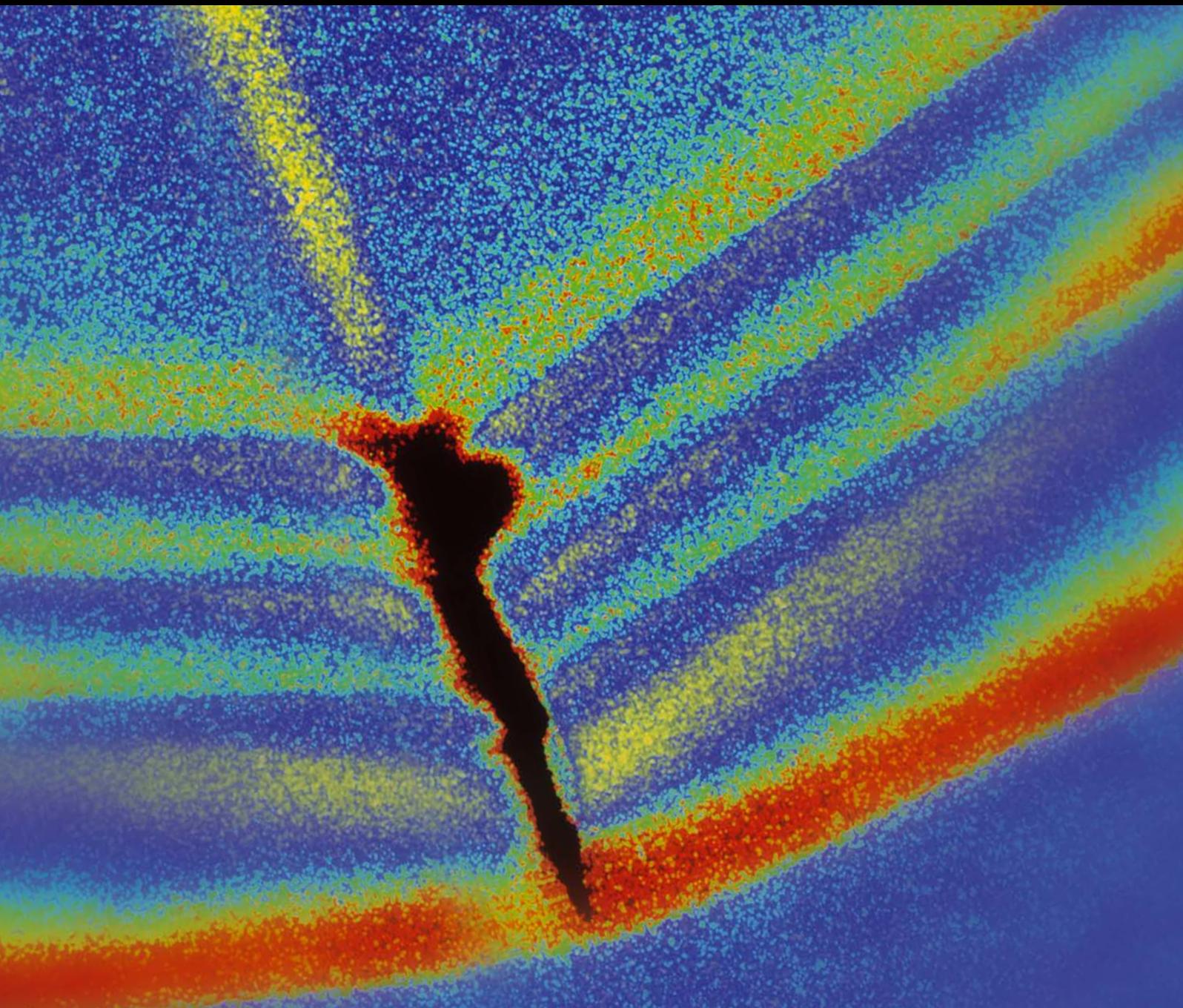


# Fault Diagnosis and Prognosis of Critical Components

Guest Editors: Dong Wang, Chuan Li, Achmad Widodo, Pavan Kumar Kankar, and Wahyu Caesarendra



---



# **Fault Diagnosis and Prognosis of Critical Components**

Shock and Vibration

---

## **Fault Diagnosis and Prognosis of Critical Components**

Guest Editors: Dong Wang, Chuan Li, Achmad Widodo,  
Pavan Kumar Kankar, and Wahyu Caesarendra



---

Copyright © 2016 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "Shock and Vibration." 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.

## Editor-in-Chief

Mehdi Ahmadian, Virginia Polytechnic Institute and State University, USA

## Editorial Board

Brij N. Agrawal, USA  
Marco Alfano, Italy  
Farbod Alijani, Canada  
Sumeet S. Aphale, UK  
Mariano Artés, Spain  
Hassan Askari, Canada  
Matteo Aureli, USA  
Ranjan Banerjee, UK  
Mahmoud Bayat, Iran  
José A. Becerra Villanueva, Spain  
Marco Belloli, Italy  
Subhamoy Bhattacharya, UK  
Ivo Caliò, Italy  
Antonio Carcaterra, Italy  
Dumitru I. Caruntu, USA  
Noel Challamel, France  
Athanasios Chasalevris, UK  
Peng Chen, Japan  
Ashwin Chinnayya, France  
Giorgio Dalpiaz, Italy  
Farhang Daneshmand, Canada  
Sergio De Rosa, Italy  
Dario Di Maio, UK  
Longjun Dong, China  
Lorenzo Dozio, Italy  
Mohamed El badaoui, France  
Mohammad Elahinia, USA  
Fiorenzo A. Fazzolari, UK

Francesco Franco, Italy  
Alessandro Gasparetto, Italy  
Gianluca Gatti, Italy  
Anindya Ghoshal, USA  
Nere Gil-Negrete, Spain  
Hassan Haddadpour, Iran  
M.I. Herreros, Spain  
Hamid Hosseini, Japan  
Reza Jazar, Australia  
Sakdirat Kaewunruen, UK  
Yuri S. Karinski, Israel  
Jeong-Hoi Koo, USA  
Georges Kouroussis, Belgium  
Mickaël Lallart, France  
Kenneth J. Loh, USA  
Nuno M. Maia, Portugal  
Giuseppe C. Marano, Italy  
Laurent Mevel, France  
Emiliano Mucchi, Italy  
Tony Murmu, UK  
Sundararajan Natarajan, India  
Toshiaki Natsuki, Japan  
Miguel Neves, Portugal  
Coral Ortiz, Spain  
Aleksandar Pavic, UK  
Evgeny Petrov, UK  
Antonina Pirrotta, Italy  
Vitomir Racic, Italy

Carlo Rainieri, Italy  
Didier Rémond, France  
Francesco Ripamonti, Italy  
Salvatore Russo, Italy  
Edoardo Sabbioni, Italy  
Jerzy T. Sawicki, USA  
Onome E. Scott-Emuakpor, USA  
Vadim V. Silberschmidt, UK  
Kumar V. Singh, USA  
Isabelle Sochet, France  
Alba Sofi, Italy  
Jussi Sopanen, Finland  
Stefano Sorace, Italy  
Narakorn Srinil, UK  
Salvatore Strano, Italy  
Chao Tao, China  
Mario Terzo, Italy  
Tai Thai, Australia  
Carlo Trigona, Italy  
Federica Tubino, Italy  
Nerio Tullini, Italy  
Marcello Vanali, Italy  
Jörg Wallaschek, Germany  
Zaili L. Yang, UK  
Stana Živanović, UK  
Lei Zuo, USA

# Contents

---

## **Fault Diagnosis and Prognosis of Critical Components**

Dong Wang, Chuan Li, Achmad Widodo, Pavan Kumar Kankar, and Wahyu Caesarendra  
Volume 2016, Article ID 9597656, 3 pages

## **A Hybrid Prognostic Approach for Remaining Useful Life Prediction of Lithium-Ion Batteries**

Wen-An Yang, Maohua Xiao, Wei Zhou, Yu Guo, and Wenhe Liao  
Volume 2016, Article ID 3838765, 15 pages

## **Trace Ratio Criterion-Based Kernel Discriminant Analysis for Fault Diagnosis of Rolling Element Bearings Using Binary Immune Genetic Algorithm**

Wen-An Yang, Maohua Xiao, Wei Zhou, Yu Guo, Wenhe Liao, and Gang Shen  
Volume 2016, Article ID 8631639, 15 pages

## **Gearbox Fault Diagnosis Using Complementary Ensemble Empirical Mode Decomposition and Permutation Entropy**

Liye Zhao, Wei Yu, and Ruqiang Yan  
Volume 2016, Article ID 3891429, 8 pages

## **Sparse Signal Representations of Bearing Fault Signals for Exhibiting Bearing Fault Features**

Wei Peng, Dong Wang, Changqing Shen, and Dongni Liu  
Volume 2016, Article ID 1835127, 12 pages

## **Fault Diagnosis for a Multistage Planetary Gear Set Using Model-Based Simulation and Experimental Investigation**

Guoyan Li, Fangyi Li, Yifan Wang, and Dehao Dong  
Volume 2016, Article ID 9263298, 19 pages

## **Application of Artificial Neural Network for Damage Detection in Planetary Gearbox of Wind Turbine**

Marcin Strączkiewicz and Tomasz Barszcz  
Volume 2016, Article ID 4086324, 12 pages

## **Application of Reassigned Wavelet Scalogram in Wind Turbine Planetary Gearbox Fault Diagnosis under Nonstationary Conditions**

Xiaowang Chen and Zhipeng Feng  
Volume 2016, Article ID 6723267, 12 pages

## **Research on the Sparse Representation for Gearbox Compound Fault Features Using Wavelet Bases**

Chunyan Luo, Changqing Shen, Wei Fan, Gaigai Cai, Weiguo Huang, and Zhongkui Zhu  
Volume 2015, Article ID 560171, 11 pages

## **Cyclostationary Analysis for Gearbox and Bearing Fault Diagnosis**

Zhipeng Feng and Fulei Chu  
Volume 2015, Article ID 542472, 12 pages

## **Feature Extraction of Faulty Rolling Element Bearing under Variable Rotational Speed and Gear Interferences Conditions**

Dezun Zhao, Jianyong Li, and Weidong Cheng  
Volume 2015, Article ID 425989, 9 pages

**Two General Architectures for Intelligent Machine Performance Degradation Assessment**

Yanwei Xu, Aijun Xu, and Tancheng Xie

Volume 2015, Article ID 676959, 5 pages

**In Situ Measurement of Seeking Speed and Seeking Induced Head-Disk Interface Instability in Hard Disk Drives**

Yu Wang, Xiongfei Wei, Yanyang Zi, and Kwok-Leung Tsui

Volume 2015, Article ID 989854, 7 pages

**A New Transmissibility Based Indicator of Local Variation in Structure and Its Application for Damage Detection**

X. Z. Li, Z. K. Peng, X. J. Dong, W. M. Zhang, and G. Meng

Volume 2015, Article ID 850286, 18 pages

**Gearbox Fault Identification and Classification with Convolutional Neural Networks**

ZhiQiang Chen, Chuan Li, and René-Vinicio Sanchez

Volume 2015, Article ID 390134, 10 pages

**Multifault Diagnosis of Rolling Element Bearings Using a Wavelet Kurtogram and Vector Median-Based Feature Analysis**

Phuong H. Nguyen and Jong-Myon Kim

Volume 2015, Article ID 320508, 14 pages

**Condition Monitoring and Fault Diagnosis for an Antifalling Safety Device**

Guangxiang Yang and Hua Liang

Volume 2015, Article ID 286781, 12 pages

**Rolling Element Bearing Fault Diagnosis Based on Multiscale General Fractal Features**

Weigang Wen, Zhaoyan Fan, Donald Karg, and Weidong Cheng

Volume 2015, Article ID 167902, 9 pages

**Screw Performance Degradation Assessment Based on Quantum Genetic Algorithm and Dynamic Fuzzy Neural Network**

Xiaochen Zhang, Hongli Gao, and Haifeng Huang

Volume 2015, Article ID 150797, 11 pages

**Planetary Gearbox Vibration Signal Characteristics Analysis and Fault Diagnosis**

Qiang Miao and Qinghua Zhou

Volume 2015, Article ID 126489, 8 pages

## Editorial

# Fault Diagnosis and Prognosis of Critical Components

**Dong Wang,<sup>1</sup> Chuan Li,<sup>2</sup> Achmad Widodo,<sup>3</sup>  
Pavan Kumar Kankar,<sup>4</sup> and Wahyu Caesarendra<sup>5</sup>**

<sup>1</sup>*City University of Hong Kong, Kowloon Tong, Hong Kong*

<sup>2</sup>*Universidad Politécnica Salesiana, 010150 Cuenca, Ecuador*

<sup>3</sup>*University of Diponegoro, Semarang 50275, Indonesia*

<sup>4</sup>*PDPM Indian Institute of Information Technology, Jabalpur 482005, India*

<sup>5</sup>*University of Wollongong, Wollongong, NSW 2522, Australia*

Correspondence should be addressed to Dong Wang; [dongwang4-c@my.cityu.edu.hk](mailto:dongwang4-c@my.cityu.edu.hk)

Received 12 November 2015; Accepted 12 November 2015

Copyright © 2016 Dong 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.

Some critical components, such as bearings, gearboxes, and impellers, are widely used in machines. Their faults may accelerate failures of other components and finally result in machine breakdowns. To prevent any unexpected machine breakdowns and accidents, early faults of critical components should be detected as soon as possible. Once early faults of critical components are diagnosed, their performance degradation assessment and remaining useful life estimation should be conducted to maximize lifetime of critical components. This special issue mainly focuses on fault diagnosis and prognosis of critical components and enables researchers and engineers to share their most recent developments and ideas so as to improve technologies adopted in the research community of machine fault diagnosis and prognosis.

We have received many submissions from different countries and regions. After rigorous peer review processes, the following manuscripts have been accepted and their contents are briefly summarized as follows.

In the paper entitled “Gearbox Fault Identification and Classification with Convolutional Neural Networks,” Z. Chen et al. proposed a deep learning convolutional neural network based intelligent method for gearbox fault diagnosis. The comparison with the classic support vector machine based diagnostic method demonstrated that their method has higher fault prediction accuracies.

In the paper entitled “Rolling Element Bearing Fault Diagnosis Based on Multiscale General Fractal Features,” W. Wen et al. proposed a multiscale general fractal dimensions based method to identify different bearing faults

under different operating conditions and they experimentally demonstrated that their method has better performance than wavelet packet and empirical mode decomposition based methods under different operating conditions.

In the paper entitled “Condition Monitoring and Fault Diagnosis for an Antifalling Safety Device,” G. Yang and H. Liang proposed a discrete wavelet transform based noise reduction method for condition monitoring and fault diagnosis of an antifalling safety device used in a construction elevator. Moreover, they experimentally found that the sym8 wavelet basis is the most effective wavelet basis among some wavelet basis candidates.

In the paper entitled “Screw Performance Degradation Assessment Based on Quantum Genetic Algorithm and Dynamic Fuzzy Neural Network,” X. Zhang et al. used quantum genetic algorithm to optimize dynamic fuzzy neural network so as to propose a screw performance degradation assessment method. The results show that the proposed method has higher prediction accuracies than backpropagation neural network and radial basis function neural network.

In the paper entitled “A New Transmissibility Based Indicator of Local Variation in Structure and Its Application for Damage Detection,” X. Z. Li et al. proposed a new transmissibility based indicator to monitor local variations of stiffness and damping for structure health condition monitoring. The results demonstrated that their proposed indicator is effective in localizing cracks of beam structures.

In the paper entitled “Planetary Gearbox Vibration Signal Characteristics Analysis and Fault Diagnosis,” Q. Miao and

Q. Zhou proposed a planetary vibration model to understand planetary vibration components. The results showed that their proposed model matches vibration data collected from a planetary test rig.

In the paper entitled “Multifault Diagnosis of Rolling Element Bearings Using a Wavelet Kurtogram and Vector Median-Based Feature Analysis,” P. H. Nguyen and J.-M. Kim proposed a wavelet kurtogram and vector median-based feature analysis method to diagnose different bearing faults under different rotational speeds and crack sizes. Compared with principle component analysis, independent component analysis, and linear discriminant analysis, their proposed analysis method has better performance for feature extraction.

In the paper entitled “Fault Diagnosis for a Multistage Planetary Gear Set Using Model-Based Simulation and Experimental Investigation,” G. Li et al. proposed a coupled lateral and torsional dynamic model to predict modulation sidebands of a two-stage compound planetary gear set. Moreover, fault characteristic frequencies were derived. Their work provided a theoretical support for fault diagnosis of a multistate planetary gear set by using modulation sideband analysis.

In the paper entitled “In Situ Measurement of Seeking Speed and Seeking Induced Head-Disk Interface Instability in Hard Disk Drives,” Y. Wang et al. investigated the instability of head-disk interface caused by the voice coil motor end crashing the crash stop in the process of magnetic head seeking and they proposed a maximum likelihood estimation and extended Kalman filter based method to seek speed. The results show that the rolling mode of air bearing will lead to a serious disk damage in the case of high seeking speed.

In the paper entitled “Cyclostationary Analysis for Gearbox and Bearing Fault Diagnosis,” Z. Feng and F. Chu derived explicit expressions of cyclic correlation and cyclic spectrum of modulation and frequency modulation signals and thoroughly investigated their applications to gearbox and bearing fault diagnosis.

In the paper entitled “Sparse Signal Representations of Bearing Fault Signals for Exhibiting Bearing Fault Features,” W. Peng et al. proposed an optimal wavelet filtering based sparse signal presentation method. The results demonstrated that their proposed method is effective in using a few wavelet coefficients to represent bearing fault features.

In the paper entitled “Feature Extraction of Faulty Rolling Element Bearing under Variable Rotational Speed and Gear Interferences Conditions,” D. Zhao et al. proposed an instantaneous dominant meshing multiply and empirical mode decomposition based on bearing fault feature extraction method under varying operation conditions. Multiple vibration mixtures were used to verify the effectiveness of their proposed method.

In the paper entitled “Two General Architectures for Intelligent Machine Performance Degradation Assessment,” Y. Xu et al. proposed two general architectures for intelligent machine performance degradation assessment. Their proposed architectures have wide applications, such as gearbox performance degradation and bearing performance degradation.

In the paper entitled “Application of Reassigned Wavelet Scalogram in Wind Turbine Planetary Gearbox Fault Diagnosis under Nonstationary Conditions,” X. Chen and Z. Feng applied reassigned wavelet scalogram to diagnose wind turbine planetary gearboxes. The experimental and in situ evaluations were conducted to verify the effectiveness of reassigned wavelet scalogram for turbine planetary gearbox fault diagnosis.

In the paper entitled “Research on the Sparse Representation for Gearbox Compound Fault Features Using Wavelet Bases,” C. Luo et al. proposed a split augmented Lagrangian shrinkage algorithm based method to extract gearbox compound fault features. Laplace wavelet optimized by correlation filtering was used to design a dictionary for constructing redundant basis functions. The results demonstrated the effectiveness of their proposed method and showed that a series of sparse coefficients are able to show gearbox compound fault features.

In the paper entitled “A Hybrid Prognostic Approach for Remaining Useful Life Prediction of Lithium-Ion Batteries,” W.-A. Yang et al. proposed a selective kernel ensemble-based relevance vector machine and particle swarm optimization based prognostic method for remaining useful life prediction. Their proposed method has potential applications to prognostics of gearbox, bearing, and so on.

In the paper entitled “Application of Artificial Neural Network for Damage Detection in Planetary Gearbox of Wind Turbine,” M. Strackiewicz, and T. Barszcz proposed an artificial neural network and linear regression based method to distinguish healthy and damaged planetary gearbox conditions. Their proposed method has great potential for condition monitoring of various rotating machinery, such as mining excavators and gas turbines.

In the paper entitled “Trace Ratio Criterion-Based Kernel Discriminant Analysis for Fault Diagnosis of Rolling Element Bearings Using Binary Immune Genetic Algorithm,” W.-A. Yang et al. proposed a trace ratio criterion-based kernel discriminant analysis based method for bearing fault diagnosis. The parameters of their proposed method were optimized by binary immune genetic algorithm. The results showed that their proposed method is better than some popular intelligent methods for bearing fault diagnosis.

In the paper entitled “Gearbox Fault Diagnosis Using Complementary Ensemble Empirical Mode Decomposition and Permutation Entropy,” L. Zhao et al. proposed a complementary ensemble empirical mode decomposition and permutation entropy based method for gearbox fault diagnosis. The results showed that their proposed method is able to improve prediction accuracies of support vector machines.

We sincerely hope that this special issue could provide useful technologies and information for researchers and engineers in the research community of fault diagnosis and prognosis.

## **Acknowledgments**

We would like to thank all authors for submitting their works to this special issue. We also would like to thank all reviewers and editors for their valuable comments on all manuscripts.

*Dong Wang*  
*Chuan Li*  
*Achmad Widodo*  
*Pavan Kumar Kankar*  
*Wahyu Caesarendra*

## Research Article

# A Hybrid Prognostic Approach for Remaining Useful Life Prediction of Lithium-Ion Batteries

Wen-An Yang,<sup>1</sup> Maohua Xiao,<sup>2</sup> Wei Zhou,<sup>3</sup> Yu Guo,<sup>1</sup> and Wenhe Liao<sup>1</sup>

<sup>1</sup>College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

<sup>2</sup>College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

<sup>3</sup>Nanjing Surveying and Mapping Instrument Factory, Nanjing 210003, China

Correspondence should be addressed to Wen-An Yang; [dreamflow@nuaa.edu.cn](mailto:dreamflow@nuaa.edu.cn)

Received 3 July 2015; Revised 30 October 2015; Accepted 1 November 2015

Academic Editor: Chuan Li

Copyright © 2016 Wen-An Yang 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.

Lithium-ion battery is a core component of many systems such as satellite, spacecraft, and electric vehicles and its failure can lead to reduced capability, downtime, and even catastrophic breakdowns. Remaining useful life (RUL) prediction of lithium-ion batteries before the future failure event is extremely crucial for proactive maintenance/safety actions. This study proposes a hybrid prognostic approach that can predict the RUL of degraded lithium-ion batteries using physical laws and data-driven modeling simultaneously. In this hybrid prognostic approach, the relevant vectors obtained with the selective kernel ensemble-based relevance vector machine (RVM) learning algorithm are fitted to the physical degradation model, which is then extrapolated to failure threshold for estimating the RUL of the lithium-ion battery of interest. The experimental results indicated that the proposed hybrid prognostic approach can accurately predict the RUL of degraded lithium-ion batteries. Empirical comparisons show that the proposed hybrid prognostic approach using the selective kernel ensemble-based RVM learning algorithm performs better than the hybrid prognostic approaches using the popular learning algorithms of feedforward artificial neural networks (ANNs) like the conventional backpropagation (BP) algorithm and support vector machines (SVMs). In addition, an investigation is also conducted to identify the effects of RVM learning algorithm on the proposed hybrid prognostic approach.

## 1. Introduction

Lithium-ion batteries are significant energy solution for many systems (e.g., satellite, spacecraft, and electric vehicles) due to their high energy density, high galvanic potential, lightness of weight, and long lifetime compared to lead-acid, nickel-cadmium, and nickel-metal-hydride cells [1]. Their failure can lead to reduced capability, downtime, and even catastrophic breakdowns. For example, in November 2006, The National Aeronautics and Space Administration's Mars Global Surveyor stopped working after the radiator for its batteries was positioned towards the sun causing an increase in the temperature of the batteries, which resulted in lost charge capacity [2]. Battery health management would greatly enhance the reliability of such systems. Thus, this raises the challenging issue of remaining useful life (RUL) prediction in relation to lithium-ion batteries.

In the past few years, much research effort has been devoted to developing approaches to lithium-ion battery degradation modeling and RUL prediction. In general, these approaches can be classified into categories of model-based and/or data-driven methodologies. The model-based methodologies attempt to constitute physical models of the lithium-ion battery for RUL prediction. Recently, various Bayesian filtering models such as Kalman filter [3], extended Kalman filter [4–6], particle filter [7–9], and unscented particle filter [10] have been extensively used to construct exhaustive models of deteriorating lithium-ion batteries. However, uncertainty due to assumptions and simplifications in the models may impose severe limitations upon their applicability in practical applications. In order to overcome the aforementioned problems that can occur with the model-based methodologies, intensive research has been conducted into the utilization of various data-driven methodologies,

for example, autoregressive moving average (ARMA) models [11], artificial neural networks (ANNs) [12], and support vector machines (SVMs) [13], to model lithium-ion battery degradation and to predict the RUL of lithium-ion batteries. Data-driven techniques utilize monitored operational data related to lithium-ion battery health. Compared with the model-based methodologies, the data-driven methodologies may be more appropriate when the understanding of first principles of system operation is not comprehensive or when the system is so complex such that developing an accurate model is prohibitively expensive but sufficient data are available for constructing a map of the performance degradation space. Furthermore, rapid development has recently been achieved in automatic data collection and processing of real-time field data, which hugely facilitate the continuous monitoring of the state of health of operating lithium-ion batteries and the lean management of the related large amount of reference data. The most natural data-driven methodology for RUL prediction is to fit a curve of the available data of the lithium-ion battery degradation evolution using regression models and then to extrapolate the curve to the criteria indicating failure. In practice, however, the lithium-ion battery degradation history available may be short and incomplete and even differ significantly because of different operating conditions, so that a common extrapolation may lead to large errors and unreliable results. The same problem arises when employing ARMA models, although the method can handle the situation in which more run-to-failure data are unavailable or insufficient. With respect to ANNs, they have the advantages of superior learning, noise suppression, and parallel computation abilities. However, despite their advantages, ANNs also have some disadvantages: (1) design and training often lead to a complex and time-consuming task, in which architecture and many training parameters must be tuned; (2) minimization of the training errors can result in poor generalization performance; and (3) performance can be degraded when working with low-sized datasets. With respect to SVMs, they are powerful in solving problems with small samples, nonlinearities, and local minimum. However, despite their advantages, SVMs also have some disadvantages: (1) by assuming an explicit loss function (usually, the  $\epsilon$ -insensitive loss function), one assumes a fixed distribution of the residuals; (2) the soft margin parameter  $C$  must be tuned usually through cross-validation methods, which result in time-consuming tasks; (3) the kernel function used in SVM must satisfy Mercer's theorem to be valid; and (4) sparsity is not always achieved and a high number of support vectors are thus obtained.

More recently, some researchers have attempted to combine model-based and data-driven methods for RUL prediction of lithium-ion batteries in order to leverage the strength from both data-driven methodology and model-based methodology and have obtained promising results [14]. Most of the combination of model-based and data-driven methods in literature has focused on the utilization of relevance vector machines (RVMs) in place of ANNs or SVMs as the prognostic technique. RVM, a general Bayesian probabilistic framework of SVM, can efficiently

alleviate some of these shortcomings of SVMs [15]. Saha et al. employed a RVM to find the most representative relevant vectors to fit the capacity degradation data of lithium-ion batteries [16]. Maio et al. combined a RVM and an exponential function to predict the RUL of bearings [17]. Zio and Maio employed a RVM to find the most representative relevant vectors to fit a crack growth model for predicting RUL [18]. Wang et al. employed a RVM to find the most representative relevant vectors to fit the three-parameter capacity degradation model to predict the RUL of lithium-ion batteries [19]. A review of the related literature also indicates that similar idea has already been investigated in the area of applying SVM to RUL prediction. Benkedjouh et al. [20] employed a SVM to find the most representative support vectors to fit a power model for RUL prediction of the cutting tool. Also based on a similar idea, Benkedjouh et al. employed a SVM to find the most representative support vectors to fit an exponential regression for bearing performance degradation assessment and RUL estimation [21]. The ability to extract the relevant vectors is very useful for making good predictions, as the relevant vectors can be used to find the representative training vectors containing the cycles of the relevant vectors and the predictive values at the cycles of the relevant vectors. A review of the related literature [16–21] also indicates that, for the hybrid prognostic approaches that are based on RVM learning algorithm, their RUL prediction performances are very sensitive to kernels choice and kernel parameters setting. A kernel (or kernel parameter setting) that works well for one situation might not be the appropriate choice for the other. However, no systematic methodology as yet has been established for determining the optimal kernel type and kernel parameters for the RVM learning algorithm. Most of the previous work in the area of applying RVM to RUL prediction determined single kernel and kernel parameters by trial and error and did not deal with automatic kernel choice and kernel parameters optimization.

According to the literature review given above, the aim of this study is to develop a hybrid prognostic approach of physical laws and data-driven modeling that integrates selective kernel ensemble-based RVM (a data-driven methodology) and exponential regression (a model-based methodology) for on-line RUL prediction of lithium-ion batteries. The choice of kernel (and kernel parameters) of RVM is evolutionarily determined via coevolutionary swarm intelligence, without the need of any human intervention. A sum of two exponential functions' model is fitted to these relevant vectors to predict the RUL of degraded lithium-ion batteries. The experimental results indicate that the proposed hybrid prognostic approach can accurately predict the RUL of degraded lithium-ion batteries. Empirical comparisons show that the proposed hybrid prognostic approach using the selective kernel ensemble-based RVM learning algorithm performs better than the hybrid prognostic approaches using popular learning algorithms of feedforward artificial neural networks (ANNs) like the conventional backpropagation (BP) algorithm and support vector machines (SVMs). The proposed hybrid prognostic approach using the selective kernel ensemble-based RVM learning algorithm outperforms

the hybrid prognostic approaches using the single kernel-based RVM learning algorithm and the Ensemble All-based RVM learning algorithm.

The rest of this study is organized as follows. Section 2 gives a review of the RVM basic framework. Section 3 presents a selective kernel ensemble-based RVM learning algorithm. Section 4 describes a hybrid prognostic approach for RUL prediction of lithium-ion batteries. Section 5 conducts an investigation to identify the effects of RVM learning on the hybrid prognostic approach. Section 6 provides an empirical comparison of the proposed hybrid prognostic approach with other existing approaches. Section 7 presents a concluding summary and suggests some directions for future research.

## 2. Review of Relevance Vector Machine

RVM is a Bayesian form representing a generalized linear model of identical functional form of SVM. Unlike SVM, RVM can provide probabilistic interpretation of its outputs [15]. As a supervised learning, RVM starts with a dataset of input-target pairs  $\{\mathbf{x}_n, t_n\}_{n=1}^N$ . The aim is to learn a model of the dependency of the targets on the inputs to make accurate prediction of  $t$  for previously unseen values of  $\mathbf{x}$ . Typically, the predictions are based on a function  $y(\mathbf{x})$  defined over the input space, and learning is the process of inferring (perhaps the parameters of) this function. In the context of SVM, this function takes the following form:

$$y(\mathbf{x}) = \sum_{n=1}^N w_n K(\mathbf{x}, \mathbf{x}_n) + w_0, \quad (1)$$

where  $\{w_n\}$  are the model “weights,”  $w_0$  is bias, and  $K(\mathbf{x}, \mathbf{x}_n)$  is a kernel function.

By considering only the scalar valued output we follow the standard probabilistic formulation and add additive noise with output samples for better data overfitting, which is described as follows:

$$t_n = y(\mathbf{x}_n) + \varepsilon_n, \quad (2)$$

where  $\varepsilon_n$  are independent samples from some noise process which is further assumed to be zero-mean Gaussian noise with variance  $\sigma^2$ .

The likelihood of the complete dataset can be written as

$$p(\mathbf{t} | \mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-(N/2)} \exp\left\{-\frac{1}{2\sigma^2} \|\mathbf{t} - \Phi\mathbf{w}\|^2\right\}, \quad (3)$$

where  $\mathbf{t} = [t_1, t_2, \dots, t_N]^T$ ,  $\mathbf{w} = [w_0, w_1, w_2, \dots, w_N]^T$ , and  $\Phi$  is the  $N \times (N + 1)$  “design” matrix with  $\Phi = [\boldsymbol{\varphi}(\mathbf{x}_1), \boldsymbol{\varphi}(\mathbf{x}_2), \dots, \boldsymbol{\varphi}(\mathbf{x}_N)]^T$ , wherein  $\boldsymbol{\varphi}(\mathbf{x}_n) = [1, K(\mathbf{x}_n, \mathbf{x}_1), K(\mathbf{x}_n, \mathbf{x}_2), \dots, K(\mathbf{x}_n, \mathbf{x}_N)]^T$ .

Maximizing likelihood prediction of  $\mathbf{w}$  and  $\sigma^2$  in (3) often leads to overfitting. Therefore, a preference for smoother functions is encoded by choosing a zero-mean Gaussian prior distribution over  $\mathbf{w}$ :

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \prod_{i=0}^N N(w_i | 0, \alpha_i^{-1}), \quad (4)$$

where  $\boldsymbol{\alpha}$  is a vector of  $(N + 1)$  hyperparameters.

Using Bayes’ rule, the posterior over all unknowns can be computed; that is,

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) = \frac{p(\mathbf{t} | \mathbf{w}, \boldsymbol{\alpha}, \sigma^2) p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t})}. \quad (5)$$

However, we cannot compute the solution of the posterior  $p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t})$  in (5) directly. But we can decompose the posterior as  $p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) = p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2 | \mathbf{t})$ , where

$$\begin{aligned} p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) &= \frac{p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha})}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)} \\ &= (2\pi)^{-(N+1)/2} |\boldsymbol{\Sigma}|^{-1/2} \\ &\cdot \exp\left\{-\frac{1}{2} (\mathbf{w} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{w} - \boldsymbol{\mu})\right\}, \end{aligned} \quad (6)$$

where the posterior covariance and mean are expressed as follows:

$$\begin{aligned} \boldsymbol{\Sigma} &= (\Phi^T \mathbf{B} \Phi + \mathbf{A})^{-1} \\ \boldsymbol{\mu} &= \boldsymbol{\Sigma} \Phi^T \mathbf{B} \mathbf{t} \end{aligned} \quad (7)$$

with  $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$  and  $\mathbf{B} = \sigma^{-2} \mathbf{I}_N$ . Thus, RVM method becomes the search for the best hyperparameters posterior mode. Predictions for new data are then made according to integration of the weights to obtain the marginal likelihood for the hyperparameters:

$$\begin{aligned} p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2) &= \int p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}) d\mathbf{w} = (2\pi)^{-N/2} \\ &\cdot |\mathbf{B}^{-1} + \Phi \mathbf{A}^{-1} \Phi^T|^{-1/2} \\ &\cdot \exp\left\{-\frac{1}{2} \mathbf{t}^T (\mathbf{B}^{-1} + \Phi \mathbf{A}^{-1} \Phi^T)^{-1} \mathbf{t}\right\}. \end{aligned} \quad (8)$$

The hyperparameters  $\boldsymbol{\alpha}$  and  $\sigma^2$  which maximize (8) are obtained by using an alternate reproduction approach [15], because values of  $\boldsymbol{\alpha}$  and  $\sigma^2$  cannot be directly calculated in closed form. Suppose that the values of  $\boldsymbol{\alpha}_{MP}$  and  $\sigma_{MP}^2$  that can maximize (8) are obtained. Then we can compute the predictive distribution for a new input  $\mathbf{x}_*$  by using (6):

$$\begin{aligned} p(t_* | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) \\ = \int p(t_* | \mathbf{w}, \sigma_{MP}^2) p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) d\mathbf{w}. \end{aligned} \quad (9)$$

Since both terms in the integral are Gaussian, one can easily compute the probability as follows:

$$p(t_* | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) = N(t_* | y_*, \sigma_*^2), \quad (10)$$

where the mean and variance of the predicted value are, respectively,

$$y_* = \boldsymbol{\mu}^T \boldsymbol{\varphi}(\mathbf{x}_*), \quad (11)$$

$$\sigma_*^2 = \sigma_{MP}^2 + \boldsymbol{\varphi}(\mathbf{x}_*)^T \boldsymbol{\Sigma} \boldsymbol{\varphi}(\mathbf{x}_*). \quad (12)$$

The variance of the predicted value (i.e., (12)) is the sum of the variance associated with noise in the training data and uncertainty associated with prediction of weights.

### 3. Selective Kernel Ensemble-Based Relevance Vector Machine

As mentioned in Section 1, kernel types and kernel parameters have significant influences on the generalization capability of the RVM learning. Generally, commonly used basic kernels for RVM learning include Gaussian kernel (i.e., (13)), Exponential kernel (i.e., (14)), Laplacian kernel (i.e., (15)), Polynomial kernel (i.e., (16)), Sigmoid kernel (i.e., (17)), Cauchy kernel (i.e., (18)), and Multiquadric kernel (i.e., (19)):

$$K_{\text{Gau}}(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma_{\text{Gau}}^2}\right), \quad (13)$$

$$K_{\text{Exp}}(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|}{2\sigma_{\text{Exp}}^2}\right), \quad (14)$$

$$K_{\text{Lap}}(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|}{\sigma_{\text{Lap}}}\right), \quad (15)$$

$$K_{\text{Poly}}(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}^T \mathbf{x}_i + 1)^d, \quad (16)$$

$$K_{\text{Sig}}(\mathbf{x}, \mathbf{x}_i) = \tanh(\xi \mathbf{x}^T \mathbf{x}_i + \eta), \quad (17)$$

$$K_{\text{Cau}}(\mathbf{x}, \mathbf{x}_i) = \frac{1}{1 + \|\mathbf{x} - \mathbf{x}_i\|^2 / \sigma_{\text{Cau}}^2}, \quad (18)$$

$$K_{\text{Mul}}(\mathbf{x}, \mathbf{x}_i) = \sqrt{\|\mathbf{x} - \mathbf{x}_i\|^2 + \tau^2}, \quad (19)$$

where  $\sigma_{\text{Gau}}$ ,  $\sigma_{\text{Exp}}$ ,  $\sigma_{\text{Lap}}$ ,  $d$ ,  $\xi$ ,  $\eta$ ,  $\sigma_{\text{Cau}}$ , and  $\tau$  are kernel parameters that need to be finely tuned. It is impossible to fully determine which one is the best kernel for all problems, because the choice of a kernel depends on the problem at hand. For example, Gaussian kernel is a local kernel and Polynomial kernel is a global kernel [22]. In the case of local kernel, only the data that are close or in the proximity of each other have an influence on the kernel values [22]. In the case of global kernel, samples that are far away from each other still have an influence on the kernel value [22]. With respect to Gaussian kernel and Polynomial kernel only, the mixture of these two basic kernels has been demonstrated to substantively improve the generalization performance of the SVM [23, 24]. However, for many existing basic kernels mentioned above, this combination of basic kernels can also be different for different problems. In one extreme case where all of the individual basic kernels are completely identical, the size of the combination can be reduced without sacrificing the generalization performance of the RVM. In addition, in some scenarios, eliminating some unacceptable basic kernels and meanwhile selecting several acceptable ones to construct a kernel ensemble may be better than combining all of those basic kernels. In this study, each kernel applied to RVM learning algorithm is a selective kernel ensemble of these basic kernels. It should be noted that although

the multikernel idea has been successfully used in several machine learning models [25–28] that assume a weighted linear sum of basic kernel and estimate the kernel weights during training, to the best of the authors' knowledge, it is the first time that a multikernel version of RVM with adaptive kernel selections, adaptive kernel combinations, and adaptive kernel parameters optimization is proposed. The selective kernel ensemble can be expressed as follows:

$$K(\mathbf{x}, \mathbf{x}_i) = \sum_{h=1}^H \lambda_h \rho_h K_h(\mathbf{x}, \mathbf{x}_i) \quad (20)$$

$$\sum_{h=1}^H \lambda_h = 1$$

$$0 \leq \lambda_h \leq 1$$

$$\rho_h = \begin{cases} 1, & \text{being selected} \\ 0, & \text{otherwise,} \end{cases}$$

where  $H$  is the number of basic kernels under consideration and equals 7 in this study,  $K_h(\mathbf{x}, \mathbf{x}_i)$  denotes the  $h$ th basic kernel,  $\lambda_h$  stands for the weight assigned to  $K_h(\mathbf{x}, \mathbf{x}_i)$ , and  $\rho_h$  represents the selection label assigned to  $K_h(\mathbf{x}, \mathbf{x}_i)$ .

**3.1. Selection of Candidate Basic Kernels.** Instead of combining all of candidate basic kernels, selective kernel ensemble tries to select an optimal subset of individual basic kernels to constitute a selective convex combination. However, selecting an optimal subset from candidate basic kernels is not an easy task since the space of possible subsets is very large ( $2^H - 1$ ) for a basic kernel population of size  $H$ . It is very difficult if not impractical to use exhaustive search to find an optimal subset if and especially when  $H$  is a large number. In this study, discrete particle swarm optimization (DPSO) [29] algorithm is used for obtaining an optimal subset from candidate basic kernels. Each dimension of a particle in DPSO is encoded by binary bit, where each element of "1" (i.e.,  $\rho_h = 1$ ) denotes an individual basic kernel appearing in the selective kernel ensemble while "0" (i.e.,  $\rho_h = 0$ ) denotes its absence,  $h = 1, 2, \dots, 7$ . The optimal subset of individual basic kernels can be obtained according to the best evolved selective label vector  $\Theta = (\rho_1, \rho_2, \dots, \rho_7)$  that can achieve the maximum fitness value. Thus, such a DPSO bit representation gets rid of the tedious trial-and-error search for an optimal subset of basic kernels.

**3.2. Determination of Kernel Parameters and Additional Weights.** Although utilization of selective kernel ensemble can relieve the influence of kernel types on the generalization capability of RVM, it involves 7 additional weight coefficients ( $\lambda_1, \lambda_2, \dots, \lambda_7$ ). In addition, more component basic kernels mean more kernel parameters. It is not easy to determine the optimal values of all these design parameters, including kernel parameters ( $\sigma_{\text{Gau}}$ ,  $\sigma_{\text{Exp}}$ ,  $\sigma_{\text{Lap}}$ ,  $d$ ,  $\xi$ ,  $\eta$ ,  $\sigma_{\text{Cau}}$ , and  $\tau$ ) and convex combination coefficients ( $\lambda_1, \lambda_2, \dots, \lambda_7$ ) that can allow the RVM to achieve the

maximum performance. In this circumstance, manual trial-and-error method is absolutely tedious and unacceptable. Moreover, manual trial-and-error method does not necessarily guarantee a good decision, because these parameters usually interact with each other nonlinearly. In this study, these 7 additional weight coefficients ( $\lambda_1, \lambda_2, \dots, \lambda_7$ ) together with kernel parameters ( $\sigma_{\text{Gau}}, \sigma_{\text{Exp}}, \sigma_{\text{Lap}}, d, \xi, \eta, \sigma_{\text{Cau}}$ , and  $\tau$ ) constitute a general real-value parameter vector  $\Psi = \{\lambda_1, \lambda_2, \dots, \lambda_7, \sigma_{\text{Gau}}, \sigma_{\text{Exp}}, \sigma_{\text{Lap}}, d, \xi, \eta, \sigma_{\text{Cau}}, \tau\}$ , which will be represented in the population of continuous particle swarm optimization (CPSO) [30]. Thus, such a CPSO real-value representation gets rid of the tedious trial-and-error search for optimal kernel parameters and additional weights.

**3.3. Coevolution of DPSO and CPSO.** As mentioned in Sections 3.1 and 3.2, one swarm population DPSO with population size  $s_{\text{DPSO}}$  and the other swarm population CPSO with population size  $s_{\text{CPSO}}$  are involved in equipping the RVM with adaptive kernel selections, adaptive kernel combinations, and adaptive kernel parameters optimization. From a pure DPSO perspective, this suffices for the design of the RVM with the best kernel selection, but without taking kernel parameters and weights in kernel combination into account; that is, only good kernel selection obtained with DPSO may not necessarily mean good RVM performance. Similarly, only good kernel parameters and weights in kernel combination may not necessarily evoke maximum RVM performance. Therefore, the evolution of kernel selections by DPSO and the evolution of kernel combinations and kernel parameters by CPSO should be taken into consideration simultaneously. Inspired by the coevolution of swarms, a coevolutionary PSO scheme is proposed in this section. In the proposed coevolutionary PSO scheme, the DPSO and the CPSO interact with each other through the fitness evaluation. Within each iteration, the DPSO is run for a certain number ( $g_{\text{DPSO}}$ ) of generations; then the CPSO is run for a certain number ( $g_{\text{CPSO}}$ ) of generations; this process is repeated until either an acceptable solution has been obtained or the maximum number ( $\text{max.i}_{\text{DPSO}}$ ) of iterations has been reached. The global best in the population of DPSO is the final solution for the selection label vector, and the global best in the population of CPSO is the final solution for the general parameter vector with regard to kernel parameters and additional weight coefficients. The procedure of coevolution of DPSO and CPSO is outlined in the following pseudocode.

*Step 1.* Initialize randomly one swarm population DPSO with population size  $s_{\text{DPSO}}$ .

*Step 2.* Initialize randomly the other swarm population CPSO with population size  $s_{\text{CPSO}}$ .

*Step 3.* Run the DPSO for  $g_{\text{DPSO}}$  generations.

*Step 4.* Reevaluate the personal best values for the CPSO if it is not the first cycle.

*Step 5.* Run the CPSO for  $g_{\text{CPSO}}$  generations.

*Step 6.* Reevaluate the personal best values for the DPSO.

*Step 7.* Go back to Step 3. Repeat this procedure until a termination criterion is reached.

