$. From (10), the expected waiting time is calculated as

$$t' = \frac{\int_0^\infty t' P(t' + t) dt'}{\int_t^\infty P(r) dr}. \quad (11)$$

To develop some intuition, we first analyze the Poisson exponential distribution. From (10), we can get

$$P(t') = \frac{\exp(-t'/R_q)}{R_q}. \quad (12)$$

Corresponding to the exponential distribution without memory, the estimation of the time t' does not depend on the elapsed time t , with $d\langle t' \rangle / dt \equiv 0$. As for the stretched exponential, from (9) and (10), we get

$$P(t'|t) = \frac{a_\gamma/R_q \exp[-b_\gamma(t' + t/R_q)^\gamma]}{\int_t^\infty a_\gamma/R_q \exp[-b_\gamma(t/R_q)^\gamma] dr}. \quad (13)$$

We then calculate (13) using the Gauss-Kronrod integration method. Figure 4 shows the distribution of $P(t')$ at time $t = R_q$, $t = 2R_q$, and $t = 5R_q$ based on the fitting results of Figure 3. It is obvious that $P(t')$ has a progressively broadening tail to the origin as t increases and lies above $P(r)$ at $t' > R_q$; that is, the probability of large numbers for the residual waiting time t' increases as elapsed time t increases. The answer to the anomalous result is positive, and this property is evidently connected with the slow decay in $P(t')$ compared with the Poisson exponential distribution.

When we further investigate the relationship between the expected time $\langle t' \rangle$ ((1)) and the elapsed time t based on the fitting results in Figure 3, one can find that the expected waiting time $\langle t' \rangle$ depends on t , compared with the exponential distribution of the memoryless data. It clearly displays the effect of different long-term correlations, where the expected residual time to the next event increases with an increasing t . At the same time, the degree of the anomalous behavior increases, and this change is enhanced by decreasing γ . This result implies that the dependence of the memory effect exists not only between adjacent return intervals but also in the unknown interval between the waiting time t' and the elapsed time t . This result confirms the finding in the previous statistical analysis [17].

4. Conclusion

This study investigates the influence of memory effect on extreme value models, based on the random data generated by the Lévy stable distribution, which is different from the previous evaluation using the Gaussian or exponential distribution in [17, 18]. Combining with non-Gaussian and memory effect, the LFSN is used to simulate the experimental data. The simulation result shows that the stretched exponential distribution provides a reliable way to estimate the scaling behavior of extreme event intervals, generalizing the previous evaluation of the stretched exponential function to analyze random data following the Gaussian and exponential distributions. Using the Bayesian conditional statistical principle

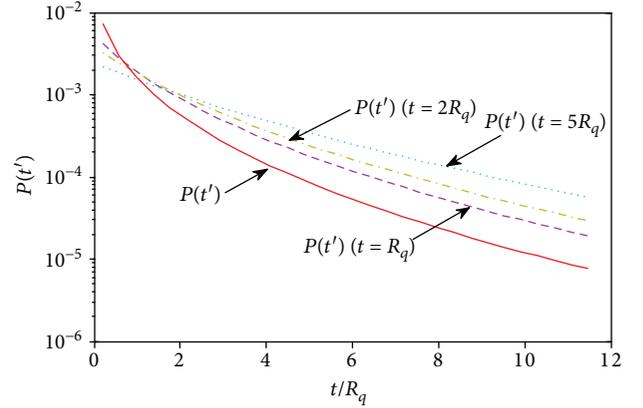


FIGURE 4: $P(t')$ is shown for $t = R_q$, $t = 2R_q$, $t = 5R_q$ together with $t = 0$ which returns to $P(r)$ distribution. At different elapsed time t , $P(t')$ exhibits varying degrees of slow decay.

in conjunction with the stretched exponential distribution, we also theoretically validate the “anomalous” behavior identified by various studies [17, 18] (where the residual waiting time can increase with an increasing elapsed time under long-term memory or the so called “anomalous residence time”), which may shed light on the real-world extreme event prediction.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant nos. 11572112, 41628202, and 41330632 and the Fundamental Research Funds for the Central Universities under Grant no. 2017B21614.

References

- [1] P. Yiou and M. Nogaj, “Extreme climatic events and weather regimes over the North Atlantic: when and where?,” *Geophysical Research Letters*, vol. 31, no. 7, article L07202, 2004.
- [2] O. Korup and J. J. Clague, “Natural hazards, extreme events, and mountain topography,” *Quaternary Science Reviews*, vol. 28, no. 11-12, pp. 977–990, 2009.
- [3] J. J. Love, “Credible occurrence probabilities for extreme geophysical events: earthquakes, volcanic eruptions, magnetic storms,” *Geophysical Research Letters*, vol. 39, no. 10, article L10301, 2012.
- [4] A. Saichev and D. Sornette, ““Universal” distribution of inter-earthquake times explained,” *Physical Review Letters*, vol. 97, no. 7, article 078501, 2006.

- [5] R. Glaser, *Klimageschichte Mitteleuropas: 1000 Jahre Wetter, Klima, Katastrophen*, Primus, 2001.
- [6] W. Soon and S. Baliunas, "Proxy climatic and environmental changes of the past 1000 years," *Climate Research*, vol. 23, no. 2, pp. 89–110, 2003.
- [7] I. Rodriguez-Iturbe and A. Rinaldo, *Fractal River Basins: Chance and Selforganization*, Cambridge University Press, Cambridge, NY, USA, 1997.
- [8] R. A. Monetti, S. Havlin, and A. Bunde, "Long-term persistence in the sea surface temperature fluctuations," *Physica A*, vol. 320, pp. 581–589, 2003.
- [9] P. Huybers and W. Curry, "Links between annual, Milankovitch and continuum temperature variability," *Nature*, vol. 441, no. 7091, pp. 329–332, 2006.
- [10] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer Science & Business Media, 2012.
- [11] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events: for Insurance and Finance*, Springer Science & Business Media, 2013.
- [12] H. V. Storch and F. W. Zwiers, *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, NY, USA, 1999.
- [13] M. R. Leadbetter and H. Rootzén, "Extremal theory for stochastic processes," *The Annals of Probability*, vol. 16, no. 2, pp. 431–478, 1988.
- [14] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, "Statistics of return intervals in long-term correlated records," *Physical Review E*, vol. 75, no. 1, article 011128, 2007.
- [15] E. G. Altmann and H. Kantz, "Recurrence time analysis, long-term correlations, and extreme events," *Physical Review E*, vol. 71, no. 5, article 056106, 2005.
- [16] J. Lavergnat and P. Gole, "A stochastic model of raindrop release: application to the simulation of point rain observations," *Journal of Hydrology*, vol. 328, no. 1-2, pp. 8–19, 2006.
- [17] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, "Long-term memory: a natural mechanism for the clustering of extreme events and anomalous residual times in climate records," *Physical Review Letters*, vol. 94, no. 4, article 048701, 2005.
- [18] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, "Extreme value statistics in records with long-term persistence," *Physical Review E*, vol. 73, no. 1, article 016130, 2006.
- [19] J. Laherrère and D. Sornette, "Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 2, no. 4, pp. 525–539, 1998.
- [20] D. Sornette, *Springer Series in Synergetics, Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-Organization and Disorder: Concepts and Tools*, Springer Series in Synergetics, Springer, Berlin, 2004.
- [21] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Physical Review E*, vol. 49, no. 2, pp. 1685–1689, 1994.
- [22] E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H. J. Schellnhuber, "Indication of a universal persistence law governing atmospheric variability," *Physical Review Letters*, vol. 81, no. 3, pp. 729–732, 1998.
- [23] J. D. Pelletier and D. L. Turcotte, "Self-affine time series: II. Applications and models," *Advances in Geophysics*, vol. 40, pp. 91–166, 1999.
- [24] R. O. Weber and P. Talkner, "Spectra and correlations of climate data from days to decades," *Journal of Geophysical Research: Atmospheres*, vol. 106, no. D17, pp. 20131–20144, 2001.
- [25] J. F. Eichner, E. Koscielny-Bunde, A. Bunde, S. Havlin, and H.-J. Schellnhuber, "Power-law persistence and trends in the atmosphere: a detailed study of long temperature records," *Physical Review E*, vol. 68, no. 4, article 046133, 2003.
- [26] P. M. Davis, D. D. Jackson, and Y. Y. Kagan, "The longer it has been since the last earthquake, the longer the expected time till the next?," *Bulletin of the Seismological Society of America*, vol. 79, no. 5, pp. 1439–1456, 1989.
- [27] S. N. Ward and S. D. B. Goes, "How regularly do earthquakes recur? A synthetic seismicity model for the San Andreas fault," *Geophysical Research Letters*, vol. 20, no. 19, pp. 2131–2134, 1993.
- [28] S. D. B. Goes and S. N. Ward, "Synthetic seismicity for the San Andreas fault," *Annals of Geophysics*, vol. 37, no. 6, pp. 1495–1513, 1994.
- [29] D. Sornette and L. Knopoff, "The paradox of the expected time until the next earthquake," *Bulletin of the Seismological Society of America*, vol. 87, no. 4, pp. 789–798, 1997.
- [30] A. Bunde, J. F. Eichner, S. Havlin, and J. W. Kantelhardt, "Return intervals of rare events in records with long-term persistence," *Physica A: Statistical Mechanics and its Applications*, vol. 342, no. 1-2, pp. 308–314, 2004.
- [31] V. V. Uchaikin and V. M. Zolotarev, *Chance and Stability: Stable Distributions and Their Applications*, Walter de Gruyter, 1999.
- [32] R. Weron, "Computationally intensive value at risk calculations," in *Papers/Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE)*, Humboldt University Berlin, 2004.
- [33] J. P. Nolan, "Parameterizations and modes of stable distributions," *Statistics & Probability Letters*, vol. 38, no. 2, pp. 187–195, 1998.
- [34] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A method for simulating stable random variables," *Publications of the American Statistical Association*, vol. 71, no. 354, pp. 340–344, 1976.
- [35] Y. Liang and W. Chen, "A survey on computing Lévy stable distributions and a new MATLAB toolbox," *Signal Processing*, vol. 93, no. 1, pp. 242–251, 2013.
- [36] I. I. Eliazar and M. F. Shlesinger, "Fractional motions," *Physics Reports*, vol. 527, no. 2, pp. 101–129, 2013.
- [37] S. A. Stoev, *Stable Self-Similar and Locally Self-Similar Random Processes: Stochastic Properties, Parameter Estimation, and Simulation*, Boston University, 2005.
- [38] G. F. Newell and M. Rosenblatt, "Zero crossing probabilities for Gaussian stationary processes," *The Annals of Mathematical Statistics*, vol. 33, no. 4, pp. 1306–1313, 1962.
- [39] I. Blake and W. Lindsey, "Level-crossing problems for random processes," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 295–315, 1973.
- [40] A. Saichev and D. Sornette, "Theory of earthquake recurrence times," *Journal of Geophysical Research: Solid Earth*, vol. 112, no. B4, article B04313, 2007.
- [41] J. B. Bassingthwaite, L. S. Liebovitch, and B. J. West, "Properties of fractal phenomena in space and time," in *Fractal Physiology, Methods in Physiology Series*, pp. 11–44, Springer, New York, NY, USA, 1994.

- [42] R. Hardstone, S. S. Poil, G. Schiavone et al., “Detrended fluctuation analysis: a scale-free view on neuronal oscillations,” *Frontiers in Physiology*, vol. 3, p. 450, 2012.
- [43] A. Eke, P. Herman, L. Kocsis, and L. R. Kozak, “Fractal characterization of complexity in temporal physiological signals,” *Physiological Measurement*, vol. 23, no. 1, pp. R1–R38, 2002.

Research Article

A Modified Time Reversal Method for Guided Wave Detection of Bolt Loosening in Simulated Thermal Protection System Panels

Guan-nan Wu,^{1,2} Chao Xu ,¹ Fei Du,¹ and Wei-dong Zhu²

¹School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China

²Mechanical Engineering Department, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

Correspondence should be addressed to Chao Xu; chao_xu@nwpu.edu.cn

Received 12 February 2018; Revised 7 May 2018; Accepted 13 May 2018; Published 2 July 2018

Academic Editor: Chen Lu

Copyright © 2018 Guan-nan Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this work, a modified time reversal method is proposed for guided wave detection and localizing loosened bolt in a complicated multibolt-jointed structure. Different from the traditional time reversal guided wave method, the response signal due to a tone burst input received at the healthy state is time reversed and recorded as a standard reemitting signal. In the detection process, this recorded standard signal is used for all damage cases to yield time reversal-focalized reconstruction signals. This largely improves the sensitivity of the focalized signal to damage state. In this paper, the peak amplitude of the focalized wave packet in the reconstructed signal is calculated and utilized as tightness index. By bonding PZT transducers at different joint locations inside the structure, multiple tightness indices, where each tightness index presents the correlation between the current joint condition to its healthy condition at the joint, can be obtained. To analyze a large number of tightness indices, a principle component analysis method is introduced, and a neural network-based loosening detection method is proposed. The proposed method is experimentally validated in a simulated double-layer bolt-jointed thermal protection system panel. Experimental results illustrate that the proposed method is effective to identify and localized the bolt loosening in complicated multibolt-jointed structure. The detection and identification of the location of multibolt loosening is realized.

1. Introduction

In aerospace engineering, a thermal protection system (TPS) is necessary for protecting an aerospace vehicle from harsh heating due to high reentry temperatures. It plays a critical role in keeping structural integrity and safety. A typical TPS design is a multilayer configuration, where a cover panel made of high-temperature material, such as carbon-carbon composite, is attached to the main airframe panel of the aerospace vehicle through mechanically bolted joints. The extreme thermal, vibration, and shock excitations during the flight may cause the pretightened bolts to loosen, which further cause detachment gap in a TPS. Then, hot air can penetrate through the gap and cause catastrophic results. Therefore, the detection of loosening in bolted joints plays an underlying role to ensure a TPS functionality, integrity, and reusability.

Structural health monitoring (SHM) is referred as the process of implementing a damage detection and

characterization strategy for structures [1]. In the past decade, a number of SHM approaches have been developed to detect bolt loosening in TPS structures. At the early stage, the SHM methods based on experimental modal analysis were proposed to realize the objective. For example, Vandawaker et al. [2] proposed to detect damage in a TPS tile by comparing mode shapes and frequencies from healthy and damaged structures. However, since high-frequency vibration modes are difficult to obtain in practice, the modal analysis-based methods mainly rely on the low-frequency vibration modes, which only represent the properties of the whole structure. Like Todd et al. [3] pointed out an assembled structure which consists of many bolts and one or less number of loosened bolts cannot cause the whole structural properties to change significantly. Consequently, the modal analysis-based methods are relatively insensitive to detect local bolt loosening for a complicated jointed structure.

To detect local and small structural damage, a number of local SHM methods have been developed. Among them,

electromechanical impedance-based and guided wave-based methods are very popular [4, 5]. Basically, the electromechanical impedance-based method monitors variations in mechanical impedance due to damage, which is coupled with electrical impedance of a piezoelectric transducer (PZT) [6, 7]. A low-cost miniaturized impedance measurement device was developed by Peairs and Inman [8] for detection of bolt loosening. A system that includes PZTs and a wireless impedance device for data acquisition and communication was built by Mascarenas et al. [9] to detect bolt preload loss. Recently, an electromechanical impedance method for health monitoring of aircraft-bolted joints was presented by Kuznetsov et al. [10]. While the previous studies have shown the feasibility of using impedance-based approaches for detection of bolt loosening, the effective detection area of a transducer is very small and a large number of transducers are required for detecting a structure with thousands of bolts.

Guided wave-based SHM techniques have been intensively developed over the last two decades. They are very attractive and commonly utilized due to their ability to inspect a large structure over long distance with a small number of transducers. In recent years, several guided wave-based SHM methods have been proposed for bolt loosening detection. Yang and Chang [11] developed a guided wave-based method to identify bolt loosening in a TPS panel. In their study, a piezoelectric- (PZT-) embedded smart washer was developed to generate and acquire ultrasonic wave signals. The features of wave energy dissipation (WED) and specific damping capacity were extracted. However, a large number of smart washers are required in applying the detection method to a TPS with hundreds of bolts, because each bolt should be installed with a smart washer. Similar wave energy dissipation-based methods were studied by Montoya and Maji [12] and Wang et al. [13]. It is known that the nonlinear contact mechanics on the joint interface can cause nonlinear transmission of guided waves. The SHM method based on nonlinear ultrasonic wave phenomenon was proposed by Doyle et al. [14] to localize loosened bolts in a satellite panel. In their work, the acoustoelastic behavior due to the joint was used as an indicator of structural integrity. Furthermore, the feature of contact acoustic nonlinearity in bolted joints was utilized for preload monitoring [15–20]. Amerini and Meo [15] studied both linear and nonlinear ultrasonic methods and proposed several tightening state indices, including the first-order acoustic moment, high-harmonic generation, and sideband modulation indices. Bao et al. [16] established a nonlinear contact element model to simulate the strip lap joint specimen. On this basis, the high-harmonic generation-based index was studied by Shen et al. [17]. After that, Zhang et al. [18] proposed to use subharmonic resonance to detect bolt looseness, and Zhou et al. [19] utilized nonlinear-modulation method to detect bolt looseness in frame structure. Zhang et al. [20] carried out a comparative study of the WED-based method and the vibroacoustic modulation-based method and found that the sensitivity of WED-based methods is closely related to the wave energy transmitted across joints.

In recent years, new guided wave detection methods based on the time reversal principle in modern acoustics have been developed and the effectiveness was demonstrated. Fundamentally, the time reversal concept lies in reconstruction of an input signal at an excitation point, while a response signal measured at another point is reemitted to the original excitation point after being reversed in time domain, taking into account that there is no damage in the wave path. As a result, changes between the reconstructed signal and the input signal can be inferred as presence of damage in the wave path. Wang et al. [21] carried out an experimental work to study the applicability of the time reversal concept to guided waves in plates. Then, Park et al. [22] utilized the time reversal technique to improve the detect ability of local defects in composite plate. A detailed theoretical investigation on the effect of multimodal and reflections on time reversal progress was given in [23]. It was pointed out that narrowband input signal can enhance time reversal progress. In recent years, the technique of time reversal guided wave SHM has been applied to a variety of structures, such as metallic plates [24], composite plates [25], and rebar-reinforced concrete beams [26]. Mustapha and Ye proposed a very effective and innovative time reversal-based imaging algorithm to detect debonding in sandwich composite plates [27] and then extended it to detect multiple debonding in complex tapered composite sandwich panels [28]. However, the time reversal guided wave method has not been widely used for detection of bolt loosening, except the works done by Tao et al. [29] and Parvasi et al. [30], where only very simple single bolt lap-jointed structures were considered. In their work, the aim is to identify the preload loss in single bolt. The problem of detection of loosening in a multibolt-jointed structure was not considered.

In this paper, a modified time reversal (MTR) guided wave method is proposed to detect bolt loosening in TPS panels. The conventional time reversal process is modified, and a new implemented process is proposed to localize loosening bolts in a complicated multibolt-jointed structure. In the proposed method, a standard reemitted signal which recorded for a nominal healthy structure is used for all damage cases. The peak amplitude of the refocused wave packet obtained by the MTR method is normalized and utilized as the tightness index. By bonding PZT sensors at different joint locations inside the structure, multiple tightness indices can be obtained, where each tightness index presents the correlation between the current joint condition to its healthy condition at the joint. To analyze a large number of tightness indices, a principle component analysis method is introduced, and a neural network-based loosening detection method is proposed. The feasibility and effectiveness of the proposed method to identify and localize bolt loosening is experimentally investigated in a simulated TPS structure.

2. Theory Background

2.1. The Conventional Time Reversal Guided Wave Method. Figure 1 shows the work principle of the conventional time reversal guided wave method in a simple lap joint. First, a

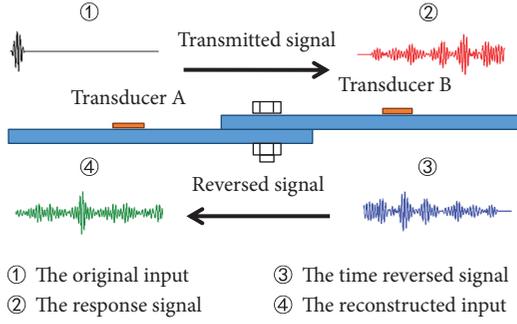


FIGURE 1: Illustration of the time reversal principle in a lap joint.

tone burst input $e(t)$ is applied to transducer A, which acts as an actuator and activates wave propagation in the structure. A wave response signal $u(t)$ is captured by the transducer B. The response signal at the transducer B can be defined as

$$u(t) = h(t) * e(t), \quad (1)$$

where $h(t)$ is the impulse response function (IRF) of the structure between points A and B; $*$ denotes the convolution. After that, the recorded response signal is reversed in time domain before reemitted at transducer B. Thus, the reemitted signal is

$$u(-t) = h(-t) * e(-t). \quad (2)$$

After the signal is reversed in time domain, the wave packets that arrive at transducer B early will be reemitted late, and the wave packets that arrive at transducer B late will be reemitted early. Therefore, wave packets in the time-reversed signal arrive at transducer A at the same time and overlap to a focalized wave packet. Note that, in practical applications, the reversed signal is amplified before feeding it into PZT B. The reconstructed signal received at the transducer A can be written as

$$\begin{aligned} s(t) &= h(t) * u(-t) = h(t) * h(-t) * e(-t) \\ &= e(-t) * \int_{-\infty}^{+\infty} h(\tau)h(\tau - t)d\tau = e(-t) * R_{11}(t), \end{aligned} \quad (3)$$

where $R_{11}(t)$ is a autocorrelation function and also called the time reversal operator. $R_{11}(t)$ gets its maximum value at time point $t = 0$ and is equal to the energy of IRF of the structure between points A and B. It is noted that the energy of IRF is decided by the wave energy transmits through the joint interface. Bolt loosening causes the decrease of contact area and consequently transmitting wave energy and the IRF energy decreases. Thus, the peak amplitude of reconstructed signal $s(t)$ is proportional to the IRF energy, and the peak amplitude of $s(t)$ can be extracted as a tightness index (TI) to detect bolt loosening.

This principle is used by Tao et al. [29] and Parvasi et al. [30] for detecting preload loss in a single bolted lap joint. However, the results show that a saturation phenomenon of the TI happens when the preload reaches to a critical value [29, 30]. The reason is that the feature of peak amplitude is

related to the IRF and the wave energy transmitted across joint, which are determined by the contact area. When the preload is high, the contact area changes very little due to the change of preload. Thus, in this situation, the IRF nearly holds constant. This causes the conventional time reversal guided method which has a low sensitivity at the early stage of loosening.

2.2. The Modified Time Reversal Guided Wave Method. The basic idea of the MTR method is to introduce a “standard” reemitted signal (SRS) for all damage cases, where damage is referred to bolt loosening in this work. As shown in Figure 2, the time-reversed reemitted signal (the symbol ③ signal in Figure 1) recorded at the nominal healthy structure is chosen as the “standard” reemitted signal for all damage cases. When the SRS is reemitted back to the structure, damage affects the reconstructed signal in extensive patterns comparing to the conventional time reversal method. Thus, additional damage information that can be extracted to reveal the extent of damage is contained in the reconstructed signal. The detailed explanation of the proposed modified method is presented below.

Fundamentally, when a SRS is reemitted back to the structure, if the bolt in the wave path is not loosened, which means the structure can be regarded as the same as the nominal healthy structure, the MTR is therefore equal to the TR method. The velocities for every wave packet in the SRS keep the same and a focalized wave packet can be obtained by the backward travelling of SRS, due to the in-phase overlapping of wave packets in the SRS. However, if the bolt in the wave path is loosened, the wave velocity for each wave packet may also be changed. According to Zagrai et al. [31], the main reason of the change of wave velocity is due to acoustoelastic effect. With the change of bolt preload, the static stress level changes and the wave speed of the transmitted guided wave signals changes. Therefore, the in-phase overlapping of wave packets in the SRS cannot be realized to a focalized wave packet. This indicates that in the MTR method, not only the transmitted wave energy is related to the peak amplitude of the reconstructed signal, but also the impaired focalizing ability of SRS can cause the decrease of the peak amplitude.

If the IRF of the healthy structure between a actuator and a sensor is $h_1(t)$, therefore, as the same as (1), the “standard” reemitted signal can be written as

$$u_{\text{stand}}(-t) = h_1(-t) * e(-t). \quad (4)$$

This signal is recorded and used for all damage cases. If there exists damage in the wave path, it is obvious the IRF changes. It is assumed that the IRF of the damaged structure is $h_2(t)$, and then the reconstructed signal $s_2(t)$ due to the SRS can be obtained, defined as

$$\begin{aligned} s_2(t) &= h_2(t) * u_{\text{stand}}(-t) = h_2(t) * h_1(-t) * e(-t) \\ &= e(-t) * \int_{-\infty}^{+\infty} h_1(\tau)h_2(\tau - t)d\tau = e(-t) * R_{21}(t), \end{aligned} \quad (5)$$

where $R_{21}(t)$ is a cross-correlation function of $h_1(t)$ and $h_2(t)$ and represents the degree of the correlation of $h_1(t)$ and

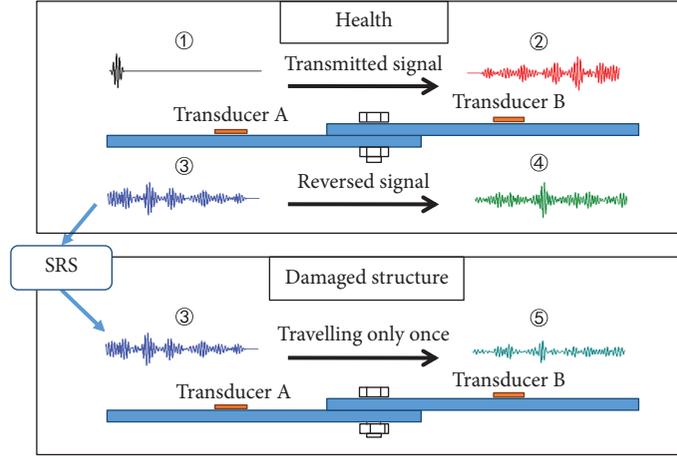


FIGURE 2: A schematic of the proposed method for bolt loosening detection.

$h_2(t)$. R_{21} gets its maximum value at the time instant $t = \Delta t$. Comparing (5) to (3), it can be seen that the MTR method actually modifies the time reversal operator. In the conventional time reversal method, the autocorrelation function of the IRF is used, while in the modified time reversal method, by introducing a reference IRF, the cross-correlation function is used as the time reversal operator. Meanwhile, it can be found that in (5), similar to (3), the peak amplitude of the focalized wave packet is proportional to the maximum value of R_{21} . Therefore, the principle of the proposed method is to change the time reversal operator from the autocorrelation function of structural states to the cross-correlation function of structural states.

It should be noted that in the conventional time reversal method, for any signal path, both transducers need to work as actuators and sensors. This often means a large number of hardware, such as PZT amplifier and wiring are required. In [25], a simplified time reversal method is proposed, which can largely reduce the hardware requirement by using only one transducer A to actuate signals and another transducer B to act as a sensor for a signal path. The underlying reason is that the impulse response function (IRF) of the structure from point A to B is assumed the same as that from point B to A. In this paper, the same simplified technique is used to reduce hardware requirement and experimental cost.

2.3. Loosening Localization Method for Multibolt Structures.

It is noted that the above-proposed principle uses a single bolt loosening detection problem as an example. However, the proposed method can be easily extended to a multibolt structure. In this paper, the aim is to detect the loosening location for a multibolt TPS structure. To this end, each joint with attached bolts is considered as a possible loosening location. At each possible loosening location, a PZT sensor is bonded to act as a wave signal receiver. In addition, a PZT transducer bonded at the central location of bottom panel is used to generate wave signal to sensors. The detailed sensor configuration is given in the next section.

In this paper, the peak amplitude of the reconstructed focalized wave packet is extracted as a TI for detection of bolt

loosening in a multibolt-jointed TPS structure. The TI can be written as

$$TI_i = \frac{A_{ij}}{A_{i0}}, \quad (6)$$

where A_{ij} is the peak amplitude of reconstructed signal received by sensor i at the current structure, and A_{i0} is the peak amplitude received by sensor i at the healthy structure. By installing PZT sensors around every possible loosening bolts, predictive variables TI_i s can be defined. Each TI_i represents the correlation coefficient of a structural state of a joint to its healthy state. For instance, bolt loosening happening at joint 1 changes the IRF of actuator to sensor 1 significantly; according to (5) and (6), the corresponding TI_1 is therefore significantly decreased. On the other hand, the loosening at joint 1 has less effect to the IRF of actuator to sensor 2 which is far from joint 1. Hence, the TI_2 corresponding to sensor 2 shows a smaller change than TI_1 when loosening occurs at joint 1. Depending on this feature, the bolt loosening in the structure can be detected by using predictive variables TI_i s.

To implement the proposed MTR method, each bolt joint location needs one sensor to form a signal path. Hence, in real application, a great number of transducers may be needed, and a lot of variables need to be processed, correspondingly. However, too many predictive variables cause the difficulty in the visual description of detection results. Therefore, it is important to keep the independence of each predictive variable for the proposed method. In this paper, dimension reduction of predictive variables TI_i s before the establishment of damage index database is performed. The principal component analysis (PCA) [32] is used for dimension reduction of TI_i s. First, TI_i s should be scandalized to mean equals to zero and the standard deviation equals to one. Assuming TI_i is a $n \times 1$ vector and n is the number of data. \mathbf{D} is the $n \times m$ matrix composed by TI_i s, where m is the number of the predictive variable. Then, the normalized matrix \mathbf{X} can be obtained by

$$\mathbf{X} = (\mathbf{V}^{1/2})^{-1}(\mathbf{D} - \mu), \quad (7)$$

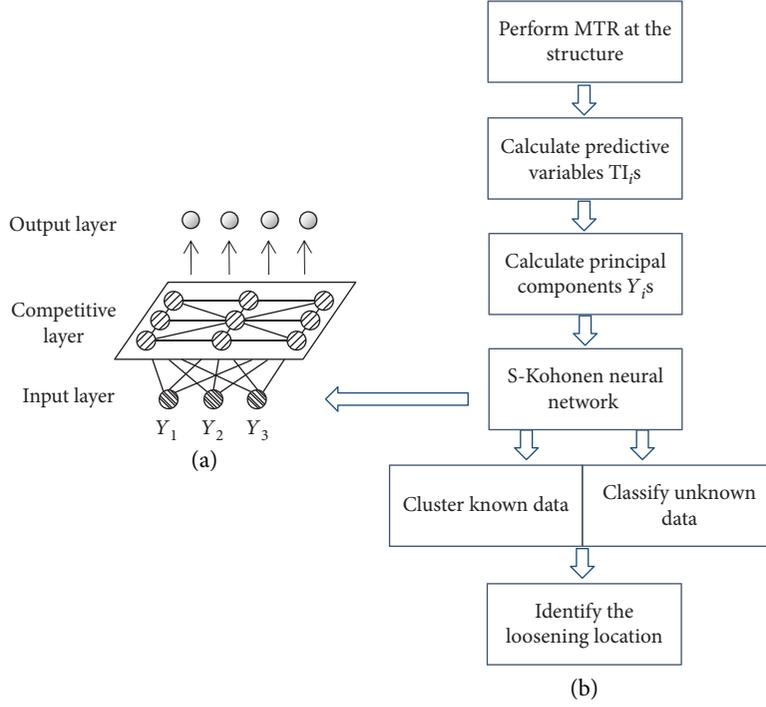


FIGURE 3: The schematic of the proposed method: (a) the topological structure of S-Kohonen network and (b) the flow chart of the loosening detection process.

where index “ -1 ” means calculating inverse, μ is the means of \mathbf{D} , and $\mathbf{V}^{1/2}$ is a $m \times m$ matrix of standard deviation, defined as

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{mm} \end{bmatrix}, \quad (8)$$

where σ_{ii} is the standard deviation of TI_i . Calculating the covariance matrix $\text{Cov}(\mathbf{X})$ of \mathbf{X} , and we can get the eigenvalues and eigenvectors $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_m, e_m)$ of $\text{Cov}(\mathbf{X})$. Then, the number i principal component Y_i can be obtained by

$$Y_i = (TI_1, TI_2, \dots, TI_m)e_i. \quad (9)$$

The number of principal component can be determined by the data variation ratio that the obtained Y_i can explain. Generally, 90% of data variation should be explained by the k principal components, that is,

$$\sum_{i=1}^k a_i \geq 90\%, \quad (10)$$

where a_i is the data variation explained by Y_i and $a_i = \lambda_i/m$.

Finally, the S-Kohonen neural network [33] is used in this work for clustering analysis of the known data and classifying of the unknown data. The S-Kohonen neural network is an extended Kohonen neural network (KNN) that contains an added output layer. As shown in Figure 3(a), a

typical KNN contains an input layer and a competitive layer, the node number of input layer equals to the predictive variables, and the competitive layer is a 2D array that usually contains much more nodes than the target variables. Furthermore, the S-Kohonen network has an added output layer after the competitive layer to connect target variables directly to the clustering result. By performing the above steps, a loosening detection flow can be built, as shown in Figure 3(b).

3. Experimental Study

3.1. The Simulated TPS Structure. As shown in Figure 4, a simulated multibolt-jointed TPS structure is considered for validating the proposed MTR method. A Q235 steel thin plate simulating the cover panel is attached to a Q235 steel thin base plate to simulate the airframe panel via four steel brackets. M6 bolts are used as fasteners. The dimension of the base plate is $450 \text{ mm} \times 385 \text{ mm} \times 3 \text{ mm}$ and the upper plane is $335 \text{ mm} \times 335 \text{ mm} \times 3 \text{ mm}$. PZT patches are used to actuate and receive wave signals. One PZT actuator is bonded on the center of the base plate, and each bracket is bonded with one sensor. The PZT sensor location and the bolt numbers are shown in Figure 4.

Considering the allowable tensile load of M6 bolt and the yielding strength of Q235 steel, $10 \text{ N}\cdot\text{m}$ is selected as the standard bolt torque. For each bracket, totally, 6 bolt loosening cases are taken into account as shown in Table 1. During the experiments, each bolt loosening case is repeated four times to get reliable results. In this work, the bolt loosening means a totally loose condition. In each bolt loosening detection experiment, the TPS panel specimen is assembled

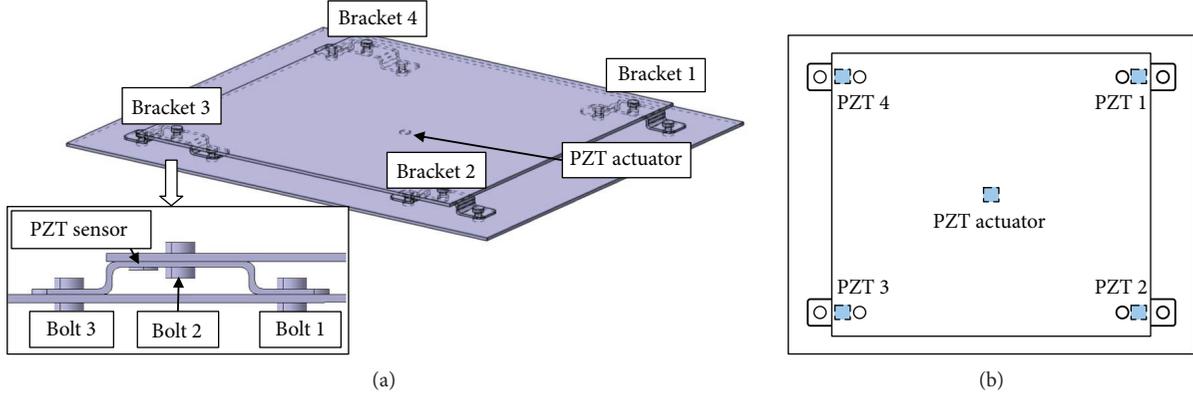


FIGURE 4: A schematic of the simulated PTS panel: (a) the simulated PTS panel and its brackets and (b) relative locations of the PZT transmitters.

TABLE 1: Bolt loosening conditions for each bracket.

Case number	1	2	3	4	5	6
Loosened bolt	Bolt 1	Bolt 2	Bolt 3	Bolts 1 and 2	Bolts 2 and 3	Bolts 1 and 3

to the normal tightening condition firstly, and then it is adjusted to a bolt loosening case by loosening corresponding bolts. A standard assembly sequence is regulated to guarantee experimentally repeatability. Each bolt is tightened to 70% standard torque firstly and then tightened to 100% in sequence to prevent unnecessary internal stress.

3.2. Experimental Setup. Figure 5 gives the experimental setup. The multifunction data acquisition (DAQ) system NI USB-6366 is used to generate and receive guided wave signals. Before the input signal is sent to PZT actuator, a signal amplifier PINTEK HA-400 is used to amplify the signal amplitude to 50 Vpp (peak to peak). A LabVIEW program is coded and runs on the computer to control the DAQ system. In addition, the structure is placed on a foam support.

A 5-cycle tone burst is used as the original input of MTR process. The central frequency of the input is chosen as 150 kHz based on the group velocity dispersion plot of 3 mm steel plate, as shown in Figure 6. It can be seen that at 150 kHz, only the first S_0 symmetrical mode and antisymmetrical A_0 mode are generated, and the velocity dispersion of these modes is minimized. Response signals of the tone burst input at each sensor in the healthy structure are recorded. After reversed in time domain, they are saved as SRSs. The generated and recorded signals in the MTR method for the healthy state are shown in Figure 7. To reduce the effect of noise, each recorded signal is the average of 32 measurements. Moreover, a high-pass filter is used to cut off the low-frequency noise; the low-pass cutoff is set to 10 kHz. It can be clearly found that a focalized wave packet occurs in the reconstructed signal.

4. Results and Discussion

4.1. Validating the Modified Time Reversal Process. By implementing the proposed time reversal method, the

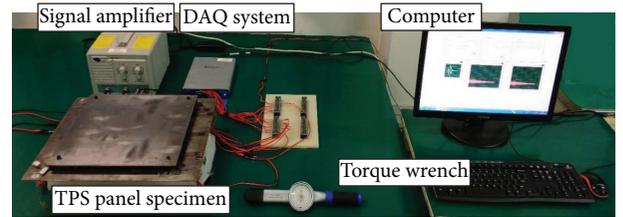


FIGURE 5: The experimental platform.

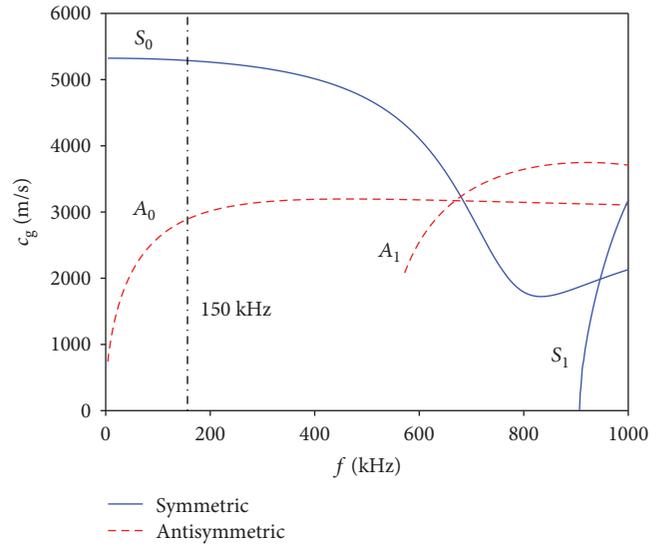


FIGURE 6: The group velocity dispersion plot of 3 mm Q235 steel plate.

reconstructed signals recorded at each bracket sensor for the healthy structure are given in Figure 8(a). It can be seen for the healthy state that each sensor received a clear

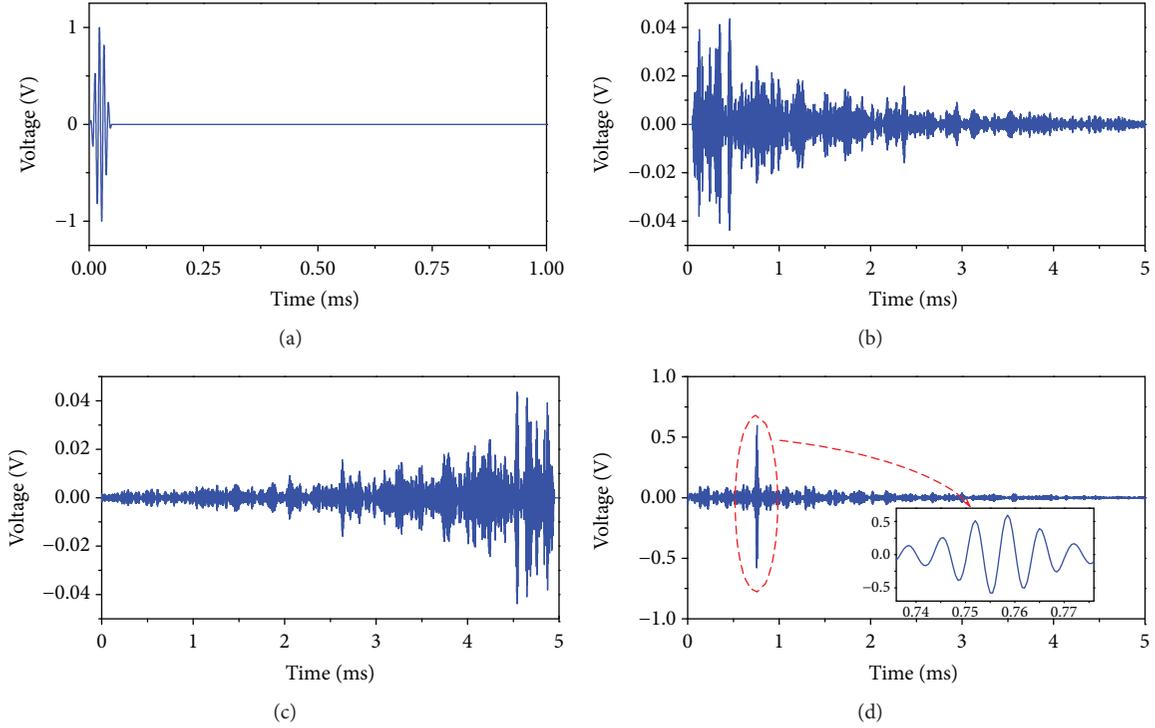


FIGURE 7: Input/output signals of the time reversal process: (a) the tone burst input, (b) response signal due to the tone burst input, (c) the time-reversed reemitting signal, and (d) the reconstructed signal received at healthy structure.

reconstructed signal with a focalized wave packet. This result corresponds with the conventional time reversal guided wave method. However, it can be found that the peak amplitudes of the focalized wave packet for each sensor are slightly different. This is due to the structural assembly error and the deviation of bonding condition of PZT patches.

The reconstructed signals in Figure 8(b) were normalized by the corresponding reconstructed signals for the healthy structure. Then, the peak amplitudes of the reconstructed signals for the healthy structure were normalized to 1. Figure 8(b) shows the comparisons of normalized peak amplitudes for damaged and healthy structures at different PZT locations. It is found that a focalized wave packet still exists in the reconstructed signal for each sensor, but the peak amplitude of reconstructed wave packet at each sensor changes obviously. Since the loosening of bolt 2 in bracket 2 has changed the wave propagation behavior of the whole structure significantly, the peak amplitude of each PZT patch is affected by different extents. It can be seen in Figure 8(b) that the peak amplitude decreases the largest at the PZT patch bonded on bracket 2 compared to the other three results.

4.2. Predictive Variables TI_i s. The predictive variables TI_i s for each bracket at different bolt loosening cases are calculated and shown in Figure 9. First, it can be seen that in each bolt loosening case, each TI_i is less than 1, and the loosening bracket corresponding TI_i is smaller than the other TI_i s. Therefore, based on TI_i s, the bolt loosening location can be identified by directly comparing the values of predictive variable. Second, the robustness of the MTR method is

examined by three repeating experiments at each bracket. It can be seen that at each bolt loosening case, the predictive variable TI_i at each bracket appears similar character, and from three repeating experiments, good repeatability of the measurement can be observed.

In addition, it can be seen that the mismatched TI_i s of the loosening bracket is nearly unchanged at each case, but the corresponding TI_i changes with the change of case number. On the one hand, comparing the results of case 1/2/3, it can be found that the loosening of bolt 3 has the most significant influence to the corresponding TI_i . On the other hand, the change of TI_i becomes more distinct with the increase of the number of loosening bolt. Checking on Figure 4, it can be found that the bolt 1 is the farthest bolt to the PZT sensor, and bolt 2 is the nearest one. Since the actuator PZT patch is bonded on the base plate, the loosening of bolt 1 and bolt 3 has direct influence to the wave propagation path, comparing to bolt 2 which connects the bracket with the cover plate. Overall, considering these factors, the reason of why bolt 3 has the most significant influence to TI_i can be explained: (1) the loosening of bolt 1 and bolt 3 can directly affect the wave propagation path, and (2) bolt 3 is more closer to the PZT sensor than bolt 1.

4.3. Clustering Results. The PCA of TI_i s is performed in this step. The correlation of TI_i s is shown in Figure 10, and it can be seen that predictive variables TI_1 , TI_2 , TI_3 , and TI_4 show the same correlation with each other. For instance, the values of TI_1 and TI_2 are high when the bolt loosening happens at bracket 3 and bracket 4, and when it happens at

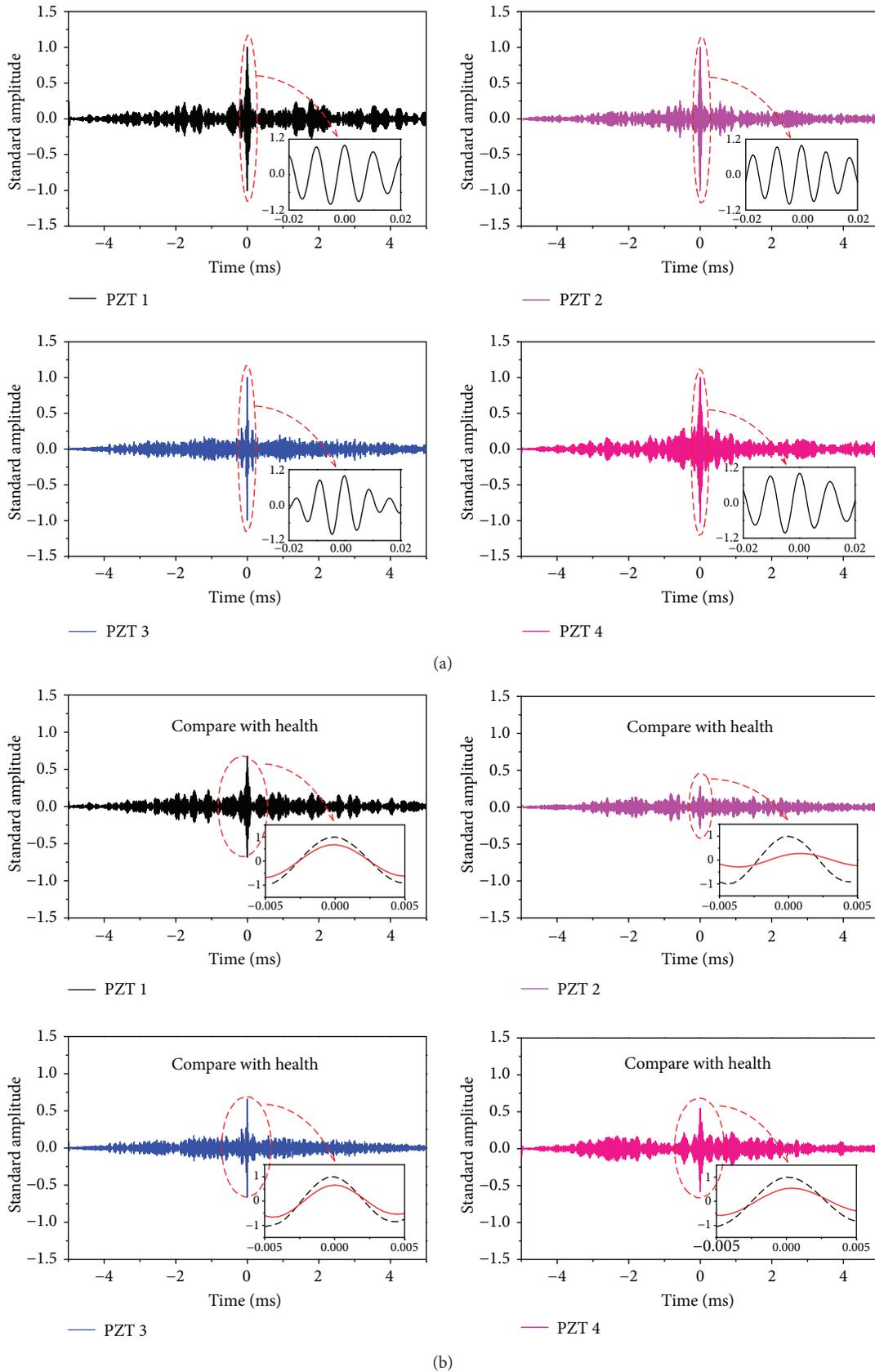
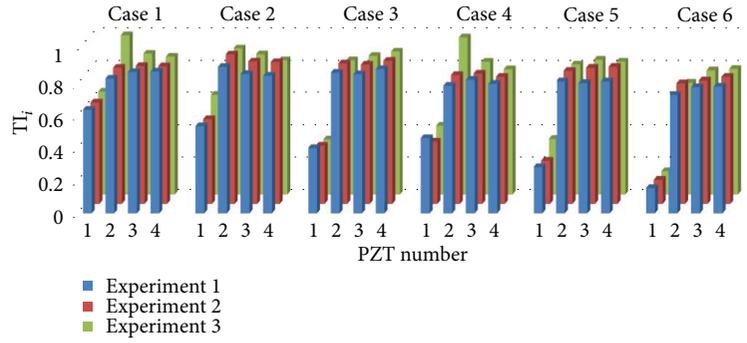
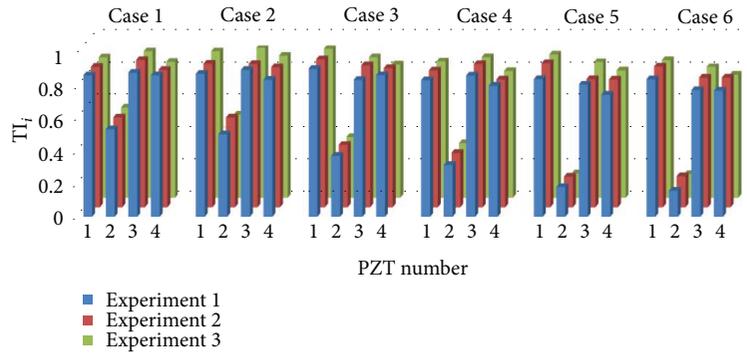


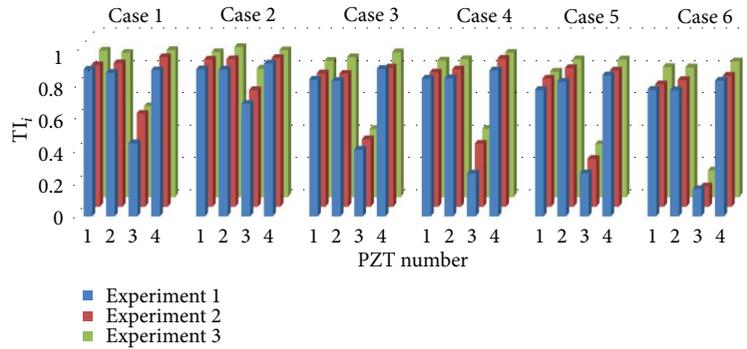
FIGURE 8: Experimental results using the proposed method: (a) normalized reconstructed signals for healthy structure and (b) reconstructed signals for damaged structure with joint loosening at the bracket 2.



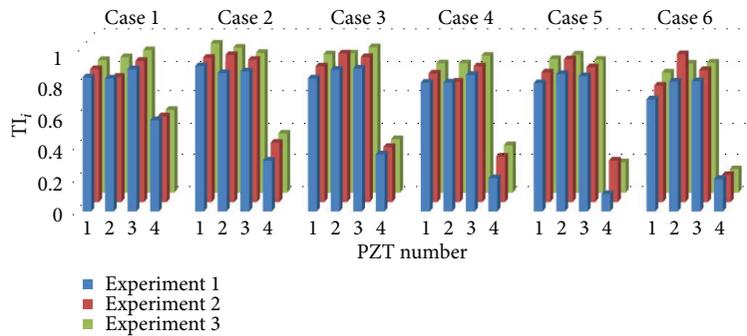
(a)



(b)



(c)



(d)

FIGURE 9: The predictive variables TI_i s at different bolt loosening cases: (a) results of bracket 1, (b) results of bracket 2, (c) results of bracket 3, and (d) results of bracket 4.

bracket 1 and bracket 2, TI_1 and TI_2 get a low value, respectively. The scatter diagram of TI_1 and TI_2 appears a curvilinear correlation with the loosening at different positions.

Furthermore, it can be found that this relationship is kept in the correlation of other predictive variables. Since a same curvilinear correlation is found between any two TI_i s, it

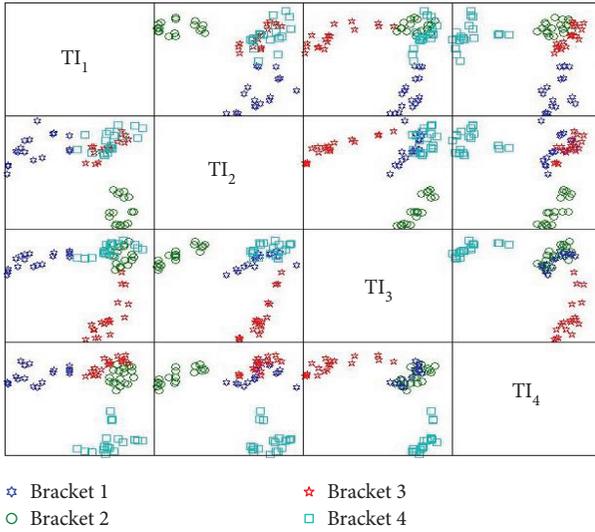


FIGURE 10: The correlations of predictive variables TI_i s; each symbol represents a bracket number where bolt loosening happens.

indicates that the PCA method can be performed properly to extract principal components.

This study got 96 TI_i s from 4 repeating experiments of 6 bolt loosening cases at 4 brackets, respectively. The 72 data from the first 3 times repeating experiments were used to establish a database, and the last 24 data were utilized as test. The PCA for the 72 data was performed, and the covariance matrix $Cov(\mathbf{X})$ was calculated. Then, the eigenvalues and eigenvectors of $Cov(\mathbf{X})$ were calculated as $\lambda_1 = 1.3938$, $\lambda_2 = 1.2234$, $\lambda_3 = 0.9615$, and $\lambda_4 = 0.4213$. Based on (10), it can be found that when $k = 3$, the principal components can explain 89.47% of the data variation. A 3D scatter plot of principal components Y_1 , Y_2 , and Y_3 is shown in Figure 11; it can be found that the data at the same loosened bracket is close to each other and the distance of data from different loosening brackets is large. This distribution of data guarantees the accuracy of the clustering analysis and the possibility for bolt loosening locating.

4.4. Tests of the Loosening Detection Database. Figure 12 shows the distribution of winning nodes at the competitive layer of the 72 Y_i s data in one running result of S-Kohonen neural network. Accordingly, the node number of input layer is 3, and the node number of competitive layer is $4 \times 4 = 16$. The learning rate is set to [0.01, 0.1], and the learning diameter is set to [0.4, 2]; number of cycles is 1000. In Figure 12, the symbol \circ represents bolt loosening cases at bracket 1, the symbol $\textcircled{2}$ represents bolt loosening cases at bracket 2, and the rest can be done in the same manner. The blank represents the node not belonging to any cluster. It can be seen that the winning nodes distribute as blocks; the effect of clustering analysis is good.

Based on (9), the last 24 TI_i s data can be transferred to Y_i s. The classifying results of S-Kohonen network is shown in Figure 13. The node number of the output layer is 4, and learning rate between the output layer and the competitive layer is [0.5, 1]. It can be seen that all of the

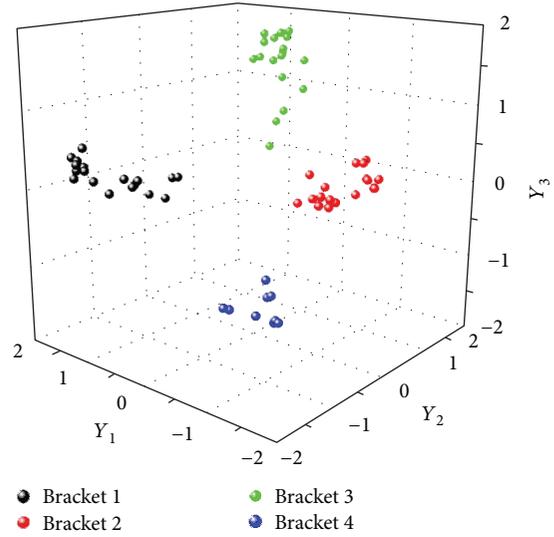


FIGURE 11: The 3D scatter plot of Y_1 , Y_2 , and Y_3 .

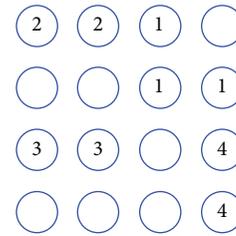


FIGURE 12: The distribution of winning nodes at the competitive layer.

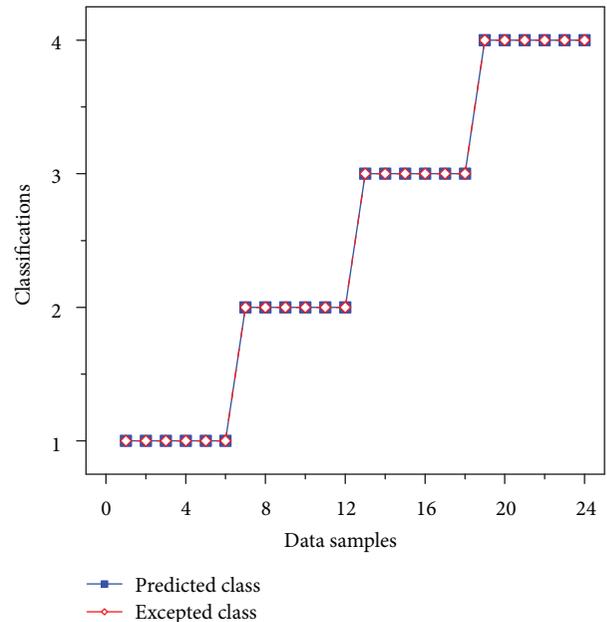


FIGURE 13: The classifying result of S-Kohonen network.

24 data have been classified correctly; the accuracy is 100%. Therefore, it can be concluded that the proposed MTR method can correctly recognize the bolt loosening location and the recognizing accuracy is extremely high.

5. Conclusion

The problem of detecting bolt loosening for a complicated multibolt-jointed structure is investigated. To this end, a modified time reversal guided wave method is proposed in this paper. Different with the conventional time reversal guided wave method, the response signal due to a tone burst input received at the healthy structure is time reversed and saves as standard reemitting signal. The peak amplitude of the focalized wave packet in the reconstructed signal due to the SRS is extracted as tightness index. By bonding PZT sensors at different joint locations inside the structure, multiple tightness indices can be obtained, where each tightness index presents the correlation between the current joint condition to its healthy condition at the joint. To analyze a large number of tightness indices, a principle component analysis method is introduced, and a neural network-based loosening detection method is proposed.

Experiments at a simulated TPS panel are conducted to verify the feasibility and effectiveness of the MTR method for bolt loosening identification. Different bolt loosening cases are considered, and repeating experiments have been done. Conclusions can be extracted from the analysis of experiment results. It is found that the proposed modified time reversal method can be used to identify and localize the bolt loosening in a simulated TPS panel. By analyzing principal components Y_i s, the loosened bracket can be found out. Moreover, the change of TI_i becomes more distinct with the increase of the number of loosening bolt. This work can be extended to similar bolt loosening detection problems in variety multibolt-jointed structures.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors wish to acknowledge the financial support from the China NSAF Project (Grant no. U1530139), China Science Challenge Project (Grant no. TZ2018007), and National Natural Science Foundation of China (Grant no. 51705422).

References

- [1] C. R. Farrar and K. Worden, "An introduction to structural health monitoring," *Philosophical Transactions of the Royal Society A Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 303–315, 2010.
- [2] R. M. Vandawaker, A. N. Palazotto, and R. G. Cobb, "Damage detection through analysis of modes in a partially constrained plate," *Journal of Aerospace Engineering*, vol. 20, no. 2, pp. 90–96, 2007.
- [3] M. D. Todd, J. M. Nichols, C. J. Nichols, and L. N. Virgin, "An assessment of modal property effectiveness in detecting bolted joint degradation: theory and experiment," *Journal of Sound and Vibration*, vol. 275, no. 3–5, pp. 1113–1126, 2004.
- [4] M. Mitra and S. Gopalakrishnan, "Guided wave based structural health monitoring: a review," *Smart Materials & Structures*, vol. 25, no. 5, article 053001, 2016.
- [5] H. Sohn, C. R. Farrar, and D. J. Inman, "Overview of piezoelectric impedance-based health monitoring and path forward," *The Shock and Vibration Digest*, vol. 35, no. 6, pp. 451–463, 2003.
- [6] V. Giurgiutiu, A. Zagrai, and J. B. Jing, "Piezoelectric wafer embedded active sensors for aging aircraft structural health monitoring," *Structural Health Monitoring*, vol. 1, no. 1, pp. 41–61, 2002.
- [7] A. V. G. Madhav and S. C. Kiong, "Application of electromechanical impedance technique for engineering structures: review and future issues," *Journal of Intelligent Material Systems and Structures*, vol. 21, no. 1, pp. 41–59, 2010.
- [8] D. M. Peairs and D. J. Inman, "Improving accessibility of the impedance-based structural health monitoring method," *Journal of Intelligent Material Systems and Structures*, vol. 15, no. 2, pp. 129–139, 2004.
- [9] D. L. Mascarenas, G. Park, K. M. Farinholt, M. D. Todd, and C. R. Farrar, "A low-power wireless sensing device for remote inspection of bolted joints," *Proceedings of the Institution of Mechanical Engineers Part G Journal of Aerospace Engineering*, vol. 223, no. 5, pp. 565–575, 2009.
- [10] S. Kuznetsov, I. Pavelko, T. Panidis, V. Pavelko, and I. Ozolinsh, "Bolt-joint structural health monitoring by the method of electromechanical impedance," *Aircraft Engineering and Aerospace Technology: An International Journal*, vol. 86, no. 3, pp. 207–214, 2014.
- [11] J. Yang and F. K. Chang, "Detection of bolt loosening in C–C composite thermal protection panels: I. Diagnostic principle," *Smart Materials & Structures*, vol. 15, no. 2, pp. 581–590, 2006.
- [12] A. C. Montoya and A. K. Maji, "An assessment of joint rigidity with ultrasonic wave energy," *Journal of Nondestructive Evaluation*, vol. 30, no. 3, pp. 122–129, 2011.
- [13] T. Wang, G. Song, Z. Wang, and Y. Li, "Proof-of-concept study of monitoring bolt connection status using a piezoelectric based active sensing method," *Smart Materials & Structures*, vol. 22, no. 8, article 087001, 2013.
- [14] D. Doyle, A. Zagrai, B. Arritt, and H. Cakan, "Damage detection in bolted space structures," *Journal of Intelligent Material Systems and Structures*, vol. 21, no. 3, pp. 251–264, 2010.
- [15] F. Amerini and M. Meo, "Structural health monitoring of bolted joints using linear and nonlinear acoustic/ultrasound methods," *Structural Health Monitoring*, vol. 10, no. 6, pp. 659–672, 2011.
- [16] J. Bao, Y. Shen, and V. Giurgiutiu, "Linear and nonlinear finite element simulation of wave propagation through bolted lap joint," in *54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, pp. 1–12, Boston, MA, USA, 2013.

- [17] Y. Shen, J. Bao, and V. Giurgiutiu, "Health monitoring of aerospace bolted lap joints using nonlinear ultrasonic spectroscopy: theory and experiments," *International Workshop on Structural Health Monitoring*, vol. 745, pp. 2333–2340, 2013.
- [18] M. Zhang, Y. Shen, L. Xiao, and W. Qu, "Application of subharmonic resonance for the detection of bolted joint looseness," *Nonlinear Dynamics*, vol. 88, no. 3, pp. 1643–1653, 2017.
- [19] W. Zhou, Y. Shen, L. Xiao, and W. Qu, "Application of nonlinear-modulation technique for the detection of bolt loosening in frame structure," *Journal of Testing and Evaluation*, vol. 44, no. 2, pp. 967–975, 2016.
- [20] Z. Zhang, M. Liu, Z. Su, and Y. Xiao, "Quantitative evaluation of residual torque of a loose bolt based on wave energy dissipation and vibro-acoustic modulation: a comparative study," *Journal of Sound and Vibration*, vol. 383, pp. 156–170, 2016.
- [21] C. H. Wang, J. T. Rose, and F. K. Chang, "Computerized time-reversal method for structural health monitoring," in *Proceedings SPIE, Nondestructive Evaluation and Health Monitoring of Aerospace Materials and Composites II*, vol. 5046, no. 2, pp. 48–58, San Diego, CA, USA, August 2003.
- [22] H. W. Park, H. Sohn, K. H. Law, and C. R. Farrar, "Time reversal active sensing for health monitoring of a composite plate," *Journal of Sound and Vibration*, vol. 302, no. 1-2, pp. 50–66, 2007.
- [23] H. W. Park, S. B. Kim, and H. Sohn, "Understanding a time reversal process in lamb wave propagation," *Wave Motion*, vol. 46, no. 7, pp. 451–467, 2009.
- [24] B. Poddar, A. Kumar, M. Mitra, and P. M. Mujumdar, "Time reversibility of a lamb wave for damage detection in a metallic plate," *Smart Materials & Structures*, vol. 20, no. 2, article 025001, 2011.
- [25] R. Watkins and R. Jha, "A modified time reversal method for lamb wave based diagnostics of composite structures," *Mechanical Systems & Signal Processing*, vol. 31, no. 8, pp. 345–354, 2012.
- [26] S. Mustapha, Y. Lu, J. Li, and L. Ye, "Damage detection in rebar-reinforced concrete beams based on time reversal of guided waves," *Structural Health Monitoring*, vol. 13, no. 4, pp. 347–358, 2014.
- [27] S. Mustapha and L. Ye, "Propagation behaviour of guided waves in tapered sandwich structures and debonding identification using time reversal," *Wave Motion*, vol. 57, pp. 154–170, 2015.
- [28] S. Mustapha, L. Ye, D. Wang, and Y. Lu, "Debonding detection in composite sandwich structures based on guided waves," *AIAA Journal*, vol. 50, no. 8, pp. 1697–1706, 2012.
- [29] W. Tao, L. Shaopeng, S. Junhua, and L. Yourong, "Health monitoring of bolted joints using the time reversal method and piezoelectric transducers," *Smart Materials & Structures*, vol. 25, no. 2, article 025010, 2016.
- [30] S. M. Parvasi, S. C. M. Ho, Q. Kong, R. Mousavi, and G. Song, "Real time bolt preload monitoring using piezoceramic transducers and time reversal technique—a numerical study with experimental verification," *Smart Materials & Structures*, vol. 25, no. 8, article 085015, 2016.
- [31] A. Zagrai, D. Doyle, and B. Arritt, "Embedded nonlinear ultrasonics for structural health monitoring of satellite joints," in *Proceedings-SPIE, Health Monitoring of Structural and Biological Systems*, vol. 6935, San Diego, CA, USA, March 2008.
- [32] N. Kambhatla and T. Leen, "Dimension reduction by local principal component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1493–1516, 1997.
- [33] V. Lopes and S. D. Silva, "Structural health monitoring algorithms for smart structures," in *Damage Prognosis: for Aerospace, Civil and Mechanical Systems*, John Wiley & Sons, Ltd., 2005.

Research Article

Multiplicative Fault Estimation-Based Adaptive Sliding Mode Fault-Tolerant Control Design for Nonlinear Systems

Ali Ben Brahim , Slim Dhahri, Fayçal Ben Hmida, and Anis Sellami

Engineering Laboratory of Industrial Systems and Renewable Energies, National Higher School of Engineers of Tunis, Tunis University, Tunis, Tunisia

Correspondence should be addressed to Ali Ben Brahim; alibenibrahimm@yahoo.com

Received 12 January 2018; Accepted 24 April 2018; Published 26 June 2018

Academic Editor: Chen Lu

Copyright © 2018 Ali Ben Brahim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article deals with the sliding mode fault-tolerant control (FTC) problem for a nonlinear system described under Takagi-Sugeno (T-S) fuzzy representation. In particular, the nonlinear system is corrupted with multiplicative actuator faults, process faults, and uncertainties. We start by constructing the separated FTC design to ensure robust stability of the closed-loop nonlinear system. First, we propose to conceive an adaptive observer in order to estimate nonlinear system states, as well as robust multiplicative fault estimation. The novelty of the proposed approach is that the observer gains are obtained by solving the multiobjective linear matrix inequality (LMI) optimization problem. Second, an adaptive sliding mode controller is suggested to offer a solution to stabilize the closed-loop system despite the occurrence of real fault effects. Compared with the separated FTC, this paper provides an integrated sliding mode FTC in order to achieve an optimal robustness interaction between observer and controller models. Thus, in a single-step LMI formulation, sufficient conditions are developed with multiobjective optimization performances to guarantee the stability of the closed-loop system. At last, nonlinear simulation results are given to illustrate the effectiveness of the proposed FTC to treat multiplicative faults.

1. Introduction

In the last few years, there has been a growing interest in fault-tolerant control (FTC) design based on a fault estimation (FE) technique (location, occurrence time, and magnitude). Active FTC takes a primordial place in modern control application, system reliability, and supervision of industrial technology.

The main challenge of active FTC is to conceive a robust controller such that the closed-loop system is stable with acceptable performances even with the presence simultaneously of faults and uncertainties. In the literature, several approaches have been proposed to explore this powerful issue (see for instance [1–6] and the references herein). In practical applications, most of the systems are complex and usually having hard nonlinearities, so it is significant to study FE and FTC for nonlinear systems. Since their excellent ability to describe a nonlinear system, very interesting approaches have represented nonlinear systems under the T-S fuzzy form [7]. Actually, in the presence of system

uncertainties, several attempts have been oriented to the fault diagnosis and FTC of nonlinear systems (see for instance [8–13] and the references herein).

Popular FE approaches have been developed in a precise and effective way for nonlinear systems, where fault is modelled as additive changes appearing in actuators or sensors [14–18]. The major drawback of the preceding approaches resides primarily on treating actuator and sensor faults with additive terms. However, in practical engineering, it is often the case when some actuator faults and component faults occur in a multiplicative form. Thus, multiplicative faults are mixed with the inputs and outputs of the system. In this way, estimation of the characteristic and magnitude of unknown multiplicative faults has been a growing interest in modern control theory in recent years. It is practically important to decouple their eventual parameter or structure effects in the system or in the process model subject to ameliorate FTC design for a large class of nonlinear systems. Special attentions have already been made in the application of observer design to achieve multiplicative FE for linear and

nonlinear systems [19, 20]. In [21–24], robust observers have been used to estimate unknown fault for linear systems. The developed schemes are based on treating multiplicative faults with additive terms. Nevertheless, this approach may not be practical in all situations. For this purpose, it is not an easy task; real effect multiplicative fault detection, location, and estimation especially for nonlinear systems have been a significant research activity in the past decade. Zhang et al. [25] considers the problem of fault diagnosis for a class of nonlinear systems with unstructured modeling uncertainty. The proposed approach addresses the detection and isolation of nonlinear fault function that are modeled as measurable signals. More recently, in [26], a robust adaptive observer FE approach is discussed in order to extract the real component fault effects for Lipschitz nonlinear systems subject to unknown disturbances.

1.1. Contributions. Regarding the fact that multiplicative faults have not yet been fully tackled, an FE-based adaptive sliding mode FTC scheme for a class of uncertain nonlinear systems, approximated by the T-S representation, is of great interest in this paper. Thus, sliding mode control has been widely studied and employed in industrial applications based on its computational simplicity and in particular strong robustness against uncertainties or disturbances. The main contributions of the present paper are divided as follows:

- (1) In the first scheme, we propose a separated sliding mode FTC for the closed-loop nonlinear system subject to both multiplicative faults and uncertainties. More precisely, it should be pointing that we consider multiplicative faults a partial loss of actuator effectiveness and parameter changes in the nonlinear system state matrix.
- (2) In the second scheme, this paper provides an integrated sliding mode FTC in order to achieve an optimal robustness interaction between observer and controller models. Thus, in a single-step LMI formulation, sufficient conditions are developed with H_∞ optimization performances to guarantee the stability of the closed-loop system. In particular, the fault nonlinear function satisfies a Lipschitz condition. In this study, we use a multiobjective LMI optimization approach in which the Lipschitz constant and uncertainty attenuation level are maximized simultaneously.

The remainder of this paper is organized as follows: Section 2 gives the description of the nonlinear system. In Section 3, we describe the proposed T-S adaptive observer design. Section 4 presents the sliding mode controller structure. Sections 5 and 6 propose, respectively, the design of separated and integrated sliding mode FTC schemes to stabilize the closed-loop system. The simulation example is given in Section 7 based on nonlinear simulation illustrating the effectiveness of the proposed schemes. Finally, Section 8 presents some concluding remarks.

2. Problem Formulation

Consider an uncertain nonlinear system governed by the following equations:

$$\begin{aligned}\dot{x}(t) &= \varphi_1(x(t), u(t), \xi(x, t), f(x, u, t)), \\ y(t) &= \varphi_2(x(t)),\end{aligned}\quad (1)$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}^m$ represents the control inputs, $y(t) \in \mathbb{R}^p$ denotes the measurement output vector, and $\xi(x, t) \in \mathbb{R}^l$ stands for the uncertainty vector. In the present paper, $f(x, u, t) \in \mathbb{R}^q$ represents the component and/or actuator gain fault which is described as

$$f(x, u, t) = \sum_{j=1}^q \theta_j(t) h_j(x, u, t), \quad (2)$$

where $\theta_j(t)$ is a vector of unknown function reflecting the magnitude of the time-varying or constant multiplicative faults. $h_j(x, u, t)$ represents the functional structure of the j th multiplicative faults and usually mixes with system states and/or inputs. Before starting the main results of this paper, we will make the following assumptions.

Assumption 1. The fault vector $\theta_j(t)$ is assumed to be unknown but bounded as $\|\theta_j(t)\| \leq \|\theta_{\max}\| = \rho$, $\forall j = 1, \dots, q$, where $\theta_{\max} \in \mathbb{R}^q$ is a known constant vector and ρ is a known positive constant.

Assumption 2. It is assumed that the nonlinear system states and inputs are all bounded before and after the occurrence of a fault, and fault nonlinear function structure $h_j(x, u, t)$ satisfies a Lipschitz condition locally on a set $M \subset \mathbb{R}^n$ in which

$$\|h_j(x_1, u, t) - h_j(x_2, u, t)\| \leq \gamma_j \|x_1 - x_2\|, \quad \forall x_1, x_2 \in M, \quad (3)$$

where $\gamma_j > 0$ is called a Lipschitz constant and $j = 1, \dots, q$.

Design Objective. This paper features a robust estimation of real effect factor $\theta_j(t)$, $\forall j = 1, \dots, q$, for the uncertain nonlinear system (1) subject both to multiplicative fault $f(x, u, t)$ given by (2) and to uncertainties $\xi(x, t)$. It was the main purpose of the paper to solve two problems by (i) estimating multiplicative fault magnitude using a robust adaptive observer and (ii) stabilizing the closed-loop nonlinear system, after the occurrence of multiplicative fault $f(x, u, t)$, using a robust adaptive sliding mode controller. To treat this powerful issue, this article introduces two different approaches: separated and integrated FE-based adaptive sliding mode FTC design.

We will make the following definition, notation, and lemma in obtaining the main results.

Definition. For an arbitrary matrix $X \in \mathbb{R}^{n \times m}$, if $X^+ \in \mathbb{R}^{m \times n}$ verifies $X^+X = I_m$, then $X^+ = (X^T X)^{-1} X^T$ is said to be the left-inverse of X .

Notation. The notation $(*)$ corresponds to the symmetry matrix block, $He(X)$ signifies $X + X^T$, and $\|\cdot\|$ stands for the standard norm symbol.

Lemma 1. For matrices \mathbf{X} and \mathbf{Y} with appropriate dimensions, the following condition holds:

$$X^T Y + Y^T X \leq \varepsilon^{-1} X^T X + \varepsilon Y^T Y, \quad (4)$$

where ε is the positive scalar.

3. T-S Adaptive Observer Design

Referring to an interpolation mechanism with the convex sum properties [7], the system (1) can be approximated by T-S fuzzy representation with multiplicative faults as follows:

$$\begin{aligned} \dot{\hat{x}}(t) = & \sum_{i=1}^k \mu_i(\zeta(t)) \{ \mathbf{A}_i \hat{x}(t) + \mathbf{B}_i u(t) + \mathbf{D}_i \xi(x, t) \\ & + \mathbf{M}_i \sum_{j=1}^q \theta_j(t) h_j(x, u, t) \}, \end{aligned} \quad (5)$$

$$y(t) = \sum_{i=1}^k \mu_i(\zeta(t)) \{ \mathbf{C}_i \hat{x}(t) \},$$

where \mathbf{A}_i , \mathbf{B}_i , \mathbf{D}_i , \mathbf{M}_i , and \mathbf{C}_i are real known matrices with appropriate dimensions. We assume that the pair $(\mathbf{A}_i, \mathbf{B}_i)$ is controllable and the pair $(\mathbf{A}_i, \mathbf{C}_i)$ is observable. Let $\mu_i(\zeta(t))$ be the normalized fuzzy membership functions which satisfy the properties of the sum convex.

$$\sum_{i=1}^k \mu_i(\zeta(t)) = 1, 0 \leq \mu_i(\zeta(t)) \leq 1, \quad \forall i \in [1, \dots, k]. \quad (6)$$

Through the approximation of the nonlinear system (1) by augmented T-S fuzzy representation (5), we construct a multiplicative fault estimation adaptive observer as

$$\begin{aligned} \dot{\hat{x}}(t) = & \sum_{i=1}^k \mu_i(\zeta(t)) \{ \mathbf{A}_i \hat{x}(t) + \mathbf{B}_i u(t) + \mathbf{G}_{l,i} e_y(t) \\ & + \mathbf{M}_i \sum_{j=1}^q \hat{\theta}_j(t) h_j(\hat{x}, u, t) \}, \end{aligned} \quad (7)$$

$$\hat{y}(t) = \sum_{i=1}^k \mu_i(\zeta(t)) \{ \mathbf{C}_i \hat{x}(t) \}, \quad (8)$$

$$\hat{\theta}_j(t) = \sigma_j h_j^T(\hat{x}, u, t) \Gamma e_y(t), \quad j = 1, \dots, q, \quad (9)$$

where σ_j are positive scalars, $\hat{x}(t)$ is the observer state, $\hat{y}(t)$ represents the observer output, $\hat{\theta}_j(t)$ denotes the estimated fault magnitude, and $e_y(t) = y(t) - \hat{y}(t)$ is the output estimation error. $\mathbf{G}_{l,i}$ are appropriate gain matrices, which can be obtained using LMIs as discussed later. Γ is a design matrix representing the learning rate.

From now on, we assume that the state and fault estimation errors are defined, respectively, as $e(t) = x(t) - \hat{x}(t)$ and $e_{\theta_j}(t) = \theta_j(t) - \hat{\theta}_j(t)$. It remains to deduce that

$$\begin{aligned} \dot{e}(t) = & \sum_{i=1}^k \mu_i(\zeta(t)) \{ [\mathbf{A}_i - \mathbf{G}_{l,i} \mathbf{C}_i] e(t) + \mathbf{D}_i \xi(x, t) \\ & + \mathbf{M}_i \sum_{j=1}^q [\theta_j(t) h_j(x, u, t) - \hat{\theta}_j(t) h_j(\hat{x}, u, t)] \}, \end{aligned} \quad (10)$$

$$\dot{e}_{\theta_j}(t) = \sum_{i=1}^k \mu_i(\zeta(t)) \{ -\sigma_j h_j^T(\hat{x}, u, t) \Gamma e_y(t) \}. \quad (11)$$

The objective is to derive the gains of the robust adaptive observer ((7), (8), and (9)) in order to estimate multiplicative fault magnitudes.