In the above coevolutionary PSO scheme, when one PSO is running, the other PSO serves as its ecological environment; that is, for each PSO its ecological environment has varied from iteration to iteration. Therefore, the personal best obtained in the previous iteration has to be reevaluated in accordance with the new ecological environment before playing its coevolving role. It is also worth noting that, in each generation of the coevolution, the real weights are normalized so that the selected individual basic kernels are combined using a weighted average. Hence, this study uses a quite simple normalization scheme as follows:

$$\tilde{\lambda}_h = \frac{\lambda_h \rho_h}{\sum_{h=1}^7 \lambda_h \rho_h}. \quad (21)$$

## 4. Hybrid Prognostic Approach for RUL Prediction

As a lithium-ion battery ages, its maximum capacity begins to deteriorate over time. If the maximum capacity falls below 80% of its initial rated capacity, the battery is considered to be unable to provide reliable power supplies and needs to be replaced. In the current academia/industry practices, reliability of a lithium-ion battery for providing reliable power supplies is ensured via the prediction of the remaining maximum capacity. In this study, a hybrid prognostic approach that integrates selective kernel ensemble-based RVM learning algorithm and exponential regression is proposed for RUL prediction of lithium-ion batteries. Figure 1 shows an overall flowchart of the proposed hybrid prognostic approach.

**4.1. Capacity Degradation Data Collection.** To develop the degradation model, four lithium-ion batteries under test went through the full charge and discharge procedure. These four lithium-ion batteries in the following text are referred to as A1, A2, A3, and A4, respectively. Noting that, these batteries have a graphite anode and a lithium cobalt oxide cathode which were verified using electron dispersive spectroscopy (EDS). The rated capacity of the tested lithium-ion battery was 0.9 Ah. Multiple charge-discharge tests were performed with an Arbin BT2000 battery testing system under ambient temperature (around 25°C). The discharge current was 0.45 A. Cut-off voltage was 2.5 V. The failure threshold of the lithium-ion batteries was 0.72 Ah. The discharge capacity was recorded after each full charge-discharge cycle. Herein,  $N$  successive capacity degradation measurements are denoted as  $\mathbf{Q} = (q_1, q_2, \dots, q_N)^T$  and their corresponding cycles are  $\mathbf{I} = (1, 2, \dots, N)^T$ . Therefore, the battery capacity degradation condition can be monitored through the measurements of the input-target pairs  $\{l, q_l\}$ , where  $l = 1, 2, \dots, N$ . Inspection of the battery capacity degradation state is made at the predefined inspection cycles of  $(C_1, C_2, \dots, C_J)$ , respectively. At each predefined inspection cycle  $C_i$ , along the developing lithium-ion battery degradation-to-failure trajectory, the last input-target pair of  $\{l, q_l\}$  ( $l = C_i$ ) is recorded and appended to the vector of the input-target pairs  $\{l, q_l\}$  ( $l = 1, 2, \dots, C_i - 1$ )

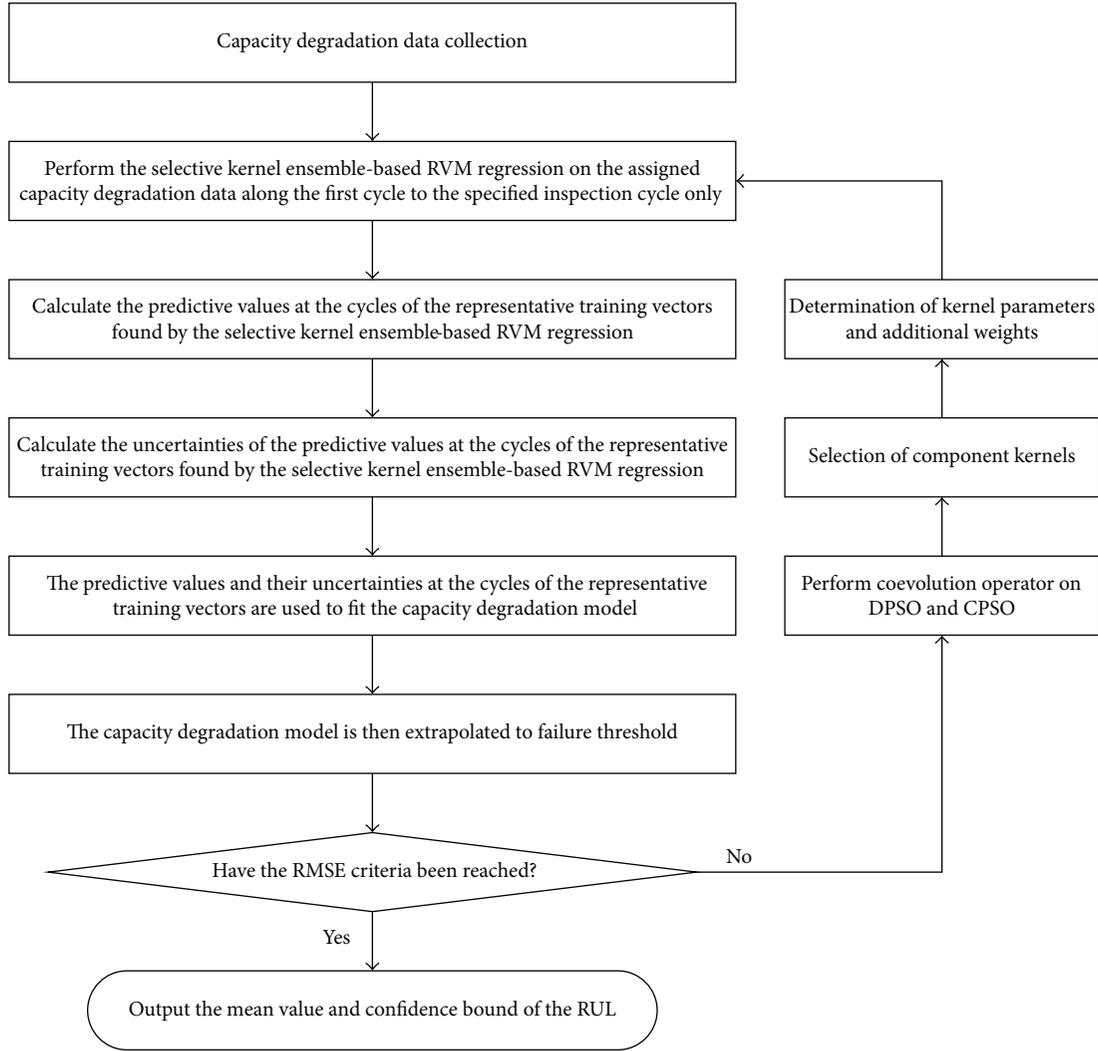


FIGURE 1: Flowchart of the proposed hybrid prognostic approach for RUL prediction of lithium-ion batteries.

collected at the previous  $(C_i - 1)$  inspections, so that the capacity degradation data used for degradation model development at the inspection cycle  $C_i$  is collected.

**4.2. Degradation Model Formulation.** At each inspection cycle  $C_i$ , the selective kernel ensemble-based RVM learning algorithm is performed on the  $N$  available input-target pairs of data  $\{l, q_l\}$ , where  $l = 1, 2, \dots, N$ , and thus the  $M \leq N$  most representative input-target pairs of data  $\{l_m^*, q_{lm}^*\}$  identified by the RVM regression, that is, the relevant vectors whose corresponding basis functions are associated with the remaining nonzero weights, are collected in a sparse dataset, where  $m = 1, 2, \dots, M$  and  $y_{lm}^*$  is the capacity estimate provided by the RVM in correspondence with  $l_m^*$ . For convenience of expression, the sparse dataset is referred to as  $\{\bar{l}_M^*, \bar{q}_{lM}^*\}$ . Then, fitting to the sparse dataset  $\{\bar{l}_M^*, \bar{q}_{lM}^*\}$  is performed to identify the unknown parameters of the model adopted. Finally, the fitted model is extrapolated up to the predefined failure threshold of 0.72 Ah to predict the RUL at inspection cycle  $C_i$ ,  $RUL(C_i)$ .

An important issue in developing a capacity degradation model is determining the fitted model, which influences substantially the prognostics performance of the proposed hybrid prognostic approach. The appropriate fitted model depends on the battery under consideration. Goebel et al. [31] used a sum of two exponential functions to model the increase of internal impedance due to solid-electrolyte interface thickening with time. As battery capacity degradation is closely related to the internal impedance increase, potential models for capacity degradation can also be exponential models. Following up Goebel et al. [31] work, He et al. [2] have experimentally demonstrated that the sum of two exponential functions can well describe the capacity degradation trend of many different batteries:

$$q_l = a \cdot \exp(b \cdot l) + c \cdot \exp(d \cdot l), \quad (22)$$

where  $q_l$  is the capacity of the battery at the cycle  $l$ ;  $a$  and  $b$  are the parameters associated with the internal impedance; and  $c$  and  $d$  are the parameters associated with the aging rate. To demonstrate the suitability of the model in (22)

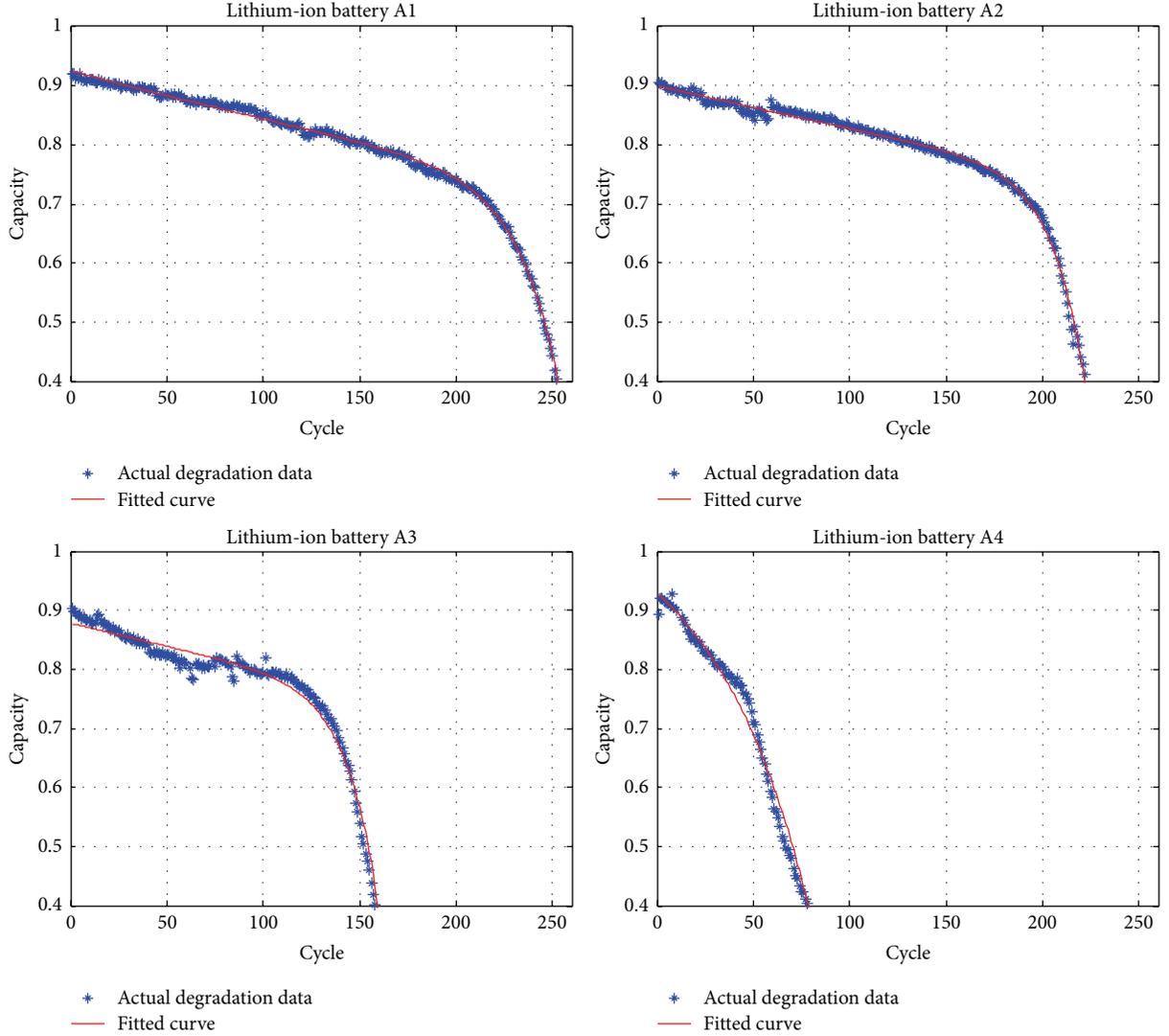


FIGURE 2: Degradation data and fitting curves of the four lithium-ion batteries A1, A2, A3, and A4.

in depicting battery degradation being addressed, Figure 2 shows the curve fitting result (solid line) of (22) to the capacity data of these four batteries, which indicates that the sum of two exponential functions closely agrees with the values of the measured capacity degradation data. Hence, in this study, the sum of two exponential functions was used to fit the degradation curves of the lithium-ion batteries on the basis of the sparse dataset  $\{\bar{l}_M^*, \bar{q}_{lM}^*\}$ .

**4.3. RUL Prediction.** The RUL of lithium-ion batteries can be obtained by extrapolating the fitted model to a predefined failure threshold. The predicted RUL at the inspection cycle  $C_i$  is then derived by projecting the  $M$  state estimates, namely,  $\boldsymbol{\mu}^T \boldsymbol{\varphi}(l_1^*), \boldsymbol{\mu}^T \boldsymbol{\varphi}(l_2^*), \dots, \boldsymbol{\mu}^T \boldsymbol{\varphi}(l_{M-1}^*)$  and  $\boldsymbol{\mu}^T \boldsymbol{\varphi}(l_M^*)$ , into the future until the future cycle  $C_\theta$  at which the predictive future capacity degradation value hits the predefined failure threshold. Thus, the predicted RUL at the inspection cycle  $C_i$

can be expressed as the differences between the inspection cycle  $C_i$  and the future cycle  $C_\theta$  and can be calculated as

$$\text{RUL}(C_i) = C_\theta - C_i. \quad (23)$$

**4.4. Experiment and Results.** In order to demonstrate the performance of the proposed hybrid prognostic approach for on-line prediction of lithium-ion battery RUL, four lithium-ion batteries A1, A2, A3, and A4 are employed in this experiment. These four illustrative batteries are in exactly the same experimental environments. For detailed information on these four lithium-ion batteries, please refer to Section 4.1. The battery capacity data used in this study is provided by the Center for Advanced Life Cycle Engineering, University of Maryland [32].

**4.4.1. Relative Parameter Settings.** This study uses the root mean squared error (RMSE) as a measure of accuracy

index to evaluate the performance of the proposed hybrid prognostic approach. The closer the value of the RMSE is to 0, the better the performance of the fitted model (i.e., (22)) is. Let  $RMSE_{\Theta\Psi}$  denote the RMSE of the fitted model enabled by the sparse dataset  $\{\bar{l}_M^*, \bar{q}_{LM}^*\}$  which was found by the selective kernel ensemble-based RVM with the selection vector  $\Theta$  and the general parameter vector  $\Psi$ . Thus,  $f(\Theta, \Psi) = 1/RMSE_{\Theta\Psi}$  is taken as the fitness function of the proposed hybrid prognostic approach.

In order to apply the developed coevolution of DPSO and CPSO for equipping the RVM with adaptive kernel selections, adaptive kernel combinations, and adaptive kernel parameters optimization, after a small number of simple trials relative parameter settings are determined. For the step of selection of candidate basic kernels, parameters of DPSO are set as follows: s\_DPSO, 30; acceleration coefficients  $c_1$  and  $c_2$ , 1.0 and 0.5; initial and final inertia weight, 0.9 and 0.2; initial and final inertia velocity, 4 and  $-4$ ; fitness function,  $f(\Theta, \Psi) = 1/RMSE_{\Theta\Psi}$ . For the step of determination of kernel parameters and additional weight coefficients, parameters of CPSO are set as follows: s\_CPSO, 60; acceleration coefficients  $c_1$  and  $c_2$ , 1.0 and 0.5; initial and final inertia weight, 0.9 and 0.4; initial and final inertia velocity, 4 and  $-4$ ; fitness function,  $f(\Theta, \Psi) = 1/RMSE_{\Theta\Psi}$ . For the step of coevolution of DPSO and CPSO, parameters of coevolutionary PSO are set as follows: g\_DPSO: 10; g\_CPSO: 15; i\_max\_DPSO: 50.

**4.4.2. Experimental Results.** To show the robustness of the proposed hybrid prognostic approach, we will perform four independent experiments for each battery. For this purpose, four different inspection cycles corresponding to 60%, 70%, 80%, and 90% data partition rates have been used for on-line prediction of lithium-ion battery RUL. That is, the inspection cycle was set at the cycle steps that separate the whole battery capacity degradation data available into two parts, where the first 60%, 70%, 80%, and 90% are used for RVM learning. Here we assume that the RVM learning with less than 60% battery capacity degradation data is inadequate.

The first case refers to the lithium-ion battery A1. The proposed hybrid prognostic approach is applied to the lithium-ion battery A1 capacity degradation data plotted with dots in Figures 3–6. As aforementioned in Section 4.4.1, four different inspection cycles corresponding to 60%, 70%, 80%, and 90% data partition rates have been used for on-line prediction of lithium-ion battery RUL; that is, predictions of the RUL of the lithium-ion battery A1 are calculated at the inspection cycles of 125, 146, 167, and 188, respectively. Table 1 summarizes the predicted RUL and the actual RUL when the inspection cycles of 125, 146, 167, and 188 were chosen for battery A1. In the second case, the lithium-ion battery A2 is investigated. The proposed hybrid prognostic approach is then conducted on the lithium-ion battery A2 capacity degradation data, which are plotted as the dots in Figures 7–10. Predictions of the RUL of the lithium-ion battery A2 are calculated at the inspection cycles of 113, 132, 151, and 170, respectively. Table 2 summarizes the predicted RUL and the actual RUL when the inspection cycles of 113, 132, 151, and 170

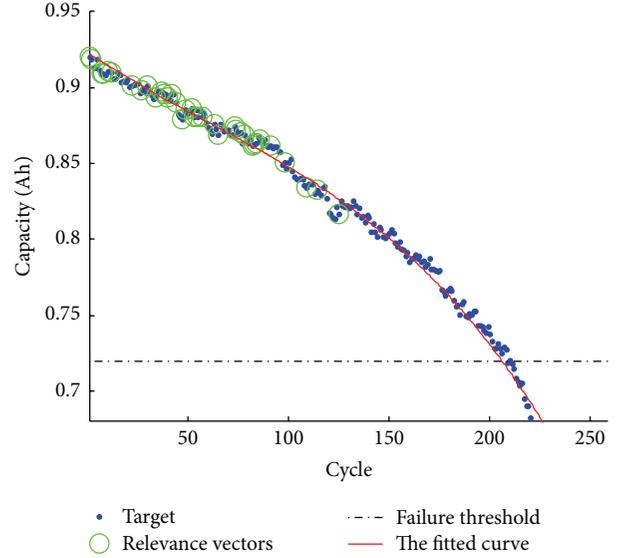


FIGURE 3: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 125 for lithium-ion battery A1.

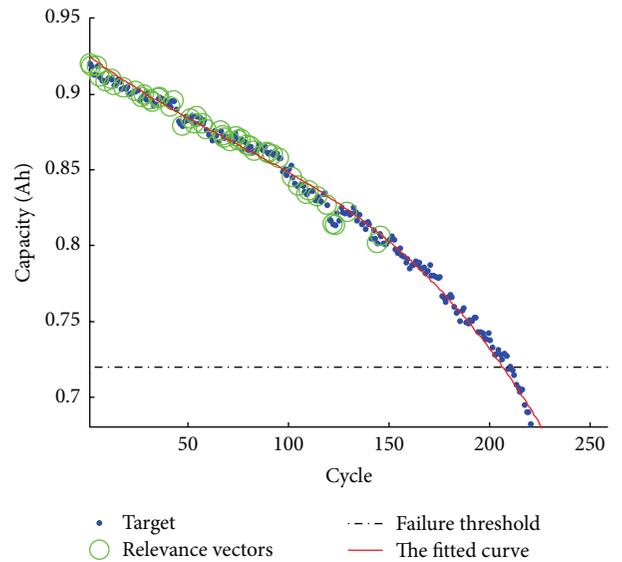


FIGURE 4: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 146 for lithium-ion battery A1.

were chosen for battery A2. In the third case, a lithium-ion battery A3 is investigated. Battery A3 capacity degradation data, plotted with the dots in Figures 11–14, are analyzed by the proposed hybrid prognostic approach. Predictions of the RUL of the lithium-ion battery A3 are calculated at the inspection cycles of 79, 92, 106, and 119, respectively. Table 3 summarizes the predicted RUL and the actual RUL when the inspection cycles of 79, 92, 106, and 119 were chosen for battery A3. In the fourth case, a lithium-ion battery A4 is investigated. Battery A4 capacity degradation data, plotted with the dots in Figures 15–18, are analyzed by the proposed hybrid prognostic approach. Predictions of the RUL of the lithium-ion battery A4 are calculated at the inspection cycles

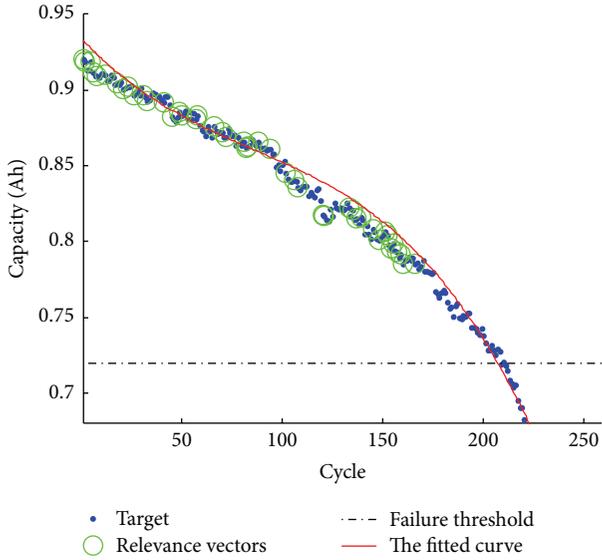


FIGURE 5: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 167 for lithium-ion battery A1.

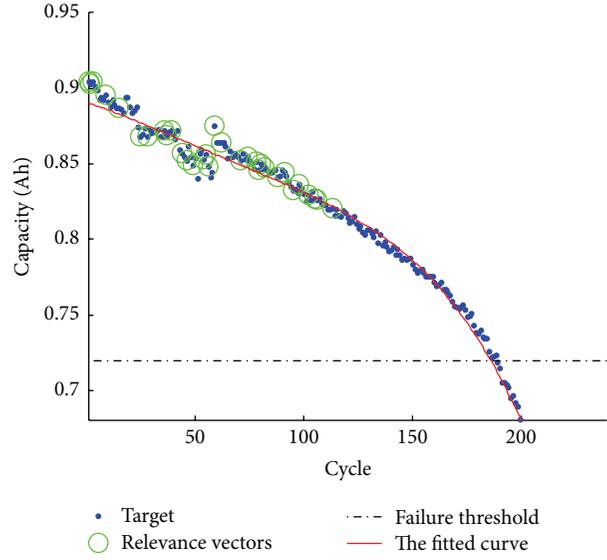


FIGURE 7: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 113 for lithium-ion battery A2.

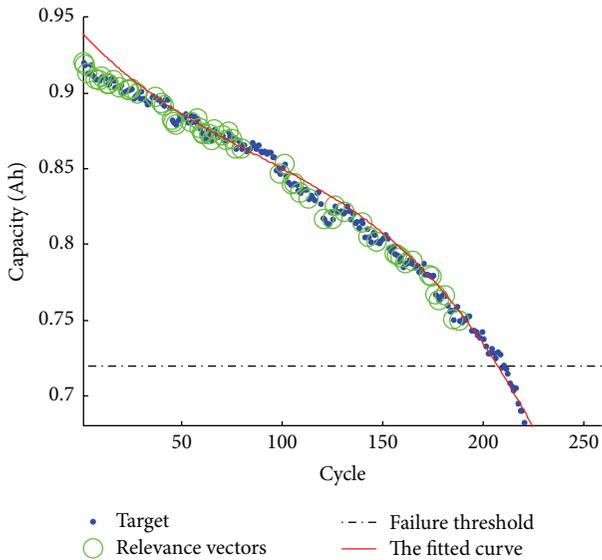


FIGURE 6: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 188 for lithium-ion battery A1.

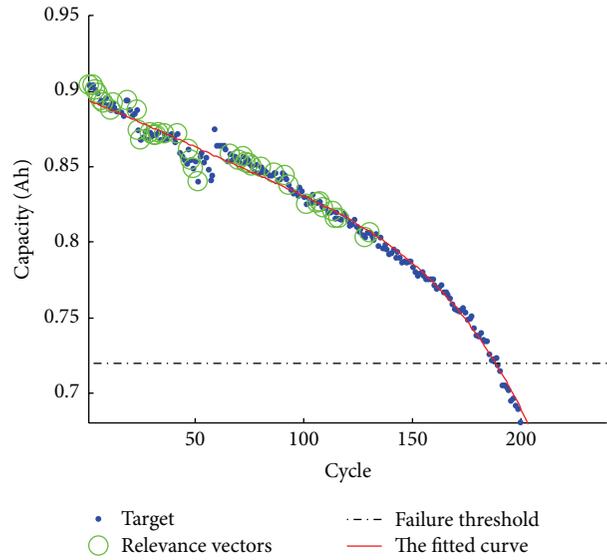


FIGURE 8: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 132 for lithium-ion battery A2.

of 29, 34, 38, and 43, respectively. Table 4 summarizes the predicted RUL and the actual RUL when the inspection cycles of 29, 34, 38, and 43 were chosen for battery A4. Note that the relevant vectors are highlighted by the circles in Figures 3–18. Also note that the parameter values are obtained by fitting the capacity degradation values predicted by the selective kernel ensemble-based RVM at the cycles of the representative training vectors. As seen in Figures 3–18, the proposed hybrid prognostic approach can effectively identify the lithium-ion battery capacity degradation trajectory, except in the third case of inspection number = 79 and the fourth case of inspection number = 29. The results in Tables 1–4 also revealed the good agreement of the predicted RUL and the

actual RUL. It can be concluded from Figures 3–18 and Tables 1–4 that the proposed hybrid prognostic approach may be a promising tool for lithium-ion battery RUL prediction.

### 5. Effects of RVM Learning on Hybrid Prognostic Approach

In order to investigate the effects of RVM learning on the proposed hybrid prognostic approach, comparison of the proposed hybrid prognostic approach using the selective kernel ensemble-based RVM learning algorithm with the hybrid prognostic approaches using the single kernel-based RVM learning algorithm and the Ensemble All-based RVM

TABLE 1: Values of the predicted RUL and the actual RUL for lithium-ion battery A1.

Inspection cycle ( $C_i$ )	Kernel mixed coefficient							Kernel parameters							Predicted RUL ( $C_i$ )	Actual RUL ( $C_i$ )	
	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\sigma_{\text{Gau}}$	$\sigma_{\text{Exp}}$	$\sigma_{\text{Lap}}$	$d$	$\xi$	$\eta$	$\sigma_{\text{cau}}$			$\tau$
125	0.4	N/A	N/A	0.4	0.2	N/A	N/A	5.56	N/A	N/A	0.98	109.2	8795	N/A	N/A	83	84
146	0.4	N/A	N/A	0.4	0.2	N/A	N/A	5.56	N/A	N/A	0.98	120.3	89.95	N/A	N/A	63	63
167	0.4	N/A	N/A	0.4	0.2	N/A	N/A	5.56	N/A	N/A	0.98	127.7	89.95	N/A	N/A	42	42
188	0.4	N/A	N/A	0.4	0.2	N/A	N/A	12.59	N/A	N/A	0.98	129.7	89.95	N/A	N/A	20	21

TABLE 2: Values of the predicted RUL and the actual RUL for lithium-ion battery A2.

Inspection cycle ( $C_i$ )	Kernel mixed coefficients							Kernel parameters							Predicted RUL ( $C_i$ )	Actual RUL ( $C_i$ )	
	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\sigma_{\text{Gau}}$	$\sigma_{\text{Exp}}$	$\sigma_{\text{Lap}}$	$d$	$\xi$	$\eta$	$\sigma_{\text{Cau}}$			$\tau$
113	0.6	N/A	N/A	0.3	0.1	N/A	N/A	2.85	N/A	N/A	9.75	100.28	81.43	N/A	N/A	74	76
132	0.6	N/A	N/A	0.3	0.1	N/A	N/A	2.98	N/A	N/A	9.78	70.28	81.43	N/A	N/A	58	57
151	0.6	N/A	N/A	0.3	0.1	N/A	N/A	2.98	N/A	N/A	8.79	70.28	71.43	N/A	N/A	39	38
170	0.6	N/A	N/A	0.3	0.1	N/A	N/A	2.98	N/A	N/A	8.79	129.7	89.95	N/A	N/A	20	19

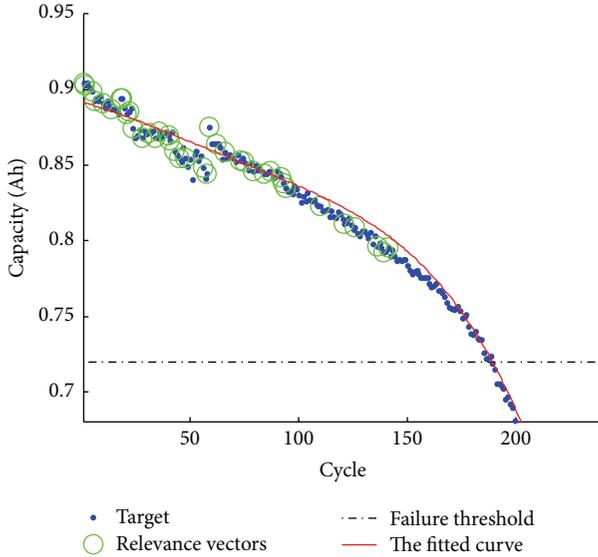


FIGURE 9: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 151 for lithium-ion battery A2.

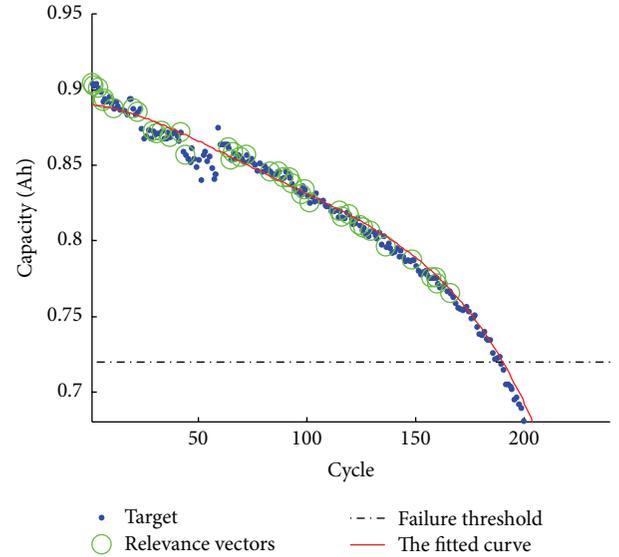


FIGURE 10: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 170 for lithium-ion battery A2.

learning algorithm was made in this section. It should be noted that herein the best kernel-based RVM learning algorithm means that the RVM only adopts the best performing component kernel among all available basic kernels for supervised learning, while the Ensemble All-based RVM learning algorithm means that the RVM adopts all of those available basic kernels for supervised learning. The results are summarized in Table 5 in terms of the predicted RUL and actual RUL. As can be seen, the proposed hybrid prognostic approach using the selective kernel ensemble-based RVM learning algorithm performed better (and in most cases substantially better) than the compared hybrid prognostic approach using the best kernel-based RVM learning algorithm. It can also be concluded from Table 5 that the proposed hybrid prognostic approach outperformed the

compared hybrid prognostic approach using the Ensemble All-based RVM learning algorithm, even though selective kernel ensemble-based RVM learning algorithm only uses a far smaller number of basic kernels. Taking the lithium-ion battery A1, for example, the size of the selective kernel ensemble-based RVM learning algorithm is about only 43% (3.0/7.0), 43% (3.0/7.0), 43% (3.0/7.0), and 43% (3.0/7.0) of the size of the Ensemble All-based RVM learning algorithm for four inspection cycles of 125, 146, 167, and 188, respectively. Significant improvement obtained not only demonstrates better generalization performance of selective kernel ensemble-based RVM learning algorithm but also proves the feasibility and necessity of removing redundant basic kernels in Ensemble All-based RVM learning algorithm. Therefore, the step of eliminating some unacceptable basic kernels and

TABLE 3: Values of the predicted RUL and the actual RUL for lithium-ion battery A3.

Inspection cycle ( $C_i$ )	Kernel mixed coefficient							Kernel parameters							Predicted RUL ( $C_i$ )	Actual RUL ( $C_i$ )	
	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\sigma_{\text{Gau}}$	$\sigma_{\text{Exp}}$	$\sigma_{\text{Lap}}$	$d$	$\xi$	$\eta$	$\sigma_{\text{Cau}}$			$\tau$
79	N/A	0.8	N/A	N/A	0.1	N/A	0.1	N/A	100.4	N/A	N/A	3.457	2.178	N/A	9.866	46	53
92	N/A	0.8	N/A	N/A	0.1	N/A	0.1	N/A	1.33	N/A	N/A	3.9209	3.987	N/A	59.91	38	40
106	0.67	0.165	N/A	N/A	0.165	N/A	N/A	0.1003	1.1895	N/A	N/A	150.2544	108.6072	N/A	N/A	24	26
119	0.67	0.165	N/A	N/A	0.165	N/A	N/A	0.10	0.1081	N/A	N/A	110.254	108.608	N/A	N/A	13	13

TABLE 4: Values of the predicted RUL and the actual RUL for lithium-ion battery A4.

Inspection cycle ( $C_i$ )	Kernel mixed coefficient							Kernel parameters							Predicted RUL ( $C_i$ )	Actual RUL ( $C_i$ )	
	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\sigma_{\text{Gau}}$	$\sigma_{\text{Exp}}$	$\sigma_{\text{Lap}}$	$d$	$\xi$	$\eta$	$\sigma_{\text{Cau}}$			$\tau$
29	0.8	N/A	N/A	0.1	0.1	N/A	N/A	1.79	N/A	N/A	1.29	0.58	2.178	N/A	N/A	16	19
34	0.8	N/A	N/A	0.1	0.1	N/A	N/A	0.42	N/A	N/A	1.68	0.59	0.84	N/A	N/A	14	14
38	0.8	N/A	N/A	0.1	0.1	N/A	N/A	0.45	N/A	N/A	0.48	0.59	0.84	N/A	N/A	10	10
43	0.8	N/A	N/A	0.1	0.1	N/A	N/A	0.21	N/A	N/A	0.58	0.49	0.53	N/A	N/A	5	5

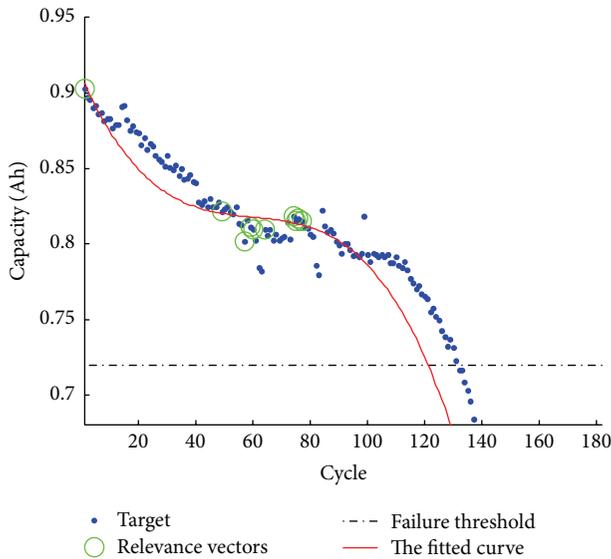


FIGURE 11: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 79 for lithium-ion battery A3.

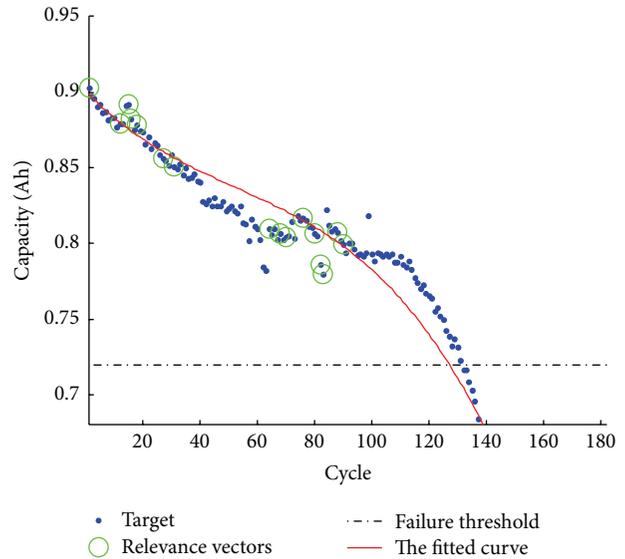


FIGURE 12: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 92 for lithium-ion battery A3.

meanwhile selecting several acceptable ones to construct a basic kernel combination plays a crucial role in enhancing the generalization capability of RVM.

## 6. Comparison with Existing Approaches

In this section, performances of the proposed hybrid prognostic approach using the selective kernel ensemble-based RVM learning algorithm were compared with those of the hybrid prognostic approaches using the popular algorithms of feedforward ANNs like the conventional BP algorithm and SVMs on four lithium-ion batteries A1, A2, A3, and A4. Although there are many variants of BP algorithm, a faster BP algorithm called Levenberg-Marquardt algorithm is used in ANNs. The activation function used is a simple sigmoidal

function  $g(x) = 1/(1 + \exp(-x))$ , whereas the kernel function used in SVMs is Gaussian.

Table 6 presents the results in the test set for the hybrid prognostic approaches using the selective kernel ensemble-based RVM learning algorithm, ANN, and SVM. We show the RUL prediction result. Also, we include the number of relevant/support vectors in the learning algorithms, which is related to prognostic model structural complexity (sparsity). As seen in Table 6, the proposed hybrid prognostic approach using the selective kernel ensemble-based RVM learning algorithm can provide more accurate RUL prediction of degraded lithium-ion batteries than the hybrid prognostic approaches using ANN and SVM can. This is mainly because (1) the selective kernel ensemble-based RVM learning algorithm can overcome some shortcomings of the ANN, such

TABLE 5: Comparison of the hybrid prognostic approach's performances using different RVM learning algorithms.

Battery name	Inspection cycle ( $C_i$ )	Predicted RUL ( $C_i$ )			Actual RUL ( $C_i$ )
		SKE	BK	EA	
A1	125	83	88	90	84
	146	63	60	67	63
	167	42	43	46	42
	188	20	23	25	21
A2	113	74	81	78	76
	132	58	61	60	57
	151	39	42	44	38
	170	20	23	24	19
A3	79	46	48	57	53
	92	38	36	35	40
	106	24	23	30	26
	119	13	10	16	13
A4	29	16	14	15	19
	34	14	16	12	14
	38	10	12	10	10
	43	5	4	6	5

SKE: the proposed hybrid prognostic approach that integrates selective kernel ensemble-based RVM and exponential regression; BK: another hybrid prognostic approach that integrates the best kernel-based (i.e., the best performing component kernel among all available basic kernels) RVM with exponential regression; EA: the other hybrid prognostic approach that integrates the Ensemble All-based (i.e., combining all of those available basic kernels) RVM with exponential regression.

TABLE 6: Comparison of the predicted RUL and sparse results for four lithium-ion batteries.

Battery name	Inspection cycle ( $C_i$ )	Predicted RUL ( $C_i$ )			Actual RUL ( $C_i$ )	Relevant/support vectors		
		SKE-RVM	SVM	BP		SKE-RVM	SVM	BP
A1	125	83	89	96	84	42	54	—
	146	63	66	71	63	46	62	—
	167	42	47	51	42	61	70	—
	188	20	25	27	21	60	75	—
A2	113	74	83	90	76	30	51	—
	132	58	61	66	57	40	55	—
	151	39	41	44	38	61	83	—
	170	20	25	27	19	52	86	—
A3	79	46	61	66	53	11	32	—
	92	38	45	48	40	23	38	—
	106	24	22	24	26	4	17	—
	119	13	15	16	13	4	15	—
A4	29	16	14	15	19	5	12	—
	34	14	16	18	14	4	14	—
	38	10	8	13	10	4	15	—
	43	5	6	7	5	6	18	—

SKE-RVM: selective kernel ensemble-based RVM learning algorithm.

as overfitting and local minima; (2) the selective kernel ensemble-based RVM learning algorithm used the selective convex combination and thus the RVM has stronger generalization capability; (3) the evolution of kernel parameters and kernel weights via CPSO can improve the generalization performance of the selective kernel ensemble-based RVM learning algorithm. Moreover, the selective kernel ensemble-based RVM learning algorithm also provides the capability

of yielding a decision function that is much sparser than SVM; that is, the number of relevant vectors is much smaller than that of support vectors, while maintaining the prediction accuracy. This can lead to a significant reduction in prognostic model structural complexity, thereby making it more suitable for on-line real-time RUL prediction. In addition, RVM does not need the tuning of a soft margin parameter ( $C$ ) as in SVM during the training phase.

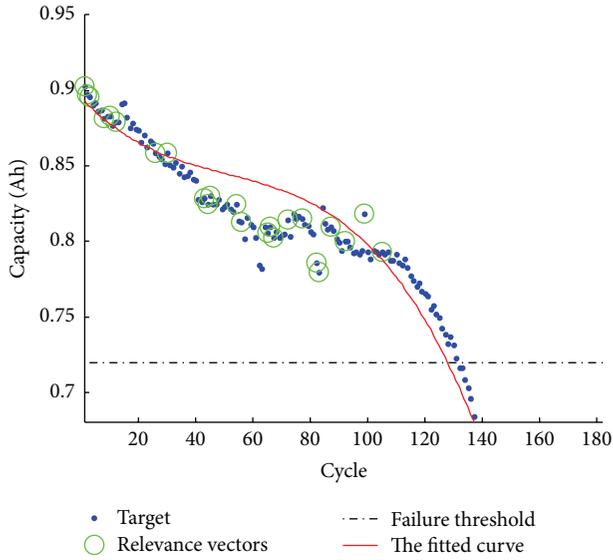


FIGURE 13: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 106 for lithium-ion battery A3.

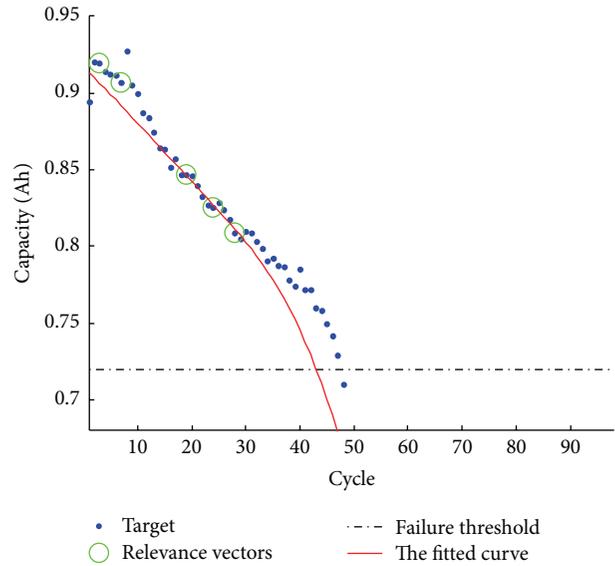


FIGURE 15: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 29 for lithium-ion battery A4.

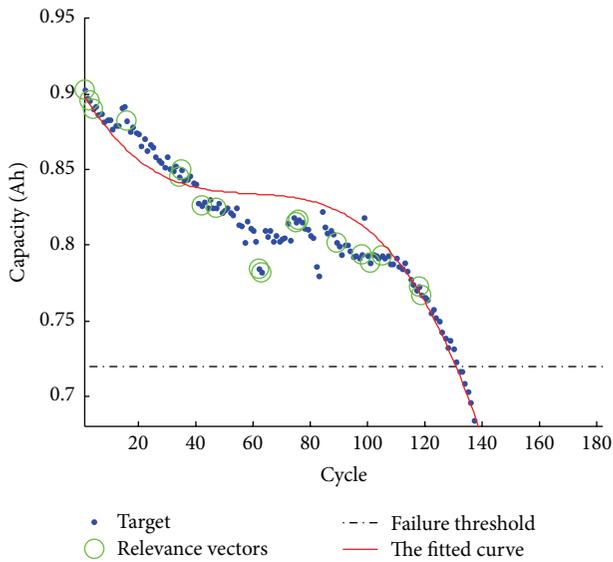


FIGURE 14: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 119 for lithium-ion battery A3.

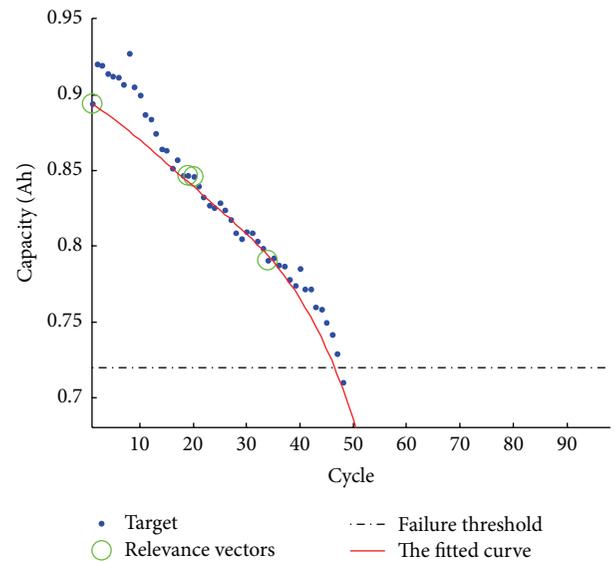


FIGURE 16: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 34 for lithium-ion battery A4.

### 7. Conclusions

Lithium-ion battery is a core component of many systems and is critical to the systems' functional capabilities. Battery failure could lead to reduced performance, operational impairment, and even catastrophic failure, especially in aerospace systems. Therefore, RUL prediction of degraded lithium-ion batteries is very helpful for preventing long-term breakdown or catastrophic failure. Using selective kernel ensemble-based relevance vector machine and exponential regression together, an automatic, effective but simpler-to-use hybrid prognostic approach was proposed for RUL prediction of degraded lithium-ion batteries. Four lithium-ion batteries,

namely, A1, A2, A3, and A4, were considered in this study. The experimental result indicated that the proposed hybrid prognostic approach can accurately predict the RUL of degraded lithium-ion batteries. Empirical comparisons showed that the proposed hybrid prognostic approach performed better than the hybrid prognostic approaches using the popular algorithms of feedforward ANNs like the conventional BP algorithm and SVMs. This study also demonstrates that the hybrid prognostic approach using the selective kernel ensemble-based RVM learning algorithm outperformed the hybrid prognostic approaches using the single kernel-based RVM learning algorithm and the Ensemble All-based RVM learning algorithm.

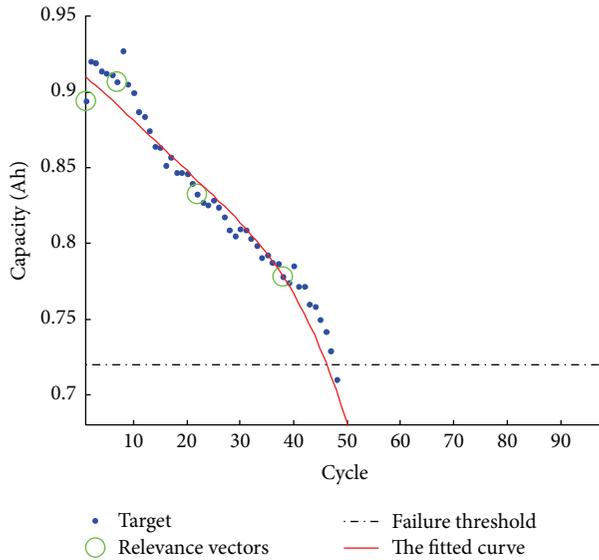


FIGURE 17: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 38 for lithium-ion battery A4.

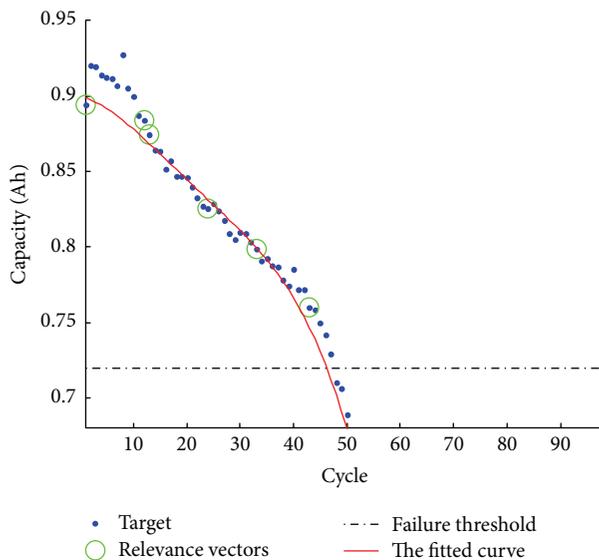


FIGURE 18: Predictive results obtained by the hybrid prognostic approach at inspection cycle of 43 for lithium-ion battery A4.

Three research directions are worth pursuing. First, although this study considers the application of lithium-ion batteries, the proposed hybrid prognostic approach can be modified and extended to other types of battery (e.g., li-polymer battery [33], lead-acid battery [34]). Second, bearings, gearboxes, and oil sand pumps are core components of all kinds of machinery, and maintenance of bearings, gearboxes, and oil sand pumps is essential. The proposed hybrid prognostic approach can also be extended to deal with gear, bearing, and oil sand pump performance degradation assessment and RUL estimation [35–38]. Third, in this study, only seven basic kernels were adopted. Hence, including other types of basic kernels (e.g., Wavelet kernel [39], Cauchy

kernel [40], and Rational Quadratic kernel [41]) in the RVM learning is another further research direction that is also worth pursuing.

## Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

## Acknowledgments

The research is funded partially by the National Science Foundation of China (51405239), National Defense Basic Scientific Research Program of China (A2620132010, A2520110003), Fundamental Research Funds for the Central Universities (1005-YAH15055), Jiangsu Provincial Natural Science Foundation of China (BK20150745, BK20140727), Jiangsu Province Science and Technology Support Program (BE2014134), and Jiangsu Postdoctoral Science Foundation of China (1501024C). The authors would like to express sincere appreciation to Professor Pecht and Center for Advanced Life Cycle Engineering, University of Maryland, for their efforts to make battery dataset available and permission to use dataset.

## References

- [1] I.-S. Kim, "A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer," *IEEE Transactions on Power Electronics*, vol. 25, no. 4, pp. 1013–1022, 2010.
- [2] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method," *Journal of Power Sources*, vol. 196, no. 23, pp. 10314–10321, 2011.
- [3] B.-H. Seo, T. H. Nguyen, D.-C. Lee, K.-B. Lee, and J.-M. Kim, "Condition monitoring of lithium polymer batteries based on a sigma-point Kalman filter," *Journal of Power Electronics*, vol. 12, no. 5, pp. 778–786, 2012.
- [4] B. S. Bhangu, P. Bentley, D. A. Stone, and C. M. Bingham, "Non-linear observers for predicting state-of-charge and state-of-health of lead-acid batteries for hybrid-electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 3, pp. 783–794, 2005.
- [5] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries using extended Kalman filtering," in *Proceedings of the International Microelectronics and Packaging Society (IMAPS) Advanced Technology Workshop on High Reliability Microelectronics for Military Applications*, pp. 17–19, Linthicum Heights, Md, USA, September 2011.
- [6] D. Andre, A. Nuhic, T. Soczka-Guth, and D. U. Sauer, "Comparative study of a structured neural network and an extended Kalman filter for state of health determination of lithium-ion batteries in hybrid electricvehicles," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 3, pp. 951–961, 2013.
- [7] B. Saha and K. Goebel, "Modeling Li-ion battery capacity depletion in a particle filtering framework," in *Proceedings of the Annual Conference of the Prognostics and Health Management Society (PHM '09)*, pp. 1–10, San Diego, Calif, USA, October 2009.
- [8] M. Dalal, J. Ma, and D. He, "Lithium-ion battery life prognostic health management system using particle filtering framework,"

- Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 225, no. 1, pp. 81–90, 2011.
- [9] Y. J. Xing, E. W. M. Ma, K.-L. Tsui, and M. Pecht, “An ensemble model for predicting the remaining useful performance of lithium-ion batteries,” *Microelectronics Reliability*, vol. 53, no. 6, pp. 811–820, 2013.
  - [10] Q. Miao, L. Xie, H. J. Cui, W. Liang, and M. Pecht, “Remaining useful life prediction of lithium-ion battery with unscented particle filter technique,” *Microelectronics Reliability*, vol. 53, no. 6, pp. 805–810, 2013.
  - [11] B. Long, W. Xian, L. Jiang, and Z. Liu, “An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries,” *Microelectronics Reliability*, vol. 53, no. 6, pp. 821–831, 2013.
  - [12] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, “Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods,” *Journal of Power Sources*, vol. 239, no. 1, pp. 680–688, 2013.
  - [13] C. Weng, Y. Cui, J. Sun, and H. Peng, “On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression,” *Journal of Power Sources*, vol. 235, pp. 36–44, 2013.
  - [14] C. C. Chen and M. Pecht, “Prognostics of lithium-ion batteries using model-based and data-driven methods,” in *Proceedings of the 3rd Annual IEEE Prognostics and System Health Management Conference (PHM '12)*, pp. 1–6, Beijing, China, May 2012.
  - [15] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
  - [16] B. Saha, K. Goebel, S. Poll, and J. Christophersen, “Prognostics methods for battery health monitoring using a Bayesian framework,” *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 2, pp. 291–296, 2009.
  - [17] F. Di Maio, K.-L. Tsui, and E. Zio, “Combining relevance vector machines and exponential regression for bearing residual life estimation,” *Mechanical Systems and Signal Processing*, vol. 31, pp. 405–427, 2012.
  - [18] E. Zio and F. Di Maio, “Fatigue crack growth estimation by relevance vector machine,” *Expert Systems with Applications*, vol. 39, no. 12, pp. 10681–10692, 2012.
  - [19] D. Wang, Q. Miao, and M. Pecht, “Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model,” *Journal of Power Sources*, vol. 239, pp. 253–264, 2013.
  - [20] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, “Remaining useful life prediction based on nonlinear feature reduction and support vector regression,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 7, pp. 1751–1760, 2013.
  - [21] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, “Health assessment and life prediction of cutting tools based on support vector regression,” *Journal of Intelligent Manufacturing*, vol. 26, no. 2, pp. 213–223, 2015.
  - [22] A. J. Smola, *Learning with kernels [Ph.D. thesis]*, TU Berlin, 1998.
  - [23] G. F. Smits and E. M. Jordaán, “Improved SVM regression using mixtures of kernels,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '02)*, vol. 3, pp. 2785–2790, IEEE, Honolulu, Hawaii, USA, May 2002.
  - [24] A. T. Quang, Q.-L. Zhang, and X. Li, “Evolving support vector machine parameters,” in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 548–551, Beijing, China, November 2002.
  - [25] M. Girolami and S. Rogers, “Hierarchic bayesian models for kernel learning,” in *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, pp. 241–248, August 2005.
  - [26] M. Gönen and E. Alpaydın, “Multiple kernel learning algorithms,” *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.
  - [27] S. R. Gunn and J. S. Kandola, “Structural modelling with sparse kernels,” *Machine Learning*, vol. 48, no. 1–3, pp. 137–163, 2002.
  - [28] M. Hu, Y. Chen, and J. T.-Y. Kwok, “Building sparse multiple-kernel SVM classifiers,” *IEEE Transactions on Neural Networks*, vol. 20, no. 5, pp. 827–839, 2009.
  - [29] J. Kennedy and R. C. Eberhart, “A discrete binary version of the particle swarm algorithm,” in *Proceedings of the IEEE International Conference on Computational Cybernetics and Simulation*, pp. 4104–4108, IEEE Press, Orlando, Fla, USA, October 1997.
  - [30] J. Kennedy and R. C. Eberhart, “Particle swarm optimization,” in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, IEEE, Perth, Australia, November–December 1995.
  - [31] K. Goebel, B. Saha, A. Saxena, J. R. Celaya, and J. P. Christophersen, “Prognostics in battery health management,” *IEEE Instrumentation and Measurement Magazine*, vol. 11, no. 4, pp. 33–40, 2008.
  - [32] M. Pecht, *CALCE Battery Data*, University of Maryland, College Park, Md, USA, 2014.
  - [33] B. Saha, E. Koshimoto, C. C. Quach et al., “Predicting battery life for electric UAVs,” in *Proceedings of the AIAA Infotech at Aerospace Conference and Exhibit*, St. Louis, Mo, USA, March 2011.
  - [34] E. Frisk, M. Krysander, and E. Larsson, “Data-driven lead-acid battery prognostics using random survival forests,” in *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, Fort Worth, Tex, USA, September 2014.
  - [35] D. Wang and P. W. Tse, “Prognostics of slurry pumps based on a moving-average wear degradation index and a general sequential Monte Carlo method,” *Mechanical Systems and Signal Processing*, vol. 56, pp. 213–229, 2014.
  - [36] D. Wang, Q. Miao, Q. H. Zhou, and G. W. Zhou, “An intelligent prognostic system for gear performance degradation assessment and remaining useful life estimation,” *Journal of Vibration and Acoustics*, vol. 137, no. 2, Article ID 021004, 2015.
  - [37] D. Wang and C. Shen, “An equivalent cyclic energy indicator for bearing performance degradation assessment,” *Journal of Vibration and Control*, 2014.
  - [38] D. Wang, P. W. Tse, W. Guo, and Q. Miao, “Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis,” *Measurement Science and Technology*, vol. 22, no. 2, Article ID 025102, 2011.
  - [39] L. Zhang, W. Zhou, and L. Jiao, “Wavelet support vector machine,” *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, vol. 34, no. 1, pp. 34–39, 2004.
  - [40] J. Basak, “A least square kernel machine with box constraints,” in *Proceeding of the IEEE 19th International Conference on Pattern Recognition (ICPR '08)*, pp. 1–4, IEEE, Tampa, Fla, USA, December 2008.
  - [41] D. K. Duvenaud, *Automatic model construction with Gaussian processes [Ph.D. thesis]*, University of Cambridge, 2014.

## Research Article

# Trace Ratio Criterion-Based Kernel Discriminant Analysis for Fault Diagnosis of Rolling Element Bearings Using Binary Immune Genetic Algorithm

Wen-An Yang,<sup>1</sup> Maohua Xiao,<sup>2</sup> Wei Zhou,<sup>3</sup> Yu Guo,<sup>1</sup> Wenhe Liao,<sup>1</sup> and Gang Shen<sup>4</sup>

<sup>1</sup>College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

<sup>2</sup>College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

<sup>3</sup>Nanjing Surveying and Mapping Instrument Factory, Nanjing 210003, China

<sup>4</sup>School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China

Correspondence should be addressed to Wen-An Yang; [dreamflow@nuaa.edu.cn](mailto:dreamflow@nuaa.edu.cn)

Received 3 July 2015; Revised 19 October 2015; Accepted 27 October 2015

Academic Editor: Dong Wang

Copyright © 2016 Wen-An Yang 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 rolling element bearing is a core component of many systems such as aircraft, train, steamboat, and machine tool, and their failure can lead to reduced capability, downtime, and even catastrophic breakdowns. Due to misoperation, manufacturing deficiencies, or the lack of monitoring and maintenance, it is often found to be the most unreliable component within these systems. Therefore, effective and efficient fault diagnosis of rolling element bearings has an important role in ensuring the continued safe and reliable operation of their host systems. This study presents a trace ratio criterion-based kernel discriminant analysis (TR-KDA) for fault diagnosis of rolling element bearings. The binary immune genetic algorithm (BIGA) is employed to solve the trace ratio problem in TR-KDA. The numerical results obtained using extensive simulation indicate that the proposed TR-KDA using BIGA (called TR-KDA-BIGA) can effectively and efficiently classify different classes of rolling element bearing data, while also providing the capability of real-time visualization that is very useful for the practitioners to monitor the health status of rolling element bearings. Empirical comparisons show that the proposed TR-KDA-BIGA performs better than existing methods in classifying different classes of rolling element bearing data. The proposed TR-KDA-BIGA may be a promising tool for fault diagnosis of rolling element bearings.

## 1. Introduction

The rolling element bearing is a core component of many systems such as aircraft, train, steamboat, and machine tool, and their failure can lead to reduced capability, downtime, and even catastrophic breakdowns [1–6]. Due to misoperation, manufacturing deficiencies, or the lack of monitoring and maintenance, it is often found to be the most unreliable component within these systems. Therefore, effective and efficient fault diagnosis of rolling element bearings has an important role in ensuring the continued safe and reliable operation of their host systems.

Over the past few years, much research effort has been devoted to developing approaches to fault diagnosis of rolling element bearings. When faults occur in rolling element

bearings, vibration signals in the relevant time/frequency-domain have been demonstrated to deviate from their normal ones because of the increased friction and impulsive forces [7–10]. Usually, several dozens or even hundreds of time/frequency-domain features are calculated from the bearing vibration signals to represent the different health status. In the current study, 9 time-domain features and 6 time-frequency-domain features are extracted from the bearing vibration signals to jointly construct a 15-dimension feature vector. In that way, fault diagnosis of rolling element bearings is usually solved as a high-dimensional pattern recognition problem. However, for high-dimensional data, the intrinsic dimension may be small. For example, the number of features responsible for a certain type of fault pattern may be small. Moreover, projection of high-dimensional data onto 2- or

3-dimension subspace can provide real-time visualization, which is convenient for the user to monitor the health status of rolling element bearings. In addition, projection of high-dimensional data onto low dimension subspace also plays a part of data compression, which is helpful for efficient storage and retrieval. Thus, dimensionality reduction techniques are often used to project the high-dimensional feature space to a lower-dimensional space while preserving most of “intrinsic information” contained in the data properties [11–15]. Upon performing dimensionality reduction on the data, its compact representation can be utilized for succeeding tasks (e.g., visualization and classification). Among various dimensionality reduction methods [16–24], principal component analysis (PCA) and linear discriminant analysis (LDA) are the two most common methods [21]. The former is an unsupervised method, which pursues the direction of maximum variance for optimal reconstruction. The latter is a supervised method, which aims to maximize the between-class scatter while minimizing the within-class scatter. Owing to the utilization of labeled information, the latter generally outperforms the former if sufficient labeled samples are provided [21]. In the past few years, a series of studies have been conducted to formulate the LDAs for pattern recognition by Fukunaga [21], Wang et al. [22], Sun and Chen [25], Guo et al. [26], Zhao et al. [27], Jin et al. [28], Jia et al. [29], and so on. Generally, the formulation of LDAs is based on the ratio trace criterion but not trace ratio criterion, because the ratio trace problem is more tractable than the trace ratio problem. Nevertheless, as pointed out by Wang et al. [22], solutions obtained based on ratio trace criterion may deviate from the original intent of the trace ratio problems. To improve the behaviour of LDA implementation, Wang et al. [22], Guo et al. [26], Zhao et al. [27], Jin et al. [28], and Jia et al. [29] presented various trace ratio criterion-based LDAs (TR-LDAs), in which the numerator and denominator of the criterion directly reflect Euclidean distances between of inter- and intraclass samples. Another advantage of trace ratio criterion is that the calculated projection matrix is orthogonal, which can eliminate the redundancy between different projection directions. In addition, the orthogonal projection can thus preserve such similarities without any change when using Euclidean distance to evaluate the similarity between data points [22]. Although the above TR-LDA formulation methods have the aforementioned advantages, they are criticized due to their incapability of dealing with the redundancy among eigenvectors. For example, if the most discriminative eigenvector is duplicated several times, the above TR-LDA formulation methods are prone to selecting all of them. This is problematic for selection of an optimal subset of eigenvectors because other discriminative and complementary eigenvectors will be missed. A classifier with the eigenvectors selected in this way can give rise to poor classification performance. Therefore, the issue of TR-LDA formulation has remained unresolved.

A review of the related literature also indicates that most of the previous work in the area of applying LDA or TR-LDA to fault diagnosis assumed that samples in each class follow a linear distribution. However, in many

fault diagnosis practices, samples in each class that may follow a nonlinear distribution cannot satisfy the assumption. Without this assumption, the separation of different classes may not be well characterized by the scatter matrices, causing the classification results to be degraded [21]. To solve this problem, kernel trick [30–32], which is to extend many linear methods to its nonlinear kernel version, can be used to extend TR-LDA to handle nonlinear problem. Thus, this study develops a nonlinear kernel version of TR-LDA, that is, trace ratio criterion-based kernel discriminant analysis (TR-KDA), for fault diagnosis of rolling element bearings. However, like many other TR-LDA models, the TR-KDA model presented in this study shares the trace ratio problem in the formulation of projection matrix. Although the above TR-LDA formulation methods have the aforementioned advantages, they are criticized due to the inability to handle redundancy in eigenvector selection. For example, if the most discriminative eigenvector is duplicated several times, the above TR-LDA formulation methods are prone to selecting all of them. This is problematic for selecting the best set of eigenvectors because other discriminative and complementary eigenvectors will be missed. A classifier with the eigenvectors selected in such a way can lead to a poor classification performance. Fortunately, immune genetic algorithm (IGA), a novel evolutionary computation technique developed by Jiao and Wang [33], has the potential to determine a set of discriminative and mutually irredundant eigenvectors. In this study, we propose a method called TR-KDA-BIGA that uses binary IGA (BIGA) to formulate TR-KDA for dimensionality reduction of statistical and wavelet features extracted from the vibration signals and gives rise to effective and efficient fault diagnosis of rolling element bearings. In particular the contributions are to

- (i) use immune evolutionary computation technique such as BIGA to obtain a reduced set of discriminative and mutually irredundant eigenvectors for TR-KDA-BIGA formulation,
- (ii) provide the capability of two-dimensional representation of bearing data that is very useful for the practitioners to monitor the health status of bearings,
- (iii) build a TR-KDA-BIGA model architecture for the vibration measurements for effective and efficient fault diagnosis of rolling element bearings.

The rest of this study is structured as follows. Section 2 briefly reviews the basic concepts of TR-LDA and kernel extension. Section 3 presents a TR-KDA-BIGA method. Section 4 discusses its convergence and initialization. Section 5 conducts performance evaluations of the proposed TR-KDA-BIGA on benchmark problems. Section 6 describes an overall flowchart of the proposed TR-KDA-BIGA for fault diagnosis of rolling element bearings. Section 7 summarizes the conclusions drawn from this study.

## 2. Review of TR-LDA and Kernel Extension

*2.1. Review of TR-LDA.* Suppose we are given a set of  $n$   $d$ -dimensional samples  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , belonging to  $l$  different classes. The goal of LDA tries to obtain a linear projection matrix  $\mathbf{W} \in \mathbf{R}^{d \times k}$  that can map the original  $d$ -dimensional data  $\mathbf{x}_i$  onto the  $k$ -dimensional data  $\mathbf{y}_i$  (usually  $k \ll d$ ) by maximizing the between-class scatter and meanwhile minimizing the within-class scatter. The between-class scatter matrix  $\mathbf{S}_B$  and the within-class scatter matrix  $\mathbf{S}_W$  are expressed as follows:

$$\begin{aligned} \mathbf{S}_B &= \sum_{i=1}^l n_i (\mathbf{m}^{(i)} - \mathbf{m})(\mathbf{m}^{(i)} - \mathbf{m})^T, \\ \mathbf{S}_W &= \sum_{i=1}^l \left( \sum_{j=1}^{n_i} (\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)})(\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)})^T \right), \end{aligned} \quad (1)$$

where  $\mathbf{m}$  represents the total sample mean vector,  $n_i$  represents the number of samples in the  $i$ th class,  $\mathbf{m}^{(i)}$  represents the average vector of the  $i$ th class, and  $\mathbf{x}_j^{(i)}$  represents the  $j$ th sample in the  $i$ th class. The new mapped feature vectors  $\mathbf{y}_i \in \mathbf{R}^k$  can then be expressed as  $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$ . The original LDA formulation, known as the Fisher LDA [21], only handles binary classification problems. However, many practical applications involve multiclass classification. In order to overcome this issue, a number of researchers have proposed optimization criteria for extending the Fisher LDA to handle multiclass classification problems. The first optimization criterion is in a ratio trace form (referred to as RT-LDA):

$$\mathbf{W}^* = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr} \left( \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}} \right), \quad (2)$$

where  $\text{Tr}[\cdot]$  denotes the matrix trace;  $\mathbf{I}$  is an identity matrix. In order to achieve a set of orthogonal normalized vectors, it usually adds the constraint  $\mathbf{W}^T \mathbf{W} = \mathbf{I}$  to (2). The second optimization criterion is in a trace ratio form (referred to as TR-LDA):

$$\mathbf{W}^* = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{Tr}(\mathbf{W}^T \mathbf{S}_B \mathbf{W})}{\text{Tr}(\mathbf{W}^T \mathbf{S}_W \mathbf{W})}. \quad (3)$$

The optimization problem in (3) can be solved directly through the generalized eigenvalue decomposition (GED) method [22]:

$$\mathbf{S}_B \mathbf{w}_i = \tau_i \mathbf{S}_W \mathbf{w}_i, \quad (4)$$

where  $\tau_i$  is the  $i$ th largest eigenvalue,  $\mathbf{w}_i$  is the eigenvector corresponding to  $\tau_i$ , and  $\mathbf{w}_i$  constitutes the  $i$ th column vector of the matrix  $\mathbf{W}$ . Although a closed-form solution for (3) can be approximately obtained with the GED, it does not necessarily guarantee best trace ratio optimization. Thus, this approximation of ratio trace optimization to trace ratio optimization may lead to classification capability loss of the derived optimal low-dimensional feature space. Moreover, the physical meaning of the trace ratio form is clearer than that of

the ratio trace form. However, the optimization problem in (3) is generally nonconvex and a closed-form solution for it does not exist. Fortunately, a recent study conducted by Guo et al. [26] showed that, using the trace difference function  $z(\lambda) = \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr}[\mathbf{W}^T (\mathbf{S}_B - \lambda \mathbf{S}_W) \mathbf{W}]$ , the trace ratio problem can be solved equivalently by finding zero points of the equation  $z(\lambda) = 0$ . Following up Guo et al.'s work, Wang et al. presented an iterative method named ITR algorithm to solve the trace ratio problem [22]. The ITR algorithm optimizes the objective function  $\lambda = \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr}[\mathbf{W}^T \mathbf{S}_B \mathbf{W}] / \text{Tr}[\mathbf{W}^T \mathbf{S}_W \mathbf{W}]$  in an iterative and incremental manner. The  $\mathbf{W}$  in the  $t$ th iteration step (referred to as  $\mathbf{W}_t$ ) is obtained through solving the trace difference problem  $\arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr}[\mathbf{W}^T (\mathbf{S}_B - \lambda_t \mathbf{S}_W) \mathbf{W}]$ , where  $\lambda_t$  represents the trace ratio value derived from the  $\mathbf{W}$  in the previous iteration step (referred to as  $\mathbf{W}_{t-1}$ ). However, the initialization for the  $\mathbf{W}$  influences substantially the convergence performance of the ITR algorithm. A good initialization can generally make the ITR algorithm yield a quick convergence. A bad initialization usually increases the number of iterations. Moreover, in ITR algorithm, although it seems that the  $\mathbf{W}$  formed with these eigenvectors corresponding to the  $k$  largest eigenvalues of  $\mathbf{S}_B - \lambda_t \mathbf{S}_W$  can maximize the trace difference  $\text{Tr}[\mathbf{W}^T (\mathbf{S}_B - \lambda_t \mathbf{S}_W) \mathbf{W}]$ , it cannot necessarily maximize the trace ratio  $\text{Tr}[\mathbf{W}^T \mathbf{S}_B \mathbf{W}] / \text{Tr}[\mathbf{W}^T \mathbf{S}_W \mathbf{W}]$ . On the other hand, from the perspective of fault diagnosis, the aim is mainly to find a set of projection vectors that can pose the highest levels of discrimination in the different fault patterns. Thus, these eigenvectors with the largest eigenvalues are not necessarily representative for discriminating one class from others as previously mentioned in Section 1. To overcome the above shortcomings, this study presents a BIGA-based solution method for trace ratio criterion (to be in detail discussed in Section 3).

*2.2. Kernel Extension.* In some applications, it is insufficient to model the data using the TR-LDA, which is a linear discriminating method. To address the issue of nonlinearities in the data, this section presents a nonlinear discriminating method using kernel trick [30–32], that is, TR-KDA. The so-called kernel trick is to map the original data to a high-dimensional Hilbert space through a nonlinear mapping function  $\phi$ . Let  $\phi(\mathbf{X})$  denote the data matrix in the Hilbert space:  $\phi(\mathbf{X}) = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_n)]$ . The function form of the mapping does not need to be known since it is implicitly defined by the choice of kernel function  $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ , that is, the inner product in the kernel-induced feature space. The kernel function  $K$  may be any positive kernel satisfying Mercer's condition. Radial basis function (RBF) kernel function, one of the most popular kernel functions employed in various kernelled learning algorithms, is adopted in this study. Then, (3) in Hilbert space can be written as follows:

$$\mathbf{W}^{\phi*} = \arg \max_{(\mathbf{W}^{\phi})^T \mathbf{W}^{\phi} = \mathbf{I}} \frac{\text{Tr}((\mathbf{W}^{\phi})^T \mathbf{S}_B^{\phi} \mathbf{W}^{\phi})}{\text{Tr}((\mathbf{W}^{\phi})^T \mathbf{S}_W^{\phi} \mathbf{W}^{\phi})}, \quad (5)$$

where  $\mathbf{W}^{\phi}$ ,  $\mathbf{S}_B^{\phi}$ , and  $\mathbf{S}_W^{\phi}$  are the matrices in Hilbert space corresponding to  $\mathbf{W}$ ,  $\mathbf{S}_B$ , and  $\mathbf{S}_W$  in (3), respectively. Notably,

we can show that matrices  $\mathbf{S}_B$  and  $\mathbf{S}_W$  in (3) can be essentially expressed as  $\mathbf{S}_B = \mathbf{X}\mathbf{L}_B\mathbf{X}^T$  and  $\mathbf{S}_W = \mathbf{X}\mathbf{L}_W\mathbf{X}^T$  through simple manipulation, respectively.  $\mathbf{X}$  is the vector, where  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ . Matrices  $\mathbf{L}_B$  and  $\mathbf{L}_W$  are the graph Laplacian matrices [34] of the weighted undirected graphs reflecting the between-class and within-class relationship of the samples. Consider

$$\begin{aligned}
\mathbf{S}_B &= \sum_{i=1}^l n_i (\mathbf{m}^{(i)} - \mathbf{m}) (\mathbf{m}^{(i)} - \mathbf{m})^T \\
&= \left( \sum_{i=1}^l n_i \mathbf{m}^{(i)} (\mathbf{m}^{(i)})^T \right) - \mathbf{m} \left( \sum_{i=1}^l n_i (\mathbf{m}^{(i)})^T \right) \\
&\quad - \left( \sum_{i=1}^l n_i \mathbf{m}^{(i)} \right) \mathbf{m}^T + \left( \sum_{i=1}^l n_i \right) \mathbf{m} \mathbf{m}^T \\
&= \left( \sum_{i=1}^l \frac{1}{n_i} (\mathbf{x}_1^{(i)} + \mathbf{x}_2^{(i)} + \dots + \mathbf{x}_{n_i}^{(i)}) \right. \\
&\quad \cdot \left. (\mathbf{x}_1^{(i)} + \mathbf{x}_2^{(i)} + \dots + \mathbf{x}_{n_i}^{(i)})^T \right) - 2n\mathbf{m}\mathbf{m}^T + n\mathbf{m}\mathbf{m}^T \\
&= \left( \sum_{i=1}^l \sum_{j,q=1}^{n_i} \frac{1}{n_i} \mathbf{x}_j^{(i)} (\mathbf{x}_q^{(i)})^T \right) - n\mathbf{m}\mathbf{m}^T = \mathbf{X}\mathbf{G}\mathbf{X}^T \\
&\quad - n\mathbf{m}\mathbf{m}^T = \mathbf{X}\mathbf{G}\mathbf{X}^T - \mathbf{X} \left( \frac{1}{n} \mathbf{e}\mathbf{e}^T \right) \mathbf{X}^T = \mathbf{X} \left( \mathbf{G} - \frac{1}{n} \right. \\
&\quad \cdot \left. \mathbf{e}\mathbf{e}^T \right) \mathbf{X}^T,
\end{aligned} \tag{6}$$

where  $\mathbf{e} = [1, 1, \dots, 1]^T$  is an  $n$ -dimensional vector. We can simplify the above equation even further by defining that

$$(\mathbf{G})_{ij} = \begin{cases} \frac{1}{n_q} & \text{if samples } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the } q\text{th class} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$\mathbf{L}_B = \mathbf{G} - \frac{1}{n} \mathbf{e}\mathbf{e}^T. \tag{8}$$

Thus, we get

$$\mathbf{S}_B = \mathbf{X}\mathbf{L}_B\mathbf{X}^T. \tag{9}$$

Then, the matrix  $\mathbf{S}_W$  can similarly be computed as follows:

$$\begin{aligned}
\mathbf{S}_W &= \sum_{i=1}^l \left( \sum_{j=1}^{n_i} (\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)}) (\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)})^T \right) \\
&= \sum_{i=1}^l \left( \sum_{j=1}^{n_i} \mathbf{x}_j^{(i)} (\mathbf{x}_j^{(i)})^T - \mathbf{m}^{(i)} (\mathbf{x}_j^{(i)})^T - \mathbf{x}_j^{(i)} (\mathbf{m}^{(i)})^T \right. \\
&\quad \left. + \mathbf{m}^{(i)} (\mathbf{m}^{(i)})^T \right) = \sum_{i=1}^l \left( \sum_{j=1}^{n_i} \mathbf{x}_j^{(i)} (\mathbf{x}_j^{(i)})^T \right.
\end{aligned}$$

$$\begin{aligned}
&\quad \left. - n_i \mathbf{m}^{(i)} (\mathbf{m}^{(i)})^T \right) = \sum_{i=1}^l \left( \mathbf{X}_i \mathbf{X}_i^T - \frac{1}{n_i} (\mathbf{x}_1^{(i)} + \mathbf{x}_2^{(i)} \right. \\
&\quad \left. + \dots + \mathbf{x}_{n_i}^{(i)}) (\mathbf{x}_1^{(i)} + \mathbf{x}_2^{(i)} + \dots + \mathbf{x}_{n_i}^{(i)})^T \right) = \sum_{i=1}^l \left( \mathbf{X}_i \mathbf{X}_i^T \right. \\
&\quad \left. - \frac{1}{n_i} \mathbf{X}_i (\mathbf{e}_i \mathbf{e}_i^T) \mathbf{X}_i^T \right) = \sum_{i=1}^l \left( \mathbf{X}_i \left( \mathbf{I} - \frac{1}{n_i} \mathbf{e}_i \mathbf{e}_i^T \right) \mathbf{X}_i^T \right) \\
&= \sum_{i=1}^l \mathbf{X}_i \mathbf{L}_i \mathbf{X}_i^T,
\end{aligned} \tag{10}$$

where  $\mathbf{X}_i = [\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{n_i}^{(i)}]$  is a  $d \times n_i$  matrix,  $\mathbf{e}_i = [1, 1, \dots, 1]^T$  is an  $n_i$ -dimensional vector,  $\mathbf{I}$  is the identity matrix,  $\mathbf{L}_i = \mathbf{I} - 1/n_i \mathbf{e}_i \mathbf{e}_i^T$  is an  $n_i \times n_i$  matrix, and  $\mathbf{X}_i \mathbf{L}_i \mathbf{X}_i^T$  is the data covariance matrix of the  $i$ th class. Based on (7), the above equation can be simplified similarly by defining

$$\mathbf{L}_W = \mathbf{I} - \mathbf{G}. \tag{11}$$

Thus, we get

$$\mathbf{S}_W = \mathbf{X}\mathbf{L}_W\mathbf{X}^T. \tag{12}$$

Using the definitions in (9) and (12), (5) can be rewritten as follows:

$$\begin{aligned}
\mathbf{W}^{\phi^*} &= \arg \max_{(\mathbf{W}^{\phi})^T \mathbf{W}^{\phi} = \mathbf{I}} \frac{\text{Tr} \left( (\mathbf{W}^{\phi})^T (\phi(\mathbf{X}) \mathbf{L}_B (\phi(\mathbf{X}))^T) \mathbf{W}^{\phi} \right)}{\text{Tr} \left( (\mathbf{W}^{\phi})^T (\phi(\mathbf{X}) \mathbf{L}_W (\phi(\mathbf{X}))^T) \mathbf{W}^{\phi} \right)}. \tag{13}
\end{aligned}$$

In order to pursue the matrix  $\mathbf{W}^{\phi^*}$ , solving the above equation involves decomposition of  $\phi(\mathbf{X})$  into an orthogonal matrix  $\mathbf{Q}$  (satisfying  $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ ) and a right triangular matrix  $\mathbf{R}$  such that  $\mathbf{Q} * \mathbf{R} = \phi(\mathbf{X})$ . We have

$$(\phi(\mathbf{X}))^T \phi(\mathbf{X}) = \mathbf{R}^T \mathbf{R}. \tag{14}$$

Let us map  $\mathbf{W}^{\phi}$  into the span of  $\mathbf{Q}$ .  $\mathbf{Q}$  is currently an orthogonal basis of  $\phi(\mathbf{X})$ , so we have

$$\mathbf{W}^{\phi} = \mathbf{Q}\mathbf{V}^{\phi}, \tag{15}$$

where  $\mathbf{V}^{\phi} = \mathbf{R}^{k \times n}$  is an orthogonal matrix satisfying  $(\mathbf{V}^{\phi})^T \mathbf{V}^{\phi} = \mathbf{I}$ . Using the definitions in (14) and (15), (13) can be further rewritten as follows:

$$\mathbf{V}^{\phi^*} = \arg \max_{(\mathbf{V}^{\phi})^T \mathbf{V}^{\phi} = \mathbf{I}} \frac{\text{Tr} \left( (\mathbf{V}^{\phi})^T \mathbf{R} \mathbf{L}_B \mathbf{R}^T \mathbf{V}^{\phi} \right)}{\text{Tr} \left( (\mathbf{V}^{\phi})^T \mathbf{R} \mathbf{L}_W \mathbf{R}^T \mathbf{V}^{\phi} \right)}. \tag{16}$$

Let  $\tilde{\mathbf{S}}_B = \mathbf{R} \mathbf{L}_B \mathbf{R}^T$  and  $\tilde{\mathbf{S}}_W = \mathbf{R} \mathbf{L}_W \mathbf{R}^T$ ; then (16) can be further rewritten as follows:

$$\mathbf{V}^{\phi^*} = \arg \max_{(\mathbf{V}^{\phi})^T \mathbf{V}^{\phi} = \mathbf{I}} \frac{\text{Tr} \left( (\mathbf{V}^{\phi})^T \tilde{\mathbf{S}}_B \mathbf{V}^{\phi} \right)}{\text{Tr} \left( (\mathbf{V}^{\phi})^T \tilde{\mathbf{S}}_W \mathbf{V}^{\phi} \right)}. \tag{17}$$

After the matrix  $\mathbf{V}^{\phi*}$  is obtained with the BIGA-based solution method (to be in detail discussed in Section 3), the output points in the reduced data space can thus be expressed as

$$\mathbf{Y}^{\phi*} = (\mathbf{W}^{\phi*})^T \phi(\mathbf{X}) = (\mathbf{V}^{\phi*})^T \mathbf{Q}^T \mathbf{Q} \mathbf{R} = (\mathbf{V}^{\phi*})^T \mathbf{R}. \quad (18)$$

### 3. The Proposed TR-KDA-BIGA

As previously mentioned, construction of TR-KDA needs to select  $k$  out of  $d$  eigenvectors to form the matrix  $\mathbf{V}^{\phi}$  for dimensionality reduction. However, finding a subset of eigenvectors based on the trace ratio criterion is not an easy task since the space of possible subsets is very large especially when  $d$  is a large number. Thus, it is not impractical to use exhaustive search to find an optimal subset of  $k$  eigenvectors. Instead, in this study, the BIGA is utilized to select  $k$  out of  $d$  eigenvectors of  $\mathbf{V}_t^{\phi}$  as the bases for projection matrix formulation based on the trace ratio criterion such that the trace ratio value  $\lambda = \text{Tr}[(\mathbf{V}^{\phi})^T \tilde{\mathbf{S}}_B \mathbf{V}^{\phi}] / \text{Tr}[(\mathbf{V}^{\phi})^T \tilde{\mathbf{S}}_W \mathbf{V}^{\phi}]$  can be maximized. Immune genetic algorithm, originally developed by Jiao and Wang [33], is a novel genetic algorithm based on the biological immune theory, which combined the immune mechanism with the evolutionary mechanism. In what follows, further discussion of the proposed TR-KDA-BIGA is carried out.

**3.1. Chromosome Encoding.** Encoding a solution of a problem into a chromosome is an important issue when using BIGAs. In this study, every chromosome in a BIGA corresponds to a discrete binary selector  $\mathbf{u} = [u_1, u_2, \dots, u_d]$ , where each gene in the *chromosome* is “1,” indicating an eigenvector  $\mathbf{v}_i^{\phi}$  ( $i = 1, 2, \dots, d$ ) of  $\tilde{\mathbf{S}}_B - \lambda_t \tilde{\mathbf{S}}_W$  appearing in forming the projection matrix  $\mathbf{V}^{\phi}$  of the  $t$ th step, while “0” denotes its absence. Thus, the length of the chromosome is  $d$ .

**3.2. Genetic Operators.** Genetic operators give every chromosome the chance to become the fittest chromosome of its generation. If it is difficult to reach the target of trace ratio optimization, crossover and mutation may introduce degeneracy into generations of chromosomes.

**3.2.1. Crossover Operator.** Crossover operator in a BIGA is employed to generate two new children chromosomes based on two existing parent chromosomes selected from the current population in terms of a prespecified crossover rate. In this study, “one-point” crossover operator was adopted to randomly select a cut point to exchange the parts between the cut point and the end of the string of the parent chromosomes. Specifically, suppose that two parent chromosomes  $P_1$  and  $P_2$  selected randomly from the population are undergoing the crossover operation at a randomly selected crossover point  $g$  ( $1 \leq g \leq d$ ), where

$$\begin{aligned} P_1 &= (u_{i,1}, u_{i,2}, \dots, u_{i,d}), \\ P_2 &= (u_{j,1}, u_{j,2}, \dots, u_{j,d}). \end{aligned} \quad (19)$$

Consequently, the offspring is generated by one-point crossover on the genes of two parents selected randomly from the population. We can thus get the two offspring chromosomes  $C_1$  and  $C_2$ :

$$\begin{aligned} C_1 &= (u_{i,1}, u_{i,2}, \dots, u_{i,g-1}, u_{i,g}, u_{j,g+1}, \dots, u_{j,d}), \\ C_2 &= (u_{j,1}, u_{j,2}, \dots, u_{j,g-1}, u_{j,g}, u_{i,g+1}, \dots, u_{i,d}). \end{aligned} \quad (20)$$

However, the exchange procedure is not simply exchanging their genetic information between gene segments after the crossover points. We must keep the number of eigenvectors to be included in the subset equal to  $k$ . In this study, therefore, a simple but effective crossover operator strategy in this study is performed in order to ensure that the crossover operator does not change the total number of “1” genes in chromosomes.

Let

$$\begin{aligned} n_{P_1} &= \sum_{q=g+1}^d u_{i,q}, \\ n_{P_2} &= \sum_{q=g+1}^d u_{j,q}. \end{aligned} \quad (21)$$

When  $n_{P_1}$  is not equal to  $n_{P_2}$ , the following retention criterion will be conducted:

- (1) If  $n_{P_1}$  is larger than  $n_{P_2}$ , randomly select  $(n_{P_1} - n_{P_2})$  genes with “0-bit” from the current offspring chromosome  $C_1$  and reset these  $(n_{P_1} - n_{P_2})$  selected genes to “1-bit,” and then randomly select  $(n_{P_1} - n_{P_2})$  genes with “1-bit” from the current offspring chromosome  $C_2$  and reset these  $(n_{P_1} - n_{P_2})$  selected genes to “0-bit.”
- (2) If  $n_{P_1}$  is smaller than  $n_{P_2}$ , randomly select  $(n_{P_2} - n_{P_1})$  genes with “1-bit” from the current offspring chromosome  $C_1$  and reset these  $(n_{P_2} - n_{P_1})$  selected genes to “0-bit,” and then randomly select  $(n_{P_2} - n_{P_1})$  genes with “0-bit” from the current offspring chromosome  $C_2$  and reset these  $(n_{P_2} - n_{P_1})$  selected genes to “1-bit.”

**3.2.2. Mutation Operator.** Mutation operator in a BIGA is used primarily as a mechanism for maintaining diversity in the population. For each gene in a chromosome that is undergoing the mutation, a real-valued number is randomly selected within the range of  $[0, 1]$ . If the real-valued number is less than the prespecified mutation rate, then the gene will change from “0-bit” to “1-bit” and vice versa. Upon adding (or removing) one eigenvector in that way, we shall randomly remove (or add) a different one such that the number of eigenvectors to be included in the subset is equal to  $k$ . The mutation operator helps the chromosomes to guide the search in new areas.

**3.3. Immune Operators.** The immune ability of BIGAs is realized through two kinds of immune operators: a vaccination and an immune selection. The vaccination is responsible for improving individuals’ overall fitness levels. The immune selection is responsible for prevention of deterioration.