4. Sliding Mode Controller Design

4.1. Adaptive Sliding Mode Controller Structure. The proposed sliding mode controller with adaptive law is assigned to provide a corrective action in order to compensate multiplicative fault effects and stabilize the nonlinear system described by T-S fuzzy representation. Before starting FTC design, we assume the following:

Assumption 3. $\text{rank}(C_i B_i) = \text{rank}(B_i), \forall i \in [1, \dots, k]$. As the first step, one can define the sliding surface \mathbb{S} , when the sliding motion will take place on it, as

$$\mathbb{S} = \{ y(t) \in \mathbb{R}^p : S_c(t) = 0 \}. \quad (12)$$

$S_c(t) \in \mathbb{R}^m$ is a linear switching function, based on the output feedback information, described as

$$S_c(t) = \sum_{i=1}^k \mu_i(\zeta(t)) \{ N_{c,i} y(t) \}, \quad (13)$$

where $N_{c,i} = [C_i B_i]^+ - h [I_p - C_i B_i C_i B_i^+]$ with an arbitrary matrix $h \in \mathbb{R}^{m \times p}$. As mentioned above that (A_i, B_i) is controllable, a nonlinear control input is given by

$$u(t) = u_l(t) + u_n(t), \quad (14)$$

where $u_l(t)$ designs the linear part which is defined as

$$u_l(t) = \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j(\zeta(t)) \{ -K_j \hat{x}(t) - F_{a,i} \hat{f}(x, u, t) \}, \quad (15)$$

where $-F_{a,i} \hat{f}(x, u, t)$ is designed to compensate multiplicative fault influence. It is assumed that $K_j \in \mathbb{R}^{m \times n}$ and $F_{a,i} = B_i^+ M_i$.

As may be seen below, the nonlinear part $u_n(t)$, capable of inducing the sliding motion on the sliding surface \mathbb{S} , is proposed with adaptive law as

$$u_n(t) = \begin{cases} -\eta_c(t) \frac{S_c(t)}{\|S_c(t)\|}, & \text{if } S_c(t) = 0, \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

where $\eta_c(t) = \hat{\rho}_c + \varrho_c$, where $\varrho_c > 0$ is a small constant and $\hat{\rho}_c$ is used to determinate $\eta_c(t)$ such that we will make the following adaptive term as

$$\dot{\hat{\rho}}_c = \epsilon_c \|S_c(t)\|, \quad \hat{\rho}_c(0) \geq 0, \quad (17)$$

where ϵ_c is the positive gain.

4.2. Reaching Condition. As mentioned earlier, it must be proven, using the nonlinear part structure of $u_n(t)$, that the system will be forced to reach and slide onto the corresponding sliding mode surface \mathbb{S} in a finite time. In this way, we design a Lyapunov function as

$$V_c(t) = \frac{1}{2} S_c^T(t) S_c(t) + \frac{1}{2\epsilon_c} \tilde{\rho}_c^2, \quad (18)$$

where $\tilde{\rho}_c = \rho_c - \hat{\rho}_c$ is the estimated error of ρ_c .

The derivative of (18) with respect to time gives

$$\begin{aligned} \dot{V}_c(t) &= \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j (\zeta(t)) \{ (N_{c,i} C_i A_i - K_j) \|x(t)\| \\ &+ \rho_c - \eta_c(t) \|S_c(t)\| - \tilde{\rho}_c \|S_c(t)\| = \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j (\zeta(t)) \\ &\cdot \{ (N_{c,i} C_i A_i - K_j) \|x(t)\| - \varrho_c - \epsilon_c \} \|S_c(t)\|. \end{aligned} \quad (19)$$

Define the subset system as

$$\Omega_c = \{x \| \|x(t)\| \leq \kappa_c\}. \quad (20)$$

The reachability condition, which guarantees to force the system to attain the sliding surface \mathbb{S} , is satisfied if the scalar ϱ_c is chosen to satisfy $\varrho_c > (N_{c,i} C_i A_i - K_j) \kappa_c$ such that

$$S_c^T(t) S_c(t) \leq -\epsilon_c \|S_c(t)\|. \quad (21)$$

Furthermore, the proposed sliding mode controller with adaptive law ensures the existence of an ideal sliding motion in finite time; that is, $S_c(t) = \dot{S}_c(t) = 0, \forall t \geq t_c$.

5. FE-Based Fault-Tolerant Control Design: A Separated Approach

For several years, great effort has been devoted to study the FE-based FTC problem in a precise and effective way. The focus of previous studies has been on the division of this issue into separate steps:

Step 1. Conceive an observer to estimate faults and state variables.

Step 2. Conceive a controller to stabilize the closed-loop systems.

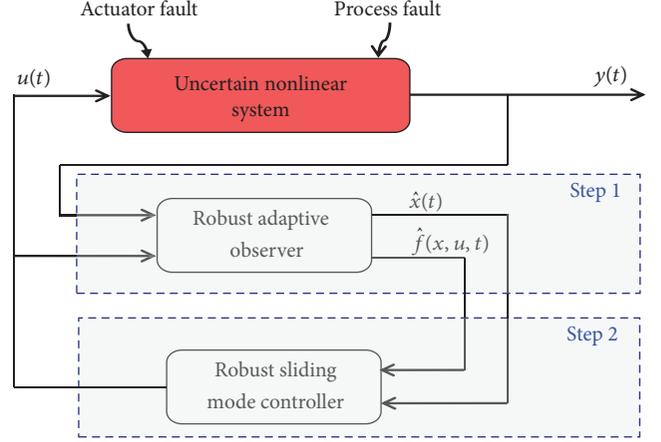


FIGURE 1: FE-based fault-tolerant control: a separated approach.

Figure 1 illustrates the separated FE-based FTC design for uncertain nonlinear systems subject to process faults and multiplicative actuator faults.

Now, we propose to design the separated FE-based FTC problem in order to compute adaptive observer gains $G_{l,i}$ and controller gains K_j such that a robust stability of the closed-loop nonlinear system, described by the T-S form, is achieved despite the presence of multiplicative faults and uncertainties.

5.1. LMI Optimization-Based Observer Stability. Theorem 1 establishes the sufficient conditions for the stability of the observer errors ((10) and (11)) with prescribed \mathcal{H}_∞ performances by using Lyapunov stability and LMI technique.

Theorem 1. *The state estimation error is robustly stable with simultaneously maximized admissible Lipschitz constant $\gamma_j > 0$ and minimized gain $\varsigma > 0$ for the system uncertainties $\xi(x, t)$, if there exist constants, $0 \leq \lambda \leq 1$, $\epsilon > 0$, and $\alpha_j > 0$, and matrices $P_e = P_e^T > 0$, W_i , such that the following multiobjective optimization problem has a solution:*

$$\min \left[\lambda \left(\epsilon + \sum_{j=1}^q \alpha_j \right) + (1 - \lambda) \varsigma \right], \quad (22)$$

subject to the following LMI

$$\begin{bmatrix} \bar{\Xi}_i + C_e^T C_e & P_e D_i & P_e M_i & I_n \\ (*) & -\varsigma I_l & 0 & 0 \\ (*) & (*) & -\epsilon I_q & 0 \\ (*) & (*) & (*) & -\sum_{j=1}^q \alpha_j \end{bmatrix} < 0, \quad (23)$$

where $\bar{\Xi}_i = A_i^T P_e + P_e A_i - W_i C_i - C_i^T W_i^T$. One can prove that the observer gain can be obtained from $G_{l,i} = P_e^{-1} W_i$. In addition, the T-S adaptive observer ((7), (8), and (9)) ensures that the estimated $\hat{x}(t)$ and $\hat{\theta}_j(t)$ converge to the

nonlinear system state $x(t)$ and the multiplicative fault magnitude $\theta_j(t)$.

Proof. One can start by investigating the following Lyapunov function:

$$V_e(t) = e^T(t)P_e e(t) + \sum_{j=1}^q \sigma_j^{-1} e_{\theta_j}^T(t) e_{\theta_j}(t) > 0, \quad (24)$$

where $P_e \in \mathbb{R}^{n \times n}$ is the design Lyapunov matrix. The time derivative of $V_e(t)$ is handled as

$$\begin{aligned} \dot{V}_e(t) &= \dot{e}^T(t)P_e e(t) + e^T(t)P_e \dot{e}(t) \\ &+ \sum_{j=1}^q \sigma_j^{-1} \left[\dot{e}_{\theta_j}^T(t) e_{\theta_j}(t) + e_{\theta_j}^T(t) \dot{e}_{\theta_j}(t) \right]. \end{aligned} \quad (25)$$

Based on state and fault estimation errors ((10) and (11)), $\dot{V}_e(t)$ can be written as

$$\begin{aligned} \dot{V}_e(t) &= \sum_{i=1}^k \mu_i(\zeta(t)) \left\{ e^T(t) [A_{i,i}^T P_e + P_e A_{i,i}] e(t) \right. \\ &+ \sum_{j=1}^q 2e^T(t) P_e M_i \left[\theta_j(t) h_j(x, u, t) - \hat{\theta}_j(t) h_j(\hat{x}, u, t) \right] \\ &+ \left. \sum_{j=1}^q 2\sigma_j^{-1} e_{\theta_j}(t) \dot{\hat{\theta}}_j(t) + 2e^T(t) P_e D_i \xi(x, t) \right\}, \end{aligned} \quad (26)$$

where $A_{i,i} = A_i - G_{i,i} C_i$. From this, one can conclude that

$$\begin{aligned} \dot{V}_e(t) &= \sum_{i=1}^k \mu_i(\zeta(t)) \left\{ e^T(t) [A_{i,i}^T P_e + P_e A_{i,i}] e(t) \right. \\ &+ \sum_{j=1}^q 2e^T(t) P_e M_i \left[\theta_j(t) h_j(x, u, t) - \theta_j(t) h_j(\hat{x}, u, t) \right] \\ &+ \left. 2e^T(t) P_e D_i \xi(x, t) \right\}. \end{aligned} \quad (27)$$

Referring to Lemma 1 and according to the Lipschitz condition (3), one can further derive

$$\begin{aligned} &2e^T(t) P_e M_i \left[\theta_j(t) h_j(x, u, t) - \theta_j(t) h_j(\hat{x}, u, t) \right] \\ &\leq \frac{1}{\varepsilon} e^T(t) P_e M_i M_i^T P_e e(t) + \varepsilon [h_j(x, u, t) - h_j(\hat{x}, u, t)]^T \\ &\quad \theta_j^T(t) \theta_j(t) [h_j(x, u, t) - h_j(\hat{x}, u, t)] \\ &\leq \frac{1}{\varepsilon} e^T(t) P_e M_i M_i^T P_e e(t) + \varepsilon \|\theta_j(t)\|^2 \\ &\quad \|h_j(x, u, t) - h_j(\hat{x}, u, t)\|^2 \\ &\leq \frac{1}{\varepsilon} e^T(t) P_e M_i M_i^T P_e e(t) + \varepsilon \rho^2 \gamma_j^2 \|e(t)\|^2 \\ &= \frac{1}{\varepsilon} e^T(t) P_e M_i M_i^T P_e e(t) + \varepsilon \tilde{\gamma}_j^2 \|e(t)\|^2. \end{aligned} \quad (28)$$

where $\tilde{\gamma}_j = \rho \gamma_j$. Now, according to (28), one can derive $V_e(t)$ as

$$\begin{aligned} \dot{V}_e(t) &\leq \sum_{i=1}^k \mu_i(\zeta(t)) \left\{ e^T(t) [A_{i,i}^T P_e + P_e A_{i,i}] e(t) \right. \\ &+ \frac{1}{\varepsilon} e^T(t) P_e M_i M_i^T P_e e(t) + \sum_{j=1}^q e^T(t) \varepsilon \tilde{\gamma}_j^2 e(t) \\ &+ \left. 2e^T(t) P_e D_i \xi(x, t) \right\}. \end{aligned} \quad (29)$$

To attain the robustness of the proposed multiplicative fault estimation adaptive observer ((7), (8), and (9)) against system uncertainties $\xi(x, t)$, we investigate the controlled estimation error $r(t)$ as

$$r(t) = C_e e(t). \quad (30)$$

Consider the following worst-case performance measure:

$$\|H\|_{\infty} = \sup_{\|\xi\|_2 \neq 0} \frac{\|r(t)\|_2^2}{\|\xi(x, t)\|_2^2} \leq \varsigma. \quad (31)$$

One can now proceed with the presence of the following variable $J_e(t)$ as

$$J_e(t) = \dot{V}_e(t) + r^T(t)r(t) - \varsigma \xi^T(x, t)\xi(x, t). \quad (32)$$

Obviously, one can write the above expression (32) by

$$\begin{aligned} J_e(t) &\leq \sum_{i=1}^k \mu_i(\zeta(t)) \left\{ e^T(t) \left[A_{i,i}^T P_e + P_e A_{i,i} + \frac{1}{\varepsilon} P_e M_i M_i^T P_e \right. \right. \\ &+ \sum_{j=1}^q \varepsilon \tilde{\gamma}_j^2 + C_e^T C_e \left. \right] e(t) + 2e^T(t) P_e D_i \xi(x, t) \\ &- \left. \varsigma \xi^T(x, t)\xi(x, t) \right\}. \end{aligned} \quad (33)$$

Let one define the following new variable as

$$\alpha_j = \frac{1}{\varepsilon \tilde{\gamma}_j^2}. \quad (34)$$

From now on, one can get

$$\tilde{\gamma}_j = \frac{1}{\sqrt{\varepsilon \alpha_j}}. \quad (35)$$

The stability of the T-S fuzzy system (5) is achieved for any fault Lipschitz nonlinear function with Lipschitz constant less than or equal to an unknown maximized constant $\tilde{\gamma}_j, \forall j = 1, \dots, q$. Maximization of $\tilde{\gamma}_j$ and minimization of ς can be accomplished by simultaneous minimization of α_j, ε ,

and $\varsigma, \forall j = 1, \dots, q$. In this way, one can obtain a multiobjective optimization design.

As is clear from (34), one has

$$\begin{aligned} J_{e(t)} \leq & \sum_{i=1}^k \mu_i(\zeta(t)) \left\{ e^T(t) \left[A_{l,i}^T P_e + P_e A_{l,i} + \frac{1}{\varepsilon} P_e M_i M_i^T P_e \right. \right. \\ & + \sum_{j=1}^q \alpha_j^{-1} + C_e^T C_e \left. \right\} e(t) + 2e^T(t) P_e D_i \xi(x, t) \\ & - \varsigma \xi^T(x, t) \xi(x, t) \} = \Phi^T(t) \Delta \Phi(t), \end{aligned} \quad (36)$$

where

$$\begin{aligned} \Phi(t) &= \begin{bmatrix} e(t) \\ \xi(x, t) \end{bmatrix}, \\ \Delta &= \sum_{i=1}^k \mu_i(\zeta(t)) \begin{bmatrix} \Xi_{ij} + C_e^T C_e & P_e D_i \\ D_i^T P_e & -\varsigma \end{bmatrix}, \\ \Xi_{ij} &= A_{l,i}^T P_e + P_e A_{l,i} + \frac{1}{\varepsilon} P_e M_i M_i^T P_e + \sum_{j=1}^q \alpha_j^{-1}. \end{aligned} \quad (37)$$

One can conclude that if $\Delta < 0$, it implies that $J_e(t) < 0$.

According to the Schur complement, the previous inequality is satisfied if and only if we checked for the presence of this relation:

$$\sum_{i=1}^k \mu_i(\zeta(t)) \Omega_{ij} < 0, \quad (38)$$

where Ω_{ij} are the same as those in (23).

In this way, the state estimation error is asymptotically stable with the attenuation level ς as follows:

$$\|r(t)\|_2^2 \leq \varsigma \|\xi(x, t)\|_2^2. \quad (39)$$

It is obvious that $e(t) = x(t) - \hat{x}(t) \rightarrow 0$. Due to the fault, nonlinear function $h_j(x, u, t)$ satisfies the Lipschitz condition; the T-S multiplicative fault estimation adaptive observer ((7), (8), and (9)) ensures that $e_{\theta_j}(t) \rightarrow 0$. From this, one can deduce that, according to (9), the estimation of multiplicative fault magnitude $\theta_j(t)$, for the uncertain nonlinear system (1) described by T-S fuzzy structure (5), can be achieved.

This completes the proof.

5.2. LMI Optimization-Based Closed-Loop Stability. Once the sliding mode is obtained, we consider to analyze the stability of the closed-loop T-S fuzzy system. Let the equivalent control $u_{\text{eq}}(t)$, such that $\dot{S}_c(t)$ is equal to zero, be

$$u_{\text{eq}}(t) = \sum_{i=1}^k \mu_i(\zeta(t)) \{ -N_{c,i} C_i [A_i x(t) + D_i \xi(x, t)] + u_i(t) \}. \quad (40)$$

The dynamic of the closed-loop system with the equivalent control law (40) takes the form

$$\dot{x}(t) = \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j(\zeta(t)) \{ (\Theta_i A_i - B_i K_j) x(t) + \bar{B}_{i,j} \phi(t) \}, \quad (41)$$

$$y(t) = \sum_{i=1}^k \mu_i(\zeta(t)) \{ C_i x(t) \}, \quad (42)$$

where $\bar{B}_{i,j} = [B_i K_j M_i \Theta_i D_i]$, $\Theta_i = I_n - B_i N_{c,i} C_i$, and $\phi(t) = \begin{bmatrix} e^T(t) & e_f^T(x, u, t) & \xi^T(x, t) \end{bmatrix}^T$, where $e_f(x, u, t) = f(x, u, t) - \hat{f}(x, u, t)$.

The objective now is to develop a sufficient condition to achieve the stability of the closed-loop T-S fuzzy system ((41) and (42)) on the sliding surface S despite the occurrence of multiplicative faults and the presence of uncertainties.

Theorem 2. *The closed-loop T-S fuzzy system ((41) and (42)) is robustly stable with the H_∞ attenuation level $\varsigma_s > 0$, if there exist the matrices $\bar{P}_x = \bar{P}_x^T > 0$, Q_j , and $Y = Y^T$, such that*

$$\min \quad [\varsigma_s] \quad (43)$$

satisfying the following LMI constraints:

$$\Delta_{\text{SepaFTC}} = \begin{bmatrix} Y_i, j & B_i Q_j & M_i & \Theta_i D_i & \bar{P}_x C_i^T \\ (*) & -2Y + \varsigma_s I_n & 0 & 0 & 0 \\ (*) & (*) & -\varsigma_s I_q & 0 & 0 \\ (*) & (*) & (*) & -\varsigma_s I_l & 0 \\ (*) & (*) & (*) & (*) & -I_p \end{bmatrix} < 0, \quad (44)$$

where $Y_i, j = \bar{P}_x A_i^T \Theta_i^T + \Theta_i A_i \bar{P}_x - B_i Q_j - Q_j^T B_i^T$, $Y = \varsigma_s \bar{P}_x$, and $K_j = Q_j \bar{P}_x^{-1}$.

Proof. Consider the following Lyapunov function for the closed-loop system as

$$V_x(t) = x^T(t) P_x x(t), \quad (45)$$

where $P_x \in \mathbb{R}^{n \times n}$ is the symmetric positive definite matrix. The time derivative of $V_x(t)$ is handled as

$$\begin{aligned} \dot{V}_x(t) &= \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j(\zeta(t)) \left\{ x^T(t) \left[(\Theta_i A_i - B_i K_j)^T P_x \right. \right. \\ &\quad \left. \left. + P_x (\Theta_i A_i - B_i K_j) \right] x(t) + 2x^T(t) P_x \bar{B}_{i,j} \phi(t) \right\}. \end{aligned} \quad (46)$$

To achieve the robustness with H_∞ performance of the closed-loop T-S fuzzy system ((41) and (42)) to $\phi(t)$, the following inequality must then hold:

$$J_x(t) = \dot{V}_x(t) + y^T(t)y(t) - \varsigma_s \phi^T(t)\phi(t) < 0. \quad (47)$$

Insertion of (46) in (47) yields

$$\begin{aligned} J_x(t) = & \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j (\zeta(t)) \left\{ x^T(t) \left[(\Theta_i A_i - B_i K_j)^T P_x \right. \right. \\ & + P_x (\Theta_i A_i - B_i K_j) + C_i^T C_i \left. \right] x(t) + 2x^T(t) P_x \bar{B}_{i,j} \phi(t) \\ & \left. - \varsigma_s \phi^T(t)\phi(t) \right\}. \end{aligned} \quad (48)$$

Consequently (in the matrix form), it remains to prove that $J_x(t) < 0$, if

$$\begin{bmatrix} \psi_{i,j} & P_x \bar{B}_{i,j} \\ \bar{B}_{i,j}^T P_x & -\varsigma_s I \end{bmatrix} < 0, \quad (49)$$

where $\psi_{i,j} = (\Theta_i A_i - B_i K_j)^T P_x + P_x (\Theta_i A_i - B_i K_j) + C_i^T C_i$. Using the Schur complement, the relation (49) can be reformulated as

$$\prod_{i,j} = \begin{bmatrix} \sum_{i,j} P_x B_i K_j & P_x M_i & P_x \Theta_i D_i & C_i^T \\ (*) & -\varsigma_s I_n & 0 & 0 & 0 \\ (*) & (*) & -\varsigma_s I_q & 0 & 0 \\ (*) & (*) & (*) & -\varsigma_s I_l & 0 \\ (*) & (*) & (*) & (*) & -I_p \end{bmatrix} < 0, \quad (50)$$

where $\sum_{i,j} = \Theta_i^T A_i^T P_x + P_x \Theta_i A_i - K_j^T B_i^T P_x - P_x B_i K_j$.

Inequality (50) contains several nonlinear terms. One can design in the next step to formulate this as an LMI problem. To effect the necessary change of variables, one will define the following matrix \mathbf{X} with the special diagonal structure as $\mathbf{X} = \text{diag}\{P_x^{-1}, P_x^{-1}, I_q, I_l, I_p\}$. Then, $\mathbf{X} \times \prod_{i,j} \times \mathbf{X}^T < 0$ is true, and it is obvious that

$$\begin{bmatrix} \Upsilon_{i,j} & B_i Q_j & M_i & \Theta_i D_i & \bar{P}_x C_i^T \\ (*) & -\varsigma_s \bar{P}_x \bar{P}_x & 0 & 0 & 0 \\ (*) & (*) & -\varsigma_s I_q & 0 & 0 \\ (*) & (*) & (*) & -\varsigma_s I_l & 0 \\ (*) & (*) & (*) & (*) & -I_p \end{bmatrix} < 0, \quad (51)$$

where $\Upsilon_{i,j} = \bar{P}_x A_i^T \Theta_i^T + \Theta_i A_i \bar{P}_x - B_i Q_j - Q_j^T B_i^T$, $\bar{P}_x = P_x^{-1}$, and $Q_j = K_j P_x^{-1}$. According to Lemma 1, it is evident to check the presence of the following relation

$$\bar{P}_x + \bar{P}_x \leq \bar{P}_x \bar{P}_x + I_n. \quad (52)$$

Obviously, (52) is true for $\varsigma_c > 0$ as

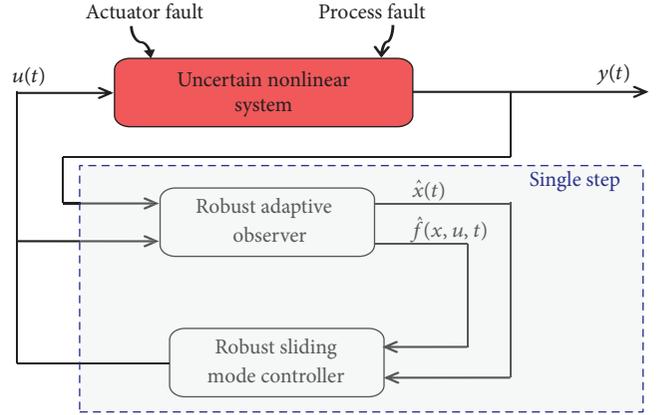


FIGURE 2: FE-based fault-tolerant control: an integrated approach.

$$-\varsigma_s \bar{P}_x \bar{P}_x \leq -2\varsigma_s \bar{P}_x + \varsigma_s I_n. \quad (53)$$

After some manipulations, one can get

$$\sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j (\zeta(t)) \Delta_{\text{SepaFTC}} < 0, \quad (54)$$

where Δ_{SepaFTC} has the same structure with (44). Clearly, a stability proof of the closed-loop T-S fuzzy system ((41) and (42)) is required with respect to the H_∞ performance level ς_s .

This completes the proof.

6. FE-Based Fault-Tolerant Control Design: An Integrated Approach

Several publications have appeared in recent years documenting FE-based FTC design with a single step in order to achieve an optimal robustness interaction between observer and controller models. Figure 2 illustrates the integrated FE-based FTC design for uncertain nonlinear systems subject to process fault and multiplicative actuator fault.

In this section, we explore the possibility of the integrated FE-based FTC design to compute, in a single step, adaptive observer gains $G_{i,i}$ and controller gains K_j in the sense that it ameliorates the robustness of the closed-loop nonlinear system despite the presence of multiplicative faults and uncertainties. Combining (10), (11), and (41) gives the following augmented closed-loop system, including fault estimation with fault compensation control, expressed as

$$\dot{x}(t) = \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j (\zeta(t)) \left\{ (\Theta_i A_i - B_i K_j) x(t) + \bar{B}_{i,j} \phi(t) \right\}, \quad (55)$$

$$\begin{aligned} \dot{e}(t) = & \sum_{i=1}^k \mu_i (\zeta(t)) \left\{ [A_i - G_{i,i} C_i] e(t) + \bar{D}_i^1 \phi(t) \right. \\ & \left. + M_i \sum_{j=1}^q \left[\theta_j(t) h_j(x, u, t) - \hat{\theta}(t) h_j(\hat{x}, u, t) \right] \right\}, \end{aligned} \quad (56)$$

$$\dot{e}_{\theta_j}(t) = \sum_{i=1}^k \mu_i(\zeta(t)) \left\{ -\sigma_j h_j^T(\tilde{x}, u, t) \Gamma_i e_y(t) \right\}, \quad (57)$$

$$y_L(t) = \sum_{i=1}^k \mu_i(\zeta(t)) \{ C_{L,i} x(t) + C_{e,i} e(t) \}, \quad (58)$$

where $\bar{B}_{i,j} = [B_i K_j \quad M_i \quad \Theta_i D_i]$, $\bar{D}_i^1 = [0 \quad 0 \quad D_i]$, $\Theta_i = I_n - B_i N_{c,i} C_i$, and $\phi(t) = \left[e(t)^T \quad e_f^T(x, u, t) \quad \xi^T(x, t) \right]^T$.

Theorem 3. Under the sliding mode input structure (14), the closed-loop T-S fuzzy system ((55), (56), (57), and (58)) is robustly stable with both maximized admissible Lipschitz constant $\gamma_{e,j} > 0$ and minimized gain $\|H_{yL\phi}\|_{\infty} < \varsigma_c$, if there exist constants, $0 \leq \lambda_c \leq 1$, $\varepsilon_e > 0$, and $\alpha_{e,j} > 0$, and matrices $P_e = P_e^T > 0$, $\bar{P}_x = \bar{P}_x^T > 0$, W_i , Q_j , and $Y = Y^T$, such that the multiobjective LMI optimization problem admits a solution as

$$\min \left[\lambda_c \left(\varepsilon_e + \sum_{j=1}^q \alpha_{e,j} \right) + (1 - \lambda_c) \varsigma_c \right], \quad (59)$$

$$\text{subject to } \bar{\Xi}_{ij} = \begin{bmatrix} \bar{\Xi}_{11,ij} & \bar{\Xi}_{12,ij} \\ (*) & \bar{\Xi}_{22,ij} \end{bmatrix} < 0,$$

where

$$\begin{aligned} \bar{\Xi}_{11,ij} &= \begin{bmatrix} \Xi_{11,\neq j} & 0 \\ (*) & \Xi_{22,i} \end{bmatrix}, \\ \bar{\Xi}_{11,ij} &= He(\Theta_i A_i \bar{P}_x) - He(B_i Q_j), \\ \bar{\Xi}_{22,i} &= He(P_e A_i) - He(W_i C_i) + C_{e,i}^T C_{e,i} W_i = P_e G_{i,i} \\ \bar{\Xi}_{12,ij} &= \begin{bmatrix} B_i Q_j & M_i & \Theta_i D_i & \bar{P}_x C_{L,i}^T & 0 & 0 \\ 0 & 0 & P_e D_i & 0 & P_e M_i & I_n \end{bmatrix}, \\ \bar{\Xi}_{22,ij} &= -\text{diag} \left[2Y - \varsigma_c I_n, \varsigma_c I_q, \varsigma_c I_l, I_p, \varepsilon_e I_q, \sum_{j=1}^q \alpha_{e,j}^{-1} \right], \\ Y &= \varsigma_c \bar{P}_x, \bar{P}_x = P_x^{-1}, Q_j = K_j P_x^{-1}. \end{aligned} \quad (60)$$

The gain matrices of the adaptive sliding mode controller K_j and observer $G_{i,i}$ are given by

$$\begin{aligned} K_j &= Q_j \bar{P}_x^{-1}, \\ G_{i,i} &= P_e^{-1} W_i. \end{aligned} \quad (61)$$

Proof. Stability analysis: In order to assure the stability of the augmented closed-loop system ((55), (56), (57), and (58)), one can start by investigating the following Lyapunov function as

$$V(t) = V_e(t) + V_x(t), \quad (62)$$

where $V_e(t) = e^T(t) P_e e(t) + \sum_{j=1}^q \sigma_j^{-1} e_{\theta_j}^T(t) e_{\theta_j}(t) > 0$ and $V_x(t) = x^T(t) P_x x(t) > 0$, and $P_x \in \mathbb{R}^{n \times n}$ is the symmetric positive definite matrix.

As first, one can proceed analogously to Theorem 1. Hence, the time derivative of $V_e(t)$ is bounded as

$$\begin{aligned} \dot{V}_e(t) &\leq \sum_{i=1}^k \mu_i(\zeta(t)) \left\{ e^T(t) \left[A_{L,i}^T P_e + P_e A_{L,i} + \frac{1}{\varepsilon_e} P_e M_i M_i^T P_e \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^q \alpha_{e,j}^{-1} \right] e(t) + 2e^T(t) P_e \bar{D}_i^1 \phi(t) \right\}. \end{aligned} \quad (63)$$

On the other hand, similar to Theorem 2 and by taking into account the closed-loop T-S fuzzy system (55), the time derivative of $V_x(t)$ is expressed as

$$\begin{aligned} \dot{V}_x(t) &= \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j(\zeta(t)) \left\{ x^T(t) \left[(\Theta_i A_i - B_i K_j)^T P_x \right. \right. \\ &\quad \left. \left. + P_x (\Theta_i A_i - B_i K_j) \right] x(t) + 2x^T(t) P_x \bar{B}_{ij} \phi(t) \right\}. \end{aligned} \quad (64)$$

Robust performance index: Let

$$J(t) = \dot{V}(t) + y_L^T(t) y_L(t) - \varsigma_c \phi^T(t) \phi(t) < 0. \quad (65)$$

The inequality (65), after substituting (63) and (64), becomes

$$\begin{aligned} J(t) &\leq \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j(\zeta(t)) \left\{ x^T(t) \left[(B_{ij}^1)^T P_x \right. \right. \\ &\quad \left. \left. + P_x B_{ij}^1 + C_{L,i}^T C_{L,i} \right] x(t) + e^T(t) \left[A_{L,i}^T P_e + P_e A_{L,i} \right. \right. \\ &\quad \left. \left. + \frac{1}{\varepsilon_e} P_e M_i M_i^T P_e + \sum_{j=1}^q \alpha_{e,j}^{-1} + C_{e,i}^T C_{e,i} \right] e(t) \right. \\ &\quad \left. + 2x^T(t) P_x \bar{B}_{ij} \phi(t) + 2e^T(t) P_e \bar{D}_i^1 \phi(t) - \varsigma_c \phi^T(t) \phi(t) \right\}. \end{aligned} \quad (66)$$

where $B_{ij}^1 = \Theta_i A_i - B_i K_j$. Equivalently, in the matrix form, one can obtain the following expression as

$$J(t) \leq \sum_{i=1}^k \sum_{j=1}^k \mu_i \mu_j(\zeta(t)) \left\{ \chi^T(t) \Gamma_{ij} \chi(t) \right\}, \quad (67)$$

where $\chi(t) = [x^T(t) \quad e^T(t) \quad \phi^T(t)]^T$. The variable Γ_{ij} is defined as

$$\Gamma_{ij} = \begin{bmatrix} \Gamma_{11,ij} & 0 & P_x B_i K_j & P_x M_i & P_x \Theta_i D_i \\ (*) & \Gamma_{22,i} & 0 & 0 & P_e D_i \\ (*) & (*) & -\varsigma_c I_n & 0 & 0 \\ (*) & (*) & (*) & -\varsigma_c I_q & 0 \\ (*) & (*) & (*) & (*) & -\varsigma_c I_l \end{bmatrix}, \quad (68)$$

such that

$$\begin{aligned}\Gamma_{11,ij} &= He\left([\Theta_i A_i - B_i K_j]^T P_x\right) + C_{L,i}^T C_{L,i}, \\ \Gamma_{22,i} &= He\left(A_{L,i}^T P_e\right) + \frac{1}{\varepsilon_e} P_e M_i M_i^T P_e + \sum_{j=1}^q \alpha_{e,j}^{-1} + C_{e,i}^T C_{e,i}.\end{aligned}\quad (69)$$

To effect the necessary change of variables, one will make the following matrix \mathbf{X} with the special diagonal structure as $\mathbf{X} = \text{diag}\{P_x^{-1}, I_n, P_x^{-1}, I_q, I_l\}$. After pre- and postmultiplying by X and its transpose in Γ_{ij} , then it is obvious that

$$\bar{\Gamma}_{ij} = \begin{bmatrix} \bar{\Gamma}_{11,ij} & 0 & B_i Q_j & M_i & \Theta_i D_i \\ (*) & \bar{\Gamma}_{22,i} & 0 & 0 & P_e D_i \\ (*) & (*) & -\zeta_c \bar{P}_x \bar{P}_x & 0 & 0 \\ (*) & (*) & (*) & -\zeta_c I_q & 0 \\ (*) & (*) & (*) & (*) & -\zeta_c I_l \end{bmatrix}, \quad (70)$$

where

$$\begin{aligned}\bar{\Gamma}_{11,ij} &= He(\Theta_i A_i \bar{P}_x) - He(B_i Q_j) + \bar{P}_x C_{L,i}^T C_{L,i} \bar{P}_x, \\ \bar{\Gamma}_{22,i} &= He(P_e A_i) + He(W_i C_i) + \frac{1}{\varepsilon_e} P_e M_i M_i^T P_e \\ &\quad + \sum_{j=1}^q \alpha_{e,j}^{-1} + C_{e,i}^T C_{e,i}, \\ \bar{P}_x &= P_x^{-1}, Q_j = K_j \bar{P}_x^{-1}, W_i = P_e G_{L,i}.\end{aligned}\quad (71)$$

After simple manipulation by using Lemma 1, it is evident to obtain the relation (59). From this, one can conclude that the augmented closed-loop T-S fuzzy system ((55), (56), (57), and (58)) is robustly stable against $e(t)$, $e_f(x, u, t)$, and $\xi(x, t)$ with respect to the H_∞ performance level ζ_c .

This completes the proof.

7. Illustrative Example

In the present section, the design of the separated and integrated sliding mode fault-tolerant control based on adaptive observer information requirement is performed by considering the nonlinear model of a single-link flexible-joint robot arm taken from [13]. Firstly, let us consider the nonlinear model without faults defined by

$$\begin{aligned}\dot{\theta}_m &= \omega_m, \\ \dot{\omega}_m &= \frac{k}{J_m} (\theta_1 - \theta_m) - \frac{B_v}{J_m} \omega_m + \frac{K_\tau}{J_m} u(t), \\ \dot{\theta}_1 &= \omega_1, \\ \dot{\omega}_1 &= \frac{k}{J_1} (\theta_1 - \theta_m) - \frac{mgh}{J_1} \sin(\theta_1),\end{aligned}\quad (72)$$

TABLE 1: Nonlinear system parameters.

System parameters	Values	Units
Motor inertia J_m	3.7×10^{-3}	kg m ²
Link inertia J_1	9.3×10^{-3}	kg m ²
Pointer mass m	0.21	kg
Link length h	0.15	m
Torsional spring k	0.18	Nm/rad
Viscous friction B_v	4.6×10^{-3}	M
Amplifier gain K_τ	8×10^{-2}	Nm/V

where θ_m and ω_m are the position and angular velocity of the DC motor, respectively, and θ_1 and ω_1 represent the position and angular velocity of the link. The values of the parameters are given in Table 1.

We choose that $x_1 = \theta_m$, $x_2 = \omega_m$, $x_3 = \theta_1$, and $x_4 = \omega_1$.

The flexible-joint robot arm system is described in the nonlinear form as follows:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + \Gamma(x, t) + Mf(x, u, t) + D\xi(x, t), \\ y(t) &= Cx(t),\end{aligned}\quad (73)$$

with

$$\begin{aligned}A &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ -48.6 & -1.25 & 48.6 & 0 \\ 0 & 0 & 0 & 1 \\ 19.5 & 0 & -19.5 & 0 \end{bmatrix}, \\ B = M &= \begin{bmatrix} 0 \\ 21.62 \\ 0 \\ 0 \end{bmatrix}, \\ D &= \begin{bmatrix} 0 \\ 0.25 \\ 0 \\ 0 \end{bmatrix}, \\ C &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\ \Gamma(x, u, t) &= \begin{bmatrix} 0 \\ 21.62u(t) \\ 0 \\ 3.33\sin(x_3(t)) \end{bmatrix},\end{aligned}\quad (74)$$

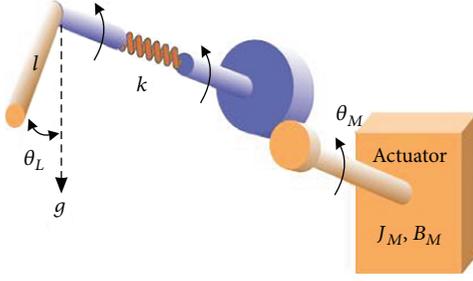


FIGURE 3: Schematic diagram of a single-link flexible-joint robot arm.

TABLE 2: LMI optimization gains.

	Integrated FTC		Separated FTC	
		Observer	Controller	
Uncertainty attenuation level	$\zeta_c = 0.1588$	$\zeta = 0.0612$	$\zeta_s = 0.7337$	
Admissible Lipschitz constant	$\gamma_1 = \gamma_2 = 0.7916$	$\gamma_{e,1} = \gamma_{e,2} = 0.3911$		

where $\Gamma(x, u, t)$ encapsulates the nonlinearities present in the DC motor. The schematic diagram of a single-link flexible-joint robot arm is shown in Figure 3.

To evaluate the performances of the proposed FE-based adaptive sliding mode fault-tolerant control, we consider the presence of two types of multiplicative faults affecting the considered nonlinear system, which are described in the following.

(1) Actuator gain fault (loss of effectiveness): multiplicative fault occurs in the actuator which is defined as a partial loss of effectiveness. We suppose that fault magnitude $\theta_1(t)$ has the following structure:

$$\theta_1(t) = \begin{cases} 0, & t < 5 \text{ sec}, \\ 0.8, & 5 \text{ sec} \leq t < 20 \text{ sec}, \\ 0, & t > 15 \text{ sec}. \end{cases} \quad (75)$$

The fault function $h_1(x, u, t)$ is expressed as

$$h_1(x, u, t) = [0 \quad u(t) \quad 0 \quad 0]^T. \quad (76)$$

In this case, the first multiplicative fault is modeled as

$$f_1(x, u, t) = [0 \quad 3\theta_1(t)u(t) \quad 0 \quad 0]^T. \quad (77)$$

(2) Abnormal friction subject to process fault: an abnormal friction appears in the DC motor where it leads to parameter changes in the nonlinear system state matrix. This multiplicative process fault has the following special structure:

$$f_2(x, u, t) = [0 \quad -5\theta_2(t)\omega_m(t) \quad 0 \quad 0]^T, \quad (78)$$

which corresponds to the structure function handled as

$$h_2(x, u, t) = [0 \quad -5\omega_m(t) \quad 0 \quad 0]^T. \quad (79)$$

We suppose that the viscous friction constant B_v increases by 80% at $t = 2.5$ sec, that is, $\theta_2(t) = 0$ at $t < 2.5$ sec and $\theta_2(t) = 0.8$ at $t \geq 2.5$ sec.

The flexible-joint robot arm system can be formulated in the T-S representation (5), where $k = 2$, with the system matrices:

$$\begin{aligned} \mathbf{A}_1 &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ -48.6 & -1.25 & 48.6 & 0 \\ 0 & 0 & 0 & 1 \\ 19.5 & 0 & -22.83 & 0 \end{bmatrix}, \\ \mathbf{B}_1 = \mathbf{M}_1 &= \begin{bmatrix} 0 \\ 21.62 \\ 0 \\ 0 \end{bmatrix}, \\ \mathbf{A}_2 &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ -48.6 & -1.24 & 48.6 & 0 \\ 0 & 0 & 0 & 1 \\ 19.5 & 0 & -18.77 & 0 \end{bmatrix}, \\ \mathbf{B}_2 = \mathbf{M}_2 &= \begin{bmatrix} 0 \\ 21.62 \\ 0 \\ 0 \end{bmatrix}, \\ \mathbf{D}_1 = \mathbf{D}_2 &= \begin{bmatrix} 0 \\ 0.25 \\ 0 \\ 0 \end{bmatrix}, \\ \mathbf{C}_1 = \mathbf{C}_2 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \end{aligned} \quad (80)$$

The parameters $\mu_i(x(t))$ are given by

$$\begin{aligned} \mu_1(x(t)) &= \frac{\vartheta(t) + 0.21}{1.21}, \\ \mu_2(x(t)) &= \frac{1 - \vartheta(t)}{1.21}, \end{aligned} \quad (81)$$

where $\vartheta(t) = \sin(x_3(t))/x_3(t)$.

Comparative simulations are given using the separated and integrated multiplicative FE-based FTC design with the same system parameters and initial conditions.

7.1. Separated Multiplicative FE-Based FTC Design

7.1.1. First Step: Adaptive Observer Design. The design parameters were chosen as $\lambda = 0.9$ and $H = I_{4 \times 4}$. By solving Theorem 1 with the MATLAB LMI Toolbox, the adaptive observer ((7), (8), and (9)) design is achieved as

$$\begin{aligned}
G_{l,1} &= 10^4 \times \begin{bmatrix} 0.0001 & -0.0021 & -0.0088 \\ 4.0422 & 0.1157 & 3.8222 \\ 0.0117 & -0.0027 & 0.0001 \\ 0.0077 & -0.0013 & -0.0021 \end{bmatrix}, \\
G_{l,2} &= 10^4 \times \begin{bmatrix} 0.0006 & 0.0023 & -0.0050 \\ -2.5615 & 1.1566 & -0.6657 \\ 0.0066 & 0.0005 & 0.0015 \\ 0.0228 & 0.0003 & -0.0170 \end{bmatrix},
\end{aligned} \tag{82}$$

where we find that

$$P_e = \begin{bmatrix} 4.1673 & 0 & 0 & 0 \\ 0 & 0.0022 & 0 & 0 \\ 0 & 0 & 3.7929 & -2.0669 \\ 0 & 0 & -2.0669 & 4.1673 \end{bmatrix}. \tag{83}$$

7.1.2. Second Step: Sliding Mode Controller Design. From Theorem 2, the sliding mode controller gains (14) are described as

$$K_1 = K_2 = [101.5736 \quad 11.6305 \quad -34.4968 \quad 11.4298], \tag{84}$$

such that

$$P_x = \begin{bmatrix} 0.2650 & -0.9467 & 0.2119 & -0.8697 \\ -0.9467 & 12.3589 & 0.0313 & 0.0634 \\ 0.2119 & 0.0313 & 0.4505 & -0.5907 \\ -0.8697 & 0.0634 & -0.5907 & 6.2204 \end{bmatrix}. \tag{85}$$

7.2. Integrated Multiplicative FE-Based FTC Design. By solving the LMI conditions given in Theorem 3, using the “mincx” function of the MATLAB LMI toolbox, the matrix gains of the adaptive observer ((7), (8), and (9)) and the sliding mode controller (14) are computed in a single step as

$$\begin{aligned}
G_{l,1} &= \begin{bmatrix} 0.5001 & 0.4705 & -4.4445 \\ 386.9460 & 475.4862 & 24.2993 \\ 4.4446 & 0.0295 & 0.9559 \\ 21.2220 & 0.0114 & -21 : 4797 \end{bmatrix}, \\
G_{l,2} &= \begin{bmatrix} 0.5001 & 0.4705 & -4.4445 \\ 386.9496 & 475.4862 & 24.3004 \\ 4.4446 & 0.0295 & 0.9559 \\ 21.2220 & 0.0114 & -17.4097 \end{bmatrix}, \\
K_1 = K_2 &= [102.6561 \quad 11.4223 \quad -35.6482 \quad 12.0718],
\end{aligned} \tag{86}$$

where we find that

$$\begin{aligned}
P_e &= 10^4 \times \begin{bmatrix} 0.9035 & 0 & 0 & 0 \\ 0 & 0.0011 & 0 & 0 \\ 0 & 0 & 1.0631 & -0.4119 \\ 0 & 0 & -0.4119 & 1.0631 \end{bmatrix}, \\
P_x &= \begin{bmatrix} 0.2689 & -0.9436 & 0.2134 & -0.8727 \\ -0.9436 & 26.4751 & 0.0344 & 0.0597 \\ 0.2134 & 0.0344 & 0.4366 & -0.5899 \\ -0.8727 & 0.0597 & -0.5899 & 5.9295 \end{bmatrix}.
\end{aligned} \tag{87}$$

Additionally, the LMI optimization gains of the integrated and separated FE-based FTC designs are listed in Table 2. As can be seen, the uncertainty attenuation level, which refers to the integrated approach, is much less than that of the separated approach. This latter loses a certain robustness degree against uncertainties illustrating the better multiplicative fault estimation and compensation using the integrated FE-based FTC approach.

We learned that the fault nonlinear function satisfies the Lipschitz condition. In this way, the admissible Lipschitz constant, which refers to the integrated approach, is greater than the one given by the separated approach, thus illustrating the superiority of the integrated FE-based FTC approach to treat a large range of fault nonlinear function.

It is worth pointing that simulation results are given with online multiplicative fault estimation and compensation for the closed-loop nonlinear system in the presence of uncertainties as $\xi(x, t) = 0.1 \times 3\sin(0.3t)$ such that

- (i) the sliding mode controller (14) is considered where $F_{a,1} = F_{a,2} = 1$, $N_1 = N_2 = [-0.7500 \quad 0.0463 \quad -0.7500]$, and $\eta_c(t) = \hat{p}_c + 0.1$ such that the adaptive term is given by $\dot{\hat{p}}_{sc} = 2 \times \|S_c(t)\|$,
- (ii) the initial conditions are $x_{10} = \Pi/15$, $x_{20} = 0.2$, $x_{30} = \Pi/12$, $x_{40} = 0$, $\hat{\theta}_1(t=0) = \hat{\theta}_2(t=0) = 0$, and $\hat{p}_c(t=0) = 0$,
- (iii) $\sigma_1 = \sigma_2 = 10$ and $\Gamma_1 = \Gamma_2 = 0.75 \times [1 \quad 1 \quad 1]$.

From Figure 4, it is quite clear to see that the state estimation errors remain zero in finite time; that is, the T-S adaptive observer proposed in this paper can robustly estimate nonlinear system states with acceptable performances. Figure 5 illustrates the first multiplicative fault estimation error (partial loss of actuator effectiveness).

The results display that the proposed adaptive observer ((7), (8), and (9)), using the integrated FE-based FTC design, is capable of estimating multiplicative actuator fault with better performances compared to the separated FE-based FTC design. Note that abrupt changes of fault can generate small peak in Figure 5 at time 5 sec. Meanwhile, an abnormal

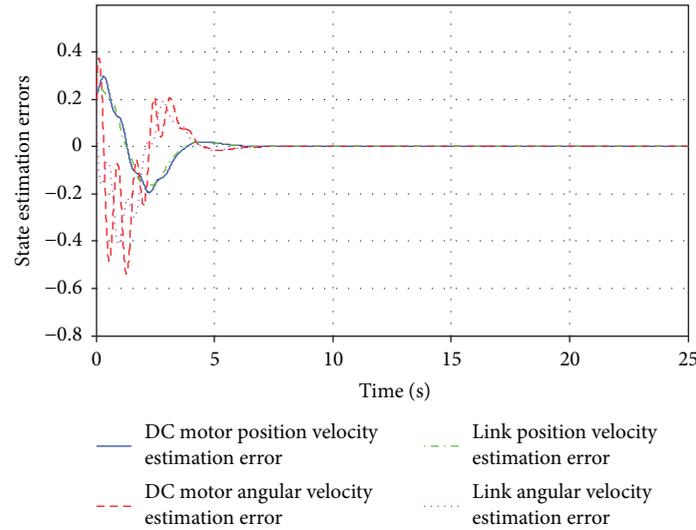


FIGURE 4: Nonlinear system state estimation errors.

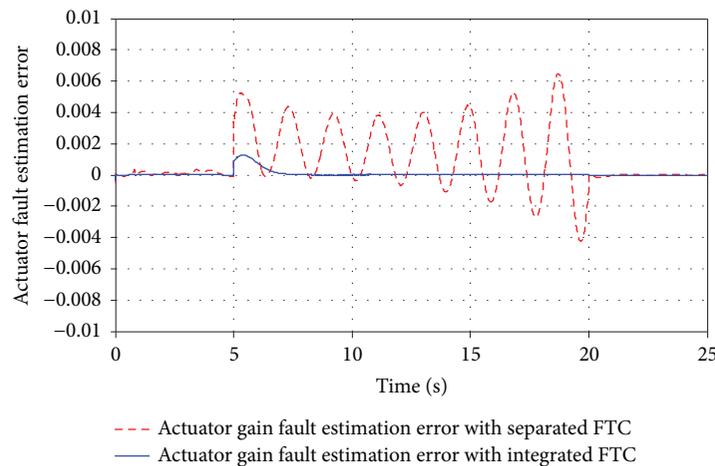


FIGURE 5: Actuator gain fault estimation error.

friction in the DC motor has been introduced, at $t = 2.5$ sec, to show the capability of the proposed FE strategy to handle process faults.

Consider Figure 6, which plots process fault estimation error generated using the separated FE-based FTC scheme against this variable using the integrated FE-based FTC scheme. It should be highlighted that the proposed adaptive observer design, using a single-step LMI formulation, can still track the considered process fault with better performances compared to the separated FE-based FTC design in terms of precision and robustness against the uncertainties.

Figures 7–9 outline a comparison between the nonlinear single-link flexible-joint robot output responses referring to two different cases: output responses with the separated FE-based FTC and output responses with the integrated FE-based FTC.

Based on the figures shown above, the conceived adaptive sliding mode controller (14) can stabilize with satisfactory

performances the closed-loop nonlinear single-link flexible-joint robot based on the integrated and separated FE-based sliding mode FTC. More precisely, as can be seen in zoomed versions from Figures 7–9, the proposed integrated FE-based FTC is capable of compensating real multiplicative faults with better performances compared to the separated FE-based FTC in terms of precision and robustness against uncertainties.

8. Conclusion

This paper has proposed two adaptive sliding mode FTC schemes for an uncertain nonlinear system subject to multiplicative and process faults. In the first scheme, the separated FE-based FTC is constructed to compensate real fault effects based on output feedback information and to ensure robust stability of the closed-loop system. In the second scheme, the integrated FE-based sliding mode FTC is conceived in

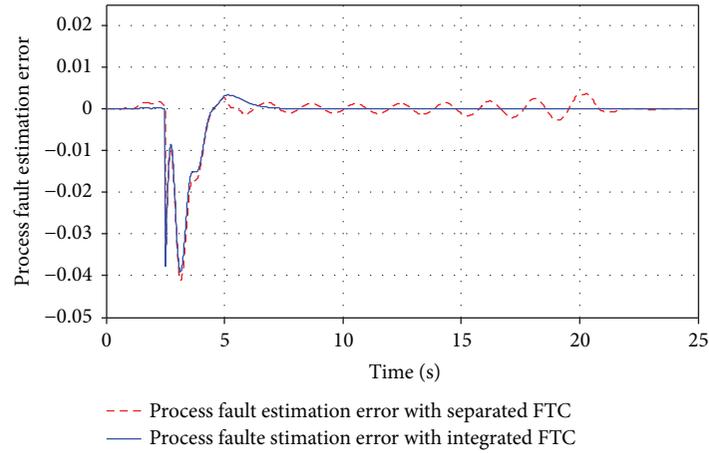


FIGURE 6: Process fault estimation error.

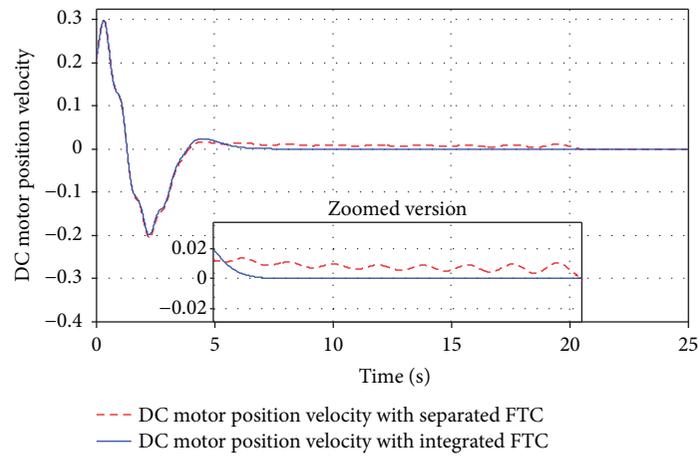


FIGURE 7: First closed-loop system output response: output response with the separated FE-based FTC (red line) and output response with the integrated FE-based FTC (blue line).

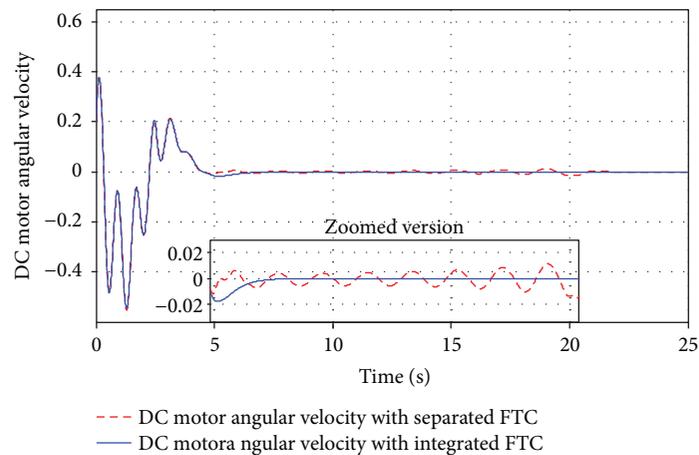


FIGURE 8: Second closed-loop system output response: output response with the separated FE-based FTC (red line) and output response with the integrated FE-based FTC (blue line).

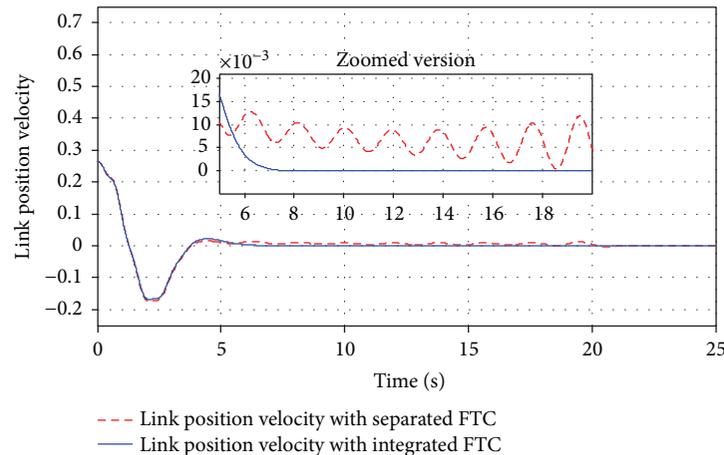


FIGURE 9: Third closed-loop system output response: output response with the separated FE-based FTC (red line) and output response with the integrated FE-based FTC (blue line).

order to achieve an optimal robustness interaction between observer and controller models. The novelty of the proposed approach is that the observer and controller gains are obtained by solving a single-step multiobjective LMI optimization problem in order to offer a solution to stabilize the closed-loop nonlinear system despite the occurrence of real fault effects. As shown in the comparative simulations of the flexible-joint robot arm system described by the T-S fuzzy model, the proposed integrated FE-based FTC improves the best robustness interactions between multiplicative fault estimation and sliding mode control. Summing up the results, we can conclude that the research into the integrated FE-based sliding mode FTC for an uncertain nonlinear system has been very successful.

Data Availability

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

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

- [1] L. Wang, M. Cai, H. Zhang, F. Alsaadi, and L. Chen, "Active fault-tolerant control for wind turbine with simultaneous actuator and sensor faults," *Complexity*, vol. 2017, Article ID 6164841, 11 pages, 2017.
- [2] M. Van, P. Franciosa, and D. Ceglarek, "Fault diagnosis and fault-tolerant control of uncertain robot manipulators using high-order sliding mode," *Mathematical Problems in Engineering*, vol. 2016, Article ID 7926280, 14 pages, 2016.
- [3] R. Isermann, *Fault-Diagnosis Applications: Model-Based Condition Monitoring: Actuators, Drives, Machinery, Plants, Sensors, and Fault-Tolerant Systems*, Springer, Science and Business Media, 2011.
- [4] H. Alwi, C. Edwards, and C. P. Tan, *Fault Detection and Fault-Tolerant Control Using Sliding Modes*, Springer, Science and Business Media, 2011.
- [5] J. Wang, "H ∞ fault-tolerant controller design for networked control systems with time-varying actuator faults," *International Journal of Innovative Computing Information and Control*, vol. 11, no. 4, pp. 1471–1481, 2015.
- [6] F. Li, H. Zhenggao, and G. Zhao, "Fault estimation and adaptive fault tolerant control for dynamic systems based on the second-order sliding mode observer," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 230, no. 3, pp. 222–230, 2016.
- [7] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 1, pp. 116–132, 1985.
- [8] S. Aouaouda, M. Chadli, V. Cocquempot, and M. Tarek Khadir, "Multi-objective H $_2$ /H ∞ fault detection observer design for Takagi-Sugeno fuzzy systems with unmeasurable premise variables: descriptor approach," *International Journal of Adaptive Control and Signal Processing*, vol. 27, no. 12, pp. 1031–1047, 2013.
- [9] H. Wang, D. Ye, and G. H. Yang, "Actuator fault diagnosis for uncertain T-S fuzzy systems with local nonlinear models," *Nonlinear Dynamics*, vol. 76, no. 4, pp. 1977–1988, 2014.
- [10] D. Ichalal, B. Marx, J. Ragot, and D. Maquin, "Fault detection, isolation and estimation for Takagi-Sugeno nonlinear systems," *Journal of the Franklin Institute*, vol. 351, no. 7, pp. 3651–3676, 2014.
- [11] G. H. Yang and H. Wang, "Fault detection and isolation for a class of uncertain state-feedback fuzzy control systems," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 1, pp. 139–151, 2015.
- [12] A. B. Brahim, S. Dhahri, F. B. Hmida, and A. Sellami, "An H ∞ sliding mode observer for Takagi-Sugeno nonlinear systems with simultaneous actuator and sensor faults An," *International Journal of Applied Mathematics and Computer Science*, vol. 25, no. 3, pp. 547–559, 2015.
- [13] A. Ben Brahim, S. Dhahri, F. Ben Hmida, and A. Sellami, "Simultaneous actuator and sensor faults reconstruction based on robust sliding mode observer for a class of nonlinear systems," *Asian Journal of Control*, vol. 19, no. 1, pp. 362–371, 2017.

- [14] R. Raoufi, H. J. Marquez, and A. S. I. Zinober, " \mathcal{H}_∞ sliding mode observers for uncertain nonlinear Lipschitz systems with fault estimation synthesis," *International Journal of Robust and Nonlinear Control*, vol. 20, pp. 1785–1801, 2010.
- [15] S. Dhahri, A. Sellami, and F. B. Hmida, "Robust H_∞ sliding mode observer design for fault estimation in a class of uncertain nonlinear systems with LMI optimization approach," *International Journal of Control, Automation and Systems*, vol. 10, no. 5, pp. 1032–1041, 2012.
- [16] J. Zhang, S. K. Nguang, and A. K. Swain, "Detection and isolation of incipient sensor faults for a class of uncertain nonlinear systems," *IET Control Theory and Applications*, vol. 6, no. 12, pp. 1870–1880, 2012.
- [17] A. Valibeygi, A. Toudeshki, and K. Vijayaraghavan, "Observer-based sensor fault estimation in nonlinear systems," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 230, no. 8, pp. 759–777, 2016.
- [18] J. Lan and R. J. Patton, "Integrated fault estimation and fault-tolerant control for uncertain Lipschitz nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 27, no. 5, pp. 761–780, 2017.
- [19] C. Lu, Y. Cheng, H. Liu, and Z. Wang, "An approach to fault detection and isolation for control components in the aircraft environmental control system," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 228, no. 7, pp. 1202–1214, 2014.
- [20] S. Yin, H. Luo, and S. X. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2402–2411, 2014.
- [21] S. X. Ding, P. M. Frank, and E. L. Ding, "An approach to the detection of multiplicative faults in uncertain dynamic systems," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002*, pp. 4371–4376, Las Vegas, NV, USA, 2002.
- [22] C. P. Tan and C. Edwards, "Multiplicative fault reconstruction using sliding mode observers," in *2004 5th Asian Control Conference (IEEE Cat. No.04EX904)*, pp. 957–962, Melbourne, VIC, Australia, 2004.
- [23] K. Zhang, B. Jiang, and P. Shi, "Fast fault estimation and accommodation for dynamical systems," *IET Control Theory & Applications*, vol. 3, no. 2, pp. 189–199, 2009.
- [24] J. Lan and R. J. Patton, "A new strategy for integration of fault estimation within fault-tolerant control," *Automatica*, vol. 69, pp. 48–59, 2016.
- [25] X. Zhang, M. M. Polycarpou, and T. Parisini, "Fault diagnosis of a class of nonlinear uncertain systems with Lipschitz nonlinearities using adaptive estimation," *Automatica*, vol. 46, no. 2, pp. 290–299, 2010.
- [26] C. Gao and G. Duan, "Robust adaptive fault estimation for a class of nonlinear systems subject to multiplicative faults," *Circuits, Systems, and Signal Processing*, vol. 31, no. 6, pp. 2035–2046, 2012.

Research Article

Rigorous Solution of Slopes' Stability considering Hydrostatic Pressure

Chengchao Li , Pengming Jiang, and Aizhao Zhou

Department of Civil and Architecture Engineering, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China

Correspondence should be addressed to Chengchao Li; justlcc2015@163.com

Received 10 December 2017; Revised 23 April 2018; Accepted 8 May 2018; Published 12 June 2018

Academic Editor: Rafał Burdzik

Copyright © 2018 Chengchao Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

According to characteristics of soils in failure, a sliding mechanism of slopes in limit state is divided into five parts, for building a slip line field satisfying all possible boundary conditions. An algorithm is built to obtain the rigorous solution approaching upper and lower bound values simultaneously, which satisfies the static boundary and the kinematical boundary based on the slip line field, while stress discontinuity line and velocity discontinuity line are key points. This algorithm is compared with the Spencer method to prove its feasibility with a special example. The variation of rigorous solution, including an ultimate load and a sliding belt the rigid body sliding along rather than a single slip surface for friction-type soils, is achieved considering hydrostatic pressure with soil parameters changing.

1. Introduction

The stability of slopes has been regarded as a classic and difficult problem for engineers because of less boundary constraints, compared with the earth pressure of retaining wall and the bearing capacity of foundation. Over the past few years, many investigators have evaluated slope stability, thereby developing many methods for meeting engineering requirements, such as the limit equilibrium method (LEM) [1, 2], the finite element method (FEM) [3, 4], and the limit analysis method (LAM) [5–7]. The LEM captures the static equilibrium of rigid blocks on a particular slip surface, while not considering the plastic deformation of soils. The equilibrium equation accounts for the whole slice but not guarantees each point in soils. The strength reduction method (SRM) [8–10] is the main finite element slope stability method currently employed, by which stress field and displacement of soils in slopes can be calculated with an elastic-plastic constitutive model to get the safety factor. Although the displacement mutation, the numerical calculation of nonconvergence, and other criteria can determine the

slope instability, the slope displacement calculated by the SRM is not the actual displacement of soils and the safety factor required is an approximation. For the slope stability, it is sometimes not necessary to get the variation of stress field and strain field, only to get the ultimate load. Based on the extremum principal [11], the lower bound (LB) [12, 13] solution can be got by static analysis for limit equilibrium problems and the upper bound (UB) [14–17] solution can be got by dynamic analysis. If the lower bound solution satisfies all the kinematic conditions or the upper bound solution satisfies all the static conditions, the solution will be the rigorous one. Compared with obtaining the safety factor, the solution calculated by the upper and lower bound theorems is closer to the real condition because the ultimate load approaches the upper bound solution and the lower bound solution at the same time, satisfying all possible boundary conditions.

When a slope is on the verge of collapse, for $c - \varphi$ type, a sliding belt is emerged within the slope due to the friction between soils. And the sliding belt is not a single slip surface, but a thin shear zone. This paper builds the slip line field satisfying the static boundary condition and the velocity

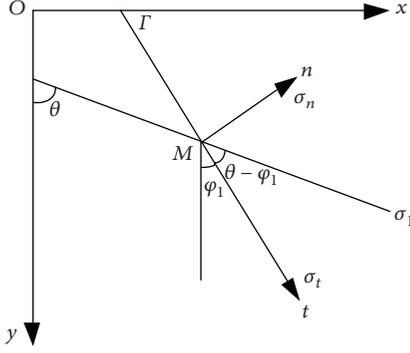


FIGURE 1: Basic coordinate system and noncharacteristic line stress boundary.

boundary condition according to the characteristics of the stress discontinuity line and the velocity discontinuity line and compiles an algorithm to gain the distribution of the sliding belt and the ultimate capacity of the slope considering the hydrostatic pressure. It also indicates the variation of the ultimate bearing capacity and the sliding belt with different parameters. The application of this algorithm is proved by comparing with the Spencer's method.

2. Slip Line Field

To analyze the slope stability based on the LAM, the slip line field is constructed according to the stress boundary conditions; one of which is a noncharacteristic line stress boundary with normal stress and shear normal stress, and the other one is the interface of a rigid region and a plastic region or a stress discontinuity surface [18]. Under the limit state, only the noncharacteristic line boundary can be used to construct the slip line field, while the other is unknown. Basic coordinate and noncharacteristic line stress boundary is shown in Figure 1. θ is the angle between the direction of the maximum principal stress σ_1 , and the y -axis and Γ is the noncharacteristic line stress boundary. φ_1 ($-\pi < \varphi_1 \leq \pi$) is the angle between the tangential direction t of the boundary Γ and the y -axis, so the angle between t and σ_1 is $\theta - \varphi_1$. According to σ_n and τ_n , two Mohr's circles tangent to the Coulomb failure line can be drawn (Figure 2). When soils are in the extreme state, σ_t , σ_n , and τ_n on the boundary Γ are expressed, respectively, as follows:

$$\sigma_t = p \pm R \cos 2(\theta - \varphi_1), \quad (1.1)$$

$$\sigma_n = p \mp R \cos 2(\theta - \varphi_1), \quad (1.2)$$

$$\tau_n = R \sin 2(\theta - \varphi_1), \quad (1.3)$$

$$R = (p + \sigma_c) \sin \varphi, \quad (1.4)$$

where σ_n , τ_n , and σ_t are normal stress, shear stress, and tangent normal stress on the boundary Γ , respectively. p is average stress; R is radius of Mohr's circle; c is soil cohesion; and φ is internal friction angle. σ_c is cohesive internal stress, which is given by: $\sigma_c = c \cot \varphi$.

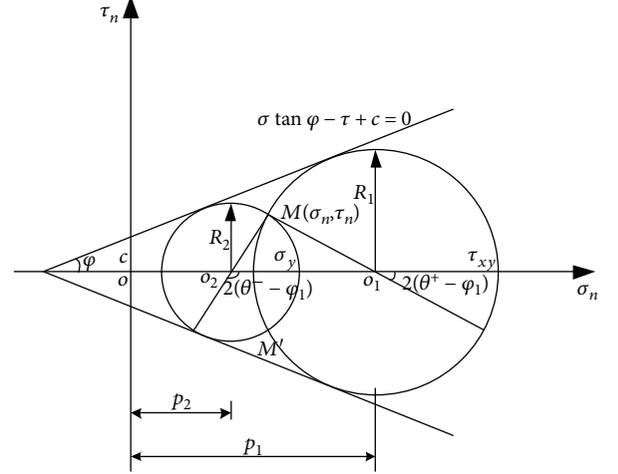


FIGURE 2: Mohr's circles on noncharacteristic line stress boundary.

Because M is a point on the Mohr's circle, R can be rewritten as follows:

$$R = \sqrt{(\sigma_n - p)^2 + \tau_n^2}. \quad (2)$$

According to (1.4) and (2):

$$p^2 \cos^2 \varphi - 2(\sigma_n + c \sin \varphi \cos \varphi)p + \sigma_n^2 + \tau_n^2 - c^2 \cos^2 \varphi = 0. \quad (3.1)$$

Solving the unary quadratic equation of p :

$$p_{1,2} = \sec^2 \varphi \left[\sigma_n + c \sin \varphi \cos \varphi \pm \sqrt{(\sigma_n \sin \varphi - c \cos \varphi)^2 - \tau_n^2 \cos^2 \varphi} \right]. \quad (3.2)$$

When $\sigma_t > \sigma_n$, according to (1.3):

$$\theta^+ = \varphi_1 + \frac{1}{2} \arcsin \frac{\tau_n}{R} = \varphi_1 + \frac{1}{2} \arcsin \frac{\tau_n}{(p_1 + \sigma_c) \sin \varphi}. \quad (4)$$

When $\sigma_t < \sigma_n$, (1.3) should be rewritten as follows:

$$\begin{aligned} \tau_n &= R \sin [\pi - 2(\theta - \varphi_1)] = R \sin 2 \left[\frac{\pi}{2} - (\theta - \varphi_1) \right], \\ \theta^- &= \varphi_1 - \frac{1}{2} \arcsin \frac{\tau_n}{R} + \frac{\pi}{2} \\ &= \varphi_1 - \frac{1}{2} \arcsin \frac{\tau_n}{(p_2 + \sigma_c) \sin \varphi} + \frac{\pi}{2}, \end{aligned} \quad (5)$$

where θ^+ and θ^- are the maximum principal direction angles.

According to the above analysis, the analytic expressions of noncharacteristic line stress boundary conditions can be uniformly written as follows:

$$p_{1,2} = \sec^2 \varphi \left[\sigma_n + c \sin \varphi \cos \varphi \pm \sqrt{(\sigma_n \sin \varphi + c \cos \varphi)^2 - \tau_n^2 \cos^2 \varphi} \right],$$

$$\theta^{+,-} = \varphi_1 \pm \frac{1}{2} \arcsin \frac{\tau_n}{(p_{1,2} + \sigma_c) \sin \varphi} \mp \frac{\pi}{4} + \frac{\pi}{4},$$

$$\sigma_{t1,2} = \sigma_n \pm 2 \sqrt{(p_{1,2} + \sigma_c)^2 \sin^2 \varphi - \tau_n^2}. \quad (6)$$

With θ on the noncharacteristic line stress boundary, (6) can be used to draw the stress line field near the stress boundary:

$$\alpha = \theta - \mu \quad (\text{along line } \alpha),$$

$$\beta = \theta + \mu \quad (\text{along line } \beta), \quad (7)$$

where α and β are the angles between stress characteristic lines through the point M and the y -axis; μ is the angle between the stress characteristic line and the direction of major principal stress and given as follows:

$$\mu = \frac{\pi}{4} - \frac{\varphi}{2}. \quad (8)$$

In this case, stress boundary conditions are achieved. In order to facilitate the following calculations, new variables σ_e and θ are used instead of σ_x , σ_y , and τ_{xy} (Figure 3). σ_e is the effective stress, and θ is the direction angle of the maximum principal stress, written as the (9). In addition, σ_e and θ of each point in slip line field can be calculated by solving the basic boundary value problems. In the theory of hyperbolic partial differential equations, there are three basic boundary value problems [19, 20], of which Cauchy problem is the key issue the others can be solved by. Here, only the Cauchy problem is introduced and the rest of the problems (Riemann problem and mixed boundary value problem) can be referred to the literature [19]. Taking into account of the hydrostatic pressure, the soils above the ground water take unit natural weight γ and the soils below ground water take unit floating weight γ' .

$$\sigma_e = \frac{1}{2} (\sigma_x + \sigma_y) + \sigma_c,$$

$$\left. \begin{array}{l} \sigma_x \\ \sigma_y \end{array} \right\} = \sigma_e (1 \pm \sin \varphi \cos 2\theta) - \sigma_c, \quad (9)$$

$$\tau_{xy} = \sigma_e \sin \varphi \sin 2\theta.$$

As shown in Figure 4(a), OA is a smooth and continuous stress boundary line, a nonslip line, in which the coordinates (x, y) of each point, effective stress σ_e , and the direction angle θ are known. By solving the Cauchy problem, the distribution of slip line field in the triangular OAB surrounded by the lines α and β through the points of O and A , respectively, could be obtained.

Such as x , y , σ_e , and θ on the point 22 can be obtained by the points of 21 and 32, along line α :

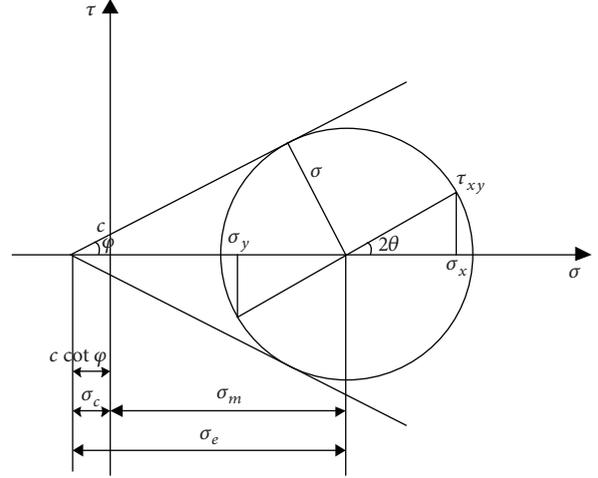


FIGURE 3: Effective stress of Mohr's circle.

$$\begin{aligned} & (\sigma_{e22} - \sigma_{e21}) - (\sigma_{e22} + \sigma_{e21} + 2\sigma_c) \tan \varphi (\theta_{22} - \theta_{21}) \\ &= \frac{-\gamma^* \sin [(\theta_{22} + \theta_{21})/2 + \mu] (y_{22} - y_{21})}{\cos \varphi \cos [(\theta_{22} + \theta_{21})/2 - \mu] (y_{22} - y_{21})} \quad (10.1) \\ &= (x_{22} - x_{21}) \cot \left[\frac{\theta_{22} + \theta_{21}}{2 - \mu} \right], \end{aligned}$$

along line β :

$$\begin{aligned} & (\sigma_{e22} - \sigma_{e32}) - (\sigma_{e22} + \sigma_{e32} + 2\sigma_c) \tan \varphi (\theta_{22} - \theta_{32}) \\ &= \frac{-\gamma^* \sin [(\theta_{22} + \theta_{32})/2 + \mu] (y_{22} - y_{32})}{\cos \varphi \cos [(\theta_{22} + \theta_{32})/2 + \mu] (y_{22} - y_{32})} \quad (10.2) \\ &= (x_{22} - x_{32}) \cot \left[\frac{\theta_{22} + \theta_{32}}{2 + \mu} \right], \end{aligned}$$

when soils are above ground water, $\gamma^* = \gamma$. When soils are below ground water, $\gamma^* = \gamma'$.

If the ground water table is between two points, it should be simplified to calculate (Figure 4(b)). Setting an allowable value Δh , if $\Delta h_1 \leq \Delta h$, the γ^* of the points 21 and 22 all take γ' . If $\Delta h_2 \leq \Delta h$, the γ^* of the points 21 and 22 all take γ . If the $\Delta h_1 > \Delta h$ and $\Delta h_2 > \Delta h$, γ^* of the points 21 and 22 take γ and γ' , respectively. In fact, the ground water table is not a straight line. The smaller the grid in the slip line field is, the more similar to a straight line the water line is.

The above two sets of nonlinear equations ((10.1) and (10.2)) need to be solved by an iterative method. It can be assumed that the initial values of point 22 is

$$\begin{aligned} x_{22} &= \frac{x_{21} + x_{32}}{2}, \\ y_{22} &= \frac{y_{21} + y_{32}}{2}, \\ \text{or } \theta_{22} &= \frac{\theta_{21} + \theta_{32}}{2}, \\ \sigma_{e22} &= \frac{\sigma_{e21} + \sigma_{e32}}{2}. \end{aligned} \quad (11)$$

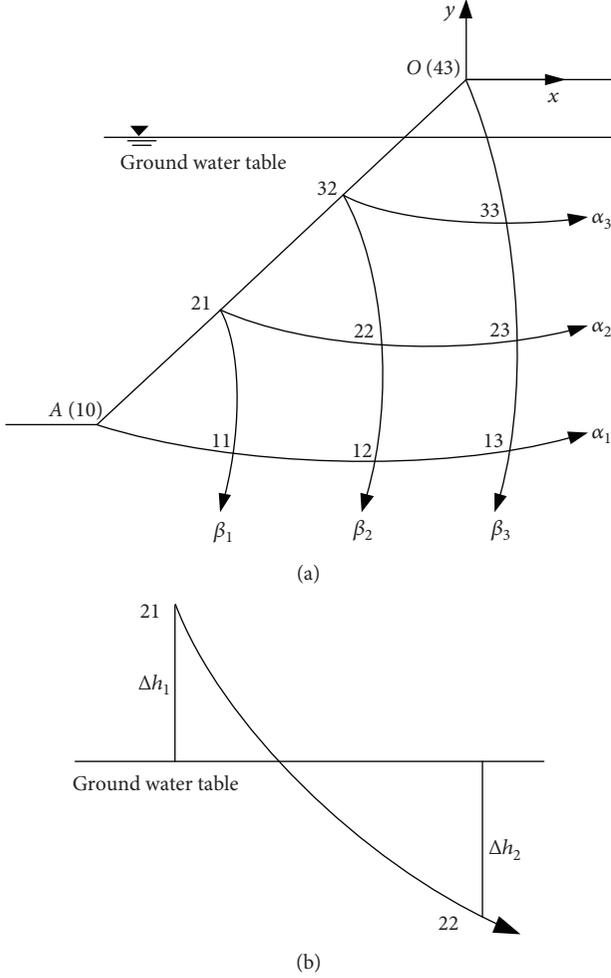


FIGURE 4: Cauchy boundary problem considering the impact of the ground water table.

Substituting x_{22} and y_{22} in (11) into (10.2), the approximations of θ_{22} and σ_{e22} can be obtained. Similarly, substituting θ_{22} and σ_{e22} in (11) into (10.1), the approximations of x_{22} and y_{22} can be obtained. Then, substituting the first approximations of θ_{22} and σ_{e22} into (10.1), the second approximations of x_{22} and y_{22} can be gained, so repeatedly until the accuracy of the calculation to meet the precision. And x , y , σ_e , and θ of each point can be achieved in the entire OAB . Finally, according to the (9), σ_x , σ_y , and τ_{xy} in the slip line field could be gained.