**3.3.1. Vaccination Operator.** Given a chromosome  $\mathbf{u}$ , vaccination operation in a BIGA is employed to modify the genes on some bits according to a priori knowledge such that individuals with higher fitness have a greater probability of being selected. Let  $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_0})$  be a population; the vaccination operation on  $\mathbf{U}$  means that the operation is performed on  $n_\alpha = \alpha n_0$  chromosomes selected from  $\mathbf{U}$  according to the proportion of  $\alpha$ , where  $n_0$  represents the population size of a BIGA. A vaccine is abstracted from the prior knowledge of the pending problem, whose information amount and validity play an important role in the performance of the algorithm.

**3.3.2. Immune Selection Operator.** The immune selection operation consists of the following two steps. The first step is the immunity test: if the fitness of a chromosome  $\mathbf{u}$  is smaller than that of the parent chromosome, which indicates that degeneration occurred during crossover and mutation, then the parent chromosome will be used for the next competition. The second step is the annealing selection [35]: a chromosome  $\mathbf{u}_i$  is selected from the current offspring population  $\mathbf{U}_k$  to join with the new parents with the probability as follows:

$$P(\mathbf{u}_i) = \frac{\exp(f(\mathbf{u}_i)/T_k)}{\sum_{i=1}^{n_0} \exp(f(\mathbf{u}_i)/T_k)}, \quad (22)$$

where  $f(\mathbf{u}_i)$  is the fitness of the individual  $\mathbf{u}_i$  and  $\{T_k\}$  is the temperature-controlled series tending towards 0.

**3.4. Fitness Evaluation.** Fitness evaluation plays a critical role in selecting offspring chromosomes from the current population for the next generation. In this study, the fitness function for eigenvector selection is defined as

$$f(\mathbf{u}) = \frac{u_1 h_1 + u_2 h_2 + \dots + u_d h_d}{u_1 g_1 + u_2 g_2 + \dots + u_d g_d} = \frac{\mathbf{u}\mathbf{h}^T}{\mathbf{u}\mathbf{g}^T}, \quad (23)$$

where  $h_i$  denotes the  $\mathbf{v}_i^T \tilde{\mathbf{S}}_B \mathbf{v}_i$  value for the  $i$ th eigenvector,  $g_i$  denotes the  $\mathbf{v}_i^T \tilde{\mathbf{S}}_W \mathbf{v}_i$  value for the  $i$ th eigenvector,  $\mathbf{h} = [h_1, h_2, \dots, h_d]$ ,  $\mathbf{g} = [g_1, g_2, \dots, g_d]$ ,  $\mathbf{u} = [u_1, u_2, \dots, u_d]$ ,  $u_i \in \{0, 1\}$ ,  $\mathbf{u}\mathbf{1}^T = k$ , and  $i = 1, 2, \dots, d$ . Notably,  $\mathbf{u}$  is called the binary selector and  $k$  is the desired lower feature dimension. Finally, according to the evolved binary selector  $\mathbf{u}$ , we can thus form the projection matrix  $\mathbf{V}^\phi$  of the  $t$ th step by choosing the  $k$  eigenvectors with  $u_i = 1$  ( $i = 1, 2, \dots, d$ ). The procedures of the proposed TR-KDA-BIGA are summarized in the procedures of the proposed TR-KDA-BIGA part. The computational flow of the BIGA obtained using the aforementioned genetic and immune operators is also provided in the computational flow of the BIGA part.

*The Procedures of the Proposed TR-KDA-BIGA.* The procedures are as follows:

- (1) Construct the kernel matrix  $\mathbf{K} = (\phi(\mathbf{X}))^T \phi(\mathbf{X})$ .
- (2) Perform Cholesky decomposition to the kernel matrix  $\mathbf{K} = \mathbf{R}^T \mathbf{R}$ .
- (3) Form the kernel scatter matrixes as  $\tilde{\mathbf{S}}_B = \mathbf{R}\mathbf{L}_B \mathbf{R}^T$  and  $\tilde{\mathbf{S}}_W = \mathbf{R}\mathbf{L}_W \mathbf{R}^T$ .

- (4) Set iterations number  $t$  to 1.
- (5) Set the initial trace ratio value  $\lambda_t$  to  $\text{Tr}(\tilde{\mathbf{S}}_B)/\text{Tr}(\tilde{\mathbf{S}}_W)$ .
- (6) Compute the eigendecomposition of  $\tilde{\mathbf{S}}_B - \lambda_t \tilde{\mathbf{S}}_W$  as  $(\tilde{\mathbf{S}}_B - \lambda_t \tilde{\mathbf{S}}_W) \mathbf{v}_i^\phi = \tau_i \mathbf{v}_i^\phi$ , where  $\mathbf{v}_i^\phi$  ( $i = 1, 2, \dots, d$ ) is the eigenvector of  $\tilde{\mathbf{S}}_B - \lambda_t \tilde{\mathbf{S}}_W$ .
- (7) Calculate  $h_i = (\mathbf{v}_i^\phi)^T \tilde{\mathbf{S}}_B \mathbf{v}_i^\phi$  and  $g_i = (\mathbf{v}_i^\phi)^T \tilde{\mathbf{S}}_W \mathbf{v}_i^\phi$  for each eigenvector  $\mathbf{v}_i^\phi$  ( $i = 1, 2, \dots, d$ ).
- (8) Generate a population of BIGA selectors.
- (9) Evolve the population where the fitness of a BIGA selector  $\mathbf{u}$  is measured as  $f(\mathbf{u}) = \mathbf{u}\mathbf{h}^T/\mathbf{u}\mathbf{g}^T$ .
- (10)  $\mathbf{u}^*$  is the evolved best BIGA selector.
- (11) Form the projection matrix  $\mathbf{V}_t^\phi$  by choosing the  $k$  eigenvectors  $\mathbf{v}_i^\phi$  with  $u_i^* = 1$  ( $i = 1, 2, \dots, d$ ).
- (12) Update the trace ratio value  $\lambda_{t+1} = \text{Tr}[(\mathbf{V}_t^\phi)^T \tilde{\mathbf{S}}_B \mathbf{V}_t^\phi] / \text{Tr}[(\mathbf{V}_t^\phi)^T \tilde{\mathbf{S}}_W \mathbf{V}_t^\phi]$ ,  $t = t + 1$ , and go to step (6). Repeat this procedure until a convergence condition was established when the trace ratio value does not increase in consecutive 5 iterations.
- (13) Output  $\mathbf{Y}^{\phi*} = (\mathbf{V}^{\phi*})^T \mathbf{R}$ .

*The Computational Flow of the BIGA.* The computational flow is as follows:

- (1) Set  $l$  (time of generation) to 1.
- (2) Initialize randomly the original population  $A_l$ .
- (3) Evaluate each chromosome in the original population  $A_l$ .
- (4) Abstract vaccines according to the prior knowledge.
- (5) Check for termination criteria. If the fixed number of generations is not reached or the optimal chromosome found thus far is not satisfied, then go to the next step. Otherwise, output the optimal chromosome as the final solutions for further decision-making.
- (6) Perform crossover operation on the  $A_l$  and then generate the population  $B_l$ .
- (7) Perform mutation operation on the  $B_l$  and then generate the population results  $C_l$ .
- (8) Perform vaccination operation on the  $C_l$  and then generate the population  $D_l$ .
- (9) Perform immune selection operation on the  $D_l$  and then generate the next generational population  $A_{l+1}$ . Go to step (3).

## 4. The Convergence of the Proposed TR-KDA-BIGA

In this section, we analyze the convergence of the proposed TR-LDA-BIGA. Before doing this task, it should be worth noting that the BIGA is convergent. It has been demonstrated by Jiao and Wang [33] that as long as enough iteration has

been completed, the immune genetic population converges towards the true optimum with probability one.

Recall the trace difference function

$$z(\lambda) = \max_{(\mathbf{V}^\phi)^\top \mathbf{V}^\phi = \mathbf{I}} \text{Tr} \left[ (\mathbf{V}^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda \tilde{\mathbf{S}}_W) \mathbf{V}^\phi \right]; \quad (24)$$

it follows that

$$\begin{aligned} z(\lambda_{t+1}) &= \max_{(\mathbf{V}^\phi)^\top \mathbf{V}^\phi = \mathbf{I}} \text{Tr} \left[ (\mathbf{V}^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}^\phi \right] \implies \\ z(\lambda_{t+1}) &= \text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_{t+1}^\phi \right]. \end{aligned} \quad (25)$$

Since  $\lambda_{t+1} = \text{Tr}[(\mathbf{V}_t^\phi)^\top \tilde{\mathbf{S}}_B \mathbf{V}_t^\phi] / \text{Tr}[(\mathbf{V}_t^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_t^\phi]$  as previously mentioned, we get

$$\text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right] = 0. \quad (26)$$

Consider the inequality

$$\begin{aligned} \text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_{t+1}^\phi \right] \\ \geq \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right], \end{aligned} \quad (27)$$

and the equation

$$\text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right] = 0, \quad (28)$$

and we have

$$\begin{aligned} z(\lambda_{t+1}) &\geq 0, \\ \text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top \tilde{\mathbf{S}}_B \mathbf{V}_{t+1}^\phi \right] - \lambda_{t+1} \text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_{t+1}^\phi \right] \\ &\geq 0. \end{aligned} \quad (29)$$

Consequently,

$$\begin{aligned} \frac{\text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top \tilde{\mathbf{S}}_B \mathbf{V}_{t+1}^\phi \right]}{\text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_{t+1}^\phi \right]} &\geq \lambda_{t+1} \implies \\ \lambda_{t+2} &\geq \lambda_{t+1}. \end{aligned} \quad (30)$$

Substituting the subscript  $t + 1$  by  $t$  yields

$$\lambda_{t+1} \geq \lambda_t. \quad (31)$$

So we obtain the following inequality which gives the first expression of convergence of the proposed TR-KDA-BIGA.

Further, suppose that  $\lambda^*$  is the optimal trace ratio value; it follows that

$$z(\lambda^*) = \text{Tr} \left[ (\mathbf{V}^{\phi*})^\top (\tilde{\mathbf{S}}_B - \lambda^* \tilde{\mathbf{S}}_W) \mathbf{V}^{\phi*} \right] = 0, \quad (32)$$

TABLE 1: Specification of benchmark problems.

Data set	#samples	#dim.	#class
Heart-statlog	270	13	2
Ionosphere	351	34	2
Iris	150	4	3
Wine	178	13	3
Waveform	5000	40	3
Balance	625	4	3
SCCTS	600	60	6

where  $\mathbf{V}^{\phi*}$  is the optimal projection matrix. We therefore have

$$\begin{aligned} z(\lambda^*) &= \text{Tr} \left[ (\mathbf{V}^{\phi*})^\top (\tilde{\mathbf{S}}_B - \lambda^* \tilde{\mathbf{S}}_W) \mathbf{V}^{\phi*} \right] \\ &\geq \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda^* \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right] \\ &= \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right] \\ &\quad + (\lambda_{t+1} - \lambda^*) \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_t^\phi \right] \\ &= z(\lambda_{t+1}) + (\lambda_{t+1} - \lambda^*) \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_t^\phi \right]. \end{aligned} \quad (33)$$

Consider  $z(\lambda^*) = 0$ ,  $z(\lambda_{t+1}) \geq 0$  and  $\tilde{\mathbf{S}}_W$  is semipositive definite; we have

$$\lambda_{t+1} - \lambda^* \leq 0. \quad (34)$$

So we obtain the following inequality which gives the second expression of convergence of the proposed TR-KDA-BIGA:

$$\lambda_{t+1} \leq \lambda^*. \quad (35)$$

We conclude therefore that, for a particular initial trace ratio value  $\lambda_t$ , the updated value  $\lambda_{t+1}$  can always satisfy (1)  $\lambda_{t+1} \geq \lambda_t$  and (2)  $\lambda_{t+1} \leq \lambda^*$ .

## 5. Performance Evaluation on Benchmark Problems

In order to extensively verify the performance of the proposed TR-KDA-BIGA, it is first tested on wide types of commonly used benchmark problems taken from the UCI machine learning repository and evaluated with the classification rate (i.e., the number of correctly identified training examples/total number of training examples) by comparison with other existing methods such as PCA, LDA, KPCA [30, 31], KDA [32], and TR-LDA. These data sets include Heart-statlog, Ionosphere, Iris, Wine, Waveform, Balance, and Synthetic Control Chart Time Series (SCCTS) data sets (Table 1), which are of small sizes, low dimensions, large sizes, and/or high dimensions. For comparative study, we randomly select 50% data points from each data set as training set and the rest of the data points as test set. All methods use training set

TABLE 2: Results of the classification rate for the benchmark problems (mean  $\pm$  derivation).

Data set	PCA	KPCA	LDA	KDA	TR-LDA	TR-KDA-BIGA
Heart-statlog	59.7 $\pm$ 4.6	60.0 $\pm$ 4.2	75.5 $\pm$ 3.3	74.2 $\pm$ 2.7	76.1 $\pm$ 3.1	75.8 $\pm$ 2.3
Ionosphere	74.4 $\pm$ 4.0	74.9 $\pm$ 3.8	74.7 $\pm$ 5.7	75.1 $\pm$ 4.7	75.2 $\pm$ 5.9	76.7 $\pm$ 4.5
Iris	91.0 $\pm$ 4.0	91.0 $\pm$ 4.2	91.0 $\pm$ 3.5	91.8 $\pm$ 3.0	92.3 $\pm$ 3.2	92.9 $\pm$ 3.6
Wine	68.3 $\pm$ 5.2	69.0 $\pm$ 4.9	77.7 $\pm$ 8.4	77.0 $\pm$ 6.9	78.4 $\pm$ 8.5	78.6 $\pm$ 6.3
Waveform	72.9 $\pm$ 1.7	73.1 $\pm$ 1.5	73.3 $\pm$ 1.7	73.0 $\pm$ 2.3	74.5 $\pm$ 1.3	74.9 $\pm$ 2.8
Balance	63.4 $\pm$ 4.2	66.4 $\pm$ 4.7	77.8 $\pm$ 4.9	78.4 $\pm$ 3.7	78.3 $\pm$ 4.2	79.2 $\pm$ 3.2
SCCTS	80.0 $\pm$ 4.8	80.1 $\pm$ 5.0	80.5 $\pm$ 5.4	81.1 $\pm$ 6.2	81.2 $\pm$ 5.6	82.9 $\pm$ 6.7

TABLE 3: Time-domain statistical features.

Feature	Eq.	Feature	Eq.
Standard deviation	$x_{\text{std}} = \sqrt{\frac{\sum_{i=1}^N (x(i) - \bar{x})^2}{N}}$	Clearance factor	$\text{CLF} = \frac{x_p}{x_{\text{rms}}}$
Peak	$x_p = \max  x(i) $	Shape factor	$\text{SF} = \frac{x_{\text{rms}}}{(1/N) \sum_{i=1}^N  x(i) }$
Skewness	$x_{\text{ske}} = \frac{(1/N) \sum_{i=1}^N (x(i) - \bar{x})^3}{x_{\text{std}}^3}$	Impact factor	$\text{IF} = \frac{x_p}{(1/N) \sum_{i=1}^N  x(i) }$
Kurtosis	$x_{\text{kur}} = \frac{(1/N) \sum_{i=1}^N (x(i) - \bar{x})^4}{x_{\text{std}}^4}$	Square mean root	$x_{\text{smr}} = \left( \frac{1}{N} \sum_{i=1}^N \sqrt{ x(i) } \right)^2$
Crest factor	$\text{CF} = \frac{x_p}{x_{\text{rms}}}$		

where  $x(i)$  is a digital signal series,  $i = 1, 2, \dots, N$ ,  $N$  is the number of elements of the digital signal, and  $\bar{x} = \sum_{i=1}^N x(i)/N$  and  $x_{\text{rms}} = \sqrt{\sum_{i=1}^N x(i)^2/N}$  are the mean value and root-mean-square value of the digital signal series, respectively.

in the output reduced space to train one nearest neighborhood (INN) classifier for evaluating the classification rate of test set. To restrict the influence of random effects, the experiments of PCA, LDA, KPCA, KDA, TR-LDA, and TR-KDA-BIGA compared on each benchmark problem are independently performed for 20 runs. Table 2 compares the classification rate for benchmark problems of the proposed TR-KDA-BIGA with that of the PCA, LDA, KPCA, KDA, and TR-LDA. As seen in Table 2, the proposed TR-KDA-BIGA can perform better than all the compared methods, except in the case of Heart-statlog.

The results obtained demonstrate the ability of the proposed TR-KDA-BIGA in classifying different classes well. Thus, the proposed TR-KDA-BIGA may be effectively employed for fault diagnosis of rolling element bearings.

## 6. The Proposed TR-KDA Using BIGA for Fault Diagnosis of Rolling Element Bearings

In this section, the proposed TR-KDA-BIGA is applied to fault diagnosis of rolling element bearings. Vibration signals resulting from rolling element bearings are first filtered by using a low-pass filter. Then, the filtered vibration signals are divided into sections of equal window length. One set of relevant features obtained from each window is used for characterizing to some extent the health status of the rolling element bearings. Most of the faults occurring in rolling element

bearings will introduce the increased friction and impulsive forces when bearings are rotating, which generally lead the vibration signals in time-domain, frequency-domain, and/or time-frequency-domain to vary (become different) from the normal ones. In this study, 9 time-domain statistical features (Table 3) are extracted from the vibration signal. All of these 9 time-domain statistical features reflect the characteristics of time series data in the time-domain. Moreover, 6 time-frequency-domain wavelet features about the percentages of energy corresponding to wavelet coefficients are extracted from the vibration signal by using Daubechies-4 (db4) wavelet to decompose the vibration signal into five levels [32]. Wavelet features extracted in such a way can to the greatest extent reflect the vibration energy distribution in the time-frequency-domain. Thus, 9 time-domain statistical features together with 6 time-frequency-domain wavelet features are used to represent each window's vibration signals.

*6.1. Experimental Setup.* In order to demonstrate the performance of the proposed TR-KDA-BIGA, rolling element bearing data obtained from the Bearing Data Centre, Case Western Reserve University [36], are used. The test rolling element bearings were SKF 6205 JEM, a type of deep groove ball bearing. Single-point faults were seeded into the drive end ball bearing using electrodischarge machining. Faults occurring in rolling element bearings introduced impact-like vibration signals when bearings were rotating. An accelerometer was mounted on the drive end of the motor housing

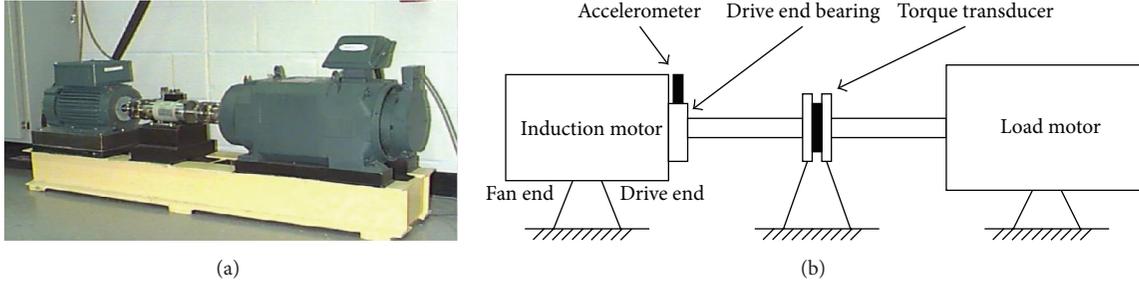


FIGURE 1: Experimental setup: (a) test rig; (b) schematic description of the test rig.

to detect such impacts that behaved like damped oscillations. Vibration signals were captured from four different health statuses of bearing, that is, normal bearings (Normal), inner race fault (IR), ball fault (BA), and outer race fault (OR). For each of the three abnormal statuses (IR, BA, and OR), there are three different levels of severity with fault diameters (0.007 inches, 0.014 inches, and 0.021 inches). All the experiments were done for three different load conditions (1 HP, 2 HP, and 3 HP). Figure 1 illustrates the experimental setup. Experimental data were collected from the drive end ball bearing of an induction motor (Reliance Electric 2 HP IQPreAlert) driven test rig. Table 4 gives a short description of rolling element bearing data.

## 6.2. Experiment Results

**6.2.1. Visualization of Bearing Data.** Visualization performances of the proposed TR-KDA-BIGA are compared with those of PCA, LDA, KPCA, KDA, and TR-LDA using simulations, where KPCA and KDA are the kernel extensions to PCA and LDA, respectively. The two-dimensional visualization results of bearing data for three different load conditions (1, 2, and 3 HP) obtained with PCA, LDA, KPCA, KDA, TR-LDA, and the proposed TR-KDA-BIGA are summarized in Figures 2, 3, and 4, respectively. As seen in Figures 2, 3, and 4, the proposed TR-KDA-BIGA outperforms all the compared methods in not only closely conglomerating bearing data belonging to the same class but also clearly separating bearing data belonging to different classes of three different load conditions (1, 2, and 3 HP). Compared with the unsupervised methods (i.e., PCA and KPCA), the supervised methods (i.e., LDA, KDA, TR-LDA, and TR-KDA-BIGA) can preserve more discriminative information embedded in bearing data and obtain clearer and less overlapped boundaries. It can also be concluded from Figures 2, 3, and 4 that the methods using kernel trick (i.e., KPCA, KDA, and TR-KDA-BIGA) performed better than the methods without using kernel trick (i.e., PCA, LDA, and TR-LDA) in separating the discriminative property—samples from different classes in the learned subspace.

**6.2.2. Classification of Bearing Data.** Classification performances of the proposed TR-KDA-BIGA are compared with those of PCA, LDA, KPCA, KDA, and TR-LDA. In order to show the robustness of the proposed TR-KDA-BIGA, we

TABLE 4: Description of rolling element bearing data.

#samples	#dimension	#class	#samples per class
800	15	10	80

perform 4 independent experiments for each load condition in terms of 4 different data partitions. In this study, 10, 20, 30, and 40 samples per class in bearing data set are randomly selected from each class in bearing data as the training set and the remaining samples as the test set. Then, each method uses the training set to train a INN classifier in order to classify different health status in test set. Tables 5, 6, and 7 summarize the average classification results of PCA, LDA, KPCA, KDA, TR-LDA, and the proposed TR-KDA-BIGA with various numbers of training samples for 1 HP, 2 HP, and 3 HP load conditions, respectively. It can be observed that the overall average performance of the classification of health status is fairly good. Tables 5, 6, and 7 demonstrate that the proposed TR-KDA-BIGA performs remarkably better than the compared methods (PCA, LDA, KPCA, KDA, and TR-LDA). It should be noted that the proposed model can also provide the capability of real-time visualization that is very useful for the practitioners to monitor the health status of rolling element bearings. Tables 5, 6, and 7 also demonstrate that the number of training samples does significantly affect the classification accuracy for bearing health status.

## 7. Conclusions

The rolling element bearing is a core component of many systems, and their failure can lead to reduced capability, downtime, and even catastrophic breakdowns. Effective and efficient fault diagnosis of rolling element bearings plays an extremely important role in the safe and reliable operation of their host systems. In the current study, fault diagnosis of rolling element bearings is done in a pattern recognition way by calculating a high-dimensional feature data set from vibration signals, which represents the different status of bearings. Specifically, the TR-KDA is presented for fault diagnosis of rolling element bearings and the BIGA is employed to solve the trace ratio problem in TR-KDA. The numerical results obtained using extensive simulation indicate that the proposed TR-KDA-BIGA can effectively

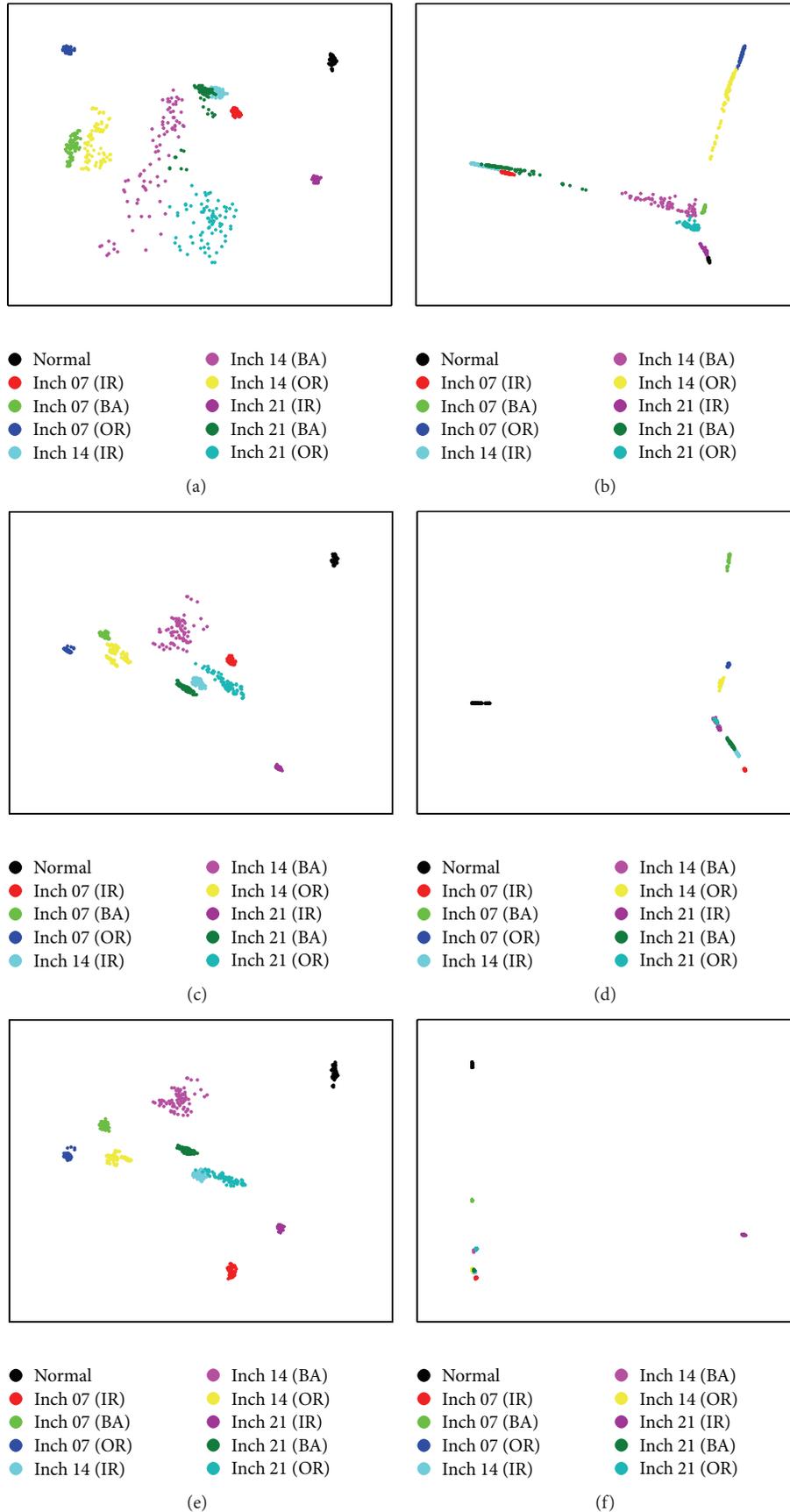


FIGURE 2: Two-dimensional representation of bearing data under 1 HP motor load by each method: (a) PCA; (b) KPCA; (c) LDA; (d) KDA; (e) TR-LDA; (f) TR-KDA-BIGA.

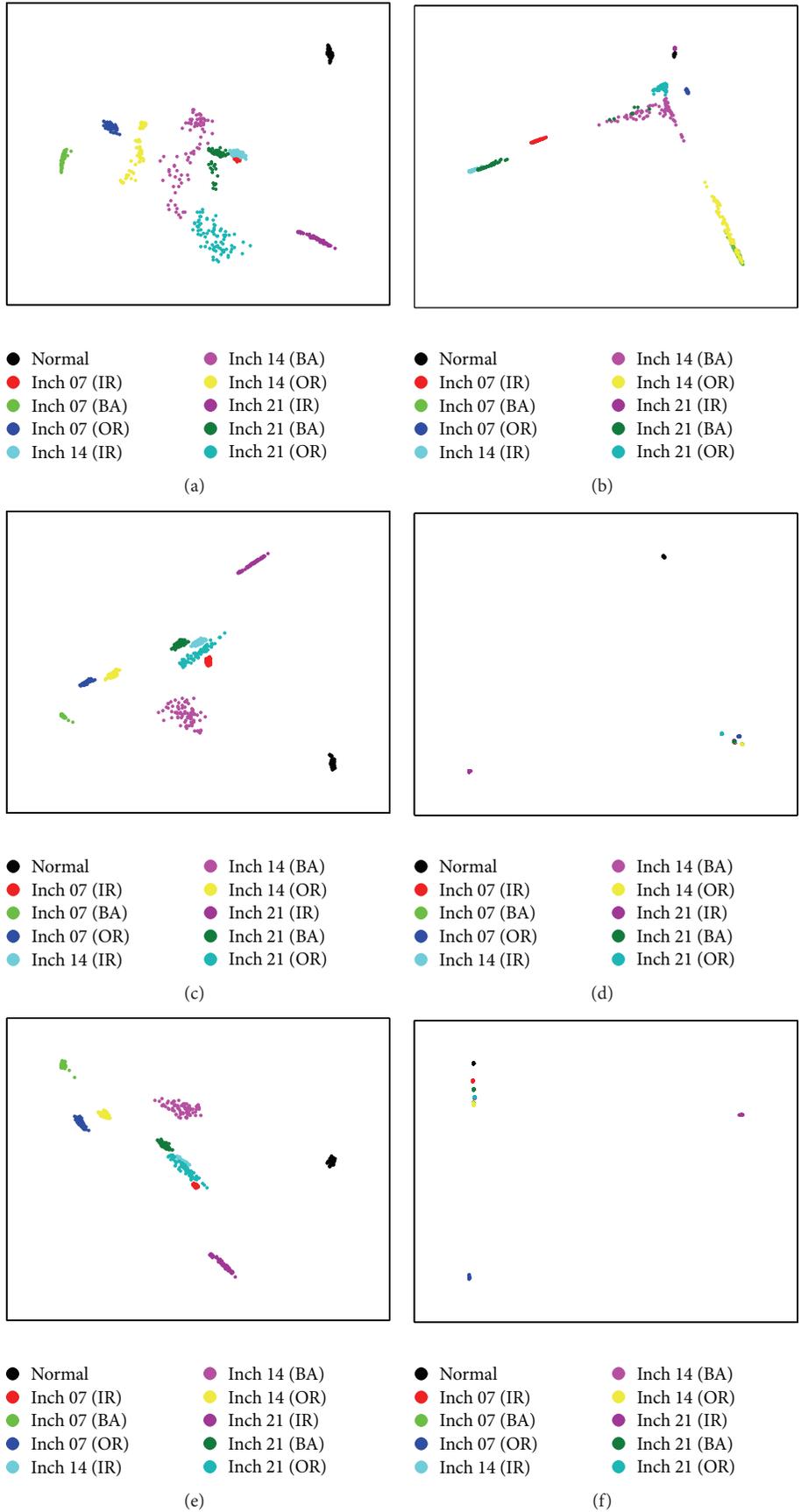


FIGURE 3: Two-dimensional representation of bearing data under 2 HP motor load by each method: (a) PCA; (b) KPCA; (c) LDA; (d) KDA; (e) TR-LDA; (f) TR-KDA-BIGA.

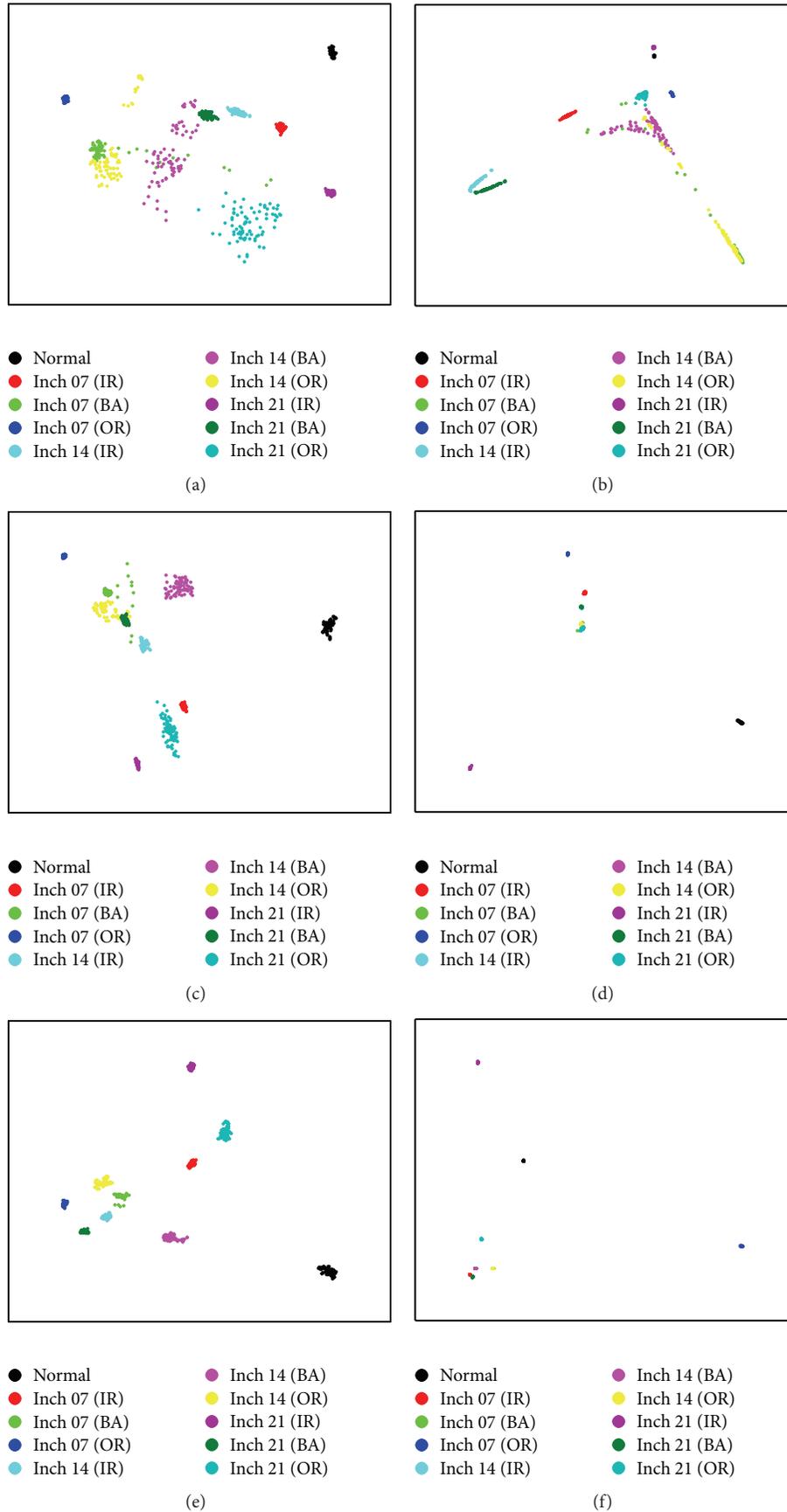


FIGURE 4: Two-dimensional representation of bearing data under 3 HP motor load by each method: (a) PCA; (b) KPCA; (c) LDA; (d) KDA; (e) TR-LDA; (f) TR-KDA-BIGA.

TABLE 5: Classification accuracy of bearing data under 1 HP motor load.

Method	Number of training samples				Average
	10	20	30	40	
PCA	94.7793	95.9016	96.7564	97.7785	96.30395
KPCA	94.9203	95.9007	96.5873	98.4867	96.47375
LDA	96.2741	97.1929	98.0017	98.8438	97.57813
KDA	96.7026	97.1957	98.1504	99.1511	97.79995
TR-LDA	97.1914	97.9510	99.0587	99.7006	98.47543
TR-KDA-BIGA	97.5793	98.8971	99.5416	99.8068	98.95620

TABLE 6: Classification accuracy of bearing data under 2 HP motor load.

Method	Number of training samples				Average
	10	20	30	40	
PCA	94.4654	95.4452	96.3670	97.6571	95.98368
KPCA	94.2155	96.5207	97.3084	98.7273	96.69298
LDA	96.0477	96.9640	97.6314	99.0035	97.41165
KDA	96.3333	97.5962	97.2495	99.4105	97.64738
TR-LDA	97.3188	98.0867	98.7036	99.7777	98.47170
TR-KDA-BIGA	98.9123	98.6209	99.7737	99.9311	99.30950

TABLE 7: Classification accuracy of bearing data under 3 HP motor load.

Method	Number of training samples				Average
	10	20	30	40	
PCA	96.1337	97.3051	97.6719	97.9738	97.27113
KPCA	96.7442	97.0369	97.6968	98.2854	97.44083
LDA	97.0421	98.5900	99.2875	99.4996	98.60480
KDA	97.4820	98.8408	99.1516	99.6082	98.77065
TR-LDA	98.8961	99.3457	99.9190	99.6082	99.44225
TR-KDA-BIGA	99.2918	99.9153	99.9103	99.9291	99.76163

classify different classes of rolling element bearing data, while also providing the capability of real-time visualization that is very useful for the practitioners to monitor the health status of rolling element bearings. Empirical comparisons show that the proposed TR-KDA-BIGA performs better than existing methods in classifying different rolling element bearing data. The proposed TR-KDA-BIGA may be a promising tool for fault diagnosis of rolling element bearings.

Three research directions are worth pursuing. First, although this study considers the specific fault diagnosis of rolling element bearings, the proposed method can be modified and extended to address the fault diagnosis of gearboxes [37, 38] and cutting tools [39, 40]. Second, frequency-domain information can be utilized for fault diagnosis of rolling element bearings [41, 42]; it would thus be interesting to integrate frequency-domain features to time-domain and time-frequency-domain features. Third, empirical mode decomposition is a very powerful tool for nonlinear and nonstationary signal processing [43–45]; it would be also interesting to employ the empirical mode decomposition to extract periodic components and random transient components from the bearing vibration signal mixture, which may be very helpful for extraction of fault signatures from a collected bearing vibration signal.

## Conflict of Interests

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

## Acknowledgments

The research is funded partially by the National Science Foundation of China (51405239), National Defense Basic Scientific Research Program of China (A2620132010, A2520110003), Jiangsu Provincial Natural Science Foundation of China (BK20150745, BK20140727), Jiangsu Province Science and Technology Support Program (BE2014134), Fundamental Research Funds for the Central Universities (1005-YAH15055), and Jiangsu Postdoctoral Science Foundation of China (1501024C). The authors would like to express sincere appreciation to Professor KA Loparo and Case Western Reserve University for their efforts to make bearing data set available and permission to use data set.

## References

- [1] X. Jin, E. W. M. Ma, L. L. Cheng, and M. Pecht, "Health monitoring of cooling fans based on mahalanobis distance with mRMR feature selection," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2222–2229, 2012.

- [2] X. Jin and T. W. S. Chow, "Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis-Taguchi system," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5787–5795, 2013.
- [3] J. Zarei, "Induction motors bearing fault detection using pattern recognition techniques," *Expert Systems with Applications*, vol. 39, no. 1, pp. 68–73, 2012.
- [4] J.-B. Yu, "Bearing performance degradation assessment using locality preserving projections," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7440–7450, 2011.
- [5] D. Wang, P. W. Tse, and Y. L. Tse, "A morphogram with the optimal selection of parameters used in morphological analysis for enhancing the ability in bearing fault diagnosis," *Measurement Science and Technology*, vol. 23, no. 6, Article ID 065001, 2012.
- [6] W. Wang and M. Pecht, "Economic analysis of canary-based prognostics and health management," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 7, pp. 3077–3089, 2011.
- [7] 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.
- [8] 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.
- [9] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4568–4579, 2010.
- [10] W. He, Z.-N. Jiang, and K. Feng, "Bearing fault detection based on optimal wavelet filter and sparse code shrinkage," *Measurement*, vol. 42, no. 7, pp. 1092–1102, 2009.
- [11] 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.
- [12] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [13] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3398–3407, 2013.
- [14] M. B. Zhao, Z. Zhang, and T. W. S. Chow, "Trace ratio criterion based generalized discriminative learning for semi-supervised dimensionality reduction," *Pattern Recognition*, vol. 45, no. 4, pp. 1482–1499, 2012.
- [15] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [16] K. Feng, Z. Jiang, W. He, and B. Ma, "A recognition and novelty detection approach based on Curvelet transform, nonlinear PCA and SVM with application to indicator diagram diagnosis," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12721–12729, 2011.
- [17] Q. Jiang, M. Jia, J. Hu, and F. Xu, "Machinery fault diagnosis using supervised manifold learning," *Mechanical Systems and Signal Processing*, vol. 23, no. 7, pp. 2301–2311, 2009.
- [18] E. G. Strangas, S. Aviyente, and S. S. H. Zaidi, "Time-frequency analysis for efficient fault diagnosis and failure prognosis for interior permanent-magnet AC motors," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4191–4199, 2008.
- [19] Y. Wang, E. W. M. Ma, T. W. S. Chow, and K.-L. Tsui, "A two-step parametric method for failure prediction in hard disk drives," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 419–430, 2014.
- [20] J. Yu, "Local and nonlocal preserving projection for bearing defect classification and performance assessment," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 5, pp. 2363–2376, 2012.
- [21] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1990.
- [22] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, Minneapolis, Minn, USA, June 2007.
- [23] M. B. Zhao, R. H. M. Chan, P. Tang, T. W. S. Chow, and S. W. H. Wong, "Trace ratio linear discriminant analysis for medical diagnosis: a case study of dementia," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 431–434, 2013.
- [24] L. Zhou, L. Wang, and C. H. Shen, "Feature selection with redundancy-constrained class separability," *IEEE Transactions on Neural Networks*, vol. 21, no. 5, pp. 853–858, 2010.
- [25] T. Sun and S. Chen, "Class label versus sample label-based CCA," *Applied Mathematics and Computation*, vol. 185, no. 1, pp. 272–283, 2007.
- [26] Y.-F. Guo, S.-J. Li, J.-Y. Yang, T.-T. Shu, and L.-D. Wu, "A generalized Foley-Sammon transform based on generalized fisher discriminant criterion and its application to face recognition," *Pattern Recognition Letters*, vol. 24, no. 1–3, pp. 147–158, 2003.
- [27] M. B. Zhao, X. H. Jin, Z. Zhang, and B. Li, "Fault diagnosis of rolling element bearings via discriminative subspace learning: visualization and classification," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3391–3401, 2014.
- [28] X. H. Jin, M. B. Zhao, T. W. S. Chow, and M. S. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441–2451, 2014.
- [29] Y. Q. Jia, F. P. Nie, and C. S. Zhang, "Trace ratio problem revisited," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 729–735, 2009.
- [30] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.
- [31] C. Zhang, F. Nie, and S. Xiang, "A general kernelization framework for learning algorithms based on kernel PCA," *Neurocomputing*, vol. 73, no. 4–6, pp. 959–967, 2010.
- [32] S. Ji and J. Ye, "Kernel uncorrelated and regularized discriminant analysis: a theoretical and computational study," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 10, pp. 1311–1321, 2008.
- [33] L. C. Jiao and L. Wang, "A novel genetic algorithm based on immunity," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 30, no. 5, pp. 552–561, 2000.
- [34] F. R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, No. 92, American Mathematical Society, 1997.

- [35] J. S. Zhang, Z. B. Xu, and Y. Liang, "The whole annealing genetic algorithms and their sufficient and necessary conditions of convergence," *Science of China*, vol. 27, no. 2, pp. 154–164, 1997.
- [36] K. A. Loparo, "Bearings vibration data set," Case Western Reserve University, <http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>.
- [37] D. Wang, Q. Miao, and R. Kang, "Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition," *Journal of Sound and Vibration*, vol. 324, no. 3–5, pp. 1141–1157, 2009.
- [38] D. Wang, P. W. Tse, W. Guo, and Q. Miao, "Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis," *Measurement Science and Technology*, vol. 22, no. 2, Article ID 025102, 2011.
- [39] F. J. Alonso and D. R. Salgado, "Analysis of the structure of vibration signals for tool wear detection," *Mechanical Systems and Signal Processing*, vol. 22, no. 3, pp. 735–748, 2008.
- [40] K. P. Zhu, G. S. Hong, and Y. S. Wong, "A comparative study of feature selection for hidden Markov model-based micro-milling tool wear monitoring," *Machining Science and Technology*, vol. 12, no. 3, pp. 348–369, 2008.
- [41] D. Wang, Q. Miao, X. F. Fan, and H.-Z. Huang, "Rolling element bearing fault detection using an improved combination of Hilbert and wavelet transforms," *Journal of Mechanical Science and Technology*, vol. 23, no. 12, pp. 3292–3301, 2010.
- [42] Y. G. Lei, M. J. Zuo, Z. J. He, and Y. Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1419–1430, 2010.
- [43] D. Wang, W. Guo, and P. W. Tse, "An enhanced empirical mode decomposition method for blind component separation of a single-channel vibration signal mixture," *Journal of Vibration and Control*, 2014.
- [44] Y. G. Lei, M. J. Zuo, and M. R. Hoseini, "The use of ensemble empirical mode decomposition to improve bispectral analysis for fault detection in rotating machinery," *Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science*, vol. 224, no. 8, pp. 1759–1769, 2010.
- [45] Y. G. Lei, Z. J. He, and Y. Y. Zi, "Application of the EEMD method to rotor fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1327–1338, 2009.