3. Stress Discontinuity Line and Velocity Discontinuity Line

When soils reach the limit state, the stress discontinuity field and the velocity discontinuity field will be generated under complex boundaries [21]. With different boundary conditions, the stress discontinuity line and the velocity discontinuity line will also change. Therefore, the distribution and the characteristic of the two discontinuity lines are significant for the slope stability.

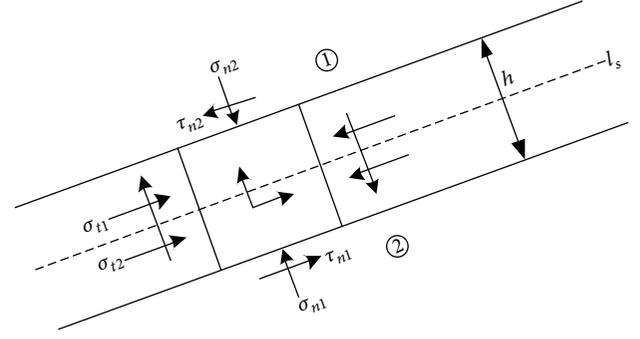


FIGURE 5: Stress discontinuity line.

3.1. Characteristics of Stress Discontinuity Line

3.1.1. Stress Equation. The stress discontinuity line is actually a thin transition zone, where the stress changes rapidly. It divides the element of the discontinuity line crossing through two plastic regions ① and ②, while only the tangential stress can be interrupted and the normal stress and the shear stress stay the same on both sides (Figure 4). In the plastic regions ① and ②, stresses on the discontinuity line should obey the Mohr-Coulomb yield criterion.

$$\sigma_{t1} \neq \sigma_{t2}, \sigma_{n1} = \sigma_{n2}, \tau_{n1} = \tau_{n2}, \quad (12)$$

$$\omega = \theta^+ + \delta^+ = \theta^- + \delta^-. \quad (13)$$

where θ^+ and θ^- are the maximum principal direction angles of the plastic regions ① and ②; ω is the tangential direction angle of the discontinuity line; and δ^+ and δ^- are the angles between the maximum principal stress direction and tangential direction of stress discontinuity line and are given as follows (Figure 5):

$$\delta^+ = \arctan \left[\left(-1 \pm \sqrt{1 + \frac{\cos^2 \varphi}{\tan^2(\theta^+ - \theta^-)}} \right) \frac{\tan(\theta^+ - \theta^-)}{1 + \sin \varphi} \right],$$

$$\delta^- = \arctan \left[\left(1 \pm \sqrt{1 + \frac{\cos^2 \varphi}{\tan^2(\theta^+ - \theta^-)}} \right) \frac{\tan(\theta^+ - \theta^-)}{1 + \sin \varphi} \right]. \quad (14)$$

The relationship of equivalent stresses on both sides of the discontinuity line is

$$\sigma_e^+ = \eta \sigma_e^-, \quad (15)$$

where σ_e^+ and σ_e^- are equivalent stresses on both sides of the discontinuity line; η is given as follows:

$$\eta = \left(\sqrt{1 + \frac{\cos^2 \varphi}{\tan^2(\theta^+ - \theta^-)}} + \sin \varphi \right)^2 \frac{\sin^2(\theta^+ - \theta^-)}{\cos^2 \varphi}. \quad (16)$$

3.1.2. Geometric Condition. As shown in Figure 6, the angle $\angle BAD$ of the principal stress element is 2γ . The angle between the stress discontinuity line l_s and the principal stress plane in the zone ① is ζ , and the angle in the zone ②

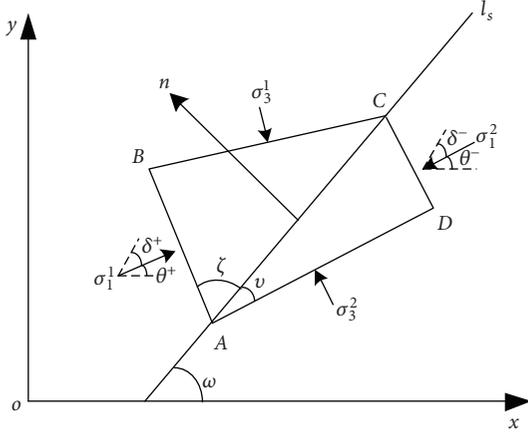


FIGURE 6: Stress elements on both sides of stress discontinuity line.

is v . From the geometric relationship shown in Figure 6, it can be obtained:

$$2\gamma = \zeta + v, \quad (17)$$

$$\theta^+ = \frac{\pi}{2} - \zeta + \omega, \quad (18.1)$$

$$\theta^- = v + \omega. \quad (18.2)$$

According to the geometric relationship shown in Figure 2, (12) can be rewritten as follows:

$$\begin{aligned} \sigma_{e1} [1 - \sin \varphi \cos 2(\theta^+ - \omega)] &= \sigma_{e2} [1 - \sin \varphi \cos 2(\theta^- - \omega)], \\ \sigma_{e1} \sin 2(\theta^+ - \omega) &= \sigma_{e2} \sin 2(\theta^- - \omega). \end{aligned} \quad (19)$$

The relationship of θ on both sides of a stress discontinuity line can be obtained eliminating σ_{e1} and σ_{e2} from (19).

$$\cos (2\omega - \theta^+ - \theta^-) + \sin \varphi \cos (\theta^- - \theta^+) = 0. \quad (20)$$

Substituting (18.1) and (18.2) into (20):

$$\sin (\zeta - v) - \sin \varphi \sin (\zeta + v) = 0. \quad (21)$$

The position of the stress discontinuity line can be determined according to (15) and (21) in the slip line field.

3.1.3. Kinematic Equation. Because the stress discontinuity line is emerged by the rigid region reducing to the limit state, the tangential velocity on the discontinuity line remains unchanged [21]:

$$dv_x \cos \omega + dv_y \sin \omega = 0, \quad (22.1)$$

where v_x and v_y are velocity components in the x and y directions.

Assuming that v_α and v_β are the velocity components in the direction of the characteristic line, a rigid kinematic equation of the stress discontinuity line will be obtained:

$$\begin{aligned} v_x &= [v_\alpha \sin (\theta + \mu) - v_\beta \sin (\theta - \mu)] \sec \varphi, \\ v_y &= [-v_\alpha \sin (\theta + \mu) + v_\beta \sin (\theta - \mu)] \sec \varphi. \end{aligned} \quad (22.2)$$

Substituting (22.2) into (22.1):

$$\begin{aligned} \sin (\theta + \mu - \omega) dv_\alpha - \sin (\theta - \mu - \omega) dv_\beta \\ = [-v_\alpha \cos (\theta + \mu - \omega) + v_\beta \cos (\theta - \mu - \omega)] d\theta. \end{aligned} \quad (23)$$

3.2. Characteristics of Velocity Discontinuity Line. Obeying the associated flow rule, the velocity discontinuity line is a slip line or an envelope of slip line. The boundary between a rigid region and a plastic region is a velocity discontinuity line. So the velocity discontinuity line is the slip line at both ends of the stress discontinuity line, dividing the slope into rigid zone and plastic zone. For the Mohr-Coulomb material, the angle between the velocity direction and the slip line is φ [22, 23].

4. Algorithm

For slope stability problems, the equilibrium equation and the plastic flow equation should be settled simultaneously to obtain the rigorous solution, which is difficult to achieve in mathematics and only depends on numerical methods. If a statically admissible stress field σ_{ij}^0 has been got, the strain rate field $\dot{\epsilon}_{ij}^*$ and the kinematically admissible velocity field v_i will be obtained by the stress field based on the associated flow rule. If the strain rate field and the velocity field are nothing less than the kinematical field, in this case, the plastic region corresponding to such a statically admissible stress field and a kinematically admissible velocity field must be the sliding belt in the extreme state and the external load must be the ultimate bearing capacity. Based on the upper and lower bound theorems, the numerical algorithm is established to get the distribution of the sliding belt and the ultimate bearing capacity considering the hydrostatic pressure.

When a slope is on the verge of collapse (shear failure), not all the points in sliding mass reach to the yield state, but the mixture of the plastic region and the rigid region is emerged. As shown in Figure 7, the entire slope is divided into five areas. The stress discontinuity line divides the slope foot into the strong plastic region and the weak plastic region, passive, and active, respectively. The boundary between two regions is the velocity discontinuity line, which is the slip line at both ends of the stress discontinuity line. So the stress discontinuity line and the velocity discontinuity line determine the distribution of the plastic region and the rigid region. The stress and deformation in each region are different as well as kinematic features. Only to meet all the possible kinematically admissible conditions, the solution will be exact.

The calculation process (shown in Figure 8) is consisted of two parts: a statically admissible field and a kinematically admissible field. The specific calculation process is as follows.

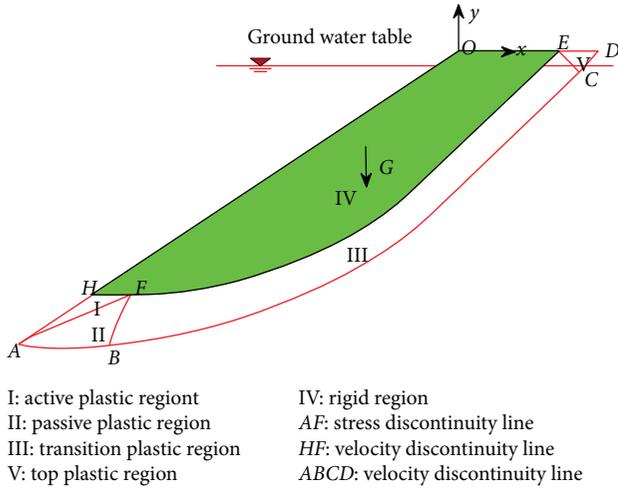


FIGURE 7: Calculation model.

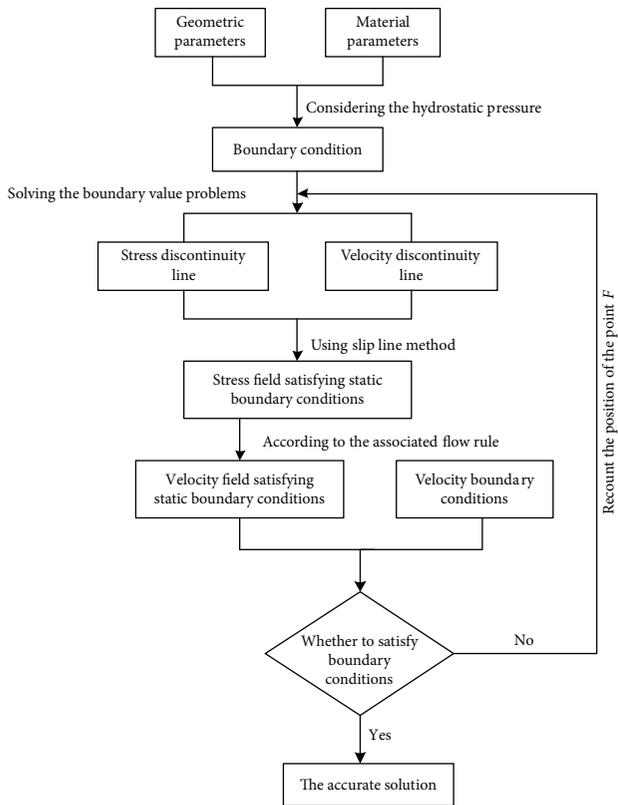


FIGURE 8: Flow chart of rigorous solution for slope stability.

4.1. Static Calculation

- (1) The maximum principal stress on the noncharacteristic line stress boundary OA is 0, and its direction angle is θ_1 , which is perpendicular to the slope surface. From the boundary OA , the slip line of the active region I is calculated in a very dense grid by solving the Cauchy problem.
- (2) To get the stress discontinuity line in the active region, the specific approach is to firstly select the point F in

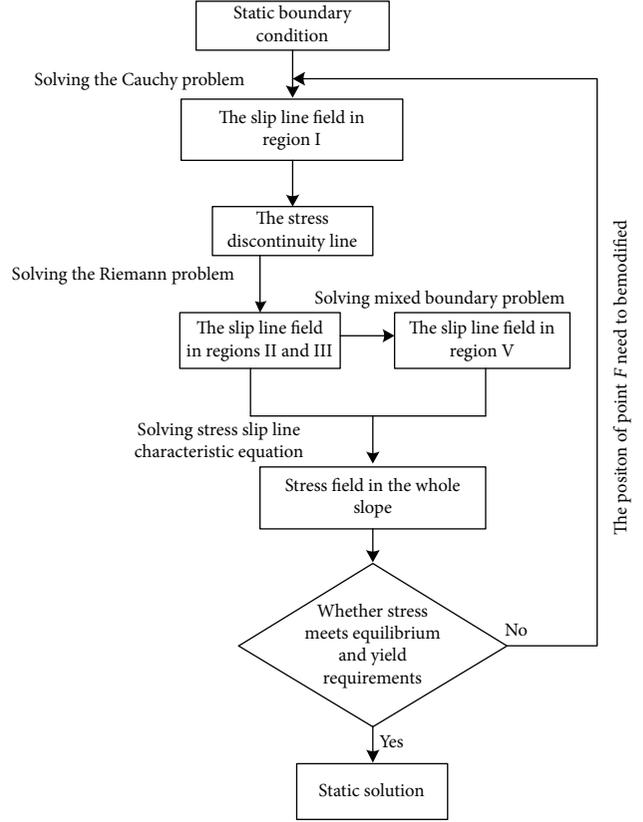


FIGURE 9: Flow chart of static calculation.

the grid. The stress discontinuity line is drawn by (15) and (21). A group of the maximum principal stress direction angles of the stress discontinuity line in the active region is assumed as θ^+ . Then, the direction angle θ^- , the tangential direction angle of the discontinuity line w , and the coordinate of each point in the line AF in the passive region can be obtained by the (13) and (15). And the slip line field of the region II can be obtained by solving the Cauchy problem.

- (3) According to the direction angle of the maximum principal stress θ artificially assumed and the coordinate of each point in the line FE , the effective stress σ_e is obtained. From the lines FB and FE , the slip line field of region III can be got by solving the Riemann problem. On the basis of maximum principal stress direction angle $\theta_2 = 90^\circ$, the slip line field of region V can be got by solving the mixed boundary value problem.
- (4) The numerical integration is carried out along the slip lines HF and FE to get the forces in the x and y directions and the moments of each force to the point O . Then the vertical stress on the boundary ED is calculated.
- (5) At this point, the static calculation is over. Then, check whether the force and the moment of the rigid block satisfy the equilibrium condition and whether the points within the rigid block satisfy the yield

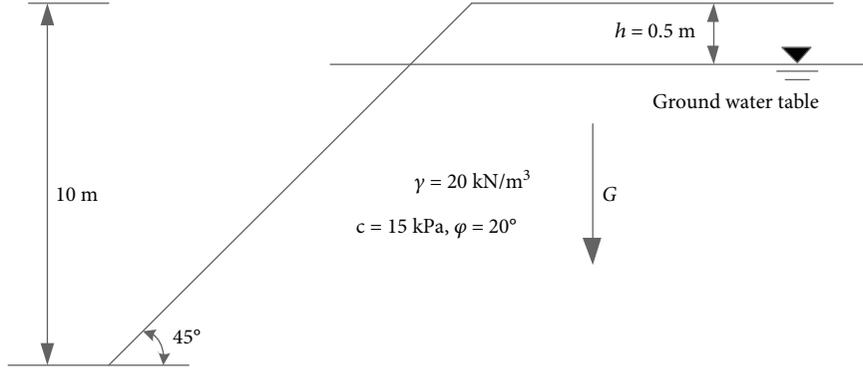


FIGURE 10: Parameters of a calculation model.

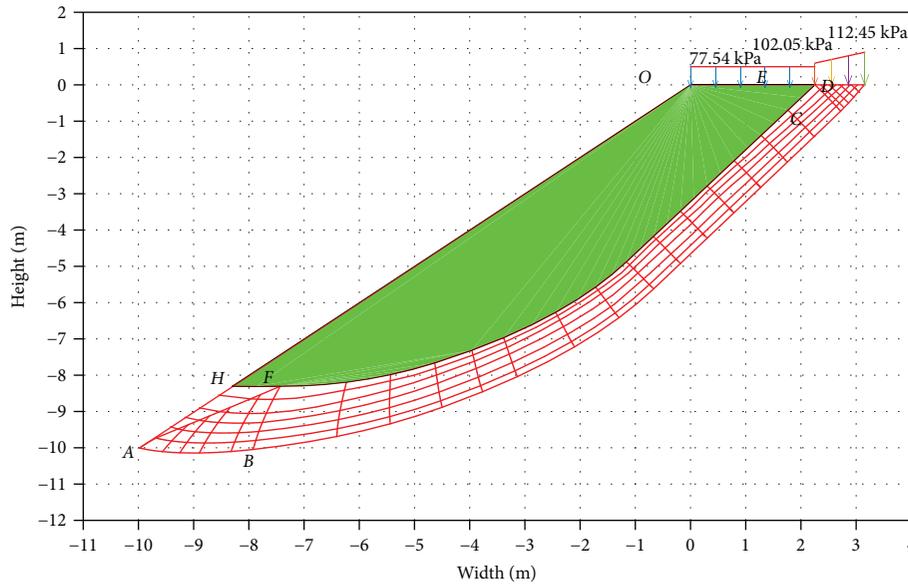


FIGURE 11: Distribution of the sliding belt and ultimate loads under the limit state.

condition. If not satisfied, the position of point F is modified to recalculate. The whole calculation process is shown in Figure 9.

4.2. Velocity Field Analysis. According to the associated flow rule, the dynamic analysis is carried out on the basis of the slip line field got by the static calculation.

- (1) The absolute velocity of point D on slope top is assumed as 1 m/s. The velocity component v_α and v_β in the slip line $ABCD$ is obtained according to the velocity characteristic equation.

$$\begin{aligned} v_\alpha &= v_x \cos(\theta - \mu) + v_y \sin(\theta - \mu), \\ v_\beta &= v_x \cos(\theta + \mu) + v_y \sin(\theta + \mu). \end{aligned} \quad (24)$$

- (2) The velocity field of region II can be obtained by solving the mixed boundary value problem for the condition that the tangential velocity on the discontinuity line remains unchanged (23). From the slip lines BCD and BF , the velocity field in regions III and V

and the velocity of each point in the line FE can be got by solving the Riemann problem. Then, the velocity v_α^- and v_β^- of each point in the line AF in the passive region II is translated into the velocity v_α^+ and v_β^+ in the active region I. Finally, the velocity field in the region I and the velocity in line FH are obtained by solving the Cauchy problem from the line AF .

- (3) At this point, the velocity analysis is over. Then, check whether the velocity in the lines HF and FE satisfy the kinematically rigid condition of the sliding block IV. If not satisfied, the input values of θ on the slip line FE and the position of the point F are modified to recalculate all.

5. Example Verification

Geometric parameters and material properties are shown in Figure 10. The Fortran 95 compiler is utilized to compute the calculation program for the rigorous solution satisfying the static equilibrium and the kinematical requirement.

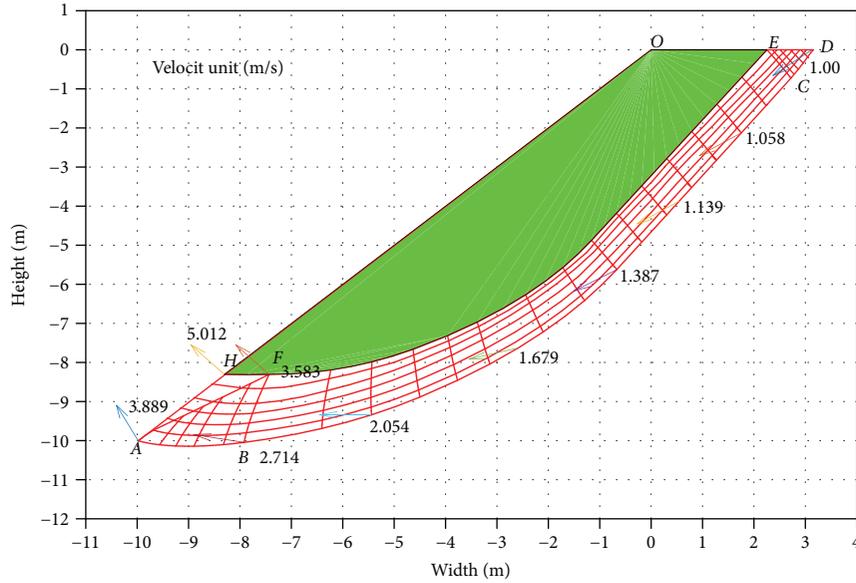


FIGURE 12: Velocity field of the slope under the limit state.

TABLE 1: Variation of the ultimate loads and the distribution of the sliding belt with the depth h .

Ground water depth h (m)	Ultimate load (kPa)			
	Rigid region	Plastic region (average)	OE (m)	ED (m)
0	77.80	107.69	2.32	0.87
0.5	77.54	107.25	2.32	0.87
1.0	76.69	106.74	2.31	0.87
1.5	73.34	105.39	2.31	0.86
2.0	70.78	104.38	2.30	0.86
2.5	64.09	101.67	2.29	0.86
3.0	59.83	99.93	2.28	0.85
4.0	43.89	93.51	2.27	0.85
No ground water	33.85	86.98	2.25	0.85

When a slope reaches the limit state, the ultimate bearing capacity is composed of two parts (Figure 11). While the load on the rigid region reaches 77.54 kPa, the other on the plastic region changes from 102.05 kPa to 112.45 kPa, considering the hydrostatic pressure. The rigid body rotates around point E , slipping along the sliding belt. The velocity of each point in the velocity discontinuity line $ABCD$ increases gradually from top to foot (Figure 12). The angle between a velocity discontinuity line and a slip line is φ . The position of points E and D calculated determines the general distribution of the sliding belt. In order to facilitate the appearance, the figures (Figures 11 and 12) show only part of the slip lines and the actual existence is hundreds of groups.

The limit state can be seen as a status with safety factor F_s equal to 1.0, and the external load at this time is the ultimate bearing capacity. When considering the hydrostatic pressure, the soils below ground water take floating weight resulting in the slide force falling and the slide resistance remains unchanged. From the Table 1, the ultimate load has become

TABLE 2: Variation of the ultimate loads and the distribution of the sliding belt with different soil parameters.

c (kPa), φ ($^\circ$), γ (kN/m^3), h (m)	Ultimate load (kPa)		OE (m)	ED (m)	
	Rigid region	Plastic region			
$\gamma = 10$	83.23	84.86	2.25	1.14	
$c = 15$	$\gamma = 15$	79.99	93.97	2.29	0.99
$\varphi = 20$	$\gamma = 20$	77.54	107.25	2.32	0.87
$h = 0.5$	$\gamma = 25$	71.91	125.62	2.34	0.77
$c = 10$	53.52	78.61	2.33	0.85	
$\gamma = 20$	$c = 15$	77.54	107.25	2.32	0.87
$\varphi = 20$	$c = 20$	110.57	191.98	2.31	0.89
$h = 0.5$	$c = 25$	198.88	331.25	2.30	0.92
$\varphi = 15$	54.22	75.89	2.39	0.79	
$\gamma = 20$	$\varphi = 20$	77.54	107.25	2.32	0.87
$c = 15$	$\varphi = 25$	108.52	198.55	2.28	1.03
$h = 0.5$	$\varphi = 30$	163.35	387.97	2.26	1.12

larger compared with the load without considering the ground water. As the depth increases causing the slide force to rise, the ultimate loads in both regions decrease, of which the one on the rigid region is influenced greater than the other by the ground water. While changing the depth, the distribution of sliding belt hardly alters. As the water level rises, the width of the sliding belt increases just a little.

From Table 2, when weight of soils increases gradually, the ultimate load on the rigid region decreases corresponding to the other on plastic region increasing. The distribution of sliding belt undergoes great changes due to the reducing width of ED . For $\gamma = 0 \text{ kN/m}^3$, the rigid region in the slope disappears under the limit state and the whole slip field becomes plastic. Changes of cohesive only bring about the increase of ultimate loads on both regions, and the sliding

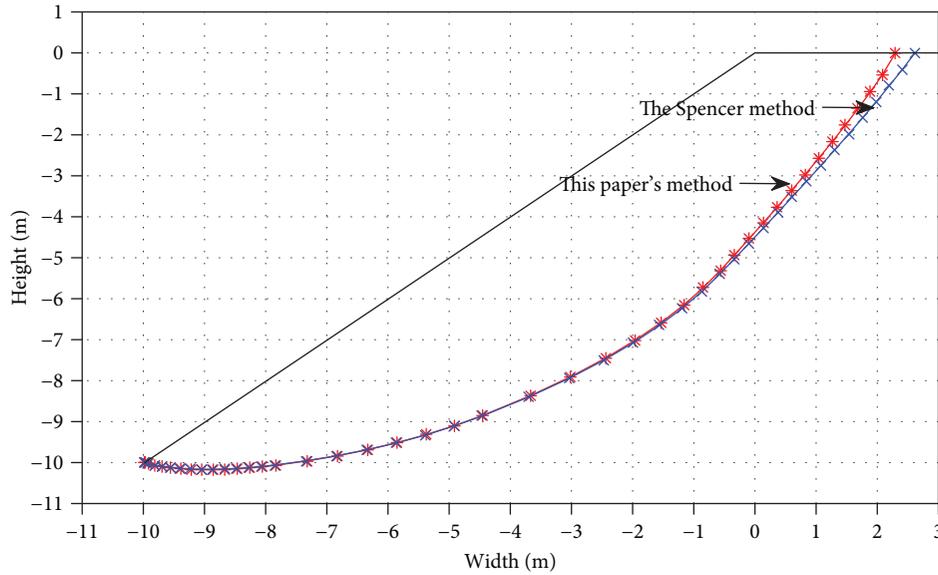


FIGURE 13: Comparison between this paper's method and the Spencer method.

belt has a little change. Internal friction angle has a great impact on the load on the plastic region and the sliding belt. With friction angle rising, the load on plastic region increases sharply and the width of sling belt fills out gradually. When the angle φ reduces to 0 resulting the plastic region disappeared, the sliding belt transforms into a signal slip surface and the bearing capacity of a slope depends on the cohesive. At this time, the ultimate load calculated by the algorithm is considered as the external load on the slope top and this status can be calculated by the Spencer method to get the slip surface and the safety factor. As shown in Figure 13, the critical slip surface obtained by this paper's method is basically consistent with the Spencer method. The safety factor F_s calculated by the Spencer method is 0.99, when the ultimate load is 65.53 kPa. Generally speaking, the comparison proves the feasibility of the algorithm proposed by this paper.

With the hydrostatic pressure, the difference between the two loads on both regions becomes smaller compared with the state of no ground water. The variable amplitude of the ultimate loads and the sliding belt with the different parameters is also diminished.

6. Discussion

During the process of searching the rigorous solution of slopes stability, it is necessary to find a slip line field satisfying all possible static and kinematic conditions. Although the calculation is so difficult that the iterative computation takes thousands of times, the author used the Fortran 95 compiler to compile the calculation program, which only takes less than one minute to obtain the rigorous solution and greatly simplified the calculation. Since the grid does not meet the requirements of infinite subdivision, it is not each point on the boundary but the intersection of the grid and the boundary that satisfies the stress boundary condition and the velocity boundary condition.

With the calculation model, the program is compiled based on the algorithm developed to obtain the ultimate bearing capacity and the distribution of the sliding belt. The influence of the ground water table on the ultimate bearing capacity and the sliding belt of a slope is discussed by an example. The results show that the ultimate load increases considering the hydrostatic pressure, while decreasing with the descending water level. However, the sliding belt remains constant. The soil weight and the internal friction angle have a tremendous impact on the belt, especially for the later. The ultimate load is influenced by the three parameters, of which the friction angle mainly controlled the load on the plastic region. When the internal friction angle reduces to zero, the sliding belt will translate into a traditional slip surface. Considering the ultimate load in this status as the external load on the slope top, the critical slip surface obtained by the Spencer method is basically consistent with the one got by this paper and the safety factor obtained by the Spencer method is 0.99 very close to 1.0. So the feasibility of this algorithm is verified by the specific example. Due to the hydrostatic pressure, the difference of loads on the two regions and the variation of sliding belt become smaller compared with not considering the ground water.

On the basis of this study, the next research focus is to add the seepage field into the slip field, which are not completely coincide. The authors hope that others can be motivated to consider adding the penetration force into the equilibrium equation of slip line to gain the numerical solution and the influence of seepage on the sliding belt and the ultimate load.

7. Conclusion

Based on the upper and lower bound theorems, the rigorous solution satisfying all static boundaries and kinematic boundaries is obtained, which approaches lower and upper bound solutions at the same time and is considered as the

actual solution. An algorithm for calculating the ultimate bearing capacity and the distribution of sliding belt has been developed, and the computer program has been accomplished. The application of this algorithm is verified by a specific example (Figure 13).

Considering the hydrostatic pressure, the variation of the ultimate load and the sliding belt under different parameters has been gained. The ultimate load decreases with the rising water level and become the minimum with no ground water. The sliding belt is controlled by the internal friction angle and the soil weight, while cohesive has a obvious impact on the ultimate loads. The difference of ultimate loads on the two regions and the variation of sliding belt become smaller because of the soils taking the floating weight.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (Grant no. 51579119).

References

- [1] X.-L. Chen, C.-G. Liu, Z.-F. Chang, and Q. Zhou, "The relationship between the slope angle and the landslide size derived from limit equilibrium simulations," *Geomorphology*, vol. 253, pp. 547–550, 2016.
- [2] D. Dong-ping, L. Liang, W. Jian-feng, and Z. Lian-heng, "Limit equilibrium method for rock slope stability analysis by using the Generalized Hoek–Brown criterion," *International Journal of Rock Mechanics and Mining Sciences*, vol. 89, pp. 176–184, 2016.
- [3] H. Zheng, D. F. Liu, and C. G. Li, "Slope stability analysis based on elasto-plastic finite element method," *International Journal for Numerical Methods in Engineering*, vol. 64, no. 14, pp. 1871–1888, 2005.
- [4] S.-H. Jiang, D.-Q. Li, L.-M. Zhang, and C.-B. Zhou, "Slope reliability analysis considering spatially variable shear strength parameters using a non-intrusive stochastic finite element method," *Engineering Geology*, vol. 168, pp. 120–128, 2014.
- [5] D.-Q. Li, T. Xiao, Z.-J. Cao, K.-K. Phoon, and C.-B. Zhou, "Efficient and consistent reliability analysis of soil slope stability using both limit equilibrium analysis and finite element analysis," *Applied Mathematical Modelling*, vol. 40, no. 9-10, pp. 5216–5229, 2016.
- [6] W. F. Chen and X. L. Liu, *Limit Analysis in Soil Mechanics*, Elsevier, Amsterdam, Netherlands, 1990.
- [7] F.-T. Liu, Y.-H. Fan, and J.-H. Yin, "The use of QP-free algorithm in the limit analysis of slope stability," *Journal of Computational and Applied Mathematics*, vol. 235, no. 13, pp. 3889–3897, 2011.
- [8] W. Fu and Y. Liao, "Non-linear shear strength reduction technique in slope stability calculation," *Computers and Geotechnics*, vol. 37, no. 3, pp. 288–298, 2010.
- [9] Y. Tu, X. Liu, Z. Zhong, and Y. Li, "New criteria for defining slope failure using the strength reduction method," *Engineering Geology*, vol. 212, pp. 63–71, 2016.
- [10] Q. Xu, H. Yin, X. Cao, and Z. Li, "A temperature-driven strength reduction method for slope stability analysis," *Mechanics Research Communications*, vol. 36, no. 2, pp. 224–231, 2009.
- [11] D. C. Drucker and W. Prager, "Soil mechanics and plastic analysis or limit design," *Quarterly of Applied Mathematics*, vol. 10, no. 2, pp. 157–165, 1952.
- [12] D. Li and Y. Cheng, "Lower bound limit analysis using nonlinear failure criteria," *Procedia Earth and Planetary Science*, vol. 5, pp. 170–174, 2012.
- [13] J. Zhou, Q. Chen, and J. Wang, "Rigid block based lower bound limit analysis method for stability analysis of fractured rock mass considering rock bridge effects," *Computers and Geotechnics*, vol. 86, pp. 173–180, 2017.
- [14] S. W. So and W. A. Strauss, "Upper bound on the slope of steady water waves with small adverse vorticity," *Journal of Differential Equations*, vol. 264, no. 6, pp. 4136–4151, 2018.
- [15] N. Deusdado, A. N. Antão, M. V. d. Silva, and N. Guerra, "Application of the upper and lower-bound theorems to three-dimensional stability of slopes," *Procedia Engineering*, vol. 143, pp. 674–681, 2016.
- [16] I. B. Donald and Z. Chen, "Slope stability analysis by the upper bound approach: fundamentals and methods," *Canadian Geotechnical Journal*, vol. 34, no. 6, pp. 853–862, 1997.
- [17] K. Krabbenhoft, A. V. Lyamin, M. Hjiaj, and S. W. Sloan, "A new discontinuous upper bound limit analysis formulation," *International Journal for Numerical Methods in Engineering*, vol. 63, no. 7, pp. 1069–1088, 2005.
- [18] M.-c. Yang and Y.-r. Zhen, "The analysis expression of non-characteristic line stress boundary condition in the slip line field theory for geotechnical materials," *Rock and Soil Mechanics*, vol. 22, no. 4, pp. 395–398, 2011.
- [19] N. Louat, "Solution of boundary value problems in plane strain," *Nature*, vol. 196, no. 4859, pp. 1081–1082, 1962.
- [20] S. Momani, S. Abusad, and Z. Oibat, "Variational iteration method for solving nonlinear boundary value problems," *Applied Mathematics and Computation*, vol. 183, no. 2, pp. 1351–1358, 2006.
- [21] R. T. Shield, "Stress and velocity fields in soil mechanics," *Journal of Mathematics and Physics*, vol. 33, no. 1–4, pp. 144–156, 1954.
- [22] O. C. Zienkiewicz, C. Humpheson, and R. W. Lewis, "Associated and non-associated visco-plasticity and plasticity in soil mechanics," *Geotechnique*, vol. 25, no. 4, pp. 671–689, 1975.
- [23] H. Liu, "Unified sand modeling using associated or non-associated flow rule," *Mechanics Research Communications*, vol. 50, pp. 63–70, 2013.

Review Article

Research Progress on Monitoring and Separating Suspension Particles for Lubricating Oil

Ziping Wang , Xian Xue, He Yin, Zhengxuan Jiang, and Yefei Li

Faculty of Civil Engineering and Mechanics, National Center for International Research on Structural Health Management of Critical Components, Jiangsu University, Zhenjiang, Jiangsu Province 212013, China

Correspondence should be addressed to Ziping Wang; wzpxx2004@126.com

Received 8 January 2018; Revised 12 April 2018; Accepted 8 May 2018; Published 5 June 2018

Academic Editor: Gangbing Song

Copyright © 2018 Ziping Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lubricant failure or irrational lubrication is the root cause of industrial equipment failure. By monitoring the distribution of the suspended particles in lubricants, it is possible to discover hidden lubrication problems. After taking the lubricating oil samples of industrial equipment, the oil monitoring technology is used to analyze the particle size distribution and the type and content of the abrasive particles by electrical, magnetic, and optical monitoring techniques. It is necessary to separate the suspended particles in oils with impurities by some method to eliminate potential safety hazards and ensure the reuse efficiency of the lubricant. In this paper, the principles, advantages, and disadvantages of several important oil monitoring methods are described, and new developments in various methods are analyzed. Several typical methods for separation of the suspended particles in purified oils were introduced. The advantages and disadvantages of each process were summarized. The development direction of lubricant monitoring technology was pointed out, and guidance was provided for the separation and online monitoring of the suspended particles in lubricants. Finally, compared with similar review papers, this paper specifically figured out that ultrasonic separation method has the advantages of real time, high efficiency, and no pollution and has important application value for micron-scale particle separation of large precision machines.

1. Introduction

Lubricants are mainly used in machinery and vehicles to reduce the friction between their parts. It reduces noise and plays a role in cooling. In some heavy industries and manufacturing industries, the economic losses caused by maintenance and downtime are as high as 50% of the operating costs [1]. The survey data from Shell of the United States showed that [2] about 35% of diesel engine operation failures and 38.5% of gear failures are due to wear, and almost 40% of rolling bearing failures are due to improper lubrication. The degree of wear and tear of the mechanical components can be inferred by the detection and study of the particle size and composition of the particles in the lubricant to control the wear rate, extend the service life of the equipment, and avoid catastrophic accidents. Lubricant quality testing and dynamic analysis have become one of the important means for the diagnosis and health assessment of mechanical equipment. People are constantly striving to find more accurate

methods, especially for various power mechanical systems that need long-term nonstop operation. The traditional method of lubricating oil testing is laboratory analysis after sampling. Laboratory analysis can provide comprehensive information on the wear of equipment components. The test results have a certain degree of accuracy. However, with too high technology, strict environment, and long testing time, it is susceptible to uncertainties, even for an experienced engineering analyst. The test results are still relatively discrete. And this offline laboratory test cannot provide real-time information on the health of the machine equipment. The lag in laboratory testing information has increased the risk of accidents in operating equipment. Research shows that [3], whether it is the production of equipment parts of wear particles or other pollution-generated particles, with the size of the particles in the 20–30 μm maximum impact on the device, less than 1 μm particles have little effect on the wear and tear, and particles larger than 100 μm can be collected and removed by magnetic plug inspection.

Therefore, solid particles with a particle size of 1–60 μm suspended in lubricating oil become the focus of monitoring.

Some online lubricating oil monitoring devices have been developed abroad, and real-time diagnosis of mechanical equipment becomes possible. Hager [4], Flanagan et al. [5], Wu et al. [6], and Martin et al. [7] used acoustic emission detection techniques to judge the quality of lubricants by reflecting the amplitude changes of sound waves, but this method is susceptible to mechanical background noise and lube oil temperature changes. Khandaker et al. [8], Keller and Saba [9], and Flanagan et al. [10] used a capacitive sensor to detect changes in the dielectric constant of lubricating oils, and the test results were often affected by changes in oil properties and oil ambient temperatures. It becomes very complicated, and the measurement of the dielectric constant cannot determine the size and concentration of the particles. Flynn and Whittington [11] further improved the resistive capacitive sensing method, which can not only detect iron particles but also nonferrous metal particles; however, only particles with a particle diameter greater than 100 μm can be detected, and particles smaller than 100 μm cannot be detected. In 1995, Liu et al. [12] confirmed that scatter counting optical methods can detect particles in lubricating oil, but the accuracy of the measurement is affected by the optical properties of the particles. Reintles et al. [13] studied the relationship between lubricating oil and particulate vibration and judged the wear status of the device by comparing with the vibration spectrum. The judgment result depends on the vibration spectrum of the previous study. Peng et al. [14] used the real-time measurement of lubricant wear debris for quantitative assessment of wear, developed an online particle counter to quantitatively assess equipment wear, and considered the total amount of debris as the quality loss of the test sample during the assessment process, but ignoring the contaminants and combustion products in the lubricant exaggerated the degree of wear. Iwai et al. [15] used a series of expansion and expansion methods to separate out 9.9 μm particles, but focused only on fluids with flow rates which are slower than 200 mL/min. Yilmaz and Morton [16] used oil fragment magnetic field sensors with seven channels arranged in parallel to monitor metal fragments in lubricants and successfully separated particles with particle sizes of 75 μm –105 μm and 125 μm –150 μm in different flow rates. The output detection is 7 times larger than the single channel, and the particle size is monitored. Du and Zhe [17] proposed a new asymmetric sharpening edge method to monitor high-flux particle concentration. This method has low sensitivity to oil flow and can separate particles of 9.94 μm . The method needs to arrange a series of sharp corners, and the monitoring structure is more complicated. Fan et al. [18] used frequency division multiplexing techniques to use multiple channel impedance-pulsed sensor shunts, but each frequency must correspond to a single channel. In recent years, the ultrasonic standing wave particle separation techniques [19–24] appeared. With the use of ultrasound standing wave field to move the transverse acoustic radiation generated by the suspended particles in fluids, to move the micron-sized suspended particles in continuous fluids, to achieve particle continuous separation mode using the Coulter counting

method, to separate metal particles, and to meet the real-time and nondisruptive requirements of online monitoring, Zhe et al. [25–27] have preliminary applied ultrasonic separation technology to the suspended particles in lubricating oil separation and online monitoring. Ren et al. [28] used a curved interdigital transducer to produce a stronger and more concentrated surface acoustic wave, reducing the energy loss of the surface acoustic wave during propagation. At the same time, this technique has achieved experimental success in the separation of polystyrene particles/polyamide (about 5 μm) [29] and cells in the blood (10–100 μm) [30] in liquids, but the effect of viscous on acoustic radiation is not considered, and only two particle sizes need to be separated. Our research team has also begun to study the role of aerosol particles in the microfluidic channel between the acoustic radiation force and the particles [31–34]. Zhu et al. [35] studied a single attribute sensor such as a wear sensor and a monitoring sensor [36] for online lubricating oil condition monitoring.

A variety of special function oil sensors provided the conditions for comprehensive monitoring. The above works present new ideas and methods for the researches on monitoring and separating lubrication oil theory. The aim of the paper is to analyze the progress of traditional and advanced monitoring and separating methods at present. Combined with the research results of this research group, the research focus and direction of the oil separation work are proposed which provide a guidance for corresponding research workers.

2. Main Monitoring Methods for the Suspended Particles in Lubricants

At present, the most widely used and effective oil monitoring technologies are lubricant's physical and chemical index analysis [37, 38] and wear particle analysis [39]. The physical and chemical index analysis of lubricants is to monitor the lubrication status of the equipment by monitoring the degree of changes in the physicochemical properties of the oil due to the loss of additives and the decay of the base oil. According to the different oil indicators, this method mainly includes atomic spectroscopy and infrared spectroscopy methods. The wear particle analysis method monitors the service life of lubricating oil and diagnosis equipment failure by changing the parameters such as the appearance, size, quantity, and color of wear particles carried in the lubricating oil, thereby determining the degree of contamination of the lubricating oil. Finally, the purpose of monitoring the equipment friction and fault diagnosis is achieved. Based on the different physical parameters of the monitored particles, this method mainly includes particle counting method, ferrography method, and magnetic plug and magnetic detection methods. The common lubricating oil monitoring methods and their advantages and disadvantages are shown in Table 1.

From the literature analysis, the online monitoring of lubricants and the real-time diagnostics of equipment health are gradually changing from qualitative to quantitative, the indirect analysis of the quality of lubricants to the accurate measurement of debris particles, and static detection to dynamic monitoring of large flows. From the perspective of

TABLE 1: Main monitoring methods for the suspended particles in lubricants.

Project	Monitoring principle	Advantages	Disadvantages
Particle counting method [40]	When the light path illuminates the sample, the light path is blocked, and the photoelectric receiver receives the change of the photoelectric intensity, which is converted into a voltage pulse signal. The number of particles of different sizes is recorded through different voltage valves.	The particle counter is simple to operate and has a fast counting speed. It is suitable for on-site monitoring.	Particles smaller than 100 μm in diameter cannot be monitored, and the particles cannot be qualitatively analyzed.
Atomic spectroscopy [41]	The content of this element is calculated by detecting the number of photons consistent with the characteristic frequency.	Easy to operate and no need for on-site treatment of lubricants.	Cannot detect the suspended particles larger than 10 μm .
Infrared spectroscopy method [42]	Based on the characteristic absorption peaks, numbers, and relative intensities of different substances, the presence of particles in the oil sample was deduced and its molecular structure was determined.	It can analyze oil degradation and pollution status quickly and efficiently.	Insensitive to metal particles, unable to analyze wear particle size and morphology, and no qualitative analysis.
Ferrography method [43]	A high-gradient magnetic field device was designed to separate the metal from the oil sample and to deposit it on a transparent substrate in order of size and then qualitatively and quantitatively analyze the abrasive particles.	Abrasive particle detection range size is from 1 to 1000 μm . It can perform the quantitative and qualitative analysis of abrasive particles at the same time.	The analysis is slow, and the analysis results strongly depend on the experience of the engineer.
Magnetic plug and magnetic detection methods [44, 45]	The metal's magnetic principle is used to analyze the morphology, number, and size of the captured metal abrasive particles.	Can detect larger abrasive particles (100–1000 μm).	Cannot detect aerosols below 100 μm .

the development of the method, the appearance of the new method mainly follows two different paths. One is to improve the existing methods, and the other is to propose new concepts or principles. The key issue in the development of online monitoring technology for lubricating oils is the accurate determination of the suspended particles in lubricating oil, which facilitates the capture of real-time information on the operation of machinery. The measurement accuracy depends on the collection and analysis of particles. The manipulation of these suspended particles can be based on other relatively mature technologies, combined with the characteristics of lubricating oil and the suspended particles, and theoretically researched and experimentally verified, forming a new online monitoring theory and method.

3. Separation Method for Lubricating Oil

In the aerospace and mechanical fields, foreign particles in the engine oil can cause engine wear and reduce the life of the engine and even cause major accidents. This problem requires more efficient removal or detection of foreign particles in the oil. According to the different purification methods, the traditional method of purifying oil is mainly divided into physical methods, chemical methods, and conjunction methods [46]. The physical methods mainly include sedimentation [47], filtration [48], distillation [49], and centrifugation [50]. The chemical method is mainly sulfuric acid-bearing clay refining technology [51]. The conjunction

methods mainly refer to combing the advantages of integrated physical methods and chemical methods, which can reduce pollution and increase efficiency [52].

3.1. Sulfuric Acid-Bearing Clay Refining and Separation. Sulfuric acid-bearing clay refining technology [53–55] regenerates waste oils with deeper deterioration, which can remove oxygenates, sulfur compounds, and nitrogen compounds in waste oils, and gums, asphaltenes, and asphalt are produced during use. The oil quality after regeneration is improved, reaching the standard of the base oil, but this technology produces a large amount of SO_2 , acid slag, acid water, and white clay slag during the regeneration of waste lubricating oil and brings about serious environmental pollution. In response to solve these problems, the IFP process [56] is used by means of adding propane to purify oil before the acidification of sulfuric acid, thus reducing the amount of sulfuric acid and clay and decreasing the production cost and environmental pressure. The white clay high-temperature refining process has disadvantages such as large amount of white clay, low oil recovery rate, severe equipment corrosion, and harsh operating conditions. In view of the problem of waste clay soil pollution, hydrogenation supplement refining technology came into being, replacing the original sulfuric acid-bearing clay refining process and has become the current mainstream process for the regeneration of waste lubricants. The use of hydrogenation supplement refining technology has advantages such as high oil recovery rate and good

product quality in waste lubricant regeneration, but this technology requires harsh operating conditions, huge equipment investment, and a suitable source of hydrogen; thus, its application is limited. Currently, the waste lubricant regeneration technologies mainly used is modified sulfuric acid clay technology [57], such as recycling acid slag and changing the feed rate of sulfuric acid. Meanwhile, the refining of lubricating oils from waste lubricating oil was examined, utilizing a novel blend of solvent extraction and activated alumina adsorbent [58], which has confirmed that solvent mixture can give good efficiency with the highest percent.

3.2. Vacuum Distillation. Decompression distillation method [59–62] removes water by decompression distillation firstly and obtains a certain amount of lube oil fraction through a metal element-containing additive whose boiling point is generally higher than that of a lube oil fraction, such as light oil and pitch, but its flash point, viscosity, and acid value are still noncompliant. Therefore, a second step is required for refining. NMP or furfural is used to remove undesirable components such as colloids and acidic oxides; thereby, better-quality base oil is extracted. However, this method requires high degree of operating vacuum and high temperature in the rectification process, and the oil is prone to cracking, resulting in equipment corrosion caused by refined additives. Though many variables have been studied in this research, such other variables as mixing, pressure, settling time, and temperature will affect the purification results. Further research is required to take this process to the compensation.

3.3. Centrifugation Separating Method. Centrifugation is a method of extracting the pure oil by separating the suspended particles in the lubricant by a centrifugal separator [63–65]. It uses a liquid separation aid having a density higher than the density of the oil to attract and combine the suspended particles. As shown in Figure 1, the contaminated oil is supplied to a rotary centrifuge separation chamber. The central outlet of the purified oil separation chamber is discharged, and the liquid separation aid and the separated particles are discharged through the outlet of the outer separation chamber located radially at the middle outlet. Figure 1 shows a mathematical model of a tube centrifuge. “ r ” is the radius of the area that determines the settling velocity. “ r_1 ” and “ r_2 ” are the radius of the liquid layer surface in the cylinder and the cylinder’s inner wall, respectively. The liquid phase can be continuously introduced into the inlet of the bottom. The rotation of the cylinder causes the liquid phases to rotate at the same angular velocity, and the fluid is uniform in the axial direction and the outer layer is formed. A layer next to the cylinder and the inner layer is a “liquid phase surface” that is not in contact with air. Under the action of the acceleration of the centrifuge, the movement of the input material is constrained, and the particles suspended in the continuous liquid phase are dispersed under the strong shearing force and lose the focused state. However, its main weakness that lies in the separation effect has a close relation with the radius of the centrifuge rotor and the composition of the solvent and the particles; thus, it is mainly suitable for offline separation.

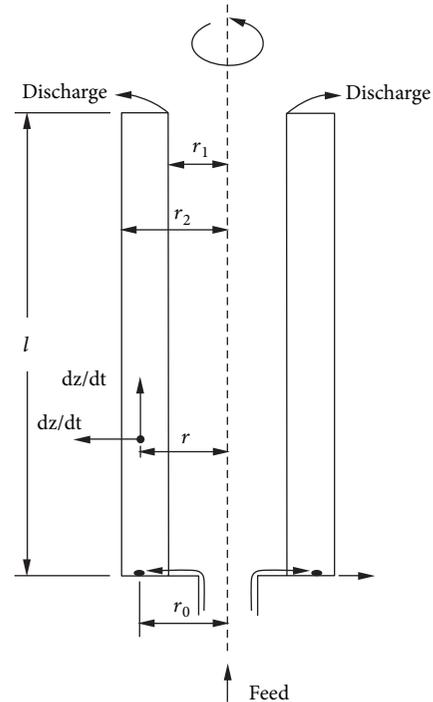


FIGURE 1: Schematic diagram of tube centrifuge [64].

3.4. Membrane Separation. Membrane separation technology [66–68] uses selective permeation membrane as the separation medium. Under the action of the driving force on both sides of the membrane, the components on the raw material side can selectively permeate the membrane to achieve the concentration, separation, and purification of lubricating oil. Its specific processing technology is shown in Figure 2. Membrane separation technology [69] is widely used in advanced water quality and water reuse. Owing to the large lubricating viscosity of impurities, the amount of particle passing through the membrane is very small. In the application process, ultrafiltration is usually used to remove the water-soluble salts produced from the consumption of the same additives, colloidal particles, asphaltenes, and carbon black in waste lubricating oils. In recent years, polymer organic membranes developed by Miyagi et al. [70] can reduce 14% of polar substances and 32% of oxides in waste lubricating oils, effectively improving the quality of oil products. Bart Van der Bruggen [71] used drone membrane technology to remove the content of metal particles, cleaner than the US national base oil standards. In addition, increasing the temperature and adding organic solvents, supercritical fluid technology, and so on are often used in reducing the viscosity of the lubricating oil, thereby increase the amount of membrane filtration of waste lubricants. Although the membrane treatment technology has the advantages of less pollution, simple operation, and low energy consumption, blockage and inefficient separation are usually caused by large viscosity of waste lubricants and low filtration amount.

Traditional separation method can only remove the impurity particles in oil to a certain degree. However, for large-scale precision machines (such as vehicle engines and fighter jets), the impurity particles in the engine oil are

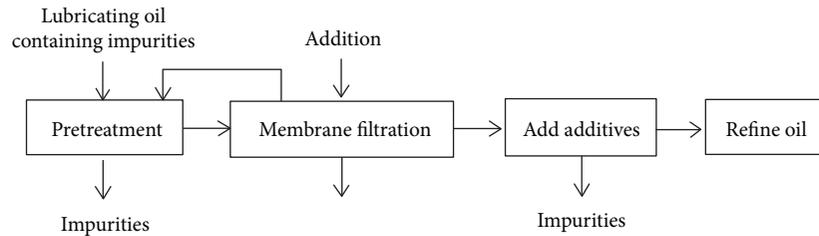


FIGURE 2: Membrane separation process.

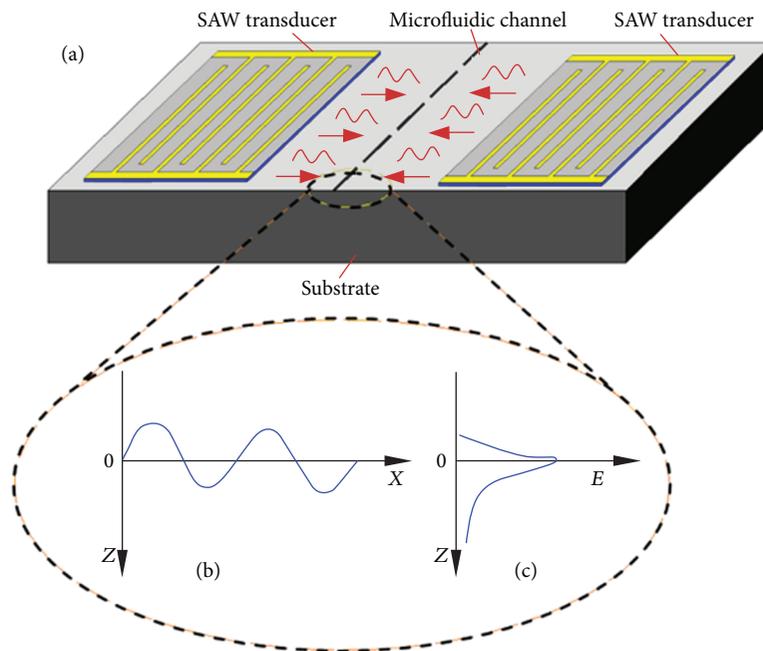


FIGURE 3: Schematic diagram of ultrasonic separation method.

required to be within several micrometers; thus, traditional separation methods are difficult to purify lubricating oil in a real sense. Ultrasonic separation or aggregation techniques have quite possibility to obtain higher selective, accurate, and reliable analytical results. Therefore, it is of great significance for oil sample to employ pretreatment techniques such as separation or aggregation.

3.5. Ultrasonic Separation Technology. Compared with the traditional purifying method, the ultrasonic separation technology has broad application prospects in this field for its advantages of continuous, high efficiency, low contact, and low pollution. However, the ultrasonic separation technology cannot be applied without the microfluidic chip and the theory of acoustic radiation force; thus, it is necessary to develop a transducer that can generate an ultrasonic field. In order to prepare a transducer capable of generating a qualified surface acoustic wave standing wave field and separating micron-sized particles in the oil, our team conducted related research and experiments [31–34].

The theory that particles can be moved by the acoustic radiation force in the ultrasonic field was firstly proposed by King [72] who calculated the acoustic radiation forces and related conditions of rigid spheres in fluids. This theory

was later promoted by Yosioka and Kawasima [73] for elastic spheres. At present, ultrasonic separation technology has been widely used in biomedical [74–76] and chemical industry fields [77–79], and with the increasing influence of biomedical and microchemical technologies [80] on human society, this technology will appear huge potential application value.

Figure 3 shows a schematic diagram of a device for separating particles using acoustic surface standing waves and a working mechanism in our research group [31–34]. Two identical interdigital (IDT) electrodes are mounted on a piezoelectric substrate. The width of the microfluidic channel is equal to half of a wavelength. The microfluidic channel is installed between two IDTs so that only one node of surface acoustic wave (SAW) field is formed inside the microfluidic channel. The oil containing the foreign particles is then fed into the pipe from one end using a pressure transmitter or a peristaltic pump. The particles are evenly distributed in the channel when the SAW is not excited. When the same RF signal is applied to a pair of IDTs, two columns of the SAWs with the same amplitude and frequency but opposite propagation directions are generated. In the microchannel region, when two wave arrays are superimposed, a single-node acoustic surface standing wave field is

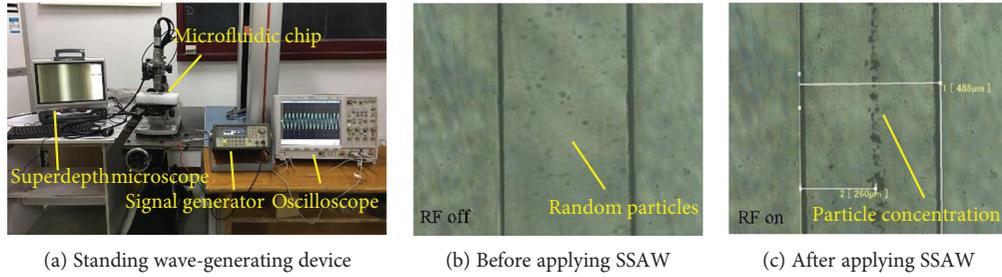


FIGURE 4: Particle concentration experiment device and result graph [31].

generated. The particles in the standing wave field will be concentrated by the ultrasonic radiation force to the node, to separate the impurities. The experimental setup that is shown in Figures 4(a)–4(c) demonstrates particle concentration effects of particles observed before and after the application of voltage using ultra-well-depth microscopes, respectively, providing technical platform for online monitoring of lubricating oil.

Compared with the traditional method, the method of separating the foreign particles by the acoustic radiation force received by the particles in the ultrasonic standing wave field has its unique advantages such as the suspended particles in the fluid which can be separated easily, continuously, and efficiently and has the characteristics of no contact and low pollution. In addition, the engine oil due to a high degree of cleanliness has a protective function for the operation of large-scale equipment (automobiles, airplanes, etc.) engines; thus, it must be replaced on schedule. Ultrasonic separation technology can not only test the quality of lubricant oil, timely check the content of impurity particles in the lubricant, but also can replace the replaced lubricant for secondary cleaning and reuse, and the cleaned lubricant can fully meet the use of skill requirements. This will not only reduce the wear and tear on the mechanical equipment of the impurity particles, extend the working time of the equipment, but also reduce waste, protect the environment, reduce pollution, and greatly improve social and economic benefits.

4. Outlook and Challenges

In recent years, with the development of microfabrication processes and the development of a variety of separating methods for the suspended particles, particle separation technology has achieved breakthroughs in terms of separation accuracy and application range. Ultrasonic standing wave separation methods perform in both biological and industrial fields. Great potentials for application prospects of the key techniques for monitoring and separating suspension particles for lubricating oil are as follows.

- (i) Traditional physical and chemical separating methods for oil are generally cost-effective and environment-friendly due to their destroying original oil quality and affection online operation, which can be used as supplement and comparison method to separate smaller wear debris.

- (ii) The wideband IDT has the advantages of small power loss, more accurate positioning, and stronger propagation capability and can be used to generate a variety of frequency-adjustable standing waves. Therefore, designing and fabricating an IDT integrated with the substrate can improve the separation efficiency of particles with different particle sizes in the microfluidic channel and provide guarantees for the separation of smaller particles.
- (iii) Due to the large perturbation of tiny particles in viscous fluids, the forces experienced by particles in different flow regimes vary greatly. The forces and motion models of particles in viscous fluids need to be refined, so as to accurately and comprehensively analyze the particles. The force and motion state in the microfluidic channel are of great significance for achieving the aggregation and separation of microparticles.
- (iv) Develop a highly integrated, low-energy, and low-cost aerosol particle control device to establish a mechanism that can be used on a large scale in industrial production, which will greatly increase the service life of the engine and will have important implications for environmental protection.

5. Conclusion

The conventional methods for separating the suspended solids from lubricating oils are reviewed in this paper. However, due to the shortcomings of these conventional methods, it is more popular at home and in abroad to use ultrasonic separation methods for particle separation. Based on the abovementioned several common methods for separation of particles, the future methods for separation of particles are prospected. And especially the efficiency improvement and the achievement on separation of smaller particles are prospected, which will provide a monitoring basis for online monitoring of lubricants. With its environment-friendly characteristics, it will be widely used in a large-scale production of impurities in engine oil.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 11402101 and no. 11520101001), the Postdoctoral Program of Jiangsu Province (no. 1501109B), and the Jiangsu University Foundation (no. 14JDG022).

References

- [1] X. F. Jiang, Y. W. Fei, K. Qian, and Y. Q. Hu, "The application of lubricant oil monitoring to aero-engine preventive maintenance," *Lubrication Engineering*, vol. 2, pp. 56–58, 2004.
- [2] M. H. Jones, *Tribology—a Key Element in Condition Monitoring. Proc of Inter Conf on Condition Monitoring*, Oxford University Press, Oxford, UK, 2001.
- [3] Y. T. Zhang and G. Q. Ren, "The present condition and development of oil monitoring technique," *Lubrication Engineering*, vol. 2, pp. 65–66, 2002.
- [4] H. E. Hager, "Fluid property evaluation by piezoelectric crystals operating in the thickness shear mode," *Chemical Engineering Communications*, vol. 43, no. 1–3, pp. 25–38, 1986.
- [5] I. M. Flanagan, J. R. Jordan, and H. W. Whittington, "Wear-debris detection and analysis techniques for lubricant-based condition monitoring," *Journal of Physics E: Scientific Instruments*, vol. 21, no. 11, pp. 1011–1016, 1988.
- [6] T. H. Wu, H. K. Wu, Y. Du, and Z. X. Peng, "Progress and trend of sensor technology for on-line oil monitoring," *Science China Technological Sciences*, vol. 56, no. 12, pp. 2914–2926, 2013.
- [7] S. J. Martin, A. J. Ricco, T. M. Niemczyk, and G. C. Frye, "Characterization of SH acoustic plate mode liquid sensors," *Sensors and Actuators*, vol. 20, no. 3, pp. 253–268, 1989.
- [8] I. I. Khandaker, E. Glavas, and G. R. Jones, "A fibre-optic oil condition monitor based on chromatic modulation," *Measurement Science and Technology*, vol. 4, no. 5, pp. 608–613, 1993.
- [9] M. Keller and C. Saba, "Monitoring of ester based lubricant by dielectric constant," *Lubrication Engineering*, vol. 45, pp. 347–351, 1989.
- [10] I. M. Flanagan, J. R. Jordan, and H. W. Whittington, "An inductive method for estimating the composition and size of metal particles," *Measurement Science and Technology*, vol. 1, no. 5, pp. 381–384, 1990.
- [11] B. W. Flynn and H. W. Whittington, "Improved transducer design for machine wear debris monitoring," *Electronics Letters*, vol. 31, no. 3, pp. 177–179, 1995.
- [12] Y. Liu, Z. Liu, Y. Xie, and Z. Yao, "Research on an on-line wear condition monitoring system for marine diesel engine," *Tribology International*, vol. 33, no. 12, pp. 829–835, 2000.
- [13] J. Reintles, R. Mahon, M. D. Duncan, and A. Schultz, "Optical debris monitoring," in *Proceedings of the 49th Meeting of the Society for Machinery Failure Prevention Technology*, pp. 263–272, Taiwan, 1995.
- [14] Z. Peng, N. J. Kessissoglou, and M. Cox, "A study of the effect of contaminant particles in lubricants using wear debris and vibration condition monitoring techniques," *Wear*, vol. 258, no. 11–12, pp. 1651–1662, 2005.
- [15] Y. Iwai, T. Honda, T. Miyajima, S. Yoshinaga, M. Higashi, and Y. Fuwa, "Quantitative estimation of wear amounts by real time measurement of wear debris in lubricating oil," *Tribology International*, vol. 43, no. 1–2, pp. 388–394, 2010.
- [16] N. Yilmaz and B. Morton, "Effects of preheating vegetable oils on performance and emission characteristics of two diesel engines," *Biomass and Bioenergy*, vol. 35, no. 5, pp. 2028–2033, 2011.
- [17] L. Du and J. Zhe, "Parallel sensing of metallic wear debris in lubricants using undersampling data processing," *Tribology International*, vol. 53, pp. 28–34, 2012.
- [18] L.-L. Fan, Y. Han, X.-K. He, L. Zhao, and J. Zhe, "High-throughput, single-stream microparticle focusing using a microchannel with asymmetric sharp corners," *Microfluidics and Nanofluidics*, vol. 17, no. 4, pp. 639–646, 2014.
- [19] A. K. Patel, *Analysis of magnetic field and electromagnetic forces in transformer and superconducting magnets*, MS thesis, National Institute of Technology, Rourkela, India, 2015.
- [20] J. J. Hawkes, W. T. Coakley, M. Gröschl et al., "Single half-wavelength ultrasonic particle filter: predictions of the transfer matrix multilayer resonator model and experimental filtration results," *The Journal of the Acoustical Society of America*, vol. 111, no. 3, pp. 1259–1266, 2002.
- [21] M. Wiklund, C. Günther, R. Lemor, M. Jäger, G. Fuhr, and H. M. Hertz, "Ultrasonic standing wave manipulation technology integrated into a dielectrophoretic chip," *Lab on a Chip*, vol. 6, no. 12, pp. 1537–1544, 2006.
- [22] S. Kapishnikov, V. Kantsler, and V. Steinberg, "Continuous particle size separation and size sorting using ultrasound in a microchannel," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, no. 1, article P01012, 2006.
- [23] J. Shi, X. Mao, D. Ahmed, A. Colletti, and T. J. Huang, "Focusing microparticles in a microfluidic channel with standing surface acoustic waves (SSAW)," *Lab on a Chip*, vol. 8, no. 2, pp. 221–223, 2008.
- [24] L. Y. Yeo and J. R. Friend, "Surface acoustic wave microfluidics," *Annual Review of Fluid Mechanics*, vol. 46, no. 1, pp. 379–406, 2014.
- [25] S. Murali, A. V. Jagtiani, X. Xia, J. Carletta, and J. Zhe, "A microfluidic Coulter counting device for metal wear detection in lubrication oil," *Review of Scientific Instruments*, vol. 80, no. 1, article 016105, 2009.
- [26] L. Du, J. Zhe, J. Carletta, R. Veillette, and F. Choy, "Real-time monitoring of wear debris in lubrication oil using a microfluidic inductive Coulter counting device," *Microfluidics and Nanofluidics*, vol. 9, no. 6, pp. 1241–1245, 2010.
- [27] J. Zhe, L. Du, J. E. Carletta, and R. J. Veillette, "Metal wear detection apparatus and method employing microfluidic electronic device," US Patent US8522604B2, 2013.
- [28] L. Ren, Y. Chen, P. Li et al., "A high-throughput acoustic cell sorter," *Lab on a Chip*, vol. 15, no. 19, pp. 3870–3879, 2015.
- [29] A. Nilsson, F. Petersson, H. Jönsson, and T. Laurell, "Acoustic control of suspended particles in micro fluidic chips," *Lab on a Chip*, vol. 4, no. 2, pp. 131–135, 2004.
- [30] X. Ding, S. C. S. Lin, B. Kiraly et al., "On-chip manipulation of single microparticles, cells, and organisms using surface acoustic waves," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 28, pp. 11105–11109, 2012.
- [31] J. Z. Gu, K. Wang, Y. Luo, and C. G. Xu, "Preliminary study on separation in viscous fluid based on ultrasonic wave guide," *Piezoelectrics & Acousto-optics*, vol. 36, pp. 178–181, 2014.
- [32] Z. Wang, Y. Luo, G. Zhao, and F.-G. Yuan, "Design and optimization of an OPFC ultrasonic linear phased array

- transducer," *International Journal of Mechanics and Materials in Design*, vol. 13, no. 1, pp. 57–69, 2017.
- [33] Z. P. Wang, X. Xue, Y. Luo, and F. G. Yuan, "Standing surface acoustic wave technology applied for micro-particle concentration in oil," in *Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XII*, p. 105991G, Denver, CO, USA, 2018.
- [34] Z. Wang, Z. Jiang, L. Chen, Y. Li, M. Li, and S. Wang, "Standing wave performance test of IDT-SAW transducer prepared by silk-screen printing," *AIP Advances*, vol. 8, no. 5, article 055303, 2018.
- [35] X. Zhu, C. Zhong, and J. Zhe, "Lubricating oil conditioning sensors for online machine health monitoring – a review," *Tribology International*, vol. 109, pp. 473–484, 2017.
- [36] J. Zhu, J. M. Yoon, D. He, Y. Z. Qu, and E. Bechhoefer, "Lubrication oil condition monitoring and remaining useful life prediction with particle filtering," *International Journal of Prognostics and Health Management*, vol. 4, pp. 124–138, 2013.
- [37] A. Rezasoltani and M. Khonsari, "On monitoring physical and chemical degradation and life estimation models for lubricating greases," *Lubricants*, vol. 4, no. 3, p. 34, 2016.
- [38] B. Rahimi, A. Semnani, A. Nezamzadeh-Ejhi, H. Shakoori Langeroodi, and M. Hakim Davood, "Monitoring of the physical and chemical properties of a gasoline engine oil during its usage," *Journal of Analytical Methods in Chemistry*, vol. 2012, Article ID 819524, 8 pages, 2012.
- [39] B. Fan, S. Feng, Y. Che, J. Mao, and Y. Xie, "An oil monitoring method of wear evaluation for engine hot tests," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9–12, pp. 3199–3207, 2018.
- [40] W. W. Seifert and V. C. Westcott, "A method for the study of wear particles in lubricating oil," *Wear*, vol. 21, no. 1, pp. 27–42, 1972.
- [41] R. Q. Aucélio, R. M. de Souza, R. C. de Campos, N. Miekeley, and C. L. P. da Silveira, "The determination of trace metals in lubricating oils by atomic spectrometry," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 62, no. 9, pp. 952–961, 2007.
- [42] F. R. van de Voort, J. Sedman, R. A. Cocciardi, and D. Pinchuk, "FTIR condition monitoring of in-service lubricants: ongoing developments and future perspectives," *Tribology Transactions*, vol. 49, no. 3, pp. 410–418, 2006.
- [43] O. P. Sondhiya and A. K. Gupta, "Wear debris analysis of automotive engine lubricating oil using by ferrography," *International Journal of Engineering and Innovative Technology*, vol. 2, no. 5, pp. 46–54, 2012.
- [44] S. Kumar, P. S. Mukherjee, and N. M. Mishra, "Online condition monitoring of engine oil," *Industrial Lubrication and Tribology*, vol. 57, no. 6, pp. 260–267, 2005.
- [45] L. Zeng, H. Zhang, Q. Wang, and X. Zhang, "Monitoring of non-ferrous wear debris in hydraulic oil by detecting the equivalent resistance of inductive sensors," *Micromachines*, vol. 9, no. 3, p. 117, 2018.
- [46] H. Yin, K. M. Solval, J. Huang, P. J. Bechtel, and S. Sathivel, "Effects of oil extraction methods on physical and chemical properties of red salmon oils (*Oncorhynchus nerka*)," *Journal of the American Oil Chemists Society*, vol. 88, no. 10, pp. 1641–1648, 2011.
- [47] E. G. Latondress, "Oil-solids separation in edible oil processing," *Journal of the American Oil Chemists Society*, vol. 60, no. 2Part1, pp. 257–261, 1983.
- [48] Y. P. Galaguz and G. L. Safina, "Modeling of particle filtration in a porous medium with changing flow direction," *Procedia Engineering*, vol. 153, pp. 157–161, 2016.
- [49] A. H. Ammar, A.-H. Meniai, and F. Zagrouba, "Experimental study and modeling of essential oil extraction from plants by hydrodistillation," *Chemical Engineering & Technology*, vol. 37, no. 7, pp. 1235–1242, 2014.
- [50] A. U. Hahn and K. L. Mittal, "Mechanism of demulsification of oil-in-water emulsion in the centrifuge," *Colloid & Polymer Science*, vol. 257, no. 9, pp. 959–967, 1979.
- [51] W. Jiang, "Technology for producing industrial sulfuric acid from refinery sour gas," *Petroleum Processing & Petrochemicals*, vol. 3, 2001.
- [52] I. M. Atadashi, M. K. Aroua, and A. A. Aziz, "High quality biodiesel and its diesel engine application: a review," *Renewable & Sustainable Energy Reviews*, vol. 14, no. 7, pp. 1999–2008, 2010.
- [53] R. R. Mohammed, I. A. R. Ibrahim, A. H. Taha, and G. Mckay, "Waste lubricating oil treatment by extraction and adsorption," *Chemical Engineering Journal*, vol. 220, pp. 343–351, 2013.
- [54] A. S. Abdulkareem, A. S. Afolabi, S. O. Ahanonu, and T. Mokrani, "Effect of treatment methods on used lubricating oil for recycling purposes," *Energy Sources*, vol. 36, no. 9, pp. 966–973, 2014.
- [55] A. Amiri and M. Ghanbarzadeh, "Regeneration of spent clay in used oil re-refining process by solvent extraction," *ISWA World Congress*, vol. 7, pp. 1–15, 2013.
- [56] B. P. Xu, N. Jiang, and M. Lei, "On-spot evaluation of entropy production ratio under heat transfer intensification. Anonymous, IFP launches a new process for producing propylene," *Petroleum Refining and Chemical Industry*, vol. 31, no. 6, pp. 46–50, 2000.
- [57] J. D. Udonne, "A comparative study of recycling of used lubrication oils using distillation, acid and activated charcoal with clay methods," *Journal of Petroleum and Gas Engineering*, vol. 2, no. 2, pp. 12–19, 2011.
- [58] D. I. Osman, S. K. Attia, and A. R. Taman, "Recycling of used engine oil by different solvent," *Egyptian Journal of Petroleum*, vol. 3, 2017.
- [59] M.-. S. Kim, J.-. S. Hwang, and H.-. R. Kim, "Re-refining of waste lube oils by vacuum distillation with petroleum atmospheric residuum," *Journal of Environmental Science and Health . Part A: Environmental Science and Engineering and Toxicology*, vol. 32, no. 4, pp. 1013–1024, 1997.
- [60] Y. Elkasabi, C. A. Mullen, and A. A. Boateng, "Distillation and isolation of commodity chemicals from bio-oil made by tail-gas reactive pyrolysis," *ACS Sustainable Chemistry & Engineering*, vol. 2, no. 8, pp. 2042–2052, 2014.
- [61] B. Doshi, M. Sillanpää, and S. Kalliola, "A review of bio-based materials for oil spill treatment," *Water Research*, vol. 135, pp. 262–277, 2018.
- [62] I. Hamawand, T. Yusaf, and S. Rafat, "Recycling of waste engine oils using a new washing agent," *Energies*, vol. 6, no. 2, pp. 1023–1049, 2013.
- [63] C. J. Breeding and R. T. Marshall, "Crystallization of butter oil and separation by filter-centrifugation," *Journal of the American Oil Chemists Society*, vol. 72, no. 4, pp. 449–453, 1995.
- [64] H. E. Lun, Z. Jiang, J. Sun, and L. I. Yi, "The application of centrifugal oil purifier in shield construction," *Tunnel Construction*, pp. 1–5, 2009.

- [65] J. Cheng, "Centrifugal separation," *The Science and Technology of Gelatin*, vol. 16, pp. 1–5, 2011.
- [66] M. Padaki, R. Surya Murali, M. S. Abdullah et al., "Membrane technology enhancement in oil–water separation. A review," *Desalination*, vol. 357, pp. 197–207, 2015.
- [67] Y. Zhu, D. Wang, L. Jiang, and J. Jin, "Recent progress in developing advanced membranes for emulsified oil/water separation," *NPG Asia Materials*, vol. 6, no. 5, article e101, 2014.
- [68] C. Miguel, L. Matos, F. Magalhães, L. Madeira, and A. Mendes, "Membrane ultrafiltration for oil-from-water separation: multidisciplinary lab experiment," *International Journal of Engineering Education*, vol. 30, pp. 254–262, 2014.
- [69] A. Kupareva, P. Mäki-Arvela, and D. Y. Murzin, "Technology for rerefining used lube oils applied in Europe: a review," *Journal of Chemical Technology and Biotechnology*, vol. 88, no. 10, pp. 1780–1793, 2013.
- [70] A. Miyagi, M. Nakajima, H. Nabetani, and R. Subramanian, "Feasibility of recycling used frying oil using membrane process," *European Journal of Lipid Science and Technology*, vol. 103, no. 4, pp. 208–215, 2001.
- [71] B. Van der Bruggen, "Integrated membrane separation processes for recycling of valuable wastewater streams: nanofiltration, membrane distillation, and membrane crystallizers revisited," *Industrial & Engineering Chemistry Research*, vol. 52, no. 31, pp. 10335–10341, 2013.
- [72] L. V. King, "On the acoustic radiation pressure on spheres," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 147, no. 861, pp. 212–240, 1934.
- [73] K. Yosioka and Y. Kawasima, "Acoustic radiation pressure on a compressible sphere," *Acta Acustica united with Acustica*, vol. 5, pp. 167–173, 1955.
- [74] H. Yang, H. Sun, Y. Zhang, and Y. Zhang, "Design and fabrication of an ultrasonic microdevice for microparticles separation," in *IEEE International Conference on Nano/micro Engineered and Molecular Systems*, pp. 1159–1162, Bangkok, Thailand, 2007.
- [75] A. Włoch, H. Czyż, and T. Jasiński, "Ultrasonic methods of the cells separation in human blood," *Acta Physica Polonica A*, vol. 128, no. 2, pp. 234–236, 2015.
- [76] T. Dung Luong and N. Trung Nguyen, "Surface acoustic wave driven microfluidics – a review," *Micro and Nanosystemse*, vol. 2, no. 3, pp. 217–225, 2010.
- [77] A. Chappellaz, M. Alexander, and M. Corredig, "Phase separation behavior of caseins in milk containing flaxseed gum and κ -carrageenan: a light-scattering and ultrasonic spectroscopy study," *Food Biophysics*, vol. 5, no. 2, pp. 138–147, 2010.
- [78] B. Lipkens, J. Dionne, M. Costolo, A. Stevens, and E. Rietman, "Separation of bacterial spores from flowing water in macro-scale cavities by ultrasonic standing waves," *The Journal of the Acoustical Society of America*, vol. 128, no. 4, p. 2338, 2010.
- [79] J. Shi, D. Ahmed, X. Mao, and T. J. Huang, "Surface acoustic wave (SAW) induced patterning of micro beads in microfluidic channels," in *2008 IEEE 21st International Conference on Micro Electro Mechanical Systems*, pp. 26–29, Wuhan, China, 2008.
- [80] P. Sajeesh and A. K. Sen, "Particle separation and sorting in microfluidic devices: a review," *Microfluidics and Nanofluidics*, vol. 17, no. 1, pp. 1–52, 2014.

Research Article

Contactless Modal Phenomena Based Approach to Detecting, Identifying, and Diagnosing of Electrical Connections

Pavel Orlov  and Talgat Gazizov 

Tomsk State University of Control Systems and Radioelectronics, Lenin Ave. 40, Tomsk 634050, Russia

Correspondence should be addressed to Pavel Orlov; praetorian281@gmail.com

Received 11 January 2018; Revised 4 April 2018; Accepted 15 April 2018; Published 30 May 2018

Academic Editor: Gangbing Song

Copyright © 2018 Pavel Orlov and Talgat Gazizov. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a unified description of a new approach for contactless detection, identification, and diagnostics of electrical connections and describes an idea and principles of using modal probing for these tasks. Simulation and experimental results on pulsed signal propagation through flat cables demonstrate the modal decomposition of a pulsed signal, which varies depending on the state of the probed wire. It is shown that the presented tasks can be solved by modal probing. The article also considers the analysis of modal distortions in frequency domain and gives the formula for its practical use. This formula can be useful when the pulse duration time is longer than the minimum of mode delay difference. In conclusion, we present further development ideas of the modal probing.

1. Introduction

To provide error-free and stable performance of electronic and electrical systems it is important to control their functioning. In this case, detection, identification, and diagnostics of electrical connections become urgent [1], especially for such areas as aeronautics and aerospace [2]. One of the ubiquitous techniques is reflectometry. The development of this technique increases its functionality and applications [3–5]. Unfortunately, the reflectometry response itself is not always self-sufficient to identify and locate the defects in the electrical connections and it is a reason why a solution of the inverse problem may also enhance the applicability of reflectometry [6, 7]. However, wiring network development strengthens the requirements for probing devices, which necessitates creation of the devices based on other principles. The impedance spectroscopy can be applicable for wire fault diagnostics [8, 9]. Particularly, the development of contactless methods is important [10].