## Research Article

# Gearbox Fault Diagnosis Using Complementary Ensemble Empirical Mode Decomposition and Permutation Entropy

Liye Zhao,<sup>1,2</sup> Wei Yu,<sup>1,2</sup> and Ruqiang Yan<sup>1</sup>

<sup>1</sup>*School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China*

<sup>2</sup>*Key Laboratory of Micro Inertial Instrument and Advanced Navigation Technology, Ministry of Education, Nanjing 210096, China*

Correspondence should be addressed to Ruqiang Yan; ruqiang@seu.edu.cn

Received 7 July 2015; Revised 26 September 2015; Accepted 15 October 2015

Academic Editor: Dong Wang

Copyright © 2016 Liye Zhao 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 paper presents an improved gearbox fault diagnosis approach by integrating complementary ensemble empirical mode decomposition (CEEMD) with permutation entropy (PE). The presented approach identifies faults appearing in a gearbox system based on PE values calculated from selected intrinsic mode functions (IMFs) of vibration signals decomposed by CEEMD. Specifically, CEEMD is first used to decompose vibration signals characterizing various defect severities into a series of IMFs. Then, filtered vibration signals are obtained from appropriate selection of IMFs, and correlation coefficients between the filtered signal and each IMF are used as the basis for useful IMFs selection. Subsequently, PE values of those selected IMFs are utilized as input features to a support vector machine (SVM) classifier for characterizing the defect severity of a gearbox. Case study conducted on a gearbox system indicates the effectiveness of the proposed approach for identifying the gearbox faults.

## 1. Introduction

Gears can be considered as significant subassembly in machines for power or rotation transmission from one shaft to another. Their fault may cause unexpected breakdown of the machine systems and lead to significant economic loss or even personnel casualties [1, 2]. Since structural defect-caused vibration signals often reflect changes of the dynamic characteristics related to the gearbox, many researches focus on transient feature extraction of the vibration signal and fault recognition of the defective gearboxes using vibration signal analysis [3]. Nonetheless, a number of factors related to structural transformation, friction, velocity shear, and strike affect the vibration-oriented signal study and reduce the effectiveness of defective diagnosis. Consequently, a number of conventional linear approaches might not operate well in detection of dynamic changes [4, 5].

Aiming at avoiding restrictions of conventional techniques, permutation entropy (PE) is used to characterize vibration signals for the purpose of fault diagnosis. The PE only uses the order of entropy for signal characterization and can overcome nonlinear distortion which existed in

the signal. It has been applied in various applications. For instance, permutation entropy is proved to offer an efficient evaluation to monitor rolling bearings [5]. By integrating support vector machine (SVM) with multiscale PE, the operating condition of rolling bearing can be identified [6, 7]. Another study combined optimized SVM, ensemble empirical mode decomposition, and PE to detect and classify motor bearing faults [8]. The effectiveness of the PE has also been proved in detecting dynamic changes in rotating machines when comparing with that of other features, like Lyapunov exponent and fractal dimensions [9]. Furthermore, background noise which existed in real world applications always disturbs the result of the fault diagnosis. Therefore, performing noise reduction in the vibration signals is necessary before the PE method is executed.

Empirical mode decomposition (EMD), as an approach of adaptive signal treatment in the field of time frequency analysis, can decompose a signal into sets of intrinsic mode functions (IMFs) based on its features [10, 11]. The approach of EMD could be applied in pretreatment of the signals of vibration. For instance, a novel approach for extracting fault feature with combined AR model and EMD algorithm has

been applied in processing bearing vibration signals [5]. Nonetheless, when the EMD is used to process a signal with intermittent components, the signal could not be fully decomposed because of the existence of mode mixture phenomenon [12]. Further development with ensemble empirical mode decomposition (EEMD) was proposed by Huang et al. for EMD performance improvement. EEMD becomes more precise and efficient for decomposition of signals in comparison with the original EMD by adding noise to the original signal and continuously calculating the IMF means [12]. Though the approach of EEMD has efficiently solved the issue of mode-mixing, it takes lots of time to implement the large amount of ensemble mean. In other words, the efficiency of algorithm will be decreased. In order to resolve this issue, the complementary approach of EEMD (CEEMD) has been put forward [13]. Through complementary integration of IMFs and both positive and negative added white noises in the CEEMD, residual of the noises could be extracted out from the combination of white noises and data. The approach of CEEMD demonstrates similar effectiveness to that of the EEMD with improved computational efficiency.

By making full use of characteristics of the PE and CEEMD, this paper proposes a hybrid approach to diagnose gearbox faults. The CEEMD is utilized as the preprocessing to filter signals and extract IMFs that are closely associated with the filtered signal. Subsequently, PE value of each chosen IMF would be calculated. The PE value of the chosen IMFs is utilized as the feature vector to a classifier in which the support vector machine (SVM) is applied for identifying gearbox defect. The remaining parts of this paper are arranged as follows. Overview of the gearbox fault diagnosis approach is shown in Section 2. Experimental verification is conducted on automobile transmission gearbox system in Section 3. Last but not least, the last section presents the summary and comment.

## 2. Theoretical Framework

*2.1. Complementary Ensemble Empirical Mode Decomposition.* CEEMD is developed based upon EEMD. Originally, the EMD approach deals with a given signal  $x_i(t)$  into the form presented in (1) through recursive elimination of the mean of the lower and upper envelope related to the maximum and minimum of the signal [14]:

$$x_i(t) = \sum_{i=1}^N C_{i,N}(t) + r_{i,n}(t), \quad (1)$$

where  $N$  refers to the number of IMFs,  $C_{i,N}(t)$  refers to the component of IMF which covers a certain frequency band, and  $r_{i,n}(t)$  refers to the mean trend of the signal residue. The EMD can be considered as adaptive local analysis approach for processing both nonlinear and nonstationary signals. However, the decomposition of EMD would generally undergo mixture of modes, which is defined as either a single IMF covering widely disparate scales or a signal existing in different IMF components.

Later, Huang et al. have proposed a noise-guided statistical approach to resolve the mode mixture issue, which is

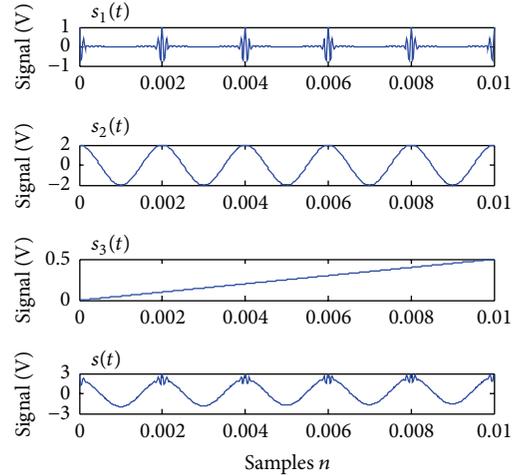


FIGURE 1: The simulated signal.

the ensemble empirical mode decomposition. However, the effect of the additional noise could only be restricted by a large amount of ensemble mean computation, causing high computational load.

Complementary ensemble mode decomposition, as an improved and noise enhanced data analysis approach, has been developed for reducing computational burden [13]. The procedure of CEEMD for the signal  $x(t)$  is illustrated in the following steps.

*Step 1.* A pair of white Gaussian noises with the same amplitude is added to  $x(t)$ . Thus, two signals,  $x_1 = x + x_n$  and  $x_2 = x - x_n$ , are generated.

*Step 2.* Decompose  $x_1$  and  $x_2$  by EMD for a number of times; then  $\text{IMF}_{x_1}$  referring to ensemble means of IMF from  $x_1$  and  $\text{IMF}_{x_2}$  referring to those from  $x_2$  are obtained.

*Step 3.* The final IMF which is the ensemble of  $\text{IMF}_{x_1}$  and  $\text{IMF}_{x_2}$  is calculated as the decomposition results of CEEMD as follows:

$$\text{IMF} = \frac{(\text{IMF}_{x_1} + \text{IMF}_{x_2})}{2}. \quad (2)$$

Specifically, a simulated signal  $s(t)$  composed of  $s_1(t)$ ,  $s_2(t)$ , and  $s_3(t)$  has been adopted as an instance.  $s_1(t)$  is a Gaussian impulse interference signal,  $s_2(t)$  is a cosine signal with the frequency of 500 Hz, and  $s_3(t)$  is a trend term. Figure 1 illustrates the waveform of the simulated signal and Figure 2 illustrates the decomposed results by CEEMD.

Through comparing the result in Figure 2 with the signal waveforms in Figure 1, it is shown that there is no mode mixture. That is to say, CEEMD is more suitable for the study of signal.

*2.2. Permutation Entropy.* PE is a nonlinear dynamic parameter that characterizes a signal's complexity. Based on the

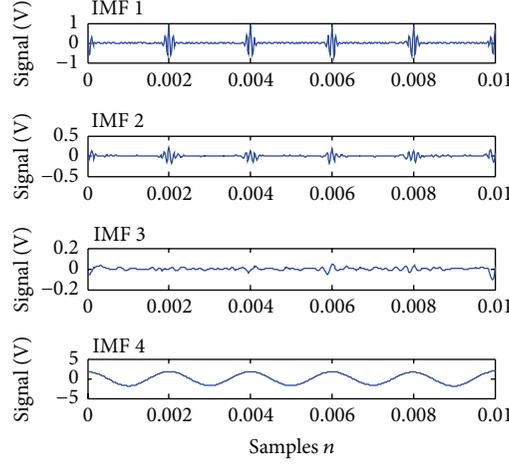


FIGURE 2: The decomposition result by CEEMD.

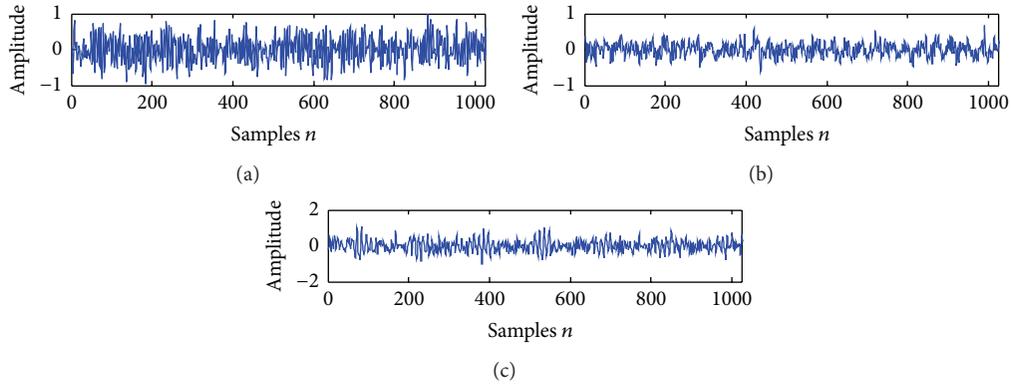


FIGURE 3: Gearbox vibration signals under various operating conditions.

principle of Takens-Maine, the phase space of time series  $\{y(i), i = 1, 2, \dots, N\}$  can be expressed as

$$Y(i) = \{y(i), y(i + \tau), \dots, y(i + (m-1)\tau)\}, \quad (3)$$

$$i = 1, 2, \dots, N - (m-1)\tau,$$

where  $m$  refers to the embedded dimension, while  $\tau$  refers to the delay of time. Furthermore,  $m$  sample points of data contained in every  $Y(i)$  could be sorted in an incremental order as

$$\{y(i + (j_1 - 1)\tau) \leq y(i + (j_2 - 1)\tau) \leq \dots \leq y(i + (j_m - 1)\tau)\}. \quad (4)$$

If  $y(i + (j_1 - 1)\tau) = y(i + (j_2 - 1)\tau)$ , the original positions could be classified as  $j_1 \leq j_2$ ,  $y(i + (j_1 - 1)\tau) \leq y(i + (j_2 - 1)\tau)$ . Thus, vector  $Y(i)$  could be shown in a set of symbols as [15, 16]

$$T(l) = (j_1, j_2, \dots, j_m), \quad (5)$$

where  $l = 1, 2, \dots, k$  and  $k \leq m!$ .  $T(l)$  refers to  $m!$  symbol permutation which has been shown in  $m$  number symbols  $(j_1, j_2, \dots, j_m)$ . If  $P_1, P_2, \dots, P_k$  are applied in denoting the possibility distribution of each symbol sequence and

$\sum_{j=1}^J P_j = 1$ , the permutation entropy of  $m$  for the time series of  $\{y(i), i = 1, 2, \dots, N\}$  could be considered as the entropy of Shannon for  $k$  symbol sequence as follows:

$$H_{PE}(m) = - \sum_{j=1}^J P_j \ln P_j. \quad (6)$$

If all the symbol sequences appear with the same possibility distribution as  $P_j = 1/m!$ , the maximum value of  $H_{PE}(m)$  could be described as  $\ln(m!)$ . Thus, the permutation entropy of order  $m$  can be standardized as

$$0 \leq H_{PE} = \frac{H_{PE}(m)}{\ln(m!)} \leq 1. \quad (7)$$

$H_{PE}$  value shows the randomness level of the time series. A large value of  $H_{PE}$  indicates high randomness of the time series. On the contrary, a small value of  $H_{PE}$  means the time series has more regular characteristics.

To demonstrate the validity of the PE algorithm, sample vibration signals of a gearbox under three different conditions are shown in Figure 3, and the corresponding single factor analysis result is shown in Figure 4. Figure 4 shows that defect severity of the gearbox could be efficiently recognized by the value of PE.

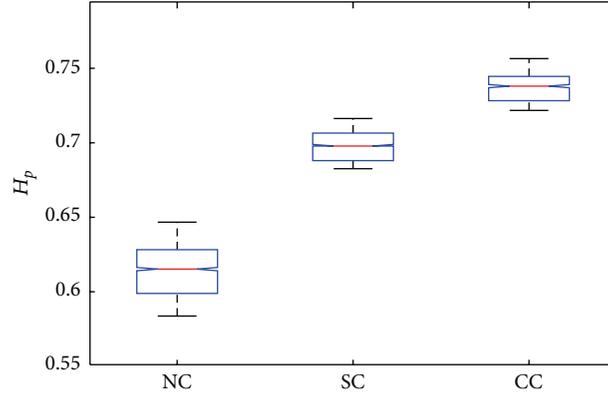


FIGURE 4: Boxplot of PE values on normal condition (NC), slight fault condition (SC), and catastrophe fault condition (CC).

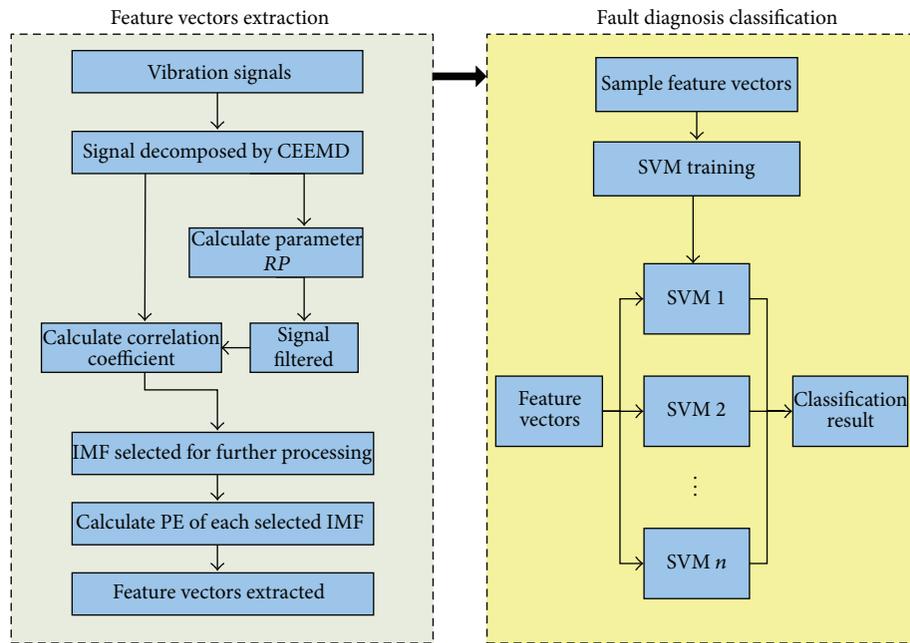


FIGURE 5: Flow chart of the proposed method.

**2.3. Fault Diagnosis Based on CEEMD and Permutation Entropy.** In this study, a gearbox fault diagnosis method has been developed using the CEEMD and PE, and Figure 5 shows the flow chart of the method. Particularly, the procedure to implement the proposed fault diagnosis method is as follows.

*Step 1.* The sampled vibration signal measured on gearbox is decomposed using CEEMD.

*Step 2.* The product  $P_j$  is calculated using (8), and the parameter  $RP_j$  is calculated by (9). The signal is filtered through comparison of the proposed threshold value and the parameter  $RP_j$  [17]. In other words, when  $RP_j \geq 1$ , it can be assured that  $P_j$  of the  $j$ th IMF can be enhanced for a number of times in comparison with the mean value of  $P_j$  which can be calculated based on the former  $j - 1$  IMFs. Thus, the previous  $j - 1$  IMFs with the term of trend can be eliminated

as noise and the residue IMFs can be considered as filtered signal:

$$P_j = E_j \times T_j, \quad (8)$$

$$RP_j = \left| \frac{P_j - (1/(j-1)) \sum_{i=1}^{j-1} P_i}{(1/(j-1)) \sum_{i=1}^{j-1} P_i} \right| \quad (j \geq 2), \quad (9)$$

where  $E_j = (1/N) \sum_{i=1}^N [A_j(i)]^2$  refers to the  $j$ th IMF's energy density,  $T_j = 2N/O_j$  refers to the mean period of the  $j$ th IMF,  $N$  represents the length of each IMF,  $A_j$  denotes the  $j$ th IMF's amplitude, and  $O_j$  refers to the overall number of extreme points in the  $j$ th IMF.

*Step 3.* The correlation coefficients between each IMF and filtered signal are calculated by (10). IMFs closely associated

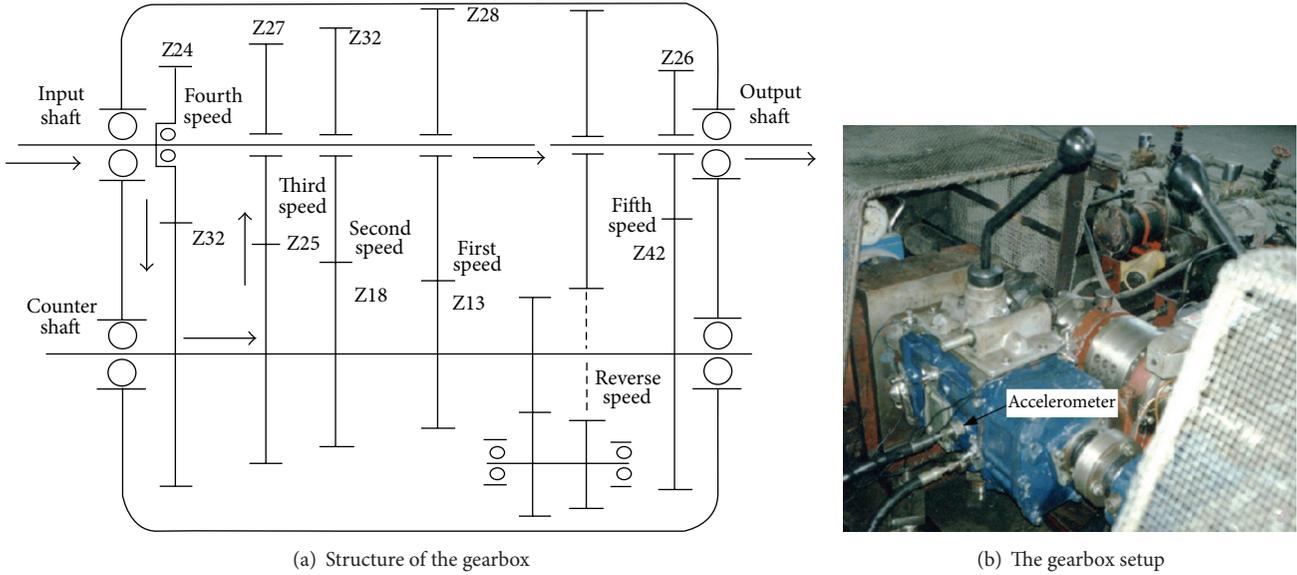


FIGURE 6: The automobile transmission gearbox.

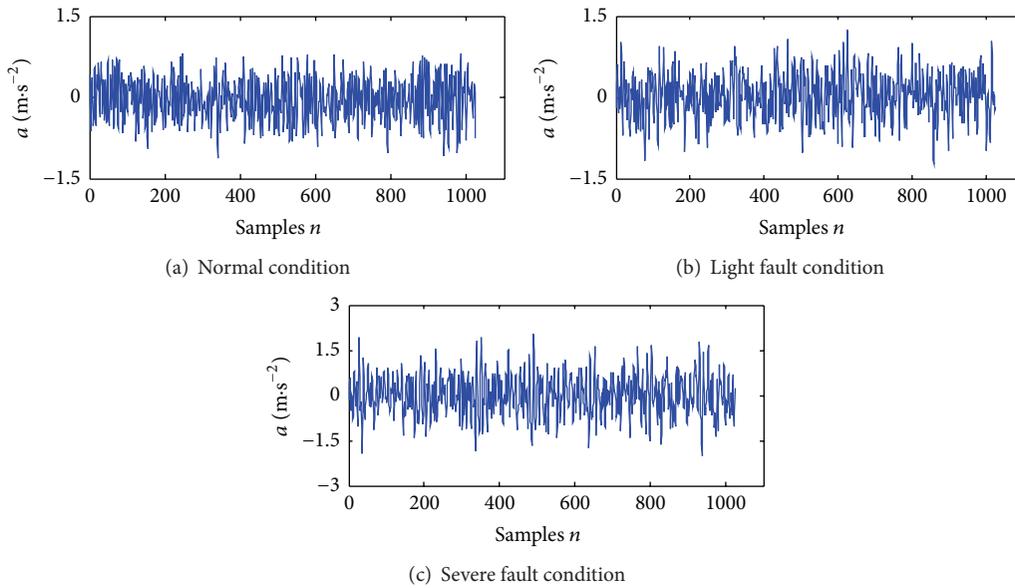


FIGURE 7: Vibration signal waveforms of the gearbox under different conditions.

with the filtered signal are chosen to calculate the PE value [18]:

$$\rho_{xy} = \frac{\sum_{k=1}^N x(k) y(k)}{\left[ \sum_{k=1}^N x(k)^2 \sum_{k=1}^N y(k)^2 \right]^{1/2}}. \quad (10)$$

*Step 4.* The PE values of all the chosen IMFs are calculated to generate a feature vector which can be utilized to train the SVM for identification of gearbox operating condition.

*Step 5.* The PE feature vector from test gearbox vibration signal is extracted and utilized as input to the well-trained SVMs. In this way, the result of classification can be realized [19, 20].

### 3. Experimental Evaluation

A series of gearbox fault signals acquired from LC5T81 type transmission were used to verify the effectiveness of the presented approach. The data was measured from the testbed presented in Figure 6. One backward speed and five forward speeds could be load on the tested gearbox. The vibration signals were collected at 3000 samples per second using the accelerometer fixed on the gearbox case. The tested gearbox is operated with the third speed of 1600 rpm and the meshing frequency of 500 Hz.

The waveforms of the vibration signals collected from the test gearbox under three conditions are shown in Figure 7.

TABLE 1: Correlation coefficients between filtered signals and each IMF.

	Correlation coefficient									
	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7	IMF 8	IMF 9	IMF 10
Normal condition	0.5235	0.7648	0.4783	0.4284	0.4036	0.1092	0.0971	-0.0056	-0.0012	-0.0078
Light fault condition	0.8594	0.4635	0.3582	0.2347	0.1921	0.0925	-0.0008	-0.0074	-0.0056	-0.0033
Severe fault condition	0.9219	0.2578	0.2215	0.1937	0.1625	0.0859	-0.0014	0.0008	-0.0015	0.0007

TABLE 2: Permutation entropy values of IMF 1~IMF 5.

	Permutation entropy value				
	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5
Normal condition	0.958	1.239	0.782	0.837	0.5036
Light fault condition	0.710	0.859	0.518	0.421	0.8421
Severe fault condition	0.539	0.709	0.876	0.659	0.2625

TABLE 3: Fault diagnosis using improved approach based on CEEMD and PE.

Fault type	Test sample	Classification results			Classification rate [%]	Overall classification rate [%]
		Normal condition	Minor fault condition	Serious fault condition		
Normal condition	20	20	0	0	100	
Light fault condition	20	1	18	1	90	95
Severe fault condition	20	0	1	19	95	

TABLE 4: Fault diagnosis using the approach based on CEEMD and ApEn.

Fault type	Test sample	Classification results			Classification rate [%]	Overall classification rate [%]
		Normal condition	Minor fault condition	Serious fault condition		
Normal condition	20	19	0	1	95	
Light fault condition	20	1	17	2	85	88.3
Severe fault condition	20	1	2	17	85	

Figure 7(a) shows the signal under the normal condition, Figure 7(b) shows the signal under the light fault condition, and Figure 7(c) shows the signal from the severe fault condition.

Figure 8 illustrates the decomposed IMFs of these signals and Table 1 shows the correlation coefficients between the filtered signal and each of the IMFs.

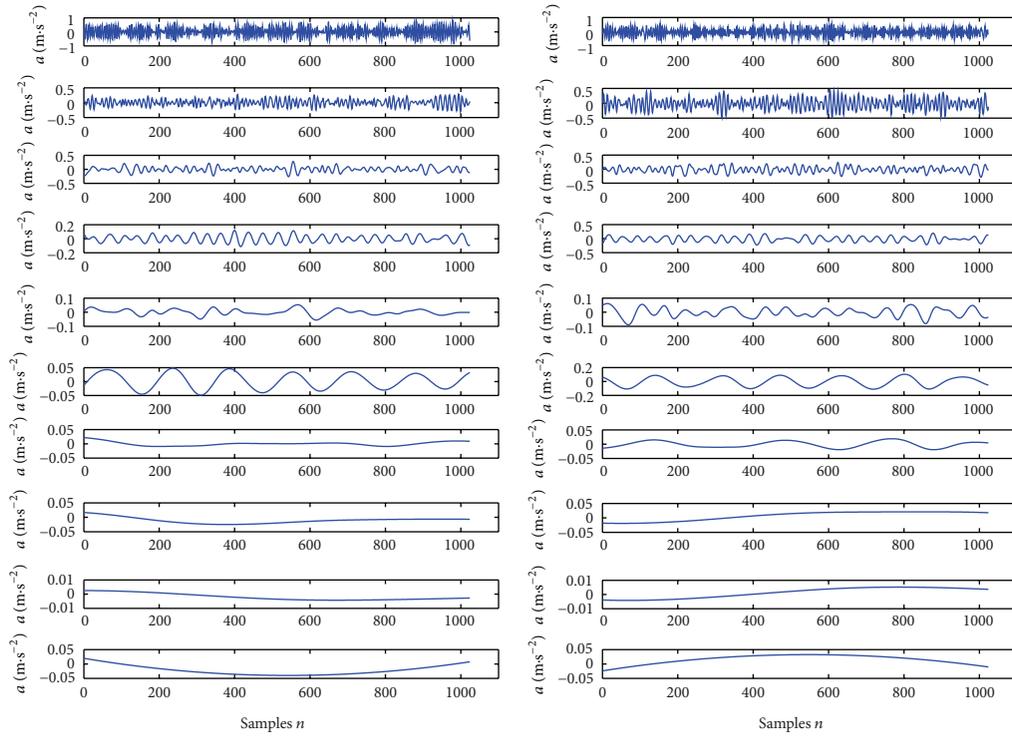
It can be seen from the table that correlation coefficients for the first 5 IMFs are all more than 0.1. They can describe the main features of the signal and thus are selected for further analysis. According to the main steps of the presented fault diagnosis approach, the permutation entropy values of these IMFs are calculated, as listed in Table 2.

In the experiment, 120 feature vectors in total were gained from three different circumstances. 50% of the feature vectors were applied into classifier training, while the rest of them were used in classification of fault. Table 3 shows the results

of classification. It shows that various working conditions can be efficiently identified. Among all the 60 groups of feature vectors, 57 groups have been classified correctly, while 3 groups have failed. The overall classification accuracy is up to 95%.

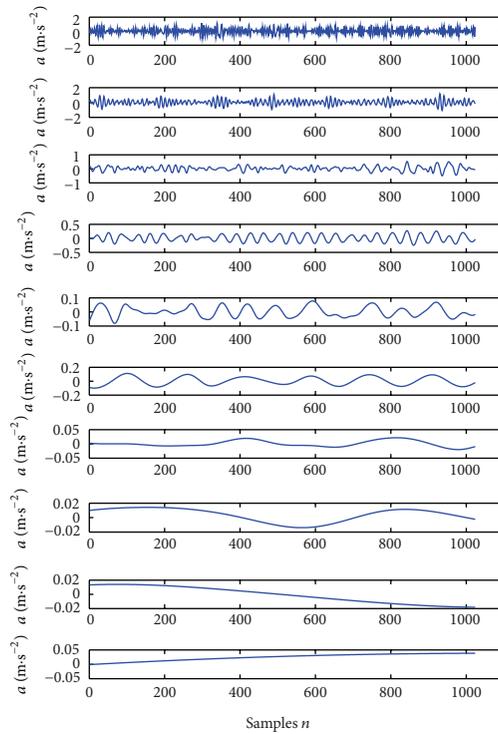
For purpose of comparison, the values of approximate entropy (ApEn) from the chosen IMFs are also calculated and applied in the SVM classifier. Table 4 shows the classification results. It can be summarized that the method is actually efficient for differentiating the gearbox faults. Furthermore, the effectiveness of the approach is compared with that of the EEMD-PE approach. It can be seen that the rates of classification in these two approaches are very similar. However, computational load of the developed approach is lower than that of the EEMD-PE approach.

To further study the effectiveness of the developed approach, a  $10 \times 10$ -fold cross validation procedure is employed



(a) Normal condition

(b) Light fault condition



(c) Severe fault condition

FIGURE 8: The decomposition result by CEEMD under different conditions.

with the selected 120 samples. The average classification rate of the  $10 \times 10$ -fold cross validation is 94.82%. The result is close to the classification result in Table 3.

#### 4. Conclusions

This study develops an integrated approach by combining PE algorithm with CEEMD to diagnose gearbox faults. With the CEEMD, gearbox vibration signals can be decomposed into sets of IMFs with low computational load. Then PE method can efficiently extract fault characteristic from the selected IMFs. Without mathematical model and the study of the fault mechanism of the system, this developed approach can directly recognize gearbox fault severity. Furthermore, the CEEMD, as a preprocessing step, can be utilized to purify the signal for PE calculation, leading to increased classification rate (e.g., 95% for experimental data). It is envisioned that the approach developed in this study could be used in a wide range of applications in the field of fault diagnosis.

#### Conflict of Interests

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

#### Acknowledgments

This paper is supported by the Nature Science Foundation of Jiangsu Province of China (no. BK2012739) and the National Natural Science Foundation of China (no. 61101163 and no. 51175080).

#### References

- [1] W. Fan, G. Cai, Z. K. Zhu, C. Shen, W. Huang, and L. Shang, "Sparse representation of transients in wavelet basis and its application in gearbox fault feature extraction," *Mechanical Systems and Signal Processing*, vol. 56–57, pp. 230–245, 2015.
- [2] D. Wang, Q. Miao, Q. Zhou, and G. Zhou, "An intelligent prognostic system for gear performance degradation assessment and remaining useful life estimation," *Journal of Vibration and Acoustics*, vol. 137, no. 2, Article ID 021004, 2015.
- [3] R. Yan and R. X. Gao, "Wavelet domain principal feature analysis for spindle health diagnosis," *Structural Health Monitoring*, vol. 10, no. 6, pp. 631–642, 2011.
- [4] D. Wang, Q. Miao, and R. Kang, "Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition," *Journal of Sound and Vibration*, vol. 324, no. 3–5, pp. 1141–1157, 2009.
- [5] R. Yan, Y. Liu, and R. X. Gao, "Permutation entropy: a non-linear statistical measure for status characterization of rotary machines," *Mechanical Systems and Signal Processing*, vol. 29, pp. 474–484, 2012.
- [6] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, and C.-C. Wang, "Bearing fault diagnosis based on multiscale permutation entropy and support vector machine," *Entropy*, vol. 14, no. 8, pp. 1343–1356, 2012.
- [7] J. Zheng, J. Cheng, and Y. Yang, "Multiscale permutation entropy based rolling bearing fault diagnosis," *Shock and Vibration*, vol. 2014, Article ID 154291, 8 pages, 2014.
- [8] X. Zhang, Y. Liang, J. Zhou, and Y. Zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.
- [9] K. Keller and M. Sinn, "Ordinal analysis of time series," *Physica A: Statistical Mechanics and its Applications*, vol. 356, no. 1, pp. 114–120, 2005.
- [10] Z. H. Wu and N. E. Huang, "Ensemble empirical mode decomposition a noise assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [11] J. Zhang, R. Yan, R. X. Gao, and Z. Feng, "Performance enhancement of ensemble empirical mode decomposition," *Mechanical Systems and Signal Processing*, vol. 24, no. 7, pp. 2104–2123, 2010.
- [12] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 108–126, 2013.
- [13] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 4144–4147, Prague, Czechia, May 2011.
- [14] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London*, vol. 454, no. 1971, pp. 903–995, 1998.
- [15] R. Tiwari, V. K. Gupta, and P. K. Kankar, "Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier," *Journal of Vibration and Control*, vol. 21, no. 3, pp. 461–467, 2015.
- [16] G. Xiong, W. Huang, and L. Zhang, "Fault severity identification of rolling bearing based on multiscale entropy," *Journal of Applied Sciences*, vol. 13, no. 13, pp. 2404–2408, 2013.
- [17] Z. H. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2046, pp. 1597–1611, 2004.
- [18] R.-X. Chen, B.-P. Tang, and J.-H. Ma, "Adaptive de-noising method based on ensemble empirical mode decomposition for vibration signal," *Journal of Vibration and Shock*, vol. 31, no. 15, pp. 82–86, 2012.
- [19] D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, and A. Alonso-Betanzos, "Automatic bearing fault diagnosis based on one-class  $\nu$ -SVM," *Computers & Industrial Engineering*, vol. 64, no. 1, pp. 357–365, 2013.
- [20] L. M. R. Baccarini, V. V. Rocha E Silva, B. R. de Menezes, and W. M. Caminhas, "SVM practical industrial application for mechanical faults diagnostic," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6980–6984, 2011.

## Research Article

# Sparse Signal Representations of Bearing Fault Signals for Exhibiting Bearing Fault Features

Wei Peng,<sup>1,2</sup> Dong Wang,<sup>3</sup> Changqing Shen,<sup>4</sup> and Dongni Liu<sup>3</sup>

<sup>1</sup>Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan 528402, China

<sup>2</sup>Institute of Reliability Engineering, School of Mechatronics Engineering, University of Electronic Science and Technology of China, Chengdu 610051, China

<sup>3</sup>Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

<sup>4</sup>School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215021, China

Correspondence should be addressed to Dong Wang; [dongwang4-c@my.cityu.edu.hk](mailto:dongwang4-c@my.cityu.edu.hk)

Received 5 May 2015; Revised 7 October 2015; Accepted 12 October 2015

Academic Editor: Peng Chen

Copyright © 2016 Wei Peng 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.

Sparse signal representations attract much attention in the community of signal processing because only a few coefficients are required to represent a signal and these coefficients make the signal understandable. For bearing faults' diagnosis, bearing fault signals collected from transducers are often overwhelmed by strong low-frequency periodic signals and heavy noises. In this paper, a joint signal processing method is proposed to extract sparse envelope coefficients, which are the sparse signal representations of bearing fault signals. Firstly, to enhance bearing fault signals, particle swarm optimization is introduced to tune the parameters of wavelet transform and the optimal wavelet transform is used for retaining one of the resonant frequency bands. Thus, sparse wavelet coefficients are obtained. Secondly, to reduce the in-band noises existing in the sparse wavelet coefficients, an adaptive morphological analysis with an iterative local maximum detection method is developed to extract sparse envelope coefficients. Simulated and real bearing fault signals are investigated to illustrate how the sparse envelope coefficients are extracted. The results show that the sparse envelope coefficients can be used to represent bearing fault features and identify different localized bearing faults.

## 1. Introduction

Rolling element bearings are commonly used in machines to support rotation shafts. Their failures may cause unexpected machine breakdown and lead to huge economic loss. A rolling element bearing consists of an outer race, an inner race, several rollers, and a cage. Once a defect is formed on the surface of either the outer race or the inner race, an impact is generated by each of the rollers striking the defect surface and thus it excites the resonant frequencies of the structures between bearings and transducers [1–4]. Therefore, to extract bearing fault features, envelope analysis is one of the most effective methods. To conduct envelope analysis, two steps are needed. The first step aims to use a band-pass filter to retain one of the resonant frequency bands for enhancing the signal to noise ratio of bearing fault signals because bearing

fault signals are often overwhelmed by strong low-frequency vibration components and heavy noises. The second step is extracting the envelope of the signals filtered by the band-pass filter [5–7]. Moreover, if the envelope signals can be represented by a few coefficients, namely, sparse envelope coefficients [8], bearing fault signals are more understandable and easily interpreted.

In this paper, a joint signal processing method for extraction of sparse envelope coefficients is proposed. Firstly, an optimal wavelet filter is tuned by particle swarm optimization (PSO). For the use of wavelet transform, the similarity between a signal and a wavelet is the most concerned and the high similarity can result in large wavelet coefficients so as to highlight hidden transients. Because the shape of a Morlet wavelet is similar with the transients caused by localized bearing faults and the Morlet wavelet has a band-pass property,

which can be used to retain one of the resonant frequency bands and enhance the signal to noise ratio of bearing fault signals, the Morlet wavelet is chosen in this paper [9–14]. To automatically tune the parameters of the Morlet wavelet, two aspects including a metric and an optimization algorithm must be determined. In the past years, some metrics, such as kurtosis, entropy, smoothness index, and sparsity measurement, have been investigated for optimization of the Morlet wavelet transform [9–14]. Their comparisons show that the sparsity measurement can generate better visual inspection performance and highlight bearing fault signatures, such as bearing fault characteristic frequency and its harmonics [15]. Therefore, the metric used in this paper is the sparsity measurement. To achieve the global optimal parameters of the Morlet wavelet transform, genetic algorithm, differential evolution, and stepwise scanning have been studied [9–15]. To explore the feasibility of an easy and simple optimization algorithm, particle swarm optimization is used in this paper to tune the sparsity measurement because the core of particle swarm optimization is based on the simple physical relationship among position, velocity, and acceleration. The use of PSO is simply introduced as follows. A number of particles move in a searching space. Then, a simple mathematic algorithm searches the best position by sharing the cognitive and social influences among all particles in the searching space [16–18]. For intelligent machine fault diagnosis, PSO was used to tune the parameters of support vector machine, artificial neural network, and proximal support vector machine, respectively [19–21]. The results show that these statistical prediction models combined with PSO have good prediction accuracies for identification of different machine faults.

Secondly, even though the optimal wavelet filtering is conducted on bearing fault signals for enhancement of bearing fault signatures, in-band noises still exist because the optimal wavelet filtering cannot remove the noises existing in the retained resonant frequency band. In recent years, an attracting method called morphological analysis (MA) is widely investigated due to its simplicity and effectiveness in extracting envelope signals [22–29]. MA aims to use Minkowski addition and subtraction to intersect the morphological features of bearing fault signals with a structuring element (SE). However, if the morphological features are overwhelmed by other strong vibration components and heavy noises, MA may fail to retain the morphological features of bearing fault signals. Therefore, MA can be used to postprocess bearing fault signals, if the signal to noise ratio of bearing fault signals is not high. In this paper, an adaptive MA with an iterative local maximum detection method is developed to automatically find the optimal parameter of MA and extract sparse envelope signals so as to exhibit bearing fault features.

The rest of this paper is outlined as follows. Section 2 introduces the fundamental algorithms used in this paper. These algorithms include wavelet transform, particle swarm optimization, and morphological analysis. In Section 3, extraction of sparse envelope coefficients from bearing fault signals is proposed. In Section 4, simulated and real bearing

faults signals are analyzed by using the proposed method. Conclusions are drawn in Section 5.

## 2. Fundamental Algorithms

*2.1. Fundamental Theory of Wavelet Transform.* Wavelet transform aims to calculate the inner product between an artificial wavelet and a signal. The mathematical formula for wavelet transform is defined as follows [30, 31]:

$$Wf(u, s) = \int_{-\infty}^{+\infty} f(t) \times \frac{1}{\sqrt{s}} \overline{\psi\left(\frac{u-t}{s}\right)} dt \quad (1)$$

$$= f(u) * \psi'_s(u),$$

$$\psi'_s(t) = \frac{1}{\sqrt{s}} \overline{\psi\left(\frac{-t}{s}\right)}, \quad (2)$$

where  $s$  is the scale parameter and  $u$  is the translation parameter.  $*$  represents the convolution operator.  $\overline{\phantom{x}}$  takes the complex conjugate of the signal. According to the properties of Fourier transform, (1) is rewritten as

$$Wf(u, s) = f(u) * \psi'_s(u) = F^{-1}\left(\widehat{f}(f) \times \sqrt{s} \widehat{\psi'_s}(sf)\right), \quad (3)$$

where  $F^{-1}$  is the inverse Fourier transform and  $\widehat{\phantom{x}}$  is the Fourier transform. As explained in Introduction, the complex Morlet wavelet is chosen in this paper. Its temporal waveform and the corresponding frequency spectrum are given as follows:

$$\psi(t) = \frac{\sigma}{\sqrt{\pi}} e^{-\sigma^2 t^2} e^{j2\pi f_c t}, \quad (4)$$

$$\vartheta(f) = e^{-\pi^2/\sigma^2 \times (f-f_c)^2}. \quad (5)$$

From (5), it is obvious that the complex Morlet wavelet has a band-pass property and its frequency support is constrained to the frequency band  $[f_c - \sigma/2, f_c + \sigma/2]$ . Because any wavelet must satisfy the admission condition, which means that the integration of a wavelet over time must be equal to zero, the following equation should be satisfied:

$$\vartheta(0) = e^{-\pi^2/\sigma^2 \times (0-f_c)^2} = 0. \quad (6)$$

It is not difficult to verify that if  $f_c/\sigma > 1.3$ ,  $\vartheta(0) \approx 0$ .

*2.2. Fundamental Theory of Particle Swarm Optimization.* PSO is a population based stochastic optimization method, which optimizes a metric by iteratively moving a number of particles in a searching space, according to some simple mathematical formulas related to the positions and velocities of all particles. Each particle represents one potential solution to the optimization problem. The movements of the particles are guided by their local best positions and the best swarm position. The basic theory of PSO is described in the following [16]. Considering the physical relationship among position, velocity, and acceleration, the following basic physic principle is listed as follows:

$$x_i(k+1) = x_i(k) + v_i(k) + \frac{1}{2} a_i(k), \quad (7)$$

where  $x_i(k+1)$  and  $x_i(k)$  mean the  $k+1$ th and the  $k$ th positions of the  $i$ th particle, respectively.  $v_i(k)$  is the  $k$ th velocity of the  $i$ th particle.  $a_i(k)$  is the  $k$ th acceleration of the  $i$ th particle. Then, the acceleration of the  $i$ th particle is divided into a cognitive acceleration, which is proportional to the distance between the current position of the  $i$ th particle and the personal best position  $p_i(k)$  of the  $i$ th particle, and a social acceleration, which is proportional to the distance between the current position of the  $i$ th particle and the global best position  $g(k)$  of the  $i$ th particle. To make these two new parts more flexible, a cognitive coefficient  $c_1$  and a social coefficient  $c_2$  are used. Consequently, (7) is reformulated as

$$\begin{aligned} x_i(k+1) = & x_i(k) + v_i(k) + \frac{1}{2}c_1(p_i(k) - x_i(k)) \\ & + \frac{1}{2}c_2(g(k) - x_i(k)). \end{aligned} \quad (8)$$

Then, in order to prevent the velocities from getting out of control, the influence of friction is considered by introducing an inertia weight  $\omega$ , which is smaller than 1, to (8). Besides, the constant 1/2 is replaced by two random numbers  $r_{1i}$  and  $r_{2i}$ , which are limited to the values between 0 and 1. Equation (8) is revised as follows:

$$\begin{aligned} x_i(k+1) = & x_i(k) + \omega v_i(k) + r_{1i}c_1(p_i(k) - x_i(k)) \\ & + r_{2i}(k)c_2(g(k) - x_i(k)). \end{aligned} \quad (9)$$

At last, (9) consists of the following velocity and position update equations:

$$\begin{aligned} v_i(k+1) = & \omega v_i(k) + r_{1i}c_1(p_i(k) - x_i(k)) \\ & + r_{2i}(k)c_2(g(k) - x_i(k)), \end{aligned} \quad (10)$$

$$x_i(k+1) = x_i(k) + v_i(k+1).$$

**2.3. Fundamental Theory of Morphological Analysis.** MA aims to extract the morphological shape of a temporal signal. It uses a structuring element to intersect with the temporal signal. Let  $f(n)$  be the one-dimensional signal over a domain  $F = (0, 1, 2, \dots, N-1)$  and let  $g(n)$  be the structuring element over a domain  $G = (0, 1, 2, \dots, M-1)$ . The basic morphology operators include the dilation operator and the erosion operator, which are related to Minkowski addition and subtraction. The equations for the erosion operator  $\ominus$  and the dilation operator  $\oplus$  are defined as follows [22]:

$$(f \ominus g)(n) = \min [f(n+m) - g(m)], \quad (11)$$

$$m \in 0, 1, 2, \dots, M-1,$$

$$(f \oplus g)(n) = \max [f(n-m) + g(m)], \quad (12)$$

$$m \in 0, 1, 2, \dots, M-1.$$

The erosion operator reduces the wave peaks and enlarges the signal minima. On the contrary, the dilation operator increases the wave valleys and enlarges signal maxima [24]. Other morphological operators are constructed based on

the combination of the above two operators. Some popular morphological operators used in bearing fault diagnosis are introduced in the following. The opening operator and the closing operator are defined as:

$$\begin{aligned} (f \circ g)(n) &= [(f \ominus g^s) \oplus g](n), \\ (f \cdot g)(n) &= [(f \oplus g^s) \ominus g](n), \end{aligned} \quad (13)$$

where  $g^s(n)$  is the reflection of  $g(n)$ . The opening operator function is smoothing the signal from the bottom by cutting wave peaks and the closing operator function is smoothing the signal from the top by filling up its wave valleys [24]. The average operator (AVG), the difference operator (DIF), the Black Top-Hat operator (BTH), and the White Top-Hat operator (WTH) are defined as [32]

$$\begin{aligned} \text{AVG}(f) &= \frac{(f \cdot g + f \circ g)}{2}, \\ \text{DIF}(f) &= f \cdot g - f \circ g, \\ \text{BTH}(f) &= f \cdot g - f, \\ \text{WTH}(f) &= f - f \circ g. \end{aligned} \quad (14)$$

The ability of the average operator lies in flattening both the positive and negative impulsive features. On the contrary, the difference operator is used to extract both the positive and negative impulsive features. The Black Top-Hat and the White Top-Hat operators are employed to extract the negative and positive impulsive features, respectively.