A new device has already been suggested for contactless (hereinafter, the term “contactless” means no need for galvanic connection with the device under test) detection, identification, and diagnostics based on the idea of employing modal distortions of pulsed signal waveform [11]. However,

the practical implementation of the device requires careful investigation of the modal phenomena in multiwire structures. A number of theoretical studies have been carried out using the software for quasistatic and electromagnetic simulation showing the ability to apply modal phenomena for detecting, identifying, and diagnosing multiwire structures and to create devices based on these phenomena [12]. Moreover, there have been some experiments carried out to confirm that one pulse can be decomposed into several pulses with lower amplitudes because of different modal delays in the structure [13], with approaches to apply modal phenomena for protecting critical equipment against the effects of UWB-pulses propagation being described [14]. Experimental results on flat cable have demonstrated the possibility of contactless wire diagnostics using modal probing even without galvanic connection to the wire [15]. The obtained results have demonstrated the possibility of developing new devices for detecting, identifying, and diagnosing electrical connections. However, the implementation of the possibility requires analyzing the signal not only in time but also in frequency domains. The first steps of modal distortion analysis in frequency domain for contactless diagnostic of electrical connections have already been described [16]. Unfortunately, generalized presentation of the obtained and some new

results on modal probing is still absent. Meantime, it can be useful to reveal the promising ways of future work.

In this paper, a generalized description of possible applications of modal phenomena for contactless detection, identification, and diagnostics of electrical connections is given for the first time.

This paper is organized as follows: Section 2 presents the theoretical background of modal probing. Section 3 describes the simulation approaches used in this paper. Section 4 describes the use of modal probing for detecting and identifying electrical connections. The diagnostics opportunities are presented in Section 5, and the analysis of modal distortion in frequency domain is presented in Section 6. The paper is concluded in Section 7.

2. Modal Probing Background

It is known that during the propagation of the pulse signal along the N -conductor transmission line (conductor $N + 1$ is the reference one) in nonhomogeneous dielectric filling, the pulse signal may be subjected to modal distortions up to decomposition into N pulses of smaller amplitude due to different modal delays. Complete decomposition of the pulse signal in the line of length l will occur if the total duration t_Σ of exciting pulse is less than the minimum modulus among the differences of modal delays, that is, under condition [17]

$$t_\Sigma < l \min |\tau_i - \tau_k|, \quad i, k = 1, \dots, N, \quad i \neq k, \quad (1)$$

where $\tau_{i(k)}$ is per-unit-length delay for $i(k)$ -th mode of structure. Indeed, according to modal theory [18], a pulse excitation of the N -conductor transmission line is considered as combination of the pulse modes, propagating in the line with own per-unit-length delays (as well as other characteristics). Each of the delays multiplied by l will give the corresponding time when the pulse arrives at the line end. In case of small values of the time, the neighboring pulses can overlap. However, if the minimal of the values is more than the total duration of the exciting pulse then the overlapping of the mode pulses will be diminished until the pulses are completely decomposed at the line end. This phenomenon can be used for detecting, identifying, and diagnosing electrical connections. In this paper, generalization of these opportunities is called modal probing. If probed conductors have different electrical and magnetic couplings with the probing line, the information about probed conductors can be obtained from the waveform of modal distortions in the probing line.

A block diagram of the device that implements the principles of modal probing is shown in Figure 1. The device operates as follows: the probing pulse from the generator propagates along a probing line. This pulse undergoes modal distortions caused by the presence of probed conductors. A transmitted signal from the probing line output and a reflected signal from the probing line input go to the receiver inputs and then to the processing unit. All units of the device function according to control unit signals. Information about the probed structure is extracted from the waveforms at the near and far ends of the probing line.

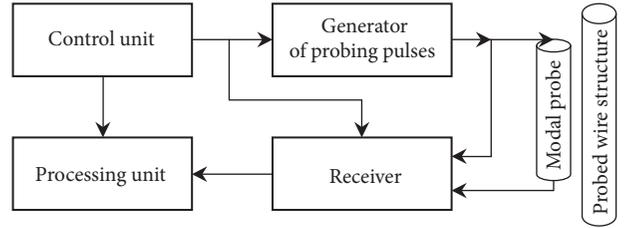


FIGURE 1: Block diagram of the modal probing device.

3. Simulation Approach

In this paper, the simulation of pulse propagation is carried out by electromagnetic and quasistatic approaches. The first is used for validation made on the basis of finite integration technique. The second is used as a main approach based on fast and accurate models implemented in the available TALGAT software [19]. The simulation is described below in more detail. The TALGAT software is based on the method of moments and allows us to make 2D quasistatic analysis. The algorithm implemented in the software allows calculating all elements of a moment matrix by using fully analytical formulae only, avoiding the time-consuming and approximate numerical integration. It can be useful for effective calculation of a capacitive matrix of 2D structures of various complexity. (Complete details of the algorithm are commonly available [20] and omitted here because of awkwardness.) We simulate a short pulse propagation along a multiconductor transmission line as a base of the structures under consideration. It is assumed in the analysis that a transmission line is uniform along its length with an arbitrary cross section. The cross section, in general, with N signal conductors and a reference, is represented by the following $N \times N$ matrices of line per-unit-length parameters: inductance (\mathbf{L}), coefficients of electrostatic induction (\mathbf{C}), resistance (\mathbf{R}), and conductance (\mathbf{G}). In paper [21] the approach based on a modified nodal admittance matrix has been presented to formulate network equations including multiconductor transmission lines, terminal, and interconnecting networks. We use the algorithm based on this approach and permitting to calculate the voltage not only at any node of the network but also at any point along any conductor of multiconductor transmission lines. (The algorithm details and various applications are not described here for the sake of brevity but can be found in [22–24].) It is the approach that is used in our research, and voltages in the time domain are obtained by applying the inverse fast Fourier transform.

4. Detection and Identification

By detection we mean the ability to detect passive (probed) conductors, while by identification we mean the ability to determine the amount of probed conductors and boundary conditions. The possibility of detecting and identifying electrical connections by modal probing is illustrated by quasistatic modeling of the trapezoidal pulse signal distortions in the microstrip structures with the length of 1.5 m (Figure 2).

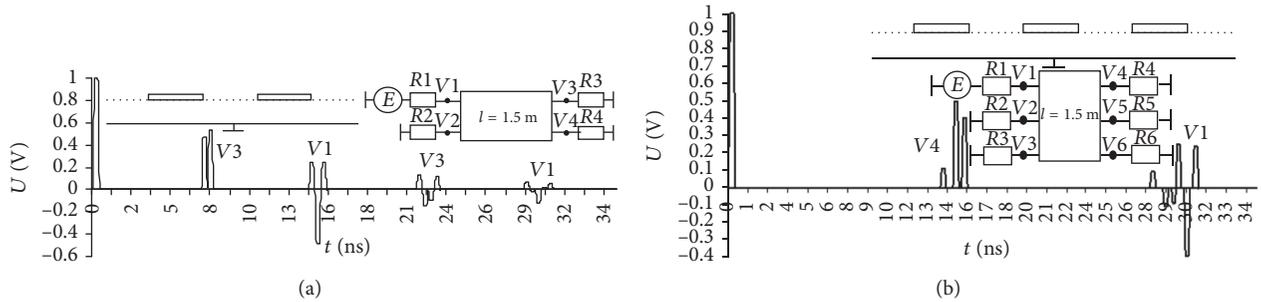


FIGURE 2: Voltage waveforms at the input and output of the microstrip line active conductor when (a) $N = 2$ and (b) $N = 3$.

TABLE 1: The matrices of the lines.

N	Matrices					
	L , nH/m			C , pF/m		
2	285	27		91	-3.8	
	27	285		-3.8	91	
3	284	27	7.2	91	-3.8	-0.05
	27	284	27	-3.8	91	-0.3
3	7.2	27	284	-0.05	-0.3	91

As the excitation source we chose a trapezoidal signal with EMF 2 V and the duration of rise, fall, and flat top of 0.1 ns. The boundary conditions at the ends of lines were chosen from the pseudomatching condition for active line and open circuit for others. The matrices of per-unit-length parameters are shown in Table 1. More detailed simulation setup and parameters of the structures are given in [17].

For $N = 2$ at the far end of the (active conductor) probing line (V3) there are two pulses instead of one (Figure 2(a)). The second pulse was caused by the presence of the probed passive conductor (and, as a consequence, by the excitation of even and odd modes), by its electric and magnetic couplings with the probing line and by the fact that the total duration of the input pulse is less than the total difference between mode delays. The difference of mode delays is caused by the inhomogeneous dielectric filling of the structure. For $N = 3$ at the far end of the probing line (V4) there are three pulses instead of one (Figure 2(b)). The appearance of three pulses is caused by the presence of two passive conductors, so three modes are excited in the structure and the delay difference between them is more than the pulse duration.

To confirm the possibility of using modal phenomena to detect electrical connections, we have carried out an experiment with the experimental PCB structure. The cross section and the photography of the experimental PCB are shown at Figure 3 and its parameters are presented in Table 2. The value of the resistance at the ends of lines with the length of 0.33 m during the experiment is 50Ω (R1–R4 on the schematic diagram in Figure 2(a)).

The experimental setup was based on a stroboscopic oscilloscope S9–11 (50Ω internal impedance, 17.8 GHz bandwidth) comprising a signal analyzer, an indicator, a generator,

a stroboscope, and a pulse shaper (Figure 4). A 20 dB attenuator was connected to the stroboscope input to protect it. The cable was connected to the pulse shaper and the attenuator via SMA connectors.

Each experiment was conducted in the following order. First, the UWB pulse from the pulse shaper propagated to an input connector, then directly (without the device under test) to the output connector, and, finally, to the attenuator and the stroboscope input to measure the signal waveform. Afterwards, the PCB or cable under test (DUT) was connected between the input and output connectors and the resulting signal waveform was measured. This experimental setup applies to all measurements with S9–11.

A voltage pulse was applied between the active trace and the reference planes of the experimental PCB. The signal parameters at the pulse shaper output under 50Ω load are listed in Table 3. The photography of the oscilloscope S9–11 with the measured input and output voltage waveforms is shown in Figure 5(a). (The input pulse was measured with a 20 dB attenuator, while other pulses were measured without it.) The comparative analysis of measured and simulated (by electromagnetic approach based on FIT method) voltage waveforms (Figure 5(b)) shows good coincidence.

Thus, these results show that according to the number of pulses at the far end of the active conductor we can determine the presence and the amount of passive conductors, that is, to solve the problem of detecting and identifying electrical connections.

Let us consider the possibility of modal probing for flat cables (Figure 6). The parameters of typical flat cables are presented in Tables 4–6. The data show that cables with air gaps have the difference of mode per-unit-length delays of more than 0.3 ns/m and cables without air gaps 0.5 ns/m. Therefore, modal probing is applicable for such cables.

To confirm the possibility of using modal phenomena to detect electrical connections, we carried out an experiment using S9–11 oscilloscope. The parameters of pulsed signals were measured at the near and far ends of the probing pair of the PUGNP 3×1.5 cable wires (Figure 6(a)), where **A** represents the active wire, **R** the reference wire, and **P** the passive wire. Thus, we considered the possibility of passive wire probing (without contact with it), based on the signal between the active and reference wires of the probing pair.

A voltage pulse was applied between the active and reference wires (a schematic diagram of the structure under

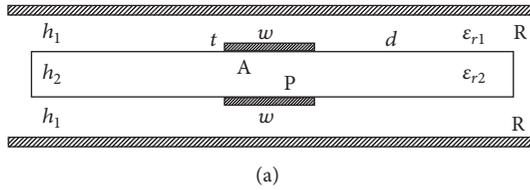


FIGURE 3: The cross section of the experimental PCB (a); photograph of the experimental PCB with the scalar network analyzer P2M-40 connected to it (b).

TABLE 2: Parameters of the experimental structure.

$w, \mu\text{m}$	$t, \mu\text{m}$	$h_2, \mu\text{m}$	Parameters				$\Delta\tau, \text{ns/m}$	Matrices				
			$h_1, \mu\text{m}$	ϵ_{r1}	ϵ_{r2}	$L, \text{nH/m}$		$C, \text{pF/m}$		Z, Ω		
185	35	130	600	4.25	10.2	2.2	524.7	310.9	243.4	-179.5	68	46
							310.9	524.7	-179.5	243.4	46	68

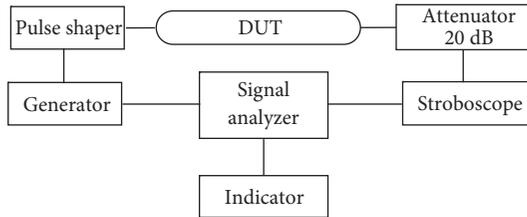


FIGURE 4: Schematic diagram of the experimental setup.

TABLE 3: Signal parameters at the pulse shaper output.

Signal type	Triangular pulse
Amplitude	320 mV
Rise time (0.1–0.9)	340 ps
Fall time (0.1–0.9)	340 ps
Duration (0.5)	240 ps

study is similar to that shown in Figure 2(a), but the length of the line (l) is 15 m) through the pulse shaper (with the output resistance $R1 = 50 \Omega$ and the maximum signal amplitude of 10 V). The input resistance of the oscilloscope was $R3 = 50 \Omega$, while $R2$ and $R4$ were changed: open circuit, short circuit, and 100Ω . The signal parameters at the pulse shaper output under 50Ω load are listed in Table 7, and the waveform is shown in Figure 7(a). (The input pulse was measured with a 20 dB attenuator, while other pulses were measured without it.)

The waveforms at the far end of the probing line are shown in Figure 7. Pulse amplitudes for different boundary conditions at the ends of the probed wire are presented in Table 8. In all these cases, there are two pulses instead of one at the far end of the probing line. The second pulse was caused by the presence of the probed passive conductor (and, as

a consequence, by the excitation of even and odd modes), by its electric and magnetic couplings with the probing line, and by the fact that the total duration of the input pulse ($\approx 0.6 \text{ ns}$) is less than the total difference between mode delays ($0.32 \text{ ns/m} \times 15 \text{ m} = 4.8 \text{ ns}$), as follows from Table 6. The oscilloscope pattern with the waveforms at the far end of the probing line (Figure 7(b)) with open circuit at the passive wire shows the ability to detect the conductor without any contact with it. To determine the boundary conditions at the ends of the passive conductor, the pulse shape should vary depending on the boundary conditions. There are two pulses at far end of the line in all oscilloscope patterns but their amplitudes depend on boundary conditions (Table 8).

Theoretical background of this dependence is based on modal theory and consists in the following. The resulting voltages and currents in the transmission lines under consideration are represented as superposition of voltages or currents of two decoupled transmission lines, having intrinsic characteristic impedance and per-unit-length delays, defined by even and odd modes. First, the exciting signal is represented as superposition of even and odd modes excitations. Then, each of the lines is excited by the intrinsic excitation. The resulting output voltage depends on transmission coefficients at the beginning and the end of a line. These coefficients are defined by boundary conditions and characteristic impedance of the modes. In case of pulse excitation, the resulting output voltage consists of two (even and odd modes) pulses of different amplitudes. In general, the values of the amplitudes are different, because they are defined by different transmission coefficients of the modes. The change of boundary conditions changes the resulting output voltages of even and odd pulses. Analytical expressions for transmission coefficients and pulse amplitudes can be easily derived similarly to [25]. However, this approach does not take into account the losses and dispersion, which may significantly

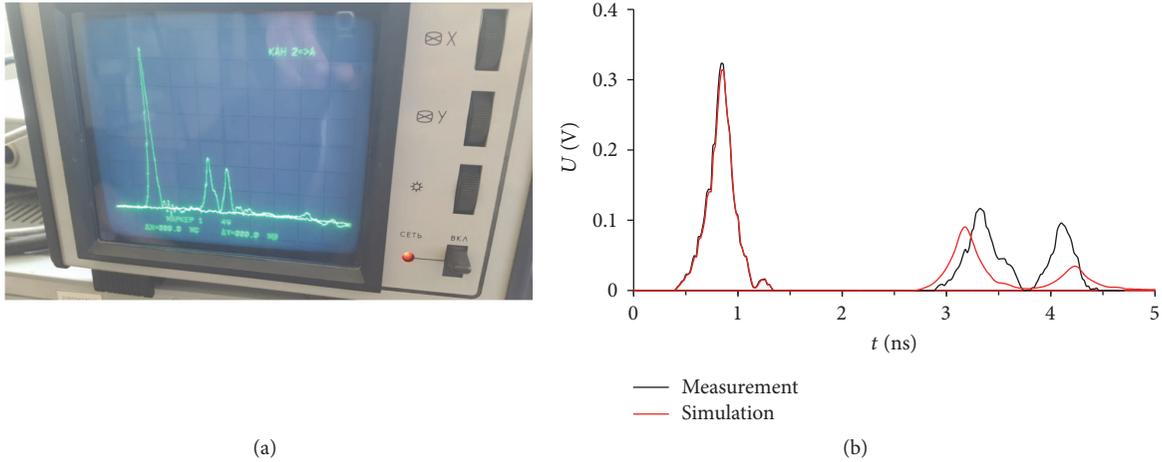


FIGURE 5: Photography of the oscilloscope S9-11 with measured input and output voltage waveforms (a); comparison of measured and simulated voltage waveforms (b) for the experimental PCB.

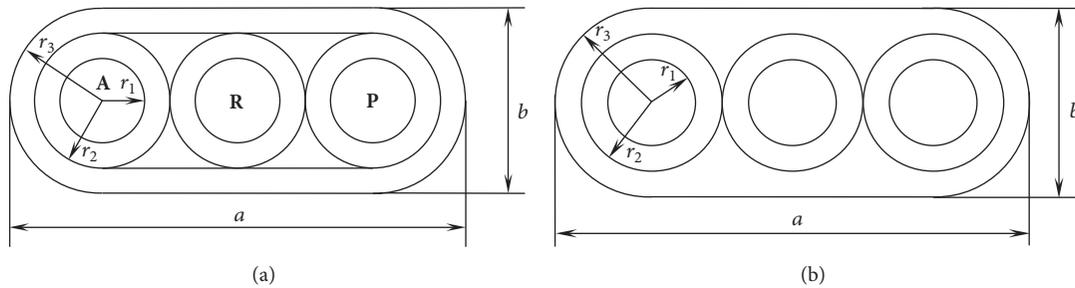


FIGURE 6: Cross section of cables with (a) and without (b) air gaps.

affect the amplitudes of the pulses. Therefore, numerical full-wave or above-mentioned quasistatic analysis is more relevant.

To verify the results of the experiment, we accomplished the quasistatic simulation. The results of the simulation are shown in Figure 8. For clarity, the comparison of the experiment and simulation results for open circuit case is shown in Figure 9. As we can see, the qualitative ratio between pulse amplitudes in the experiment and the simulation is similar, but the values of the amplitudes and the fall time are greater in the experiment. These differences are due to the difference of the actual and simulation parameters (of input signal, dielectrics, conductors, and boundary conditions), as well as measurement and simulation errors. The noncausal waveform of the simulation result is explained by ignoring the frequency dependence of ϵ_r and $\text{tg}\Delta$ during the simulation process. This effect is considered in detail in [26]. However, more detailed simulation and explanations of these results are out of the scope of this paper.

Thus, the results of the experiments and the simulations for cable of PUGNP 3×1.5 type confirm the possibility of detecting and identifying boundary conditions at the ends of a conductor by modal probing without electrical connection to the conductor. It is worth noting that none of the existing and utilized ways (pulse reflectometry, device-based induction methods) has such possibility.

5. Diagnostics

By diagnostics we mean the ability to determine passive (probed) conductor breaks. To diagnose a passive conductor with modal probing, the form of a modal distortion of the pulse signal should vary depending on the condition of the passive conductor. It is illustrated by the simulation of the pulse propagation along the flat cable of PUGNP 3×4 type (Figure 6(b)). We simulated two cases, with and without a break in a passive wire. The case with the break in the passive wire was simulated by using two sections of the cable (Figure 10), with $R_3 = R_4 = 5 \text{ G}\Omega$. The total length of the structure was equal to 2 m, and the point of the passive wire break was moved over the distances of 0.5, 1, and 1.5 m from the near end of the passive wire. (The probing pulse parameters are EMF amplitude 2 V; each of rise, fall, and flat top time values is 100 ps). The simulation results are shown in Figure 11.

Simulation results show that, in the case without a break of a passive wire (Figure 11(a)), two pulses instead of one come to the far end of the probing line, and the reason why they occur was described above. With the passive wire break, the number of pulses at the end of the probing line increases (Figures 11(b)–11(d)). When the break point is located at the distance of 0.5 m from the near end of the structure, the number of pulses at the far end of the probing line doubles. This is due to the fact that at the passive wire break point the

TABLE 4: Geometrical (mm) and material parameters of cables with air gaps.

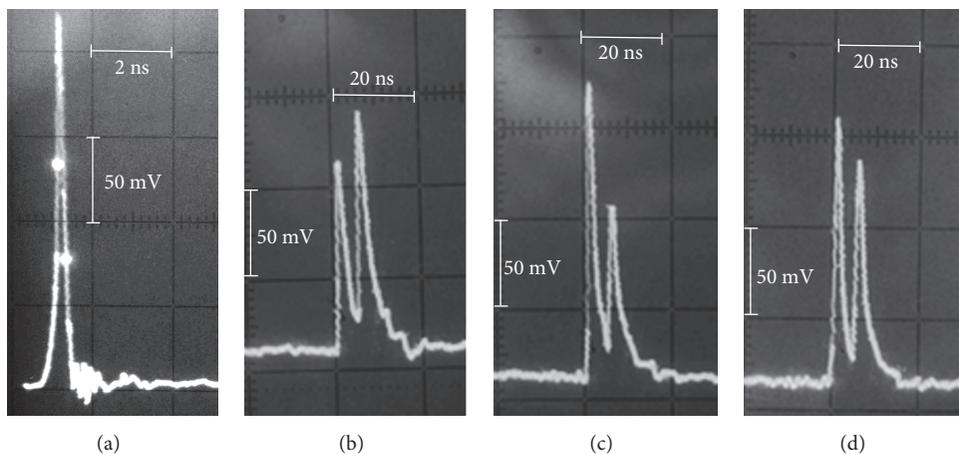
Cable type	Parameters				Material	
	Wire radii (r_1)	Wire isolation radii (r_2)	Cable isolation radii (r_3)	Cable size ($a \times b$)	Conductor	Dielectric
VVG 3 × 1.5	0.690	1.200	2.055	8.89 × 4.11	Cu	PVC
VVG 3 × 2.5	0.892	1.350	2.160	10.02 × 4.32	Cu	PVC
VVG 3 × 4	1.128	1.825	2.875	12.97 × 5.75	Al	PVC
PUGNP 3 × 1.5	0.690	1.175	1.950	8.40 × 3.90	Cu	PVC

TABLE 5: Geometrical (mm) and material parameters of cables without air gaps.

Cable type	Parameters				Material	
	Wire radii (r_1)	Wire isolation radii (r_2)	Cable isolation radii (r_3)	Cable size ($a \times b$)	Conductor	Dielectric
PUGNP 3 × 2.5	0.892	1.375	2.000	9.60 × 4.00	Cu	PVC
PUGNP 3 × 4	1.128	1.600	2.275	11.50 × 4.55	Cu	PVC

TABLE 6: L , C , and Z matrixes, mode per-unit-length delays, and their difference for the cables under consideration.

Cable type	Matrixes						Per-unit-length delays, ns/m		
	L , nH/m		C , pF/m		Z , Ω		τ_e	τ_o	$\Delta\tau$
VVG 3 × 1.5	458.53	111.59	55.56	-10.09	94.40	60.16	5.090	4.777	0.313
	111.59	458.53	-10.09	55.56	60.16	94.40			
VVG 3 × 2.5	387.65	83.41	62.22	-9.63	81.21	48.59	4.977	4.676	0.302
	83.41	387.65	-9.63	62.22	48.59	81.21			
VVG 3 × 4	422.48	96.73	58.64	-9.56	87.62	53.86	5.048	4.713	0.335
	96.73	422.48	-9.56	58.64	53.86	87.62			
PUGNP 3 × 1.5	448.25	106.95	56.05	-9.88	92.74	58.51	5.063	4.743	0.320
	106.95	448.25	-9.88	56.05	58.51	92.74			
PUGNP 3 × 2.5	397.04	86.94	73.81	-8.90	74.88	43.08	5.602	5.063	0.539
	86.94	397.04	-8.90	73.81	43.08	74.88			
PUGNP 3 × 4	351.63	70.29	82.82	-8.65	66.19	36.14	5.594	5.073	0.521
	70.29	351.63	-8.65	82.82	36.14	66.19			

FIGURE 7: Oscilloscope waveforms of input signal (a) and signal at the far end of the probing line (V4) with open circuit (b), short circuit (c), and 100 Ω (d) at both ends of the probed conductor.

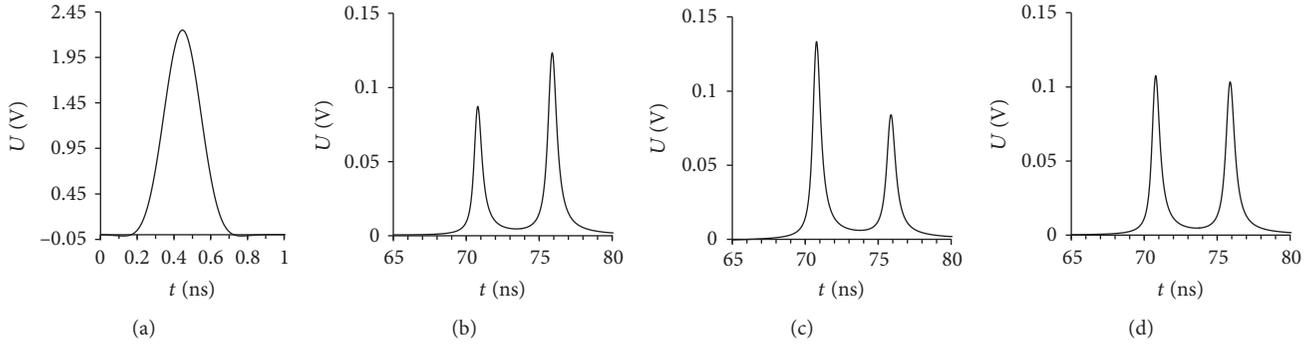


FIGURE 8: Simulated waveforms of input signal (a) and signal at the far end of the probing line (V_4) with open circuit (b), short circuit (c), and $100\ \Omega$ (d) at both ends of the probed conductor.

TABLE 7: Signal parameters at the pulse shaper output.

Signal type	Triangular pulse
Amplitude	225 mV
Rise time (0.1–0.9)	280 ps
Fall time (0.1–0.9)	280 ps
Duration (0.5)	200 ps
Horizontal scale division	2 ns/div
Vertical scale division	50 mV/div

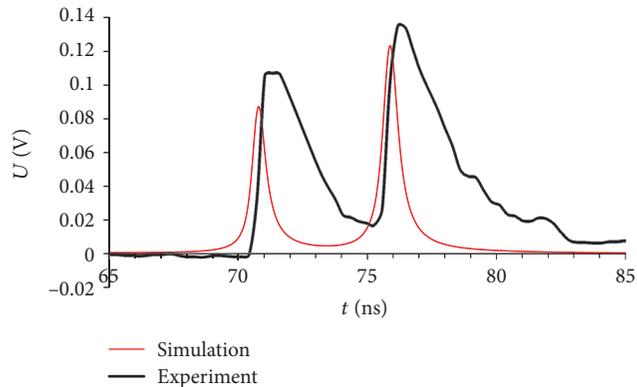


FIGURE 9: Comparison of simulated and experimental results of waveforms at the far end of the probing line (V_4) with open circuit at both ends of the probed conductor.

pulse is decomposed into two pulses because the difference of modal delays in line 1 is more than the probing pulse duration. When the break point is located at the distance of 1 m from the near end of the structure, there are three pulses at the far end of the probing line; the pulses are overlapped, so the middle pulse amplitude is rising. Thus, the information obtained from the waveform at the far end of the probing line allows us to determine a break of the passive wire. Reflections at the near end of the probing line are also informative because they allow us to locate the break.

Moreover, we obtained the experimental results confirming the diagnostics with the use of modal probing. To research the possibilities of the passive wire break diagnostics, we performed the experiment on two structures shown in

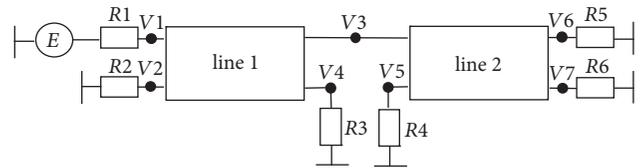


FIGURE 10: Schematic diagram of the structure under study with the break in the passive wire.

Figures 2(a) and 10. The electrical parameters of the considered cable (PUGNP 3×1.5) are presented in Table 6. The experimental conditions are thoroughly described in Section 4.

The oscillograms of the source signal and the signal at the far end of the probing line for the structure without a passive wire break (Figure 2(a)), when $R2 = R4 = 100\ \Omega$, are shown in Figure 12. The measured time delay of the fastest mode (considering the maximum permissible error of 7.5% with horizontal scale division 20 ns/div for S9–11 oscilloscope) is 72 ± 5.4 ns, which corresponds to the per-unit-length delay τ_o (Table 6) multiplied by the structure length (4.74 ns/m \times 15 m = 71.1 ns). The appearance of the second pulse is due to the presence of a passive (probed) wire, as well as to the fact that the total duration of the initial pulse (≈ 0.6 ns) is less than the full difference of the mode delay (0.32 ns/m \times 15 m = 4.8 ns), as follows from Table 6. Waveforms at the far end of the probing line under various boundary conditions at the ends of the passive wire are omitted here but were considered in more detail in [12].

The waveform at the far end of the probing line for the case with the passive wire break is shown in Figure 13(a). As can be seen, when passive wire is broken, four pulses come to the far end of the probing line instead of two pulses, as in the case without the break. The partial overlap of the pulses is due to dispersion. The reflected signal caused by the presence of the break in the passive conductor is observed too (it is circled in a white frame). The delay difference between transmitted and reflected signals is approximately 48 ± 3.6 ns (horizontal scale division 20 ns/div.) and corresponds to the per-unit-length delay τ_o (from Table 6) multiplied by twice the length of the line 1 (4.74 ns/m \times 2×5 m = 47.4 ns). Thus, the change in the number of pulses at the far end of the probing

TABLE 8: Pulse amplitudes for different boundary conditions at the ends of the probed wire for experiment and simulation.

Boundary conditions	Pulse 1, mV		Pulse 2, mV	
	Experiment	Simulation	Experiment	Simulation
Open circuit	112	85	133	121
Short circuit	160	131	93	82
100 Ω	148	107	124	101

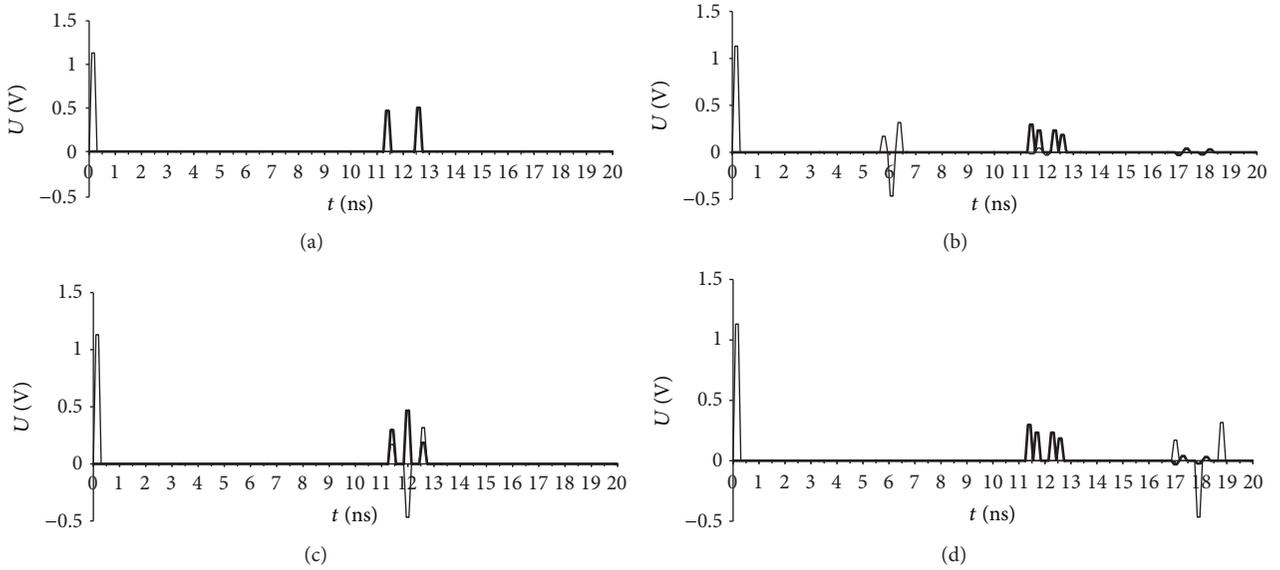
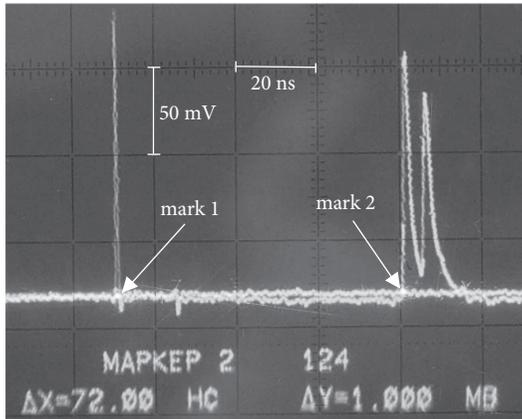


FIGURE 11: Simulated waveforms at the near (—) and far (---) ends of the probing line without the break of the passive wire (a) and with its break at distances of 0.5 (b), 1 (c), and 1.5 (d) m from the beginning.

FIGURE 12: Oscillograph pattern of a source signal (with 20 dB attenuator) and a signal at the far end of the probing line (V4) for the structure (without attenuator) from Figure 2(a) when $R_2 = R_4 = 100 \Omega$.

line experimentally confirms the possibility of contactless diagnostics of wire structures by modal probing, while the reflected signal may help to locate the wire break.

The case when there is no galvanic connection between the probing line and passive wire (i.e., open circuit condition holds at both ends of the passive wire) is of practical interest. The waveform at the far end of the probing line for this case

is shown in Figure 13(b). As in the case of 100Ω , four pulses are observed at the far end of the probing line. Thus, the presented waveforms indicate the possibility of wire break diagnostics by modal probing in the case when there is no galvanic connection either. It is noteworthy that the change in the boundary conditions at the ends of the passive wire leads to a change in the amplitude of pulses. The simulation results for this case ($R_2 = R_4 = \infty$) are shown in Figure 14. These results are consistent with the experiment and more clearly demonstrate that there are 4 pulses at the far end. Time delay between transmitted and reflected signals ($116 \text{ ns} - 69 \text{ ns} = 47 \text{ ns}$) is also consistent with the experiment (47.4 ns).

6. Analysis of the Modal Distortions in the Frequency Domain

Modal decomposition will occur only under the condition (1). However, when decomposed pulses are considerably overlapping, it becomes difficult to detect the decomposition fact when analyzing modal distortions in the time domain only. One of the approaches to solve this problem is the additional analysis in the frequency domain [16].

Frequency of the first minimum in a spectrum of a signal at the far end of an active conductor of a matched multiconductor transmission line is obtained with the following expression:

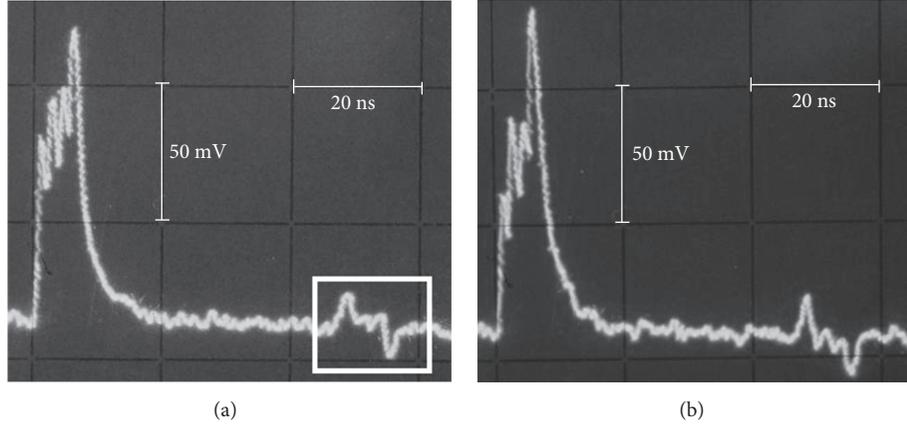


FIGURE 13: Waveform at the far end of the probing line (V_4) of the structure with a passive wire break when $R_2 = R_4 = 100 \Omega$ (a) and $R_2 = R_4 = \infty$ (b).

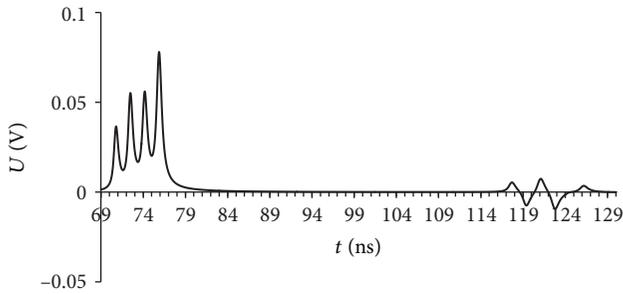


FIGURE 14: Simulated waveform at the far end of the probing line (V_4) of the structure with a passive wire break when $R_2 = R_4 = \infty$.

$$f_0 \approx \frac{1}{2l(\max(\tau_i) - \min(\tau_i))}, \quad i = 1, 2, \dots, N, \quad (2)$$

where l is length of a structure, τ_i is a per-unit-length delay for i -th mode of a structure, and N is the number of wires (excluding a reference one). For $N = 2$ (2) takes the form

$$f_0 = \frac{1}{2l|\tau_e - \tau_o|}, \quad (3)$$

where τ_e and τ_o are per-unit-length delays for even and odd modes. Then from (3) we obtain

$$l = \frac{1}{2f_0|\tau_e - \tau_o|}. \quad (4)$$

Now, using (4), we can determine the distance to the wire break point. It can be confirmed by the experimental results in a time domain described in detail above. The waveform and the spectrum of the signal at the far end of the probing line for the structure (of total length 15 m) without the passive wire break are shown in Figure 15, and with the break (at 5 m from the near end) in Figure 16. From Figure 15(a) it is possible to find the difference between the mode delays through the delay difference between the two pulse peaks ($20 \text{ ns/div.} \times 0.26 \text{ div.} = 5.2 \text{ ns}$). Then the per-unit-length delay difference would be 0.35 ns/m .

Let us refer to the data in the frequency domain. Taking the first frequency of the minimum value (83 MHz) measured by markers (Figure 15(b)) in the spectrum, we obtain (4)

$$l = \frac{1}{2 \cdot 83 \text{ MHz} \cdot 0.35 \text{ ns/m}} = 17.2 \text{ m}. \quad (5)$$

The difference between the calculated (17.2 m) and real (15 m) length values is $\pm 7\%$. Similarly, for the structure with a break (Figure 16(b)) we obtain (4)

$$l = \frac{1}{2 \cdot 126.9 \text{ MHz} \cdot 0.35 \text{ ns/m}} = 11.25 \text{ m}. \quad (6)$$

The difference between the calculated (11.25 m) and real (10 m) length values is $\pm 6\%$. For structures with a wire break, (4) gives the longest segment length. The difference between the real and calculated values is caused by a measurement error.

Consistency of experimental and simulated results is observed not only in time but also in the frequency domain. The simulation results of signal spectrum at the far end for the cases with and without a wire break are shown in Figure 17. There is a slight difference between the first resonance frequencies in comparison with the experiment (Figures 15 and 16). The values of these frequencies and their differences in the percentage terms are given in Table 9. The reasons of these differences are similar to the ones described in Section 4.

A similar experiment was conducted for the experimental PCB. However, the measurements were carried out with the scalar network analyzer R2M-40 (50Ω internal impedance, 40 GHz band-width). The experimental setup is shown in Figure 18. We measured S-parameters. The measured and simulated frequency dependencies of $|S_{21}|$ are shown in Figure 19. Taking the first frequency of the $|S_{12}|$ minimum (545 MHz) measured by R2M-40 (Figure 19), we obtain (4)

$$l = \frac{1}{2 \cdot 545 \text{ MHz} \cdot 2.2 \text{ ns/m}} = 0.41 \text{ m}. \quad (7)$$

The difference between the calculated (0.41 m) and real (0.33 m) length values is $\pm 11\%$.

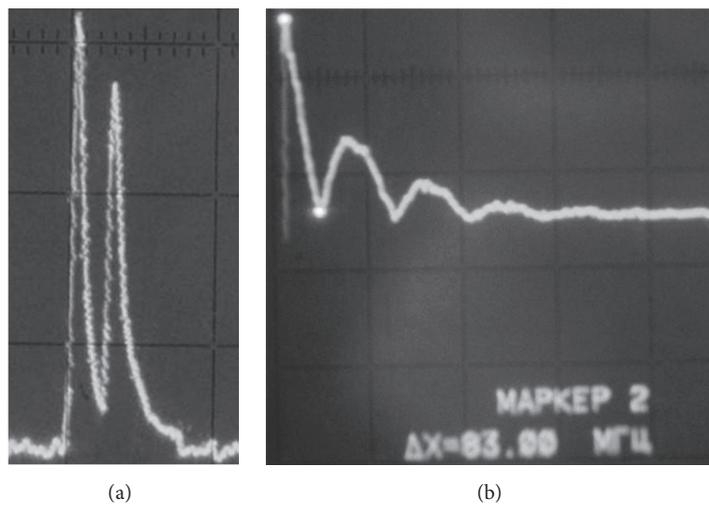


FIGURE 15: The waveform (a) and the spectrum (b) of the signal at the far end of a structure without a wire break.

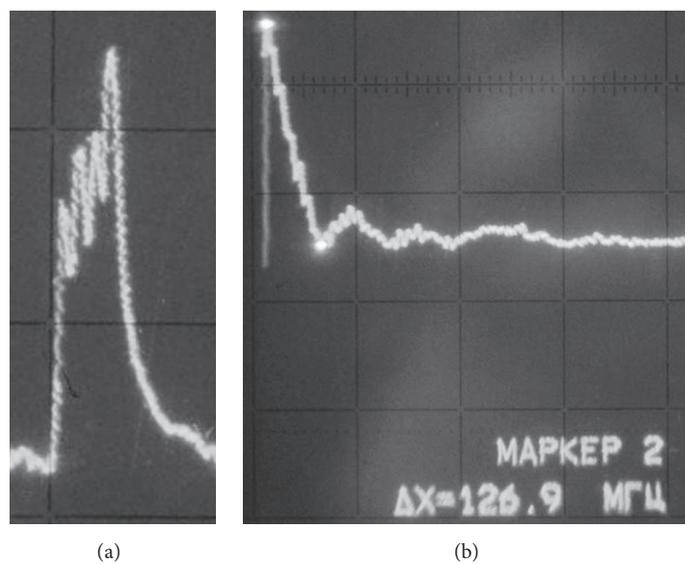


FIGURE 16: The waveform (a) and the spectrum (b) of the signal at the far end of a structure with a wire break.

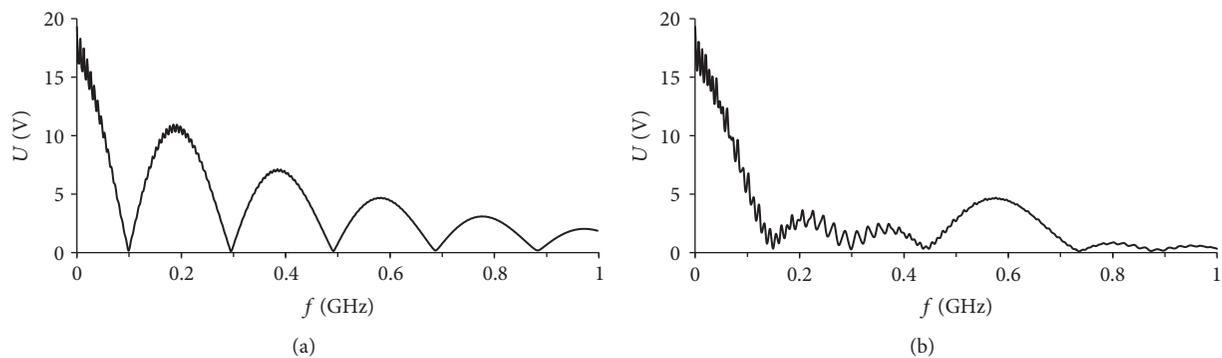


FIGURE 17: Simulated spectrum of the signal at the far end of a structure without (a) and with (b) wire break.

TABLE 9: Calculated and measured values of the first resonance frequency spectrum of the signal.

Case	First resonance frequency, MHz		$(f_S - f_E) / (f_S + f_E) \cdot 100\%$
	Simulation	Experiment	
Without a break	96.7	83	$\pm 7.6\%$
With a break	135	127	$\pm 3.05\%$

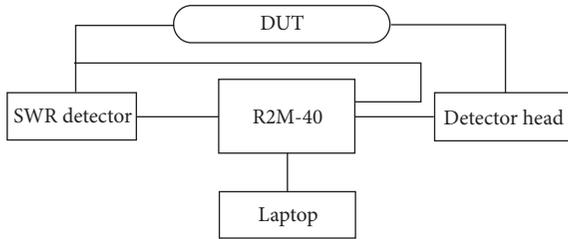
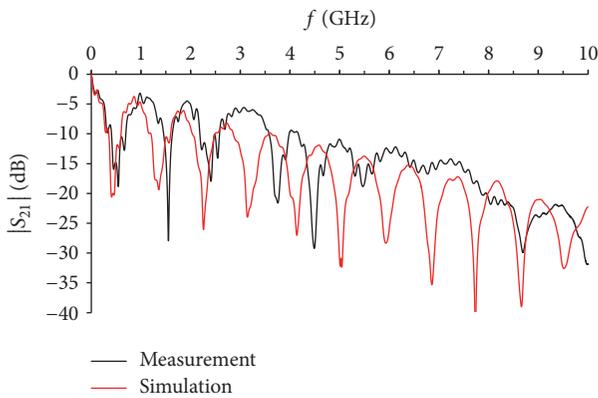


FIGURE 18: Schematic diagram of the experimental setup.

FIGURE 19: The measured and simulated frequency dependencies of $|S_{21}|$ for the experimental PCB.

Thus, the experimental and simulated results have allowed us to test the formula that relates the length of the longest segment of a structure with the difference of modal delays and the frequency of the first minimum in the spectrum of the signal at the far end of the structure. It can be used to diagnose electrical connections. Therefore, the results indicate the possibility of applying frequency analysis in modal probing.

7. Conclusion

In this paper, we described the principles of employing modal phenomena to detect, identify, and diagnose electrical connections. The proposed approach is based on a complex analysis of signal modal distortion and has two principal advantages: firstly, the galvanic connection to the probed line is not necessary and, secondly, the informativeness is inherent in other methods requiring the connection to the probed line (e.g., pulse reflectometry).

The results of the simulations and the experiments have confirmed the possibility of using modal probing. The first steps in the research of modal distortions in the frequency domain have yielded a formula that permits us to find the

length of the longest segment in the structure and the existence of a wire break in the structure, even when condition (1) is not fulfilled and there is no complete modal decomposition. However, it is required to pursue more in-depth research into applying modal phenomena for detecting, identifying, and diagnosing electrical connections by modal probing.

It should be noted that this paper does not fully describe the modal probing possibilities. Particularly, further development of the method requires a detailed research into modal distortion for structures with $N > 2$. Moreover, research into the effect of lumped elements along passive conductors on modal distortion is also high-potential. Simulation, design, and implementation of models and prototypes for various options of devices based on the modal probing are necessary to take its benefits.

Data Availability

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

Disclosure

All authors are with Department of Television and Control of Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia.

Conflicts of Interest

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

Acknowledgments

This research was supported by the Ministry of Education and Science of the Russian Federation (Project 8.9562.2017/8.9).

References

- [1] F. Auzanneau, "Wire troubleshooting and diagnosis: review and perspectives," *Progress in Electromagnetics Research B*, vol. 49, pp. 253–279, 2013.
- [2] M. W. Stavnes and A. N. Hammoud, "Assessment of Safety in Space Power Wiring Systems," *IEEE Aerospace and Electronic Systems Magazine*, vol. 9, no. 1, pp. 21–27, 1994.
- [3] A. Lelong, L. Sommervogel, N. Ravot, and M. O. Carrion, "Distributed reflectometry method for wire fault location using selective average," *IEEE Sensors Journal*, vol. 10, no. 2, pp. 300–310, 2010.
- [4] W. Ben Hassen, F. Auzanneau, F. Peres, and A. P. Tchangani, "Diagnosis sensor fusion for wire fault location in CAN bus

- systems,” in *Proceedings of the 12th IEEE Sensors Conference*, pp. 1–4, IEEE, Baltimore, MD, USA, November 2013.
- [5] S. Naik, C. M. Furse, and B. Farhang-Boroujeny, “Multicarrier reflectometry,” *IEEE Sensors Journal*, vol. 6, no. 3, pp. 812–818, 2006.
 - [6] M. K. Smail, L. Pichon, M. Olivas, F. Auzanneau, and M. Lambert, “Detection of defects in wiring networks using time domain reflectometry,” *IEEE Transactions on Magnetics*, vol. 46, no. 8, pp. 2998–3001, 2010.
 - [7] S. Schuet, D. Timuçin, and K. Wheeler, “A model-based probabilistic inversion framework for characterizing wire fault detection using TDR,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1654–1663, 2011.
 - [8] Q. Shi and O. Kanoun, “Automated wire fault location using impedance spectroscopy and genetic algorithm,” in *Proceedings of the 2012 IEEE Sensor Conference*, pp. 1–4, Taipei, Taiwan, October 2012.
 - [9] Q. Shi and O. Kanoun, “Automated wire fault location using impedance spectroscopy and Differential Evolution,” in *Proceedings of the 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC '13)*, pp. 359–364, IEEE, Minneapolis, MN, USA, May 2013.
 - [10] S. Wu, C. Furse, and C. Lo, “Noncontact probes for wire fault location with reflectometry,” *IEEE Sensors Journal*, vol. 6, no. 6, pp. 1716–1721, 2006.
 - [11] T. R. Gazizov, P. E. Orlov, A. M. Zabolotsky et al., “The device detection, identification and diagnosis of multiconductor transmission lines,” *RF Patent*, Article ID 2386964, 2009.
 - [12] P. E. Orlov and T. R. Gazizov, *New Approaches to Improve The Electrical Connections of Spaceborne Equipment*, p. 184, Tomsk, Russia, 2013.
 - [13] T. R. Gazizov, A. M. Zabolotsky, and I. E. Samotin, “Experimental results on ultra wide band pulse propagation in three-conductor power cables of flat and circular cross sections,” in *Proceedings of the International Siberian Conference on Control and Communications (SIBCON'09)*, pp. 264–269, IEEE, Tomsk, Russia, March 2009.
 - [14] T. R. Gazizov, A. M. Zabolotsky, A. O. Melkozerov, E. S. Dolganov, and P. E. Orlov, “Improved design of modal filter for electronics protection,” in *Proceedings of the 2012 International Conference on Lightning Protection (ICLP)*, pp. 1–4, Vienna, Austria, September 2012.
 - [15] P. E. Orlov, T. R. Gazizov, and A. M. Zabolotsky, “Experimental confirmation of the possibility for contactless diagnostics of multiconductor structures using modal probing,” *Russian Physics Journal*, vol. 56, no. 6, pp. 652–656, 2013.
 - [16] P. E. Orlov, T. R. Gazizov, and A. M. Zabolotsky, “Frequency analysis of modal distortions and its application to diagnostics of electric connections,” *Russian Physics Journal*, vol. 56, no. 9, pp. 1099–1101, 2014.
 - [17] A. M. Zabolotsky, T. R. Gazizov, A. G. Bova, and W. A. Radasky, “Dangerous pulse excitation of coupled lines,” in *Proceedings of the 17th International Zurich Symposium on Electromagnetic Compatibility*, pp. 164–167, IEEE, Singapore, February 2006.
 - [18] C. R. Paul, *Analysis of Multiconductor Transmission Lines*, p. 780, Wiley-IEEE Press, 2nd edition, 2007.
 - [19] S. Kuksenko, T. Gazizov, A. Zabolotsky et al., “New developments for improved simulation of interconnects based on method of moments/advances in intelligent systems research (ISSN 1951-6851),” in *Proceedings of the 2015 International Conference on Modeling, Simulation and Applied Mathematics (MSAM'15)*, pp. 293–301, Phuket, Thailand, August 2015.
 - [20] T. R. Gazizov, “Analytic expressions for MOM calculation of capacitance matrix of two dimensional system of conductors and dielectrics having arbitrarily oriented boundaries,” in *Proceedings of the 2001 International Symposium on Electromagnetic Compatibility (EMC 2001)*, pp. 151–155, IEEE, Montreal, Que., Canada, 2001.
 - [21] J. R. Griffith and M. S. Nakhla, “Time-Domain Analysis of Lossy Coupled Transmission Lines,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 38, no. 10, pp. 1480–1487, 1990.
 - [22] R. R. Gazizov, A. M. Zabolotsky, and T. R. Gazizov, “Ultrashort pulse maximum localization in multiconductor structures,” in *Proceedings of the 2016 Dynamics of Systems, Mechanisms and Machines, Dynamics 2016*, pp. 1–5, IEEE, Omsk, Russia, November 2016.
 - [23] R. R. Gazizov, A. O. Belousov, and T. R. Gazizov, “Influence of ultrashort pulse duration on localization of crosstalk peak values in PCB of spacecraft autonomous navigation system,” in *Proceedings of the 2017 International Siberian Conference on Control and Communications, SIBCON 2017*, pp. 1–6, IEEE, Astana, Kazakhstan, June 2017.
 - [24] R. R. Gazizov, R. S. Ryabov, and T. T. Gazizov, “Influence of crossover and mutation coefficients on GA optimization of ultrashort pulse duration by criteria of peak voltage maximization in PCB bus,” in *Proceedings of the 2017 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, pp. 415–420, Novosibirsk, September 2017.
 - [25] S. Park, F. Xiao, and Y. Kami, “Analytical approach for crosstalk characterization of multiconductor transmission lines using mode decomposition technique in the time domain,” *IEEE Transactions on Electromagnetic Compatibility*, vol. 52, no. 2, pp. 436–446, 2010.
 - [26] A. R. Djordjević, R. M. Biljić, V. D. Likar-Smiljanić, and T. K. Sarkar, “Wideband frequency-domain characterization of FR-4 and time-domain causality,” *IEEE Transactions on Electromagnetic Compatibility*, vol. 43, no. 4, pp. 662–667, 2001.

Research Article

A General Purpose Adaptive Fault Detection and Diagnosis Scheme for Information Systems with Superheterodyne Receivers

Dengwei Song,^{1,2} Hongmei Liu ,¹ Le Qi ,^{1,2} and Bo Zhou^{1,2}

¹School of Reliability and Systems Engineering, Beihang University, Xueyuan Road No. 37, Haidian District, Beijing, China

²Science & Technology on Reliability & Environmental Engineering Laboratory, Xueyuan Road No. 37, Haidian District, Beijing, China

Correspondence should be addressed to Hongmei Liu; liuhongmei@buaa.edu.cn

Received 3 February 2018; Accepted 14 March 2018; Published 24 April 2018

Academic Editor: Minvydas Ragulskis

Copyright © 2018 Dengwei Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A superheterodyne receiver is a type of device universally used in a variety of electronics and information systems. Fault detection and diagnosis for superheterodyne receivers are therefore of critical importance, especially in noise environments. A general purpose fault detection and diagnosis scheme based on observers and residual error analysis was proposed in this study. In the scheme, two generalized regression neural networks (GRNNs) are utilized for fault detection, with one as an observer and the other as an adaptive threshold generator; faults are detected by comparing the residual error and the threshold. Then, time and frequency domain features are extracted from the residual error for diagnosis. A probabilistic neural network (PNN) acts as a classifier to realize the fault diagnosis. Finally, to mimic electromagnetic environments with noise interference, simulation model under different fault conditions with noise interferences is established to test the effectiveness and robustness of the proposed fault detection and diagnosis scheme. Results of the simulation experiments proved that the presented method is effective and robust in simulated electromagnetic environments.

1. Introduction

With the advent of the era of big data, data transmission is playing an increasingly important role. As a significant form of data in the wireless transmission of electronic information system, analog signal is irreplaceable in many specific circumstances because of its high accuracy of rate and simplicity of signal processing.

In information system, the superheterodyne receiver is a typical kind of analog signal receiving apparatus widely used in radars and all kinds of signal receivers. The superheterodyne receiver has the capability of converting high-frequency signals. Given its high sensitivity, frequency stability, and simpler structure, the superheterodyne receiver has essentially replaced all previous receiver designs and become a standard configuration for virtually all modern radio receivers, taking the use of frequency measuring receivers in electronic intelligence reconnaissance as an example [1]. Particularly in the military field, the superheterodyne receiver plays rather a more important role.

Serious receiver performance degradation is probably caused by faults of components in receivers. Due to the significance of superheterodyne receivers, the performance of whole information system and the quality of communication will be critically affected once superheterodyne receivers failed. Hence effective fault detection and diagnosis methods for superheterodyne receivers are vital for the enhancement of performance and the mission success rate. Accurate fault detection techniques can help to improve the availability of different superheterodyne receivers and fault detection is essential for initiating maintenance action to prevent total failure of the system. Meanwhile, efficient diagnosis is instrumental to disassembly and replacement malfunction component, and then the efficiency of equipment maintenance is improved. Finally, through synthetically consideration information on detection and diagnosis, utilization of superheterodyne receivers and the support of spare parts are rationally planned to guarantee satisfactory performance of information system.

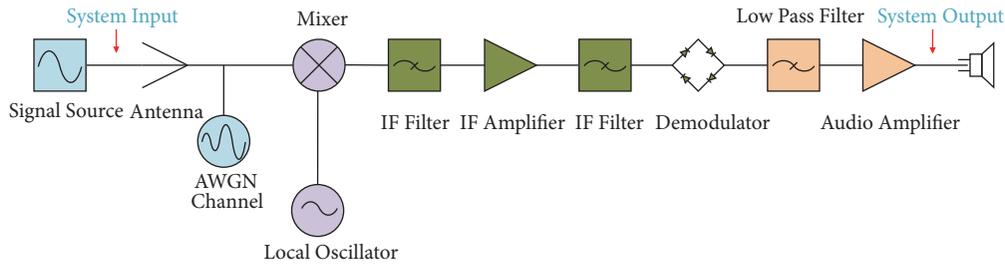


FIGURE 1: The block diagram of a superheterodyne receiver.

The implementation of fault detection and diagnosis in electromagnetic environment for superheterodyne receivers is a practical problem confronting electronic information equipment. However, there is little research known so far on fault detection and diagnosis for superheterodyne receivers under noise circumstance. Some relevant studies are as follows. In Binu and Kariyappa's survey, the approach of diagnosis of analog circuits was given and the machine learning-based approaches using artificial neural network as implementation were proposed [2]. Khanafseh and Pervan developed a general methodology to mitigate single receiver failures for architectures of carrier phase navigation [3]. Chen et al. proposed a fault diagnosis model for a radar receiver, built by applying multisignal flow graphs [4]. Mohsen and El-Yazeed invented a diagnosis approach of analog circuits based on dictionary and the excellent effectiveness of proposed algorithm was demonstrated by soft fault simulation [5]. Nho et al. designed an algorithm and an operation boundary for fault detection of an onboard GNSS receiver [6].

However, most of these studies are just suitable for a single kind of receiver or some particular receiver types and therefore not applicable to receivers on a broader sense. To overcome this disadvantage, this study discusses a novel fault detection and diagnosis method which can be applied for most commonly used superheterodyne receiver types.

The method was inspired by fault detection schemes for control systems. In their work [7], Jayakumar and Das proposed a technique of fault detection for a flight control system based on Luenberger observer. Keliris et al. developed a nonlinear observer-based approach for distributed fault detection of a class of interconnected input-output nonlinear systems [8]. Pröll et al. presented a fault isolation method combining structural diagnostic ability analysis with observer-based residual generation [9]. These studies indicate that observer-based methods tend to have excellent performance in the fault detection problem for complex systems. Therefore, in this study, an observer-based fault detection and diagnosis method using adaptive thresholds were adopted in the information system with superheterodyne receivers to explore the application of the technique in a new field.

Observer is established to obtain the estimated output, and then the difference between the actual and estimated output is taken as the residual error, which carries a great deal of information of system. Fault detection is carried out by comparing the residual error with the adaptive threshold. A type of neural network is effective to meet the nonlinear requirements of observer and adaptive threshold generator.

Considering the nonlinear fitting capability of generalized regression neural network (GRNN), it is employed in this paper. Moreover, GRNN requires shorter training time to find optimal solution than BPNN [10].

From the point of view of pattern recognition, the process of fault diagnosis of a superheterodyne receiver is pattern classification for its operating status. Feature extraction of the residual error is the key point of fault diagnosis which determines whether the fault diagnosis is successful. To further process the residual error signal, time domain and frequency domain analysis are applied to extract fault features and then a classifier diagnoses the fault. A proper classifier also has a better effect on the result of fault diagnosis. Since the probabilistic neural network (PNN) is a widely used algorithm for classification applicable to nonlinear problems and high dimension applications with short training time [11], the diagnosis is realized using PNN.

To solve the aforementioned problems, a method that combines observer, adaptive threshold generator, and classifier based on PNN is proposed in order to realize the accurate fault detection and effective diagnosis, which is beneficial to improve the operation performance of the information system with superheterodyne receivers. Moreover, the method proposed in this study is featured with stronger engineering applicability and generality. It is more practical in communication fields compared with previous approaches since the only monitoring data required by the observer are the system input/output, while, on the other hand, the GRNN, employed as the fault observer, is able to describe virtually all kinds of nonlinear systems. In general, the method is robust to various noises under electromagnetic environments with noise interference, due to its adaptive nature.

The structure of this paper is organized as follows. The simulation model of a typical superheterodyne receiver is given in Section 2. In Section 3, the fault detection and diagnosis scheme is elaborated. In Section 4, the validity of the proposed method is proved using simulation data seeded with faults and the results of the experiments were given subsequently, and Section 5 concludes the paper with some perspectives.

2. Modeling of a Typical Superheterodyne Receiver

Figure 1 is the block diagram of a typical superheterodyne receiver consisting of local oscillator, mixer, bandpass filter,

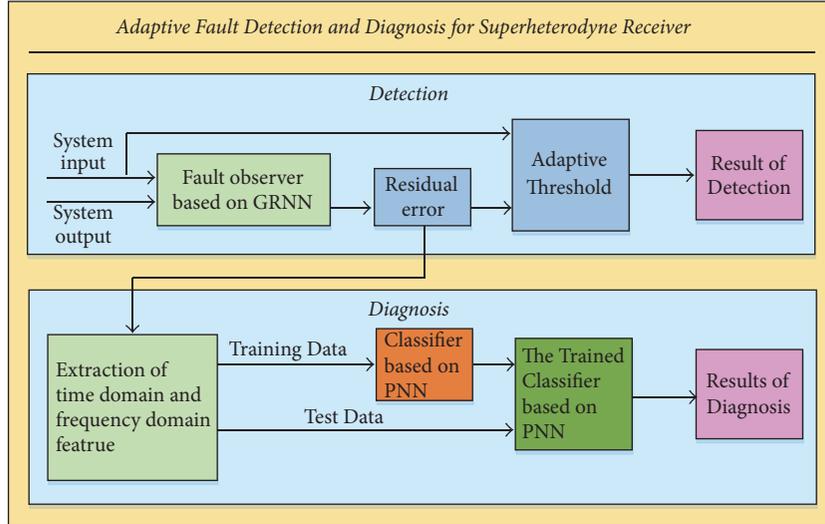


FIGURE 2: The adaptive fault detection and diagnosis for the superhetrodyne receiver.

amplifier, demodulator, and additive white Gaussian noise (AWGN) channel.

In signal source system, the module is packaged into a subsystem, including original signal, carrier, and attenuator, to simulate the reception signal. The antenna collects the radio signal. Then the radio signal is processed as follows. First, the received signal is fed into a mixer where it is mixed with a sine wave known as the local oscillation signal. A local oscillator provides the mixing frequency that is variable for tuning the receiver to different stations. The frequency mixer does the actual heterodyning, which changes the incoming radio frequency signal. Second, the signal obtained by mixing is processed by intermediate frequency (IF) band-pass filter and amplifier successively. The IF band-pass filter provides the narrowband filter and amplifier provides most of the gain for the radio signal. Then the other IF filter processes the signal again to further eliminate the band noise. Third, the envelope detector demodulates signals from the filter and provides an output which is the envelope of the original signal. Last, the extracted signal is sent into a low-pass filter. Signal from filter is the system output after being amplified by audio amplifier.

Based on the composition and the principle of the superhetrodyne receiver, the simulation model is established in Matlab/Simulink simulation environment. The details of the simulation model are as follows:

- In the signal source, the carrier frequency is 1000 kHz. The parameter of attenuator module is 0.1. The module simulates attenuation caused by the transmission distance from the transmitter to the receiver.
- In AWGN channel, the means of random white noise is 0, and the variance of random white noise is 0.001. The noise is used to simulate the interference of electromagnetic environment.
- The local oscillation signal is from the voltage-controlled oscillator controlled by the input voltage.

TABLE 1: Parameter of IF filter.

Parameter	Unit	Value
Filter order	Null	1
Center frequency	Hz	465k
Bandwidth	Hz	12k

Set up a slider gain module, and then the frequency of the local oscillator is controllable. The quiescent frequency of voltage-controlled oscillator is 465 kHz. The input sensitivity of voltage-controlled oscillator is 1000 Hz/V.

- Parameters of IF filters are shown in Table 1. The parameter of the IF amplifier is 20. The upper and lower limits of envelope detector are set to inf and 0. The parameter of the audio amplifier is 2. The bandwidth of the low pass filter is 6 kHz.

3. Fault Detection and Diagnosis for Superhetrodyne Receivers

Figure 2 shows the schematic diagram of the adaptive fault detection and diagnosis.

As is shown in the figure, in the phase of fault detection, an observer based on GRNN is fed with the system input and the system output. After training, the observer outputs an estimated value as a reference. Then residual error is obtained by comparing it with the actual output. Meanwhile, the other GRNN generates threshold that changes accordingly. The detection is finally realized by comparing the residual error and the adaptive threshold.

After fault detection, fault diagnosis is conducted based on the residual error generated by the fault observer. Three time domain parameters and three frequency domain parameters are extracted from the residual error signal. Finally, PNN is employed as the classifier to realize fault diagnosis.

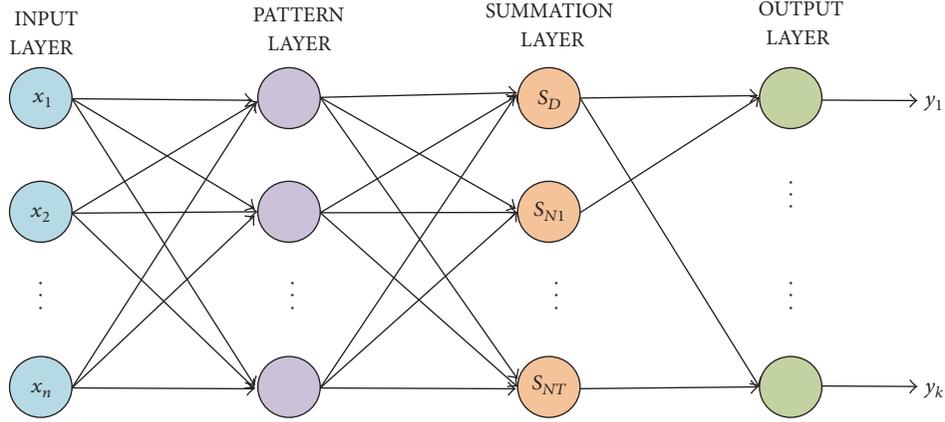


FIGURE 3: General configuration of GRNN.

3.1. GRNN Observer Based Adaptive Fault Detection

3.1.1. Generalized Regression Neural Network and Its Structure. Note that the fault detection is conducted in real-time electromagnetic environment where noise and interference will impact superheterodyne receivers' performance. GRNN is developed as an alternative to traditional neural network. As a feed-forward neural network based on nonlinear regression theory, GRNN has a simple and straightforward training algorithm which is significantly different from BPNN [12]. As shown in Figure 3, GRNN neural network includes four layers: an input layer, a pattern layer, a summation layer, and an output layer [13].

Assume that the network input is $X = [x_1, x_2, \dots, x_n]^T$, and the corresponding network output is $Y = [y_1, y_2, \dots, y_k]^T$. Each layer is presented in detail as follows.

(1) *Input Layer.* The number of neurons in the input layer is equal to the dimension of input feature vector. Each neuron is a simple distribution unit. The input of networks is transferred to the pattern layer directly.

(2) *Pattern Layer.* The number of neurons in the pattern layer is the total number of training samples n . Each neuron is related to different sample. The neural transfer function in the pattern layer is

$$P_i = \exp \left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2} \right] \quad i = 1, 2, \dots, n. \quad (1)$$

In (1), X is the system input. X_i is the learning sample corresponding to the i th neuron. Therefore, the output of neuron i is the squared Euclidean distance between the input and the corresponding training sample.

(3) *Summation Layer.* There are two summation types in the summation layer.

The function of the first type is

$$\sum_{i=1}^n \exp \left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2} \right]. \quad (2)$$

This function sums out the output of all pattern layer neurons. The transfer function is

$$S_D = \sum_{i=1}^n P_i. \quad (3)$$

The function of the second type is

$$\sum_{i=1}^n Y_i \left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2} \right]. \quad (4)$$

Outputs of all pattern layer neurons were weighted and summed. y_{ij} is the j th element of the i th output of training sample Y_i . The transfer function is

$$S_{Nj} = \sum_{i=1}^n y_{ij} P_i \quad j = 1, 2, \dots, k. \quad (5)$$

(4) *Output Layer.* The number of neurons in the output layer is equal to the dimension of output feature vector from training sample, which is k . The output of summation layer is divided in each neuron. The output of neuron j is the j th element of the network output:

$$y_j = \frac{S_{Nj}}{S_D} \quad j = 1, 2, \dots, k. \quad (6)$$

3.1.2. Design of the Fault Observer and Adaptive Threshold Generator. Design of a fault observer and an adaptive threshold generator using the method described in [14] is shown in Figure 4. Two GRNNs are utilized in this method. The first GRNN acted as an observer, which outputs the estimated system output. Then define the residual error by obtaining the deviation value between the estimated output and actual output. The other trained neural network is used as the adaptive threshold generator. The output datasets are composed of adaptive threshold values. Result of comparison between residual error and adaptive threshold can be applied to judge whether the system has fault or not.

In the training process, the system input and output are sent to the GRNN to obtain the observer. The residual error

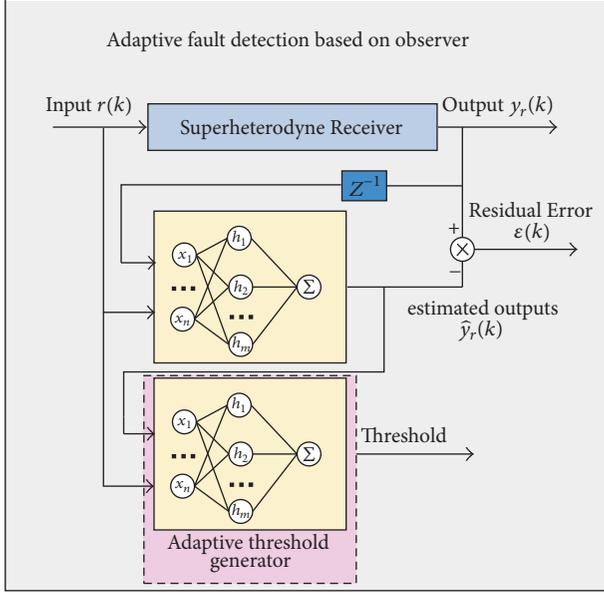


FIGURE 4: Fault detection based on GRNN observer.

is close to 0 when system works normally, while residual error increased when system is abnormal.

Aim to construct the relationship between the system input, output, and threshold. The training input datasets of the second neural network consist of the system input and output, and the training output data are the expected adaptive threshold calculated using (7).

$$\text{threshold} = \varepsilon_0(k) + \beta. \quad (7)$$

In (7), $\varepsilon_0(k)$ is the baseline residual error, which is acquired from observer under normal circumstances, and β is the correction coefficient.

The observer and adaptive threshold generator can be employed to realize the fault detection after training. The residual error is below the threshold in normal status. While the system is in faulty status, the residual error exceeds the threshold.

3.2. Residual Error Analysis Based Fault Diagnosis. Residual error is the difference between the actual output and the estimated output; therefore the residual error carries large amount of fault information. Fault diagnosis is developed based on residual error.

3.2.1. Feature Extraction in Time and Frequency Domain. Feature extraction is a key issue in fault diagnosis. Properly extracted features can improve both diagnosis speed and accuracy.

The time domain analysis can characterize the energy and stability of signal over time. The frequency analysis can decompose the signal in detail. Considering the simplicity of time domain analysis and the adaptive ability of frequencydomain analysis, in this study, the feature of residual error is extracted by combining the time domain and

frequency domain. Thereby the performance of system will be characterized as more comprehensive.

Three time domain parameters of signal, including peak value, root mean square, and average absolute value, are extracted. Suppose that a set of discrete data signal obtained by sampling is x_1, x_2, \dots, x_n ; equations of these parameters are as follows.

Peak value is

$$\alpha = X_{\max} = \max \{|x_i|\}. \quad (8)$$

Root mean square is

$$\beta = X_{\text{rms}} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}. \quad (9)$$

Average absolute value is

$$\gamma = |\bar{X}| = \frac{1}{N} \sum_{i=1}^N |x_i|. \quad (10)$$

Wavelet transformation (WT) and singular value decomposition (SVD) are general methods to extract the frequency domain feature. WT is an effective way to deal with signal. It can project a signal from time domain space to several frequency ranges, and the features of the dataset are implied in these frequency ranges. Signals are decomposed into a feature matrix and different frequency ranges can be obtained by using the WT. And the features usually appear in different frequency ranges [15]. SVD can be used to extract the prominent feature from all the frequency ranges. It can decompose a matrix into simple and meaningful pieces, which may contribute to the subsequent analysis. Matrix A decomposed by SVD can be expressed as follows:

$$A = USV^T, \quad (11)$$

where U represents the left singular vector, V denotes the right singular vector, and S indicates diagonal matrix whose elements in the main diagonal are singular values.

The samples belonging to the same category will have the similar singular values and samples with different categories decomposed by SVD will obtain diverse singular values [16]. Hence, the singular values can be utilized to represent a signal.

After the analysis in time and frequency domain, the eigenvector of superheterodyne receivers is composed of time domain and frequency domain features.

3.2.2. PNN Based Fault Diagnosis. In this study, PNN is employed as classifier for fault diagnosis. As a radial basis function neural network, a PNN neural network is a powerful algorithm for classification with its theory based on Bayes minimum risk criteria (Bayesian decision theory) [17]. Due to its concise training and strong classification ability, the PNN has widely applications for fault diagnosis in practical applications [18]. Compared with BP neural network, the advantage of a PNN is that the topology, connection weights,

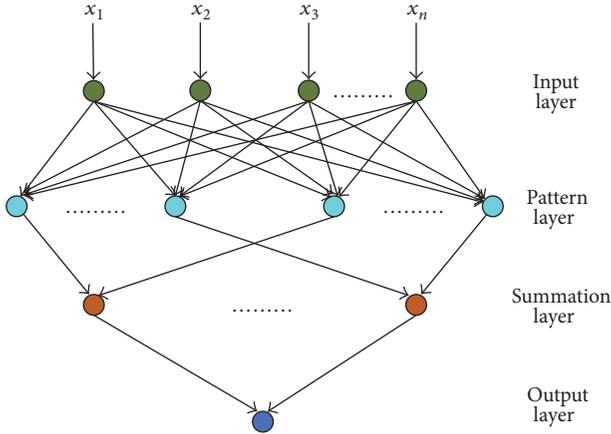


FIGURE 5: General configuration of PNN.

TABLE 2: Injected fault modes in simulation model.

Number	Fault	Details
1	Amplifier fault	Amplifier gain reduction fault
2	Oscillator fault	Local oscillator bias fault
3	Filter fault	Filter bandpass reduction fault

and thresholds can be set immediately when training samples are attainable [19].

The general structure of a PNN model is shown in Figure 5. The model contains 4 layers: input layer, the pattern layer, the summation layer, and the output layer [20].

The input layer introduced the eigenvector into network. Pattern layer calculates the distance between the unknown input and the training sample. Then distance transformed by activation function, which is Gaussian function, is the output of pattern layer. The number of neurons in summation layer is equal to the number of failure modes. Each neuron sums the probability of one kind of fault. According to the estimated probability, the output layer chooses a neuron in summation layer with the highest probability as the output of the neuro network.

In the process of diagnosis, the training data is used to train the PNN; the weights of neuro network and other parameters are created in the learning stage. In the recalling stage, the PNN is utilized as a classifier to identify the system status.

4. Case Study

4.1. Simulation Parameters and Fault Injection. A sine signal is used as the input signal in the simulation. The amplitude is 1 and the system input frequency is 100 Hz.

According to the statistics of historical maintenance data, main fault types of a superheterodyne receiver include amplifier fault, oscillator fault, and filter fault, which are fed into the simulation model to test the effectiveness of the proposed approach.

The fault modes listed in Table 2 were introduced into the simulation model by changing several specific parameters of

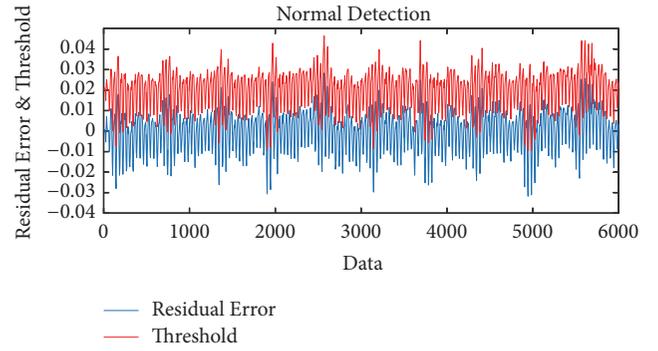


FIGURE 6: Detection result under normal status.

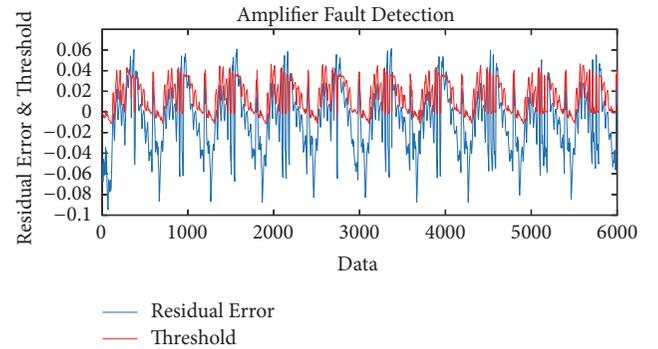


FIGURE 7: Detection result for amplifier fault.

the fault components. The details of fault injection are listed in Table 3.

The simulation duration is 0.01 s, and the sampling rate is 600 K/s. Data of normal status and three kinds of fault status were collected. In each case there are 6000 data points of system input and 6000 data points of system output.

4.2. Fault Detection and Results Analysis. For fault detection, normal data of system input and output are used to train the observer based on the first GRNN. The system input and the system output of the previous step are taken as the input sample of training data. The output sample of training data is the actual system output. The function spread of the first GRNN neural network is set to 0.01.

Normal data are used to train the second GRNN to generate the adaptive threshold as well. The system input and the estimated system output are taken as the input sample of training data. The output sample of training data is the expected adaptive threshold which is the summation of the baseline residual and β . Through multiple tests combined experience, the value of β is set to 0.1. The function spread of the second GRNN neural network is set to 0.01.

The structure and parameters of the GRNN neural networks are determined after training.

Four tests were carried out to validate the effectiveness of the fault detection scheme. The results are shown in Figures 6–9. In each figure, the blue curve shows the residual error, and the red curve is the adaptive threshold.

TABLE 3: Fault injection details.

Test number	Fault mode	Fault component	Changed parameter for fault injection (unit)	Parameter (normal)	Parameter (fault)
1	Normal	-	-	-	-
2	Amplifier fault	Electronic amplifier	Gain	20	17
3	Oscillator fault	Local oscillator	Quiescent frequency (kHz)	465	480
4	Filter fault	Intermediate frequency filter	Bandwidth (kHz)	12	1

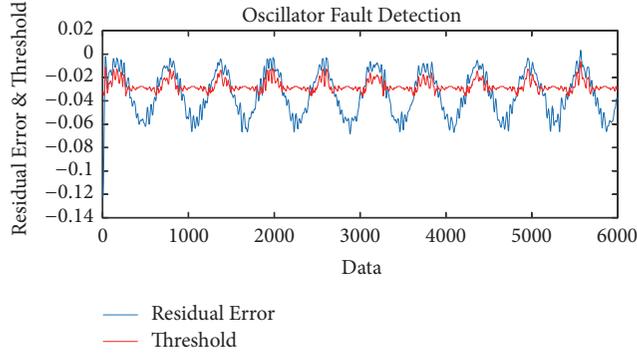


FIGURE 8: Detection result for oscillator fault.

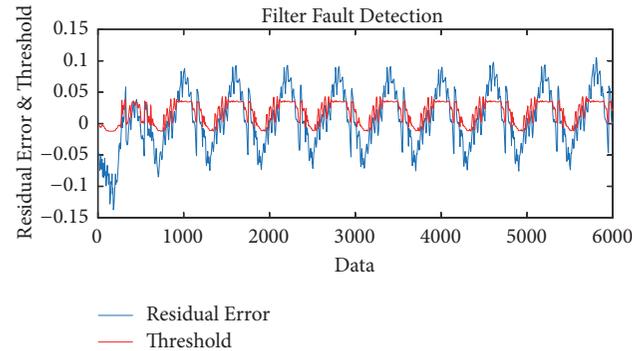


FIGURE 9: Detection result for filter fault.

In test 1, the superheterodyne receiver works normally. The result of detection is shown in Figure 6. The residual error was close to zero and below the threshold; therefore the results indicate that the system is under normal circumstances.

In test 2, an electronic amplifier fault was injected to the system. The result of detection is shown in Figure 7. The residual error became greater and exceeded the threshold; therefore, the results indicate that fault occurred in the system.

In test 3, an oscillator fault was injected to the system. The result of detection is shown in Figure 8. The residual error is beyond the threshold; therefore, the results indicate that fault occurred in the system.

In test 4, a filter fault was injected to the system. The result of detection is shown in Figure 9. The residual error increased significantly and exceeded the threshold; therefore, the results indicate that fault occurred in the system.

TABLE 4: Training eigenvectors for fault diagnosis.

Pattern	Sample size	Label data
Normal	210 eigenvectors	1
Filter fault	210 eigenvectors	2
Amplifier fault	210 eigenvectors	3
Oscillator fault	210 eigenvectors	4

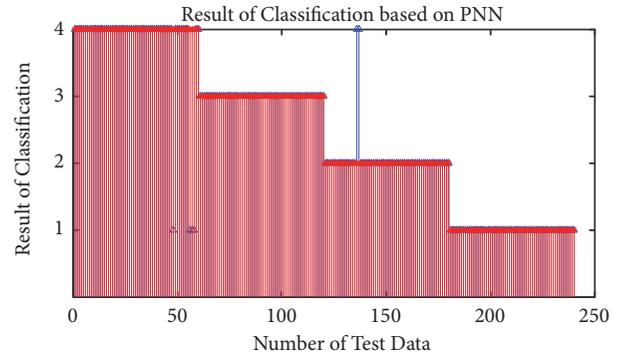


FIGURE 10: Result of Classification based on PNN.

As indicated by the simulation results, the residual error exceeded the threshold when the system worked abnormally under the interference of environment noise factors.

4.3. Fault Diagnosis and Results Analysis. For fault diagnosis, the time domain features, including the average absolute value, the root mean square, the peak value, and three frequency domain features processed by WT-SVD, are obtained from residual error. Data were preprocessed and one eigenvector including 6 features was extracted from each of 600 samples of residual error signals. 270 eigenvectors were extracted from each residual error signal and totally 1080 eigenvectors were obtained.

To train the PNN, 840 eigenvectors of each fault mode were selected to form the training data, and their labels were set as shown in Table 4. The function spread of PNN was set to 0.075.

After the training of PNN, the other 240 eigenvectors were employed as the test data to verify the effectiveness of the proposed method. Figure 10 shows the results of fault diagnosis. In Figure 10, the red points are the actual labels of the test data, while the blue points are the predicted labels by PNN.

Number 1 to number 60 are the normal data. Number 61 to number 120 are the filter fault data. Number 121 to number 180 are the amplifier fault data. Number 181 to number 240 are

the oscillator fault data. In the test of normal data, the PNN classification result of 4 test samples did not match the actual label. In the test of amplifier fault data, the PNN classification result of all test samples matched the actual label. In the test of oscillator fault data, the PNN classification result of 2 test samples did not match the actual label. In the test of filter fault data, the PNN classification result of all test samples matched the actual label.

In total, there are 6 test samples of the misjudgment of the fault diagnosis out of 240 test samples, and the corresponding diagnosis accuracy was 97.5%.

5. Conclusion

In this paper, a general purpose fault detection and diagnosis method was proposed for information systems with superheterodyne receivers. By establishing the observer and adaptive threshold generator based on GRNNs, the faults of superheterodyne receiver are detected by the variation of residual error. The residual error is then used to extract fault features and PNN is employed as the fault classifier to diagnose the fault type. The proposed method was verified by simulation experiments considering noise interferences. The result indicates that the proposed method can effectively detect superheterodyne receiver faults and diagnose fault modes accurately. Meanwhile, the adaptive threshold has strong adaptive ability for noise interferences.

The advantage of the proposed approach lies in that it is applicable not only to the superheterodyne receiver but also to more complex signal receiving systems of similar designs in which the transfer functions are difficult to obtain. It also outweighs other previous methods due to its strong robustness to environmental noises and interferences and enables a higher mission success rate.

However, future work needs to be done to make the proposed method more practical. First, the data from a lab testing of actual superheterodyne receiving system will be applied instead of simulation model to further verify the effectiveness of the proposed method. Second, the classifier needs to be developed to realize the diagnosis in the situation with multiple failures.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China [Grants nos. 51605014, 51105019, and 51575021], the Technology Foundation Program of National Defense [Grant no. Z132013B002], and the Fundamental Research Funds for the Central Universities [Grant no. YWF-16-BJ-J-18].

References

- [1] Y. Qiang, B. Daping, H. Bing, and X. Lianghao, "Jamming technology study of superheterodyne receiver," *Fire Control & Command Control*, vol. 10, pp. 56–59, 2015.
- [2] D. Binu and B. S. Kariyappa, "A survey on fault diagnosis of analog circuits: Taxonomy and state of the art," *AEÜ - International Journal of Electronics and Communications*, vol. 73, pp. 68–83, 2017.
- [3] S. Khanafseh and B. Pervan, "Detection and mitigation of reference receiver faults in differential carrier phase navigation systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2391–2404, 2011.
- [4] S.-J. Chen, K. Lian, and H.-J. Wang, "Fault diagnosis method of radar receiver using multi-signal flow graphs model," *Dianzi Keji Daxue Xuebao/Journal of the University of Electronic Science and Technology of China*, vol. 38, no. 1, pp. 87–91, 2009.
- [5] A. A. K. Mohsen and M. F. A. El-Yazeed, "Selection of input stimulus for fault diagnosis of analog circuits using ARMA model," *AEÜ - International Journal of Electronics and Communications*, vol. 58, no. 3, pp. 212–217, 2004.
- [6] H. Nho, J. Ahn, S. Sung, H. Jun, C. Yeom, and Y. Lee, "Research on Algorithm and Operation Boundary for Fault Detection of Onboard GNSS Receiver," *Journal of the Korean Society for Aeronautical & Space Sciences*, vol. 40, no. 2, pp. 171–177, 2012.
- [7] M. Jayakumar and B. B. Das, "Fault detection, isolation and reconfiguration in presence of incipient sensor faults in an electromechanical flight control actuation system," in *Proceedings of the 2006 IEEE International Conference on Industrial Technology, ICIT*, pp. 92–97, India, December 2006.
- [8] C. Keliris, M. M. Polycarpou, and T. Parisini, "A robust nonlinear observer-based approach for distributed fault detection of input-output interconnected systems," *Automatica*, vol. 53, pp. 408–415, 2015.
- [9] S. Pröll, F. Jarmolowitz, and J. Lunze, "A comprehensive observer-based fault isolation method with application to a hydraulic power train," *IFAC-PapersOnLine*, vol. 49, no. 11, pp. 547–554, 2016.
- [10] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 2, no. 6, pp. 568–576, 1991.
- [11] D. F. Specht, "Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques for Classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 1, no. 1, pp. 111–121, 1990.
- [12] R. Rooki, "Application of general regression neural network (GRNN) for indirect measuring pressure loss of Herschel-Bulkley drilling fluids in oil drilling," *Measurement*, vol. 85, pp. 184–191, 2016.
- [13] H. Bendu, B. B. V. L. Deepak, and S. Murugan, "Application of GRNN for the prediction of performance and exhaust emissions in HCCI engine using ethanol," *Energy Conversion and Management*, vol. 122, pp. 165–173, 2016.
- [14] H. Yuan, C. Lu, J. Ma, and Z.-H. Chen, "Neural network-based fault detection method for aileron actuator," *Applied Mathematical Modelling*, vol. 39, no. 19, pp. 5803–5815, 2015.
- [15] O. Daoud, Q. J. Hamarsheh, and S. Saraireh, "PAPR reduction based on entropy wavelet transform for Sniffer Mobile Robot," *AEÜ - International Journal of Electronics and Communications*, vol. 68, no. 12, pp. 1165–1172, 2014.
- [16] Q. Zhang, Y. Wang, M. D. Levine, X. Yuan, and L. Wang, "Multisensor video fusion based on higher order singular value decomposition," *Information Fusion*, vol. 24, pp. 54–71, 2015.
- [17] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.

- [18] H. Malik and S. Mishra, "Application of Probabilistic Neural Network in Fault Diagnosis of Wind Turbine Using FAST, TurbSim and Simulink," in *Proceedings of the 2nd International Symposium on Computer Vision and the Internet, VisionNet '15*, pp. 186–193, India, August 2015.
- [19] S. R. Mohanty, P. K. Ray, N. Kishor, and B. K. Panigrahi, "Classification of disturbances in hybrid DG system using modular PNN and SVM," *International Journal of Electrical Power & Energy Systems*, vol. 44, no. 1, pp. 764–777, 2013.
- [20] J. Jing, H. Liu, and C. Lu, "Fault diagnosis of electro-mechanical actuator based on WPD-STFT time-frequency entropy and PNN," in *Proceedings of the 28th International Conference on Vibroengineering*, pp. 130–135, China, October 2017.

Research Article

Statistical Identification of Parameters for Damaged FGM Structures with Material Uncertainties in Thermal Environment

Yalan Xu ¹, Yu Qian,² and Kongming Guo¹

¹School of Mechano-Electronic Engineering, Xidian University, Xi'an 710071, China

²School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Yalan Xu; ylxu@mail.xidian.edu.cn

Received 12 December 2017; Accepted 20 February 2018; Published 29 March 2018

Academic Editor: Minvydas Ragulskis

Copyright © 2018 Yalan Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Considering that the statistic numerical characteristics are often required in the probability-based damage identification and safety assessment of functionally graded material (FGM) structures, an stochastic model updating-based inverse computational method to identify the second-order statistics (means and variances) of material properties as well as distribution of constituents for damaged FGM structures with material uncertainties is presented by using measurable modal parameters of structures. The region truncation-based optimization method is employed to simplify the computational process in stochastic model updating. In order to implement the forward propagation of uncertainties required in the stochastic model updating and avoid large error resulting in the nonconvergence of the iteration process, an algorithm is proposed to compute the covariance between the modal parameters and the identified parameters for damaged FGM structures. The proposed method is illustrated by a numerically simulated damaged FGM beam with continuous spatial variation of material properties and verified by comparing with the Monte-Carlo simulation (MCS) method. The influences of the levels and sources of measured data uncertainties as well as the boundary conditions on the identification results are investigated. The numerical simulation results show the efficiency and effectiveness of the presented method for the identification of material parameter variability by using the measurable modal parameters of damaged FGM structures.

1. Introduction

Functionally graded materials (FGMs) have found wide application in modern industries including aerospace engineering, military application, and mechanical engineering, due to resistant high temperature gradient and strong mechanical performance [1], and a great deal of research has already been done into both deterministic [2–6] and stochastic [7–12] mechanical behavior of FGM structures. However, damage such as crack and fracture often occurs in FGM structures as most of FGM structures work in harsh environment such as high temperature gradient and corrosion. Damage problems in FGM structures have been widely investigated in literatures [13–19].

As is well-known, accurate and reliable mechanical behavior analysis of FGM structures with initial damage is vital for the early condition assessment and damage prognosis in order to guarantee the safe performance and prevent the possible disastrous failure of FGM structures [20, 21].

What is more, the reliable analysis of overall mechanical behavior of any FGM structure, undamaged or damaged, relies on a precise knowledge of the material properties and, especially, the distribution of constituents. Traditionally, mechanical properties and volume fraction distribution of heterogeneous materials can be determined by starting from indentation tests [22, 23]. Owing to the inhomogeneous nature of FGMs, however, experimental characterization of materials constants and volume distribution is cumbersome and time-consuming since a large number of property parameters need to be determined, and nondestructive techniques have been developed to evaluate the material property parameters and the volume fraction index for FGMs by utilizing complex relationships between the structural behavior and the material properties. Liu et al. [24] suggested a progressive neural network process for characterizing the material properties of functionally graded material (FGM) plate by using elastic wave. Han and Liu [25] presented an inverse method for determining the material properties and

distribution in the thickness direction of FGM plates by using uniform crossover microgenetic algorithm. Rahmani et al. [26] presented a regularized finite element model updating for identification of elastic constitutive parameters identification of 2D composites from full-field measured displacement data, in which mechanical constraints are used as regularization factors in the optimization algorithm. Sun et al. [27] developed a strategy to identify the temperature-dependent properties of a thermoelastic structure in thermal environment taking into time-varying material properties and thermal stresses account. Mishra and Chakraborty [28, 29] dealt with a modal analysis based inverse identification of material properties of fiber reinforced plastics composite plates with rotational flexibility at boundaries and panels having elastically restrained boundary from the experimental modal testing using finite element model updating.

All of the above studies focus on the identification of deterministic quantities of material constants or volume distribution of composite materials and structures. In reality, however, both constituent material properties and volume distribution in FGMs present inherent fluctuation due to the typicality and technical variety in the manufacturing and fabrication of FGMs and environmental temperature [30, 31]. From this point of view, therefore, the inherent uncertainties need to be incorporated in parameter identification, which is an important issue in mechanical behavior analysis and safety evaluation of FGM structures, especially in the early stage of damage.

In the present work, considering that the statistic numerical characteristics, for example, second-order statistics (means and variances) of material properties, are often required in evaluating the mechanical behavior and safety of FGM structures, an inverse computational procedure is presented to identify the second-order statistics (means and variances) of material properties and volume fraction index of FGM structures with initial damage by using stochastic finite element model updating [32], which is implemented by minimizing the differences between the analytical and actual structural modal parameters [33] easily obtained by measuring the structural response signal. It is worth noting that, as an inverse problem, both the efficient optimization method with good convergence and the forward uncertainty propagation with relatively small errors are two key issues for the statistical identification of FGM parameters based on stochastic model updating. Instead of utilizing the trust region method with expensive computation, the region truncation-based optimization method has been proposed to simplify the computational process without harming the convergence of the iteration process in this work. On the other hand, the forward uncertainty propagation needs to be carried out in every iteration step, in which not only the means and variances of the dynamic characteristics [7–12] but also the covariance between the dynamic characteristics and the identification parameters needs to be computed, and the computation of first-order derivatives is involved. Finite difference approximation is traditionally used for obtaining the involved derivatives for structures with a large number of degrees, and, however, it is obviously not suitable to be utilized in computing the derivatives with respect to

volume fraction index for FGM structures due to large errors which will accumulate in every iteration step and lead to the nonconvergences of the results, since the effective material properties of FGMs are assumed as power functions of volume fraction index. On account of the fact mentioned above, an algorithm is developed for computing the first derivative of dynamic characteristics with respect to random variables, which is employed to compute the covariance between the dynamic characteristics and the identification parameters. In addition, in consideration of the difficult of random experiment test for the damaged FGM structure with enough samples required for the random analysis, the actual modal data will be obtained by numerical simulation method in this study.

2. Volume Distribution of Material Constituents in FGM Beam

An FGM rectangular beam composed of ceramics (top surface/left surface) and metal (bottom surface/right surface) having the length l , width b , and thickness h is considered in this work. The effective material properties of the beam are assumed to vary continuously through its thickness direction or along its axial direction complying with power law distribution [11] and can be, respectively, expressed as

$$P(z) = \left[(P_c - P_m) \left(\frac{z}{h} + \frac{1}{2} \right)^n + P_m \right],$$

$$-\frac{h}{2} \leq z \leq \frac{h}{2}, \quad 0 \leq n < \infty, \quad (1)$$

$$P(x) = \left[(P_c - P_m) \left(1 - \frac{x}{l} \right)^n + P_m \right],$$

$$0 \leq x \leq l, \quad 0 \leq n < \infty,$$

where z , x are the coordinates along the thickness (h) and length (l) of the FGM beam as shown in Figure 1. $P(\bullet)$ denotes the effective material properties such as Young's modulus, Poisson's ratio, and thermal expansion efficient along z coordinate for the FGM beam through its thickness direction or x coordinate for the FGM beam along its axial direction. P_c , P_m denote the corresponding material properties of ceramics and metal, respectively. n denotes the volume fraction index.

Each material property of ceramics and metal can be expressed as a function of temperature [30]:

$$P = P_0 (P_{-1} T^{-1} + 1 + P_1 T + P_2 T^2 + P_3 T^3), \quad (2)$$

where $P_0, P_{-1}, P_1, P_2, P_3$ are the coefficients of temperature T and change with the constituent materials as well as the temperature.

3. Stochastic Model Updating

3.1. Updating the Means and Variances of Material Parameters. The model updating can often be posed as a minimization problem of the objective function which is a sum of square difference between the analytic and actual data,

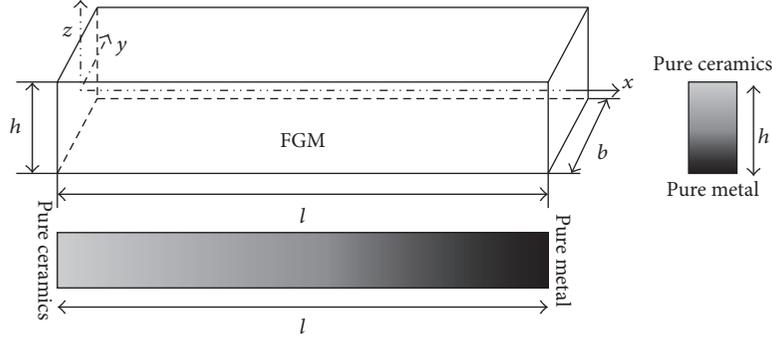


FIGURE 1: Configuration of an FGM beam.

typically dynamic characteristics (modal frequencies and mode shapes):

$$\min_{\theta} f(\theta) = \|\mathbf{z}_{ak}(\theta) - \mathbf{z}_m\|^2, \quad \underline{\theta} \leq \theta \leq \bar{\theta}, \quad (3)$$

where the updating parameter vector θ is composed of material properties and volume fraction index of FGMs to be identified, that is, $\theta = [\mathbf{P}_c^T \ \mathbf{P}_m^T \ n]^T$. The vector \mathbf{z}_{ak} is the analytic modal parameters (modal frequencies and mode shapes), that is, $\mathbf{z}_{ak} = [\mathbf{f}_{ak}^T \ \{\boldsymbol{\varphi}_{ak}^T\}]^T$, which is a nonlinear function of the updating parameters vector θ . The vector $\mathbf{z}_m = [\mathbf{f}_m^T \ \{\boldsymbol{\varphi}_m^T\}]^T$ is the actual modal parameters. The vectors $\underline{\theta}$, $\bar{\theta}$ denote the upper and lower bounds on the updating parameters.

In the stochastic model updating, the parameters can be updated by the following iterative expression [32]:

$$\begin{aligned} \mathbf{z}_{ak} &= \bar{\mathbf{z}}_{ak} + \tilde{\mathbf{z}}_{ak}, \\ \mathbf{z}_m &= \bar{\mathbf{z}}_m + \tilde{\mathbf{z}}_m, \end{aligned} \quad (4)$$

$$\boldsymbol{\theta}_k = \bar{\boldsymbol{\theta}}_k + \tilde{\boldsymbol{\theta}}_k,$$

$$\bar{\boldsymbol{\theta}}_{k+1} = \bar{\boldsymbol{\theta}}_k + \bar{\mathbf{T}}_k (\bar{\mathbf{z}}_{ak} - \bar{\mathbf{z}}_m),$$

$$\bar{\mathbf{T}}_k = \left[\bar{\mathbf{S}}_k^T \mathbf{W}_\varepsilon \bar{\mathbf{S}}_k + \mathbf{W}_\vartheta \right]^{-1} \bar{\mathbf{S}}_k^T, \quad \bar{\mathbf{S}}_k = \left[\frac{\partial \mathbf{z}_{ak}(\theta)}{\partial \theta} \right]_{\theta = \bar{\boldsymbol{\theta}}_k}, \quad (5)$$

$$\begin{aligned} \text{Cov}(\tilde{\boldsymbol{\theta}}_{k+1}, \tilde{\boldsymbol{\theta}}_{k+1}) &= \text{Cov}(\tilde{\boldsymbol{\theta}}_k, \tilde{\boldsymbol{\theta}}_k) + \text{Cov}(\tilde{\boldsymbol{\theta}}_k, \tilde{\mathbf{z}}_{ak}) \bar{\mathbf{T}}_k^T \\ &\quad + \bar{\mathbf{T}}_k \text{Cov}(\tilde{\mathbf{z}}_m, \tilde{\mathbf{z}}_m) \bar{\mathbf{T}}_k^T \end{aligned}$$

$$\begin{aligned} &- \bar{\mathbf{T}}_k \text{Cov}(\tilde{\mathbf{z}}_{ak}, \tilde{\boldsymbol{\theta}}_k) \\ &+ \bar{\mathbf{T}}_k \text{Cov}(\tilde{\mathbf{z}}_{ak}, \tilde{\mathbf{z}}_{ak}) \bar{\mathbf{T}}_k^T, \end{aligned} \quad (6)$$

where subscript k denotes the k th iteration step. The vectors $\bar{\boldsymbol{\theta}}_k$, $\bar{\mathbf{z}}_{ak}$, $\bar{\mathbf{z}}_m$ are the means of the updating parameters, the analytical modal parameters, and the actual modal parameters. The vectors $\tilde{\boldsymbol{\theta}}_k$, $\tilde{\mathbf{z}}_{ak}$, $\tilde{\mathbf{z}}_m$ are the corresponding zero-mean random parts with the same variances of $\sigma_{\tilde{\boldsymbol{\theta}}_k}^2$, $\sigma_{\tilde{\mathbf{z}}_{ak}}^2$, $\sigma_{\tilde{\mathbf{z}}_m}^2$. \mathbf{W}_ε , \mathbf{W}_ϑ is a diagonal weighting matrix to allow for regularization of ill-posed sensitivity equations. $\bar{\mathbf{T}}_k$ is a transformation matrix and $\bar{\mathbf{S}}_k$ is the sensitivity matrix of modal parameters with respect to updating parameters. In (6), the actual modal data and updating parameters are assumed to be uncorrelated; that is, $\text{Cov}(\tilde{\mathbf{z}}_m, \tilde{\boldsymbol{\theta}}_k) = 0$, $\text{Cov}(\tilde{\mathbf{z}}_m, \tilde{\mathbf{z}}_{ak}) = 0$, and the covariance matrix $\text{Cov}(\tilde{\mathbf{z}}_{ak}, \tilde{\boldsymbol{\theta}}_k)$ between the analytical modal parameters and the updating parameters as well as the covariance matrix $\text{Cov}(\tilde{\mathbf{z}}_{ak}, \tilde{\mathbf{z}}_{ak})$ of the analytic modal parameters can be evaluated by forward propagation of uncertainty.

3.2. Optimization Algorithm by Region Truncation. Region truncation-based method is employed in the optimization process of model updating, since the widely used trust region method is less straightforward and more expensive in computational though it is powerful and reliable in convergence [34]. In region truncation-based optimization algorithm, the constrained minimization problem (3) with the constraints is converted to the unconstrained problem and the updating parameters are limited within a controlled region in each iteration step. Thus, (5) can be replaced with the following iterative process:

$$\begin{aligned} \bar{\boldsymbol{\theta}}_{k+1} &= \begin{cases} \bar{\boldsymbol{\theta}}_k + \bar{\boldsymbol{\tau}}_k, & \Delta \boldsymbol{\theta}_k \geq \bar{\boldsymbol{\tau}}_k, \\ \bar{\boldsymbol{\theta}}_k + \left((\mathbf{S}^T(\boldsymbol{\theta}_k) \mathbf{W}_\varepsilon \mathbf{S}(\boldsymbol{\theta}_k))^{-1} + \mathbf{W}_\vartheta \right) \mathbf{S}^T(\boldsymbol{\theta}_k) (\bar{\mathbf{z}}_{ak} - \bar{\mathbf{z}}_m), & \underline{\boldsymbol{\tau}}_k \leq \Delta \boldsymbol{\theta}_k \leq \bar{\boldsymbol{\tau}}_k, \\ \bar{\boldsymbol{\theta}}_k + \underline{\boldsymbol{\tau}}_k, & \Delta \boldsymbol{\theta}_k \leq \underline{\boldsymbol{\tau}}_k, \end{cases} \\ \Delta \bar{\boldsymbol{\theta}}_{k+1} &= \bar{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_{k-1}, \end{aligned}$$

$$\begin{aligned}
\bar{\boldsymbol{\tau}}_k &= \beta \boldsymbol{\kappa}_k, \\
\underline{\boldsymbol{\tau}}_k &= -\beta \boldsymbol{\kappa}_k, \\
\beta &\in (0, 1), \\
\boldsymbol{\kappa}_k &= \min \left\{ \left| \bar{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_k \right|, \left| \underline{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}}_k \right| \right\},
\end{aligned} \tag{7}$$

where $\bar{\boldsymbol{\tau}}_k, \underline{\boldsymbol{\tau}}_k$ are boundary vectors of the truncated region. $\boldsymbol{\kappa}_k$ is the distance vector of the current iterative point from the nearest constraints, and the size of truncation region depends on the parameter β . It is suggested that β should be set small for a large sensitivity of modal parameters to the updating parameters and large for a small sensitivity.

4. Finite Element Model for Damaged FGM Structures

4.1. Random Effective Material Properties. Recalling (1), the uncertainty in the effective material properties of an FGM comes from the uncertainty in both material parameters and volume fraction of constituents. Taking into account the low variability in both the physical properties (such as the Young modulus and mass density of each constituent material) and volume fraction index, and neglecting high-order terms, the effective material properties for the FGM beam can be expanded by Taylor's series:

$$P(\cdot) = \bar{P}(\cdot) + \tilde{P}(\cdot), \quad \tilde{P}(\cdot) = \sum_{i=1}^l \frac{\partial \bar{P}(\cdot)}{\partial \theta_i} \tilde{\theta}_i, \tag{8}$$

where $\bar{P}(\cdot)$ is the mean value of each effective material property, and $\tilde{P}(\cdot)$ is the corresponding zero-mean random part with the variance of $\sigma_{P(\cdot)}^2$. θ_i is the i th element of the updating parameter vector $\boldsymbol{\theta}$, and $\tilde{\theta}_i$ is the corresponding zero-mean random part.

4.2. Finite Element Model of Undamaged FGM Beam. According to the third-order shear deformation theory, the displacement field of the FGM beam can be expressed as follows [11, 12]:

$$\begin{aligned}
u(x, z, t) &= u_0(x, t) \\
&+ z \left[\phi_x(x, t) - \frac{4}{3} \left(\frac{z}{h} \right)^2 \left(\phi_x(x, t) + \frac{\partial w_0(x, t)}{\partial x} \right) \right], \tag{9} \\
w(x, t) &= w_0(x, t),
\end{aligned}$$

where u, v are the axial and transverse displacements at any point of the FGM beam in the x, z directions, respectively. u_0, w_0 are those on the mid-plane. ϕ_x is the cross-sectional rotation about x -axis.

By assuming the small deformation, the following strain field can be obtained:

$$\begin{aligned}
\varepsilon_x(x, z, t) &= \frac{\partial u_0}{\partial x} \\
&+ z \left[\frac{\partial \phi_x}{\partial x} - \frac{4}{3} \left(\frac{z}{h} \right)^2 \left(\frac{\partial \phi_x}{\partial x} + \frac{\partial^2 w_0}{\partial x^2} \right) \right], \tag{10} \\
\gamma_{xz}(x, z, t) &= \phi_x + \frac{\partial w_0}{\partial x} - 4 \left(\frac{z}{h} \right)^2 \left(\phi_x + \frac{\partial w_0}{\partial x} \right),
\end{aligned}$$

where ε_x and γ_{xz} are the normal and shear strains.

The constitutive relation in thermal environment can be expressed as

$$\begin{aligned}
&\begin{Bmatrix} \sigma_x(x, z, t) \\ \tau_{xz}(x, z, t) \end{Bmatrix} \\
&= \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{55} \end{bmatrix} \left(\begin{Bmatrix} \varepsilon_x \\ \gamma_{xz} \end{Bmatrix} - \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \alpha(x, z, T) \Delta T(x, z) \right), \tag{11}
\end{aligned}$$

where σ_x and τ_{xz} are the normal and shear stresses. Q_{11} and Q_{55} are the elastic coefficients, that is, $Q_{11} = E(x, z, T)/(1 - \nu^2)$, $Q_{55} = E(x, z, T)/2(1 + \nu)$. $E(x, z, T) = \bar{E}(x, z, T) + \tilde{E}(x, z, T)$ are the effective Young's elastic modulus, which is the function of z for the FGM beam through its thickness direction, and the function of x for the FGM beam along its axial direction. $\bar{E}(x, z, T)$ is the mean and $\tilde{E}(x, z, T)$ is the corresponding zero-mean random part. Poisson's ratio ν is assumed to be constant. $\alpha(x, z, T)$ is the thermal expansion efficient. $\Delta T(x, z)$ is the temperature change distribution for the FGM beam.

The two-node shear deformable beam element with four degrees of freedom in each node is employed in the development of the finite element model. The displacement vector at the mid-plane of a beam element can be expressed as

$$\begin{aligned}
&\begin{Bmatrix} u_0 & w_0 & \frac{\partial w_0}{\partial x} & \phi_x \end{Bmatrix}^T = [\mathbf{N}] \{\mathbf{q}^e\}, \\
&\{\mathbf{q}^e\} = \left\{ u_0^{e_1}, w_0^{e_1}, \frac{\partial w_0^{e_1}}{\partial x}, \phi_x^{e_1}, u_0^{e_2}, w_0^{e_2}, \frac{\partial w_0^{e_2}}{\partial x}, \phi_x^{e_2} \right\}^T, \tag{12}
\end{aligned}$$

where $[\mathbf{N}]$ is the shape function matrix and $\{\mathbf{q}^e\}$ is the nodal displacement vector of beam element.

The strain and kinetic energies of the beam element are

$$U = \iiint_{V_e} \frac{1}{2} (\varepsilon_x Q_{11} \varepsilon_x + \gamma_{xz} Q_{55} \gamma_{xz} - \varepsilon_x Q_{11} \alpha \Delta T) dx dy dz, \quad (13)$$

$$T = \iiint_{V_e} \frac{1}{2} \rho(x, z, T) (\dot{u}^2 + \dot{w}^2) dx dy dz,$$

where V_e is the volume of a beam element. $\rho(x, z, T) = \bar{\rho}(x, z, T) + \tilde{\rho}(x, z, T)$ are the effective Young's modulus, which is the function of z for the FGM beam through its thickness, and the function of x for the FGM beam along its axial direction. $\bar{\rho}(x, z, T)$ is the mean and $\tilde{\rho}(x, z, T)$ is the corresponding zero-mean random part.

The following stochastic finite element equation of vibration for the beam element in thermal environment can be obtained by using Hamilton's principle:

$$[\mathbf{M}_e] \{\ddot{\mathbf{q}}_e\} + [\mathbf{K}_e] \{\mathbf{q}_e\} = \{\mathbf{F}_e\}, \quad (14)$$

where $[\mathbf{K}_e] = [\bar{\mathbf{K}}_e] + [\tilde{\mathbf{K}}_e]$ is the element stiffness matrix, $[\mathbf{M}_e] = [\bar{\mathbf{M}}_e] + [\tilde{\mathbf{M}}_e]$ is the element mass matrix, and $\{\mathbf{F}_e\} = [\bar{\mathbf{F}}_e] + [\tilde{\mathbf{F}}_e]$ is the thermal force. Each matrix is composed of the mean matrix and the zero-mean random part.

Recalling (8), the means $[\bar{\mathbf{K}}_e]$, $[\bar{\mathbf{M}}_e]$, $[\bar{\mathbf{F}}_e]$ and the zero-mean random parts $[\tilde{\mathbf{K}}_e]$, $[\tilde{\mathbf{M}}_e]$, $[\tilde{\mathbf{F}}_e]$ can be computed, respectively, by

$$[\bar{\mathbf{K}}_e] = \iint \bar{E}(x, z, T) [\bar{\mathbf{K}}_e] dx dz,$$

$$[\bar{\mathbf{M}}_e] = \iint \bar{\rho}(x, z, T) [\bar{\mathbf{M}}_e] dx dz, \quad (15)$$

$$\{\bar{\mathbf{F}}_e\} = \iint \bar{E}(x, z, T) \{\bar{\mathbf{F}}_e\} dx dz,$$

$$[\tilde{\mathbf{K}}_e] = \iint \left(\sum_{i=1}^l \frac{\partial \bar{E}(x, z, T)}{\partial \bar{\theta}_i} \bar{\theta}_i \right) [\tilde{\mathbf{K}}_e] dx dz,$$

$$[\tilde{\mathbf{M}}_e] = \iint \left(\sum_{k=1}^l \frac{\partial \bar{\rho}(x, z, T)}{\partial \bar{\theta}_k} \bar{\theta}_k \right) [\tilde{\mathbf{M}}_e] dx dz, \quad (16)$$

$$\{\tilde{\mathbf{F}}_e\} = \iint \left(\sum_{k=1}^l \frac{\partial \bar{E}(x, z, T)}{\partial \bar{\theta}_k} \bar{\theta}_k \right) \{\tilde{\mathbf{F}}_e\} dx dz,$$

where the matrices $[\bar{\mathbf{K}}_e]$, $[\bar{\mathbf{M}}_e]$, $\{\bar{\mathbf{F}}_e\}$ can be computed by

$$[\bar{\mathbf{K}}_e] = \frac{b}{1 - \nu^2} \left([\mathbf{B}_\varepsilon]^T [\mathbf{B}_\varepsilon] + [\mathbf{B}_\gamma]^T [\mathbf{B}_\gamma] \right),$$

$$[\bar{\mathbf{M}}_e] = b [\mathbf{D}]^T [\mathbf{D}],$$

$$\{\bar{\mathbf{F}}_e\} = \frac{b}{1 - \nu^2} [\mathbf{B}_\varepsilon]^T \alpha(x, z, T) \Delta T(x, z),$$

$$[\mathbf{B}_\varepsilon] = [1 \ 0 \ -c_1 z^3 \ z(1 - c_1 z^2)] \frac{\partial [\mathbf{N}]}{\partial x},$$

$$[\mathbf{B}_\gamma] = [0 \ 0 \ 1 - 3c_1 z^2 \ 1 - 3c_1 z^2] [\mathbf{N}],$$

$$[\mathbf{D}] = \begin{bmatrix} 1 & 0 & -c_1 z^3 & z(1 - c_1 z^2) \\ 0 & 1 & 0 & 0 \end{bmatrix} [\mathbf{N}], \quad (17)$$

where $[\mathbf{B}_\varepsilon]$, $[\mathbf{B}_\gamma]$, and $[\mathbf{D}]$ are the matrices mainly related to the shape function matrix $[\mathbf{N}]$ and $c_1 = 4/3h^2$.

4.3. Finite Element Model of Damaged FGM Beam. For the damaged FGM beam, a local damage often leads to the reduction in the local stiffness parameter. In this section, damage is introduced by reducing effectiveness elastic moduli of the corresponding elements [16]. Introducing a damage factor λ_e indicating the damage severity for the e th damaged element, effective Young's modulus of the e th damaged element can be expressed as

$$\bar{E}_d(x, z, T) = \bar{E}(x, z, T) (1 - \lambda_e), \quad (18)$$

where $\bar{E}(x, z, T)$ is the effective Young's elastic modulus for the undamaged FGM beam and $\bar{E}_d(x, z, T)$ is for the damaged FGM beam.

Recalling (15), the stiffness matrix of the damaged element can be expressed as

$$[\mathbf{K}_{e,d}] = [\mathbf{K}_e] (1 - \lambda_e), \quad (19)$$

where λ_e is the damage factor of the e th damaged element and $[\mathbf{K}_{e,d}]$ is the stiffness element for the e th damaged element.

With (18) and (19), the element matrices in (14) can be assembled to obtain the global stochastic finite element equation of vibration for the damaged FGM beam.

$$[\mathbf{M}] \{\ddot{\mathbf{q}}\} + [\mathbf{K}] \{\mathbf{q}\} = \{\mathbf{F}\},$$

$$[\mathbf{K}] = [\bar{\mathbf{K}}] + [\tilde{\mathbf{K}}],$$

$$[\mathbf{M}] = [\bar{\mathbf{M}}] + [\tilde{\mathbf{M}}],$$

$$\{\mathbf{F}\} = \{\bar{\mathbf{F}}\} + \{\tilde{\mathbf{F}}\},$$

$$[\bar{\mathbf{K}}]$$

$$= \sum_e^{n_u} [\mathbf{T}_e]^T [\bar{\mathbf{K}}_e] [\mathbf{T}_e]$$

$$+ \sum_e^{n_d} (1 - \lambda_e) [\mathbf{T}_e]^T [\bar{\mathbf{K}}_e] [\mathbf{T}_e],$$

$$[\bar{\mathbf{M}}] = \sum_e^{n_u + n_d} [\mathbf{T}_e]^T [\bar{\mathbf{M}}_e] [\mathbf{T}_e],$$

$$\{\bar{\mathbf{F}}\} = \sum_e^{n_u} [\mathbf{T}_e]^T \{\bar{\mathbf{F}}_e\} + \sum_e^{n_d} (1 - \lambda_e) [\mathbf{T}_e]^T \{\bar{\mathbf{F}}_e\},$$

$$\begin{aligned}
& [\bar{\mathbf{K}}] \\
&= \sum_e^{n_u} [\mathbf{T}_e]^T [\bar{\mathbf{K}}_e] [\mathbf{T}_e] \\
&\quad + \sum_e^{n_d} (1 - \lambda_e) [\mathbf{T}_e]^T [\bar{\mathbf{K}}_e] [\mathbf{T}_e], \\
& [\bar{\mathbf{M}}] = \sum_e^{n_e} [\mathbf{T}_e]^T [\bar{\mathbf{M}}_e] [\mathbf{T}_e], \\
& \{\bar{\mathbf{F}}\} = \sum_e^{n_u} [\mathbf{T}_e]^T \{\bar{\mathbf{F}}_e\} + \sum_e^{n_d} (1 - \lambda_e) [\mathbf{T}_e]^T \{\bar{\mathbf{F}}_e\},
\end{aligned} \tag{20}$$

where $[\mathbf{K}]$ is the global random stiffness matrix and $[\mathbf{M}]$ is the global random mass matrix. $\{\mathbf{F}\}$ is the global thermal force and $[\mathbf{T}_e]$ is a transformation matrix that transforms the local coordinate of the e th element to the global coordinate. n_u is the number of undamaged elements and n_d is the number of damaged elements.

5. Forward Uncertainty Propagation

The j th modal frequency f_j and mode shape $\{\boldsymbol{\varphi}\}_j$ are governed by the following:

$$([\mathbf{K}] - (2\pi f_j)^2 [\mathbf{M}]) \{\boldsymbol{\varphi}\}_j = 0, \quad j = 1, 2, \dots, N. \tag{21}$$

The j th modal frequency and mode shape can also be represented as the sum of its mean and a zero-mean random part:

$$\begin{aligned}
f_j &= \bar{f}_j + \tilde{f}_j, \\
\{\boldsymbol{\varphi}\}_j &= \{\bar{\boldsymbol{\varphi}}\}_j + \{\tilde{\boldsymbol{\varphi}}\}_j
\end{aligned} \tag{22}$$

and further the random part of the j th normalized mode shape can be written as

$$\{\tilde{\boldsymbol{\varphi}}\}_j = \sum_{r=1}^N \varepsilon_{rj} \{\bar{\boldsymbol{\varphi}}\}_r, \tag{23}$$

where ε_{rj} , $r = 1, 2, \dots, N$ are the small coefficients to be determined. Substituting (22) into (21), we have

$$([\bar{\mathbf{K}}] - (2\pi \bar{f}_j)^2 [\bar{\mathbf{M}}]) \{\bar{\boldsymbol{\varphi}}\}_j = 0, \tag{24}$$

$$\begin{aligned}
& [\bar{\mathbf{K}}] \sum_{r=1}^N \varepsilon_{rj} \{\bar{\boldsymbol{\varphi}}\}_r + [\bar{\mathbf{K}}] \{\bar{\boldsymbol{\varphi}}\}_j \\
&= (2\pi \bar{f}_j)^2 [\bar{\mathbf{M}}] \sum_{r=1}^N \varepsilon_{rj} \{\bar{\boldsymbol{\varphi}}\}_r + (2\pi \bar{f}_j)^2 [\bar{\mathbf{M}}] \{\bar{\boldsymbol{\varphi}}\}_j \\
&\quad + 8\pi^2 \bar{f}_j \tilde{f}_j [\bar{\mathbf{M}}] \{\bar{\boldsymbol{\varphi}}\}_j.
\end{aligned} \tag{25}$$

Using the orthogonality of mode shapes, the random part of the j th modal frequency and the coefficients ε_{ij} ($i \neq j$)

can be obtained by premultiplying (25) by $\{\bar{\boldsymbol{\varphi}}\}_j^T$ and $\{\bar{\boldsymbol{\varphi}}\}_i^T$, respectively,

$$\begin{aligned}
\tilde{f}_j &= \frac{1}{8\pi^2 \bar{f}_j} \{\bar{\boldsymbol{\varphi}}\}_j^T ([\bar{\mathbf{K}}] - (2\pi \bar{f}_j)^2 [\bar{\mathbf{M}}]) \{\bar{\boldsymbol{\varphi}}\}_j, \\
\varepsilon_{ij} &
\end{aligned} \tag{26}$$

$$= \frac{1}{4\pi^2 (\bar{f}_j^2 - \bar{f}_i^2)} \{\bar{\boldsymbol{\varphi}}\}_j^T ([\bar{\mathbf{K}}] - (2\pi \bar{f}_j)^2 [\bar{\mathbf{M}}]) \{\bar{\boldsymbol{\varphi}}\}_i.$$

Recalling the orthogonality condition $\{\boldsymbol{\varphi}\}_j^T [\mathbf{M}] \{\boldsymbol{\varphi}\}_j = 1$ and using (20) and (22), the coefficient ε_{jj} can be obtained by

$$\varepsilon_{jj} = -\frac{1}{2} \{\bar{\boldsymbol{\varphi}}\}_j^T [\bar{\mathbf{M}}] \{\bar{\boldsymbol{\varphi}}\}_j. \tag{27}$$

Neglecting the high-order terms, the random parts of modal parameters in (22) can be obtained by the Taylor's expansion:

$$\tilde{f}_j = \sum_{i=1}^l \frac{\partial \bar{f}_j}{\partial \theta_i} \tilde{\theta}_i, \tag{28}$$

$$\{\tilde{\boldsymbol{\varphi}}\}_j = \sum_{k=1}^l \frac{\partial \{\bar{\boldsymbol{\varphi}}\}_j}{\partial \theta_i} \tilde{\theta}_i.$$

Using (20), (26)–(28), we have the following equation for the solution of the unknown $\partial \bar{f}_j / \partial \theta_i$ and $\partial \{\bar{\boldsymbol{\varphi}}\}_j / \partial \theta_i$ by equating the coefficients of each $\tilde{\theta}_i$:

$$\begin{aligned}
\sum_{i=1}^l \frac{\partial \bar{f}_j}{\partial \theta_i} \tilde{\theta}_i &= \frac{1}{8\pi^2 \bar{f}_j} \{\bar{\boldsymbol{\varphi}}\}_j^T ([\bar{\mathbf{K}}] - (2\pi \bar{f}_j)^2 [\bar{\mathbf{M}}]) \{\bar{\boldsymbol{\varphi}}\}_j, \\
\sum_{i=1}^l \frac{\partial \{\bar{\boldsymbol{\varphi}}\}_j}{\partial \theta_i} \tilde{\theta}_i &= -\frac{1}{2} \{\bar{\boldsymbol{\varphi}}\}_j^T [\bar{\mathbf{K}}] \{\bar{\boldsymbol{\varphi}}\}_j \{\bar{\boldsymbol{\varphi}}\}_j \\
&\quad + \sum_{i=1}^N \left\{ \frac{1}{4\pi^2 (\bar{f}_j^2 - \bar{f}_i^2)} \{\bar{\boldsymbol{\varphi}}\}_j^T \right. \\
&\quad \cdot ([\bar{\mathbf{K}}] - (2\pi \bar{f}_j)^2 [\bar{\mathbf{M}}]) \{\bar{\boldsymbol{\varphi}}\}_j \{\bar{\boldsymbol{\varphi}}\}_i \left. \right\},
\end{aligned} \tag{29}$$

$$\begin{aligned}
[\bar{\mathbf{K}}] &= \sum_e^{n_u} [\mathbf{T}_e]^T \iint \left(\sum_{i=1}^l \frac{\partial \bar{E}(x, z, T)}{\partial \tilde{\theta}_i} \tilde{\theta}_i \right) [\bar{\mathbf{K}}_e] dx dz \\
&\quad \cdot [\mathbf{T}_e] + \sum_e^{n_d} [\mathbf{T}_e]^T \iint (1 - \lambda_e) \left(\sum_{i=1}^l \frac{\partial \bar{E}(x, z, T)}{\partial \tilde{\theta}_i} \tilde{\theta}_i \right) \\
&\quad \cdot [\bar{\mathbf{K}}_e] dx dz [\mathbf{T}_e],
\end{aligned}$$

$$\begin{aligned}
[\bar{\mathbf{M}}] &= \sum_{e=1}^{n_u+n_d} [\mathbf{T}_e]^T \iint \left(\sum_{k=1}^l \frac{\partial \bar{\rho}(x, z)}{\partial \tilde{\theta}_i} \tilde{\theta}_i \right) [\bar{\mathbf{M}}_e] dx dz \\
&\quad \cdot [\mathbf{T}_e].
\end{aligned}$$

TABLE 1: The following SuS304/Si3N4 type of FGM properties are used for computation [30].

Material	Properties	P_0	P_{-1}	P_1	P_2	P_3
Ceramics (Si ₃ N ₄)	E (Pa)	$348.43e + 9$	0	$-3.070e - 4$	$2.160e - 7$	$-8.964e - 11$
	α (1/K)	$5.8723e - 6$	0	$9.095e - 4$	0	0
	ρ (kg/m ³)	2370	0	0	0	0
Metal (SUS304)	E (Pa)	$201.04e + 9$	0	$3.079e - 4$	$-6.534e - 7$	0
	α (1/K)	$12.330e - 6$	0	$8.086e - 4$	0	0
	ρ (kg/m ³)	8166	0	0	0	0

Further, the variances of the j th modal frequency and normalized mode shape as well as the covariance between the modal parameters and the random material parameters can be obtained:

$$\begin{aligned} \text{Var}(f_j) &= \sum_{k=1}^l \sum_{q=1}^l \frac{\partial \bar{f}_j}{\partial \theta_k} \frac{\partial \bar{f}_j}{\partial \theta_q} \rho_{kq} \sigma_{\bar{\theta}_k} \sigma_{\bar{\theta}_q}, \\ \text{Var}(\{\boldsymbol{\varphi}\}_j) &= \sum_{k=1}^l \sum_{q=1}^l \text{diag} \left(\frac{\partial \{\boldsymbol{\varphi}\}_j}{\partial \theta_k} \frac{\partial \{\boldsymbol{\varphi}\}_j^T}{\partial \theta_q} \right) \rho_{kq} \sigma_{\bar{\theta}_k} \sigma_{\bar{\theta}_q}, \\ \text{Cov}(f_j, \theta_i) &= \sum_{k=1}^l \frac{\partial \bar{f}_j}{\partial \theta_k} \rho_{ki} \sigma_{\bar{\theta}_k} \sigma_{\bar{\theta}_i}, \\ \text{Cov}(\{\boldsymbol{\varphi}\}_j, \theta_i) &= \sum_{k=1}^l \frac{\partial \{\boldsymbol{\varphi}\}_j}{\partial \theta_k} \rho_{ki} \sigma_{\bar{\theta}_k} \sigma_{\bar{\theta}_i}, \end{aligned} \quad (30)$$

where $\sigma_{\bar{\theta}_k}$, $\sigma_{\bar{\theta}_q}$, $\sigma_{\bar{\theta}_i}$ are the standard deviations of the random material parameters, that is, constituent material properties and volume fraction index, and ρ_{kq} is the correlation coefficient of random parameters $\bar{\theta}_k$, $\bar{\theta}_q$.

6. Numerical Simulation

In order to demonstrate the stochastic model updating-based parameter identification approach outlined in the previous sections for damaged FGM structures with random material properties in thermal environment, a damaged FGM beam is considered for the numerical simulation. The results obtained from the presented identification approach have been validated by comparing with the given values. The FGM is composed of ceramics (Si₃N₄) and metal (SUS304) with the mean values of material properties given in Table 1. In the FGM, material properties are assumed to vary, either through its thickness direction or along its axial direction, according to power law distribution. Therefore, the bottom surface is pure metal and the top surface is pure ceramics for the FGM beam through its thickness direction, while the right side is pure metal and the left side is pure ceramics for the FGM beam along its axial direction. The FGM beam has the geometric parameters of 20 m in length, 1 m in thickness, and 0.8 m in width. In this section, Young's elastic moduli of two materials and volume fraction index are chosen as the parameters to be identified, since the former may significantly change with the environmental temperature and the latter is hard to be determined by traditional experiments.

In the finite element, the FGM beam was discretized into eight two-node shear deformable beam elements with four degrees of freedom in each node. It is assumed that the damage is located at the 8th element with a reduction in the effective Young's elastic modulus by 10%, that is, $\lambda_8 = 0.1$. Analytically, modal analysis is first carried out to obtain the initial analytical modal parameters by using finite element model of the damaged FGM beam with initial material properties and distribution. Modal analysis is then again carried out on the damaged FGM beam with given material properties and distribution to obtain the actual modal parameters due to the difficulty of random experiment with large samples.

The actual modal data is generated in a simulated way assuming that the uncertainty of modal data is caused by 5% COV (coefficient of variation, the ratio of the square root of the variance to the mean) in elastic moduli of two constituents as well as volume fraction index and 0.1% COV in measurement noise. Figure 2 shows the scatter of the actual modal frequencies set compared to the initial modal frequencies in the temperature of 300 K for the damaged FGM beam through the thickness direction ($n = 2$) compared to the undamaged case and Figure 3 shows those for the FGM beam along the axial direction ($n = 2$), in which the clamped-free (CF) boundary condition is firstly considered.

The stochastic finite element model updating is conducted using the region truncation-based optimization algorithm, and, as a result, the means and variances of elastic moduli of two materials and volume fraction index are identified by the stochastic model updating-based method (SMUM).

The statistical identification of material parameters for a damaged FGM structure can be implemented by the following steps.

Step 1. Develop the finite element model of the damaged FGM structure and determine the parameters $\boldsymbol{\theta}$ to be identified, which are chosen as the updating parameters of stochastic model updating.

Step 2. Compute the mean vector $\bar{\mathbf{z}}_m$ and covariance matrix $\text{Cov}(\bar{\mathbf{z}}_m, \bar{\mathbf{z}}_m)$ of the actual modal data.

Step 3. Initialize the means and variances of the updating parameters $\boldsymbol{\theta}$.

Step 4. Obtain the mean vector $\bar{\mathbf{z}}_{ak}$ of the analytical modal data after carrying out the modal analysis, and determine the

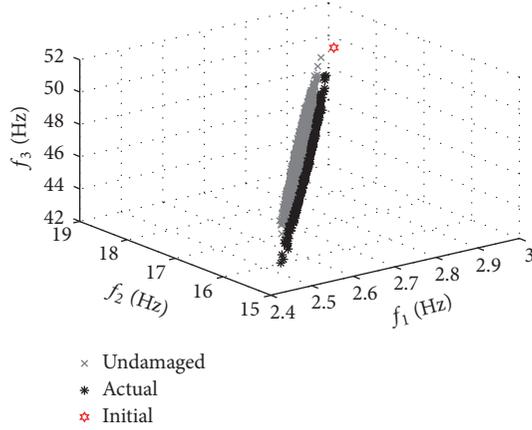


FIGURE 2: First three actual modal frequencies set (3000 samples) and initial modal frequencies (thickness).

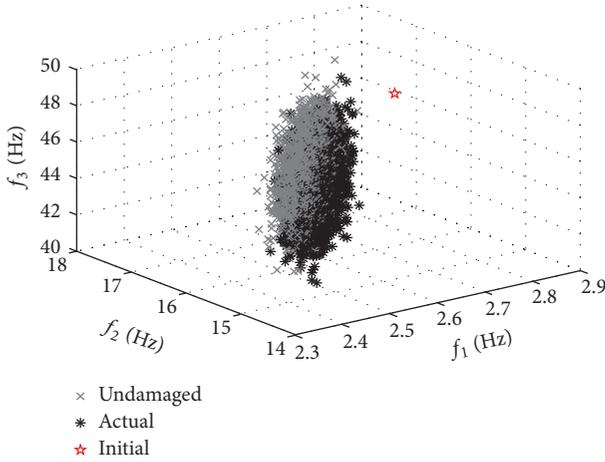


FIGURE 3: First three actual modal frequencies set (3000 samples) and initial modal frequencies (axial).

covariance matrices $\text{Cov}(\bar{\mathbf{z}}_{ak}, \tilde{\boldsymbol{\theta}}_k)$ and $\text{Cov}(\bar{\mathbf{z}}_{ak}, \bar{\mathbf{z}}_{ak})$ using the forward uncertainty propagation (30).

Step 5. Update the mean values $\bar{\boldsymbol{\theta}}_{k+1}$ and variances $\text{Cov}(\tilde{\boldsymbol{\theta}}_{k+1}, \tilde{\boldsymbol{\theta}}_{k+1})$ of the updating parameters using (7) and (6), respectively.

Step 6. Go to Step 4 until $\|\mathbf{z}_{ak}(\boldsymbol{\theta}) - \mathbf{z}_m\|^2$ is small enough and the second-order statistics (means and variances) of the identified parameters can be obtained.

The convergences of the mean values and variances of elastic moduli (E_c -elastic modulus of ceramics and E_m -elastic modulus of metal of metal) and volume fraction index (n) for the damaged FGM beam with material properties varying through its thickness in the temperature of 300 K and 600 K are shown in Figure 4. The results show that, for different temperatures, the means and variances of elastic moduli of two constituents and volume fraction index can fast converge by the proposed method. It is also seen that

the means of elastic moduli of two materials converge to different values for different temperatures since the elastic modulus of each material changes with the environmental temperature. The situation does not occur to volume fraction index, since volume fraction index is not affected by the temperature environment and depends on the manufacturing and fabrication process. The temperature has no obvious effect on the identification results of parameter variances.

For the comparison of convergence between the trust region method and the proposed method, Figure 5 gives the iteration process of mean values and variances of E_c , E_m , and n for the damaged FGM beam with material properties varying through its thickness direction in the temperature of 600 K by two methods. It is seen that mean values and variances of E_c , E_m , and n can converge faster by the proposed method than those by the trust region method. In fact, the CPU run time for the proposed method is about 6.356 seconds, and the CPU run time for the trust region method is about 13.251 seconds.

Compared to the Monte-Carlo simulation (MCS) method with 3000 samples, the mean values and COVs (coefficient of variation, the ratio of the square root of the variance to the mean) of elastic moduli of two materials and volume fraction index for both the damaged FGM beam through its thickness direction and the FGM beam along its axial direction with different volume fraction indices are shown in Tables 2 and 3, respectively.

It is seen from Tables 2 and 3 that, for the mean values or the COVs, the identification results by the stochastic model updating-based method (SMUM) proposed in this work agree favorably with those from the MCS. It is also observed that, compared to the given values, the means of parameters can be identified with good accuracy in the case of different volume fraction indices, while the identification results of the COVs are obviously affected by the volume fraction index and present relatively large errors, especially in the case of relatively small volume fraction indices or relatively large volume fraction indices; that is, one of constituents plays an dominant role in the FGM. In fact, the identified COV of E_c will be far away from the given value when the volume fraction index is large enough (e.g., $n = 10$) so that the FGM is rich in metal and few ceramics are in it. In contrast, the identified COV of E_m will be far away from the given value when the volume fraction index is small enough (e.g., $n = 0.5$) so that the FGM is rich in ceramics and few metal is in it. However, the situation does not happen to the FGM beam along its axial direction for these two cases (i.e., $n = 0.5$ and $n = 10$) and the volume fraction has no significant influence on the identification results of COVs. This can be explained that the material properties vary along the axial direction from the value of E_m to the value of E_c much more gradually than those along the thickness since the length (8 m) of beam is much larger than its thickness (1 m).

Figure 6 gives the convergences of the mean values and variances of E_c , E_m , and n for the FGM beam with material properties varying through its thickness direction in different cases of boundary conditions ($T = 600$ K). Four boundary conditions of clamped-free (CF), clamped-clamped (CC), clamped-simply (CS), and simply-simply (SS) are considered.

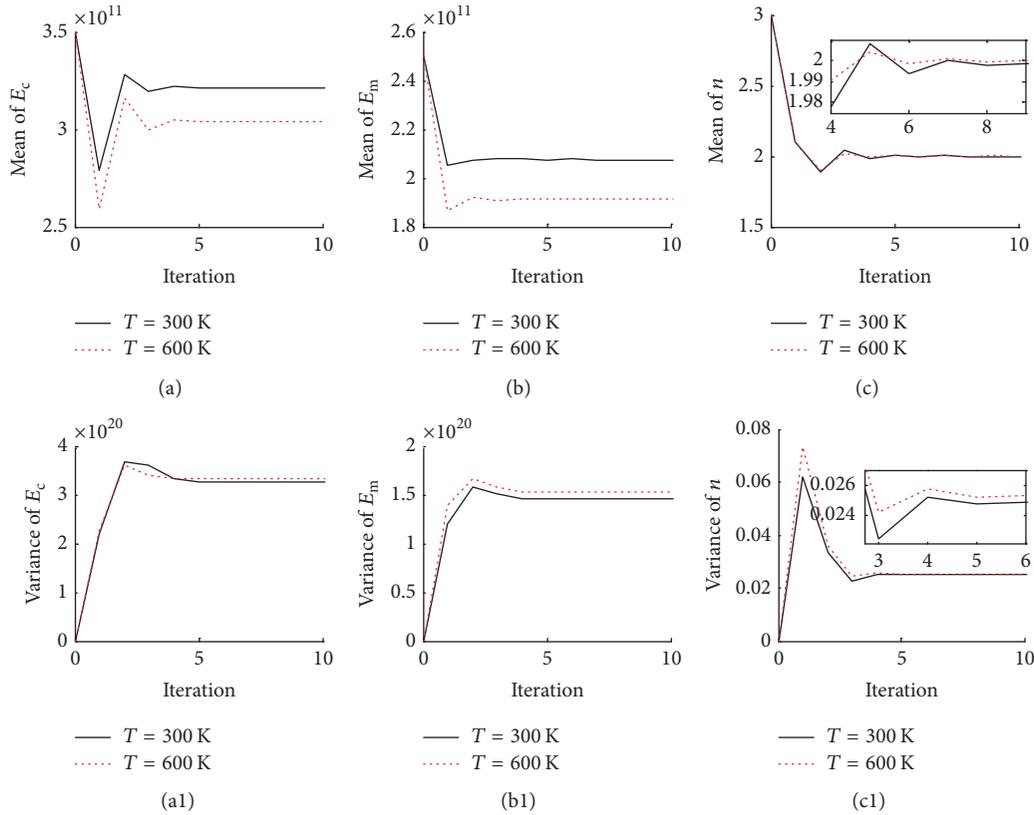


FIGURE 4: Iterative process of means and variances of material elastic moduli and volume fraction index.

It is observed that the means and variances of elastic moduli of two constituents and volume fraction index can converge by the proposed method no matter what kind of boundary conditions, which means that the boundary condition has no influence on the convergence of iterative process. It is also seen that the boundary condition has more influences on the identification result of the variance than that of the mean, and this means that the identification of the dispersion (i.e., variance or COV) is more susceptible to the external disturbance compared to the identification of the mean.

In order to further investigate the influence of the boundary conditions on the means and COVs of identified parameters, the means and COVs of material elastic moduli and volume fraction index of the FGM beam through its thickness direction or along its axial direction ($n = 2$) for different cases of boundary conditions are shown in Table 4. From Table 4, under the same case of parameter randomness and measurement noise, boundary conditions have no significant effects on the identified results of means and COVs of parameters with the exception of small fluctuation in values.

In order to investigate the influence of the uncertainty source of actual modal data on the means and COVs of identified parameters, the means and COVs of material elastic moduli and volume fraction index of the FGM beam through its thickness direction or along its axial direction ($n = 2$) for different cases of uncertainty are shown in Tables 5 and 6. These cases include that the uncertainty of actual modal data is caused by the structural material

(i.e., material elastic moduli and volume fraction index) randomness (SR) only, the measurement noise (MN) only, and the combination of two sources (both SR and MN). With structural material randomness (SR) and measurement noise (MN), four combinations of levels of uncertainty are considered. These combinations are (1) Level 1: 1% COV for SR (elastic moduli and volume fraction) and 0.1% for MN; (2) Level 2: 1% for SR and 0.5% for MN; (3) Level 3: 5% for SR and 0.1% for MN; (4) Level 4: 5% for SR and 0.5% for MN.

The results from Tables 5 and 6 show that the uncertainty does not bring too much changes in the means of identified parameters no matter what sources the uncertainty of actual modal data comes from, and, unlikely, the uncertainty contributes to the COV results, whose errors increase with the uncertainty level. Taking the FGM beam through its thickness direction with the environmental temperature of 300 K as an example, the COVs of identified parameters range from 2.1465% to 11.5036% for metal constituent elastic modulus, 3.0398% to 15.3656% for ceramics constituent modulus, and 4.5745% to 22.8086% for constituent volume fraction index when the level of uncertainty varies from Level 1 (1% SR, 0.1% MN) to Level 4 (5% SR, 0.5% MN). The same situation occurs to the FGM beam along its axial direction. It is worth noting that the accuracy of identified COVs is much more sensitive to the measurement noise, for example, ranging from 6.1553% to 15.3656% for E_c , 6.2834% to 11.5036% for E_m , and 9.9623% to 22.8086% for n when the level of MN varies from 0.1% (Level 3) to 0.5% (Level 4) with the same SR of 5%. It is

TABLE 2: Means of material elastic moduli and volume fraction index for SMUM and MCS.

Temperature	Parameters (given values)	$n = 0.5$			$n = 2$			$n = 5$			$n = 10$		
		SMUM	MCS	SMUM	MCS	SMUM	MCS	SMUM	MCS	SMUM	MCS	SMUM	MCS
FGM beam through its thickness													
T = 300 K	E_c (3.22272e11)	3.2174e112	3.2236e11	3.2226e + 11	3.2233e11	3.2163e11	3.2261e11	3.2079e11	3.2287e11				
	E_m (2.07788e11)	2.07768e11	2.0778e11	2.0731e + 11	2.0783e11	2.0759e11	2.0789e11	2.0758e11	2.0750e11				
	n (0.5/2/5/10)	0.4978	0.5027	1.9910	1.9986	4.9583	5.0508	9.7925	9.9449				
T = 600 K	E_c (3.04610e11)	3.0388e11	3.0579e11	3.0465e11	3.0437e11	3.0396e11	3.0608e11	3.0265e11	3.0242e11				
	E_m (1.90890e11)	1.9059e11	1.9176e11	1.9084e11	1.9061e11	1.9027e11	1.9101e11	1.9008e11	1.9069e11				
	n (0.5/2/5/10)	0.4985	0.5086	1.9943	2.0049	4.9412	5.0509	9.7276	9.7003				
FGM along its axial direction													
T = 300 K	E_c (3.22272e11)	3.2177e11	3.2210e11	3.2158e11	3.2247e11	3.2139e11	3.2220e11	3.2125e11	3.2231e11				
	E_m (2.07788e11)	2.0832e11	2.0752e11	2.0775e11	2.0734e11	2.0759e11	2.0768e11	2.0769e11	2.0789e11				
	n (0.5/2/5/10)	0.4986	0.5013	1.9965	1.9973	5.0012	4.9949	9.9779	10.0179				
T = 600 K	E_c (3.04610e11)	3.0386e11	3.0418e11	3.0396e11	3.0454e11	3.0383e11	3.0478e11	3.0489e11	3.0521e11				
	E_m (1.90890e11)	1.9149e11	1.8969e11	1.9125e11	1.9035e11	1.9069e11	1.9069e11	1.9105e11	1.9072e11				
	n (0.5/2/5/10)	0.4987	0.5014	1.9989	1.9998	4.9951	5.0155	9.9955	10.0383				

TABLE 3: COVs of material elastic moduli and volume fraction index for SMUM and MCS.

Temperature	Parameters (given values)	$n = 0.5$			$n = 2$			$n = 5$			$n = 10$		
		SMUM	MCS	SMUM	MCS	SMUM	MCS	SMUM	MCS	SMUM	MCS	SMUM	MCS
FGM beam through its thickness													
T = 300 K	E_c (5%)	6.9139%	6.3355%	6.1553%	6.0225%	5.8767%	5.4019%	8.4827%	4.1443%				
	E_m (5%)	8.4283%	7.4319%	6.2834%	5.3610%	5.7577%	5.0743%	5.9237%	5.1053%				
	n (5%)	11.9278%	12.2767%	9.9623%	6.8676%	9.5556%	10.4554%	26.9936%	18.9727%				
T = 600 K	E_c (5%)	7.3813%	6.4775%	6.4379%	5.9881%	6.2478%	5.2932%	8.9632%	8.8812%				
	E_m (5%)	9.0423%	7.2826%	6.8224%	5.3459%	6.3822%	5.2756%	6.5709%	6.5534%				
	n (5%)	11.3814%	11.8691%	9.7212%	6.7731%	8.2936%	9.5587%	26.0417%	28.2942%				
FGM beam along its axial direction													
T = 300 K	E_c (5%)	5.6619%	5.0117%	5.6988%	5.1429%	6.2028%	5.6765%	6.4789%	6.0394%				
	E_m (5%)	5.9169%	5.8154%	5.9243%	5.0552%	5.5639%	5.0349%	5.0739%	5.1134%				
	n (5%)	6.2618%	6.2209%	5.9821%	5.3645%	6.0643%	5.8934%	6.6918%	6.1129%				
T = 600 K	E_c (5%)	5.9368%	4.9651%	6.0989%	5.3348%	6.5508%	5.5821%	6.7065%	5.9315%				
	E_m (5%)	6.5070%	6.5424%	6.2054%	5.4021%	6.0851%	5.1826%	6.0854%	5.0826%				
	n (5%)	6.2779%	6.1971%	5.7743%	5.5224%	6.1823%	6.0234%	6.6069%	5.9734%				

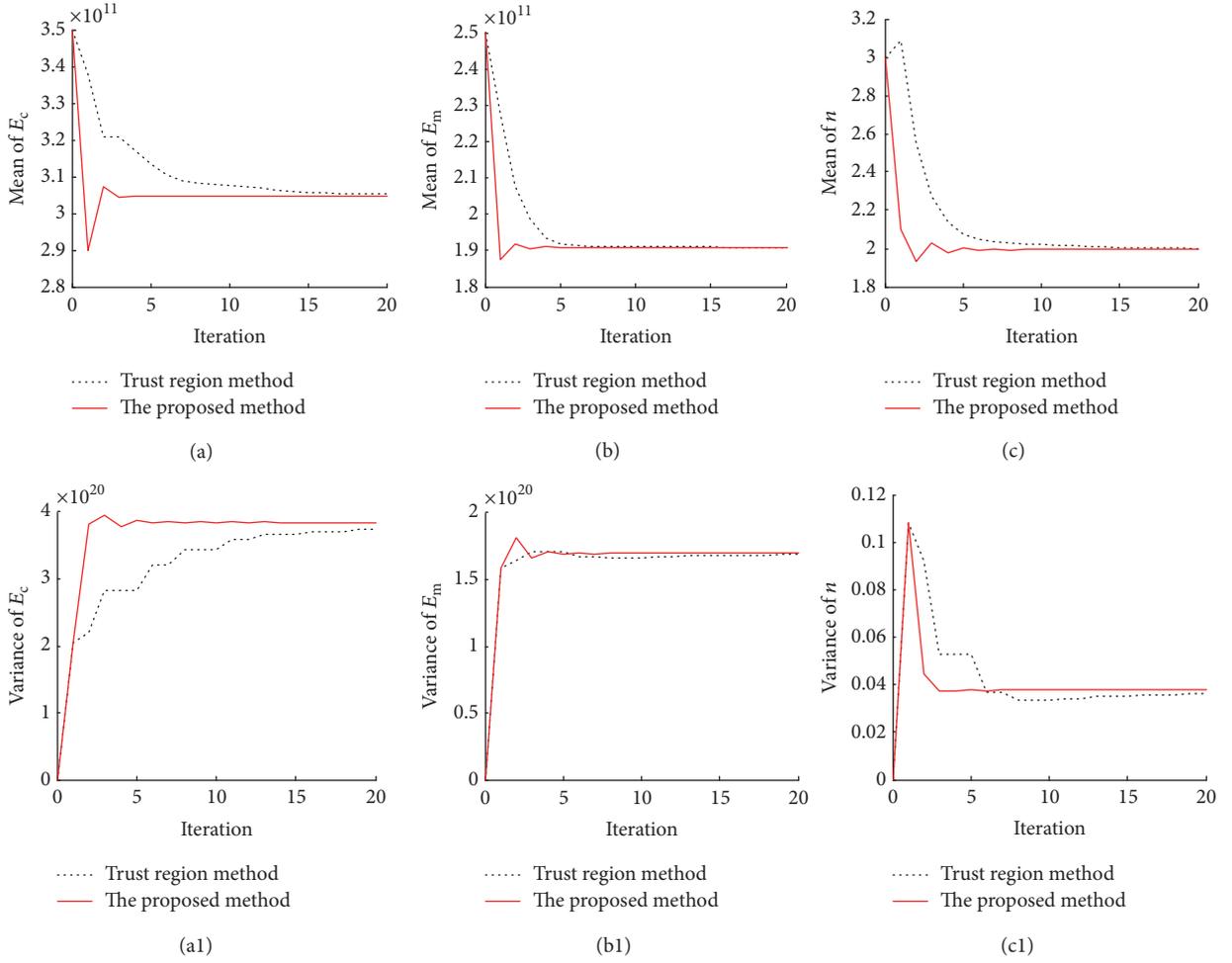


FIGURE 5: Iterative process of means and variances of E_c , E_m , and n for two methods ($T = 600$ K).

also observed that the influence of the measurement noise with a large value on the identified results is much more larger than that of structural material randomness although the identified COVs for each parameter obtained in the case of the combination of two uncertainty sources are higher than those obtained in the case of SR only or MN only; for example, the COVs is 15.3656% for E_c of the FGM through its thickness direction ($T = 300$ K) in the case of 5% SR and 0.5% MN compared to 5.5865% in the case of 5% SR only and 14.2591% in the case of 0.5% MN only, which means that the identified results of COVs when both SR and MN are included are not the simple linear superposition of the results with respect to SR only or MN only concerned, and, in fact, the measurement noise has a pronounced effect on the accuracy of the identified results. It is also seen that, for the parameters with the smaller dispersion, the higher level of measurement noise can lead to the results with much larger errors, for example, 3.0398% COVs for E_c in the case of 1% SR and 0.1% MN (level 1) compared to 6.1553% in the case of 5% SR and 0.1% MN (Level 3), which means an error of 203.98% for Level 1 compared to an error of 23.11% for Level 3. In fact, in the case of a relatively small level of SR combined with a relatively large level of MN, the identified

results of COVs could be totally annihilated by measurement noise.

The same situation occurs to the case of FGM beam along the axial direction. Compared to the given value, Figures 7, 8, and 9 show the mean values and COVs of E_c , E_m , and n for the FGM beam along its axial direction under different cases and levels of uncertainty. For each identified parameter, the result shows that the mean can be identified with good accuracy in the presence of uncertainty, while the identified result of COV is susceptible to the measurement noise. In fact, the increase of measurement noise may result in a large error in the identified result of COV compared to the given value.

Compared to the given probability density functions (PDFs), Figures 10(a), 11(a), and 12(a) show the identified PDFs of E_c , E_m , and n for the FGM beam (CF) through its thickness direction or along its axial direction with volume fraction index of $n = 2$ in the thermal environment of 300 K and 600 K for the uncertainty level of 5% SR and 0.1% MN, in which A3 and A6 denote the given PDFs for 300 K and 600 K, respectively, PT3 and PT6 denote the identified PDFs for 300 K and 600 K, respectively, when the FGM beam through its thickness direction is considered, and PA3 and PA6 denote

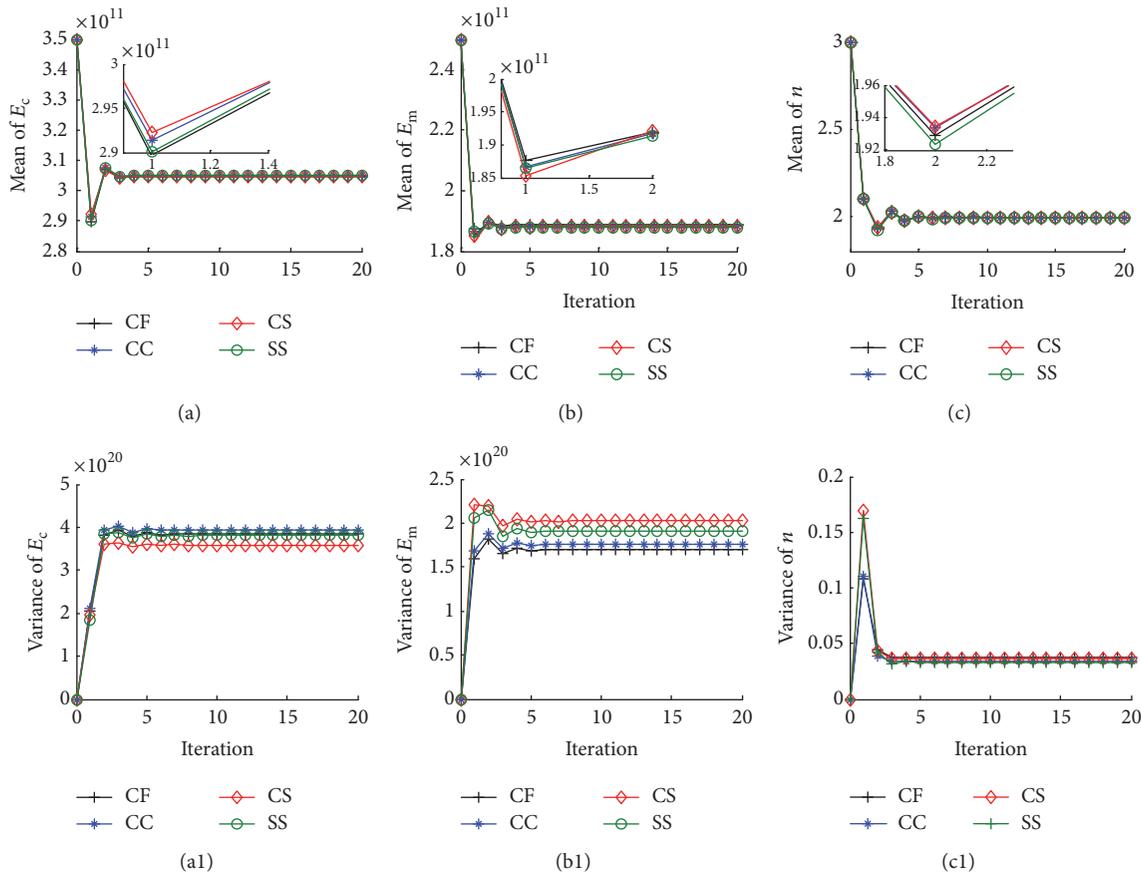


FIGURE 6: Iterative process of means and variances of E_c , E_m , and n for different boundary conditions ($T = 600$ k).

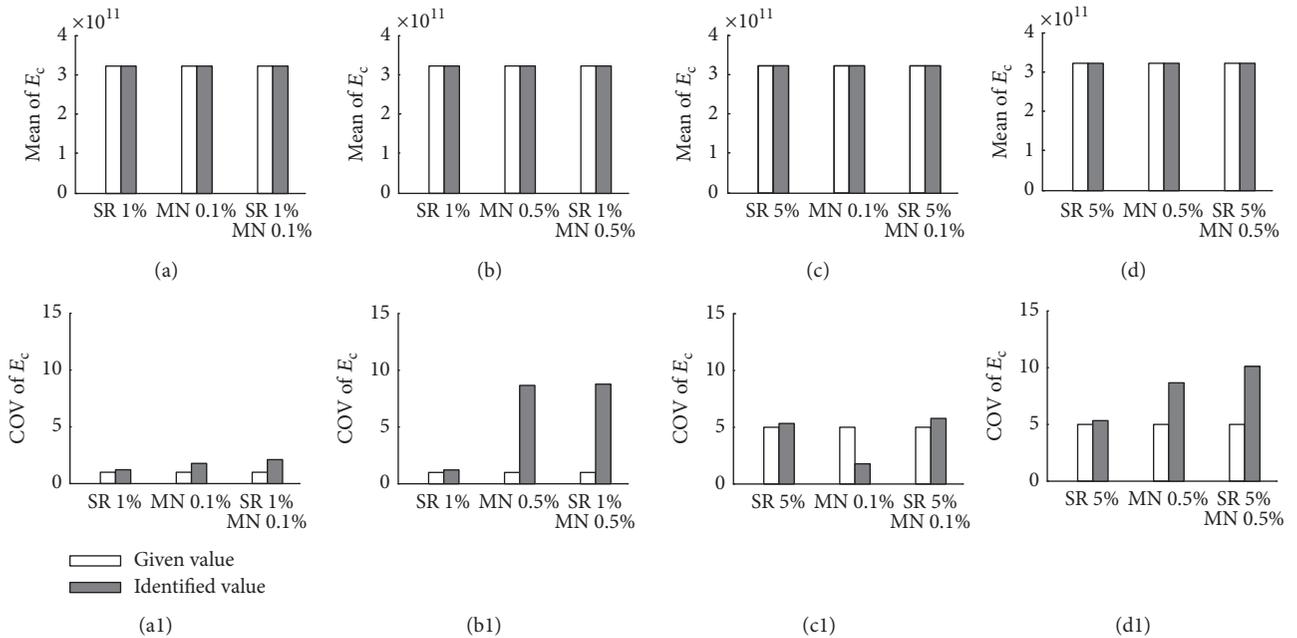


FIGURE 7: Means and COVs of elastic moduli of ceramics (E_c) for FGM along axial direction with different uncertainties.