### 3. Extraction of Sparse Envelope Coefficients for Exhibiting Bearing Fault Features

It is not difficult to find that the morphological features extracted by using MA fully depend on the shape of a temporal signal. Because of the interruption from strong low-frequency periodic components and heavy noises, the morphological features of bearing fault signals are prone to be overwhelmed. Therefore, it is necessary to enhance weak bearing fault signals prior to the use of MA. As illustrated in the previous sections, the complex Morlet wavelet optimized by PSO is used to preprocess bearing fault signals and an adaptive MA is developed to postprocess bearing fault signals and to extract sparse envelope coefficients for exhibiting bearing fault features. The flowchart of the proposed method is shown in Figure 1. Each step used in Figure 1 is detailed in the following paragraphs.

*Step 1.* Load an original bearing fault vibration signal, which is collected by using a transducer attached to a bearing housing.

*Step 2.* The parameters of the complex Morlet wavelet are tuned by PSO. First, the parameters of particle swarm optimization must be initialized. Each particle corresponds to two-dimensional coordinates, which are used to represent the center frequency and bandwidth of the complex Morlet

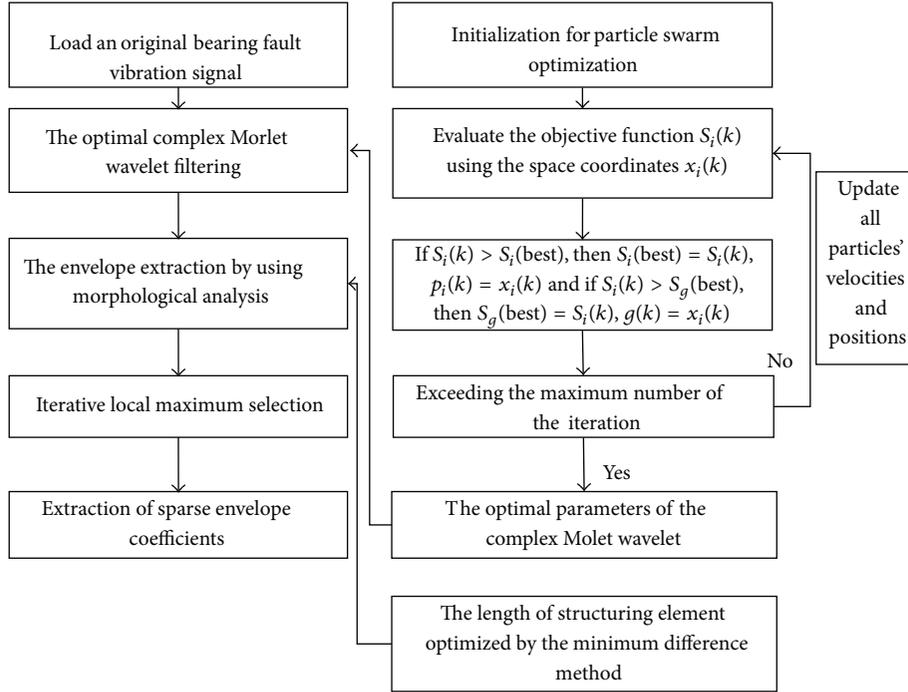


FIGURE 1: The flowchart of the developed method for extracting sparse envelope coefficients.

wavelet, respectively. The searching range for the center frequency was set to the frequency band ranging from  $0.1 \times F_s$  to  $0.4 \times F_s$  for avoiding the interruption caused by the low-frequency vibration components. The searching range for the bandwidth is from  $3 \times F_f$  to  $0.2 \times F_s$ . Here, to contain enough fault signatures, the minimum bandwidth should be three times larger than the inner race fault characteristic frequency.  $F_s$  is the sampling frequency and  $F_f$  is the inner race fault characteristic frequency, which will be formulated later. According to the literature review [17, 18], the inertia weight, the number of particles, and the maximum iteration were empirically set to 0.9, 24, and 50, respectively, because these parameters are sufficiently large for solving an optimization problem. For all particles, the initial positions are randomly generated within the specified searching ranges. The metric, namely, sparsity measurement, is defined as [15]

$$S_i(k) = \frac{\sqrt{\sum_{n=1}^L b_i^k(n)^2}}{\sum_{n=1}^L |b_i^k(n)|} = \frac{\|b_i^k(n)\|_2}{\|b_i^k(n)\|_1}, \quad (15)$$

$$1 \leq i \leq 24, \quad 1 \leq k \leq 50,$$

where  $\|b_i^k(n)\|_2$  and  $\|b_i^k(n)\|_1$  are  $L_2$  norm and  $L_1$  norm, respectively.  $b_i^k(n)$  represents the signal filtered by the complex Morlet wavelet transform, the parameters of which are the two-dimensional coordinates  $x_i(k)$  of the  $i$ th particle at the  $k$ th iteration.  $L$  is the length of the signal. At the  $k$ th iteration, if the  $S_i(k)$  of the  $i$ th particle is larger than the personal largest sparsity value  $S_p(\text{best})$  of the  $i$ th particle (recorded before the  $k$ th iteration), the personal best position of the  $i$ th particle is updated by its current position  $x_i(k)$ .

Then, if the  $S_i(k)$  of the  $i$ th particle is larger than the global largest sparsity value  $S_g(\text{best})$  of all particles (recorded before the  $k$ th iteration), the global best position of all particles is updated by the current position  $x_i(k)$  of the  $i$ th particle. At last, repeat the above steps until the maximum iteration is reached. The optimal parameters of the complex Morlet wavelet are automatically established by PSO with the maximum sparsity. Here, the real part of the filtered signal obtained by the optimal complex Morlet wavelet is denoted by  $b_{\text{opt}}^{\text{real}}(n)$ .

*Step 3.* Once the weak bearing fault signal is enhanced by the optimal complex Morlet wavelet, MA is performed to get sparse envelope coefficients. As introduced in Section 2.3, there are some available morphological operators. To select a proper morphological operator, their comparisons for processing a simulated signal are shown in Figure 2. The structuring element is the flat element with a length of 30 samples because the flat element is a simple and effective element to process a one-dimensional signal [22, 24, 27, 28]. From the results shown in Figure 2, the closing and opening operators can be used to extract the positive and negative envelopes of the signal, respectively. In this paper, only the closing operator is used because the positive envelope is used for further analyses. After both morphological operator and the structuring element are determined, it is necessary to determine the length of the flat element to extract the morphological features of bearing fault signals and suppress in-band noises.

Reference [22] reported a method to empirically decide the length of the flat element and indicated that the optimal

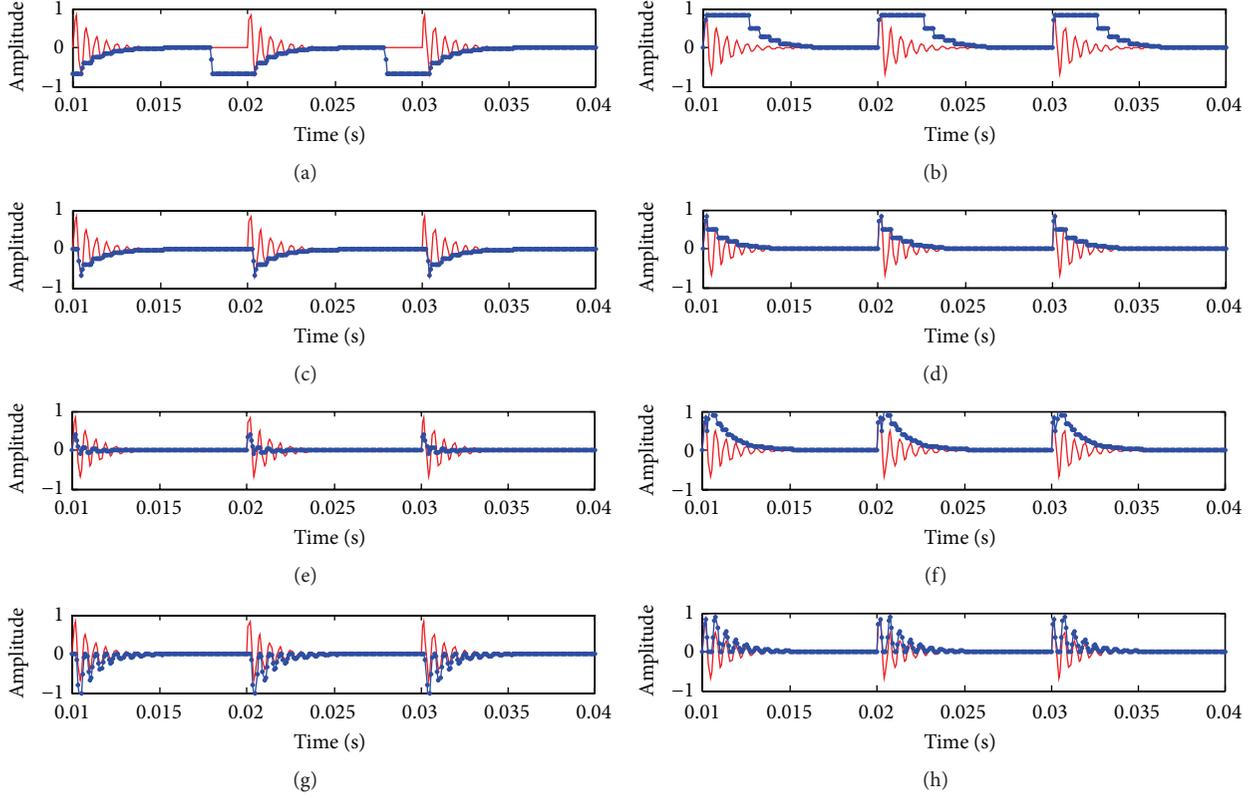


FIGURE 2: The signals obtained by using various morphological operators: (a) the signal filtered by the erosion operator; (b) the signal filtered by the dilation operator; (c) the signal filtered by opening operator; (d) the signal filtered by the closing operator; (e) the signal filtered by the AVG operator; (f) the signal filtered by the DIF operator; (g) the signal filtered by the BTH operator; and (h) the signal filtered by the WTH operator.

length should be 0.6 to 0.7 of the bearing fault period. However, in our previous research [26], it is found that the empirical optimal length is not always effective for extracting bearing fault features. This paper develops an adaptive morphological analysis with an iterative local maximum detection method to automatically retain sparse envelope coefficients.  $b_{\text{opt}}^{\text{real}}(n)$  is processed by the closing operator with various flat element lengths and their corresponding results are denoted by  $mm_{\text{opt}}^{\text{real}}(n, l)$ , where  $l$  is the length of the flat element used in the closing operator. The maximum length of the flat element does not exceed the desired bearing fault period  $T$  (unit: samples), which will be defined later. First, find the local maxima  $lm(j)$ ,  $j = 1, 2, \dots, N$  of the signal  $mm_{\text{opt}}^{\text{real}}(n, l)$  and calculate its local maxima number  $N$ . Subtract the theoretical impulsive number  $Z$  from the local maxima number  $N$  and obtain a difference (considering the influence of the negative difference, the absolute value of the difference is used in this paper). The theoretical impulsive number can be calculated as

$$Z = \text{round}\left(\frac{L}{T}\right), \quad (16)$$

where the  $\text{round}(\ast)$  function is taking the element to the nearest integer.

The bearing fault period  $T$  can be represented by the outer race fault characteristic period  $T_o$ , the inner race fault characteristic period  $T_I$ , and rolling element fault characteristic period  $T_B$ . The calculations of these periods are given by

$$\begin{aligned} T_o &= \frac{1}{((P \times f_s)/2)(1 - (d/D) \cos \alpha)^r}, \\ T_I &= \frac{1}{((P \times f_s)/2)(1 + (d/D) \cos \alpha)^r}, \\ T_B &= \frac{1}{((D \times f_s)/d)(1 - (d^2/D^2) \cos \alpha)^r}, \end{aligned} \quad (17)$$

where  $f_s$  is the shaft rotation frequency in Hz.  $d$  and  $D$  are diameters of the rolling element and the pitch, respectively.  $P$  is the number of rolling elements and  $\alpha$  is the contact angle.

If the difference between the local maxima number  $N$  and the theoretical impulsive number  $Z$  is the smallest, it means that the extracted signal by using the closing operator is the best one because the signal filtered by the complex Morlet transform has a high signal to noise ratio and then the local maxima of the filtered signal are the most possible to be the local peaks of the impulses caused by localized bearing faults. Therefore, the optimal length of the flat element can be decided. However, it should be pointed out that sometimes

the amplitudes of random noises may affect the number of local maxima. Therefore, the iterative local maximum selection is developed to remove the pseudo- and abnormal locations of the local maxima.

Locate all local maxima and denote them by  $llm(j)$ ,  $j = 1, 2, \dots, N$ . Then, calculate the adjacent distances between two successive locations  $llm(j+1) - llm(j)$ . There are  $N - 1$  distances  $dis(j)$ ,  $j = 1, 2, \dots, N - 1$ .

Assume that all the calculated distances are subject to a normal distribution  $N(\mu, \sigma^2)$ . Here,  $\mu$  and  $\sigma^2$  are the mean and the variance, respectively. In statistics and probability theory, standard deviation  $\sigma$  can be used to measure the diversity of samples. It is expected that all distances  $dis(j)$ ,  $j = 1, 2, \dots, N - 1$  tend to be close to the mean and have a low standard deviation. Therefore, a high standard deviation can be employed to reject the outlier (abnormal distance). In this paper, two standard deviations  $2\sigma$  are used. Then, it is believed that about the 95 percent of the all distances are within two standard deviations ( $\mu \pm 2\sigma$ ). Moreover, only distances that are smaller than  $(\mu - 2\sigma)$  are rejected because all distances are always positive.

Find the abnormal distances, which are denoted by  $llm(z+1) - llm(z)$ . Calculate the distances  $llm(z+2) - llm(z+1)$  and  $llm(z) - llm(z-1)$ , respectively. If  $llm(z+2) - llm(z+1) > llm(z) - llm(z-1)$ ,  $llm(z)$  is discarded. Otherwise,  $llm(z+1)$  is discarded.

Repeat the iterative local maximum selection method until all distances are within two standard deviations. Finally, sparse envelope coefficients are extracted.

## 4. Case Studies

*4.1. A Simulated Fault Signal Analyzed by the Proposed Method.* In the first case study, a simulated signal is used for analyses. The simulated signal contains the impulses with an exponential decay, two sinusoidal signals, and noises. Here, the two sinusoidal signals can be regarded as two strong interruptions caused by two low-frequency components:

$$y(k) = \sum_r \exp\left(-\alpha \times \frac{(k-r) \times F_s / f_m - \tau_r}{F_s}\right) \cdot \sin\left(2\pi f_1 \times \frac{(k-r) \times F_s / f_m - \tau_r}{F_s}\right) + 0.9 \sin\left(\frac{2\pi f_2 k}{F_s}\right) + \sin\left(\frac{2\pi f_3 k}{F_s}\right) + nm(k), \quad (18)$$

where  $\alpha$  is equal to 900,  $nm(k)$  is the noise term, and  $f_m$  is the modulating frequency (equal to 100 Hz).  $F_s$  is the sampling frequency set to 12000 Hz.  $f_1$  is the carrier frequency, equal to 3500 Hz.  $f_2$  and  $f_3$  are sinusoidal frequencies, equal to 70 Hz and 140 Hz, separately. 3600 samples are used. Normally distributed heavy noises with a mean of 0 and a variance of 0.2 are used as the noises. The simulated signal, the noise signal, and the mixed signal are shown in Figures 3(a), 3(b), and 3(c), respectively. From the mixed signal shown in Figure 3(c), it is hard to identify the periodic impulsive signal (the periodic intervals are equal to 10 ms).

The proposed method is employed to analyze the mixed signal. First, the optimal parameters of the complex Morlet wavelet are automatically determined by using PSO with the maximum sparsity measurement. The optimal center frequency, the optimal bandwidth, and the frequency spectrum of the complex Morlet wavelet are shown in Figure 4(b). For a comparison, the frequency spectrum of the mixed signal is plotted in Figure 4(a). It is found that the optimal complex Morlet wavelet is correct to locate the simulated resonant frequency band around 3500 Hz. Then, the mixed signal is filtered by the optimal complex Morlet wavelet and its corresponding frequency spectrum is given in Figure 4(c). The real part of the filtered signal is depicted in Figure 5(a), where it is seen that the in-band noises in the resonant frequency band still exist in the filtered signal. On the other hand, to validate the correction of the filtered signal, the power spectrum of the envelope of the filtered signal is plotted in Figure 5(b), where the fundamental frequency 100 Hz and its first two harmonics are identified. The global best values by using PSO are shown in Figure 6, where it is seen that sparsity value reaches the global optimal value quickly.

At last, the morphological analysis with the closing operator is employed to extract the cyclic bearing fault characteristics. The obtained temporal signal by morphological analysis is given in Figure 7(a). The absolute difference between the theoretical impulsive number and the actual impulsive number is shown in Figure 7(b), where the optimal length of 46 is found for morphological analysis. The local maximum locations of the signal shown in Figure 7(a) are displayed in Figure 7(c). It is obvious that there is a pseudolocation. The iterative local maximum selection method is applied to process these local maximum locations shown in Figure 7(c). Finally, the revised local maximum locations are plotted in Figure 7(d). Compared with the signal shown in Figure 5(a), the result plotted in Figure 7(d) can be more understandable. Moreover, it is clear to see the signal having a cyclic interval of 10 ms by picking up the interval between two successive spikes. Besides, the final signal shown in Figure 7(d) is the sparse envelope coefficients of the simulated signal mixed with noise.

*4.2. Experimental Fault Signals Analyzed by the Proposed Method.* In this paper, the real motor bearing data picked up with a sampling frequency of 12 kHz by an accelerometer at the drive end of the motor housing are used to validate the proposed method. A single-point defect was, respectively, introduced to the outer race and inner race of a normal bearing using electrodischarge machining. The fault diameter is 0.007 inches and the fault depth is 0.0011 inches. The motor load was 0 HP and the motor speed was 1797 rpm [33]. Outer race and inner race fault characteristic frequencies were calculated as 107 Hz and 162 Hz, respectively.

The original outer race fault signal and its frequency spectrum are shown in Figures 8(a) and 8(b), respectively. The optimal parameters of the complex Morlet wavelet and its frequency spectrum are shown in Figure 8(c). Then, the optimal complex Morlet wavelet is used to retain one of the resonant frequency bands. The frequency spectrum of the signal obtained by the optimal wavelet filtering is displayed

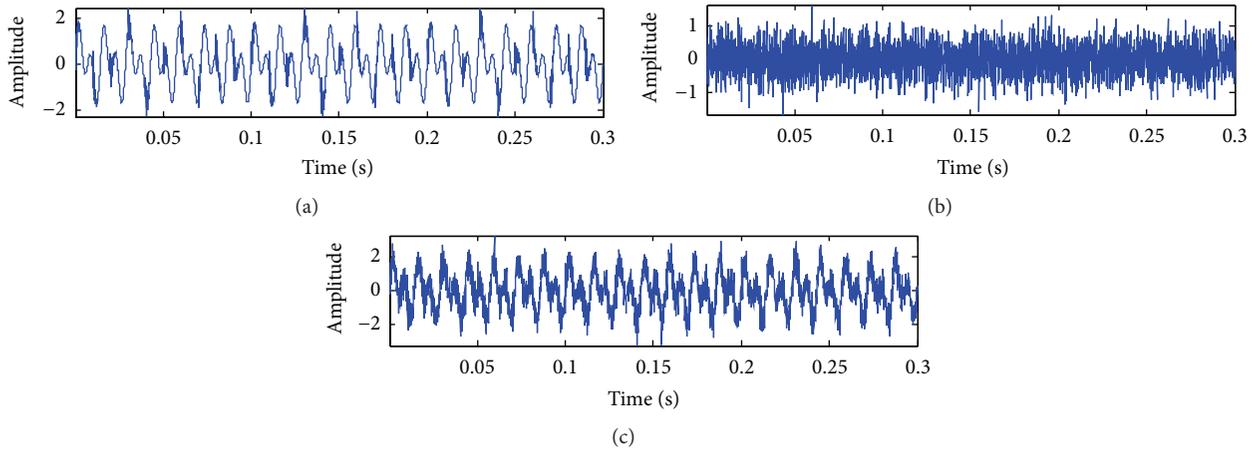


FIGURE 3: The simulated temporal signals: (a) the pure simulated signal consisting of impulses and two sinusoidal signals; (b) the noise signal; and (c) the pure simulated signal mixed with noises.

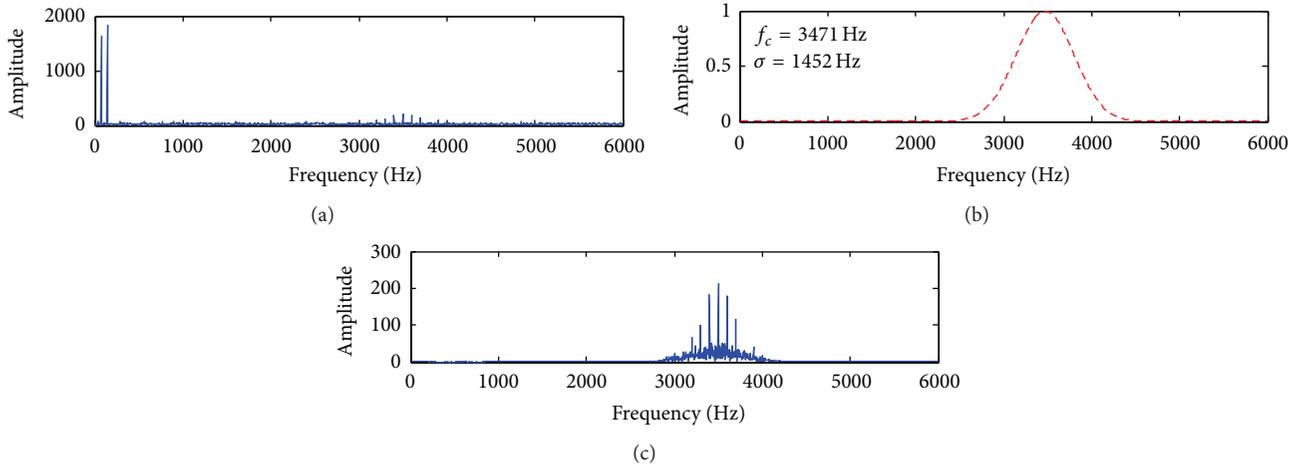


FIGURE 4: Frequency spectra: (a) the frequency spectrum of the mixed signal; (b) the frequency spectrum of the optimal complex Morlet wavelet; and (c) the frequency spectrum of the signal filtered by the optimal complex Morlet wavelet transform.

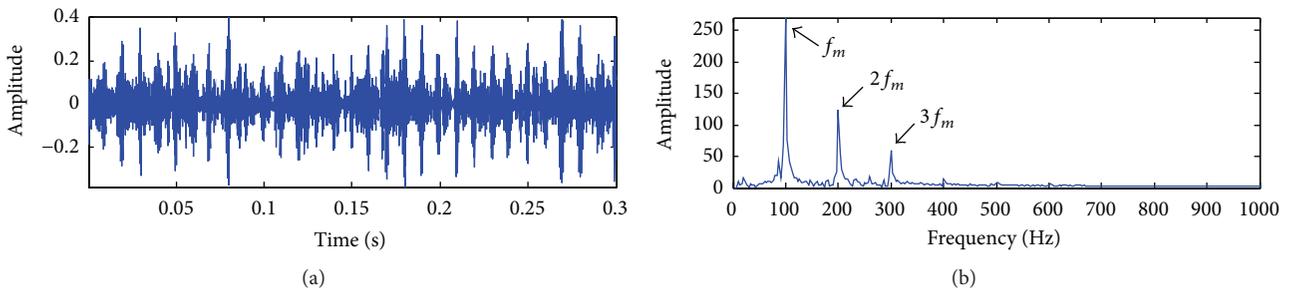


FIGURE 5: The signals obtained by the optimal complex Morlet filtering: (a) the real part of the filtered signal; (b) the power spectrum of the envelope of the filtered signal.

in Figure 8(d). It is clear to find that one of the resonant frequency bands is kept for further morphological analysis. The real part of the filtered signal and its envelope spectrum are shown in Figures 9(a) and 9(b), respectively, which

demonstrate that the optimal filtering method correctly retains the major bearing fault signatures. The global best values by using particle swarm optimization are plotted in Figure 10.

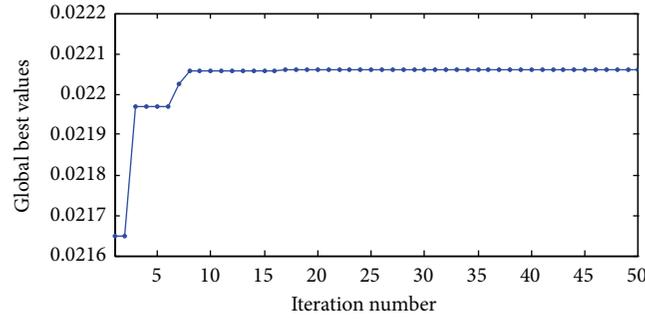


FIGURE 6: The global best values at different iterations by using PSO.

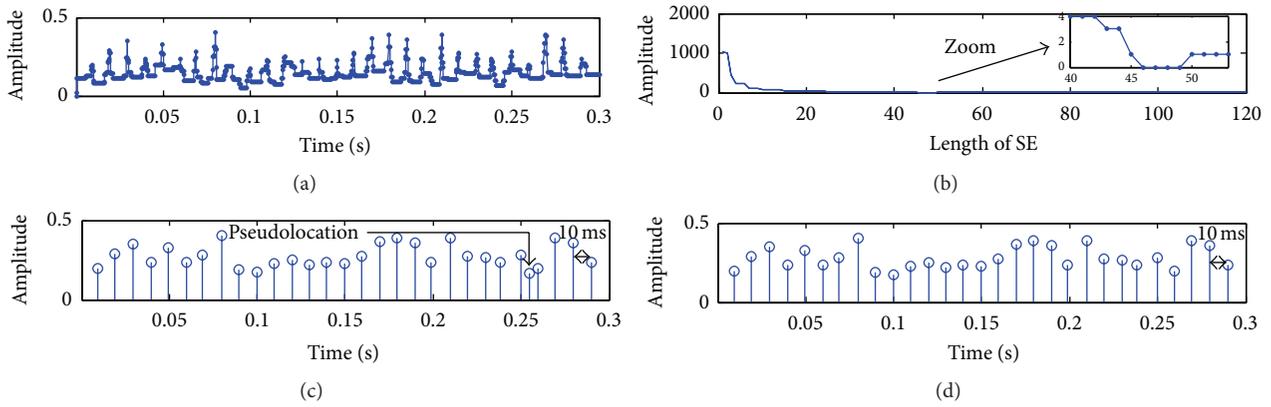


FIGURE 7: The results obtained by morphological analysis: (a) envelope extraction of the signal shown in Figure 5(a) by using the closing operator; (b) the optimal length of the SE; (c) the local maximum locations of the signal shown in Figure 7(a); and (d) the revised local maximum locations of the signal shown in Figure 7(c) by using the iterative local maximum selection.

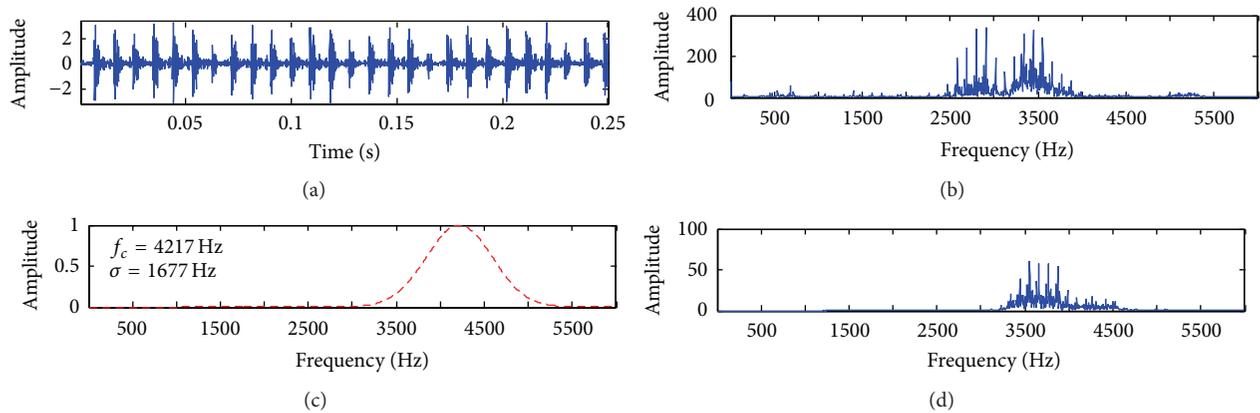


FIGURE 8: The signals: (a) the real outer race fault signal; (b) the frequency spectrum of the original outer race fault signal; (c) the frequency spectrum of the optimal complex Morlet wavelet; and (d) the frequency spectrum of the filtered signal.

Finally, morphological analysis is used to modify the shape of the signal shown in Figure 9(a). The extracted envelope by using the closing operator with the optimal length of the flat structuring element is displayed in Figure 11(a). The absolute differences are plotted in Figure 11(b) to find the optimal length of the flat structuring element. Here, the length of 29 is found. The local maximum locations of the envelope shown in Figure 11(a) are given in Figure 11(c).

It is obvious that the noises interrupt the local maximum locations. One pseudolocation is caused by an unexpected noise. Therefore, the iterative local maximum detection method is used to remove the irrelevant local maximum location. The result is shown in Figure 11(d). After removing the interruption caused by the unexpected noise, it is clear to see the pure cyclic fault characteristics with intervals of 9.3 ms.

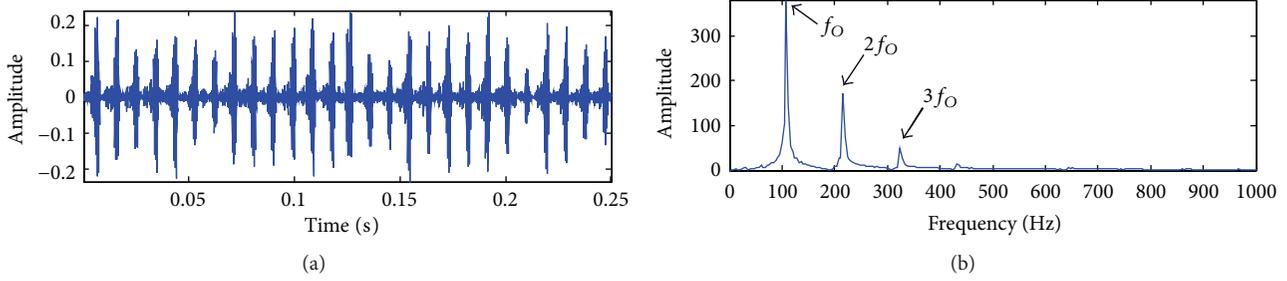


FIGURE 9: The signals obtained by the optimal complex Morlet filtering: (a) the real part of the filtered signal; (b) the power spectrum of envelope of the filtered signal.

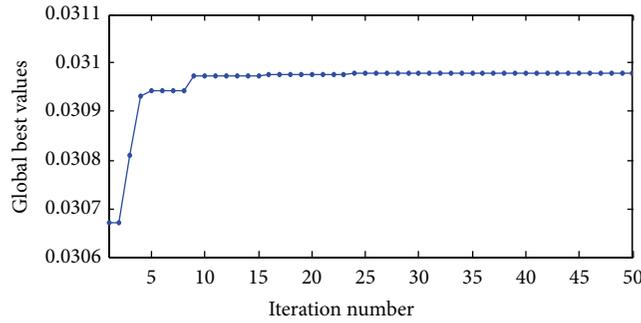


FIGURE 10: The global best values at different iterations by using PSO.

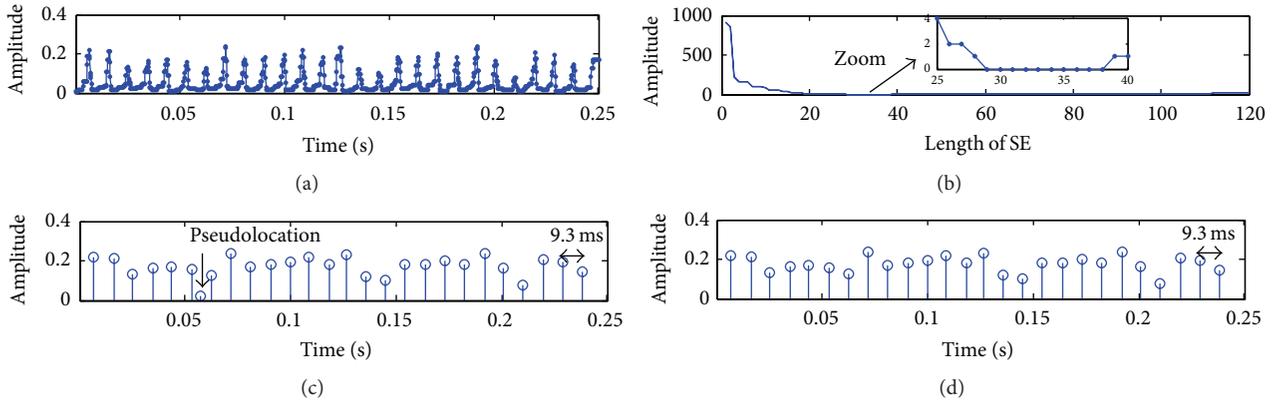


FIGURE 11: The results obtained by morphological filtering: (a) envelope extraction of the signal shown in Figure 9(a) by using the closing operator; (b) the optimal length of SE; (c) the local maximum locations of the signal shown in Figure 11(a); and (d) the revised local maximum locations of the signal shown in Figure 11(c) by using the iterative local maximum selection.

The same procedure is applied to process the inner race fault signal shown in Figure 12(a). The frequency spectrum of the inner race fault signal is plotted in Figure 12(b). The optimal parameters of the complex Morlet wavelet are given in Figure 12(c). The frequency spectrum of the filtered signal obtained by the optimal complex Morlet wavelet is displayed in Figure 12(d). Figures 13(a) and 13(b) show the real part of the filtered signal and its corresponding envelope spectrum. Obviously, the optimal complex Morlet wavelet transform reserves the inner race fault signatures. The global best values by using PSO are depicted in Figure 14.

The optimal process for the length selection of the flat structuring element is shown in Figure 15(b). The envelope obtained by using the closing operator with the optimal

length is given in Figure 15(a). After that, the local maximum locations are extracted and the result is given in Figure 15(c). The pseudolocation is caused by an unexpected noise. In order to remove the irrelevant local maximum location shown in Figure 15(c), the iterative local maximum detection method is used. The final resulting signal is shown in Figure 15(d), where a pure cyclic fault signal is generated.

### 5. Conclusion

This paper reported a method which was used to extract sparse envelope coefficients for exhibiting bearing fault features. The proposed method consisted of two steps. Firstly, a Morlet wavelet was optimized by particle swarm optimization and then sparse wavelet coefficients were extracted from

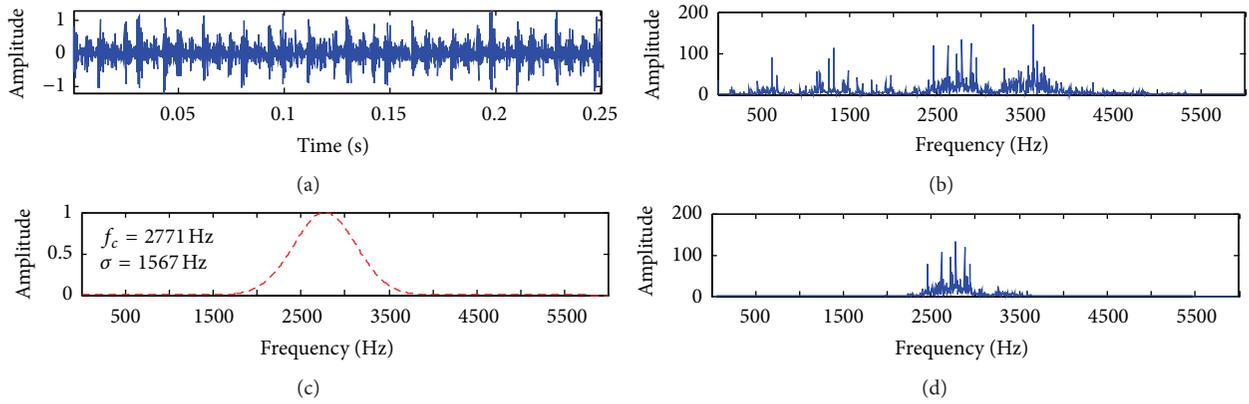


FIGURE 12: The signals: (a) the real inner race fault signal; (b) the frequency spectrum of the real inner race fault signal; (c) the frequency spectrum of the optimal complex Morlet wavelet; and (d) the frequency spectrum of the filtered signal.

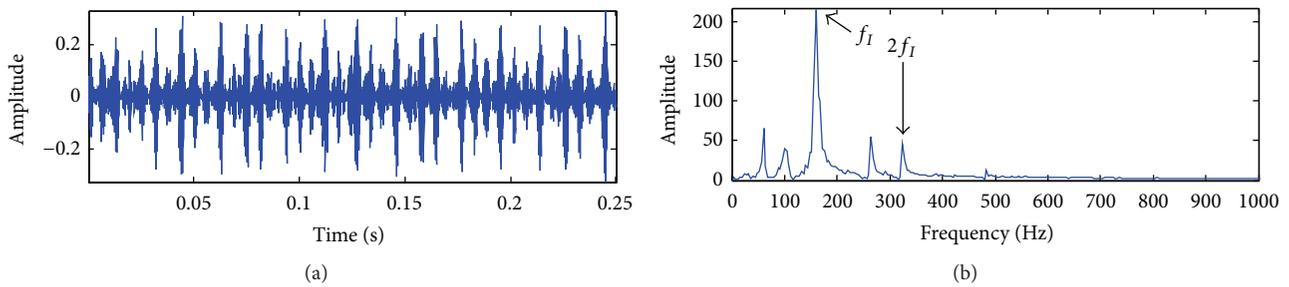


FIGURE 13: The signals obtained by the optimal complex Morlet filtering: (a) the real part of the signal filtered by the optimal Morlet wavelet transform; (b) the power spectrum of the envelope of the filtered signal.

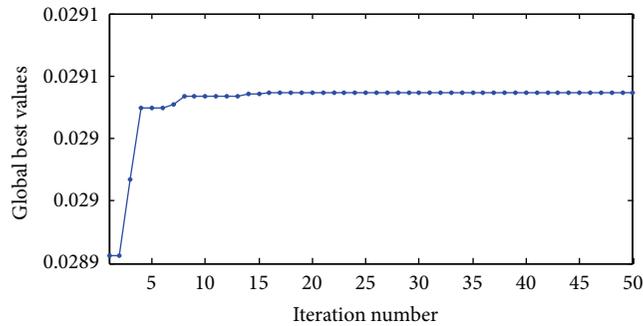


FIGURE 14: The global best values at every iteration by using particle swarm optimization.

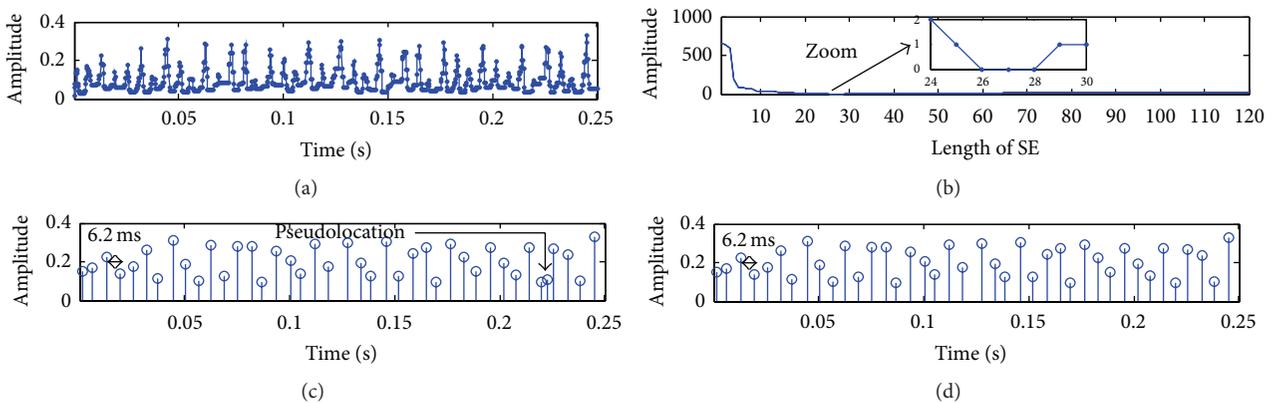


FIGURE 15: The results obtained by morphological filtering: (a) envelope extraction of the signal shown in Figure 13(a) by using the closing operator; (b) the optimal length of the flat SE; (c) the local maximum locations of the signal shown in (a); and (d) the revised local maximum locations of the signal shown in (c).

bearing fault signals. Even though sparse wavelet coefficients were useful to represent bearing fault signatures, in-band noises still existed and could not be removed by using the solely optimal wavelet transform. In this step, the sparse wavelet coefficients were not succinct enough. Secondly, to reduce in-band noises, an adaptive morphological analysis with an iterative local maximum detection method was developed to postprocess the signal obtained by the optimal wavelet filtering. The optimal flat length used in morphological analysis with the closing operator was automatically determined to retain sparse envelope coefficients and remove the pseudo- and abnormal peaks caused by unexpected noises. After the above two steps were conducted, the sparse signal representation of bearing fault signals was obtained; only a few coefficients were used to represent bearing fault features. These sparse coefficients made bearing fault signals succinct and understandable. The case studies were conducted to illustrate how the proposed method worked and the results showed that the proposed method can be used to extract the sparse representation of bearing fault signals and identify different localized bearing faults.

In the future, the following works will be conducted. Firstly, more statistical metrics will be used as objective functions so as to guide optimization of wavelet transform. Secondly, besides the popular Morlet wavelet, more wavelets that are highly similar with impulses caused by bearing defects will be investigated. Thirdly, different optimization methods will be used to find optimal wavelet parameters and their comparisons will be made. Fourthly, even though wavelet transform is effective in extracting bearing fault features, in-band noises still exist. Morphological analysis with different structure elements and operators will be studied to make bearing fault features sparse and succinct. At last, based on bearing fault features extracted by the suggested steps, intelligent bearing fault diagnosis methods will be developed to automatically identify different bearing faults without requirement of expertise.

## Conflict of Interests

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

## Acknowledgments

This research was partially supported by the National Natural Science Foundation of China (Grants nos. 51505311 and 51505307), the Science and Technology Foundation of Zhongshan City (Grant no. 20123A338), and the Natural Science Foundation of Jiangsu Province (Grant no. BK20150339). The authors would like to thank Professor K. A. Loparo for his permission to use the bearing data to validate the proposed method.

## References

- [1] S. Luo, J. Cheng, and H. Ao, "Application of LCD-SVD technique and CRO-SVM method to fault diagnosis for roller bearing," *Shock and Vibration*, vol. 2015, Article ID 847802, 8 pages, 2015.
- [2] X. Zhang, J. Kang, L. Xiao, J. Zhao, and H. Teng, "A new improved kurtogram and its application to bearing fault diagnosis," *Shock and Vibration*, vol. 2015, Article ID 385412, 22 pages, 2015.
- [3] Q. He, X. Ding, and Y. Pan, "Machine fault classification based on local discriminant bases and locality preserving projections," *Mathematical Problems in Engineering*, vol. 2014, Article ID 923424, 12 pages, 2014.
- [4] Q. Miao, L. Cong, and M. Pecht, "Identification of multiple characteristic components with high accuracy and resolution using the zoom interpolated discrete Fourier transform," *Measurement Science and Technology*, vol. 22, no. 5, Article ID 055701, 2011.
- [5] Y. Lei, J. Lin, Z. He, and Y. Zi, "Application of an improved kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1738–1749, 2011.
- [6] D. Wang, P. W. Tse, and K. L. Tsui, "An enhanced Kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 176–199, 2013.
- [7] J. Antoni, "The spectral kurtosis: a useful tool for characterising non-stationary signals," *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 282–307, 2006.
- [8] X. Chen, Z. Du, J. Li, X. Li, and H. Zhang, "Compressed sensing based on dictionary learning for extracting impulse components," *Signal Processing*, vol. 96, part A, pp. 94–109, 2014.
- [9] W. He, Z.-N. Jiang, and K. Feng, "Bearing fault detection based on optimal wavelet filter and sparse code shrinkage," *Measurement*, vol. 42, no. 7, pp. 1092–1102, 2009.
- [10] I. S. Bozchalooi and M. Liang, "A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection," *Journal of Sound and Vibration*, vol. 308, no. 1-2, pp. 246–267, 2007.
- [11] W. Su, F. Wang, H. Zhu, Z. Zhang, and Z. Guo, "Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement," *Mechanical Systems and Signal Processing*, vol. 24, no. 5, pp. 1458–1472, 2010.
- [12] H. Qiu, J. Lee, J. Lin, and G. Yu, "Robust performance degradation assessment methods for enhanced rolling element bearing prognostics," *Advanced Engineering Informatics*, vol. 17, no. 3-4, pp. 127–140, 2003.
- [13] D. Wang, W. Guo, and X. Wang, "A joint sparse wavelet coefficient extraction and adaptive noise reduction method in recovery of weak bearing fault features from a multi-component signal mixture," *Applied Soft Computing*, vol. 13, no. 10, pp. 4097–4104, 2013.
- [14] 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.
- [15] P. W. Tse and D. Wang, "The automatic selection of an optimal wavelet filter and its enhancement by the new sparsogram for bearing fault detection: part 2 of the two related manuscripts that have a joint title as 'Two automatic vibration-based fault diagnostic methods using the novel sparsity measurement—parts 1 and 2,'" *Mechanical Systems and Signal Processing*, vol. 40, no. 2, pp. 520–544, 2013.
- [16] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, IEEE, December 1995.

- [17] A. Banks, J. Vincent, and C. Anyakoha, "A review of particle swarm optimization. Part I: background and development," *Natural Computing*, vol. 6, no. 4, pp. 467–484, 2007.
- [18] A. Banks, J. Vincent, and C. Anyakoha, "A review of particle swarm optimization. Part II: hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications," *Natural Computing*, vol. 7, no. 1, pp. 109–124, 2008.
- [19] S.-F. Yuan and F.-L. Chu, "Fault diagnostics based on particle swarm optimisation and support vector machines," *Mechanical Systems and Signal Processing*, vol. 21, no. 4, pp. 1787–1798, 2007.
- [20] B. Samanta and C. Nataraj, "Use of particle swarm optimization for machinery fault detection," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 2, pp. 308–316, 2009.
- [21] B. Samanta and C. Nataraj, "Application of particle swarm optimization and proximal support vector machines for fault detection," *Swarm Intelligence*, vol. 3, no. 4, pp. 303–325, 2009.
- [22] N. G. Nlkolaou and I. A. Antoniadis, "Application of morphological operators as envelope extractors for impulsive-type periodic signals," *Mechanical Systems and Signal Processing*, vol. 17, no. 6, pp. 1147–1162, 2003.
- [23] J. Wang, G. Xu, Q. Zhang, and L. Liang, "Application of improved morphological filter to the extraction of impulsive attenuation signals," *Mechanical Systems and Signal Processing*, vol. 23, no. 1, pp. 236–245, 2009.
- [24] R. Hao, Z. Peng, Z. Feng, and F. Chu, "Application of support vector machine based on pattern spectrum entropy in fault diagnostics of rolling element bearings," *Measurement Science and Technology*, vol. 22, no. 4, Article ID 045708, 2011.
- [25] B. Li, P.-L. Zhang, Z.-J. Wang, S.-S. Mi, and Y.-T. Zhang, "Gear fault detection using multi-scale morphological filters," *Measurement*, vol. 44, no. 10, pp. 2078–2089, 2011.
- [26] B. Li, P.-L. Zhang, Z.-J. Wang, S.-S. Mi, and D.-S. Liu, "A weighted multi-scale morphological gradient filter for rolling element bearing fault detection," *ISA Transactions*, vol. 50, no. 4, pp. 599–608, 2011.
- [27] C. Li, M. Liang, Y. Zhang, and S. Hou, "Multi-scale autocorrelation via morphological wavelet slices for rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 31, pp. 428–446, 2012.
- [28] D. Wang, P. W. Tse, and Y. L. Tse, "A morphogram with the optimal selection of parameters used in morphological analysis for enhancing the ability in bearing fault diagnosis," *Measurement Science and Technology*, vol. 23, no. 6, Article ID 065001, 2012.
- [29] C. Shen, Q. He, F. Kong, and P. W. Tse, "A fast and adaptive varying-scale morphological analysis method for rolling element bearing fault diagnosis," *Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science*, vol. 227, no. 6, pp. 1362–1370, 2013.
- [30] S. Mallat, *A Wavelet Tour of Signal Processing: the Sparse Way*, Academic Press, Boston, Mass, USA, 2009.
- [31] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.
- [32] L. Zhang, J. Xu, J. Yang, D. Yang, and D. Wang, "Multiscale morphology analysis and its application to fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 22, no. 3, pp. 597–610, 2008.
- [33] X. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1077–1095, 2004.

## Research Article

# Fault Diagnosis for a Multistage Planetary Gear Set Using Model-Based Simulation and Experimental Investigation

**Guoyan Li, Fangyi Li, Yifan Wang, and Dehao Dong**

*Ministry of Education Key Laboratory of High-Efficiency and Clean Mechanical Manufacture, School of Mechanical Engineering, Shandong University, Jinan 250061, China*

Correspondence should be addressed to Fangyi Li; [lifangyi@sdu.edu.cn](mailto:lifangyi@sdu.edu.cn)

Received 27 April 2015; Revised 25 August 2015; Accepted 3 September 2015

Academic Editor: Pavan Kumar Kankar

Copyright © 2016 Guoyan 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.

The gear damage will induce modulation effects in vibration signals. A thorough analysis of modulation sidebands spectral structure is necessary for fault diagnosis of planetary gear set. However, the spectral characteristics are complicated in practice, especially for a multistage planetary gear set which contains close frequency components. In this study, a coupled lateral and torsional dynamic model is established to predict the modulation sidebands of a two-stage compound planetary gear set. An improved potential energy method is used to calculate the time-varying mesh stiffness of each gear pair, and the influence of crack propagation on the mesh stiffness is analyzed. The simulated signals of the gear set are obtained by using Runge-Kutta numerical analysis method. Meanwhile, the sidebands characteristics are summarized to exhibit the modulation effects caused by sun gear damage. At the end, the experimental signals collected from an industrial SD16 planetary gearbox are analyzed to verify the theoretical derivations. The results of experiment agree well with the simulated analysis.

## 1. Introduction

Remanufacturing is one of the best solutions to avoid resource shortage in the construction machinery industry. However, the quality of remanufactured products has long been an issue of concern. Planetary gearbox is widely used in large-scale and complex mechanical equipment such as wind turbines, helicopters, and construction machinery. The remanufacturing of planetary gearbox will create huge value for the economy development. The fault diagnosis of recycled gearboxes plays a vital role in guaranteeing the quality of remanufactured products. However, the recycled objects have experienced a service cycle, so the fault forms are complicated and various due to different service conditions and geometric structures. Therefore, the fault detection and damaged status assessment of recycled gearboxes become important research topics.

The vibration information can effectively reflect the running status of gearbox. Therefore, vibration-based fault diagnosis is the most effective technique in practice. The majority of research is focused on vibration signals analysis in recent years. The advanced signal processing methods

[1, 2] have been used to extract useful fault features [3] from vibration signals and assess the damage status. So far, the analysis of modulation sidebands has been an important and successful application in fault feature extraction. Feng et al. [4, 5] proposed the explicit equations to describe the signal models considering both the amplitude and frequency modulation due to gear damage and time-varying transfer paths and calculated the characteristic frequency of planetary gear set with local damage and distributed damage. Inalpolat et al. [6, 7] studied the amplitude modulation and frequency modulation effect due to manufacturing errors of planetary gear set through a mathematical model and dynamic model, and the amplitude modulation effect by carrier rotation was also taken account of.

Based on these researches, the vibration signals of planetary gear set have more complex spectral characteristics than fixed shaft gear trains. The damage on gear tooth surface such as wear, pitting, chipping, and cracking will induce the multiplicative amplitude modulation and frequency modulation to the meshing frequency and its harmonics. The time-varying transfer path will also result in a multiplicative amplitude modulation effect on the vibration signals. Furthermore,

the measured spectra will contain additional sidebands activity due to certain manufacturing errors, bearings damage, shafts deformations, and strong background noise. The varying operational conditions often lead to nonstationary and nonlinear signals, which make it more difficult to diagnose faults via vibration signals analysis.

Model-based dynamic analysis and simulation have been an effective method to analyze the vibration signals. More accurate dynamic models of planetary gear set have been developed in recent years considering more degrees of freedom, different nonlinear factors, damage, and so on. Lin and Parker [8] developed an analytical model of planetary gear set to investigate the natural frequency and vibration modes, which considers two translations and one rotation freedom of each compound. In another paper, Lin and Parker [9] studied the parametric instability caused by the mesh stiffness, which was modeled as rectangular waveform. The gear mesh stiffness is one of the main nonlinear parameters of the dynamic model. The sudden change of gear mesh stiffness due to gear damage will cause periodic impulses in time domain and induce the modulation phenomenon in frequency domain, which is important for fault diagnosis. Yang and Lin [10] proposed the potential energy method to calculate the time-varying mesh stiffness of a one-stage fixed shaft spur gear set, which was assumed to include three components: Hertzian contact energy, bending energy, and axial compressive energy. Later, Tian [11] redefined the potential energy method by taking the shear energy into consideration and investigated the influence of different types of gear damage on the mesh stiffness of a fixed shaft gear pair. Zhou et al. [12] and Chen and Shao [13] made the model more realistic by taking into account the deformation of gear body. Zhou et al. [12], Wu et al. [14, 15], and Tian et al. [16] investigated the effect of crack levels on the mesh stiffness based on the works of Tian [11] and used several statistical indicators to evaluate the changes induced by crack propagation in the simulated vibration signals. Liang et al. [17] extended the model to evaluate the mesh stiffness of an external-internal gear pair of the planetary gear set. The mesh stiffness model mentioned above was simplified as a cantilever beam on the base circle. Actually, the gear tooth starts from the root circle, and there is no uniform equation to describe the tooth curve between the basic circle and the root circle. Later, Liang et al. [18, 19] used straight lines to simplify this part and derived the new equations to calculate the mesh stiffness components. Wan et al. [20] also considered the potential energy stored in this part and analyzed the dynamic response of a single-stage fixed shaft spur gear system considering the influence of gear crack. More works about dynamic analysis of gear set have been reported considering more factors [20, 21], such as different gear damage, gear transmission error, bearing stiffness, and friction.

These results can provide useful information for dynamic analysis of damaged gear set. However, most of the works were limited to the simple and one-stage gear set. The gearbox in practice always has compound and multistage gear sets. In the vibration signals collected from a multistage gearbox, there are many close frequency components including

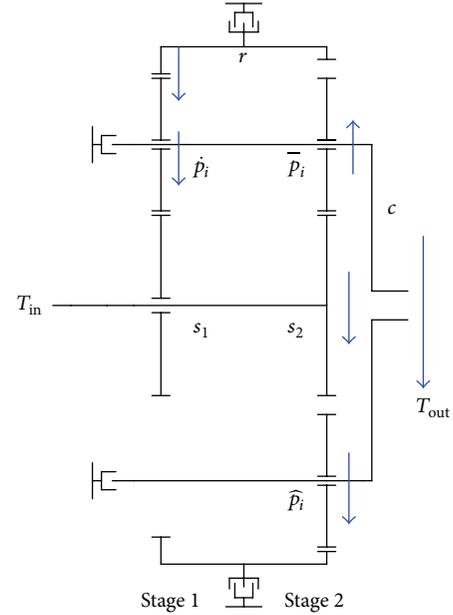


FIGURE 1: Transmission diagram of the two-stage planetary gear set.

different meshing frequencies and their harmonics. So the modulation sidebands will be more complicated and the fault features will be difficult to identify.

This paper is organized as follows: In Section 1, a review for model-based vibration signal analysis is introduced. In Section 2, a nonlinear dynamic model of a two-stage compound planetary gear set is established. Then, the mesh stiffness of each mesh pair is analytically evaluated using improved potential energy method. Further, the influence of crack propagation on the mesh stiffness is investigated when a crack appears in the sun gear of the second stage. In Section 3, the Runge-Kutta numerical analysis method is used to simulate the vibration response of the gear set in time domain. And also, the corresponding spectral analysis is applied on the simulated signals to obtain the sidebands caused by sun gear damage. The side frequencies are extracted as the useful fault features for fault diagnosis. Further, the influence of crack propagation on the vibration signal is analyzed. In Section 4, the experimental signals of an industrial SD16 planetary gearbox are analyzed to validate the dynamic model and the simulation results. Finally, conclusion and discussion are provided in Section 5. Further improvement will be provided in the future research.

## 2. Nonlinear Dynamic Model of a Two-Stage Planetary Gear Set

Figure 1 shows the structure of a two-stage planetary gear set. All the gears are standard involute spur tooth. The first stage ( $n = 1$ ) is a simple planetary gear set with  $N_1$  planets  $\hat{p}_i$  ( $N_1 = 3$ ). The second stage ( $n = 2$ ) is a meshed planet planetary gear set with  $N_2$  planet-planet mesh pairs  $\bar{p}_i - \hat{p}_i$  ( $N_2 = 3$ ). The planets of each stage distribute equidistantly and have the same parameters. The ring gear  $r$  and the carrier  $c$  are shared

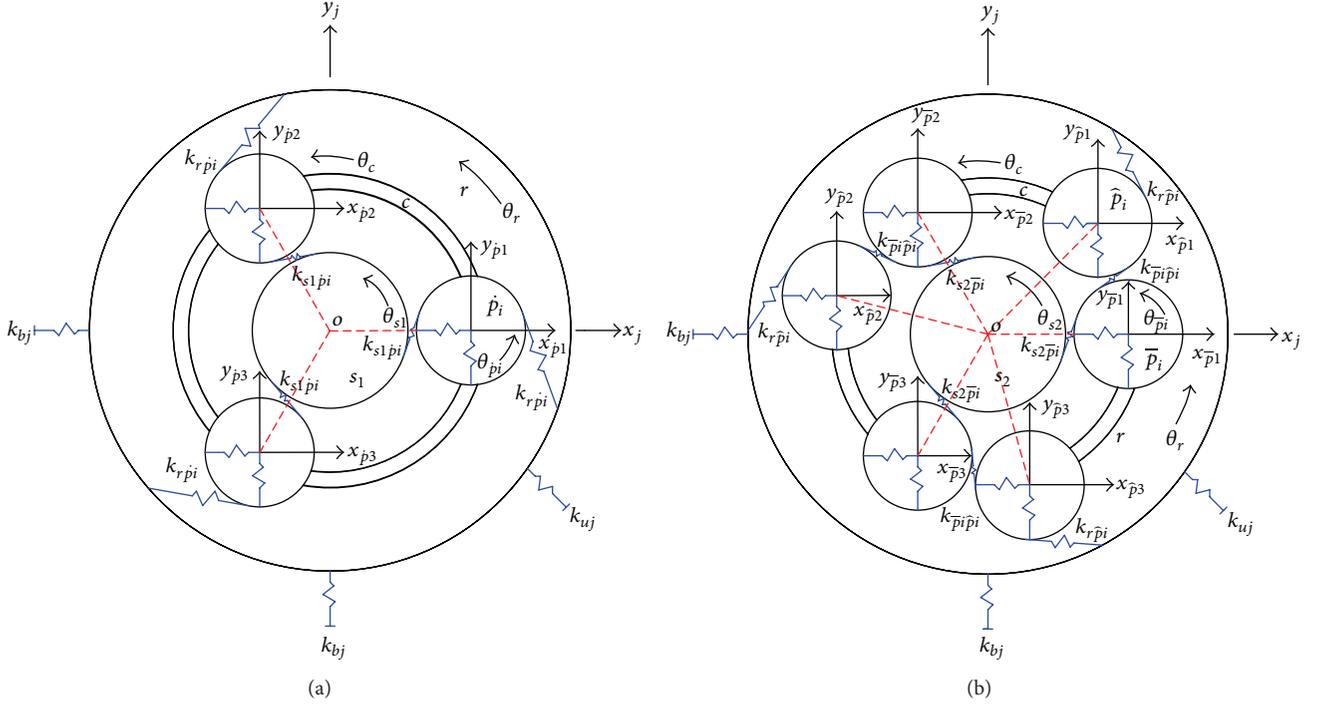


FIGURE 2: Dynamic models for (a) the first stage and (b) the second stage.

by both stages. In this structure, the sun gear  $s_1$  is fixed. The sun gear  $s_2$  and carrier  $c$  are identified as input member and output member.  $T_{in}$  and  $T_{out}$  are the input motor torque and output load torque.

The dynamic model of this two-stage planetary gear set is shown in Figure 2. A lumped-parameter model is considered in this study. The gears and carrier are treated as perfectly mounted rigid bodies with ideal geometries. The intertooth friction is ignored for simplicity. The gear mesh interactions are modeled by spring-damping structures acting along the line of action, and the periodically time-varying mesh stiffness  $k_{jpi}$  and viscous damping coefficient  $c_{jpi}$  are considered. The bearings are represented by spring-damping structures, and the bearing stiffness and damping coefficient are set to be constant. In this model, each component has three degrees of freedom: translations in  $x$  and  $y$  direction, denoted by  $x_j$  and  $y_j$ , and rotation around the axis  $\theta_j$  ( $j$  represents the center components  $c, r, s_1$ , and  $s_2$  and the planets  $p_i$  ( $\hat{p}_i, \bar{p}_i, \hat{p}_i$ )).

**2.1. Equations of Motion.** Define  $u_j = r_j \theta_j$  as the rotational coordinates, where  $r_j$  ( $j$  represents the center components  $r, s_1$ , and  $s_2$  and the planets  $p_i$  ( $\hat{p}_i, \bar{p}_i, \hat{p}_i$ )) are the base radii of the gears. According to the meshing relation, the relative gear mesh displacements  $\delta_{jpi}$  along the line of action of stage  $n$  are expressed as follows:

$$\begin{aligned} \delta_{s1\hat{p}i} = & -x_{s1} \sin \varphi_{s1\hat{p}i} + x_{\hat{p}i} \sin \varphi_{s1\hat{p}i} + y_{s1} \cos \varphi_{s1\hat{p}i} \\ & - y_{\hat{p}i} \cos \varphi_{s1\hat{p}i} + u_{s1} + u_{\hat{p}i}, \end{aligned}$$

$$\begin{aligned} \delta_{r\hat{p}i} = & -x_r \sin \varphi_{r\hat{p}i} + x_{\hat{p}i} \sin \varphi_{r\hat{p}i} + y_r \cos \varphi_{r\hat{p}i} \\ & - y_{\hat{p}i} \cos \varphi_{r\hat{p}i} + u_r - u_{\hat{p}i}, \end{aligned}$$

$$\begin{aligned} \delta_{s2\bar{p}i} = & -x_{s2} \sin \varphi_{s2\bar{p}i} + x_{\bar{p}i} \sin \varphi_{s2\bar{p}i} + y_{s2} \cos \varphi_{s2\bar{p}i} \\ & - y_{\bar{p}i} \cos \varphi_{s2\bar{p}i} + u_{s2} + u_{\bar{p}i}, \end{aligned}$$

$$\begin{aligned} \delta_{\bar{p}i\hat{p}i} = & x_{\bar{p}i} \sin \varphi_{\bar{p}i\hat{p}i} - x_{\hat{p}i} \sin \varphi_{\bar{p}i\hat{p}i} - y_{\bar{p}i} \cos \varphi_{\bar{p}i\hat{p}i} \\ & + y_{\hat{p}i} \cos \varphi_{\bar{p}i\hat{p}i} - u_{\bar{p}i} - u_{\hat{p}i}, \end{aligned}$$

$$\begin{aligned} \delta_{r\bar{p}i} = & x_r \sin \varphi_{r\bar{p}i} - x_{\bar{p}i} \sin \varphi_{r\bar{p}i} - y_r \cos \varphi_{r\bar{p}i} \\ & + y_{\bar{p}i} \cos \varphi_{r\bar{p}i} - u_r + u_{\bar{p}i}, \end{aligned}$$

(1)

where  $\varphi_{jpi} = \varphi_{pi} - \delta_j \alpha$ ,  $\delta_j = -1$  for internal  $r - \hat{p}_i$  mesh pair and external  $\bar{p}_i - \hat{p}_i$  mesh pair, and  $\delta_j = 1$  for internal  $r - \bar{p}_i$  mesh pair and external  $s_1 - \hat{p}_i$  and  $s_2 - \bar{p}_i$  mesh pairs.  $\varphi_{pi}$  is the spacing angle of the planets.  $\alpha$  is the pressure angle.

The dynamic gear mesh force is defined as follows:

$$F_{jpi} = c_{jpi} \dot{\delta}_{jpi} + k_{jpi} \delta_{jpi}; \quad (2)$$

the bearing forces are defined as follows:

$$\begin{aligned} F_{bjx} &= c_{bj} \dot{x}_j + k_{bj} x_j, \\ F_{bjy} &= c_{bj} \dot{y}_j + k_{bj} y_j, \\ F_{uj} &= c_{uj} \dot{u}_j + k_{uj} u_j, \end{aligned} \quad (3)$$

where  $c_{jpi}$  is the mesh damping coefficient,  $k_{jpi}$  is the periodically time-varying mesh stiffness,  $c_{bj}$  is the damping coefficient in radial direction,  $k_{bj}$  is the radial stiffness of the bearing,  $c_{uj}$  is the damping coefficient in torsional direction, and  $k_{uj}$  is the torsional stiffness of the bearing.

The equations of motion for sun gear can be derived as follows:

$$\begin{aligned}
m_{s1}\ddot{x}_{s1} - \sum_{i=1}^N \sin \varphi_{s1\bar{p}i} F_{s1\bar{p}i} + F_{bs1x} &= 0, \\
m_{s1}\ddot{y}_{s1} + \sum_{i=1}^N \cos \varphi_{s1\bar{p}i} F_{s1\bar{p}i} + F_{bs1y} &= 0, \\
\frac{I_{s1}}{r_{s1}^2} \ddot{u}_{s1} + \sum_{i=1}^N F_{s1\bar{p}i} + F_{us1} &= 0, \\
m_{s2}\ddot{x}_{s2} - \sum_{i=1}^N \sin \varphi_{s2\bar{p}i} F_{s2\bar{p}i} + F_{bs2x} &= 0, \\
m_{s2}\ddot{y}_{s2} + \sum_{i=1}^N \cos \varphi_{s2\bar{p}i} F_{s2\bar{p}i} + F_{bs2y} &= 0, \\
\frac{I_{s2}}{r_{s2}^2} \ddot{u}_{s2} + \sum_{i=1}^N F_{s2\bar{p}i} + F_{us2} &= \frac{T_{in}}{r_{s2}}.
\end{aligned} \tag{4}$$

The equations of motion for ring gear are written as follows:

$$\begin{aligned}
m_r \ddot{x}_r + \sum_{i=1}^N \sin \varphi_{r\bar{p}i} F_{r\bar{p}i} - \sum_{i=1}^N \sin \varphi_{r\bar{p}i} F_{r\bar{p}i} + F_{brx} &= 0, \\
m_r \ddot{y}_r - \sum_{i=1}^N \cos \varphi_{r\bar{p}i} F_{r\bar{p}i} + \sum_{i=1}^N \cos \varphi_{r\bar{p}i} F_{r\bar{p}i} + F_{bry} &= 0, \\
\frac{I_r}{r_r^2} \ddot{u}_r - \sum_{i=1}^N F_{r\bar{p}i} + \sum_{i=1}^N F_{r\bar{p}i} + F_{ur} &= 0.
\end{aligned} \tag{5}$$

The equations of motion for planets are given as follows:

$$\begin{aligned}
m_{\bar{p}i} \ddot{x}_{\bar{p}i} + \sin \varphi_{s1\bar{p}i} F_{s1\bar{p}i} + \sin \varphi_{r\bar{p}i} F_{r\bar{p}i} &= 0, \\
m_{\bar{p}i} \ddot{y}_{\bar{p}i} - \cos \varphi_{s1\bar{p}i} F_{s1\bar{p}i} - \cos \varphi_{r\bar{p}i} F_{r\bar{p}i} &= 0, \\
\frac{I_{\bar{p}i}}{r_{\bar{p}i}^2} \ddot{u}_{\bar{p}i} + F_{s1\bar{p}i} - F_{r\bar{p}i} &= 0, \\
m_{\bar{p}i} \ddot{x}_{\bar{p}i} + \sin \varphi_{\bar{p}i\bar{p}i} F_{\bar{p}i\bar{p}i} + \sin \varphi_{s2\bar{p}i} F_{s2\bar{p}i} &= 0, \\
m_{\bar{p}i} \ddot{y}_{\bar{p}i} - \cos \varphi_{\bar{p}i\bar{p}i} F_{\bar{p}i\bar{p}i} - \cos \varphi_{s2\bar{p}i} F_{s2\bar{p}i} &= 0, \\
\frac{I_{\bar{p}i}}{r_{\bar{p}i}^2} \ddot{u}_{\bar{p}i} + F_{s2\bar{p}i} - F_{\bar{p}i\bar{p}i} &= 0,
\end{aligned}$$

$$\begin{aligned}
m_{\bar{p}i} \ddot{x}_{\bar{p}i} - \sin \varphi_{\bar{p}i\bar{p}i} F_{\bar{p}i\bar{p}i} - \sin \varphi_{r\bar{p}i} F_{r\bar{p}i} &= 0, \\
m_{\bar{p}i} \ddot{y}_{\bar{p}i} + \cos \varphi_{\bar{p}i\bar{p}i} F_{\bar{p}i\bar{p}i} + \cos \varphi_{r\bar{p}i} F_{r\bar{p}i} &= 0, \\
\frac{I_{\bar{p}i}}{r_{\bar{p}i}^2} \ddot{u}_{\bar{p}i} - F_{\bar{p}i\bar{p}i} + F_{r\bar{p}i} &= 0,
\end{aligned} \tag{6}$$

where  $m_j$  and  $I_j$  represent the mass and mass moment of inertia of the gears.

For the carrier-planet subsystem,  $u_c = r_c \theta_c$  is defined as the coordinate in place of  $\theta_c$ , where  $r_c$  is the radius from the carrier center to the planet center. The relative displacements between carrier and planet in  $x_c$  and  $y_c$  directions are given as follows:

$$\begin{aligned}
\delta_{cpix} &= x_c - x_{pi} - \sin \varphi_{pi} u_c, \\
\delta_{cpiy} &= y_c - y_{pi} + \cos \varphi_{pi} u_c.
\end{aligned} \tag{7}$$

The differential equations of motion for carrier and planets are expressed as follows:

$$\begin{aligned}
m_c \ddot{x}_c + \sum_{i=1}^N (c_{bpi} \dot{\delta}_{cpix} + k_{bpi} \delta_{cpix}) + (c_{bc} \dot{x}_c + k_{bc} x_c) &= 0, \\
m_c \ddot{y}_c + \sum_{i=1}^N (c_{bpi} \dot{\delta}_{cpiy} + k_{bpi} \delta_{cpiy}) + (c_{bc} \dot{y}_c + k_{bc} y_c) &= 0, \\
\frac{I_c}{r_c^2} - \sum_{i=1}^N (c_{bpi} \dot{\delta}_{cpix} + k_{bpi} \delta_{cpix}) \sin \varphi_{pi} \\
+ \sum_{i=1}^N (c_{bpi} \dot{\delta}_{cpiy} + k_{bpi} \delta_{cpiy}) \cos \varphi_{pi} \\
+ (c_{uc} \dot{u}_c + k_{uc} u_c) &= -\frac{T_{out}}{r_c},
\end{aligned} \tag{8}$$

$$m_{pi} \ddot{x}_{pi} - (c_{bpi} \dot{\delta}_{cpix} + k_{bpi} \delta_{cpix}) = 0,$$

$$m_{pi} \ddot{y}_{pi} - (c_{bpi} \dot{\delta}_{cpiy} + k_{bpi} \delta_{cpiy}) = 0,$$

where  $m_c$  and  $I_c$  represent the mass and mass moment of inertia of the carrier.

Thus, the overall equations of motion for the two-stage planetary gear set can be constructed systematically in matrix form as follows:

$$\mathbf{M} \ddot{\mathbf{Q}}(t) + (\mathbf{C}_m + \mathbf{C}_b) \dot{\mathbf{Q}}(t) + (\mathbf{K}_m + \mathbf{K}_b) \mathbf{Q}(t) = \mathbf{T}, \tag{9}$$

where

$$\mathbf{M} = \text{diag} \left[ m_j \quad m_j \quad \frac{I_j}{r_j^2} \right],$$

$$\mathbf{Q} = [x_j \quad y_j \quad u_j]^T,$$

$$\mathbf{K}_m = \begin{bmatrix} \mathbf{K}_{cp} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{cp1} & \mathbf{K}_{cp2} & \mathbf{K}_{cp3} & \mathbf{K}_{c\bar{p}1} & \mathbf{K}_{c\bar{p}2} & \mathbf{K}_{c\bar{p}3} & \mathbf{K}_{c\bar{p}1} & \mathbf{K}_{c\bar{p}2} & \mathbf{K}_{c\bar{p}3} \\ & \mathbf{K}_{rp} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{rp1} & \mathbf{K}_{rp2} & \mathbf{K}_{rp3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{r\bar{p}1} & \mathbf{K}_{r\bar{p}2} & \mathbf{K}_{r\bar{p}3} \\ & & \mathbf{K}_{s1p} & \mathbf{0} & \mathbf{K}_{s1\bar{p}1} & \mathbf{K}_{s1\bar{p}2} & \mathbf{K}_{s1\bar{p}3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & \mathbf{K}_{s2p} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{s2\bar{p}1} & \mathbf{K}_{s2\bar{p}2} & \mathbf{K}_{s2\bar{p}3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & & \mathbf{K}_{p1} & \mathbf{0} \\ & & & & & \mathbf{K}_{p2} & \mathbf{0} \\ & & & & & & \mathbf{K}_{p3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & & & & & \mathbf{K}_{\bar{p}1} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{\bar{p}1\bar{p}1} & \mathbf{0} & \mathbf{0} \\ & & & & & & & & \mathbf{K}_{\bar{p}2} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{\bar{p}2\bar{p}2} & \mathbf{0} \\ & & & & & & & & & \mathbf{K}_{\bar{p}3} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{\bar{p}3\bar{p}3} \\ & & & & & & & & & & \mathbf{K}_{\bar{p}1} & \mathbf{0} & \mathbf{0} \\ & & & & & & & & & & & \mathbf{K}_{\bar{p}2} & \mathbf{0} \\ & & & & & & & & & & & & \mathbf{K}_{\bar{p}3} \end{bmatrix}, \quad (10)$$

$$\mathbf{K}_b = \text{diag} [\mathbf{K}_{bc} \quad \mathbf{K}_{br} \quad \mathbf{K}_{bs1} \quad \mathbf{K}_{bs2} \quad \mathbf{0} \quad \cdots \quad \mathbf{0}],$$

$$\mathbf{C}_b = \text{diag} [\mathbf{C}_{bc} \quad \mathbf{C}_{br} \quad \mathbf{C}_{bs1} \quad \mathbf{C}_{bs2} \quad \mathbf{0} \quad \cdots \quad \mathbf{0}].$$

In these equations,  $\mathbf{M}$  is the inertia matrix,  $\mathbf{Q}$  is the displacement matrix,  $\mathbf{C}_m$  is the mesh damping matrix,  $\mathbf{C}_b$  is the bearing damping matrix,  $\mathbf{K}_m$  is the periodically time-varying mesh stiffness matrix, the submatrices of  $\mathbf{K}_m$  are given in Appendix,  $\mathbf{K}_b$  is the bearing stiffness matrix, and  $\mathbf{T}$  is the applied external torque matrix.

All the parameters of this gear set are provided by the cooperative enterprise according to an industrial SD16 planetary gearbox and are listed in Table 1.

**2.2. Mesh Stiffness Evaluation.** In this section, the mesh stiffness of each mesh pair for this two-stage gear set will be calculated, respectively, based on the improved model proposed by Liang et al. [18]. The mesh stiffness model is simplified as a cantilever beam on the root circle. The potential energy method is used to analyze the mesh stiffness, which is assumed to include four components: Hertzian contact energy, bending energy, shear energy, and axial compressive energy. According to the design parameters of this gear set, the root circles of gears are smaller than the base circles for the external-external mesh pairs. The tooth curve between the base circle and the root circle is not an involute curve and there is no uniform equation to describe this part. So this part is simplified as straight lines [18], as shown in Figure 3.

Thus, for the single-tooth pair meshing duration, the total mesh stiffness can be calculated as follows:

$$k_{jpi} = \frac{1}{1/k_h + 1/k_{b1} + 1/k_{s1} + 1/k_{a1} + 1/k_{b2} + 1/k_{s2} + 1/k_{a2}}, \quad (11)$$

where subscripts 1 and 2 represent the driving gear and driven gear.  $k_h$ ,  $k_b$ ,  $k_s$ , and  $k_a$  represent the Hertzian contact stiffness, bending stiffness, shear stiffness, and axial compressive stiffness, respectively.

For the double-tooth pair mesh duration, the total mesh stiffness can be expressed as follows:

$$k_{jpi} = \sum_{i=1}^2 \frac{1}{1/k_{h,i} + 1/k_{b1,i} + 1/k_{s1,i} + 1/k_{a1,i} + 1/k_{b2,i} + 1/k_{s2,i} + 1/k_{a2,i}}, \quad (12)$$

where  $i = 1$  represents the first pair of meshing teeth and  $i = 2$  represents the second. The detail of derivations of each stiffness component was given in Liang et al. [18].

It is essential to correctly define the phasing relationships between each mesh pair. References [17–19, 22] provided an analytical calculation of the relative phase for a one-stage and simple planetary gear set. Extending the methods proposed in [17–19, 22] and combining the parameters listed in Table 1, the relative phases of this two-stage planetary gear set can be calculated. Figure 4 shows a meshing sketch for this gear set and only one of the planet branches is shown for simplicity purpose. The spacing angles of  $\bar{p}_1$  and  $\bar{p}_1$  are set to be 0.  $\bar{N}_i M_i$  and  $\bar{B}_i E_i$  ( $i = 1, 2, 3, \dots$ ) denote the theoretical meshing line and practical action line.  $P_i$  ( $i = 1, 2, 3, \dots$ ) is the pitch point of each mesh pair. As shown in Figure 1, the sun gear  $s_1$  is fixed and  $s_2$  is defined as the input member. Assuming the sun gear  $s_2$  rotates in the counter-clockwise direction, the planets  $\bar{p}_1$ ,  $\bar{p}_2$ , and  $\bar{p}_3$  are in clockwise, counter-clockwise, and counter-clockwise rotation directions, respectively.

TABLE 1: Main parameters of the gear set.

Parameters	Sun gear		Planet	Ring gear		Carrier
	$s_1$	$s_2$	$P_i (\hat{P}_i, \bar{P}_i, \hat{P}_i)$	$r$		$c$
Number of teeth	30	30	21	72	78	—
Module (mm)	3.5	3.5	3.5	3.5		—
Pressure angle (deg)	20	20	20	20		—
Pitch diameter (mm)	105	105	73.5	252	273	—
Base diameter (mm)	98.668	98.668	69.067	236.803	256.536	—
Addendum circle diameter (mm)	112	112	80.5	245.734	266.677	—
Root diameter (mm)	96.25	96.25	64.75	260.75	281.75	—
Width of teeth (mm)	40.5	40.5	38.5	92		—
Material	20 NiCrMoH	20 NiCrMoH	20 NiCrMoH	42 CrMo		ZG40Mn2
Young's modulus (Pa)	$2.06 \times 10^{11}$	$2.06 \times 10^{11}$	$2.06 \times 10^{11}$	$2.12 \times 10^{11}$		$2.02 \times 10^{11}$
Poisson's ratio	0.30	0.30	0.30	0.28		0.30
Mass (kg)	2.11	2.11	0.92	11.89		15.54
Mass moment of inertia (kg·mm <sup>2</sup> )	$0.360 \times 10^4$	$0.360 \times 10^4$	$0.081 \times 10^4$	$24.095 \times 10^4$		$12.729 \times 10^4$
Input rotation speed (rpm)			700			
Input motor torque (N·m)			96			
Output load torque (N·m)			200			
Transmission ratio			2.08			
Radial stiffness of the bearing (N/m)			$10^8$			
Damping coefficient of the bearing (N·s/m)			$10^5$			

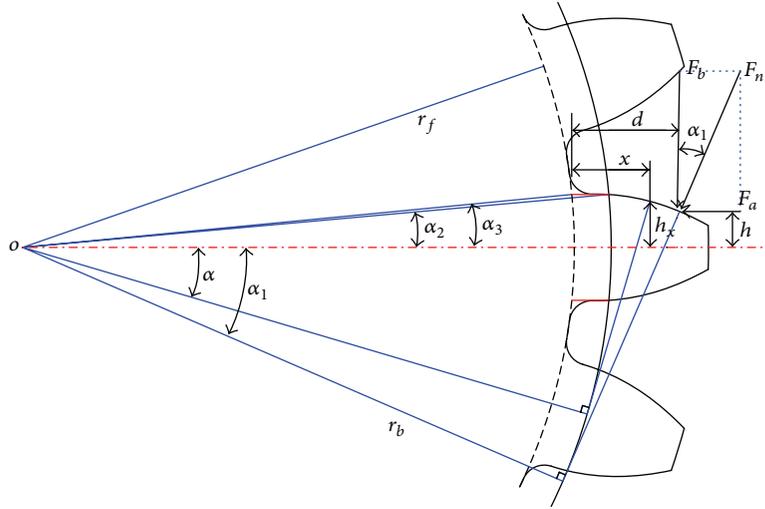


FIGURE 3: The mesh stiffness beam model for healthy gear.

Table 2 shows the relative phases of this gear set. The superscripts represent the first stage and the second stage of the gear set, respectively.  $\gamma'_{sn}$  and  $\gamma''_{sn}$  denote the relative phase between the  $n$ th sun-planet mesh pair and the first sun-planet mesh pair.  $\gamma'_{rn}$  and  $\gamma''_{rn}$  are the relative phase between the  $n$ th ring-planet mesh pair and the first ring-planet mesh pair.  $\gamma''_{pn}$  is the relative phase between the  $n$ th planet-planet mesh pair and the first planet-planet mesh pair.  $\gamma'_{rs}$  and  $\gamma''_{rs}$  are the relative phase between the  $n$ th ring-planet mesh pair and sun-planet mesh pair.  $\gamma''_{ps}$  is the relative phase between the  $n$ th planet-planet mesh pair and sun-planet mesh pair. “+” dictate the phase lag, and “-” dictate the phase lead.

We use MATLAB programs to obtain the numerical values of total mesh stiffness as a function of time  $t$ , which are plotted in Figures 5 and 6. It is observed that the mesh stiffness varies periodically. There exists a discontinuous point in a mesh period, which is the transition point from double-tooth pair to single-tooth pair. For the first stage, the mean stiffness values are  $1.676 \times 10^9$  N/m for sun-planet pair and  $1.677 \times 10^9$  N/m for ring-planet pair.  $\gamma'_{rs} = 0$  indicates the  $n$ th sun-planet mesh pair and the ring-planet mesh pair mesh at the pitch point simultaneously. For the second stage, the mean stiffness values are  $1.676 \times 10^9$  N/m for sun-planet pair,  $1.567 \times 10^9$  N/m for planet-planet pair, and  $1.573 \times 10^9$  N/m for

TABLE 2: Relative phases of the gear set.

Stage 1	$\gamma'_{s1}$	$\gamma'_{s2}$	$\gamma'_{s3}$	$\gamma'_{r1}$	$\gamma'_{r2}$	$\gamma'_{r3}$	$\gamma'_{rs}$				
	0	0	0	0	0	0	0				
Stage 2	$\gamma''_{s1}$	$\gamma''_{s2}$	$\gamma''_{s3}$	$\gamma''_{p1}$	$\gamma''_{p2}$	$\gamma''_{p3}$	$\gamma''_{r1}$	$\gamma''_{r2}$	$\gamma''_{r3}$	$\gamma''_{ps}$	$\gamma''_{rs}$
	0	0	0	0	0	0	0	0	0	-0.13	+0.38

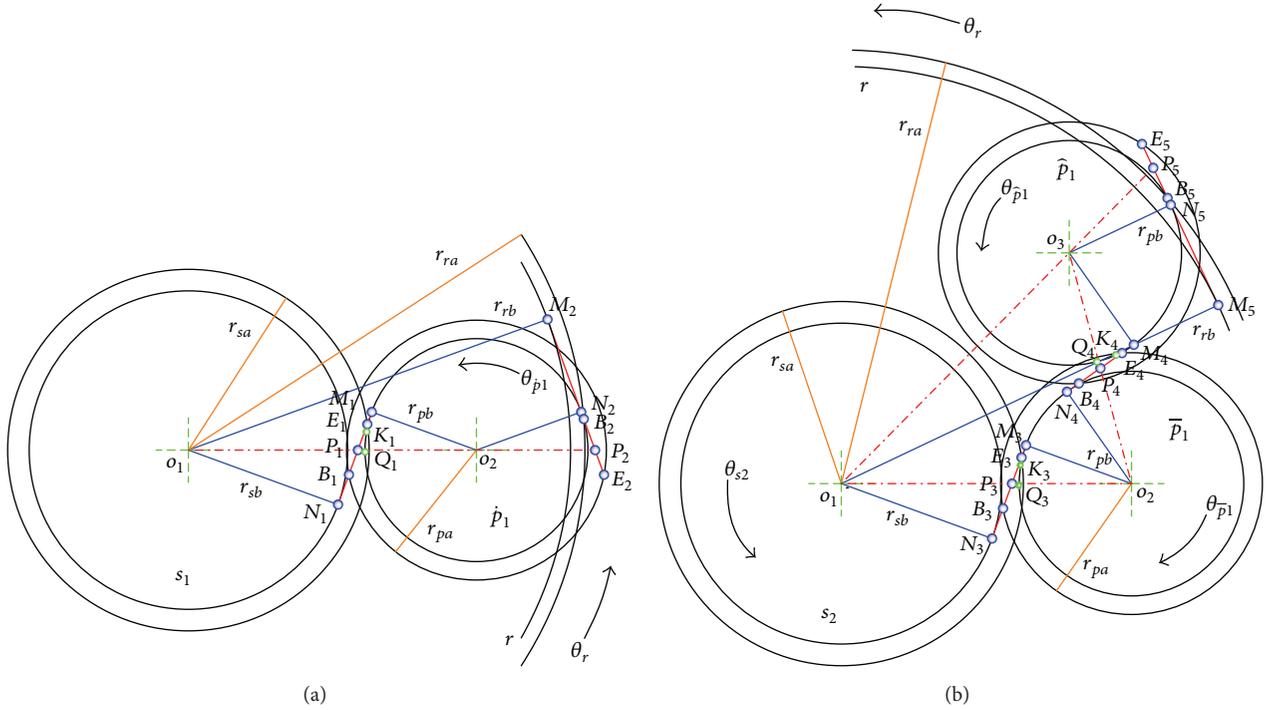


FIGURE 4: A meshing sketch for the two-stage planetary gear set: (a) the first stage and (b) the second stage.

ring-planet pair. The phase difference between the  $n$ th planet-planet mesh pair and sun-planet mesh pair is  $\gamma''_{ps} T_m$  in time, where  $T_m$  is the mesh period of the second stage. Similarly, the phase difference between the  $n$ th ring-planet mesh pair and sun-planet mesh pair can be expressed as  $\gamma''_{rs} T_m$  in time.