TABLE 4: Means and COVs of material elastic moduli and volume fraction index for different boundary conditions.

Temperature	Parameters (given means/COVs)	Parameters				Means				COV's			
		CF	CC	CS	SS	CF	CC	CS	SS	CF	CC	CS	SS
T = 300 k	E_c (3.22272e11/5%)	3.2226e + 11	3.2244e11	3.2249e + 11	3.2276e + 11	6.1553%	6.1149%	6.0565%	6.0745%	6.1553%	6.1149%	6.0565%	6.0745%
	E_m (2.07788e11/5%)	2.0731e + 11	2.07223e11	2.0734e + 11	2.0733e + 11	6.2834%	6.6021%	6.8732%	6.7610%	6.2834%	6.6021%	6.8732%	6.7610%
	n (2/5%)	1.9910	1.9954	1.9934	1.9899	9.9623%	9.4556%	10.1517%	9.4319%	9.9623%	9.4556%	10.1517%	9.4319%
T = 600 k	E_c (3.04610e11/5%)	3.0465e11	3.0511e11	3.0465e11	3.0510e11	6.4379%	6.5151%	6.2227%	6.3908%	6.4379%	6.5151%	6.2227%	6.3908%
	E_m (1.90890e11/5%)	1.9084e11	1.9032e11	1.9045e11	1.9026e11	6.8224%	6.9642%	7.47109%	7.2721%	6.8224%	6.9642%	7.47109%	7.2721%
	n (2/5%)	1.9943	1.9945	1.9934	1.9918	9.7212%	9.2149%	9.6028%	9.1508%	9.7212%	9.2149%	9.6028%	9.1508%
FGM beam along its axial direction													
T = 300 k	E_c (3.22272e11/5%)	3.2158e11	3.2064e11	3.2087e + 11	3.1971e + 11	5.6988%	5.7067%	6.0838%	5.9943%	5.6988%	5.7067%	6.0838%	5.9943%
	E_m (2.07788e11/5%)	2.0775e11	2.0754e11	2.0724e + 11	2.0732e + 11	5.9243%	6.1247%	6.0863%	6.0629%	5.9243%	6.1247%	6.0863%	6.0629%
	n (2/5%)	1.9965	1.9870	1.9845	1.9782	5.9821%	6.1728%	6.0602%	5.9818%	5.9821%	6.1728%	6.0602%	5.9818%
T = 600 k	E_c (3.04610e11/5%)	3.0396e11	3.0398e11	3.0356e + 11	3.0298e + 11	6.0989%	5.9839%	6.1146%	5.9315%	6.0989%	5.9839%	6.1146%	5.9315%
	E_m (1.90890e11/5%)	1.9125e11	1.9076e11	1.9081e + 11	1.9054e + 11	6.2054%	6.6679%	6.4078%	6.5143%	6.2054%	6.6679%	6.4078%	6.5143%
	n (2/5%)	1.9989	1.9845	1.9854	1.9790	5.7743%	6.1718%	6.0829%	6.0728%	5.7743%	6.1718%	6.0829%	6.0728%

TABLE 5: Means of material elastic moduli and volume fraction index for different cases of uncertainty.

Temperature	Parameters (Given values)	SR and MN				SR only				MN only	
		Level 1 (1% SR 0.1% MN)	Level 2 (1% SR 0.5% MN)	Level 3 (5% SR 0.1% MN)	Level 4 (5% SR 0.5% MN)	1%	5%	0.1%	0.5%		
FGM beam through its thickness											
T = 300 k	E_c (3.22272e11)	3.2223e11	3.2192e11	3.2226e + 11	3.2256e11	3.2231e11	3.2197e11	3.2231e11	3.2276e11	3.2231e11	3.2276e11
	E_m (2.07788e11)	2.0769e11	2.0779e11	2.0731e + 11	2.0774e11	2.0780e11	2.0780e11	2.0780e11	2.0780e11	2.0780e11	2.0780e11
	n (2)	1.9982	1.9964	1.9910	1.9952	1.9996	1.9927	2.0018	2.0029	2.0018	2.0029
T = 600 k	E_c (3.04610e11)	3.0465e11	3.0408e11	3.0465e11	3.0398e11	3.0461e11	3.0454e11	3.0455e11	3.0426e11	3.0455e11	3.0426e11
	E_m (1.90890e11)	1.9087e11	1.9131	1.9084e11	1.9101e11	1.9088e11	1.9074e11	1.9079e11	1.9135e11	1.9079e11	1.9135e11
	n (2)	2.0002	2.0019	1.9943	1.9965	2.0001	1.9936	1.9986	2.0028	1.9986	2.0028
FGM beam along its axial direction											
T = 300 k	E_c (3.22272e11)	3.2224e11	3.2261e11	3.2158e11	3.2236e11	3.2226e11	3.2161e11	3.2247e11	3.2261e11	3.2247e11	3.2261e11
	E_m (2.07788e11)	2.0778e11	2.0779e11	2.0775e11	2.0732e11	2.0786e11	2.0768e11	2.0787e11	2.0786e11	2.0787e11	2.0786e11
	n (2)	1.9981	2.0018	1.9965	1.9989	1.9998	1.9981	2.0019	2.0054	2.0019	2.0054
T = 600 k	E_c (3.04610e11)	3.0462e11	3.0421e11	3.0396e11	3.0357e11	3.0458e11	3.0394e11	3.0467e11	3.0432e11	3.0467e11	3.0432e11
	E_m (1.90890e11)	1.9093e11	1.9080e11	1.9125e11	1.9061e11	1.9095e11	1.9076e11	1.9087e11	1.9079e11	1.9087e11	1.9079e11
	n (2)	2.0001	1.9963	1.9989	1.9955	2.0001	1.9993	1.9999	1.9959	1.9999	1.9959

TABLE 6: COVs of material elastic moduli and volume fraction index for different cases of uncertainty.

Temperature	Parameters (given values)	SR and MN				SR only			MN only	
		Level 1 (1% SR 0.1% MN)	Level 2 (1% SR 0.5% MN)	Level 3 (5% SR 0.1% MN)	Level 4 (5% SR 0.5% MN)	1%	5%	0.1%	0.5%	
FGM through its thickness										
T = 300 k	E_c (1%/5%)	3.0398%	14.0489%	6.1553%	15.3656%	1.0956%	5.5865%	2.8725%	14.2591%	
	E_m (1%/5%)	2.1465%	9.2254%	6.2834%	11.5036%	1.1123%	5.9712%	1.7987%	9.8175%	
	η (1%/5%)	4.5745%	22.0164%	9.9623%	22.8086%	1.0342%	5.2633%	4.4465%	22.0228%	
T = 600 K	E_c (1%/5%)	3.2289%	15.0723%	6.4379%	16.0919%	1.1446%	5.6354%	2.9657%	15.0258%	
	E_m (1%/5%)	2.2965%	9.9145%	6.8224%	11.6645%	1.1962%	6.2367%	1.9336%	9.8444%	
	η (1%/5%)	4.4755%	21.9856%	9.7212%	22.7138%	1.0331%	5.3643%	4.2429%	21.9586%	
FGM beam along its axial direction										
T = 300 K	E_c (1%/5%)	2.0698%	8.6736%	5.6988%	10.1014%	1.1134%	5.3245%	1.7669%	8.6012%	
	E_m (1%/5%)	1.2578%	3.2402%	5.9243%	6.2348%	1.0989%	5.3717%	0.6229%	3.0479%	
	η (1%/5%)	2.7756%	12.6854%	5.9821%	13.5611%	1.0334%	4.8631%	2.6207%	12.6327%	
T = 600 K	E_c (1%/5%)	2.1761%	9.5164%	6.0989%	11.0243%	1.1456%	5.6354%	1.8449%	9.4465%	
	E_m (1%/5%)	1.3856%	3.6823%	6.2054%	6.91654%	1.1934%	5.9669%	0.6939%	3.4742%	
	η (1%/5%)	2.7799%	13.1743%	5.7743%	14.0605%	1.0299%	4.9781%	2.5838%	13.1392%	

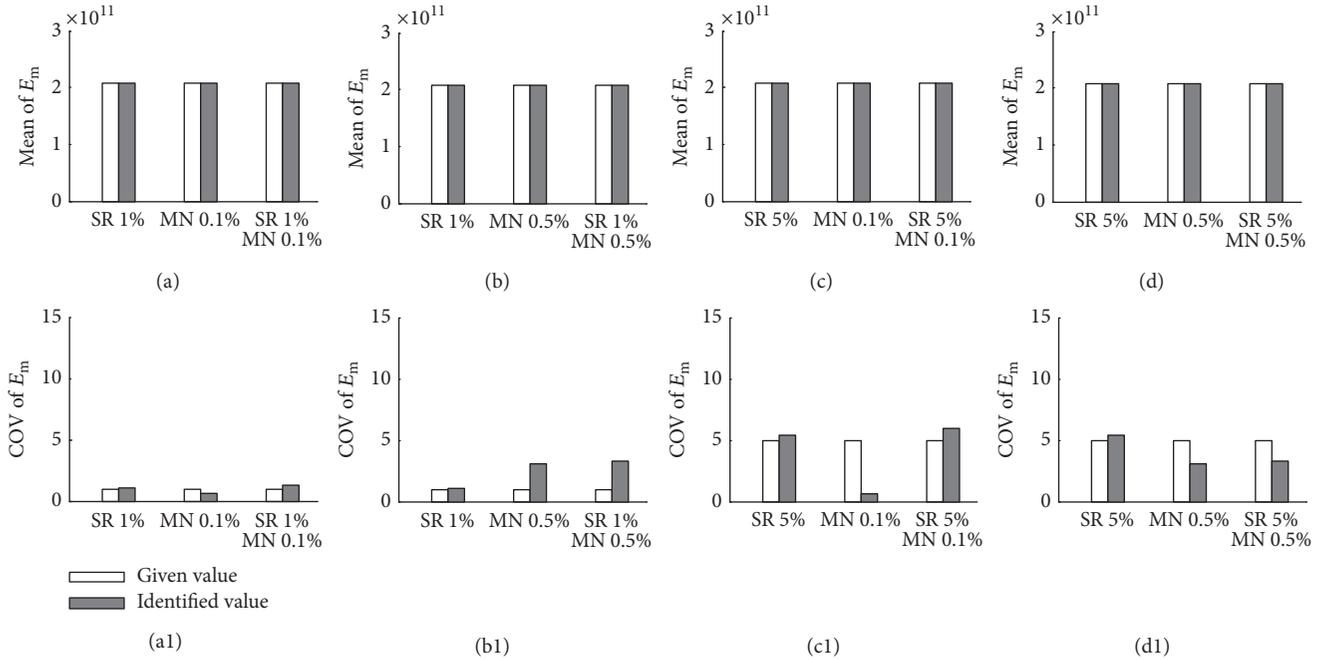


FIGURE 8: Means and COVs of elastic moduli of metal (E_m) for FGM along axial direction with different uncertainties.

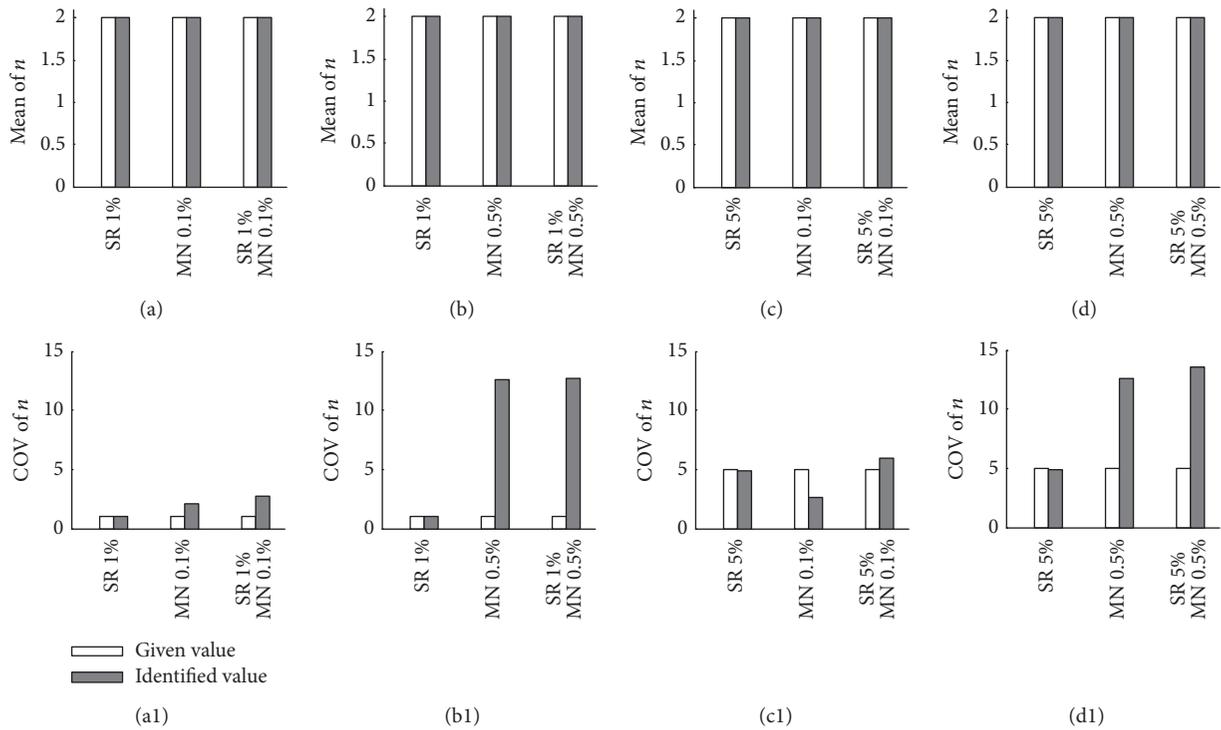


FIGURE 9: Means and COVs of volume fraction index (n) for FGM along axial direction with different uncertainties.

the identified PDFs for 300 K and 600 K, respectively, when the FGM beam along its axial direction is considered. It can be seen that the identified results for the FGM beam along its axial direction are more close to the given PDFs than those for the FGM beam through its thickness direction. Compared to the given PDFs, Figures 10(b), 11(b), and 12(b)

show the identified PDFs of the FGM beam through its thickness direction in the thermal environment of 300k with different cases of uncertainty, in which A3 denotes the given PDF, P5SR denotes the identified PDF with 5% SR (5% COVs in material elastic moduli and volume fraction index), P01MN and P05MN denote the identified PDFs for 0.1%

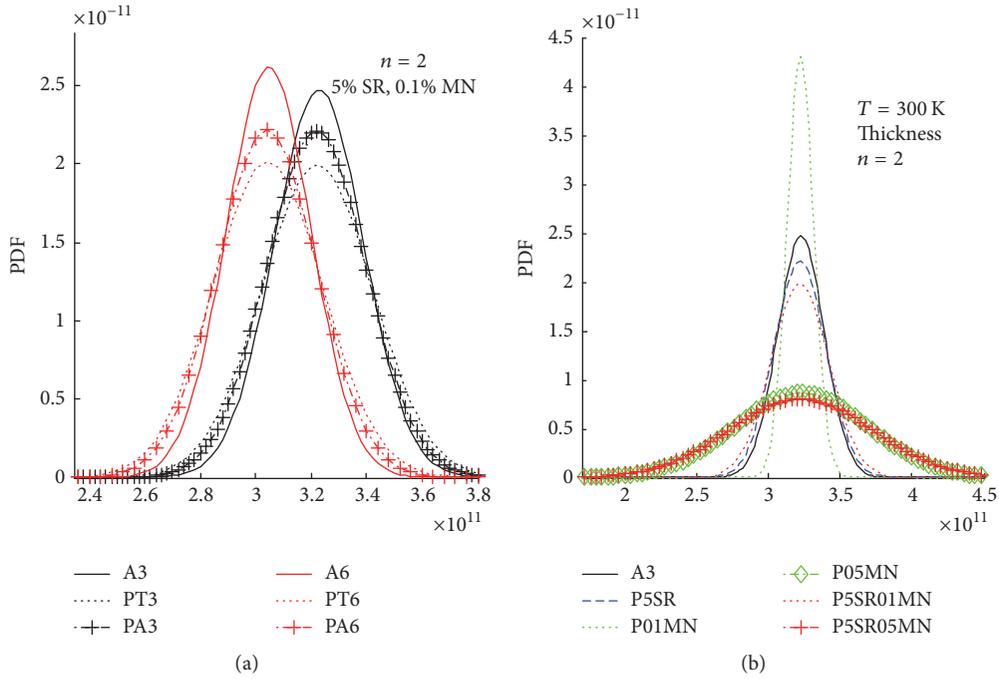


FIGURE 10: Identified PDFs of ceramics material elastic modulus for different cases.

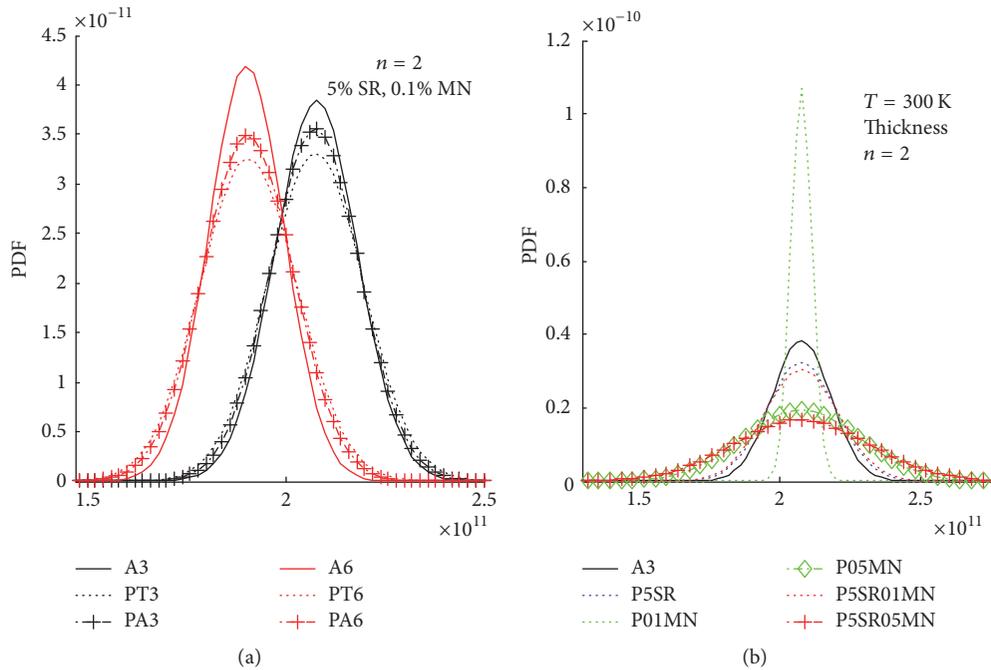


FIGURE 11: Identified PDFs of metal material elastic modulus for different cases.

MN (measurement noise) and 0.5% MN, respectively, and P5SR01MN denotes 5% SR and 0.1% MN, and P5SR05MN denotes 5% SR and 0.5% MN. It is obviously seen that the PDFs are far away from the given PDFs when MN reaches 0.5% no matter whether 5% SR is included, which also shows that the identified results of COVs are completely submerged in the measurement noise. The same situation occurs to the PDF results of all identified parameters.

7. Conclusions

Identification of the statistic numerical characteristics for constituent material properties and distribution is an important issue in mechanical behavior analysis and safety assessment of FGM structures, especially in the early stage of damage. The statistic numerical characteristics are identified for material elastic moduli and constituent volume fraction

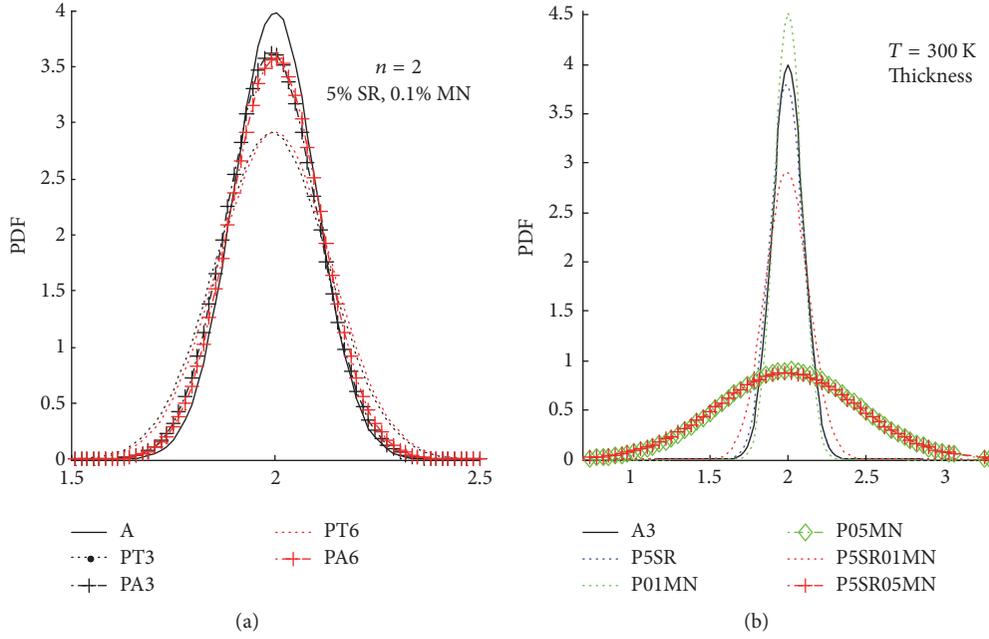


FIGURE 12: Identified PDFs of volume fraction index for different cases.

index of a random FGM beam with initial damage by using a stochastic finite element model updating-based inverse technique in this paper. A region truncation-based optimization algorithm is utilized for simplifying the optimization process and improving computational efficiency without impairing the convergence of iterative process, and a new algorithm is developed for computing the covariance between the modal parameters and the identification parameters for damaged FGM structures. The influence of structural material randomness and measurement noise on the accuracy of identification results has been investigated.

The following conclusions can be drawn from this study:

- (1) Compared to the given values and the results from MCS method, the presented inverse technique based on stochastic finite element model updating can be utilized as an alternative to identify the statistics of the material properties and volume fraction index for random damaged FGM structures, provided that the modal parameters of the corresponding FGM structures are obtained from experimental test.
- (2) Compared to the trust region method, the faster convergence of the iterative process can be guaranteed by the region truncation-based optimization. The elastic moduli of constituent materials converge to the different values in the case of different temperatures, while volume fraction indices not affected by the environment converge to the same value in the case of different temperatures.
- (3) The volume distribution has more obvious influence on the accuracy of the dispersion (COVs) of identified parameters for the FGM with material properties varying quickly (through the thickness direction)

compared to that for the FGM with material properties varying slowly (along the axial direction).

- (4) The boundary conditions have no significant influence on the convergence of iterative process as well as the identified results of means and COVs of material elastic moduli and volume fraction index.
- (5) The measurement noise plays a dominant role in the accuracy of COV identification results while the uncertainty does not bring too much change in the means of identified parameters no matter what sources the uncertainty comes from. The dispersion of material properties and volume fraction can be identified with relatively good accuracy in the presence of small measurement noise, while the COV identification results could be totally annihilated by a relatively large measurement noise.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge financial support of Science Foundation of Shaanxi Province (China) under Grant no. 2016JM1021 and National Science Foundation of China under Grant no. 11502183.

References

- [1] D. K. Jha, T. Kant, and R. K. Singh, "A critical review of recent research on functionally graded plates," *Composite Structures*, vol. 96, pp. 833–849, 2013.

- [2] H. X. Nguyen, E. Atroshchenko, H. Nguyen-Xuan, and T. P. Vo, "Geometrically nonlinear isogeometric analysis of functionally graded microplates with the modified couple stress theory," *Computers & Structures*, vol. 193, pp. 110–127, 2017.
- [3] J. N. Reddy, J. Romanoff, and J. A. Loya, "Nonlinear finite element analysis of functionally graded circular plates with modified couple stress theory," *European Journal of Mechanics - A/Solids*, vol. 56, pp. 92–104, 2016.
- [4] G. Jin, T. Ye, X. Wang, and X. Miao, "A unified solution for the vibration analysis of FGM doubly-curved shells of revolution with arbitrary boundary conditions," *Composites Part B: Engineering*, vol. 89, pp. 230–252, 2016.
- [5] A.-Y. Tang, J.-X. Wu, X.-F. Li, and K. Y. Lee, "Exact frequency equations of free vibration of exponentially non-uniform functionally graded Timoshenko beams," *International Journal of Mechanical Sciences*, vol. 89, pp. 1–11, 2014.
- [6] G. Giunta, D. Crisafulli, S. Belouettar, and E. Carrera, "A thermo-mechanical analysis of functionally graded beams via hierarchical modelling," *Composite Structures*, vol. 95, pp. 676–690, 2013.
- [7] N. L. Shegokar and A. Lal, "Stochastic nonlinear bending response of piezoelectric functionally graded beam subjected to thermoelectromechanical loadings with random material properties," *Composite Structures*, vol. 100, pp. 17–33, 2013.
- [8] D. Wu, W. Gao, K. Gao, and F. Tin-Loi, "Robust safety assessment of functionally graded structures with interval uncertainties," *Composite Structures*, vol. 180, pp. 664–685, 2017.
- [9] M. Talha and B. N. Singh, "Stochastic perturbation-based finite element for buckling statistics of FGM plates with uncertain material properties in thermal environments," *Composite Structures*, vol. 108, no. 1, pp. 823–833, 2014.
- [10] A. Lal, H. Neeranjan Singh, and N. L. Shegokar, "FEM model for stochastic mechanical and thermal postbuckling response of functionally graded material plates applied to panels with circular and square holes having material randomness," *International Journal of Mechanical Sciences*, vol. 62, no. 1, pp. 18–33, 2012.
- [11] N. L. Shegokar and A. Lal, "Stochastic finite element nonlinear free vibration analysis of piezoelectric functionally graded materials beam subjected to thermo-piezoelectric loadings with material uncertainties," *Meccanica*, vol. 49, no. 5, pp. 1039–1068, 2014.
- [12] Y. Xu, Y. Qian, J. Chen, and G. Song, "Stochastic dynamic characteristics of FGM beams with random material properties," *Composite Structures*, vol. 133, pp. 585–594, 2015.
- [13] Z. Yu and F. Chu, "Identification of crack in functionally graded material beams using the p-version of finite element method," *Journal of Sound and Vibration*, vol. 325, no. 1-2, pp. 69–84, 2009.
- [14] M. Amirpour, R. Das, and S. Bickerton, "An elasto-plastic damage model for functionally graded plates with in-plane material properties variation: Material model and numerical implementation," *Composite Structures*, vol. 163, pp. 331–341, 2017.
- [15] I. Chiong, E. T. Ooi, C. Song, and F. Tin-Loi, "Computation of dynamic stress intensity factors in cracked functionally graded materials using scaled boundary polygons," *Engineering Fracture Mechanics*, vol. 131, pp. 210–231, 2014.
- [16] Z. R. Lu, X. X. Lin, Y. M. Chen, and M. Huang, "Hybrid sensitivity matrix for damage identification in axially functionally graded beams," *Applied Mathematical Modelling*, vol. 41, pp. 604–617, 2017.
- [17] Z.-G. Zhou, B. Wang, and Y.-G. Sun, "Investigation of the dynamic behavior of a finite crack in the functionally graded materials by use of the Schmidt method," *Wave Motion*, vol. 39, no. 3, pp. 213–225, 2004.
- [18] C.-H. Xia and L. Ma, "Dynamic behavior of a finite crack in functionally graded materials subjected to plane incident time-harmonic stress wave," *Composite Structures*, vol. 77, no. 1, pp. 10–17, 2007.
- [19] R. Bagheri, M. Ayatollahi, and E. Asadi, "Dynamic fracture analysis of multiple defects in an imperfect FGM coating-substrate layers," *International Journal of Mechanical Sciences*, vol. 75, pp. 55–65, 2013.
- [20] H. Egner, M. Juchno, M. Kula, and J. Skrzypek, "Numerical analysis of FG and TBC systems based on thermo-elasto-plastic- damage model," *Journal of Thermal Stresses*, vol. 30, no. 9-10, pp. 977–1001, 2007.
- [21] D. Gayen, D. Chakraborty, and R. Tiwari, "Finite element analysis for a functionally graded rotating shaft with multiple breathing cracks," *International Journal of Mechanical Sciences*, vol. 134, pp. 411–423, 2017.
- [22] T. Nakamura, T. Wang, and S. Sampath, "Determination of properties of graded materials by inverse analysis and instrumented indentation," *Acta Materialia*, vol. 48, no. 17, pp. 4293–4306, 2000.
- [23] G. Bolzon, G. Maier, and M. Panico, "Material model calibration by indentation, imprint mapping and inverse analysis," *International Journal of Solids and Structures*, vol. 41, no. 11-12, pp. 2957–2975, 2004.
- [24] G. R. Liu, X. Han, Y. G. Xu, and K. Y. Lam, "Material characterization of functionally graded material by means of elastic waves and a progressive-learning neural network," *Composites Science and Technology*, vol. 61, no. 10, pp. 1401–1411, 2001.
- [25] X. Han and G. R. Liu, "Computational inverse technique for material characterization of functionally graded materials," *AIAA Journal*, vol. 41, no. 2, pp. 288–295, 2003.
- [26] B. Rahmani, F. Mortazavi, I. Villemure, and M. Levesque, "A new approach to inverse identification of mechanical properties of composite materials: Regularized model updating," *Composite Structures*, vol. 105, pp. 1160–125, 2013.
- [27] K. Sun, Y. Zhao, and H. Hu, "Identification of temperature-dependent thermal—structural properties via finite element model updating and selection," *Mechanical Systems and Signal Processing*, vol. 52-53, pp. 147–161, 2015.
- [28] A. K. Mishra and S. Chakraborty, "Development of a finite element model updating technique for estimation of constituent level elastic parameters of FRP plates," *Applied Mathematics and Computation*, vol. 258, pp. 84–94, 2015.
- [29] A. K. Mishra and S. Chakraborty, "Inverse detection of constituent level elastic parameters of FRP composite panels with elastic boundaries using finite element model updating," *Ocean Engineering*, vol. 111, pp. 358–368, 2016.
- [30] J. N. Reddy and C. D. Chin, "Thermomechanical analysis of functionally graded cylinders and plates," *Journal of Thermal Stresses*, vol. 21, no. 6, pp. 593–626, 1998.
- [31] F. J. Ferrante and L. L. Graham-Brady, "Stochastic simulation of non-Gaussian/non-stationary properties in a functionally graded plate," *Computer Methods Applied Mechanics and Engineering*, vol. 194, no. 12-16, pp. 1675–1692, 2005.
- [32] N. Abu Husain, H. Haddad Khodaparast, and H. Ouyang, "Parameter selection and stochastic model updating using perturbation methods with parameter weighting matrix assignment," *Mechanical Systems and Signal Processing*, vol. 32, pp. 135–152, 2012.

- [33] P. G. Bakir, E. Reynders, and G. De Roeck, "Sensitivity-based finite element model updating using constrained optimization with a trust region algorithm," *Journal of Sound and Vibration*, vol. 305, no. 1-2, pp. 211–225, 2007.
- [34] Y. Xu, Y. Qian, G. Song, and K. Guo, "Damage detection using finite element model updating with an improved optimization algorithm," *Steel and Composite Structures*, vol. 19, no. 1, pp. 191–208, 2015.

Research Article

Gear Fault Diagnosis in Variable Speed Condition Based on Multiscale Chirplet Path Pursuit and Linear Canonical Transform

Xu Shuiqing ^{1,2}, Zhang Ke ², Chai Yi,² He Yigang,¹ and Feng Li ²

¹College of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China

²College of Automation, Chongqing University, Chongqing 400044, China

Correspondence should be addressed to Zhang Ke; smeta@163.com

Received 8 November 2017; Revised 15 January 2018; Accepted 7 February 2018; Published 11 March 2018

Academic Editor: Chen Lu

Copyright © 2018 Xu Shuiqing et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The vibration signals analysis is a very effective and reliable method for detecting the gear failures. Because the vibration signals acquired from the gear in the variable speed condition often contain more useful fault information, the analysis of the gear vibration signals during the variable speed condition has been a hot research topic. In this paper, a method based on the multiscale chirplet path pursuit (MSCPP) and the linear canonical transform (LCT) has been applied to diagnose the gear fault in the variable speed condition for the first time. First, by using the MSCPP method to estimate the instantaneous meshing frequency, the suitable signal segment approximation to the acceleration or deceleration process can be selected. Then, because the LCT is a novel and efficient nonstationary signals analysis tool, the optimal LCT spectrum of the selected signal has been attained to diagnose the gear faults based on the properties of the LCT. In addition, the simulations and the experimental evaluation are provided to verify the effectiveness of the proposed method.

1. Introduction

The gear is an important device and has a wide range of applications in industry. However, owing to abrasion and other reasons, the gears may have many kinds of faults, such as pitting, chipping, and the serious crack [1–5]. When the gear causes a local fault, the amplitude and phase of the vibration signal of the gear are modulated [2, 3, 6, 7]. So the vibration signals obtained from the gear can reflect the gear's state very well. The vibration signals analysis also has been the extremely effective and reliable method for detecting the gear failures [8–11]. Meanwhile, on the one hand, the gear vibration signals obtained from the variable speed condition, that is, the acceleration or deceleration processes, often can contain more fault information compared to the stationary processes, which can detect the gear faults earlier. On the other hand, the gear vibration signals obtained from the acceleration or deceleration processes are nonstationary signals, which have low signal-to-noise ratio (SNR) in practice [11–14], which make it difficult to obtain

the gear vibration signals' fault features. Hence, the diagnosis of gear faults by analyzing the gear vibration signals obtained from the variable speed condition has been a hot research topic recently.

A number of failure diagnosis methods have been used to diagnose the gear faults in variable speed condition, for example, the traditional time frequency analysis methods, self-adaptive signal processing methods, and data driven methods [4–12, 15–25]. The traditional time frequency analysis methods, such as the short time Fourier transform [14], the Wigner-Ville distribution [10, 12], and the wavelet transform [2], will result in spectral aliasing, cross term interference, and low resolution because the vibration signals obtained from the fault gears in practice are nonstationary with low SNR [4]. The self-adaptive signal processing methods, for example, the empirical mode decomposition [5] and the local mean decomposition [17], will lead to over envelope, end effects, and distorted components, respectively. In addition, the data driven methods, such as the Elman neural network [24] and support vector machine [22], require a lot of data

for training and classification. To solve these problems, some other data driven methods have been introduced, such as the self-organizing maps (SOM) method [26–28]. In [26], the self-organizing feature map neural network has been used to diagnose the faults of the wind turbine’s converter. In [27], the SOM and minimum quantization error (MQE) method have been selected to achieve degradation assessment and fault localization. In [28], an intelligent approach based on the SOM method has been presented to machine component health prognostics. Although the SOM method does not need a lot of data, it still needs some a priori knowledge. Therefore, the diagnosis of gear faults in acceleration and deceleration processes remains an open research field and new signal analysis tools for nonstationary vibration signals are needed.

The linear canonical transform (LCT) is a generalization of the Fourier transform (FT) and the fractional Fourier transform (FRFT), which has four-parameter family of linear integral transform [29, 30]. It performs an affine mapping of the time frequency distribution of the signal because it has additional degrees of freedom [31, 32]. So, the LCT is more flexible and suitable for processing nonstationary signals, particularly in the linear frequency modulated (LFM) signals [33–37]. Simultaneously, the gear vibration signals obtained from the variable speed condition can be approximated to the multicomponent LFM signals within a short period of time. From the above analysis, because the vibration signals attained from the variable speed condition are nonstationary and the LCT has advantages in nonstationary signals processing, it is therefore worthwhile to explore the diagnosis of gear faults by using the LCT.

However, in variable speed condition, the gear vibration signals contain many different components, which are not approximated to the LFM signals. The vibration signals acquired from the acceleration or deceleration processes of the gears are often a short period of the gear vibration signals in variable speed condition. In order to apply the LCT method to diagnosis of the gear faults in variable speed condition, the vibration signals acquired from the acceleration or deceleration processes should be selected at first. In practice, it is difficult to only obtain the gear vibration signals during acceleration or deceleration processes directly. Nevertheless, the shaft rotational frequency (SRF) is time varying in the acceleration or deceleration processes, which can be seen as an indicator to select the acceleration or deceleration processes [1]. Hence, we can obtain the acceleration or deceleration time by estimating the instantaneous frequency (IF) of the gear vibration signals in variable speed condition. In addition, the optimal LCT parameters of the LFM signals are also determined by gradient of the IF [30]. For these reasons, we need to estimate the IF of the gear vibration signals in variable speed condition at first.

In this paper, the multiscale chirplet path pursuit (MSCPP) has been used in the estimation of the IF of the gear vibration signals, which is a widely used and efficient method for IF estimation [38]. The MSCPP method divides the time length of the analysis signal into a series of dynamic time support areas in binary form, finding atom with the largest correlation in each dynamic support region from the defined chirplet atom library, and the selected chirplet

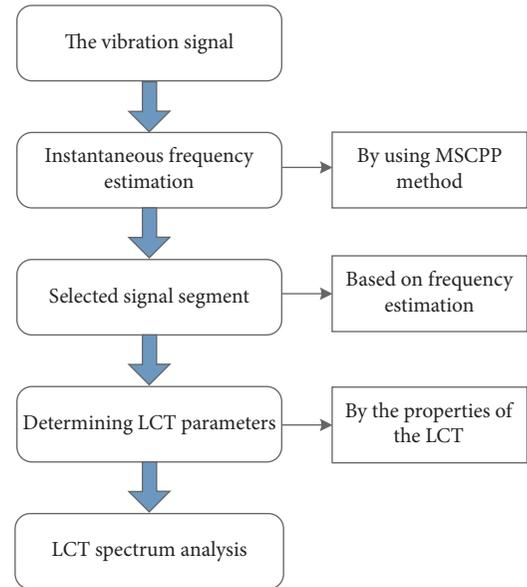


FIGURE 1: The block diagram of the proposed method.

atoms are connected one by one by using the principle of the best connection [38, 39]. Thus, it can adaptively find the instantaneous frequency trend line with the largest correlation in the signal [39]. Compared with other instantaneous frequency methods, the MSCPP method can choose shorter chirplet atoms flexibly and can effectively suppress noise interference. In addition, the MSCPP method does not have any a priori knowledge. So the instantaneous frequency of the vibration signals obtained from the variable speed condition can be estimated by the MSCPP method. Based on this, the vibration signals obtained from the acceleration or deceleration processes in a short period of time can be picked out.

In this paper, the MSCPP method and the LCT have been applied to analyze the vibration signals to diagnose the gear faults in variable speed condition for the first time. First, by using the MSCPP, the IF of the vibration signals obtained from the variable speed condition can be estimated. Then, based on the IF of the vibration signals, we can choose the suitable signal segment, which can be seen as the acceleration or deceleration processes. In addition, according to the gradient of the instantaneous frequency of the selected signal segment, the optimal LCT parameters can be obtained based on the properties of the LCT [30]. Finally, the optimal LCT spectrum of the selected signal is presented, which can be used to diagnose the gear fault. The block diagram of the proposed method is presented in Figure 1 vividly.

The remaining parts of this article are organized as follows. In Section 2, the preliminaries of the MSCPP and the LCT are introduced. In Section 3, the application of the MSCPP and the LCT to the simulated vibration signal is presented, which shows that the proposed method is effective. The experimental evaluations have been provided in Section 4. Section 5 concludes this paper.

2. Preliminaries

2.1. The Multiscale Chirplet Path Pursuit Method. The MSCPP method was first introduced in [38]; in this method, if the IF of chirplet atoms is linear, they can be used to adaptively piecewise approximate the nonstationary signals. For example, the nonstationary signal in the following is considered.

$$y(t) = A \cos [w(t) + \varphi] + n(t), \quad (1)$$

where A , φ , $n(t)$, and $w(t)$ represent the amplitude, the initial phase, the noise, and the successive and derivable instantaneous phase of the signal, respectively. In this algorithm, the chirplet atoms are obtained from the chirplet dictionary that can be written as a set of functions as follows [38].

$$f_{m,n,I}(t) = |I|^{-1/2} e^{i(mt^2/2+nt)} L_I(t), \quad (2)$$

where m is the slope coefficient and n is the offset coefficient. According to (2), we can easily have that the IF of a chirplet is linear and equal to $mt + n$. Based on the sampling theorem, we know that $mt + n$ should be less than a half of the sampling rate f_s . I is the dyadic time span, which has the binary scale division and can be defined as $I = \{I_{j,k} \mid j = 0, 1, \dots, \log_2 N - 1, k = 0, 1, \dots, 2^{j-1}\}$. In this equation, $I_{j,k} = [k2^{-j}T, (k+1)2^{-j}T]$. T and N are the total sampling time and the number of samples, respectively. j is the scale coefficients and $j = 0, 1, \dots, \log_2 N - 1$, $k = 0, 1, \dots, 2^{j-1}$. $L_I(t)$ is a rectangular window function, which is 1 when $t \in I$ and 0 when $t \notin I$. $|I|^{-1/2}$ is the normalization factor that makes $\|f_{m,n,I}(t)\|_2 = 1$.

In each time interval, by computing the maximum projection coefficient $C_{I_{j,k}}$, where $C_{I_{j,k}}$ equals $C_{I_{j,k}} = \max \langle x, f_{m,n,I_{j,k}}(t) \rangle$ ($\langle \cdot \rangle$ stands for the interior product) and includes the amplitude and initial phase information of the signal component in the time interval $I_{j,k}$, the chirplet atom which has the highest correlation to signal $y(t)$ can be obtained from the chirplet dictionary. Suppose $S_{I_{j,k}}$ is the signal component expressed by $C_{I_{j,k}}$ in the time interval $I_{j,k}$, then $S_{I_{j,k}}$ can be written in the following [38]:

$$S_{I_{j,k}}(t) = \left| C_{I_{j,k}} \right| e^{-i(mt^2+nt)+\angle C_{I_{j,k}}} I_{j,k}. \quad (3)$$

2.2. The Linear Canonical Transform. The LCT of a signal $f(t)$ with parameter $A = (a, b, c, d)$, represented as $L_f^A(u)$, is defined by [29]

$$L_f^A(u) = \begin{cases} \int_{-\infty}^{\infty} f(t) K_A(u, t) dt & b \neq 0 \\ \sqrt{d} e^{j(cdu^2/2)} f(du) & b = 0, \end{cases} \quad (4)$$

where a, b, c, d are real numbers satisfying $ad - bc = 1$. In addition, the kernel $K_A(u, t)$ is given by

$$K_A(u, t) = C_A e^{j((a/2b)t^2 - ut/b + (d/2b)u^2)} \quad (5)$$

and $C_A = \sqrt{1/j2\pi b}$. According to (4), it can easily be obtained that the LCT has three degrees of freedom and when the

LCT parameters are certain, the kernel $K_A(u, t)$ of the LCT is a LFM signal; thus the LCT of a LFM signal with suitable parameters can be a dirac function [31]. Therefore, the LCT is more flexible and suitable for processing nonstationary signals, especially for the LFM signals.

3. The Proposed Method of the Simulated Vibration Signal

When the gear causes a partial failure, the amplitude and phase of the vibration signal of the gear are modulated, which are periodic with the gear's rotation frequency [1]. The vibration signal $x(t)$ attained from a pair of meshing gears with tooth fault could be represented as [1]

$$\begin{aligned} x(t) &= \sum_{m=0}^M X_m (1 + a_m(t)) \cos(2\pi m k f_s t + \phi_m) \\ &= \sum_{m=0}^M X_m \left[1 + \sum_{n=0}^N A_{mn} \cos(2\pi n f_s t + \sigma_{mn}) \right] \\ &\quad \times \cos(2\pi m k f_s t + \phi_m), \end{aligned} \quad (6)$$

where $a_m(t)$ is the amplitude modulating function and M represents the amount of tooth meshing harmonics. X_m and ϕ_m indicate the m th meshing harmonic's amplitude and phase, separately [1]. $k f_s$ is the meshing frequency, k expresses the gear teeth's amount, and f_s stands for the SRF [1]. N represents sidebands' amount around the tooth meshing harmonics. A_{mn} and σ_{mn} are the amplitudes and phases at the n th sidebands of the amplitude modulated signals around the m th meshing harmonic, separately [2, 4].

The SRF f_s in (6) is time varying and the vibration signal can be approximated to the LFM signal when the instantaneous rotational speed is ascending and descending straight in the acceleration and deceleration, respectively. To verify the proposed algorithm is effective for the vibration signals; next, a signal is considered to simulate the vibration signal obtained from the acceleration process as follows based on (6):

$$x(t) = [1 + \cos(2\pi f_m t)] \cos(2\pi \cdot 10 \cdot f_m t + \varphi(t)), \quad (7)$$

where f_m stands for the SRF in the simulated vibration signal, which is chosen as $f_m = 2t + 30$ when $0 \leq t < 1.5$ and equal to 33 when $1.5 \leq t \leq 2$. The sampling frequency is 2000 Hz and the number of the sample points is set as 4096. The time interval is set as 0 s to 2 s. Moreover, because the vibration signals attained from the gear in variable speed condition are usually with low SNR, $\varphi(t)$ is selected as the Gaussian white noise with the SNR is -4 dB. Based on the above, Figure 2 depicts the signal shown in (7) without the noise being depicted in it. The simulated vibration signal based on (7) is presented in Figure 3. Figure 4 shows the FT spectrum of the simulated vibration signal. According to Figure 3, because the signal contains noise, any useful information can not be obtained directly. Analogously, since the simulated vibration signal is nonstationary and has noise, this leads to the FT spectrum in Figure 4 that has spectral aliasing and blurring.

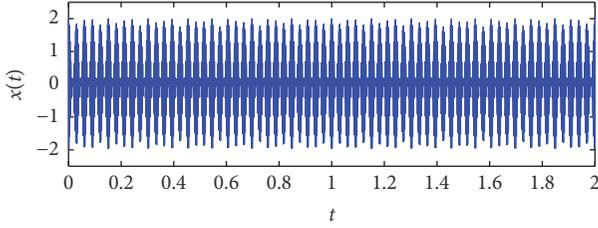


FIGURE 2: The simulated gear vibration signal without noise.

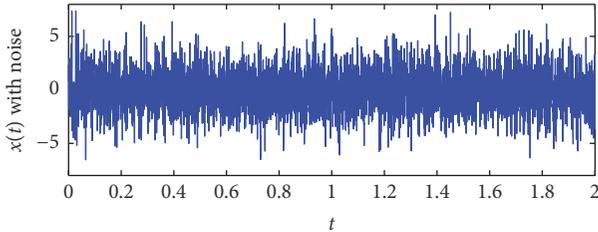


FIGURE 3: The simulated gear vibration signal with noise.

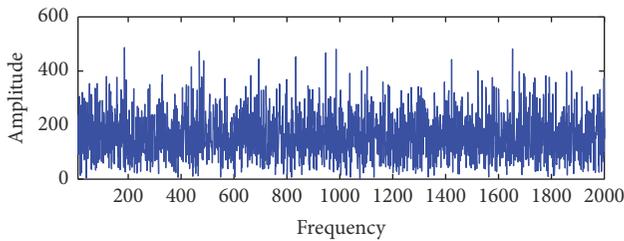


FIGURE 4: The FT of the simulated gear vibration signal with noise.

Now, to obtain the fault features from the simulated gear vibration signal and diagnose the gear faults, the MSCPP and the LCT are applied. According to the above analysis, we first use the MSCPP method to estimate the instantaneous frequency of the simulated gear vibration signal. Because the sampling frequency is 2000 Hz, we let the search range and the search step lengths of the frequency slope be -100 to 100 Hz/s and 1 Hz/s, respectively. The search range and the search step lengths of offset coefficients are 0 to 100 Hz and 1 Hz, separately. The points of any dyadic time span is 32 . Therefore, we can obtain that $j_{\max} = \log_2(N/32)$, which is equal to 7 . Based on these, by applying the MSCPP algorithm, the IF of the simulated gear vibration signal is obtained in Figure 5. In Figure 5, the blue line and the red line represent the actual and estimated instantaneous frequency of the simulated gear vibration signal, respectively. The red line is coincided with the blue line very well in Figure 5, which shows that the MSCPP algorithm can be used to estimate IF of the simulated gear vibration signal efficiently. Based on Figure 6, we can know that the time interval from 1 s to 1.5 s can be selected to approach the acceleration process of the gear in a short period of time. Next, based on the gradient of the signal segment from 1 s to 1.5 s and the properties of the LCT [30], the optimal LCT spectrum with parameters $(0.93, 0.42, 0.34, 1.89)$ of the selected signal

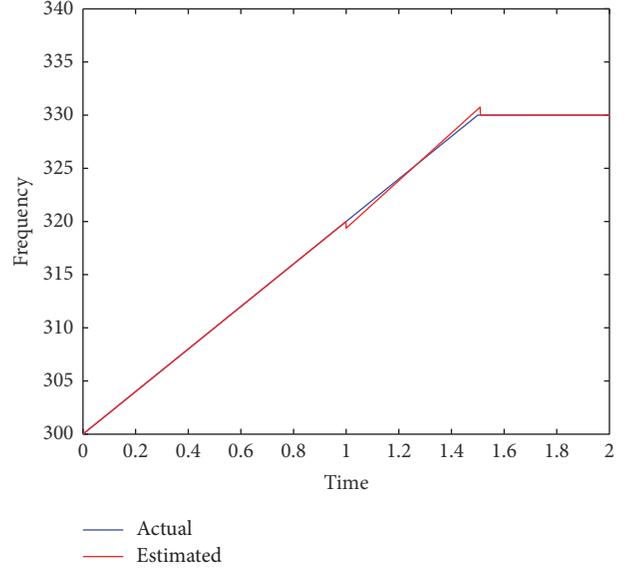


FIGURE 5: Estimated instantaneous frequency of the simulated gear vibration signal.

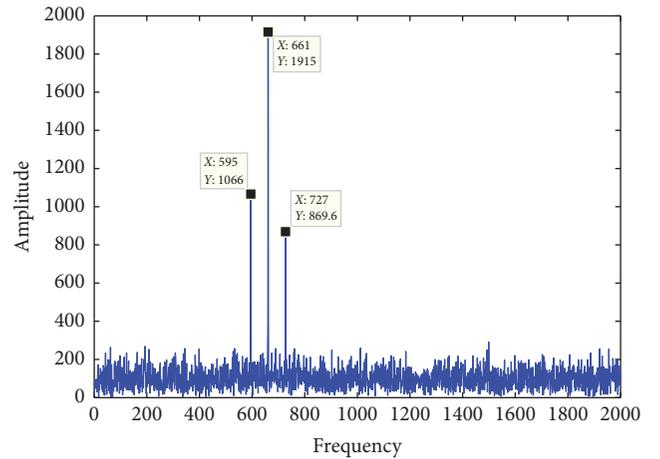


FIGURE 6: The LCT of the simulated gear vibration signal with noise.

segment is obtained in Figure 6. It is shown that Figure 6 has three peaks, which are $(595, 1066)$, $(661, 1915)$, and $(727, 869.6)$, respectively. The highest peak is $(661, 1915)$. The LCT frequency distances of each of the two adjacent peaks are 66 Hz. It is shown that the frequency of $(661, 1915)$ is almost ten times of the SRF associated with the LCT. This means that the peaks $(595, 1066)$, $(661, 1915)$, and $(727, 869.6)$ are 9 , 10 , and 11 times of the SRF associated with the LCT, separately. From the results, we can know that it is very consistent with the analysis of (7).

Moreover, since the results obtained in Figure 6 show that the range of any two adjacent peaks are equal, hence the signal presented in Figure 3 could be considered as the fault gear vibration signal based on the description in [1, 4]. Thus, the proposed method is an effective method to diagnose the gear fault according to this simulation.

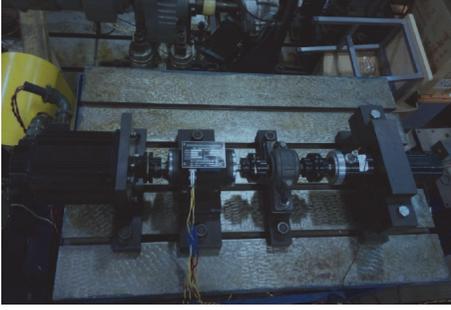


FIGURE 7: Platform of experiment.

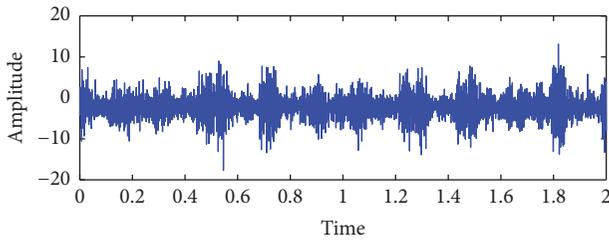


FIGURE 8: The vibration signal obtained from the normal gear.

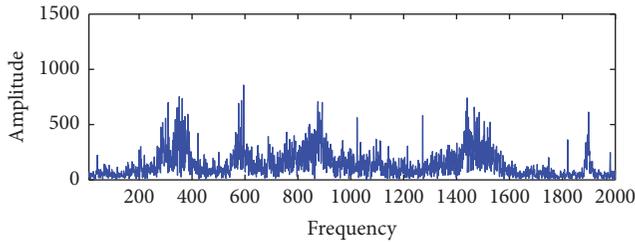


FIGURE 9: The FT of the vibration signal obtained from the normal gear.

4. Experimental Evaluation

In Section 3, the validity of the MSCPP and the LCT in gear faults diagnosis has been verified by the signal presented in (7). In order to further verify the correctness of the proposed algorithm, in this part, the gear faults of experiment setup presented in Figure 7 have been used. In this gear experiment equipment, the output gear has 75 teeth and the input pinion gear has 55 teeth. The sample frequency and the sample time of the signals obtained from this experiment equipment are 4000 Hz and 2 s, respectively.

Firstly, the vibration signal attained from a normal gear in variable speed condition is presented in Figure 8. The signal presented in Figure 8 is nonstationary since the gear signals with low SNR are obtained from variable speed condition. On this account, Figure 9 depicts the FT of the signal shown in Figure 8 which has spectrum aliasing, and it cannot be used to diagnose the gear fault. Next, by applying the method we presented in this paper, we used the MSCPP method to estimate the instantaneous frequency of the gear vibration signal at first. In the MSCPP method, the seeking

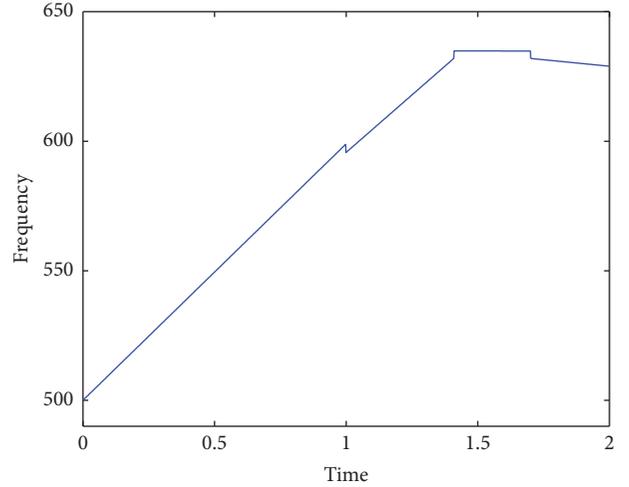


FIGURE 10: Estimated instantaneous frequency of the signal presented in Figure 13.

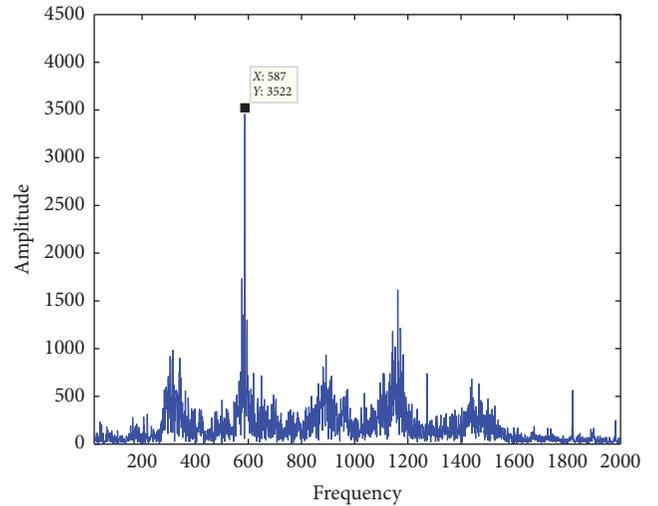


FIGURE 11: The LCT of the selected signal segment.

scope of frequency offset coefficients and slope coefficient are $[0, 2000 \text{ Hz}]$ and $[-100, 100 \text{ Hz/s}]$, respectively. The search resolutions of these are chosen as 1 Hz/s and 1 Hz, separately. The samples are 4096. Based on these, we can obtain the maximum scales of 7. Thus, the IF of the gear vibration signal is obtained in Figure 10. From Figure 10, it is easy to know that the time interval from 1 s to 1.4 s can be selected to approach the acceleration process of the gear in a short period of time. Subsequently, by utilizing the advantages of the LCT and the properties of the selected signal segment, the optimal LCT spectrum with parameters $(-0.17, 1.53, -0.98, 2.93)$ of the selected signal segment is shown in Figure 11. We can easily know that Figure 11 has only one peak. That is to say, the signal presented in Figure 8 has not amplitude modulated. Hence, the gear can be seen as a normal gear.

Then, Figure 12 shows a gear vibration signal, which is attained from the gear which has a broken tooth in variable speed condition. The FT of the signal presented

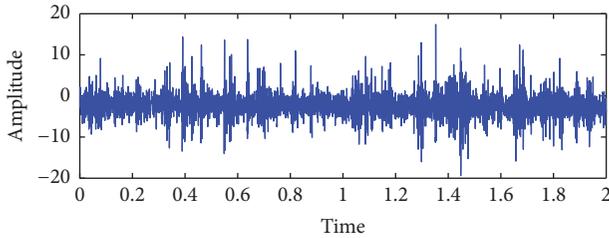


FIGURE 12: The vibration signal obtained from the fault gear.

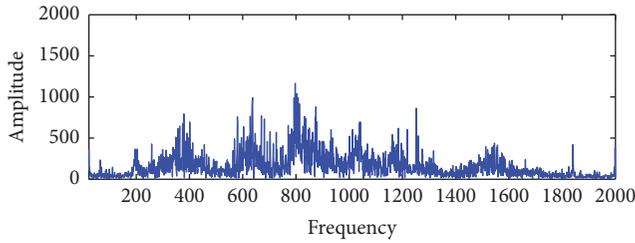


FIGURE 13: The FT of the vibration signal obtained from the fault gear.

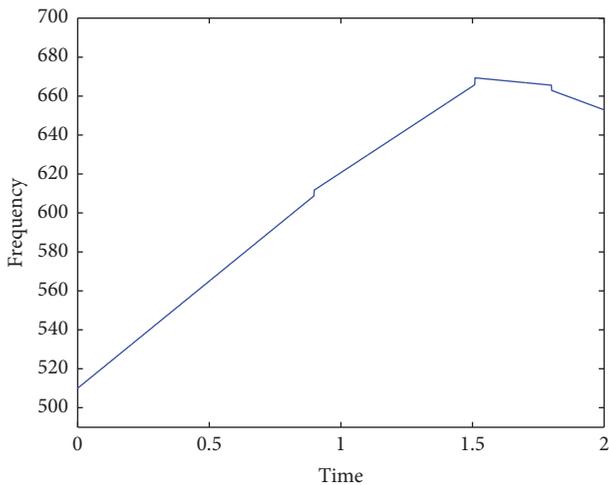


FIGURE 14: Estimated instantaneous frequency of the signal presented in Figure 17.

in Figure 12 is presented in Figure 13, which shows that it also has spectrum aliasing and cannot be used to diagnose the gear fault. Now, by applying the proposed method, the search range of frequency offset coefficients and slope coefficient are also set as $[0, 2000 \text{ Hz}]$ and $[-100, 100 \text{ Hz/s}]$. The search resolutions of these are also chosen as 1 Hz/s and 1 Hz , separately. The samples are 4096. Based on these, the estimated instantaneous frequency of the gear vibration signal can be obtained in Figure 14 based on the MSCPP method. From Figure 14, the time interval from 1 s to 1.5 s can be selected to approach the acceleration process of the gear in a short period of time. Subsequently, according to the above analysis, the optimal LCT spectrum with parameters $(-0.51, 0.87, -1.47, 0.56)$ of the selected gear vibration signal segment can be obtained in Figure 15. We can know that

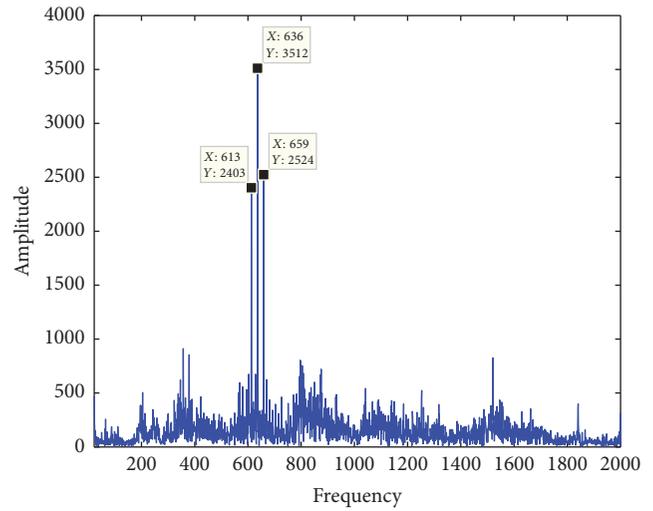


FIGURE 15: The LCT of the selected signal segment from the fault gear.

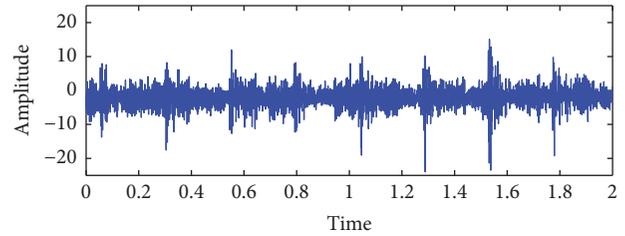


FIGURE 16: Another vibration signal obtained from the fault gear.

Figure 15 has three peaks, that is, $(613, 2403)$, $(636, 3512)$, and $(659, 2524)$. Since $(636, 3512)$ is the highest and there are 55 teeth in the pinion gear, thus the input SRF associated with the LCT can be expressed as $636 \text{ Hz}/55 = 11.56 \text{ Hz}$ [4], which is equal to half of the frequency distance. That is to say, the meshing frequency is twofold of the SRF. Based on the above analysis, the gear should be considered to have fault happened.

In addition, in order to further verify the correctness of the proposed algorithm, Figure 16 presents another gear vibration signal, which is attained from the fault gear in different variable speed condition. The FT of the signal presented in Figure 16 is presented in Figure 17, which also shows that it cannot be used to diagnose the gear fault. Now, by applying the proposed method, similar to the above cases, the largest scale is also 7. Then, the estimated IF of the gear vibration signal has been obtained in Figure 18 based on the MSCPP method. From Figure 18, the time interval from 1 s to 1.4 s can be selected to approach the deceleration process of the gear in a short period of time. Then, the optimal LCT spectrum with parameters $(1.53, 0.57, 0.42, 0.81)$ of the selected gear vibration signal segment can be obtained in Figure 19. It is shown that Figure 15 also has three peaks $(591, 2484)$, $(613, 3543)$, and $(635, 2103)$. Since $(613, 3543)$ is the highest and there are 55 teeth in the pinion gear, thus the input SRF associated with the LCT can be expressed as

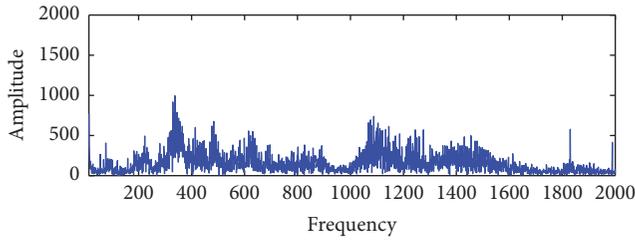


FIGURE 17: The FT of the vibration signal presented in Figure 16.

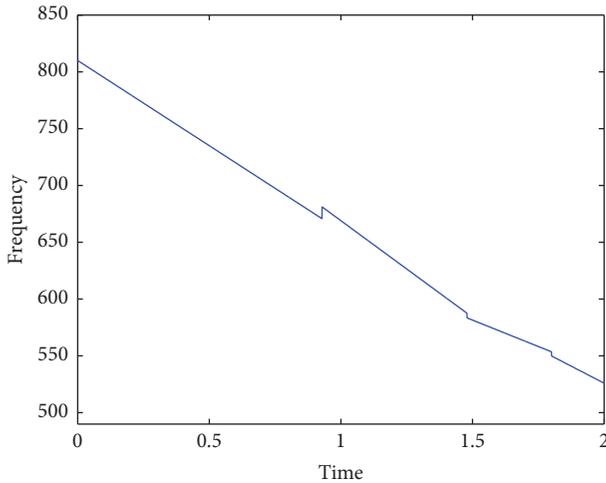


FIGURE 18: Estimated instantaneous frequency of the signal presented in Figure 16.

613 Hz/55 = 11.14 Hz. That is to say, the meshing frequency is twofold of the SRF in the LCT domain. Thus, it can be said that the gear also have fault that occurred.

Based on the normal case and two fault cases, it is shown that the method proposed in this can diagnose the gear faults in variable speed condition. However, the MSCPP and LCT method only can show that the gear faults happened, and our research directions will be the diagnosis of the types and the severity of the gear faults.

5. Conclusion

In this paper, a method based on the MSCPP and the LCT has been applied to diagnose the gear faults in the variable speed condition for the first time. Firstly, the preliminaries of the MSCPP and the LCT have been presented. Then, the proposed method of the simulated gear vibration signals have been showed. At last, in order to further verify the correctness of the proposed algorithm, the diagnosis of actual gear vibration signals also has been presented. It is indicated that the proposed method can diagnose the gear faults available. In the future, the diagnosis of early gear faults, intermittent gear faults, and multiple gear faults also will be our future research directions.

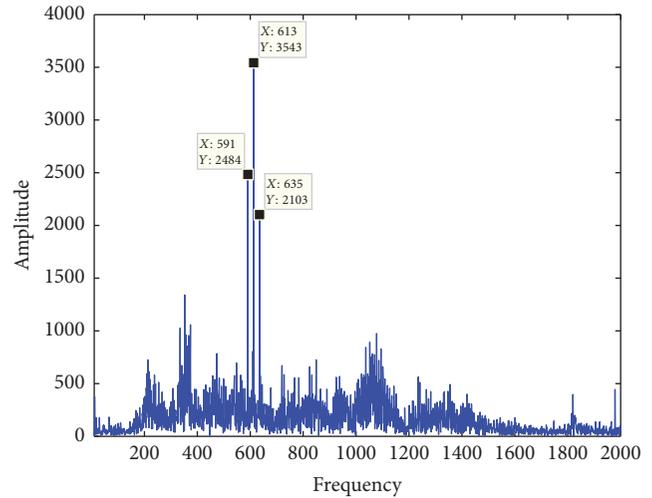


FIGURE 19: The LCT of the selected signal segment presented in Figure 16.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61374135, 61633005, 61673076, and 51637004), the National Key Research and Development Plan: Important Scientific Instruments and Equipment Development (2016YFF0102200), and Central Military Equipment Development Department Pre-Research Project (41402040301).

References

- [1] P. D. McFadden, "Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, vol. 108, no. 2, pp. 165–170, 1986.
- [2] X. Fan and M. J. Zuo, "Gearbox fault detection using Hilbert and wavelet packet transform," *Mechanical Systems and Signal Processing*, vol. 20, no. 4, pp. 966–982, 2006.
- [3] J. Luo, D. Yu, and M. Liang, "Application of multi-scale chirplet path pursuit and fractional Fourier transform for gear fault detection in speed up and speed-down processes," *Journal of Sound and Vibration*, vol. 331, no. 22, pp. 4971–4986, 2012.
- [4] P. Girdhar and C. Scheffer, *Practical Machinery Vibration Analysis and Predictive Maintenance*, vol. 46, Elsevier, 11 edition, 2004.
- [5] Y. Yang, Y. He, J. Cheng, and D. Yu, "A gear fault diagnosis using Hilbert spectrum based on MODWPT and a comparison with EMD approach," *Measurement*, vol. 42, no. 4, pp. 542–551, 2009.
- [6] M. Feldman, "Hilbert transform in vibration analysis," *Mechanical Systems and Signal Processing*, vol. 25, no. 3, pp. 735–802, 2011.
- [7] J. He, S. Yang, and C. Gan, "Unsupervised fault diagnosis of a gear transmission chain using a deep belief network," *Sensors*, vol. 17, no. 7, article no. 1564, 2017.

- [8] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Transactions on Signal Processing*, vol. 41, no. 10, pp. 3024–3051, 1993.
- [9] H. Li, J. Zhao, X. Zhang, and H. Teng, "Gear fault diagnosis and damage level identification based on Hilbert transform and Euclidean distance technique," *Journal of Vibro Engineering*, vol. 16, no. 8, pp. 4137–4151, 2014.
- [10] M. Liang and I. Soltani Bozchalooi, "An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection," *Mechanical Systems and Signal Processing*, vol. 24, no. 5, pp. 1473–1494, 2010.
- [11] J. Mei, J. Jia, R. Zeng, B. Zhou, and H. Zhao, "A multi-order FRFT self-adaptive filter based on segmental frequency fitting and early fault diagnosis in gears," *Measurement*, vol. 91, pp. 532–540, 2016.
- [12] P. F. Odgaard and J. Stoustrup, "Gear-box fault detection using time-frequency based methods," *Annual Reviews in Control*, vol. 40, pp. 50–58, 2015.
- [13] V. Sharma and A. Parey, "Gear crack detection using modified TSA and proposed fault indicators for fluctuating speed conditions," *Measurement*, vol. 90, pp. 560–575, 2016.
- [14] X. Tang, Y. Guo, Y. Ding, and H. Zheng, "Application of rolling element bearing envelope analysis based on short time Fourier transition and independent components analysis," *Journal of Mechanical Strength*, vol. 34, no. 2, pp. 1–5, 2012.
- [15] L. Shi, Y. Zhang, and W. Mi, "Application of Wigner-Ville-distribution-based spectral kurtosis algorithm to fault diagnosis of rolling bearing," *Journal of Vibration Measurement & Diagnosis*, vol. 31, no. 1, pp. 27–31, 2011.
- [16] D. Song, C. Lu, and J. Ma, "Gearbox fault diagnosis based on VMD-MSE and adaboost classifier," *Vibroengineering Procedia*, vol. 14, pp. 120–125, 2017.
- [17] C. W. A. W. Fanlei, "Gear fault diagnosis based on LCD and LME demodulation approach," *China Mechanical Engineering*, vol. 27, no. 24, p. 3332.
- [18] H. Yuan and C. Lu, "Rolling bearing fault diagnosis under fluctuant conditions based on compressed sensing," *Structural Control and Health Monitoring*, vol. 24, no. 5, Article ID e1918, 2017.
- [19] G. Cheng, X. Chen, H. Li, P. Li, and H. Liu, "Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition," *Measurement*, vol. 91, pp. 140–154, 2016.
- [20] F.-C. Zhou, G.-J. Tang, and Y.-L. He, "An Effective Gear Fault Diagnosis Method Based on Singular Value Decomposition and Frequency Slice Wavelet Transform," *International Journal of Rotating Machinery*, vol. 2016, Article ID 7458956, 8 pages, 2016.
- [21] A. Krishnakumari, A. Elayaperumal, M. Saravanan, and C. Arvindan, "Fault diagnostics of spur gear using decision tree and fuzzy classifier," *The International Journal of Advanced Manufacturing Technology*, vol. 89, no. 9-12, pp. 3487–3494, 2017.
- [22] Z. Xing, J. Qu, Y. Chai, Q. Tang, and Y. Zhou, "Gear fault diagnosis under variable conditions with intrinsic time-scale decomposition-singular value decomposition and support vector machine," *Journal of Mechanical Science and Technology*, vol. 31, no. 2, pp. 545–553, 2017.
- [23] P. V. Kane and A. B. Andhare, "Application of psychoacoustics for gear fault diagnosis using artificial neural network," *Journal of Low Frequency Noise, Vibration and Active Control*, vol. 35, no. 3, pp. 207–220, 2016.
- [24] H. Wu and C. Z. Song, "Engine gearbox fault diagnosis using modified Elman neural network and ACO algorithm," *Applied Mechanics and Materials*, vol. 190-191, pp. 982–986, 2012.
- [25] T. Waqar and M. Demetgul, "Thermal analysis MLP neural network based fault diagnosis on worm gears," *Measurement*, vol. 86, pp. 56–66, 2016.
- [26] X. You and W. Zhang, "Fault diagnosis of frequency converter in wind power system based on SOM neural network," in *Proceedings of the 2012 International Workshop on Information and Electronics Engineering, IWIEE 2012*, pp. 3132–3136, China, March 2012.
- [27] W. Zhao, D. Siegel, J. Lee, and L. Su, "An integrated framework of drivetrain degradation assessment and fault localization for offshore wind turbines," *International Journal of Prognostics and Health Management*, vol. 4, 2, pp. 462–472, 2013.
- [28] C. Lu, L. Tao, and H. Fan, "An intelligent approach to machine component health prognostics by utilizing only truncated histories," *Mechanical Systems and Signal Processing*, vol. 42, no. 1-2, pp. 300–313, 2014.
- [29] M. Moshinsky and C. Quesne, "Linear canonical transformations and their unitary representations," *Journal of Mathematical Physics*, vol. 12, pp. 1772–1780, 1971.
- [30] T. Alieva and M. J. Bastiaans, "Properties of the linear canonical integral transformation," *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 24, no. 11, pp. 3658–3665, 2007.
- [31] Y.-N. Zhang and B.-Z. Li, " ϕ -linear canonical analytic signals," *Signal Processing*, vol. 143, pp. 181–190, 2018.
- [32] S. Xu, Y. Chai, Y. Hu, C. Jiang, and Y. Li, "Reconstruction of digital spectrum from periodic nonuniformly sampled signals in offset linear canonical transform domain," *Optics Communications*, vol. 348, pp. 59–65, 2015.
- [33] R. Tao, B.-Z. Li, Y. Wang, and G. Aggrey, "On sampling of band-limited signals associated with the linear canonical transform," *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5454–5464, 2008.
- [34] S. Xu, Y. Chai, and Y. Hu, "Spectral analysis of sampled band-limited signals in the offset linear canonical transform domain," *Circuits, Systems and Signal Processing*, vol. 34, no. 12, pp. 3979–3997, 2015.
- [35] X.-N. Xu, B.-Z. Li, and X.-L. Ma, "Instantaneous frequency estimation based on the linear canonical transform," *Journal of The Franklin Institute*, vol. 349, no. 10, pp. 3185–3193, 2012.
- [36] Q. Feng and B.-Z. Li, "Convolution and correlation theorems for the two-dimensional linear canonical transform and its applications," *IET Signal Processing*, vol. 10, no. 2, pp. 125–132, 2016.
- [37] Y. Guo and B.-Z. Li, "Blind image watermarking method based on linear canonical wavelet transform and QR decomposition," *IET Image Processing*, vol. 10, no. 10, pp. 773–786, 2016.
- [38] E. J. Candès, P. R. Charlton, and H. Helgason, "Detecting highly oscillatory signals by chirplet path pursuit," *Applied and Computational Harmonic Analysis*, vol. 24, no. 1, pp. 14–40, 2008.
- [39] J. Luo, S. Zhang, M. Zhong, and Z. Lin, "Order spectrum analysis for bearing fault detection via joint application of synchrosqueezing transform and multiscale chirplet path pursuit," *Shock and Vibration*, vol. 2016, Article ID 2976389, 11 pages, 2016.

Research Article

Fault Diagnosis for Hydraulic Servo System Using Compressed Random Subspace Based ReliefF

Yu Ding ^{1,2}, Fei Wang ^{1,2}, Zhen-ya Wang ³, and Wen-jin Zhang ^{1,2}

¹*School of Reliability and Systems Engineering, Beihang University, Beijing, China*

²*Science & Technology on Reliability and Environmental Engineering Laboratory, Beijing, China*

³*Research and Development Center, China Academy of Launch Vehicle Technology, Beijing, China*

Correspondence should be addressed to Wen-jin Zhang; buaazwjok@yeah.net

Received 28 October 2017; Revised 5 January 2018; Accepted 14 January 2018; Published 18 February 2018

Academic Editor: Gangbing Song

Copyright © 2018 Yu Ding et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Playing an important role in electromechanical systems, hydraulic servo system is crucial to mechanical systems like engineering machinery, metallurgical machinery, ships, and other equipment. Fault diagnosis based on monitoring and sensory signals plays an important role in avoiding catastrophic accidents and enormous economic losses. This study presents a fault diagnosis scheme for hydraulic servo system using compressed random subspace based ReliefF (CRSR) method. From the point of view of feature selection, the scheme utilizes CRSR method to determine the most stable feature combination that contains the most adequate information simultaneously. Based on the feature selection structure of ReliefF, CRSR employs feature integration rules in the compressed domain. Meanwhile, CRSR substitutes information entropy and fuzzy membership for traditional distance measurement index. The proposed CRSR method is able to enhance the robustness of the feature information against interference while selecting the feature combination with balanced information expressing ability. To demonstrate the effectiveness of the proposed CRSR method, a hydraulic servo system joint simulation model is constructed by HyPneu and Simulink, and three fault modes are injected to generate the validation data.

1. Introduction

Hydraulic servo system plays a crucial role in electromechanical systems, like engineering machinery, metallurgical machinery, ships, and other equipment. Failures of hydraulic servo system caused by severe and complex conditions may lead to catastrophic accidents and enormous economic losses. Fault diagnosis based on monitoring and sensory signals is able to classify the current state of complex systems, which plays a key role in performance evaluation [1]. Feature set extracted from signals is an important index to reflect the fault mechanism and performance evolution laws. The quality of feature set plays a key role in improving the generalization ability of fault identification [2]. The common feature extraction methods are time-frequency index extraction, wavelet analysis, Hilbert transform, Duffing oscillator, and so on. Despite their respective applicable conditions and limitations, those methods are able to mine the health characteristics of the system from multiaspect [3, 4]. Same as

machine learning, features extracted from images, speeches, and other signals often have certain correlations and hidden mutual influences. Information expressed by a single feature is usually inadequate, which can be greatly improved when the single feature is aggregated with others [5]. Similarly, due to the nonlinearity, instability, and nonconformity of complex electromechanical systems, the expression of the information on individual feature is often one-sided. Thus, a new challenge is how to utilize those features more effectively and efficiently, in other words, how to obtain the feature set that expresses the information sufficiently by eliminating the redundant and negatively correlated features [6–9].

To tackle the challenge mentioned above, on the premise of existing feature extraction techniques, feature processing techniques including feature selection and dimension reduction have gradually become an important research focus. Both feature selection and dimension reduction can reduce the scale of feature set by obtaining a set of principal variables. Such techniques often use a variety of feature

extraction methods to integrate the features into a comprehensive representation of the signal [10]. For the purpose of enhancing the expressing ability of core information on multiclass feature sets, spatial transformation or importance measurement methods are used [11]. Such methods are able to reduce redundancy existing in features and improve learning efficiency while retaining the performance advantages. The data transformation may be linear, such as principal component analysis (PCA). But many nonlinear dimension reduction techniques also exist. The common feature processing techniques include linear dimensionality reduction methods (FDA, LPP, etc.), kernel function based dimension reduction methods (KPCA, KFDA, etc.), manifold based dimension reduction methods (Isomap, LLE, MDS, etc.), filter method based feature selection methods (Relief, Focus, information gain, CBFS, etc.), wrapper method based feature selection methods (genetic algorithm, distribution estimation, differential algorithm, etc.), and embedded method based feature selection methods (SVN-RFE, RF, etc.) [12–14].

For the nonlinear signals of complex electromechanical systems, although the dimension reduction methods can reduce the scale of input features for fault diagnosis, they change the basic attributes of the feature set. Such situation makes it difficult to give a clear understanding of the obtained feature subset. Meanwhile, existing feature selection methods sort the importance degree of the features according to the independent feature evaluation result. They ignore the interaction among features, which would lead to information loss in processing the data of electromechanical systems [15]. Aiming at the shortcomings of feature selection and dimension reduction techniques, such as low expandability, unclear evaluation indexes, and strong tendency, this study proposes compressed random space based ReliefF (CRSR) method. Based on ReliefF method, CRSR introduces ensemble strategy on feature level based on compressed random space. Furthermore, CRSR optimizes the objective function using information entropy and fuzzy theory. The main contributions are as follows:

- (i) This study analyzes the feasibility of the ReliefF based feature selection architecture for the fault diagnosis of complex electromechanical system. Meanwhile, the basic mechanism of measuring the contribution of the features based on ReliefF is also demonstrated.
- (ii) By converting the assessment process of ReliefF, which takes the entire feature set as object, into the construction and ensemble process of feature subspace, CRSR can improve the global optimization ability of ReliefF.
- (iii) Considering ReliefF based feature selection as a problem of maximizing the distance, CRSR replaces the traditional spatial distance with fuzzy membership degree, which is able to obtain a robust and steady objective function.

This paper is structured as follows: In Section 2, ReliefF method based feature selection structure is introduced. Then feature integration method based on compressed random subspace is described. Objective function optimization

method based on information entropy and fuzzy theory is also introduced. Section 3 presents the overall diagnosis procedure for hydraulic servo systems, the details of construction and fault injection of the hydraulic servo system, and analysis and comparison on feature selection results using the proposed CRSR method which are discussed numerically. Section 4 concludes the paper.

2. Related Theories

2.1. ReliefF Method Based Feature Selection Structure. ReliefF is the extension of Relief method by estimating probabilities more reliably, which is able to handle incomplete and multiclass data sets while the complexity remains the same [16–19]. By calculating the distances between the sample distributions, ReliefF can obtain the correlation weight coefficients of the features which is similar to Relief.

For a specific feature from the feature set, if its difference in same class is much smaller than that in different class, it is considered that this feature contributes to class discrimination [20]. Given a sample set $S = \{s_1, \dots, s_i, \dots, s_n\}$ ($i \in [1, n]$) with n instances, each sample, $s_i = \{s_{i1}, \dots, s_{it}, \dots, s_{im}\}$ ($t \in [1, m]$), has m -dimensional features. Meanwhile, the samples in S only belong to two classes which are tagged as $c_i \in (C_1, C_2)$. The difference between each two samples (s_i and s_j) on feature t is defined as

$$\text{diff}(t, s_i, s_j) = \begin{cases} 1, & s_{it} \neq s_{jt} \\ 0, & s_{it} = s_{jt} \end{cases} \quad (1)$$

where the attribute of feature t is discrete value. If the attribute of feature t is continuous value,

$$\text{diff}(t, s_i, s_j) = \left| \frac{s_{it} - s_{jt}}{\max_t - \min_t} \right|. \quad (2)$$

Features extracted from condition monitoring data of electromechanical system mostly are continuous data. Meanwhile \max_t and \min_t represent the maximum value and minimum value of the entire sample, respectively. The closest same-class instance of sample s_i is called “near-hit (NH),” and the closest different-class instance of sample s_i is called “near-miss (NM).” Meanwhile, the weight of feature t is denoted as ω_t , and ω_t is updated by

$$\omega_t = \omega_{t\text{-old}} - \text{diff}(t, s_i, \text{NH}) + \text{diff}(t, s_i, \text{NM}), \quad (3)$$

where the initial value of ω_t is 0, and $\omega_{t\text{-old}}$ is the last value of ω_t .

For reducing the randomness in feature evaluation, the whole process needs to repeat r times to obtain the average value being the final weight. Although Relief method is very efficient in estimating the quality of attributes, it cannot deal with incomplete data and is limited to two-class problem. Thus ReliefF method is utilized in the paper to deal with the multiclass classification and regression problems for continuous data.

For a multiclass classification problem, assuming that the samples in S belong to multiple classes and the tags for S are

$c_i \in (C_1, \dots, C_l)$, ReliefF updates ω_t on sample s_i by taking k near hits (NHS) and k near misses (NMS) into consideration,

$$\omega_t = \omega_{t\text{-old}} - \frac{\sum_{j=1}^k \text{diff}(t, s_i, \text{NHS}_j)}{(r * k)} + \frac{\sum_{c \neq \text{class}(s_i)} [(p(c_i) / (1 - p(\text{class}(s_i)))) \sum_{j=1}^k \text{diff}(t, s_i, \text{NMS}_j)]}{(r * k)}, \quad (4)$$

where $p(c_i)/(1 - p(\text{class}(s_i)))$ represents the ratio of the entire samples in class c_i to all the heterogeneous samples in s_i . Furthermore, ReliefF method equalizes the differentiation of NHs and calculates the average differences between s_i and other classes on feature t to evaluate the classification ability of the samples nearby.

2.2. Feature Integration Method Based on Compressed Random Subspace. The purpose of CRSR method is finding the balance between the differences and the correlations of features. Specifically, in the premise of fully mining the correlations of features using ReliefF method, CRSR method is applied to make each feature subset keep a certain degree of difference. Based on random subspace and feature sorting strategy, the feature sets with higher contribution can be obtained in various feature combinations [21, 22]. RS_1 and RS_2 are denoted as two random subspaces, so the difference, denoted as $D_RS(RS_1, RS_2)$, can be calculated as

$$D_RS(RS_1, RS_2) = |RS_1 \cup RS_2| - |RS_1 \cap RS_2|, \quad (5)$$

where symbol $|\cdot|$ denotes the dimension of random subspaces.

The right side of (5) obtains the noncoincident features of RS_1 and RS_2 . $D_RS(RS_1, RS_2)$ is plus one when there is an unrepeatd feature f_D . The average difference between two random subspaces on all the features is defined as

$$E(D_RS) = \sum_{q=1}^Q [P(f_D \in RS_1) P(f \notin RS_2) + P(f_D \notin RS_1) P(f \in RS_2)], \quad (6)$$

where Q is the dimension after feature ordering compression. It can be simplified as $Q = m - m_1$. Concretely, according to sorting strategy, RS_1 , the probability of the first m_0 strongly related features being selected is $2m_0/Q$, and the relatively poorly related m_0 features being selected are $m_1/(Q - m_0)$. The average difference of feature evaluation based on compressed random subspace method, which is $E(D_RS)$, can be denoted as follows:

$$\begin{aligned} E(D_RS) &= 2 \left(\left(\frac{1}{2} Q \right) \frac{m_0}{(1/2)Q} \left(1 - \frac{m_0}{(1/2)Q} \right) \right. \\ &\quad \left. + (Q - m_0) \frac{m_1}{Q - m_0} \left(1 - \frac{m_1}{Q - m_0} \right) \right) \\ &= 2 \left(\frac{m_0(Q - 2m_0)}{Q} + \frac{m_1(Q - m_1 - m_0)}{Q - m_0} \right). \end{aligned} \quad (7)$$

Equation (7) shows that the ranking result can balance the difference of ReliefF to determine the dominant features that

which is different from Relief. Similarly, the weight of feature t , which is ω_t , can be updated through

are crucial for classification, which would improve the feature selection efficiency.

Based on compressed random subspace method, this study proposes redundancy analysis method from statistics to reduce the redundancy of feature. The features are checked in pairs using redundancy analysis method [23, 24]. Firstly, two sets of feature vectors, F_1 and F_2 , are selected randomly from the feature set obtained by ReliefF. Then, the selected feature vectors are regarded as independent variable and dependent variable separately. The covariance matrixes, denoted as U and V , can be calculated, respectively, as

$$\begin{aligned} U &= E \{ (F_1 - E[F_1]) (F_1 - E[F_1])^T \}, \\ V &= E \{ (F_2 - E[F_2]) (F_2 - E[F_2])^T \}, \end{aligned} \quad (8)$$

where $E[X]$ represents mathematical expectation of vector X . Then, the correlation coefficient of F_1 and F_2 , denoted as $\rho(U, V)$, can be formulated as

$$\rho(U, V) = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U)} \sqrt{\text{Var}(V)}}, \quad (9)$$

where $\text{Cov}(U, V)$ denotes the covariance of U and V . If the correlation coefficient is greater than a presetting threshold, only the one with larger weight from F_1 and F_2 will be added to the final selected feature set. It is noticed that redundancy analysis based on matrix transformation focuses on the correlation between features instead of the similarity of data. Thus, CRSR can reduce the influence of numerical confusion existing in feature selection. Furthermore, compared with traditional methods based on data similarity, CRSR is able to obtain higher confidence level.

2.3. Objective Function Optimization Method Based on Information Entropy and Fuzzy Theory. From the aspect of maximizing the distance, Relief method can be seen as a distance optimization algorithm using feature weighting method [25]. Under this condition, the optimization objective function, denoted as $\max_{\omega} J$, can be described as [26]

$$\max_{\omega} J = \sum_{i=1}^n \rho_i(\omega), \quad (10)$$

where $\rho_i(\omega)$ denotes the distance of the i th sample. Based on (4), $\rho_i(\omega)$ can be converted as

$$\rho_i(\omega) = \sum_{t=1}^m \omega_t \cdot \left(\sum_{c \notin \text{class}(s_i)} \frac{p(c_i)}{1 - p(\text{class}(s_i))} \sum_{j=1}^k \text{diff}(t, s_i, \text{NMS}_j) - \sum_{j=1}^k \text{diff}(t, s_i, \text{NHS}_j) \right). \quad (11)$$

For complex electromechanical system signals, two problems occur when using (10) as the optimization objective function of CRSR. One problem is that the objective function concentrates the weight on one or some of the features, which leads to the result that assessment value of the remaining features tends to 0. Meanwhile, (10) regards the samples with stochastic volatility and noise similar to the normal sample. Another problem is the lack of consideration on the influence from the quality of samples on the feature selection process.

Aiming at the first problem, the information entropy theory is proposed based on compressed random subspace method, which combines the maximization of entropy together with maximization of distance to reduce the over tendency problem of the existing ReliefF method. After adding a sample estimation factor κ_i , the optimization objective function is denoted as

$$\max_{\omega, \kappa} J = \sum_{i=1}^n \kappa_i \rho_i(\omega). \quad (12)$$

Supposing that κ_i and ω_i follow probability distribution, Shannon entropy is used to adjust the sample distribution, as shown below:

$$J_e(\kappa_i) = - \sum_{i=1}^n \kappa_i \ln \kappa_i, \quad (13)$$

$$J_e(\omega_i) = - \sum_{i=1}^m \omega_i \ln \omega_i.$$

Aiming at the second problem, the fuzzy membership degree is chosen to replace the traditional nearest neighbor distance. The fuzzy membership degree has the advantage of being insensitive to sample fluctuations and noise and the ability of updating adaptively while changing the feature weights [27]. In a sample space of same class, the fuzzy membership degree of the feature s_{it} , denoted as ν_{ij} , can be calculated as [28, 29]

$$\nu_{ij} = \frac{(Q_{\nu,ij})^{1/(1-m)}}{\sum_{i=1}^{|\text{NHS}|} \left((Q_{\nu,ij})^{1/(1-m)} \right)}, \quad (14)$$

where $|\text{NHS}|$ is the sample set of same class, and $Q_{\nu,ij}$ is the fuzzy difference between feature s_{it} and feature s_{jt} , as shown below:

$$Q_{\nu i} = \sum_{j=1}^M \omega_j \text{diff}(t, s_i, s_j), \quad s_j \in \text{NHS}. \quad (15)$$

Then the same fuzzy distance, denoted as $m_diff(t, s_i, \text{NHS})$, is calculated as

$$m_diff(t, s_i, \text{NHS}) = \sum_{s_j \in \text{NHS}} \nu_{ij} \text{diff}(t, s_i, s_j). \quad (16)$$

Similarly, the fuzzy difference and the fuzzy membership degree of the heterogeneous sample sets of s_j , which are denoted as $Q_{\mu i}$ and μ_{ij} separately, can be calculated as follows:

$$Q_{\mu i} = \sum_{j=1}^M \omega_j \text{diff}(t, s_i, s_j), \quad s_j \in \text{NMS}, \quad (17)$$

$$\mu_{ij} = \frac{(Q_{\mu,ij})^{1/(1-m)}}{\sum_{i=1}^{|\text{NMS}|} \left((Q_{\mu,ij})^{1/(1-m)} \right)}.$$

Therefore, the heterogeneous fuzzy distance, denoted as $m_diff(t, s_i, \text{NMS})$, is calculated as

$$m_diff(t, s_i, \text{NMS}) = \sum_{s_j \in \text{NMS}} \mu_{ij} \text{diff}(t, s_i, s_j). \quad (18)$$

Based on the fuzzy distances obtained by (16) and (18), updated (11) can be formulated as follows:

$$\rho_{f,i}(\omega) = \sum_{t=1}^m \omega_t \cdot \left(\sum_{c \notin \text{class}(s_i)} \frac{p(c_i)}{1 - p(\text{class}(s_i))} \sum_{j=1}^k m_diff(t, s_i, \text{NMS}_j) - \sum_{j=1}^k m_diff(t, s_i, \text{NHS}_j) \right). \quad (19)$$

Based on the formulas above, the objective function, which is $\max_{\omega, \kappa} J$, can be denoted as

$$\max_{\omega, \kappa} J = \sum_{i=1}^n \kappa_i \rho_{f,i}(\omega) + \lambda_{\kappa} J_e(\kappa_i) + \lambda_{\omega} J_e(\omega_i), \quad (20)$$

where the first item on the right side is the maximum distance of ReliefF, which is meant to determine the feature set that contributes most to classification. Meanwhile, from the aspect of entropy maximization, the second and the third items denote the sample evaluation operator and Shannon entropy of feature weight, respectively, which are used to avoid the over tendency problem of objective function. λ_{κ} and λ_{ω} are the balance coefficients for adjusting the differences between features. When the maximum point of objective function is

achieved, the constraint condition of the sample evaluation operator is defined below [30]:

$$\kappa_i = \frac{\exp\left(\sum_{j=1}^m \omega_j d_{i,j}/\lambda_\kappa\right)}{\sum_{i=1}^n \exp\left(\sum_{j=1}^m \omega_j d_{i,j}/\lambda_\kappa\right)}, \quad (21)$$

where $d_{i,j}$ denotes the partial derivative of $\rho_{f,i}(\omega)$.

In the process of feature selection for mechanical system, the information entropy and fuzzy theory based optimization objective function ensures that the evaluation process of each subspace using ReliefF is adaptive and robust. Such advantage provides a new thought for feature processing of complex monitoring signal, in other words, under the premise of maintaining the high calculating efficiency of ReliefF method, reducing the bias and redundancy caused by methodological defects and external disturbances [31].

3. Method for Hydraulic Servo System Fault Diagnosis Based on CRSR

The CRSR based fault diagnosis method for hydraulic servo system consists of following successive steps: first, the average, standard deviation, skewness, and wavelet singular entropy features are extracted to form a feature matrix as the input of ReliefF model. Second, the initial contribution of the features sampled randomly is measured by calculating the inner-class distance and between-class distance, and the features of high contribution and features of low contribution are determined for fault classification. The i th iteration operation is based on the result of the previous iteration, and the iteration stops as long as i reaches the preset threshold. Third, based on the sorting result of the second step, the compressed evaluation of the features is realized using supervised sampling method in current iteration. Finally, keeping the iteration running until certain terms is satisfied to acquire the difference value as the criterion of feature selection. The detailed process of feature selection method using CRSR for hydraulic servo pump is shown in Figure 1.

Compared with the traditional feature processing methods [32–35], the CRSR method is designed to meet the robustness and accuracy requirements of fault diagnosis with smaller feature set and lower resource consumption. This study demonstrates the advantages of the proposed method by extracting a variety of categories of features from the simulation data of hydraulic servo system and optimizes the feature set extracted by CRSR method to verify the feasibility in feature selection technique.

3.1. Description of the Simulation Environment of Hydraulic Servo System. The data used in this case is generated in HyPneu and Simulink joint simulation environment. The joint simulation process of hydraulic servo system can be divided into mechanical part and control part considering the characteristics of the hydraulic components [36–38]. The mechanical hydraulic physical part, including hydraulic pumps, servo valves, and actuators, is modeled using component library in HyPneu. The failure is realized by adding the modules of fault injection for the relevant components. The dynamic control

part, including feedback sensors and electronic amplifiers, is established using the relevant model in Simulink. Meanwhile, the control of the hydraulic servo system and the simulation of fault injection can be realized by transmitting the signal data through relevant interface files. Thus, the model architectures of hydraulic servo system constructed using HyPneu and Simulink are shown in Figures 2 and 3, respectively.

As shown in Figure 3, the Simulink model includes input signals, comparison elements, control elements, amplifying elements, and HyPneu module. The gain parameter of the electronic amplifying element is 80, and the control parameter of PID is set as proportional parameter of 1500, integral parameter of 0, and differential parameter of 5. To simulate the actual operating environment of hydraulic system, random noises in the range of -0.01 to 0.01 are added on the output of the HyPneu module.

The process of joint simulation based on HyPneu and Simulink is as follows. First, the actuator receives the feedback signal converted by servo valve and right after that drives the load to do reciprocated motion. Then, the real-time displacement information of the load is collected by the sensor and transferred to the control circuit. Finally, the input signal and the amplified displacement signal are compared and inputted into the servo valve to build the negative feedback control logic of closed loop. The joint simulation model has a good ability to match the hydraulic system and to realize the fault injection of the hydraulic system effectively.

3.2. Fault Injection for Hydraulic Servo System and Multi-dimensional Feature Extraction. This study simulates four kinds of state of the hydraulic system including normal state, electronic amplifier fault, sensor constant deviation fault, and hydraulic pump wear fault. The detailed fault injection scheme is shown in Table 1.

The input of all fault modes is sinusoidal signal with amplitude being 1 and frequency being 4π rad/s. As the expression of the input signal is $y = \sin(4\pi t)$, the sampling frequency is 100 Hz and the sampling time is 70 s. The input-output relationships of the hydraulic servo system under circumstances of normal state and three fault states are collected, and every signal of the corresponding state contains 7000 points. The original output signals are shown in Figure 4.

As mentioned above, the input of the proposed CRSR method is a feature matrix, and each column of the matrix represents a feature vector extracted from the original signal. As the prerequisite of feature selection, multidimensional features should be extracted for all of the signals. It should be noted that the cycle of hydraulic servo system signal collected for the feature selection should be an integer multiple of the input waveform cycle. The cycle of the input signal is 0.5 seconds and one input-output cycle contains 50 points, which means the object points of the feature extraction should be a geometric multiple of 50. In addition, in consideration of the fact that pump wear fault is a gradual degradation process and the degree of fault in the early phase may be weak, this study takes the last 4000 points of the original signal to extract the feature matrix. Moreover, a sampling window of 500-point length is used (with 5-point interval) to obtain 700 feature vectors.

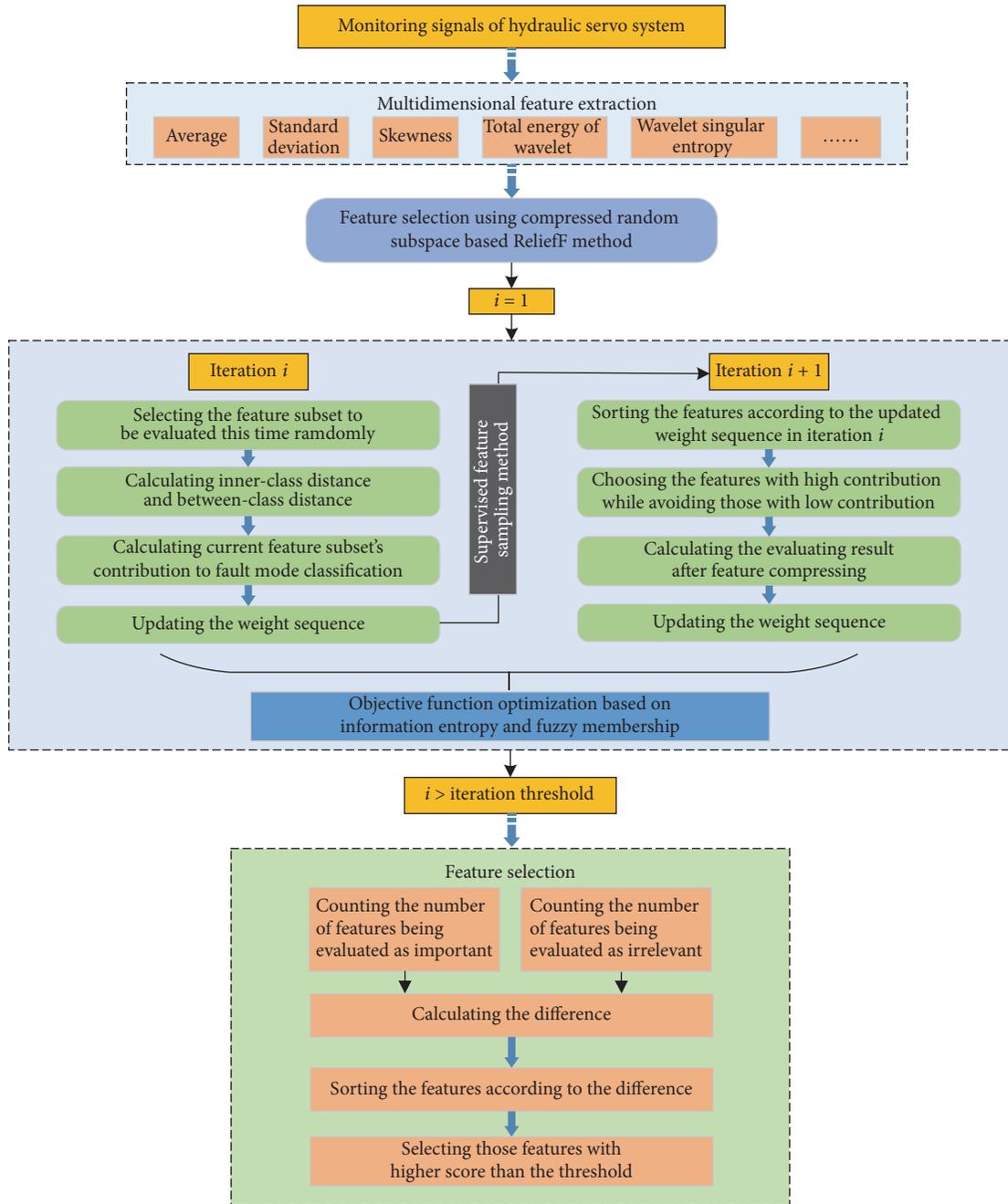


FIGURE 1: Process of CRSR based feature selection method for hydraulic servo pump.

TABLE 1: Fault injection scheme.

Fault mode	Fault phenomenon	Fault injection mode
Normal state	-	-
Electronic amplifier fault	Magnification notably increases	Multiplying the gain by 0.25
Sensor fault	Constant gain	Setting the feedback of the sensor to 0.5
Hydraulic wear fault	Output flow decreases	Setting flow-pressure coefficient of the variable orifice valve to 0.03

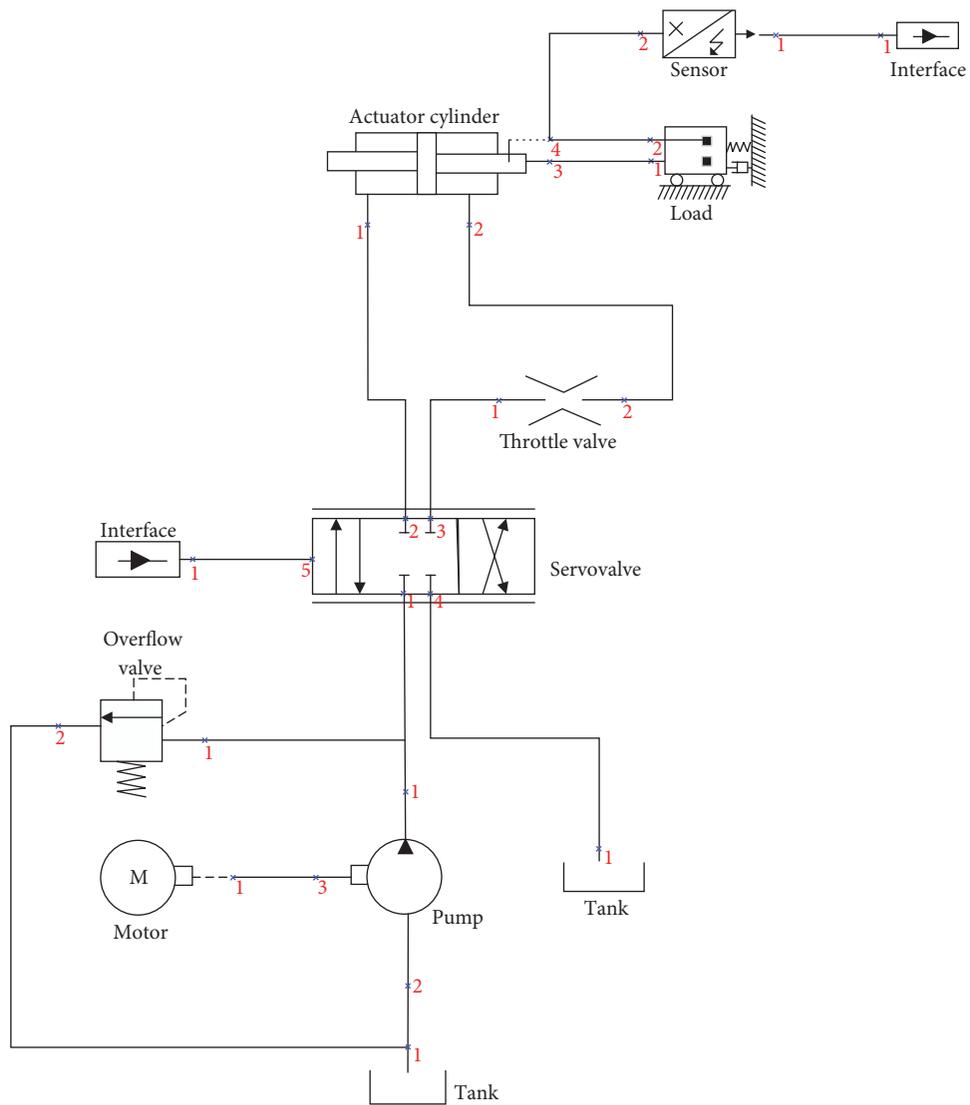


FIGURE 2: The failure injection model of mechanical hydraulic part using HyPneu.

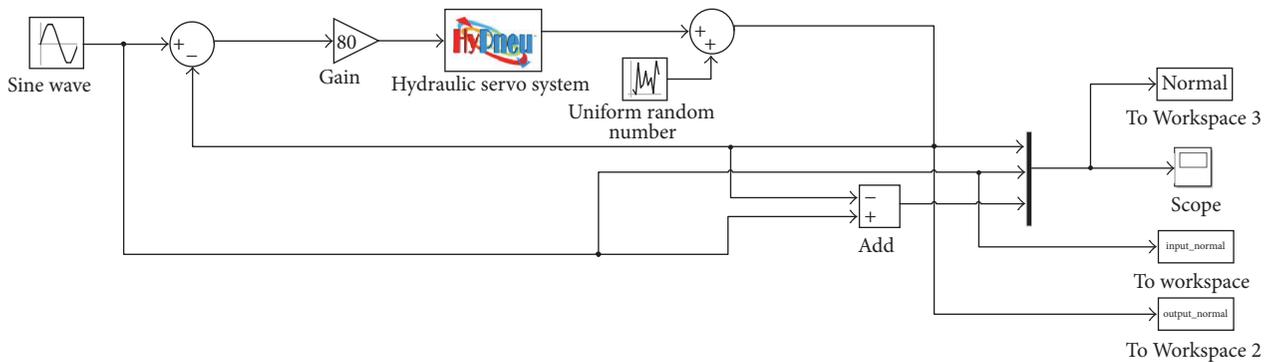


FIGURE 3: The failure injection model of dynamic control part using Simulink.

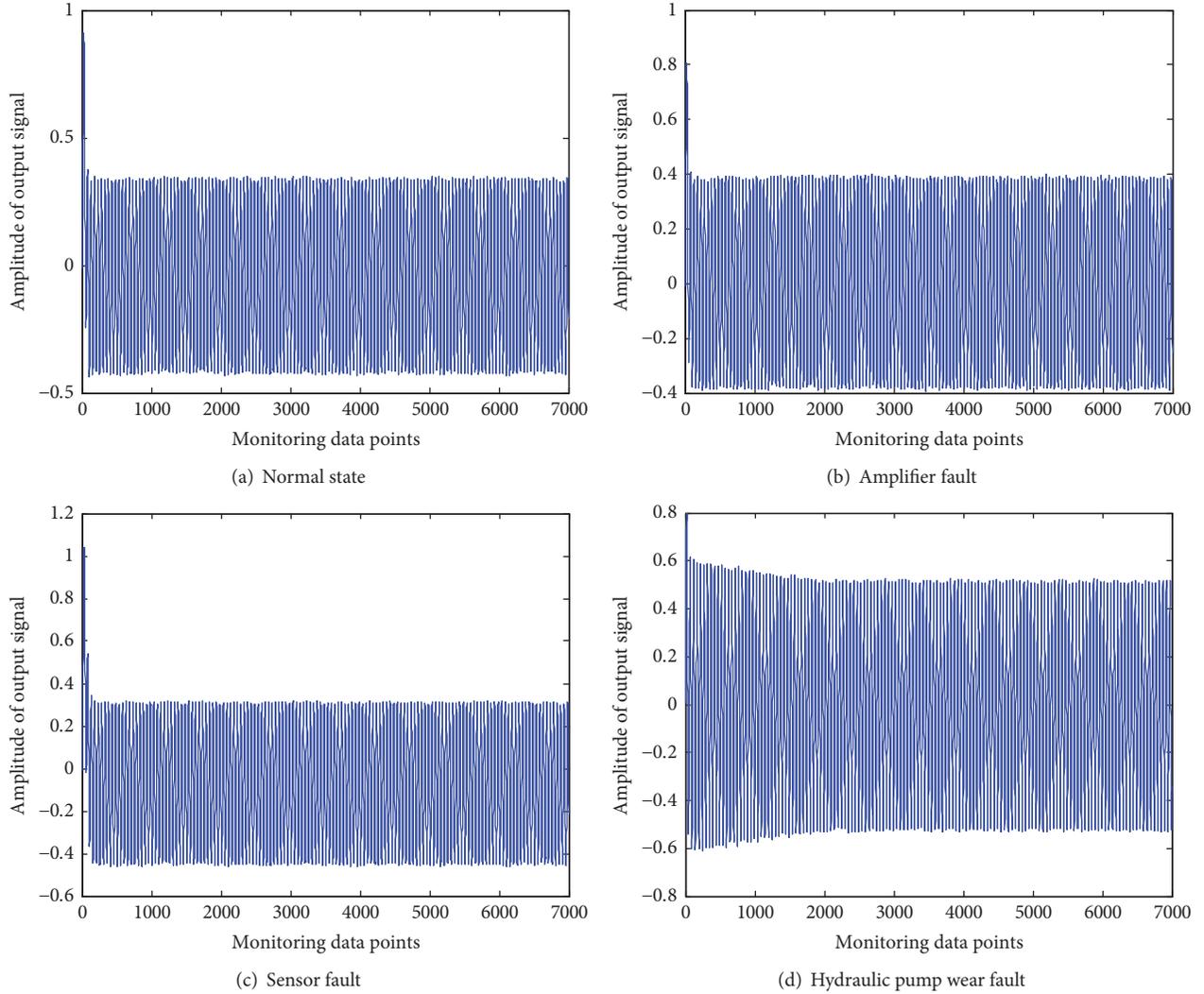


FIGURE 4: Original output signals of the hydraulic servo system.

Furthermore, this study adds 15 dB Gaussian white noise to all the original signals to validate the performance of CRSR method. The details of the extracted features are shown in Table 2.

As can be seen from Table 2, the quantitative range of different fault features fluctuates much, which could lead to the bias problem in the feature selection process of ReliefF. Therefore, this study normalizes the features to the range of zero to one. The visualization of the feature vectors of different states is shown in Figure 5.

As is shown in Figure 5, data confusion exists in the features of different fault modes, which makes it difficult to select satisfied feature subsets through artificial observation. Thus, CRSR method is designed to select the feature combination with higher contribution to the fault diagnosis adaptively.

3.3. Analysis and Comparison on Feature Selection Result. The input of CRSR feature selection model is a 9×700 matrix acquired by regularizing the 9 sets of the obtained feature vectors. One of the advantages of CRSR is the ability

to calculate the distances of either same class or different classes adaptively and iteratively. Based on the ReliefF feature selection method, the corresponding distribution vectors of feature weight can be obtained.

During the iteration process of ReliefF, feature selection constraint is established using compressed stochastic subspace method, which optimizes the performance and efficiency of CRSR. The parameters are as follows: number of iterations: 30; feature base of the subspace: 6. And the threshold condition of feature selection is that the difference between a feature being selected as “one with high contribution” and “one with low contribution” is greater than 10. With F_1-F_9 representing the feature parameters in Figure 5, the statistical result of the features being selected as “one with high contribution” and “one with low contribution” is shown in Table 3.

It can be seen from Table 3 that the feature combination selected by CRSR method contains average (F_1), standard deviation (F_2), mean square root (F_3), crest factor (F_5), and the maximum amplitude of FFT (F_6). Although wavelet

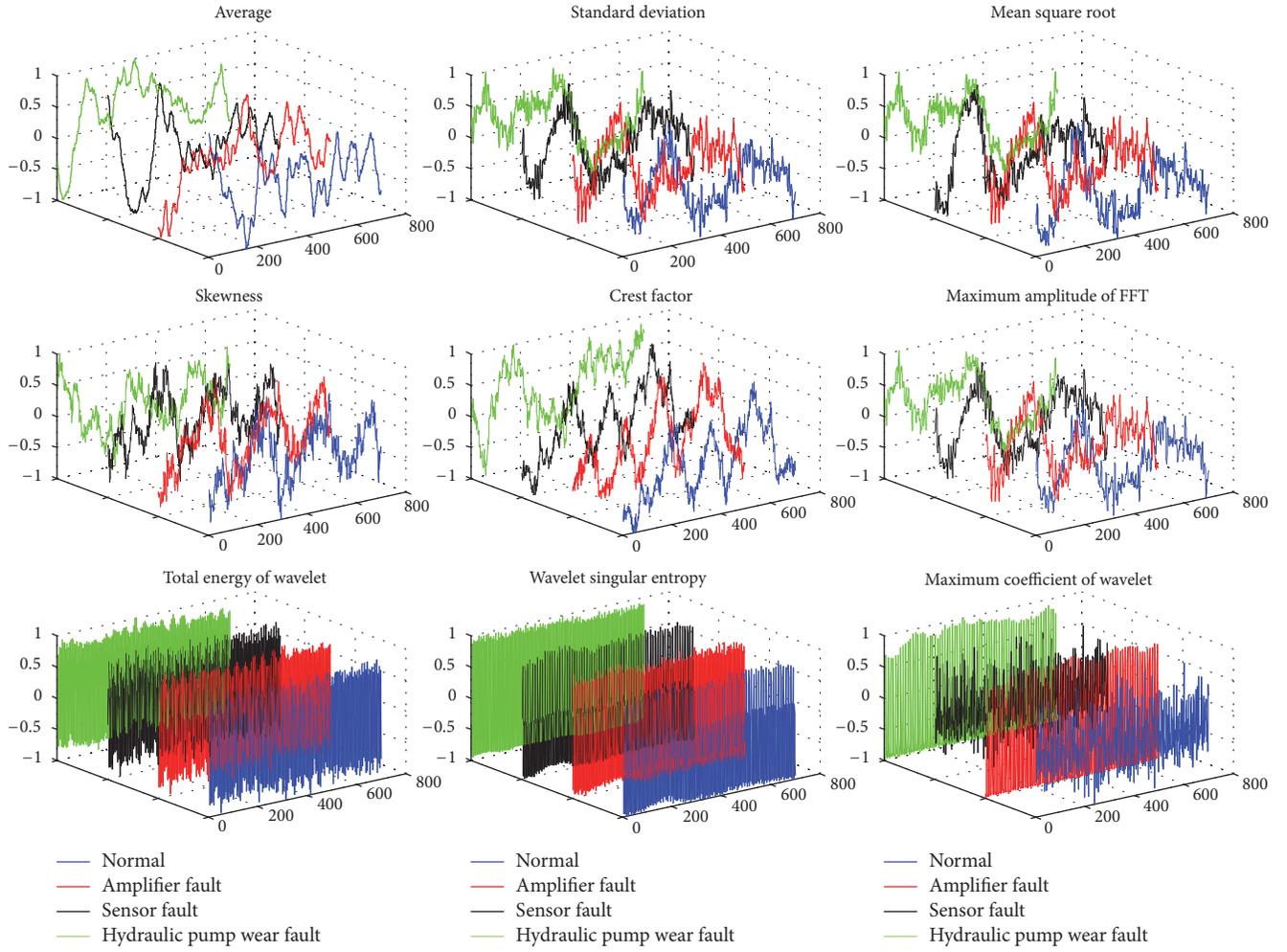


FIGURE 5: Feature vectors of the hydraulic servo system in different states.

TABLE 2: Extracted features of the output signals of hydraulic servo system.

Feature	Feature magnitude			
	Normal	Amplifier fault	Sensor fault	Hydraulic pump wear fault
Average	-0.0446~-0.0411	0.0004~0.0036	-0.0734~-0.0698	-0.0100~-0.0063
Standard deviation	0.2715~0.2728	0.2745~0.2756	0.2715~0.2727	0.3068~0.3080
Mean square root	0.2746~0.2760	0.2742~0.2754	0.2806~0.2819	0.3067~0.3079
Skewness	0.0080~0.0148	0.0028~0.0099	0.0082~0.0149	0.0085~0.0174
Crest factor	1.4603~1.4608	1.4620~1.4685	1.4599~1.4675	1.8089~1.8189
Maximum amplitude of FFT	95.8092~96.288	96.8611~97.2778	95.8319~96.2598	107.5320~107.94
Total energy of wavelet	99.4925~99.746	99.4695~99.7350	99.5243~99.7664	98.5909~99.0743
Wavelet singular entropy	0.7682~0.8193	0.7701~0.8113	0.7857~0.8401	0.8579~0.9178
Maximum coefficient of wavelet	0.0012~0.0035	0.0005~0.0669	0.0004~0.0034	0.0005~0.0543

TABLE 3: Statistical results based on compressed stochastic subspace.

Feature	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
High contribution	28	26	22	0	12	26	3	14	1
Low contribution	0	0	0	17	1	0	2	6	21
Difference	28	26	22	-17	11	26	1	8	-20

TABLE 4: Times of feature selections based on compressed random subspace.

Dependence analysis	F_1	F_2	F_3	F_5	F_6
F_1	1	0.42	0.69	0.17	0.16
F_2	-	1	0.56	0.09	0.44
F_3	-	-	1	0.31	0.61
F_5	-	-	-	1	0.82
F_6	-	-	-	-	1

TABLE 5: Fault diagnosis results with different combinations of features.

Feature combination	Fault diagnosis accuracy of the 10-fold cross validation method (%)									
$F_1 + F_2 + F_3 + F_6$	94.62	95.17	94.29	94.40	93.81	94.11	93.54	95.12	95.01	93.98
$F_1 + F_2 + F_3 + F_5 + F_6$	94.48	93.61	93.01	93.77	94.15	95.25	93.99	94.58	94.62	93.71
$F_1 + F_2 + F_3 + F_5 + F_6 + F_7 + F_8$	92.04	93.62	93.11	93.25	91.37	94.54	93.27	92.32	93.29	93.41
$F_1 + F_3 + F_4 + F_5$	93.02	93.57	92.15	92.89	92.46	94.19	93.32	93.55	92.53	92.77
$KPCA_1 \sim KPCA_4$	93.08	93.78	93.07	93.57	92.17	93.04	92.80	92.42	93.48	93.36
$LLE_1 \sim LLE_4$	94.10	93.59	94.51	94.22	93.98	94.10	93.99	93.44	93.98	94.78
$F_1 \sim F_9$	92.67	91.80	93.54	92.47	92.11	93.64	92.97	93.62	94.01	93.00

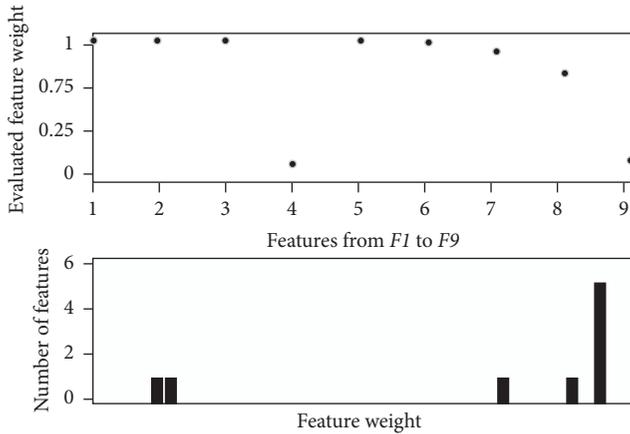


FIGURE 6: Statistical results based on compressed stochastic subspace.

singular entropy (F_8) was selected as “one with high contribution” for 14 times during the iteration, it was selected as “one with low contribution” for 6 times. Such result reveals that wavelet singular entropy is able to contribute to classification sometimes, but the ensemble performance is not stable. Even in some feature subspace matrixes, it exerts a negative impact on the classification. The final statistical result of feature weights assessed by CRSR is shown in Figure 6.

As is shown in Figure 6, if the weight basis for ReliefF is set to 0.75, the feature combination selected by CRSR method will append features F_7 (total energy of wavelet) and F_8 (wavelet singular entropy). However, according to the statistical results from the CRSR model, the contribution of these two features in the integration process is unstable. In particular, as for the total energy of wavelet, the reason why its average weight coefficient is high is that the weight coefficient reaches 0.92 and 0.88 in the 8th and 22nd iterations, respectively, while in the remaining subspaces

its contributions are lower than the average slightly. Such situation reveals that the total energy of wavelet could not meet the requirement of stability in feature selection. The result shows that the CRSR method is more reasonable than the traditional ReliefF, which reflects the advantages of ensemble learning in generalization.

Based on the feature selection result using CRSR, redundancy analysis for the feature sequence is carried out in this study. The threshold of correlation coefficient is set to 0.8, and the correlation matrix of feature is shown in Table 4.

It can be seen from Table 4 that the correlation coefficient of peak factor (F_5) and the maximum amplitude of FFT transform (F_6) are greater than the preset threshold, which indicates that redundancy exists in the fault diagnosis information provided by them. Therefore, only the maximum amplitude of FFT transform (F_6) is retained. In summary, the features selected by CRSR are average (F_1), standard deviation (F_2), mean square root (F_3), and the maximum amplitude of FFT (F_6).

The purpose of introducing CRSR method is improving the performance of fault diagnosis. In other words, for a classifier, the diagnostic performance using selected feature set should not be less than that using the original feature set. The comparative models used in this study contain the classical ReliefF algorithm, the Mean Impact Value (MIV) algorithm, the Locally Linear Embedding (LLE) algorithm, and Kernel Principal Component Analysis (KPCA). In this study, the feature sets selected by different feature selection methods are used as inputs for the classifier based on Radial Basis Function (RBF) neural network. The ratio of training samples to test samples is 50%. The fault diagnosis accuracy of the hydraulic servo system is obtained by using the 10-fold cross validation method as shown in Table 5.

In Table 5, the first row (A_1) represents the final feature set selected by CRSR; the second row (A_2) represents the feature set selected by CRSR before the redundancy analysis; the third row (A_3) represents the feature set selected by

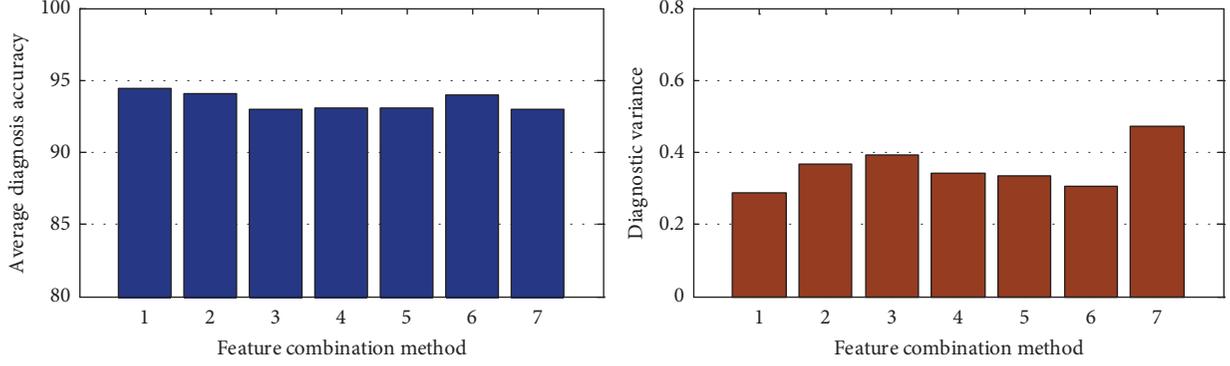


FIGURE 7: Fault diagnosis of hydraulic system under different combinations of features.

Source	SS	df	MS	Chi-sq	Prob > Chi-sq
Groups	14948.2	6	2491.36	36.1	2.63857e - 06
Error	13624.9	63	216.27		
Total	28573	69			

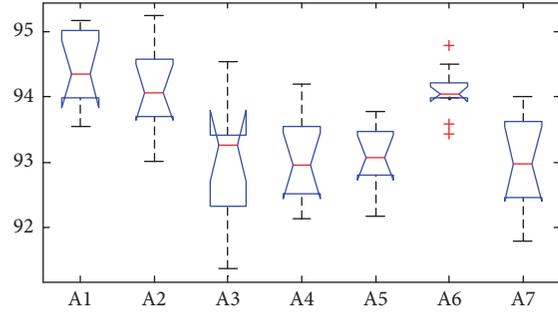


FIGURE 8: Nonparametric Kruskal-Wallis significance test results.

the classical ReliefF method; the fourth line (A_4) represents the feature set selected by MIV; the fifth row (A_5) and the sixth row (A_6) are the high-dimensional feature sequences obtained by dimension reduction using KPCA and LLE, respectively, where the dimension reduction target is set to be same as the feature numbers determined by CRSR. The last row (A_7) is the collection of original features. Taking the feature combinations mentioned above as input, the average and variance of fault diagnosis accuracy of hydraulic servo system calculated are shown in Figure 7, respectively.

According to the RBF network fault diagnostic results from Table 5 and Figure 7, the following can be obtained.

(1) Compared with using the original feature set directly, the feature selection methods are able to improve the precision of fault diagnosis based on RBF classifier, as the averages of the 10-fold cross validation from A_1 to A_7 are 94.405%, 94.217%, 93.022%, 93.045%, 93.077%, 94.069%, and 92.983%, respectively. The result indicates that the feature selection process plays a positive role in the classification task using high-dimensional data.

(2) Compared with the original ReliefF method, the introduction of compressed random subspace method eliminates those features with unstable contribution in the process of subspace integration and improves the performance and efficiency of feature selection. Eventually, higher fault diagnosis accuracy can be achieved with fewer input features.

(3) The variances calculated for A_1 (with redundancy analysis) and A_2 (without redundancy analysis) are 0.29 and 0.37, respectively. It can be seen that although the average diagnostic accuracy of them is close, the redundancy analysis

can optimize the information repetition among the features and reduce the computational resource consumption.

(4) Compared with the MIV-based feature selection method, CRSR has a great advantage in the diagnosis performance, which indicates that it has higher confidence in feature selection. Compared with KPCA and LLE, although the latter two (especially the LLE reduction algorithm) achieve high diagnostic accuracy as well, the variance of the 10-fold cross validation results for KPCA and LLE is higher than that of CRSR (0.29 and 0.31, resp.), which indicates that a certain degree of volatility exists in the results obtained by dimension reduction algorithms. In addition, due to the lack of clear interpretability and the existence of ambiguity in optimal target parameter setting, it is difficult to determine the best reduction method among the dimension reduction algorithms. Thus adaptive distance metrics based CRSR method has better applicability.

To further illustrate the rationality of CRSR method, the diagnostic results were tested by rank-based nonparametric Kruskal-Wallis test, which is based on fault diagnosis accuracy as a basic indicator in Table 5. The Kruskal-Wallis test was used to determine whether the different features sets have significant influence on the diagnostic performance. The significance threshold is set to 0.05, and the test results returned by Kruskal-Wallis function are shown in Figure 8. The p value returned by Kruskal-Wallis function is 0.00002 which is far less than 0.05. It indicates that the different combinations of features make significant impact on the fault diagnosis performance for hydraulic servo system.

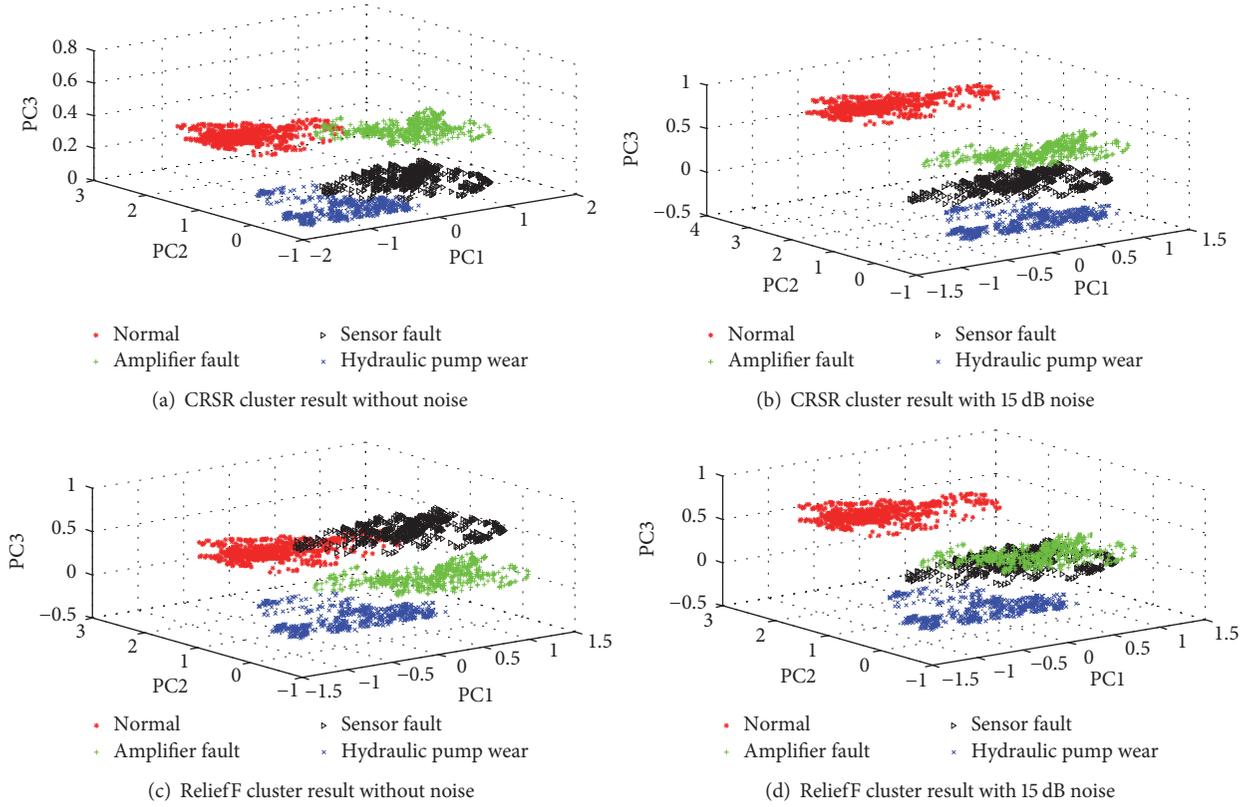


FIGURE 9: Fault feature vector clustering results under different noise conditions.

Moreover, another improvement of CRSSR is that the target optimization function is constructed based on the information entropy and the fuzzy theory. Compared with the distance measurement method used in the classical ReliefF model, the robustness of the feature selection process is improved in a complex environment. To illustrate the improvement of CRSSR mentioned above, this study analyzes the original output signal and the output signal with 15 dB noise. Meanwhile, PCA is used to project two sets of the selected feature sequences into three-dimensional space. Furthermore, the K -means method is used to cluster the feature sets of different fault modes. The results are shown in Figure 9.

It can be seen from Figure 9 that the feature qualities selected by CRSSR and ReliefF are higher before the noise is added, and the feature sets corresponding to each failure mode can be clearly distinguished by K -means method. After adding 15 dB of noise, due to the influence of external disturbances, the data points of the amplifier gain fault and the sensor constant deviation fault are obviously confused in the clustering result of ReliefF, which indicates that the feature differentiation declines. However, in the result of CRSSR method, spatial distribution boundaries between the four fault modes are still identifiable, which proves that the proposed objective function optimization method has good robustness against noise. Thus, CRSSR method is a promising technique for feature selection and subsequent fault diagnosis of hydraulic servo systems.

4. Conclusion

This study presents a fault diagnosis scheme for hydraulic servo system using compressed random subspace based ReliefF (CRSSR) method. Based on the feature selection structure of ReliefF, the proposed CRSSR method employs feature integration rules in the compressed domain and substitutes information entropy and fuzzy membership for traditional distance measure index. The advantage of the proposed method lies in the ability of determining the feature set with the better generalization performance and the less resource consumption. As a data-driven method, CRSSR could be practical and flexible in engineering.

To demonstrate the effectiveness of the proposed CRSSR method, validation data of three fault modes is generated through a hydraulic servo system joint simulation model. Comparing with existing feature reduction and feature selection methods, the result indicates that the feature selection process plays a positive role in the classification task using high-dimensional data, and CRSSR based fault diagnosis method has higher average accuracy and smaller variance. Meanwhile, the compressed random subspace method can eliminate those features with unstable contribution in the process of subspace integration and improve the performance and efficiency of feature selection. Besides, due to the robustness and stability of the information entropy and fuzzy theory based objective function optimization, the result shows that CRSSR method is more suitable for fault diagnosis problem under noisy conditions.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China [Grant nos. 51605014, 51105019, and 51575021], the Technology Foundation Program of National Defense [Grant no. Z132013B002], and the Fundamental Research Funds for the Central Universities [Grant no. YWF-16-BJ-J-18].

References

- [1] C. Lu, H. Yuan, L. Tao, and H. Liu, "Performance assessment of hydraulic servo system based on bi-step neural network and autoregressive model," *Journal of Vibroengineering*, vol. 15, no. 3, pp. 1546–1559, 2013.
- [2] X. He, "Fault diagnosis approach of hydraulic system using FARX model," in *Proceedings of the 2011 International Conference on Advanced in Control Engineering and Information Science, CEIS 2011*, pp. 949–953, China, August 2011.
- [3] H. Liu, X. Wang, and C. Lu, "Rolling bearing fault diagnosis under variable conditions using Hilbert-Huang transform and singular value decomposition," *Mathematical Problems in Engineering*, vol. 2014, Article ID 765621, 10 pages, 2014.
- [4] Z. Wang, C. Lu, Z. Wang, and J. Ma, "Health assessment of rotary machinery based on integrated feature selection and Gaussian mixed model," *Journal of Vibroengineering*, vol. 16, no. 4, pp. 1753–1762, 2014.
- [5] S.-Q. Ren, Y.-W. Fu, X. Li, and Z.-W. Zhuang, "Feature selection based on classes margin," *Journal of Software*, vol. 19, no. 4, pp. 842–850, 2008.
- [6] M. Van and H.-J. Kang, "Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 230, no. 2, pp. 291–302, 2016.
- [7] T. W. Rauber, F. de Assis Boldt, and F. M. Varejão, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 637–646, 2015.
- [8] M. Cerrada, R. V. Sánchez, D. Cabrera, G. Zurita, and C. Li, "Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal," *Sensors*, vol. 15, no. 9, pp. 23903–23926, 2015.
- [9] K. H. Hui, C. S. Ooi, M. H. Lim, M. S. Leong, and S. M. Al-Obaidi, "An improved wrapper-based feature selection method for machinery fault diagnosis," *PloS one*, vol. 12, no. 12, 2017.
- [10] J. Lin and L. Qu, "Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis," *Journal of Sound and Vibration*, vol. 234, no. 1, pp. 135–148, 2000.
- [11] I. K. Fodor, A survey of dimension reduction techniques. *Neoplasia* 7.5 (2002): 475-485.
- [12] H. Peng, X. Chen, and W. Xu, "Application of PCA feature extraction and SVM multi-classification on sensor fault diagnosis," *Journal of Data Acquisition and Processing*, vol. 25, no. 1, pp. 111–116, 2010.
- [13] Z.-B. Zhu and Z.-H. Song, "A novel fault diagnosis system using pattern classification on kernel FDA subspace," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6895–6905, 2011.
- [14] D. E. Goldberg, *Genetic Algorithm in Search, Optimization, and Machine Learning*, vol. xiii.7, 1989.
- [15] I. Iguyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [16] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "ReliefF for multi-label feature selection," in *Proceedings of the 2nd Brazilian Conference on Intelligent Systems, BRACIS 2013*, pp. 6–11, Brazil, October 2013.
- [17] Y. Huang, P. J. McCullagh, and N. D. Black, "An optimization of ReliefF for classification in large datasets," *Data & Knowledge Engineering*, vol. 68, no. 11, pp. 1348–1356, 2009.
- [18] I. Kononenko, M. Robnik, and S. U. Pompe, "ReliefF for estimation and discretization of attributes in classification, regression, and ILP problems," *Artificial Intelligence: Methodology, Systems, Applications*, 1996.
- [19] L.-X. Zhang, J.-X. Wang, Y.-N. Zhao, and Z.-H. Yang, "A novel hybrid feature selection algorithm: using ReliefF estimation for GA-Wrapper search," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, vol. 1, pp. 380–384, IEEE, Xi'an, China, November 2003.
- [20] F. Yang, W. Cheng, R. Dou, and N. Zhou, "An improved feature selection approach based on ReliefF and Mutual Information," in *Proceedings of the 2011 International Conference on Information Science and Technology, ICIST 2011*, pp. 246–250, China, March 2011.
- [21] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
- [22] T. K. Ho, "Nearest neighbors in random subspaces," in *Advances in pattern recognition (Sydney, 1998)*, vol. 1451 of *Lecture Notes in Comput. Sci.*, pp. 640–648, Springer, Berlin, 1998.
- [23] R. Duangsoithong and T. Windeatt, "Relevance and Redundancy Analysis for Ensemble Classifiers," in *Machine Learning and Data Mining in Pattern Recognition*, vol. 5632 of *Lecture Notes in Computer Science*, pp. 206–220, Springer, Berlin, Germany, 2009.
- [24] I. Kononenko, "Analysis and extension of RELIEF," in *Proceedings of the The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 1994.
- [25] Y. Sun, "Iterative RELIEF for feature weighting: algorithms, theories, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1035–1051, 2007.
- [26] N. Karayiannis, "MECA: maximum entropy clustering algorithm," in *Proceedings of the 1994 IEEE 3rd International Fuzzy Systems Conference*, pp. 630–635, Orlando, FL, USA.
- [27] D.-M. Zhao, J.-H. Wang, W. U. Jing, and J.-F. Ma, "Using fuzzy logic and entropy theory to risk assessment of the information security," in *Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC '05)*, vol. 4, pp. 2448–2453, IEEE, Guangzhou, China, August 2005.
- [28] R.-P. Li and M. Mukaidono, "Maximum-entropy approach to fuzzy clustering," in *Proceedings of the IEEE International Conference on Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and the Second International Fuzzy Engineering Symposium IEEE Xplore*, vol. 4, pp. 2227–2232, March 1995.

- [29] W. Xizhao and S. An, "Research on learning weights of fuzzy production rules based on maximum fuzzy entropy," *Journal of Computer Research Development*, vol. 43, no. 4, pp. 673–678, 2006.
- [30] P. Langley, "Selection of Relevant Features in Machine Learning," in *Proceedings of the Aaai Fall Symposium on Relevance*, 144, 140 pages, 1994.
- [31] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [32] K. Kira and L. A. Rendell, "Feature selection problem: traditional methods and a new algorithm," in *Proceedings of the Proceedings Tenth National Conference on Artificial Intelligence - AAAI-92*, pp. 129–134, July 1992.
- [33] H. Liu, J. Li, and L. Wong, "A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns," *Genome Informatics*, vol. 13, pp. 51–60, 2002.
- [34] T. Li, C. Zhang, and M. Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression," *Bioinformatics*, vol. 20, no. 15, pp. 2429–2437, 2004.
- [35] M. H. Nguyen and F. De la Torre, "Optimal feature selection for support vector machines," *Pattern Recognition*, vol. 43, no. 3, pp. 584–591, 2010.
- [36] L. Jiang and J. J. Zhang, "Dynamic Characteristics Simulation of Hydraulic Position Servo-system Based on AMESim," *Mechanical Engineering Automation*, 2007.
- [37] L. Wan X, B. Ding H, and Z. Zhou, "Study on dynamical simulation of hydraulic servo system based on AMESim and simulink," *Coal Mine Machinery*, 2007.
- [38] J. Cheng, W. Liu, and Z. Zhang, "Modeling and simulation for the electro-hydraulic servo system based on Simulink," in *Proceedings of the 2011 International Conference on Consumer Electronics, Communications and Networks, CECNet 2011*, pp. 466–469, China, April 2011.

Research Article

Damage Detection of Refractory Based on Principle Component Analysis and Gaussian Mixture Model

Changming Liu,¹ Di Zhou,² Zhigang Wang,³ Dan Yang,⁴ and Gangbing Song ⁵

¹The State Key Laboratory of Refractories and Metallurgy, Wuhan University of Science and Technology, Wuhan 430081, China

²Key Laboratory of Metallurgical Equipment and Control Technology, Wuhan University of Science and Technology, Wuhan 430081, China

³Ministry of Education, Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

⁴National Demonstration Center for Experimental Machinery Education, Wuhan University of Science and Technology, Wuhan 430081, China

⁵Smart Materials and Structures Laboratory, Department of Mechanical Engineering, University of Houston, Houston, TX 77204, USA

Correspondence should be addressed to Gangbing Song; gsong@uh.edu

Received 11 August 2017; Accepted 3 January 2018; Published 31 January 2018

Academic Editor: Michele Scarpiniti

Copyright © 2018 Changming Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acoustic emission (AE) technique is a common approach to identify the damage of the refractories; however, there is a complex problem since there are as many as fifteen involved parameters, which calls for effective data processing and classification algorithms to reduce the level of complexity. In this paper, experiments involving three-point bending tests of refractories were conducted and AE signals were collected. A new data processing method of merging the similar parameters in the description of the damage and reducing the dimension was developed. By means of the principle component analysis (PCA) for dimension reduction, the fifteen related parameters can be reduced to two parameters. The parameters were the linear combinations of the fifteen original parameters and taken as the indexes for damage classification. Based on the proposed approach, the Gaussian mixture model was integrated with the Bayesian information criterion to group the AE signals into two damage categories, which accounted for 99% of all damage. Electronic microscope scanning of the refractories verified the two types of damage.

1. Introduction

Structural health monitoring (SHM) has made significant advances in the past decades [1–8]. Monitoring of refractories, which are widely used in furnace, iron, and steel industries due to their ability to gain strength rapidly and to withstand aggressive environments and high temperature [9], receives increasing attention [10–13]. Temperature variations can lead to either interfacial separation between aggregates and matrix or microcracks, both depending on the range of coefficient of thermal expansion (CTE) mismatch between phases. Such effects modify all the thermomechanical properties of the material, especially Young's modulus (E) [10, 11, 14]. The AE technique has been developed over the last two decades as a nondestructive evaluation technique and

as a useful tool for material research [15–17]. It is an efficient method to monitor, in real time, damage growth in both structural components and laboratory specimens. This technique was often used to detect Young's modulus because it was correlated to AE activity variations considering the specific types of damage induced by CTE mismatch [12, 13]. The acoustic emission technique and the ultrasonic pulse echography technique, both carried out at high temperature, were applied as nondestructive characterization methods to monitor the damage extension within the materials submitted to thermal stress and to follow the evolution of the associated elastic properties [18, 19]. With this as a basis, the study could provide an important reference for thermal stress analysis under the AE data processing method. However, the AE signals generated by the complex structure of the refractory

are extremely complex even at normal temperature, which makes it difficult for the damage classification [20]. For this purpose, the AE signal parameters of the delay distribution, rise time, energy, and peak amplitude were selected to distinguish the effective features for different failure mechanism so that the two failure modes of fiber breakage and delamination can be distinguished [21, 22]. The related parameters can be modeled by a generative model, in particular a Gaussian mixture model (GMM) in the field of dimension processing [23, 24]. The global feature descriptor was formed by stacking the parameters of the adapted GMM (i.e., means, covariance, and weight) in a so-called supervector [25, 26]. Also, some scientists paid more attention to the parameter of the signal energy moment compared to the peak amplitude distribution in the study of the glass fiber composite materials and chose it to distinguish the fiber breakage and debonding crack. Moreover, the amplitude, ring count, and felicity ratio were found more suitable in the damage study of the B-AI composite [27]. However, much effort was put on the characterization of the overall parameters rather than on the data analysis of the damage mechanism.

Optionally, the dimensionality of the feature vectors can be reduced by a principal component analysis (PCA) [28]. The PCA was used to generate a new set of noncorrelated features to remove interference and to avoid using low variance variables (that was almost single-valued variables). Moreover, these new features were selected according to their discriminative capability. Subsequently, feature space modeling and classification were addressed by means of probabilistic self-organizing maps (SOM), a fuzzy version of classical SOM that allowed measuring the activation probability of each unit [29, 30]. Nevertheless, detecting not only an event but also the type was not a straightforward task, and previous approaches had not been able to obtain high per attack detection accuracy values. Scientists showed that the resulting GMM supervector encoding yielded an excellent representation for fuzzy parameters [31, 32]. This method was an outstanding technique for handling the description of multimodal data, making it robust with high computational efficiency [26]. Additionally, scientists employed support vector machines (SVM) to build individual classifiers per sample cluster [33, 34]. Such a SVM was a linear classifier trained by only one single positive sample and multiple negative samples; it was denoted as Exemplar-SVM. Therefore, secondly, using the features extracted by AE, the negative log likelihood was obtained by using the Bayesian GMM which was an outstanding technique for the multimodal distribution of the data with high computational efficiency [35, 36]. Among others, the PCA have been used successfully for object classification and scene classification. The PCA method is a statistical linear transformation selection from multiple variables to minor ones [28]. Meanwhile, the GMM is a Gauss probability density model, which can be used to accurately quantify matters and classify them into several models based on the Gauss probability density function [23].

Taking advantages of the PCA and GMM methods in the processing of the multidimensional models, especially the reduction of the AE parameters and pattern recognitions, this paper intends to reduce the correlation dimension of the

15 parameters of the AE signals emitted from the damage process of the materials and to obtain the two new parameters which could be used to describe the overall damage property without linear dependence. Afterwards, the GMM was used to classify the damage into two major categories. Finally, the results were verified experimentally by using scanning electron microscopy image based classification.

2. Analysis to Construct the New Characterization

The PCA method shows obvious advantages in the multiparameter dimension reduction problems and the construction process is clear to operate. The observation matrix of the sample is discussed by Shang et al. [37],

$$X = \begin{bmatrix} x & x & \cdots & x_p \\ x & x & \cdots & x_p \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_n & x_n & \cdots & x_{np} \end{bmatrix}, \quad (1)$$

where the rows of the sample matrix X represent the AE parameters and the columns correspond to different signals. The covariance matrix of the sample is

$$s = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}), \quad (2)$$

which is the estimation of x . Through the calculation of S , the characteristic quantity of the original observation matrix can be easily reconstructed so as to facilitate the sort of the features.

Step 1. The covariance matrix of the sample is constructed by S as follows:

$$S = \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_p) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \cdots & \text{cov}(X_2, X_p) \\ \vdots & \vdots & & \vdots \\ \text{cov}(X_p, X_1) & \text{cov}(X_p, X_2) & \cdots & \text{cov}(X_p, X_p) \end{bmatrix}, \quad (3)$$

where the matrix is a $P \times P$ and positive definite matrix, and there are characteristic values of P which are not equal to each other and greater than zero. Each characteristic value corresponds to a unit feature vector.

Step 2. Compute the P features and its characteristic vector. Set $\lambda_1, \lambda_2, \dots, \lambda_p$ to be eigenvalues of S . Meanwhile, $T = t_1, t_2, \dots, t_p$ are the corresponding unit feature vectors. Arranging eigenvalues in a descending order gives

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0. \quad (4)$$

Step 3. Define the contribution rate of the characteristic value

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (i = 1, 2, \dots, p) \quad (5)$$

and the accumulated contribution rate

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}. \quad (6)$$

Step 4. Based on the principle of the accumulated contribution $\sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i \geq 85\%$, the former m ($m \leq P$) principal components are picked, which means the former m mutually orthogonal eigenvector matrices are retained.

$$A = (t_1, t_2, \dots, t_m). \quad (7)$$

Step 5. Conduct linear correlation transformation between the new feature vector matrix and the original one. In this way, the original P dimension index will be reduced to m , which contains the ultimate information with mutual linear independence.

3. Classification with GMM

The GMM probability density function is set as follows:

$$p(x) = \sum_{k=1}^M \omega_k p_k(x) = \sum_{k=1}^M \omega_k N\left(x \mid \mu_k, \sum_k\right), \quad (8)$$

where M is the mixed number of the model; ω_k is the weighting coefficient of the model, and $\sum \omega_k = 1$; $N(x \mid \mu_k, \sum_k)$ is the k th single Gauss probability density function, which is depicted as

$$N(x \mid \mu, \sum) = \frac{e^{-(1/2)(x-\mu)^T \sum^{-1}(x-\mu)}}{(2\pi)^{n/2} |\sum|^{1/2}}. \quad (9)$$

The proper parameters were evaluated as

$$\theta = \left[\omega_1, \omega_2, \omega_3, \dots, \omega_M, \mu_1, \mu_2, \mu_3, \dots, \mu_M, \sum_1, \sum_2, \sum_3, \dots, \sum_M \right] \quad (10)$$

which makes the max maximum likelihood estimator of the probability density function,

$$\begin{aligned} J(\theta) &= \ln \left[\prod_{i=1}^M p(x_i) \right] = \sum_{i=1}^M \ln p(x_i) \\ &= \sum_{k=1}^M \ln \left[\omega_k N(x \mid \mu_k, \sigma_k^2) \right]. \end{aligned} \quad (11)$$

In order to obtain the maximum likelihood estimate, the GMM will be evaluated by the maximum expected value algorithm. The iteration steps are as follows.

Step 1. Initiate the parameters:

- (1) Set the mean values to be random values.
- (2) Set the covariance matrix $\sum \sum \sum \dots \sum_M$ to be the unit matrix.
- (3) Set the weighting coefficient $\omega_1, \omega_2, \omega_3, \dots, \omega_M$ of each model to be the prior probability of each model:

$$\omega_i = \frac{1}{M}, \quad (12)$$

where M was the number of GMM.

Step 2. Compute the prior probability of each item in the model:

$$\Pr(i \mid x_t, \theta^k) = \frac{\omega_k N(x_t \mid \mu_i^k, \sum_i^k)}{\sum_{k=1}^M \omega_k N(x_t \mid \mu_i^k, \sum_i^k)}. \quad (13)$$

Step 3. Update the parameters by the prior probability:

$$\begin{aligned} \omega_i^{k+1} &= \frac{1}{T} \sum_{t=1}^T \Pr(i \mid x_t, \theta^k). \\ \mu_i^{k+1} &= \frac{\sum_{t=1}^T \Pr(i \mid x_t, \theta^k) x_t}{\sum_{t=1}^T \Pr(i \mid x_t, \theta^k)} \\ \sum_i^{k+1} &= \frac{\left(\sum_{t=1}^T \Pr(i \mid x_t, \theta^k) (x_t - \mu_i^{k+1})(x_t - \mu_i^{k+1})^T \right)}{\sum_{t=1}^T \Pr(i \mid x_t, \theta^k)} \end{aligned} \quad (14)$$

Step 4. Repeat Steps 2 and 3 until the convergence:

$$|\theta^{t+1} - \theta^t| < \varepsilon, \quad (15)$$

where θ^{t+1} and θ^t are the parameters estimation of the previous and current step and ε is the set threshold, which is usually set to 10^{-5} .

4. Experimentation and Verification

4.1. Specimens and Experiment Setup. The industrial refractory tested in the study is composed of magnesia aggregates, carbon binder (phenolic resin and/or pitch), and other components. Figure 1 shows the microstructure of such a refractory without damage. The magnesia aggregates are formed by sintering of crystallites with weak interfaces. The size of magnesia grains varies from less than one half millimeters to five millimeters. The other grains impurities, such as SiO₂ and Al₂O₃, with less than 5 mm sizes, are founded scattered in the matrix, the carbon binder.

The components were mixed and shaped into bricks at low temperature (20~50°C) and under high pressure (150 MPa around) [12]. Then the bricks undergo heat treatment (100~200°C) to start the polymerization of resin and to eliminate residual water and phenols [12]. Under these sintering conditions, we directly obtain three specimens with a rectangular cross-section of 140 mm × 25 mm × 25 mm. Table 1 provides the composition of the materials after the heat treatment.

TABLE I: The composition of magnesia carbon refractory.

MgO	Chemical composition, % (weight)			Phase composition, % (volume)	
	C	SiO ₂	Al ₃ O ₃	Aggregate phase	Continuous phase
82.95	13.07	0.72	0.61	65	35

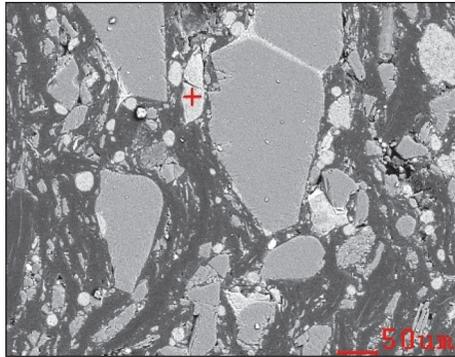


FIGURE 1: Typical microstructure of magnesia carbon refractory.

According to ASTM C1161-13 [38], the shape of the samples for 3-point bending should be rectangular and the size of the samples should be 3 by 4 by 45 to 50 mm minimum with 40 mm outer span 3-point bending. Therefore, the specimens were prepared to be 140 mm × 25 mm × 25 mm. The 3-point bending tests were performed using a HMOR/STRAIN loading machine. The crosshead speed of the machine was fixed at 0.05 mm/min. The tests were executed on three specimens of each configuration in order to ensure the accuracy of the results (Figure 2).

AE is defined as phenomena whereby transient elastic waves are generated by the rapid release of energy from localized sources within a material (or structure). The AE, which represents the generation of transient ultrasonic waves due to damage development within the material under load, is an efficient technique for structural health monitoring, as discussed elsewhere [39–41]. When a material is subjected to solicitations (such as mechanical and thermal), acoustic emission can be generated by a variety of sources, including crack nucleation and propagation, multiple dislocation slip, twinning, grain boundary sliding, phase transformations in alloys, debonding of grain in composite materials, or fracture of inclusions in alloys. This technique has been used at either the laboratory level or industrial scale. Usually, this technique is applied at room temperature as a nondestructive characterization technique in order to follow in real time the evolution of the damage of a material subjected to mechanical loading. Here, the upper surface of the sample should be slightly polished to remove the burr in order to locate the AE sensor. Then the coupling agent was coated on the polished zone and the AE sensor was fixed on the coupling agent surface with the adhesive tape. The stress wave was passed from the surface of the sample to the AE sensor through the coupling agent. The application of AE technique aims to characterize the material microdamage at a very local scale.

The device of acquisition (Figure 3) is composed of a wide band (175 kHz~1 MHz) sensor (PAC MICROPHONE μ 80), a

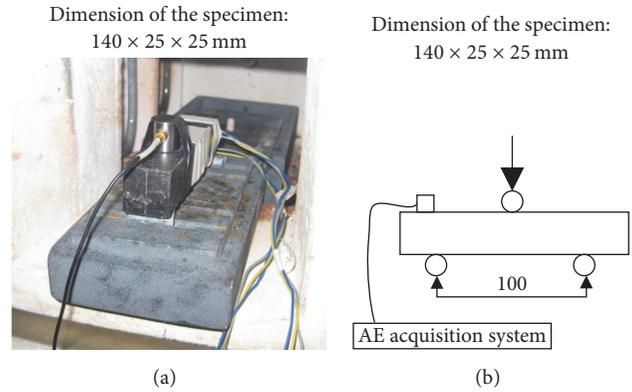


FIGURE 2: Experimental setup for three-point bend test at room temperature. (a) The physical diagram and (b) the schematic one. The AE sensor is fixed on the specimen through the adhesive tape. The pressure actuating pressure head is imposed on the middle of the specimen.

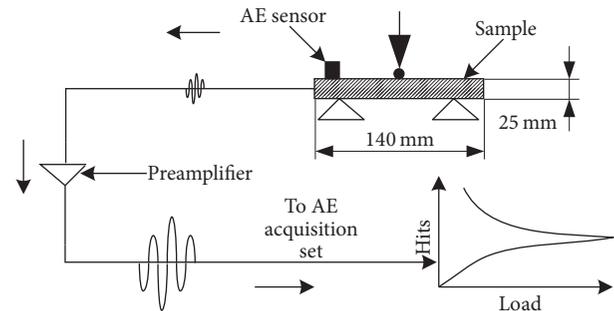


FIGURE 3: Experimental setup used for acoustic emission measurements [12].

preamplifier (EPA 1220A), and an acquisition card associated with a computer (AEDSP-32/16 MISTRAS digital system from Physical Acoustics Corporation). The AE sensor is a major element of the chain of acquisition because it collects the whole of the signals induced by the elastic waves created within the material whose amplitudes are higher than a fixed threshold in order to amplify and to record them. This system records the waveform and the main feature parameters well known in AE study such as count, hit, rise time, duration of hit, count to peak, and amplitude (in dB). Figure 4 presents different AE features extracted from the signal waveform.

4.2. PCA Parameter Reduction. The AE damage signals of the Mg-O refractory during the three-point bend test were collected and 15 parameters were directly obtained: rise time (X_1), count (X_2), energy (X_3), duration (X_4), amplitude (X_5), mean frequency (X_6), RMS (X_7), ASL (X_8), peak frequency

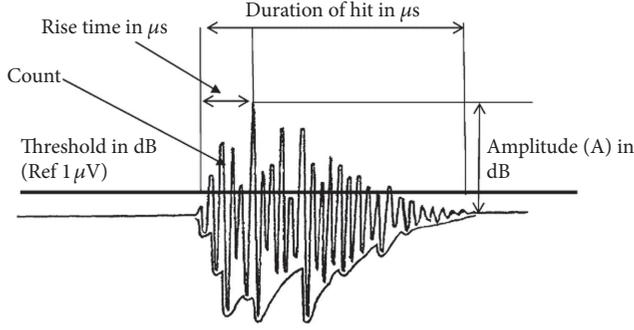


FIGURE 4: Typical AE features extracted from the recorded signal (hit) [12].

(X_9), inverse calculation frequency (X_{10}), original frequency (X_{11}), signal strength (X_{12}), absolute energy (X_{13}), centroid frequency (X_{14}), and peak frequency (X_{15}). The number of the sample signals was 11168 and the observation matrix was 11168×15 . In order to eliminate the disturbance of dimensionless parameter, the observation sample matrix was normalized before the principal component analysis and the data values were normalized to (0, 1). The covariance matrix's eigenvalues are shown in Table 2. The cumulative contribution rate of each principal component was shown in Table 3. It can be seen from the table that the cumulative contribution rate of the first two principal components is 90%, which is far greater than 85%, meaning the first two principal components are sufficient enough to replace the overall clustering index. Therefore, the new principal components are produced and the number of the parameters is reduced from 15 to 2.

4.3. Classification of the Damage Signals. For the application of GMM classification of the damage signals of the refractory, the increase in the number of the model can improve the accuracy of the model, however with increased complexity of the model, as discussed by Jiang et al. [24]. The Bayesian information criterion (BIC) has the ability to maintain the balance between the accuracy and complexity of the model; therefore, it is adopted to classify the damage.

$$\text{BIC} = -2 \ln L + k \ln T, \quad (16)$$

where L is the maximum of the likelihood function of the estimated model, T is the number of observations, and k is the number of the free parameters to be estimated in each GMM.

When the number of the model is increased from M to $M + 1$, the changing rate of the BIC is

$$\xi_{M+1} = \frac{\text{BIC}_M - \text{BIC}_{M+1}}{\text{BIC}_M} \times 100\%. \quad (17)$$

The changing rate of the BIC reflects the sensitivity of the BIC values to the number of the models. When the number of the models is increased from M to $M + 1$, the change rate of the BIC is large, which means that the number M is insufficient in the description of the original data set accuracy and should be increased to $M + 1$. When the change rate of the BIC

TABLE 2: Eigenvalues of covariance matrix.

Parameter	Value
λ_1	9.13
λ_2	6.68
λ_3	0.71
λ_4	0.36
λ_5	0.31
λ_6	0.16
λ_7	0.06
λ_8	0.031
λ_9	5.2×10^{-3}
λ_{10}	3.3×10^{-3}
λ_{11}	7.5×10^{-4}
λ_{12}	1.6×10^{-4}
λ_{13}	8.2×10^{-5}
λ_{14}	4.0×10^{-6}
λ_{15}	1.5×10^{-6}

TABLE 3: Accumulated contribution rate of each component.

Principle component index	Cumulative contribution%
1	52.3135
2	90.6189
3	94.6705
4	96.7067
5	98.5072
6	99.432
7	99.7786
8	99.9456
9	99.9756
10	99.9943
11	99.9986
12	99.9995
13	100
14	100
15	100

is small, M and $M + 1$ have little difference in the description of the original data and M is enough for the description.

The changing rate of the BIC is shown in Figure 5. It can be seen from the graph that when the number of the model is increased from 1 to 2, the change of the BIC is significant, reaching 7%. With the increase of the number of the models, the changing rate of the BIC gradually decreases (<3%). Therefore, the model number of 2 is chosen to describe the observed data set.

The GMM operation results are shown in Figure 6. It can be seen from the plot that the damage signal is divided into two categories of ω_1 and ω_2 , whose weights are 0.63 and 0.37, respectively.

4.4. Verification. The Philips scanning electron microscopy (SEM, PSEM 500) and energy spectrometer (AMETEK) were

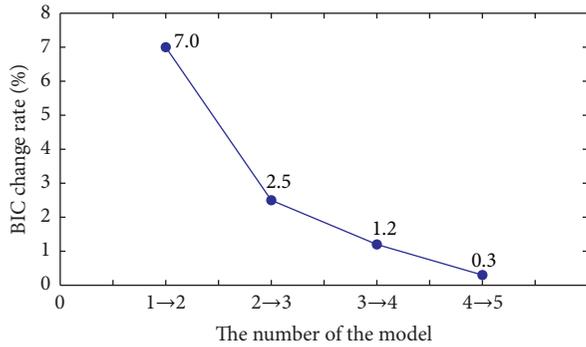


FIGURE 5: BIC change rate of each model number.

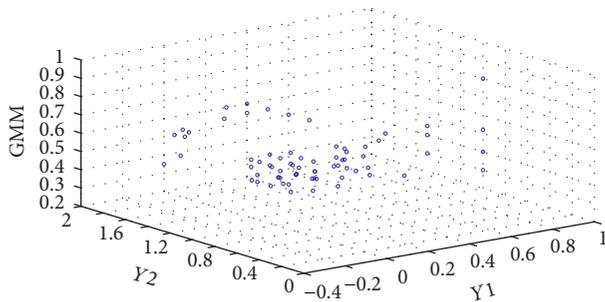


FIGURE 6: Computation results of the GMM algorithm ($\omega_1 = 0.63$, $\omega_2 = 0.37$).

used for scanning analysis of the damage sample. The microscopic scanning results reveal that the damage forms were mainly the matrix and the interphase damage, as shown in Figures 7 and 8, and the matrix damage accounted for the larger proportion. The energy spectra results of the two kinds of microdamage are shown in Figures 9 and 10, respectively. It can be seen from Figure 9 that the main component of the matrix damage area is C with the mass fraction of 97% and the crack can be regarded as the matrix crack. From Figure 10, the composition of the observation area near the interface was C with the mass fraction of 65% and O and Mg with the mass fraction of 15% and 18%, respectively, which indicates that the matrix and the particle phase existed in the area and the crack is the interfacial crack. The SEM results show that the main damage form of the MgO-C refractory is the matrix and the interface damage and the matrix phase damage accounted for the larger proportion. Therefore, the classification results using the method of the PCA and GMM are verified.

5. Conclusions

In this paper, a new AE data processing method of merging the similar parameters in the description of damage to reduce the dimension was developed. In the proposed method, the AE damage signals of the Mg-O refractory during the three-point bend test were collected and 15 parameters were directly obtained: rise time (X_1), count (X_2), energy (X_3), duration (X_4), amplitude (X_5), mean frequency (X_6), RMS (X_7), ASL (X_8), peak frequency (X_9), inverse calculation

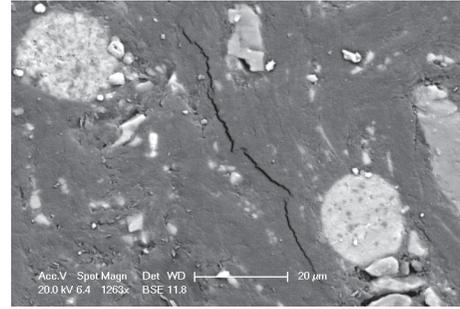


FIGURE 7: The matrix damage crack.

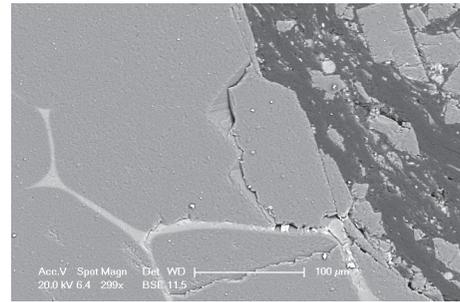


FIGURE 8: The interface damage crack.

frequency (X_{10}), original frequency (X_{11}), signal strength (X_{12}), absolute energy (X_{13}), centroid frequency (X_{14}), and peak frequency (X_{15}). The observation sample matrix was firstly normalized before the principal component analysis and the data values were normalized to (0, 1). The cumulative contribution rate of each principal component was calculated to successfully select the first two principal components of 90% contribution. Therefore, the new principal components were produced and the number of the parameters was reduced from 15 to 2.

Then the Gaussian mixture model was used to classify the damage of the refractory according to the 2 damage indexes, which could be utilized to describe the overall damage property without linear dependence. Afterwards, the damage was classified into two major categories of ω_1 and ω_2 with the damage weight of 63% and 37%, respectively. In order to verify the proposed method, the Philips scanning electron microscopy and energy spectrometer were used for scanning analysis of the sample. The scanning results showed that the damage form was indeed observed as 2 damage forms of mainly the matrix and the interphase damage. The main component of the matrix damage area was C with the mass fraction of 97%. In the interface damage crack area, C was with the mass fraction of 65%, and O and Mg were with the mass fraction of 15% and 18%, respectively, which indicated that the matrix and the particle phase existed in the area and the crack was the interfacial crack. At last, the SEM results showed that the main damage form of the MgO-C refractory was the matrix and the interface damage and the matrix phase damage accounted for the larger proportion.

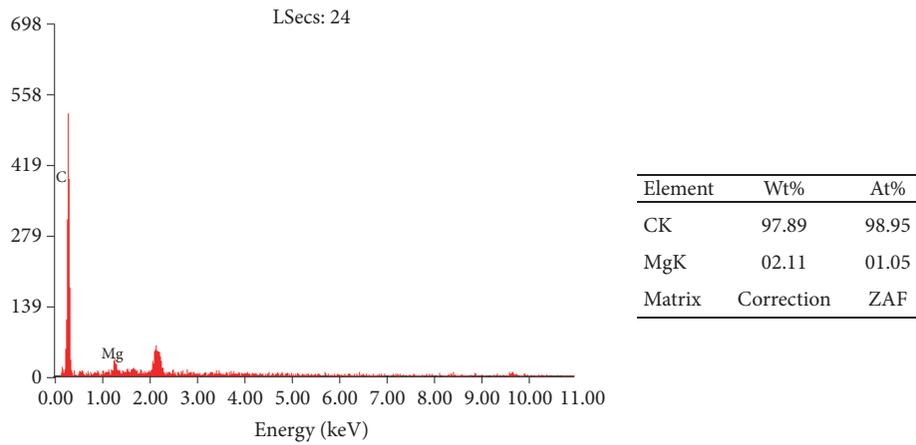


FIGURE 9: EDS analysis of the matrix damage crack.

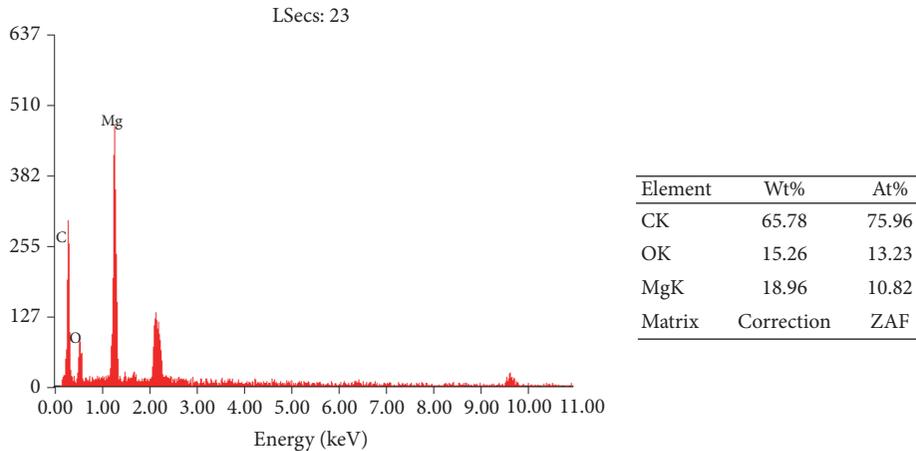


FIGURE 10: EDS analysis of the matrix damage crack.

Therefore, the classification results using the method of the PCA and GMM were verified.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Changming Liu and Zhigang Wang conceived and designed the experiments; Changming Liu performed the experiments; Dan Yang analyzed the data; Changming Liu, Zengbing Xu, and Gangbing Song contributed to materials/analysis tools development; Changming Liu, Zhigang Wang, and Gangbing Song wrote the paper.

Acknowledgments

The authors would like to thank the National Natural Science Foundation of China (51505346, 51375354, and 51405353), the Natural Science Foundation of Hubei Province (2014CFB825), and the Open Foundation of the State Key

Laboratory of Refractories and Metallurgy for financial support.

References

- [1] P. Selva, O. Cherrier, V. Budinger, F. Lachaud, and J. Morlier, "Smart monitoring of aeronautical composites plates based on electromechanical impedance measurements and artificial neural networks," *Engineering Structures*, vol. 56, pp. 794–804, 2013.
- [2] D. M. Peairs, G. Park, and D. J. Inman, "Improving Accessibility of the Impedance-Based Structural Health Monitoring Method," *Journal of Intelligent Material Systems and Structures*, vol. 15, no. 2, pp. 129–139, 2016.
- [3] S. Ritdumrongkul, M. Abe, Y. Fujino, and T. Miyashita, "Quantitative health monitoring of bolted joints using a piezoceramic actuator-sensor," *Smart Materials and Structures*, vol. 13, no. 1, pp. 20–29, 2004.
- [4] Y.-K. An and H. Sohn, "Integrated impedance and guided wave based damage detection," *Mechanical Systems and Signal Processing*, vol. 28, pp. 50–62, 2012.
- [5] G. Lu, Y. Li, T. Wang et al., "A multi-delay-and-sum imaging algorithm for damage detection using piezoceramic

- transducers,” *Journal of Intelligent Material Systems and Structures*, vol. 9, pp. 1–10, 2016.
- [6] H. Xiao, J. Zheng, and G. Song, “Severity evaluation of the transverse crack in a cylindrical part using a PZT wafer based on an interval energy approach,” *Smart Materials and Structures*, vol. 25, no. 3, article 035021, 2016.
 - [7] T. Wang, G. Song, Z. Wang, and Y. Li, “Proof-of-concept study of monitoring bolt connection status using a piezoelectric based active sensing method,” *Smart Materials and Structures*, vol. 22, article 87001, 2013.
 - [8] Q. Feng, Q. Kong, and G. Song, “Damage detection of concrete piles subject to typical damage types based on stress wave measurement using embedded smart aggregates transducers,” *Measurement*, vol. 88, pp. 345–352, 2016.
 - [9] M. Ghassemi Kakroudi, E. Yeugo-Fogaing, M. Huger, C. Gault, and T. Chotard, “Influence of the thermal history on the mechanical properties of two alumina based castables,” *Journal of the European Ceramic Society*, vol. 29, no. 15, pp. 3197–3204, 2009.
 - [10] C. Liu, Z. Wang, and Y. Li, “Innovative method on simulating the damage mechanism of the refractory,” *ISIJ International*, vol. 53, no. 7, pp. 1275–1279, 2013.
 - [11] G. Briche, N. Tessier-Doyen, M. Huger, and T. Chotard, “Investigation of the damage behaviour of refractory model materials at high temperature by combined pulse echography and acoustic emission techniques,” *Journal of the European Ceramic Society*, vol. 28, no. 15, pp. 2835–2843, 2008.
 - [12] C. Liu, Z. Wang, and Y. Li, “Damage pattern recognition and feature extraction of mgO-C refractory,” *ISIJ International*, vol. 53, no. 7, pp. 1280–1285, 2013.
 - [13] N. Schmitt, Y. Berthaud, and J. Poirier, “Tensile behaviour of magnesia carbon refractories,” *Journal of the European Ceramic Society*, vol. 20, no. 12, pp. 2239–2248, 2000.
 - [14] T. Chotard, J. Soro, H. Lemerrier, M. Huger, and C. Gault, “High temperature characterisation of cordierite-mullite refractory by ultrasonic means,” *Journal of the European Ceramic Society*, vol. 28, no. 11, pp. 2129–2135, 2008.
 - [15] C. Xu, P. Gong, J. Xie, H. Shi, G. Chen, and G. Song, “An acoustic emission based multi-level approach to buried gas pipeline leakage localization,” *Journal of Loss Prevention in the Process Industries*, vol. 44, pp. 397–404, 2016.
 - [16] W. Li, S. C. M. Ho, D. Patil, and G. Song, “Acoustic emission monitoring and finite element analysis of debonding in fiber-reinforced polymer rebar reinforced concrete,” *Structural Health Monitoring*, 2016.
 - [17] W. Li, Q. Kong, S. C. M. Ho et al., “Feasibility study of using smart aggregates as embedded acoustic emission sensors for health monitoring of concrete structures,” *Smart Materials and Structures*, vol. 25, no. 11, article 115031, 2016.
 - [18] J. Vandenbussche, P. Lee, and J. Peuteman, “Round-Off Noise of Multiplicative FIR Filters Implemented on an FPGA Platform,” *Applied Sciences*, vol. 4, no. 2, pp. 99–127, 2014.
 - [19] A. Martini, M. Troncosi, and A. Rivola, “Leak detection in water-filled small-diameter polyethylene pipes by means of acoustic emission measurements,” *Applied Sciences*, vol. 7, no. 1, article 7010002, 2017.
 - [20] C. Liu, Z. Wang, Y. Li et al., “Damage pattern recognition of refractory materials based on bp neural network,” *LNCS*, pp. 431–440, 2012.
 - [21] A.-B. A. E. Mohamad and Z. Chen, “Experimental and numerical analysis of the compressive and shear behavior for a new type of self-insulating concrete masonry system,” *Applied Sciences*, vol. 6, no. 9, article 6090245, 2016.
 - [22] G. Lacidogna, P. Cutugno, G. Niccolini, S. Invernizzi, and A. Carpinteri, “Correlation between earthquakes and AE monitoring of historical buildings in seismic areas,” *Applied Sciences (Switzerland)*, vol. 5, no. 4, pp. 1683–1698, 2015.
 - [23] V. Christlein, D. Bernecker, F. Hönig, A. Maier, and E. Angelopoulou, “Writer Identification Using GMM Supervectors and Exemplar-SVMs,” *Pattern Recognition*, vol. 63, pp. 258–267, 2017.
 - [24] Q. Jiang, B. Huang, and X. Yan, “GMM and optimal principal components-based Bayesian method for multimode fault diagnosis,” *Computers & Chemical Engineering*, vol. 84, pp. 338–349, 2016.
 - [25] S. K. Al-Jumaili, K. M. Holford, M. J. Eaton, and R. Pullin, “Parameter Correction Technique (PCT): a novel method for acoustic emission characterisation in large-scale composites,” *Composites Part B: Engineering*, vol. 75, pp. 336–344, 2015.
 - [26] R. Ahmed, A. Temko, W. Marnane, G. Lightbody, and G. Boylan, “Grading hypoxic-ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine,” *Clinical Neurophysiology*, vol. 127, no. 1, pp. 297–309, 2016.
 - [27] M. A. Wright and B. D. Intwala, “The effect of elevated temperatures on the mechanical properties of B-Al composites,” *Journal of Materials Science*, vol. 8, no. 7, pp. 957–963, 1973.
 - [28] R. Sharifi and R. Langari, “Nonlinear sensor fault diagnosis using mixture of probabilistic PCA models,” *Mechanical Systems and Signal Processing*, vol. 85, pp. 638–650, 2017.
 - [29] H. Li, Z. Chen, Y. Sun, and H. Karimi, “Stabilization for a class of nonlinear networked control systems via polynomial fuzzy model approach,” *Complexity*, vol. 21, no. 2, pp. 74–81, 2015.
 - [30] X. Yuan, Z. Ge, H. Zhang, Z. Song, and P. Wang, “Soft sensor for multiphase and multimode processes based on Gaussian mixture regression,” in *Proceedings of the 19th IFAC World Congress on International Federation of Automatic Control, IFAC 2014*, vol. 47, pp. 1067–1072, August 2014.
 - [31] H. C. Siu, J. A. Shah, and L. A. Stirling, “Classification of anticipatory signals for grasp and release from surface electromyography,” *Sensors*, vol. 16, no. 11, article no. 1782, 2016.
 - [32] S. Hamel, A. Boulkroune, and A. Bouzeriba, “Function vector synchronization based on fuzzy control for uncertain chaotic systems with dead-zone nonlinearities,” *Complexity*, vol. 21, no. S1, pp. 234–249, 2016.
 - [33] X. Zhao, W. Li, L. Zhou et al., “Application of support vector machine for pattern classification of active thermometry-based pipeline scour monitoring,” *Structural Control and Health Monitoring*, vol. 22, no. 6, pp. 903–918, 2015.
 - [34] Q. Hou, W. Jiao, L. Ren, H. Cao, and G. Song, “Experimental study of leakage detection of natural gas pipeline using FBG based strain sensor and least square support vector machine,” *Journal of Loss Prevention in the Process Industries*, vol. 32, pp. 144–151, 2014.
 - [35] C. Turgay, “Bayesian change detection based on spatial sampling and Gaussian mixture model,” *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1635–1642, 2011.
 - [36] B. Zhang, C. Zhang, and X. Yi, “Active curve axis Gaussian mixture models,” *Pattern Recognition*, vol. 38, no. 12, pp. 2351–2362, 2005.
 - [37] J. Shang, M. Chen, H. Ji, and D. Zhou, “Recursive transformed component statistical analysis for incipient fault detection,” *Automatica*, vol. 80, pp. 313–327, 2017.

- [38] *American Standard, C1161, 2002. Standard Test Method for Flexural Strength of Advanced Ceramics at Ambient Temperature*, ASTM International, West Conshohocken, PA, USA, 2002.
- [39] A. Benavent-Climent, A. Gallego, and J. M. Vico, "An acoustic emission energy index for damage evaluation of reinforced concrete slabs under seismic loads," *Structural Health and Monitoring*, vol. 11, no. 1, pp. 69–81, 2012.
- [40] D. G. Aggelis, D. V. Soulioti, E. A. Gatselou, N.-M. Barkoula, and T. E. Matikas, "Monitoring of the mechanical behavior of concrete with chemically treated steel fibers by acoustic emission," *Construction and Building Materials*, vol. 48, pp. 1255–1260, 2013.
- [41] M. K. Elbatouny, P. H. Ziehl, A. Larosche, J. Mangual, F. Matta, and A. Nanni, "Acoustic emission monitoring for assessment of prestressed concrete beams," *Construction and Building Materials*, vol. 58, pp. 46–53, 2014.

Research Article

Lithium-Ion Battery Capacity Estimation: A Method Based on Visual Cognition

Yujie Cheng, Laifa Tao, and Chao Yang

School of Aeronautic Science and Engineering, Beihang University, Beijing, China

Correspondence should be addressed to Laifa Tao; taolaifa@buaa.edu.cn

Received 18 August 2017; Accepted 16 November 2017; Published 17 December 2017

Academic Editor: Rafał Burdzik

Copyright © 2017 Yujie Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study introduces visual cognition into Lithium-ion battery capacity estimation. The proposed method consists of four steps. First, the acquired charging current or discharge voltage data in each cycle are arranged to form a two-dimensional image. Second, the generated image is decomposed into multiple spatial-frequency channels with a set of orientation subbands by using non-subsampled contourlet transform (NSCT). NSCT imitates the multichannel characteristic of the human visual system (HVS) that provides multiresolution, localization, directionality, and shift invariance. Third, several time-domain indicators of the NSCT coefficients are extracted to form an initial high-dimensional feature vector. Similarly, inspired by the HVS manifold sensing characteristic, the Laplacian eigenmap manifold learning method, which is considered to reveal the evolutionary law of battery performance degradation within a low-dimensional intrinsic manifold, is used to further obtain a low-dimensional feature vector. Finally, battery capacity degradation is estimated using the geodesic distance on the manifold between the initial and the most recent features. Verification experiments were conducted using data obtained under different operating and aging conditions. Results suggest that the proposed visual cognition approach provides a highly accurate means of estimating battery capacity and thus offers a promising method derived from the emerging field of cognitive computing.

1. Introduction

Lithium-ion (Li-ion) batteries, featuring high energy density and light in weight, are becoming more and more popular for various applications, especially in the field of aerospace and electric vehicles [1–3]. Thus, the majority of existing studies focus on ways to improve the performance of Li-ion batteries. Battery capacity, which is regarded as an important indicator of the battery performance, is highly affected by various internal and external mechanisms such as ambient temperature, aging, and usage patterns; these factors cause battery performance to gradually fade over time. Therefore, available battery capacity needs to be accurately estimated for reliability purposes and for the proper management of battery use [4].

Recent studies have reported a variety of approaches to estimating the capacity of Li-ion batteries. Most of the existing approaches are model-based methods, including electrochemical [5], equivalent circuit-based [6], and analytical [7, 8] models. These models are mostly based on complex physical and chemical processes that take into account the

dynamic behavior of batteries [9–11], and the estimation performance is highly dependent on the accuracy of the models. In particular, these types of models are usually difficult to establish given the restrictions on acquisition of knowledge of the electrochemical parameters, aging mechanisms, and properties of batteries [12]. Moreover, these models are individually dependent on the specific type of battery in terms of production processes, electrolytes, and anode and cathode materials. State-of-charge- (SOC-) open-circuit-voltage- (OCV-) based methods for in-cycle capacity estimation are widely applied in many real-world applications [13, 14]. However, the SOC-OCV-based methods rely on accurate SOC and OCV values, which are usually highly time consuming to obtain [10, 15]. Regardless of which modelling methods are used to model the battery state, the laboratory determined battery charging and discharging characteristics under different operating conditions are a source of knowledge about battery behavior. In some applications, these original data stored as discrete values are employed to create a lookup table database on the charge status of the master

battery. However, using this kind of method for Li-ion battery capacity estimation, amounts of experiments need to be performed under various operating conditions for a whole lifetime to obtain battery capacities in different life states under different operating conditions. Otherwise, database-based method will have a low accuracy with a rough database. Tao et al. [16] proposed a capacity estimation method for Li-ion battery based on similarity recognition of online data curves, which can be considered as an intelligent method based on database. Even though this method achieves a high accuracy, it takes a lot of time looking for the most similar data curve contained in the database, which restricts its real application.

A new geometric-based method was proposed in [17]; this method differs from those aforementioned and appears to be an extension of the traditional constant current-constant voltage method [18]. It estimates battery capacity by combining differential geometry and four geometric features that are sensitive to capacity fade. The four geometric features are extracted from the charging current (CC) and discharging voltage (DV) curves, including the time duration of the constant voltage (CV) curve, maximum radius of curvature of the CV stage, area under the CV curve, and slope of the voltage curve in the early stage of the discharge process. Experimental results provided in their paper demonstrate the effectiveness of the geometric-based method.

The nature of the geometric-based method aims to introduce differential geometry theory and traditional geometric features into battery capacity estimation. Inspired by the work in [17], we attempt to introduce other leading-edge interdisciplinary methods to battery capacity estimation, avoiding complicated analyses of physical-chemical processes and achieving an accurate cognition of degradation processes, thereby further enhancing the effectiveness and accuracy of battery capacity estimation.

Cognitive science is an interdisciplinary study that consists of multiple research disciplines, including psychology, artificial intelligence, philosophy, neuroscience, linguistics, and anthropology. It includes research on intelligence and behavior, especially focusing on how information is represented, processed, and transformed within nervous systems and machines [19]. Cognitive science is a large field and covers a wide array of topics relating to cognition such as language processing, artificial intelligence, and visual and auditory cognition. Among these topics, visual cognition has become the focus of many studies in cognitive science and is becoming a significant topic of interest in the twenty-first century [20]. In recent years, countries around the world have invested heavily to support research in visual cognition. In the US, the Defense Advanced Research Projects Agency launched a special research program in 2007 named “Cognitive Computing,” under which visual cognition is a key research target. In Japan, experts in the field of computer vision have been brought into the “Brain Plan” over the past decade to promote the interdisciplinary studies of brain cognitive science and visual cognition. The National Natural Science Foundation Committee of China initiated a major research project in 2008 called “Cognitive Computing Based on Visual and Auditory Information”; its purpose is to establish a new computational method based on human

visual and auditory cognitive mechanisms, thereby providing new ideas for image understanding and voice processing. Today, computing methods based on visual cognition have received extensive attention and are widely used in face recognition [21], image fusion [22], texture classification [23], and so forth. However, in the field of Li-ion battery capacity estimation, methods based on visual cognition have rarely been reported. Motivated by this, we attempt to transform CC values and DV values into a two-dimensional image and thus to further advance battery capacity estimation using a visual cognition method.

Essentially, visual cognition is a kind of bionic science; that is, it deals with the recognition of objects based on the characteristics of the human visual system (HVS). One of the well-known characteristics of the HVS is the multichannel characteristic (MCC), meaning that there are multiple spatial-frequency channels in the processing of pictorial information in the HVS, each of which further involves different number of orientation components depending upon a predetermined setting of series [24]. In this study, the authors employ MCC to extract degradation feature information from CC and DV data, which is the core of this study and also the difference distinguishing our method from other existing methods, including that in [17]. Another noted HVS characteristic is the manifold sensing characteristic (MSC). In 2000, articles published in *Science* pointed out that (1) visual information is stored as a manifold of stable neural-activity patterns in the brain, and (2) manifold learning methods can identify meaningful low-dimensional structures in high-dimensional data [25–27]. Therefore, this study utilizes the manifold learning to construct a low-dimensional intrinsic manifold, which can not only reveal the capacity degradation law that is contained in the extracted features but also reduce the computation required. Therefore, this study attempts to introduce visual cognition into Li-ion battery capacity estimation in order to establish a systematic method for capacity estimation based on MCC and MSC.

This paper is organized as follows: Section 2 describes the two HVS properties of interest, namely, MCC and MSC, as well as the corresponding computing methods derived from them, primarily NSCT and the Laplacian eigenmap (LE). Geodesic distance is also introduced, which is used in the estimation of battery capacity. Section 3 presents the entire method for battery capacity estimation based on visual cognition, including descriptions of the experimental data, image transformation, feature extraction, and capacity calculation. Typical data from NASA battery data sets are utilized to verify the proposed method; the results are reported in Section 4. Finally, Section 5 concludes the paper.

2. Related Theories

2.1. MCC of the HVS and NSCT

2.1.1. MCC and Contourlet Transform. The HVS is a crucial tool by which human beings understand and comprehend the natural world. It has been verified that the HVS possesses the ability to capture the essential information of a natural scene using a minimal number of active visual cells [28].

The receptive fields in the visual cortex are accordingly characterized as being localized, oriented, and bandpass [29]. Therefore, it is suggested that, for an image representation to be efficient, it should have the properties of being local, directional, and of multiresolution.

The contourlet transform (CT) proposed by Do and Vetterli [28] matches the MCC of the HVS well. It is composed of a Laplacian pyramid (LP) and a directional filter bank (DFB), where the LP is employed to capture the point discontinuities and the DFB is utilized to link point discontinuities to linear structures. The CT gives a flexible multiresolution, local, and directional image expansion using contour segments; thus, it can represent edges and other singularities along curves very efficiently. Unfortunately, however, the CT lacks shift invariance because of the downsampling and upsampling in both the LP and the DFB. In particular, downsampling of a filtered image may result in low-pass and high-pass frequency aliasing. These shortcomings limit the use of CT in many applications [22, 30].

2.1.2. NSCT Theory. To eliminate the frequency aliasing of CT and enhance its directional selectivity and shift invariance, da Cunha et al. [31] proposed a shift-invariant version based on non-subsampled pyramid filter banks (NSPFBs) and non-subsampled directional filter banks (NSDFBs), as shown in Figure 1(a) [31, 32].

The non-subsampled contourlet transform (NSCT), as a representative method related to the MCC, can be used to decompose an image (e.g., as transformed from a charge current or discharge voltage curve) into multiple spatial-frequency channels (a set of narrow-band frequencies), each of which further involves different number of orientation components depending upon a predetermined setting for each channel.

In NSCT, the multiscale property is obtained from a shift-invariant filtering structure that achieves a subband decomposition similar to that of the LP. The process can be implemented using two-channel non-subsampled two-dimensional (2D) filter banks. Figure 1(b) illustrates the non-subsampled pyramid decomposition with $J = 3$ stages. Such an expansion is conceptually similar to the one-dimensional (1D) non-subsampled wavelet transform computed with the *à trous* algorithm. The filters for the next stage are obtained by upsampling the filters of the previous stage with the sampling matrix:

$$D = 2I = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad (1)$$

which gives the multiscale property without the need for additional filter design. On the j th decomposition, the ideal frequency support of the low-pass filter is $[-(\pi/2^j), (\pi/2^j)]^2$. Correspondingly, the ideal support of the high-pass filter is the complement of the low-pass filter, namely, the region

$[-(\pi/2^{j-1}), (\pi/2^{j-1})]^2 \setminus [-(\pi/2^j), (\pi/2^j)]^2$. The equivalent filters of a J -level cascading NSPFB are given by

$$H_n^{\text{eq}}(z) = \begin{cases} H_1(z^{2^{n-1}}) \prod_{j=0}^{n-2} H_0(z^{2^j}), & 1 \leq n \leq J \\ \prod_{j=0}^{n-2} H_0(z^{2^j}), & n = J + 1, \end{cases} \quad (2)$$

where $H_0(z)$ and $H_1(z)$ represent the low-pass filter and the corresponding high-pass filter, respectively, at the first stage [32].

The DFB is constructed by combining critically sampled two-channel fan filter banks and resampling operations. This results in a tree-structured filter bank that splits the 2D frequency plane into directional wedges. By switching off the downsamplers/upsamplers in each two-channel filter bank in the DFB tree structure and upsampling the filters accordingly, the NSDFB is obtained. In this manner, a tree composed of two-channel NSDFBs can be obtained. Figure 1(c) illustrates a four-channel decomposition [32]. The upsampled fan filters $U_j(z^Q)$ ($j = 0, 1$) have checkerboard frequency support, where Q is the quincunx matrix:

$$Q = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \quad (3)$$

The four-channel directional decomposition can be obtained when filters $U_j(z^Q)$ are combined with the fan filters $U_i(z)$ ($i = 0, 1$). The equivalent filter in each channel $U_k(z)$ ($k = 0, 1, 2, 3$) can be given as follows:

$$U_k(z) = U_i(z) U_j(z^Q). \quad (4)$$

After J -level NSCT decomposition, one low-pass subband image and $\sum_{j=1}^J 2^{l_j}$ bandpass directional subband images can be obtained, all of which have the same size as the input image. Here, l_j is the directional decomposition level at the j th scale.

As described above, the core of NSCT is the filter design in the two-channel NSPFB and NSDFB. NSCT not only retains the characteristics of CT but also has the important property of shift invariance. Thus, this study employs NSCT to extract features from the CC and DV values of a Li-ion battery.

2.2. MSC of the HVS and LE

2.2.1. MSC and Manifold Learning. As we look at an object with conditions such as scale and illumination changing, the signals carried from the eyes to the brain by the millions of axons in the optic nerve are constantly in flux. Nevertheless, we are able to recognize that these changing signals are produced by the same object. This phenomenon was studied by Seung and Lee, who proposed a hypothesis that a visual memory is stored as a manifold of stable states or a continuous attractor [25]. Images of the same object with changes in scale, illumination, and other variable factors lie on a low-dimensional manifold, whereas images of different objects form different manifolds. From the perspective of cognitive psychology, the cognitive process of object identification

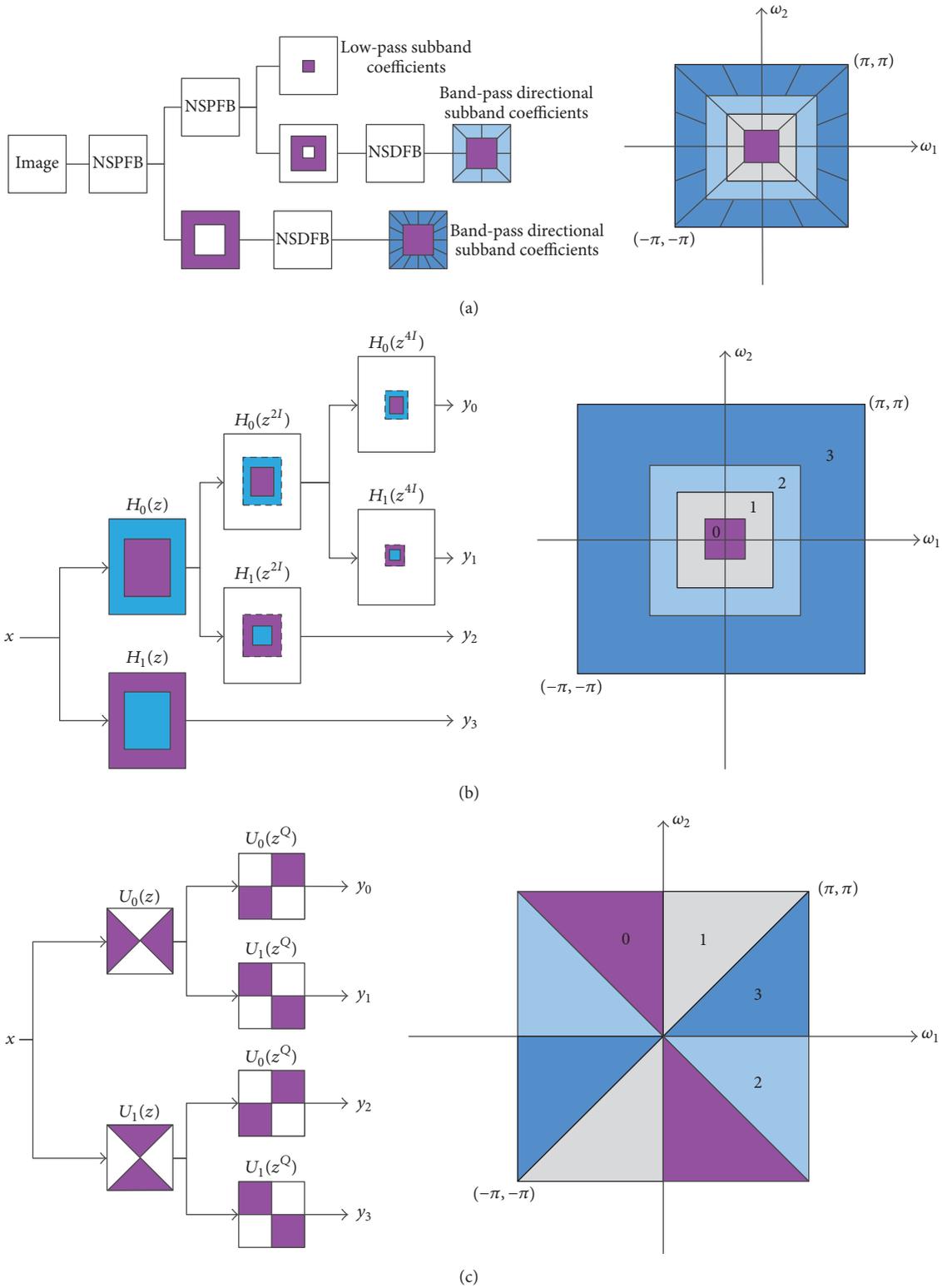


FIGURE 1: Schematic diagram of non-subsampled contourlet transform (a), three-stage non-subsampled pyramid decomposition (b), and four-channel NSDFB constructed with two-channel fan filter bank (c).

is the recognition of different low-dimensional manifolds embedded in the high-dimensional visual information. That is, the HVS has the ability to sense the manifold hidden in the brain. This characteristic of the HVS is called MSC. Similar to the MSC of the HVS, manifold learning can find meaningful low-dimensional structures hidden within high-dimensional observations; this is attracting increasing attention from scholars.

Manifold learning, also known as nonlinear dimensionality reduction, is a widely accepted method that embeds high-dimensional samples into low-dimensional feature space by preserving some local or global geometric structures [33]. Many manifold learning approaches have been proposed, such as isometric mapping [26], locally linear embedding [27], Laplacian eigenmaps [34], and Hessian eigenmaps [35]. Among these approaches, the Laplacian eigenmap (LE) is a kind of spectral graph method; this has received considerable attention from the machine learning community. In this study, the LE is employed to establish a low-dimensional intrinsic manifold and carry out the dimensionality reduction.

2.2.2. Theory of LE. LE is a typical graph-based dimensionality reduction technique. The basic mathematical notion of LE can be summarized as follows.

Assume that a d -dimensional manifold M^d (nominated as output space) embedded in an m -dimensional space $\alpha_N \in R^m$ (nominated as input space, $d < m$) can be described by a function:

$$f : C \subset M^d \longrightarrow R^m, \quad (5)$$

where C is a compact subset of M^d with open interior. A set of data points $\alpha_1, \dots, \alpha_N$, where $\alpha_i \in R^m$, are sampled with noise from the intrinsic manifold M^d ; the relationship can be represented as follows:

$$\alpha_i = f(\beta_i) + \xi_i, \quad i = 1, \dots, N, \quad (6)$$

where ξ_i denotes noise. LE can be recognized as follows: the original data set α_i 's in the higher dimensional manifold R^m are mapped (nonlinearly) to the data point β_i 's in the estimation of the unknown lower dimensional manifold M^d , with $d < m$ [36].

Given a set of N multivariate observations, for arbitrary point $A \in M^d$ with k nearest neighborhoods, a weighted adjacency graph $G = (V, E)$ can be constructed consisting of N nodes and a set of edges connecting neighboring points. We consider the problem of mapping the weighted graph G to a line such that the connected points stay as close together as possible. Let $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$ $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$, where $x_i, y_i \in R$ is a coordinate value of the i th point in R^m and M^d . A reasonable map is to choose y_i 's $\in R$ to minimize $\sum (y_i - y_j)^2 W_{ij}$ under the appropriate constraints. To avoid the heavy penalties that can occur if the neighboring points x_i and x_j are mapped far apart, the minimization is an attempt

to ensure that if points x_i and x_j are close, then y_i and y_j will be close as well. As a result, for any \mathbf{y} , we have

$$\frac{1}{2} \sum_{i,j} (y_i - y_j)^2 W_{ij} = \mathbf{y}^T L \mathbf{y}, \quad (7)$$

where $L = D - W$ is the Laplacian matrix, which is positive semidefinite. Notably, W_{ij} is symmetric, and $D_{ii} = \sum_j W_{ij}$. Thus, $\sum_{i,j} (y_i - y_j)^2 W_{ij}$ can be written as

$$\begin{aligned} & \sum_{i,j} (y_i^2 + y_j^2 - 2y_i y_j) W_{ij} \\ &= \sum y_i^2 D_{ii} + \sum y_j^2 D_{jj} - 2 \sum y_i y_j W_{ij} = 2\mathbf{y}^T L \mathbf{y}. \end{aligned} \quad (8)$$

Therefore, the minimization problem reduces to finding $\arg \min_{\mathbf{y}^T D \mathbf{y} = 1} \mathbf{y}^T L \mathbf{y}$.

The constraint $\mathbf{y}^T D \mathbf{y} = 1$ removes an arbitrary scaling factor in the embedding. Matrix D provides a natural measure on the graph vertex. The larger D_{ii} is, the more important the vertex will be. In (7), L is shown as a positive semidefinite matrix, and the vector \mathbf{y} that minimizes the objective function is given by the minimum eigenvalue solution to the generalized eigenvalue problem $L\mathbf{y} = \lambda D\mathbf{y}$ with an additional constraint of orthogonality $\arg \min_{\mathbf{y}^T D \mathbf{y} = 1, \mathbf{y}^T D \mathbf{1} = 0} \mathbf{y}^T L \mathbf{y}$.

More generally, the embedding is given by the $N \times d$ matrix $Y = [y_1, y_2, \dots, y_d]$, where the i th row, denoted by Y_i^T , provides the embedding coordinates of the i th vertex. Similarly, we need to minimize

$$\sum_{i,j} \|Y_i - Y_j\|^2 W_{ij} = \text{tr}(Y^T L Y). \quad (9)$$

This condition reduces to finding [37]

$$Y_{\text{opt}} = \arg \min_{Y^T D Y = 1} \text{tr}(Y^T L Y). \quad (10)$$

2.2.3. Time Window for Mapping Updating. The fixed set of data from high-dimensional space is mapped to low-dimensional space by LE through the mapping $g = f^{-1}$. Therefore, one can receive a corresponding low-dimensional point through the mapping when given an arbitrary point in the high-dimensional space. Considering in practice, new data are often collected and new features can be obtained in the R^m space, we need to update the mapping provided by LE to adjust to the new incoming data. Thus, a general method, the so-called "time window," is proposed, which can be set as one incoming point or any other number of incoming points with regard to a real-world application. When the number of new incoming points reaches the fixed "time window," a new updated mapping is derived.

2.3. Geodesic Distance. In mathematics, particularly differential geometry, a geodesic is a generalization of the notion of a "straight line" to curved spaces [38]. If this connection is the Levi-Civita connection induced by a Riemannian metric, then the geodesics are (locally) the shortest path between

TABLE I: Typical data under different operating conditions.

Label number	AT (°C)	CC (A)	DC (A)	EOD (V)	IC (Ah)	EOLC (%)
#5	24	1.5	2	2.7	1.8565	30
#7	24	1.5	2	2.2	1.8911	30
#29	43	1.5	4	2.0	1.8447	12.61
#54	4	1.5	2	2.2	1.1665	30

Note. AT, CC, DC, EOD, IC, and EOLC denote ambient temperature, charge current, discharge current, end-of-discharge, initial capacity, and end-of-life criteria (ratio of faded capacity to initial capacity), respectively.

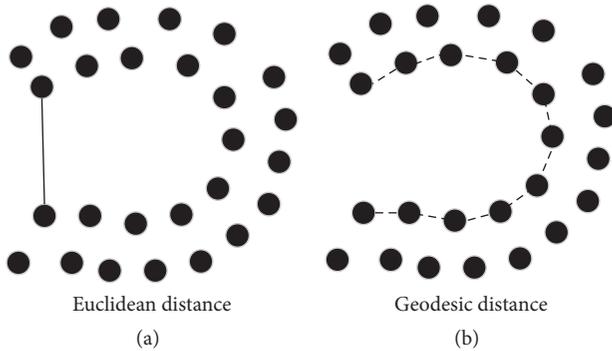


FIGURE 2: Comparison between Euclidean distance and geodesic distance.

points in the space. Thus, the geodesic distance is expected to unfold highly folded, twisted, or curved nonlinear manifolds [39].

Figure 2(a) shows the shortest path measured by Euclidean distance. According to this metric, two points on opposite sides of the horseshoe appear to be deceptively close. Figure 2(b) shows the shortest path measured by geodesic distance. In this case, the two points on opposite sides of the horseshoe are not neighbors according to the geodesic distance [39].

In this study, the geodesic distance is adopted as the geometrical metric of battery capacity on the manifold constructed by LE.

3. Method for Estimation of Li-Ion Battery Capacity Based on Visual Cognition

3.1. Description of the NASA Li-Ion Battery Experimental Data. The data used in this study were obtained from a custom-built battery setup at the NASA Ames Prognostics Center of Excellence. The experiments were conducted through three different operational profiles (charge, discharge, and impedance) at ambient temperature (AT) conditions. Charging is performed in a constant charge current mode at 1.5 A until the battery voltage reaches 4.2 V and continues in a constant voltage mode until the charge current drops to 20 mA. The discharge runs are stopped at different end-of-discharges (EODs). The experiments are conducted until the capacity decreases to specified end-of-life criteria (EOLC).

To validate the efficiency of the proposed approach, the typical data were selected (#5, #7, #29, and #54, which were also used in [16, 17]) and described in Table 1. From Table 1,

it can be seen that these data have the same charge current of 1.5 A but generally exhibit different ATs (24°C, 43°C, or 4°C), discharge currents (DCs; 2 A or 4 A), EODs (from 2.0 V to 2.7 V), initial capacities (ICs; from 1.1665 Ah to 1.8911 Ah), and EOLCs (30% or 12.61%).

3.2. Image Transformation of CC or DV Values for Visual Cognition

3.2.1. Image Transformation Method. The real state of an arbitrary battery can be identified by charging or discharging it. Consequently, the CC and DV curves obtained from the charging and discharging processes can directly reflect the real state of the battery. To discover the performance degradation law contained in these curves, the CC and DV curves for each cycle are transformed into an image for the following visual cognition. First, the CC and DV values over a lifetime of full cycles are uniformly normalized according to the linear normalization equation: $y = (x - \text{MinValue}) / (\text{MaxValue} - \text{MinValue})$, where x is the original CC or DV value, y is the normalized value, and MinValue and MaxValue are the minimum and maximum CC/DV values over the lifetime of full cycles, respectively. The normalized data points are then arranged into an $M \times N$ matrix, as shown in Figure 3. If we consider the normalized amplitude of each sample as a pixel value of an image, then the $M \times N$ matrix becomes an $M \times N$ image. The following principles are used to ensure the quality of the transformed images: (1) the transformed images should retain the most useful information of each charge/discharge cycle; (2) the CC and DV data that differ significantly from those of other cycles should be eliminated; (3) the images built based on the CC and DV data of each cycle should have the same size. To adhere to these principles, the CC and DV data of each charge/discharge cycle need to be selected and processed.

3.2.2. Data Selection and Processing. To some extent, the quality of the image transformation directly affects the visual cognition results. Therefore, appropriate selection and processing of the CC and DV data are essential to ensure high-quality transformed images.

Our experiments collect two kinds of CC/DV data containing the most useful information: (1) CC data under the constant voltage charging stage; (2) DV data during the discharging process. The following data are discarded: (A) abnormal data, (B) CC data under the constant current charging stage, (C) sensitive voltage data from the early stages of discharge, and (D) voltage recovery data; these are shown in Figure 4.

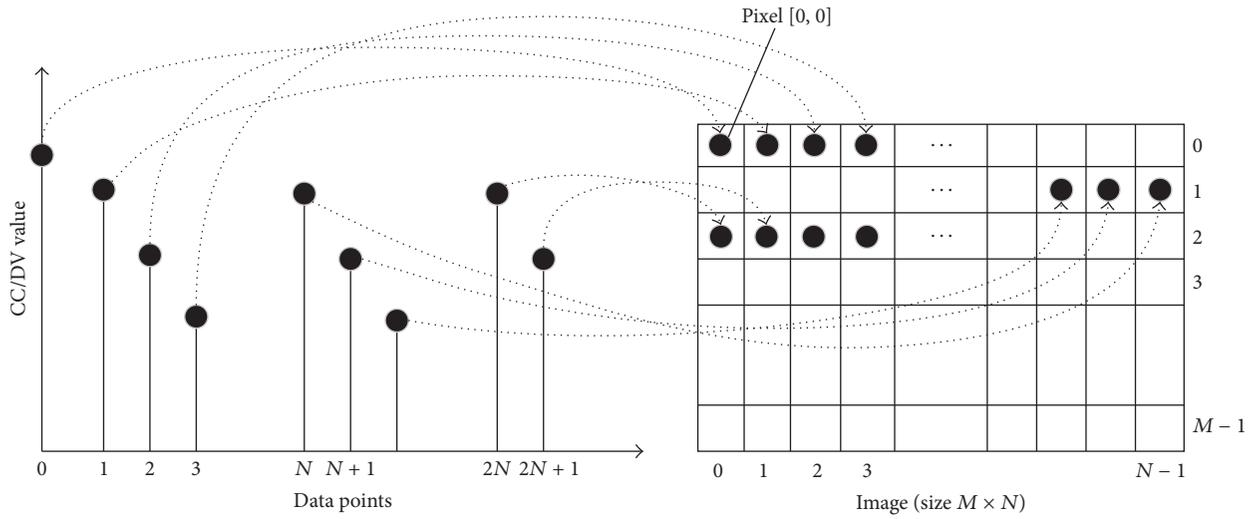


FIGURE 3: Image transformation scheme based on CC/DV values.

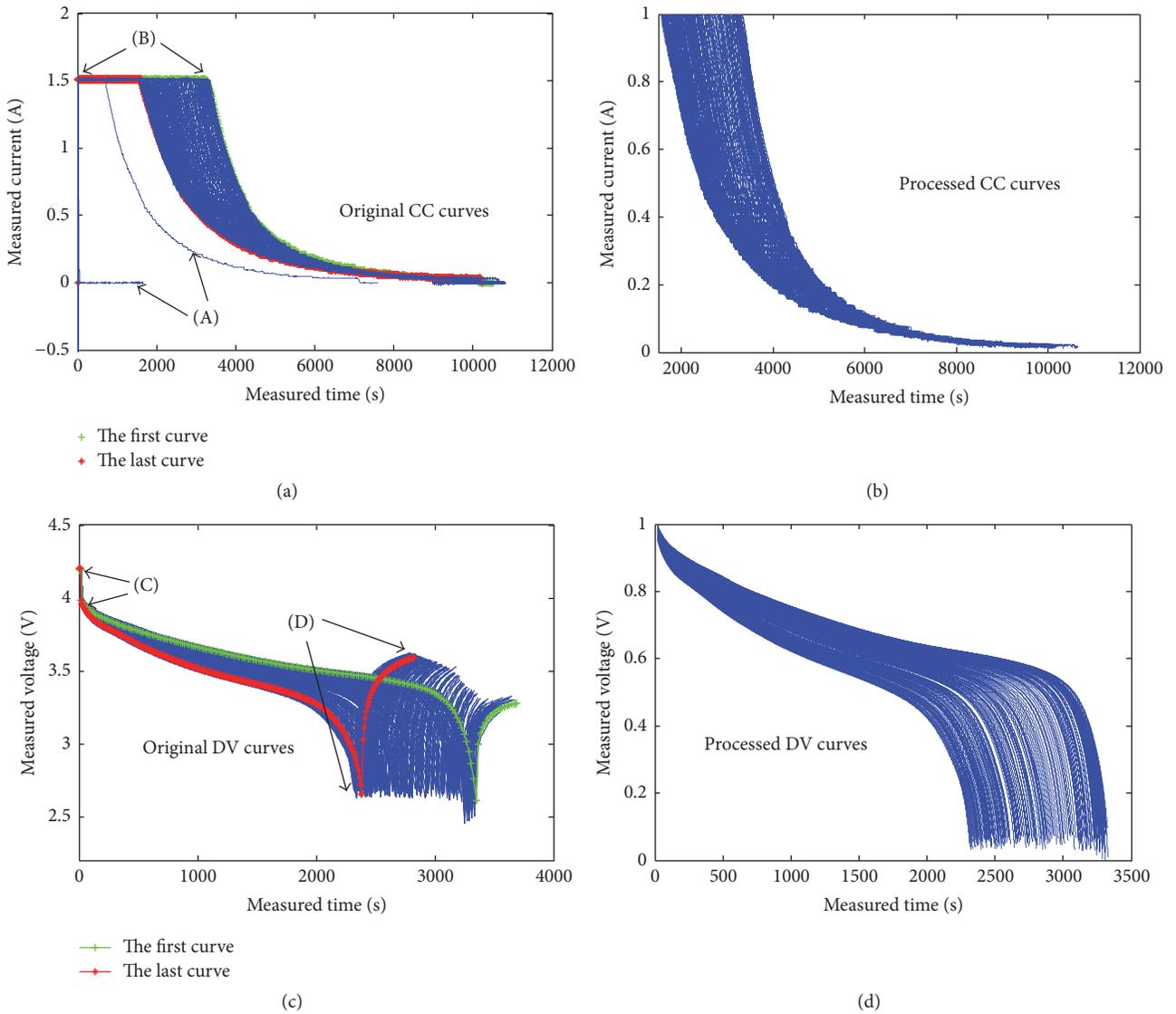


FIGURE 4: CC and DV curves: original versus processed.

The difference in sampling rate (or sampling start time) results in a different number of CC or DV data from each cycle. This causes a problem on forming images of the same size. We adopt an interpolation method to ensure the same number of data points for each cycle. If an image is too large, the computational load becomes excessive, whereas if an image is too small, it cannot reflect the characteristics of the specific charge/discharge cycle. To balance these considerations, we select $M = N = 64$. Thus, each cycle needs a total of 4096 data points to construct the image. If we obtain i ($i < 4096$) data points from a charge/discharge cycle, then the other $4096 - i$ data points are acquired by the ‘‘spline’’ interpolation algorithm. Figure 4 shows an example of the processed CC/DV data curves from battery #5.

3.3. Feature Extraction Based on NSCT and LE

3.3.1. NSCT-Based Multichannel Feature Extraction. This section describes a degradation feature extraction method based on NSCT, which is the core of this study. By utilizing NSCT, the transformed images from the CC/DV values are decomposed into multiple spatial-frequency channels with a set of orientation subbands. The subbands can be expressed as follows:

$$\{C_{i_0}, C_{i,j}\}, \quad (11)$$

$$i, i_0 = 1, \dots, n; \quad i \leq i_0; \quad j = 2, 4, 8, \dots, m; \quad n \in N, \quad m \in 2^N,$$

where i is the decomposition scale, j is the decomposition direction, C_{i_0} represents the low-frequency coefficient, and $C_{i,j}$ represents the high-frequency coefficient of the j th directional subband at the i th scale. In this study, $i_0 = 2$ and $j = \{2, 4\}$. That is, the decomposition scale is 2 and the decomposition directions at each scale are 2 and 4.

The coefficients of the low-frequency subbands reflect the outline information of the image, whereas the coefficients of the high-frequency subbands reflect the detailed information. Therefore, three time-domain indicators are extracted as feature values; these are the mean value (μ) and the variance value (σ) of the low-frequency subband coefficients as well as the energy value (E) of the high-frequency subband coefficients. The equations for calculating these three indicators are given as follows:

$$\mu = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N P(x, y),$$

$$\sigma = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N [P(x, y) - \mu]^2, \quad (12)$$

$$E = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N [P(x, y)]^2,$$

where $P(x, y)$ represents each element of the coefficients and $M \times N$ represents the size of the coefficient matrix. Thus, an eight-dimensional feature vector of each image transformed from a single charge/discharge cycle can be obtained as

$$f = [\mu, \sigma, E_{1,1}, E_{1,2}, E_{2,1}, E_{2,2}, E_{2,3}, E_{2,4}]. \quad (13)$$

3.3.2. LE-Based Intrinsic Manifold Establishment. An intrinsic manifold is established using the aforementioned LE method. The degradation law governing the battery performance is revealed by data lying on this intrinsic manifold in the M^d space, which is embedded in the high-dimensional R^m space. The R^m space is constructed by the eight-dimensional feature vectors extracted by NSCT from the images transformed from the CC/DV data. The mapping $g = f^{-1}$ from R^m to M^d gives a 2D feature matrix in the M^d space, where Li-ion battery capacity degradation can be well described. The mapping $g = f^{-1}$ is established by an analogous set of raw experimental data of full-cycle of lifetime (ASL) for each of the four typical datasets. Given an arbitrary point in R^m , the corresponding data point representing the Li-ion battery capacity in M^d can be obtained through the mapping $g = f^{-1}$.

3.4. Capacity Estimation Based on Geodesic Distance. In this study, the geodesic distance along the intrinsic manifold between the initial point and the most recent point in the degradation process is calculated to carry out the estimation of battery capacity. Denote C_{A_0} as the initial capacity, which is typically not the rated capacity, and C_{EOL} as the capacity of the final charge/discharge cycle of the ASL experimental data. We denote the geodesic distance between the initial point and the points on the intrinsic manifold M^d as geo_s , and the geodesic distance between the initial point and the last point on the intrinsic manifold of the ASL as geo_{EOL} . The capacity of each point in R^m space can then be estimated as

$$\widehat{C}_A = C_{A_0} - \frac{geo_s}{geo_{EOL}} (C_{A_0} - C_{EOL}). \quad (14)$$

4. Results and Discussion

We use battery #5 to demonstrate the effectiveness of the proposed approach. Figure 4 shows the original CC data curves during the charging process (Figure 4(a)) and the DV data curves during the discharging process (Figure 4(c)).

The corresponding processed curves derived from the stable stage of battery #5 are shown in Figures 4(b) and 4(d). Using the transformation scheme depicted in Figure 3, the normalized data for each cycle of the charging/discharging processes are transformed into an image. Figure 5 shows examples of the transformed images from a single cycle of the charging and discharging processes.

After the image transformation, the NSCT method is employed to extract features from the transformed images, thereby forming an eight-dimensional feature vector constructed by calculating the mean and variance of the low-frequency subband coefficients and the energy of the high-frequency subband coefficients. Through the mapping $g = f^{-1}$ from R^m to M^d established by LE, we construct the intrinsic manifold in the 2D space M^d . This describes the degradation law of battery capacity. Figure 6 shows the intrinsic manifold of battery #5 embedded in the eight-dimensional R^m space constructed by features extracted from the DV data.

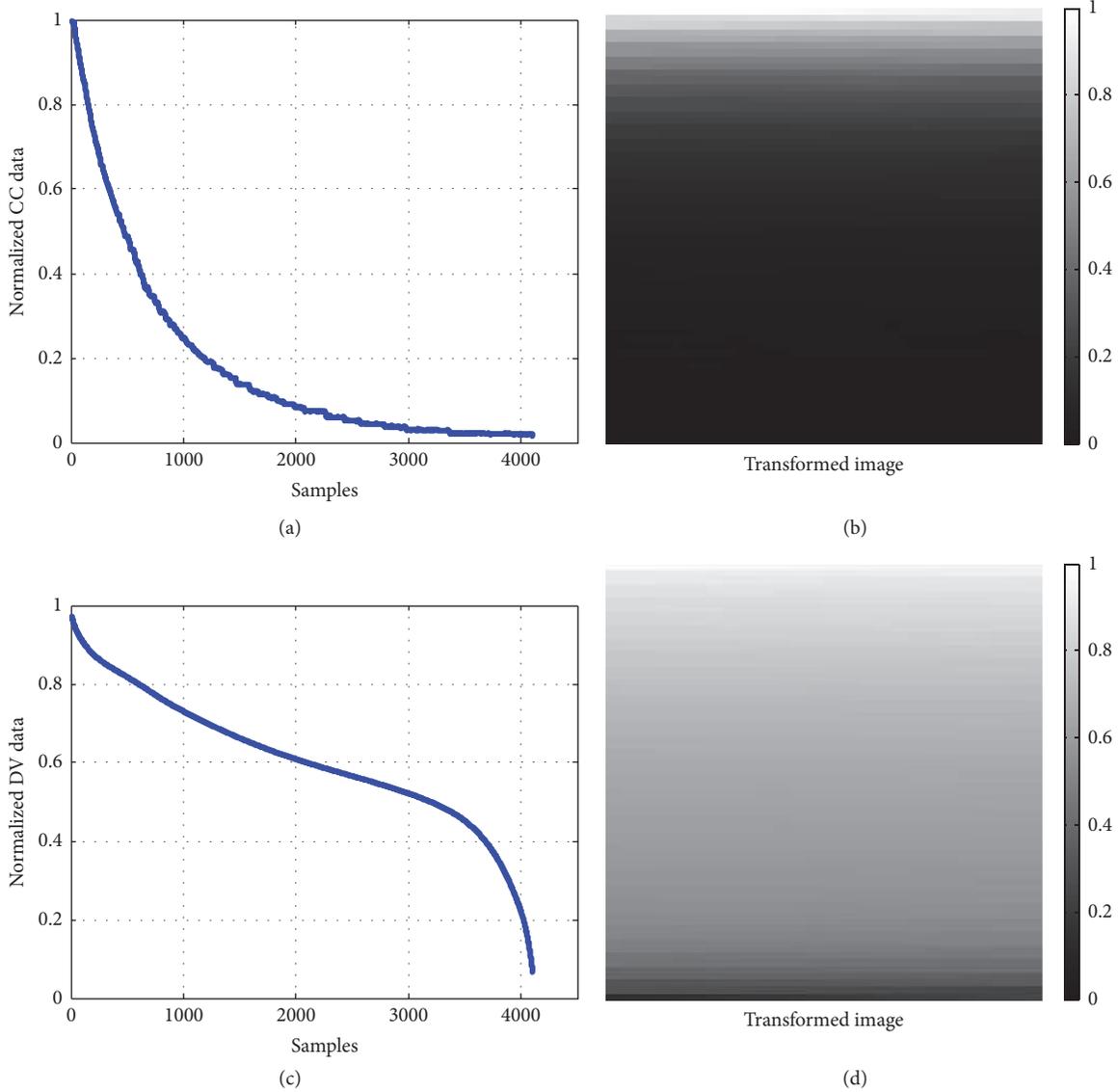


FIGURE 5: Image transformation based on CC/DV data from a single charging/discharging cycle.

In this study, the capacity is estimated using either CC or DV data based on the geodesic distance on the intrinsic manifold, as calculated by (14). The estimated results for all the demonstration data (batteries #5, #7, #29, and #54) under the different operating conditions closely track the measured capacity, as illustrated in Figure 7.

A comparison between the estimation results in this study and those of the study in [16, 17] is given in Table 2 in terms of the absolute error (AE), relative error (RE), and elapsed times (ETs). AE and RE are calculated as follows:

$$\begin{aligned}
 \text{AE} &= \text{mean} [\text{abs} (\text{Estimated capacity} \\
 &\quad - \text{Estimated capacity})], \\
 \text{RE} &= \frac{\text{abs} (\text{Estimated capacity} - \text{Estimated capacity})}{\text{Estimated capacity}}.
 \end{aligned} \tag{15}$$

Compared with [16], Table 2 shows that the proposed method based on visual cognition has approximate estimation accuracy with the similarity recognition method based on database. Using CC data for capacity estimation, the AEs and REs of batteries #7 and #29 based on the proposed method are smaller than those reported in [16], while the AEs and REs of batteries #5 and #54 are larger than those in [16]. Using DV data for capacity estimation, the AEs and REs of batteries #29 and #54 based on the proposed method are smaller than those in [16], while the AEs and REs of batteries #5 and #7 are larger than those in [16]. Even though the average AEs and REs of the proposed method are slightly larger than those in [16], however, the average ETs of the proposed method are only 11.1975 s based on CC data and 11.095 s based on DV data, more than 20 times smaller than those in [16], which makes the proposed visual cognition method more practical for real-time capacity estimation.

TABLE 2: Estimation accuracy of available capacity based on the proposed method.

Data	Items	#5	#7	#29	#54	Maximum	Average
CC	AEs (%)	2.30	2.51	0.90	3.18	3.18	2.222
	REs (%)	1.51	1.56	0.53	3.38	3.38	1.745
	ETs (s)	16.16	15.86	3.59	9.18	16.16	11.1975
CC [16]	AEs (%)	2.19	3.70	5.04	2.66	3.70	3.3975
	REs (%)	1.42	2.23	2.90	2.71	2.90	2.315
	ETs (s)	401	291	11	189	401	223
DV	AEs (%)	2.85	2.77	1.30	2.28	2.85	2.300
	REs (%)	1.76	1.69	0.75	2.43	2.43	1.658
	ETs (s)	15.70	15.41	3.82	9.45	15.70	11.095
DV [16]	AEs (%)	1.21	1.94	1.49	2.37	2.37	1.7525
	REs (%)	0.77	1.15	0.87	2.48	2.48	1.3175
	ETs (s)	460	387	10	53	460	227.5
CC + DV [17]	AEs (%)	4.48	2.42	1.85	3.71	4.48	3.115
	REs (%)	2.93	1.49	1.06	3.84	3.84	2.330

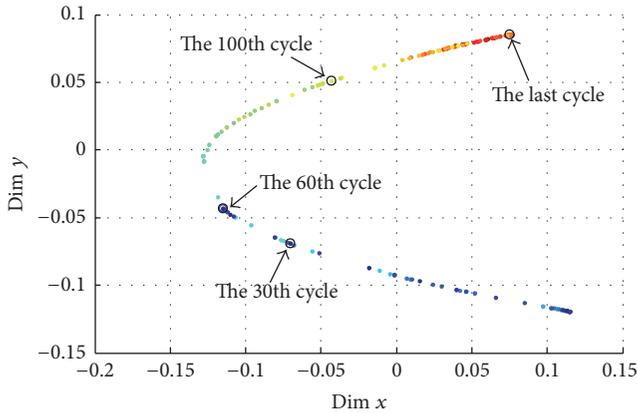


FIGURE 6: The intrinsic manifold of battery #5 that reveals the battery performance degradation.

Compared with [17], it can be seen from Table 2 that the proposed capacity estimation method based on visual cognition generally exhibits better performance than the geometric-based method [17]. The AEs and REs for batteries #5, #29, and #54 are smaller than those reported in [17], with the AE and RE of battery #7 being slightly higher. Notably, the excess part of the AE and RE of battery #7 is relatively small, because the estimation accuracy of battery #7 in [17] is already very high. Using visual cognition based on CC data, the maximum and average of the AEs are reduced by 1.3% and 0.8925%, respectively; those of the REs are reduced by 0.46% and 0.585%, respectively. Similarly, using visual cognition based on DV data, the maximum and average of the AEs are reduced by 1.63% and 0.815%, and those of the REs are reduced by 1.41% and 0.6725%.

The estimation results presented in Figure 7 and Table 2 demonstrate that the proposed visual cognition-based capacity estimation method is highly effective with either CC or

DV data in a very short time. That is, one can choose either CC or DV curves with which to estimate the battery capacity in real time with high accuracy.

5. Conclusions

This study proposes a novel method for estimating the capacity of Li-ion batteries based on visual cognition. The proposed approach transforms the collected CC or DV data from each charge/discharge cycle into an image. NSCT is then employed to extract features from the transformed image. After that, taking inspiration from the HVS manifold sensing characteristic, we utilize the LE method to establish the intrinsic manifold embedded in the high-dimensional NSCT coefficients, from which the degradation law of battery performance can be revealed. The geodesic distance on the intrinsic manifold is adopted to estimate the battery capacity.

The proposed visual cognition-based capacity estimation method can use either CC or DV data. Verification experiments were conducted using data collected from the NASA battery data sets. The results demonstrate that the proposed method can be used to perform the capacity estimation using either CC or DV data with high accuracy under different operating and aging conditions. In addition, the proposed method avoids the need to study complex electrochemical mechanisms, establish models, or conduct lengthy testing, which makes it a promising practical method for battery capacity estimation. However, further research should be conducted to ascertain the following:

- (1) the optimal number of CC/DV data for image transformation;
- (2) the choice of decomposition scale and decomposition direction in the NSCT method;
- (3) the intrinsic dimensionality of the manifold constructed by LE.

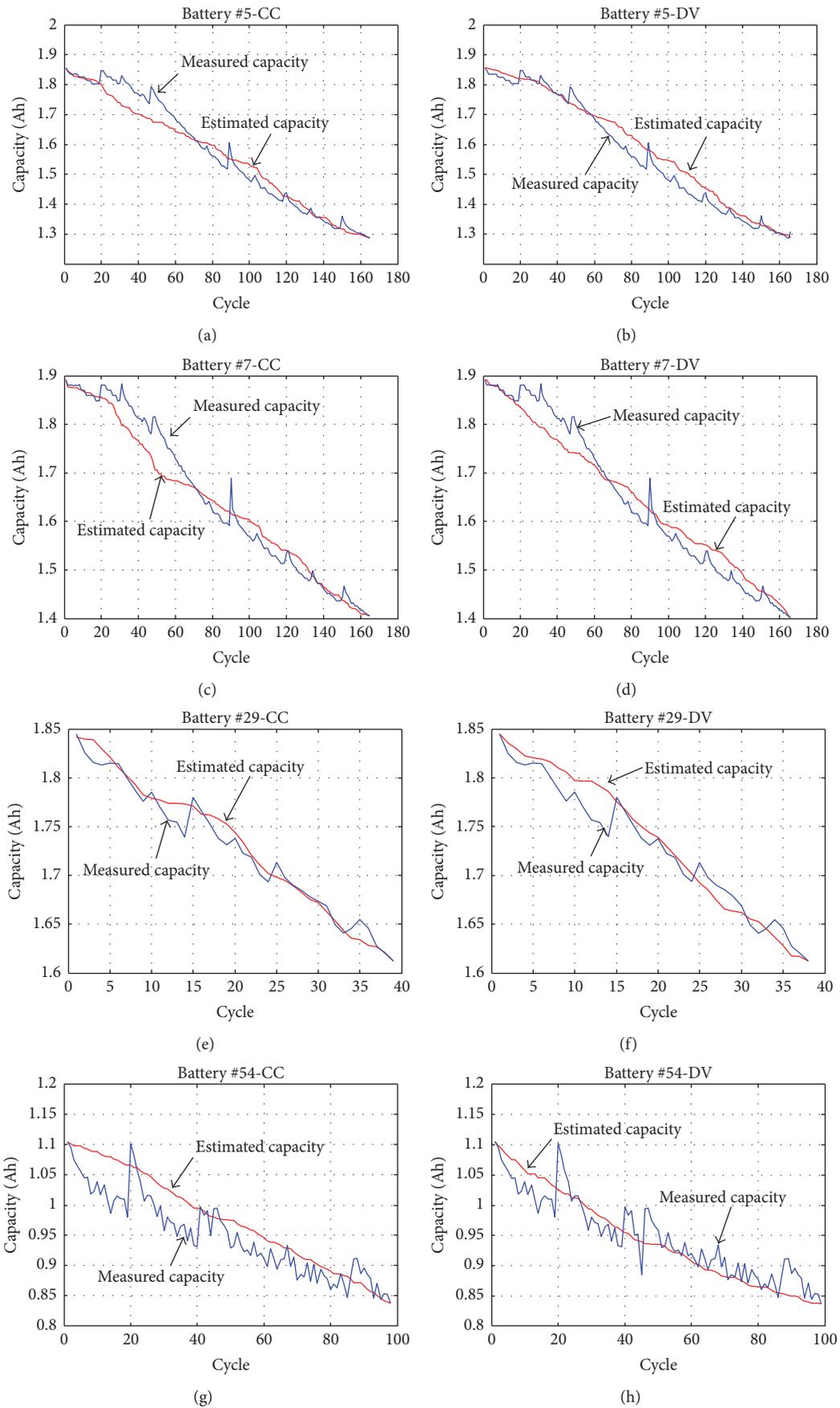


FIGURE 7: Capacity estimation results of typical data under various conditions based on CC/DV values.

Glossary

AE:	Absolute error
ASL:	An analogous set of full-cycle of lifetime
AT:	Ambient temperature
CC:	Charging current
CT:	Contourlet transform
CV:	Constant voltage
DC:	Discharge current
DFB:	Directional filter bank
DV:	Discharging voltage
EOD:	End of discharge
EOLC:	End-of-life criterion
HVS:	Human visual system
IC:	Initial capacity
LE:	Laplacian eigenmap
LP:	Laplacian pyramid
MCC:	Multichannel characteristic
MSC:	Manifold sensing characteristic
NSCT:	Non-subsampled contourlet transform
NSDFB:	Non-subsampled directional filter bank
NSFB:	Non-subsampled filter bank
NSPFB:	Non-subsampled pyramid filter bank
RE:	Relative error.

Conflicts of Interest

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

Acknowledgments

This study was supported by the Fundamental Research Funds for the Central Universities (Grant no. YWF-16-BJ-J-18) and the National Natural Science Foundation of China (Grant nos. 51575021 and 61603016), as well as the China Postdoctoral Science Foundation (Grant nos. 2017M610033 and 2017T100026).

References

- [1] L. Chen, W. Lin, J. Li, B. Tian, and H. Pan, "Prediction of lithium-ion battery capacity with metabolic grey model," *Energy*, vol. 106, pp. 662–672, 2016.
- [2] G. Dong, X. Zhang, C. Zhang, and Z. Chen, "A method for state of energy estimation of lithium-ion batteries based on neural network model," *Energy*, vol. 90, pp. 879–888, 2015.
- [3] B. Xia, C. Chen, Y. Tian, M. Wang, W. Sun, and Z. Xu, "State of charge estimation of lithium-ion batteries based on an improved parameter identification method," *Energy*, vol. 90, pp. 1426–1434, 2015.
- [4] Z. Deng, L. Yang, Y. Cai, H. Deng, and L. Sun, "Online available capacity prediction and state of charge estimation based on advanced data-driven algorithms for lithium iron phosphate battery," *Energy*, vol. 112, pp. 469–480, 2016.
- [5] L. Zheng, L. Zhang, J. Zhu, G. Wang, and J. Jiang, "Co-estimation of state-of-charge, capacity and resistance for lithium-ion batteries based on a high-fidelity electrochemical model," *Applied Energy*, vol. 180, pp. 424–434, 2016.
- [6] B. Y. Liaw, R. G. Jungst, G. Nagasubramanian, H. L. Case, and D. H. Doughty, "Modeling capacity fade in lithium-ion cells," *Journal of Power Sources*, vol. 140, no. 1, pp. 157–161, 2005.
- [7] C. Hu, B. D. Youn, and J. Chung, "A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation," *Applied Energy*, vol. 92, pp. 694–704, 2012.
- [8] R. Xiong, F. Sun, Z. Chen, and H. He, "A data-driven multi-scale extended Kalman filtering based parameter and state estimation approach of lithium-ion polymer battery in electric vehicles," *Applied Energy*, vol. 113, pp. 463–476, 2014.
- [9] Z.-W. He, M.-Y. Gao, G.-J. Ma, Y.-Y. Liu, and S.-X. Chen, "Online state-of-health estimation of lithium-ion batteries using Dynamic Bayesian Networks," *Journal of Power Sources*, vol. 267, pp. 576–583, 2014.
- [10] A. Singh, A. Izadian, and S. Anwar, "Model based condition monitoring in lithium-ion batteries," *Journal of Power Sources*, vol. 268, pp. 459–468, 2014.
- [11] J. Yi, J. Lee, C. B. Shin, T. Han, and S. Park, "Modeling of the transient behaviors of a lithium-ion battery during dynamic cycling," *Journal of Power Sources*, vol. 277, pp. 379–386, 2015.
- [12] J. Li, L. Wang, C. Lyu, L. Zhang, and H. Wang, "Discharge capacity estimation for Li-ion batteries based on particle filter under multi-operating conditions," *Energy*, vol. 86, pp. 638–648, 2015.
- [13] M. A. Roscher, J. Assfalg, and O. S. Bohlen, "Detection of utilizable capacity deterioration in battery systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 98–103, 2011.
- [14] M. Einhorn, F. V. Conte, C. Kral, and J. Fleig, "A method for online capacity estimation of lithium ion battery cells using the state of charge and the transferred charge," *IEEE Transactions on Industry Applications*, vol. 48, no. 2, pp. 736–741, 2012.
- [15] J. Zhang and J. Lee, "A review on prognostics and health monitoring of Li-ion battery," *Journal of Power Sources*, vol. 196, no. 15, pp. 6007–6014, 2011.
- [16] L. Tao, C. Lu, and A. Noktehdan, "Similarity recognition of online data curves based on dynamic spatial time warping for the estimation of lithium-ion battery capacity," *Journal of Power Sources*, vol. 293, pp. 751–759, 2015.
- [17] C. Lu, L. Tao, and H. Fan, "Li-ion battery capacity estimation: A geometrical approach," *Journal of Power Sources*, vol. 261, pp. 141–147, 2014.
- [18] B. Balagopal and M.-Y. Chow, "The state of the art approaches to estimate the state of health (SOH) and state of function (SOF) of lithium ion batteries," in *Proceedings of the 13th International Conference on Industrial Informatics, INDIN 2015*, pp. 1302–1307, UK, July 2015.
- [19] https://en.wikipedia.org/wiki/Cognitive_science.
- [20] P. Cavanagh, "Visual cognition," *Vision Research*, vol. 51, no. 13, pp. 1538–1551, 2011.
- [21] Y. Cheng, Y. Hou, C. Zhao, Z. Li, Y. Hu, and C. Wang, "Robust face recognition based on illumination invariant in nonsubsampling contourlet transform domain," *Neurocomputing*, vol. 73, no. 10–12, pp. 2217–2224, 2010.
- [22] Y. Chai, H. Li, and X. Zhang, "Multifocus image fusion based on features contrast of multiscale products in nonsubsampling contourlet transform domain," *Optik - International Journal for Light and Electron Optics*, vol. 123, no. 7, pp. 569–581, 2012.
- [23] K. Hammouche, O. Losson, and L. Macaire, "Fuzzy aura matrices for texture classification," *Pattern Recognition*, vol. 53, pp. 212–228, 2016.

- [24] T. Tan, "Texture feature extraction via visual cortical channel modelling," in *Proceedings of the Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol. IV. Conference D: Architectures for Vision and Pattern Recognition.*, pp. 607–610, The Hague, Netherlands.
- [25] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.
- [26] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [27] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [28] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.
- [29] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, pp. 106–154, 1962.
- [30] S. Lili, Y. Jiachen, and Z. Zhuoyun, "Stereo picture quality estimation based on a multiple channel HVS model," in *Proceedings of the 2009 2nd International Congress on Image and Signal Processing, CISP'09*, China, October 2009.
- [31] A. L. da Cunha, J. Zhou, and M. N. Do, "The nonsubsampling contourlet transform: theory, design, and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089–3101, 2006.
- [32] Q. Zhang and B. Guo, "Multifocus image fusion using the nonsubsampling contourlet transform," *Signal Processing*, vol. 89, no. 7, pp. 1334–1346, 2009.
- [33] T. Deng and W. Xie, "Granule-view based feature extraction and classification approach to color image segmentation in a manifold space," *Neurocomputing*, vol. 99, pp. 46–58, 2013.
- [34] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 14, pp. 585–591, 2002.
- [35] D. L. Donoho and C. Grimes, "Hessian eigenmaps: locally linear embedding techniques for high-dimensional data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 10, pp. 5591–5596, 2003.
- [36] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM Journal on Scientific Computing*, vol. 26, no. 1, pp. 313–338, 2004.
- [37] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [38] <http://en.wikipedia.org/wiki/Geodesic>.
- [39] G. Wen, L. Jiang, and J. Wen, "Using locally estimated geodesic distance to optimize neighborhood graph for isometric data embedding," *Pattern Recognition*, vol. 41, no. 7, pp. 2226–2236, 2008.