Liang et al. [18] assumed the crack propagation path to be a straight line starting from the critical area of the tooth. As shown in Figure 7, the crack propagates along the straight line until it reaches the central line. The intersection angle  $\nu$  between the crack path and the central line is set to  $45^\circ$ . Then, the crack changes its direction until a sudden breakage appears. The Hertzian and axial compressive stiffness remain the same; however, the bending and shear stiffness will change due to the crack. In this section, the influence of crack on the mesh stiffness of a sun/planet pair in the second stage is investigated when the crack appears in the sun gear. Based on the crack model provided above, five crack levels are evaluated: 10%, 30%, 50%, 70%, and 80% cracks. When the crack line  $q_1$  reaches the central line, the crack level is set to 50%, and when a sudden break occurs, the corresponding crack level is 80%. Table 3 shows the five crack levels and corresponding crack length in the sun gear  $s_2$ .

The total mesh stiffness curves of the sun-planet pair in second stage with different crack lengths are shown in

TABLE 3: Crack levels and corresponding crack length in the sun gear of second stage.

	Crack levels	Crack length (mm)
1	10%	$q_1 = 0.942$
2	30%	$q_1 = 2.826$
3	50%	$q_1 = 4.71$
4	70%	$q_1 = 4.71, q_2 = 1.884$
5	80%	Tooth missing

Figure 8. It can be observed that the mesh stiffness greatly decreases when the cracked tooth is in meshing. The influences spread over the mating duration of the cracked gear, which corresponds to 0.0089 s. As the size of crack grows, the mesh stiffness reduces correspondingly. When the sudden break appears, the mean mesh stiffness reduces to 42% of the original. This is important for dynamic analysis of the gear set.

The mesh damping coefficient is set to be proportional to the total mesh stiffness; that is,

$$c_{jpi} = \mu_{jpi} k_{jpi} \quad (13)$$

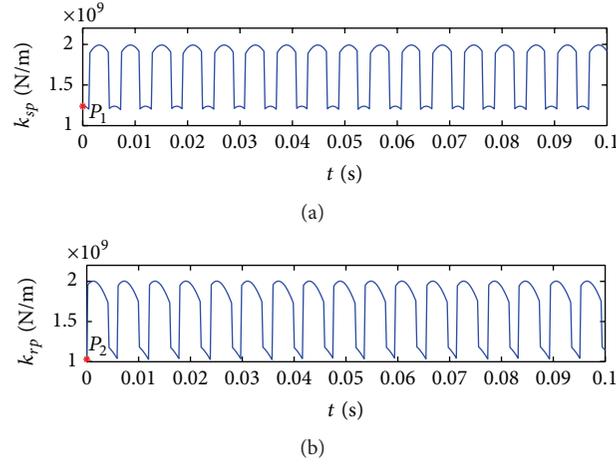


FIGURE 5: Total mesh stiffness of the perfect first stage versus time  $t$  within 0.1 s: (a) sun-planet pair and (b) ring-planet pair.

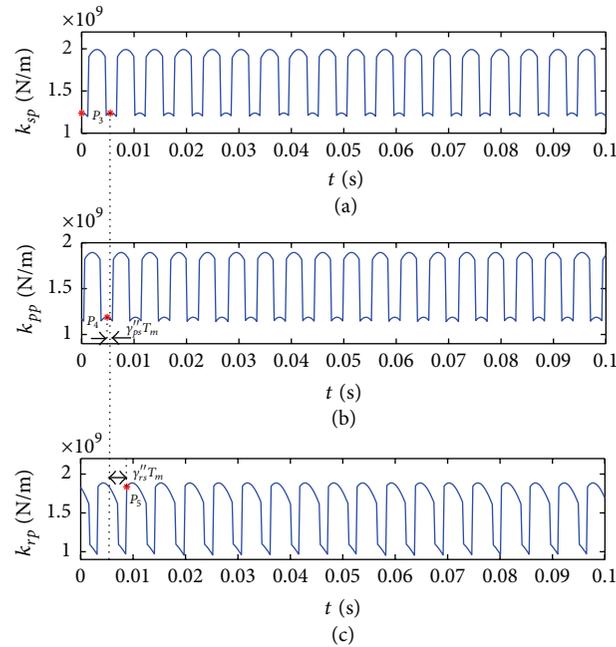


FIGURE 6: Total mesh stiffness of the perfect second stage versus time  $t$  within 0.1 s: (a) sun-planet pair, (b) planet-planet pair, and (c) ring-planet pair.

where  $\mu_{jpi}$  is the scale constant measured in seconds. For this gear set,  $\mu_{s1\bar{p}i} = \mu_{s2\bar{p}i} = 2.70 \times 10^{-6}$  s,  $\mu_{r\bar{p}i} = 2.40 \times 10^{-6}$  s,  $\mu_{r\bar{p}i} = 3.16 \times 10^{-6}$  s, and  $\mu_{r\bar{p}i} = 3.26 \times 10^{-6}$  s.

### 3. Analysis of the Simulated Signals

To illustrate the dynamic response of the system, the MATLAB's `ode15s` function is used to solve the equations of motion derived above. Then, the vibration responses of each component of this gear set as a function of time are numerically simulated. Applying the `fft` function based on the Fast Fourier Transform algorithm provided by MATLAB toolbox, the corresponding Fourier spectrum can be

obtained. According to the parameters provided in Table 1, the dominant frequencies of this gear set are estimated within 2000 Hz. Considering the sample frequency options set in the experimental equipment, the sample frequency used in this simulation is selected as 5120 Hz.

The vibration velocity responses of healthy sun gears  $s_1$  and  $s_2$  are shown, respectively, in Figure 9. The periodical impulses caused by time-varying mesh stiffness are obvious in the vibration signals. And the duration between every two impulses of  $s_1$  is equal to 0.00606 s, while the impulse duration of  $s_2$  is 0.00547 s. Figure 10 shows the Fourier spectrum of healthy gear set. As expected, the spectrum comprises four frequency components: the fundamental meshing frequency of first stage ( $f_1 = 168$  Hz), the fundamental

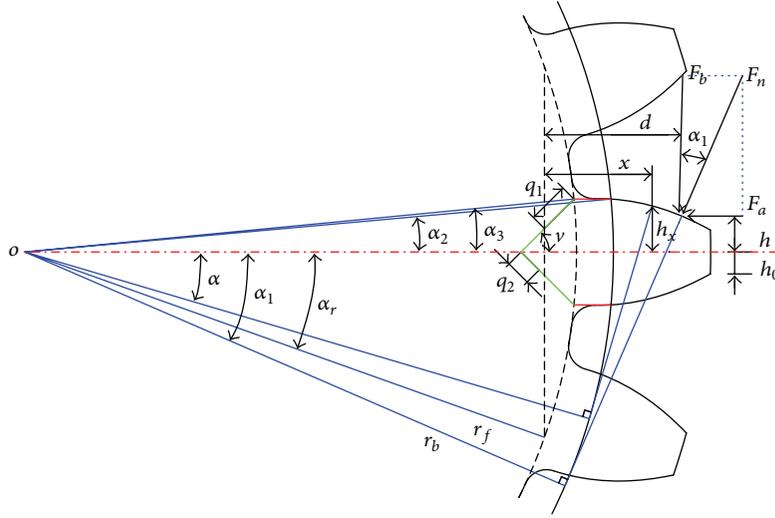


FIGURE 7: The cantilever beam model for local damaged gear.

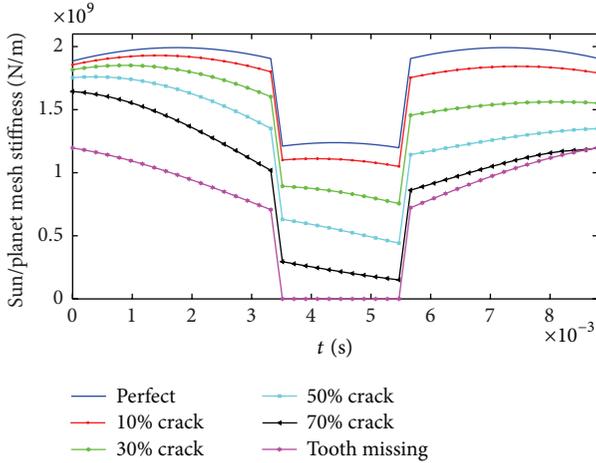


FIGURE 8: The mesh stiffness of sun-planet pair with different crack levels.

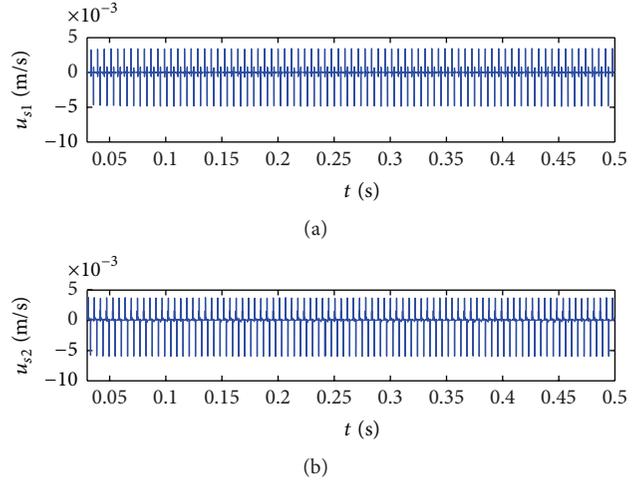


FIGURE 9: Vibration velocity responses of healthy sun gears: (a) sun gear  $s_1$  and (b) sun gear  $s_2$ .

meshing frequency of second stage ( $f_2 = 182$  Hz), and their harmonics ( $kf_1, kf_2$  ( $k = 1, 2, \dots$ )), in which the frequency component  $8f_2$  is dominant. So, for the two-stage planetary gear set, the meshing frequencies of each stage are very close to each other, which can be expressed as follows:

$$f_{\text{mesh}} = \frac{z_{\text{sun}} n_{\text{sun-carrier}}}{60}; \quad (14)$$

in this equation,  $z_{\text{sun}}$  is the number of sun gear teeth and  $n_{\text{sun-carrier}}$  is the rotation speed of sun gear relative to the carrier.

Supposing a 70% level ( $q_1 = 4.71, q_2 = 1.884$ ) crack appears on the sun gear of the first stage, the dynamic response of sun gear  $s_1$  and the corresponding Fourier spectrum of the gear set are shown in Figure 11, respectively. Compared with Figure 9(a), when the cracked sun gear contacts with the mating perfect planet, a sudden change will occur in time domain signal. As the meshing of damaged

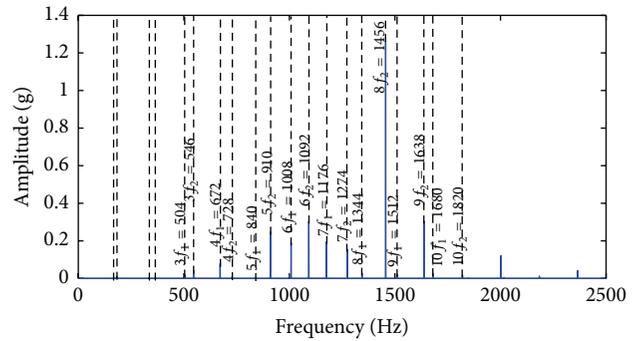


FIGURE 10: Fourier spectrum of healthy gear set.

sun gear, the impulse appears at a time period of 0.060 s. Compared with Figure 10, many spectral peaks appear as sidebands around the meshing frequency components. And

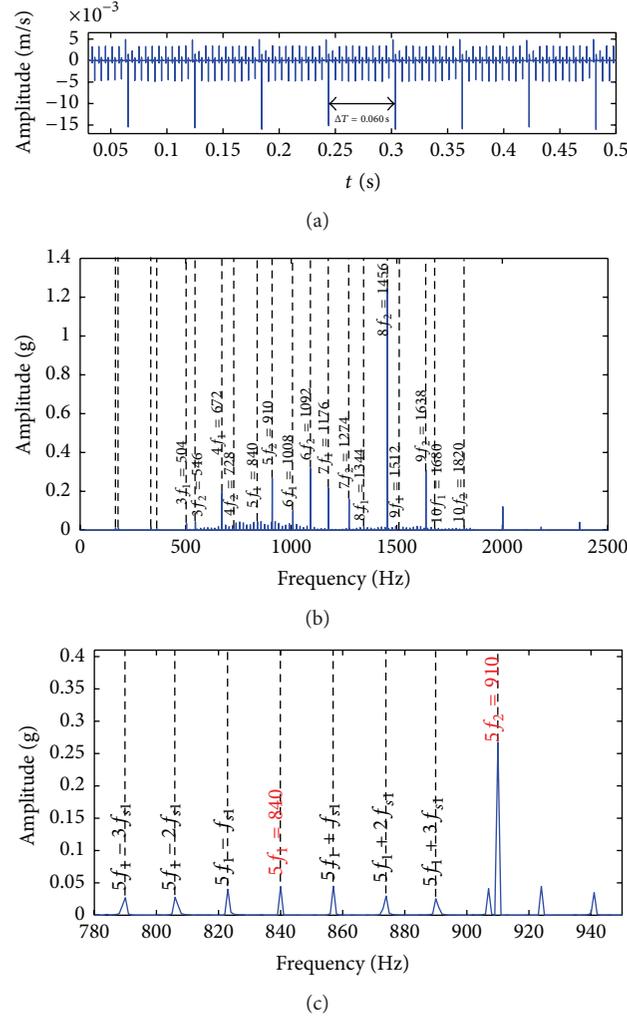


FIGURE 11: Vibration response of the gear set with 70% level crack on sun gear  $s_1$ : (a) waveform of cracked sun gear  $s_1$ , (b) Fourier spectrum, and (c) enlarged Fourier spectrum.

the amplitudes of these sidebands are small compared with the meshing frequencies. In order to analyze the details, the spectrum is enlarged in Figure 11(c). It is observed that the sidebands center around the frequency components  $kf_1$  ( $k = 1, 2, \dots$ ), at the frequencies of  $kf_1 \pm mf_{s1}$  ( $k = 1, 2, \dots$ ,  $m = 1, 2, \dots$ ), and  $f_{s1}$  is equal to  $16.8$  Hz. So the side frequency components  $mf_{s1}$  ( $m = 1, 2, \dots$ ) indicate that the gear set have fault, and the fault exists on the sun gear  $s_1$ .

Similarly, when the 70% level ( $q_1 = 4.71$ ,  $q_2 = 1.884$ ) crack exists on the surface of sun gear in the second stage, the contact of cracked sun gear tooth with mating planets leads to obvious shocks at a repeating period, which is  $\Delta T = 0.055$  s, as shown in Figure 12(a). The Fourier spectrum of the gear set with cracked sun gear  $s_2$ , as shown in Figure 12(b), has almost the same frequency components with Figure 11(b): the fundamental meshing frequency of the first stage ( $f_1 = 168$  Hz), the fundamental meshing frequency of the second stage ( $f_2 = 182$  Hz), their harmonics ( $kf_1, kf_2$  ( $k = 1, 2, \dots$ )), and the side frequencies. To illustrate the details, the spectrum is enlarged in Figure 12(c). It is observed that

almost all the side peaks appear at the frequencies of  $kf_2 \pm mf_{s2}$  ( $k = 1, 2, \dots$ ,  $m = 1, 2, \dots$ ), and  $f_{s2}$  is  $18.2$  Hz. As expected, the side frequency components  $mf_{s2}$  ( $m = 1, 2, \dots$ ) contain the fault information, and the fault exists on the sun gear  $s_2$ .

As described in Section 2, the crack grows from 10% (in the early stage) to 70% (close to a sudden breakage) with an increment percentage of 20% ( $1.884$  mm). The enlarged Fourier spectrum of the gear set with different crack levels are shown in Figure 13. Clearly, the influence of the damaged sun gear is not obvious, and the amplitudes of sidebands are too small to be identified easily when the crack is in the early stage (the crack level is lower than 10%). As the crack level increases, the amplitude of sidebands becomes stronger, but the sideband structures remain the same.

According to the analysis provided above, the contact of damaged gear tooth with mating gears will lead to periodical impulses in time domain, and the impulse period of each gear component is different and related with the system parameters. Further, the shocks will induce the multiplicative

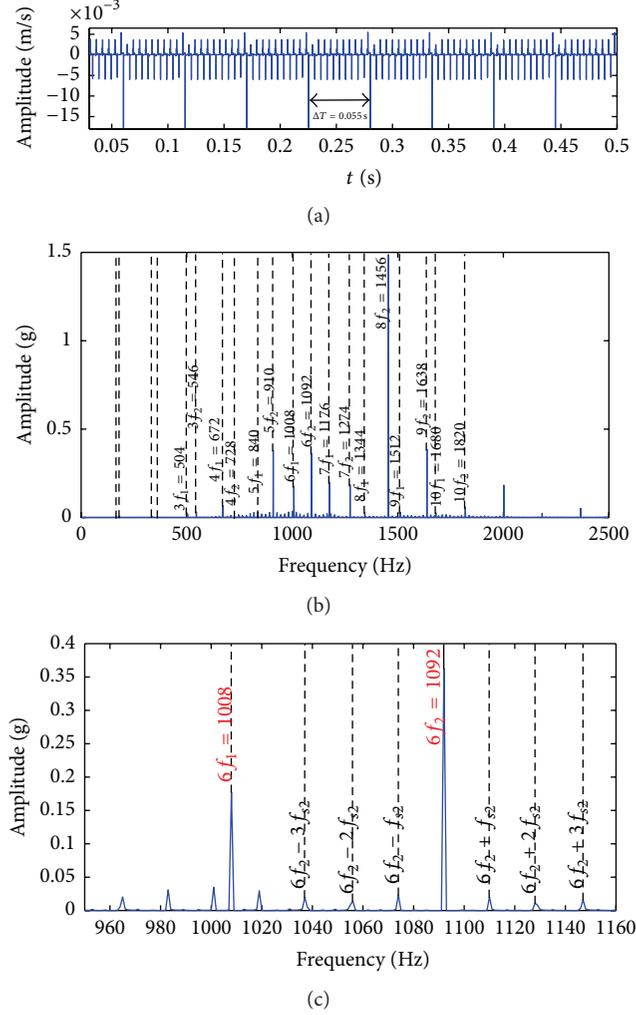


FIGURE 12: Vibration response of the gear set with 70% level crack on sun gear  $s_2$ : (a) waveform of cracked sun gear  $s_2$ , (b) Fourier spectrum, and (c) enlarged Fourier spectrum.

modulation to the meshing frequencies in frequency domain and lead to the sidebands center around the corresponding meshing frequency components. These side frequency components can provide useful information for accurately locating and reliably assessing the damage status of the system. Given the parameters and transmission structure of the planetary gear set, the fault characteristic frequency of sun gear can be derived as follows:

$$f_{\text{sun}} = \frac{N_{\text{planet}} n_{\text{sun-carrier}}}{60}. \quad (15)$$

According to (14), the equation can also be written as follows:

$$f_{\text{sun}} = \frac{N_{\text{planet}} f_{\text{mesh}}}{z_{\text{sun}}}, \quad (16)$$

where  $N_{\text{planet}}$  is the number of planets mating with corresponding sun gear.

## 4. Experimental Results

In this section, we use three experimental signals to validate the dynamic model provided above and fault characteristics of the multistage planetary gear set obtained from the simulation. One is the vibration signal collected from a new industrial SD16 planetary gearbox, which is considered as the baseline signal. The second one is the signal collected from the same gearbox with naturally broken sun gear  $s_1$ . The third one is the signal collected from the same gearbox with naturally broken sun gear  $s_2$ .

**4.1. Experimental Specification.** The planetary gearbox test rig is shown in Figure 14 and the SD16 planetary gearbox used in the experiments is shown in Figure 15(a). The fault experiments are done on the second row of this gearbox, as shown in Figure 15(b), which has the same structure and parameters as the two-stage model provided in Section 2. A three-axis accelerometer of model KD1002S is mounted on the gearbox casing. The data and vibration signals are

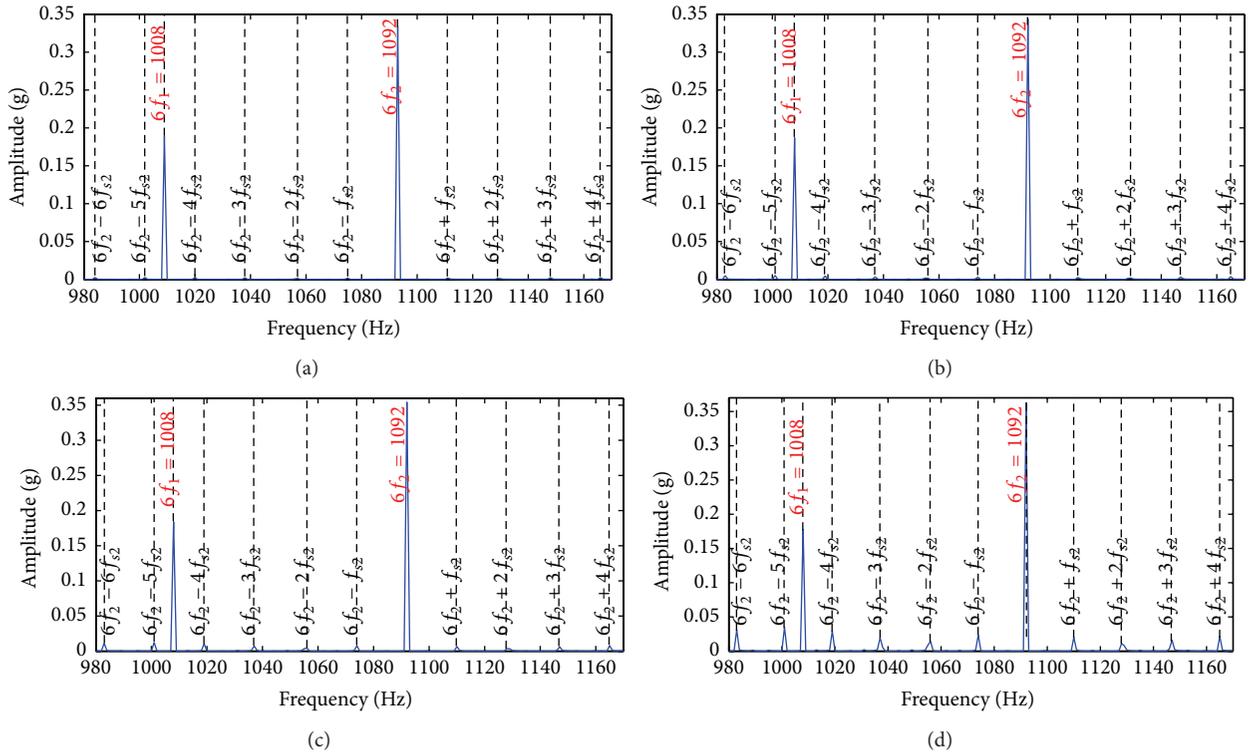
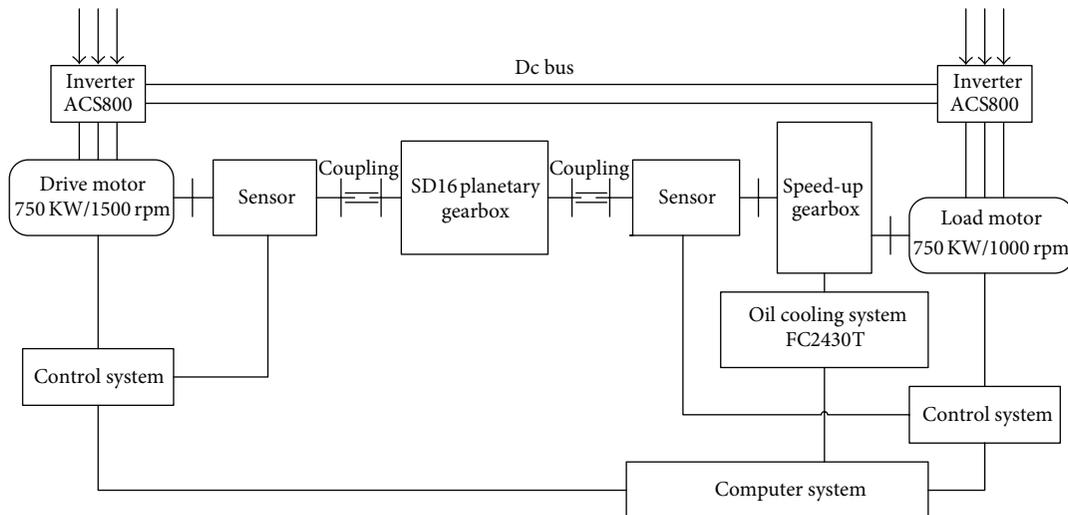


FIGURE 13: Enlarged Fourier spectrum of the gear set when the crack levels are 10%, 30%, 50%, and 70%.



(a)



(b)

FIGURE 14: (a) Planetary gearbox test rig and (b) structure diagram of the test rig.

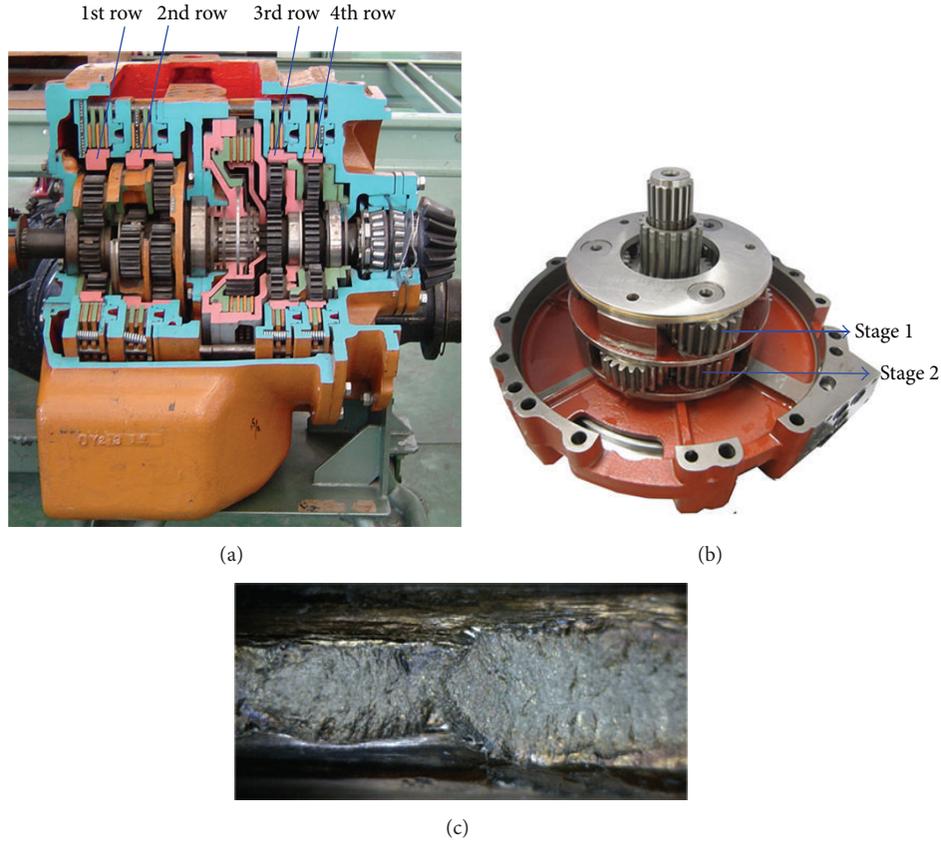


FIGURE 15: (a) SD16 planetary gearbox, (b) the second row of this gearbox, and (c) sun gear with broken tooth.

measured by CRAS vibration signal analysis system. A 12.8 s span data is recorded with a sampling frequency of 5120 Hz under each experimental condition. During the experiments, the input speed is set to 700 rpm, and a load of 200 N·m is applied to the output shaft.

For the damaged sun gear  $s_1$  experiment, the perfect sun gear of the first stage ( $s_1$ ) is replaced by the broken one to mimic the common fault status of the gearbox, as shown in Figure 15(c). Because the fault characteristic frequencies of  $s_1$  are  $mf_{s_1}$  ( $m = 1, 2, 3, \dots$ ) in simulation, these useful fault features are expected to be detected. For the damaged sun gear  $s_2$  experiment, the perfect sun gear of the second stage ( $s_2$ ) is replaced by a broken one, whereas all the other parts are normal. The fault characteristic frequencies of  $s_2$ ,  $mf_{s_2}$  ( $m = 1, 2, 3, \dots$ ) in simulation, are expected to be extracted.

**4.2. Normal Vibration Signal Analysis.** For a long time span, the input speed and load will fluctuate around the set value during the experiment. This will make the measured Fourier spectral lines deviate the theoretical values dramatically. However, from a short time view of point, the signal can be assumed to be relative stationary [4]. For this reason, we extract 1 s long signal samples from the 12.8 s span data to analyze the Fourier spectral structure, so as to obtain more accurate spectral lines.

The waveform and corresponding Fourier spectrum of the baseline signal sample collected from new SD16 planetary gearbox are shown in Figure 16. The measured spectrum

contains the fundamental meshing frequency of the first stage ( $f_1$ ), about 167.3 Hz, the fundamental meshing frequency of the second stage ( $f_2$ ), about 180.6 Hz, and their harmonics ( $kf_1, kf_2$  ( $k = 1, 2, 3, \dots$ )). Some sidebands appear around the main frequency components, but most of them are irregular and the amplitudes are small. So this does not indicate that the gearbox is damaged.

**4.3. Damaged Sun Gear Signal Analysis.** Figure 17 shows the waveform and Fourier spectrum of the signal sample (the sample length is 1 s) from the same gearbox with naturally broken sun gear  $s_1$ , respectively. Compared with Figure 16(a), the kurtosis indicator of the signal sample increases from 3.01 to 4.04. Several impulses are visible in the time domain, but the accurate impulse period is difficult to evaluate due to strong back noise. The dominant spectral peaks appear at the frequency components: the fundamental meshing frequency of the first stage ( $f_1$ ), the fundamental meshing frequency of the second stage ( $f_2$ ), their harmonics ( $kf_1, kf_2$  ( $k = 1, 2, 3, \dots$ )), and the sidebands, as the baseline spectrum, but the magnitude of the spectral peaks increases significantly. All the signs indicate the anomalies of the gearbox. As described in the simulation, the sidebands are associated with the transient impulses and contain useful information of damage location. So the zoomed-in Fourier spectrum is shown in Figure 17(c). It can be observed that the sidebands appear around the meshing frequencies of the first stage, at the frequencies of  $kf_1 \pm mf_{s_1}$  ( $k = 1, 2, \dots$ ,  $m = 1, 2, \dots$ ).

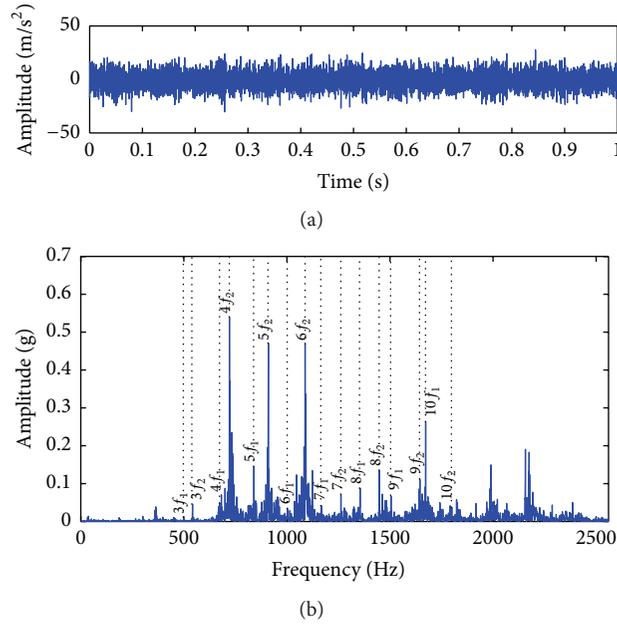


FIGURE 16: Waveform and corresponding Fourier spectrum of the baseline signal collected from the new SD16 planetary gearbox.

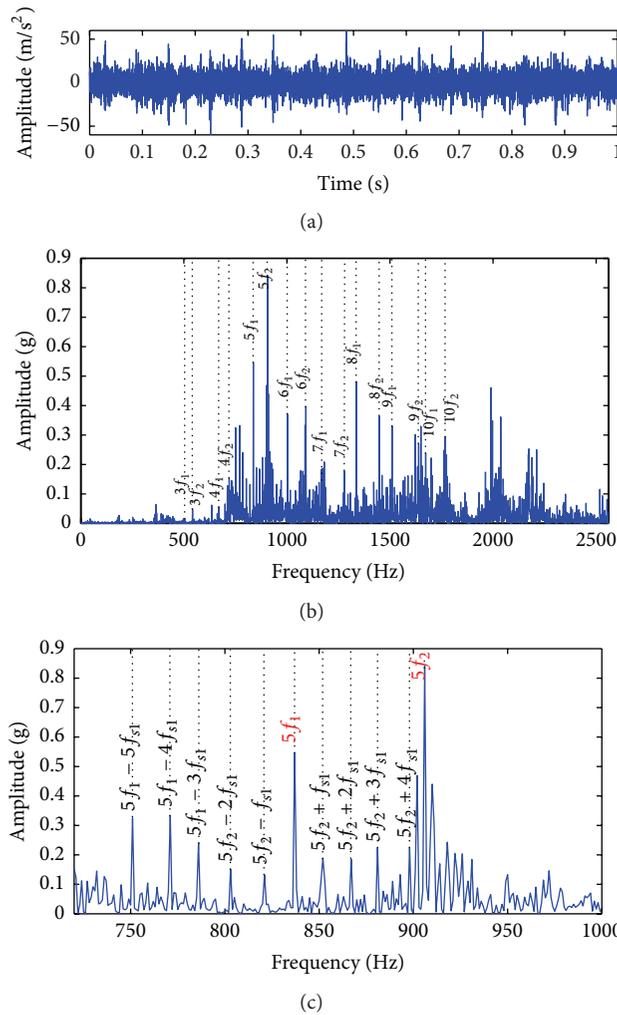


FIGURE 17: Vibration signal from the gearbox with naturally broken sun gear  $s_1$ ; (a) waveform, (b) Fourier spectrum, and (c) zoomed-in Fourier spectrum.

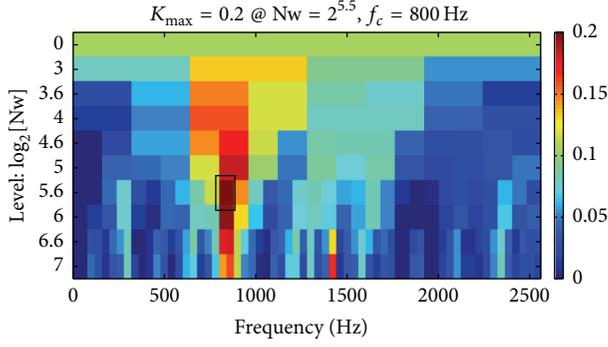


FIGURE 18: Spectral kurtosis for the signal sample in Figure 17.

To further validate the fault features, the spectral kurtosis signal processing method [23] is employed in this section to extract the transient impulses. Figure 18 presents the spectral kurtosis for the signal sample (the sample length is 5 s). The maximum spectral kurtosis appears at level 5.5 for the central frequency of 800 Hz. The filtered fault signal from the optimal band-pass filter around 800 Hz and the corresponding squared envelope spectrum of the filtered signal are shown in Figure 19. The periodic impulses are obvious in time domain, and a pattern of 0.062 s is confirmed. Frequency of 16.23 Hz and its harmonics exist in the frequency domain, which correspond with the calculation of (16) for the fault characteristic frequency of the sun gear in the first stage  $mf_{s_1}$  ( $m = 1, 2, 3, \dots$ ). So we can conclude that the gear  $s_1$  has fault. The conclusion matches the results obtained from the simulation.

Figure 20 shows the waveform and Fourier spectrum of the signal sample (the sample length is 1 s) from the broken sun gear  $s_2$  case. The kurtosis value of the signal sample increases from 3.01 to 5.24. More sidebands appear in the frequency band from 500 to 1000 Hz, and most of them have bigger amplitudes than the baseline signal. As shown in Figure 20(c), the sidebands center around the meshing frequencies of the second stage, at the frequencies of  $kf_2 \pm mf_{s_2}$  ( $k = 1, 2, \dots, m = 1, 2, \dots$ ). The corresponding spectral kurtosis is presented in Figure 21. The maximum spectral kurtosis appears at level 6.5 for the central frequency of 80 Hz. The filtered fault signal from the optimal band-pass filter around 80 Hz and the corresponding squared envelope spectrum of the filtered signal are shown in Figure 22. The periodic impulses are evident, and a pattern of 0.056 s is confirmed. Frequency of 17.67 Hz and its harmonics exist in the frequency domain, which correspond with the calculation of (16) for the fault characteristic frequency of the sun gear in the second stage  $mf_{s_2}$  ( $m = 1, 2, 3, \dots$ ). All the signs indicate that the gearbox is fault, and the damage occurs on the sun gear of the second stage. The results match well with the simulation.

## 5. Conclusions

Model-based dynamic analysis has been an important application for identifying and extracting useful fault features from vibration signals. In this study, the nonlinear dynamic model

of a two-stage planetary gear set is established. Based on the improved potential energy method, the time-varying mesh stiffness of each mesh pair is calculated, and the influence of crack propagation on the mesh stiffness of sun-planet mesh pair is presented. Then, the simulated signals of the gear set are analyzed in time and frequency domain. According to the sidebands structure, the equations of fault characteristic frequencies of sun gear components are derived. These results provide a theoretical reference for diagnosing faults via modulation sidebands analysis in practice. Further, the experimental signals collected from a SD16 planetary gearbox are analyzed, and the spectral kurtosis is employed to extract the fault characteristic frequencies of sun gear from the processed signals. In future research, more suitable signal processing methods will be developed to improve the vibration test signal from the field, and more useful fault features will be extracted to accurately locate the gear damage and reliably assess the damage status.

## Appendix

Consider the following:

$$\mathbf{K}_{cp} = \sum_{i=1}^N k_{bpi} \begin{bmatrix} 3 & 0 & -\sin \varphi_{\hat{p}i} - \sin \varphi_{\hat{p}i} - \sin \varphi_{\hat{p}i} \\ & 3 & \cos \varphi_{\hat{p}i} + \cos \varphi_{\hat{p}i} + \cos \varphi_{\hat{p}i} \\ \text{sym} & & 3 \end{bmatrix},$$

$$\mathbf{K}_{cpi} = \sum_{i=1}^N k_{bpi} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ \sin \varphi_{\hat{p}i} & -\cos \varphi_{\hat{p}i} & 0 \end{bmatrix},$$

$$\mathbf{K}_{rp}$$

$$= \sum_{i=1}^N k_{r\hat{p}i} \begin{bmatrix} \sin^2 \varphi_{r\hat{p}i} & -\sin \varphi_{r\hat{p}i} \cos \varphi_{r\hat{p}i} & -\sin \varphi_{r\hat{p}i} \\ & \cos^2 \varphi_{r\hat{p}i} & \cos \varphi_{r\hat{p}i} \\ \text{sym} & & 1 \end{bmatrix}$$

$$+ \sum_{i=1}^N k_{r\hat{p}i} \begin{bmatrix} \sin^2 \varphi_{r\hat{p}i} & -\sin \varphi_{r\hat{p}i} \cos \varphi_{r\hat{p}i} & -\sin \varphi_{r\hat{p}i} \\ & \cos^2 \varphi_{r\hat{p}i} & \cos \varphi_{r\hat{p}i} \\ \text{sym} & & 1 \end{bmatrix},$$

$$\mathbf{K}_{r\hat{p}i} = k_{r\hat{p}i} \begin{bmatrix} -\sin^2 \varphi_{r\hat{p}i} & \sin \varphi_{r\hat{p}i} \cos \varphi_{r\hat{p}i} & \sin \varphi_{r\hat{p}i} \\ & -\cos^2 \varphi_{r\hat{p}i} & -\cos \varphi_{r\hat{p}i} \\ \text{sym} & & -1 \end{bmatrix},$$

$$\mathbf{K}_{r\hat{p}i} = k_{r\hat{p}i} \begin{bmatrix} -\sin^2 \varphi_{r\hat{p}i} & \sin \varphi_{r\hat{p}i} \cos \varphi_{r\hat{p}i} & \sin \varphi_{r\hat{p}i} \\ & -\cos^2 \varphi_{r\hat{p}i} & -\cos \varphi_{r\hat{p}i} \\ \text{sym} & & -1 \end{bmatrix},$$

$$\mathbf{K}_{s1p}$$

$$= \sum_{i=1}^N k_{s1\hat{p}i} \begin{bmatrix} \sin^2 \varphi_{s1\hat{p}i} & -\sin \varphi_{s1\hat{p}i} \cos \varphi_{s1\hat{p}i} & -\sin \varphi_{s1\hat{p}i} \\ & \cos^2 \varphi_{s1\hat{p}i} & \cos \varphi_{s1\hat{p}i} \\ \text{sym} & & 1 \end{bmatrix},$$

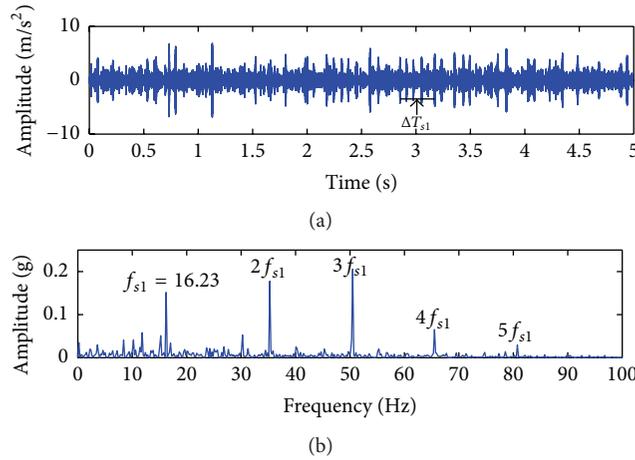


FIGURE 19: (a) Filtered fault signal from the optimal band-pass filter in the case of the damaged sun gear  $s_1$  and (b) squared envelope spectrum of the filtered signal.

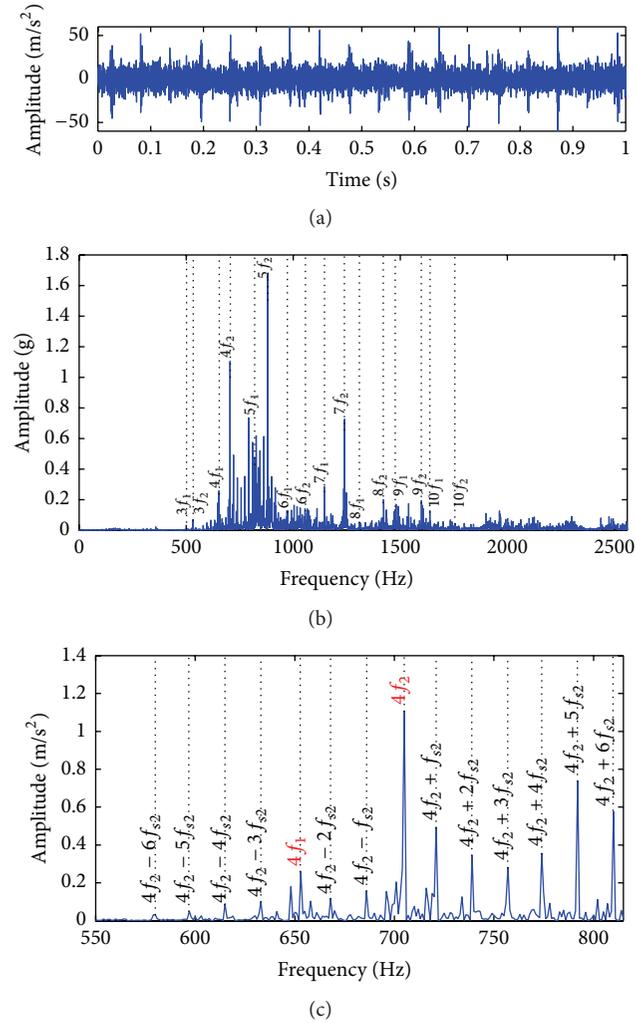


FIGURE 20: Vibration signal from the gearbox with naturally broken sun gear  $s_2$ : (a) waveform, (b) Fourier spectrum, and (c) zoomed-in Fourier spectrum.

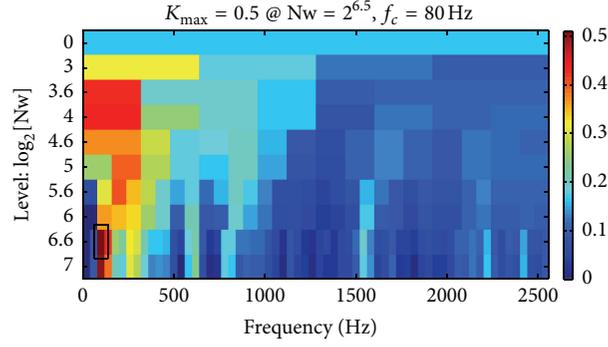
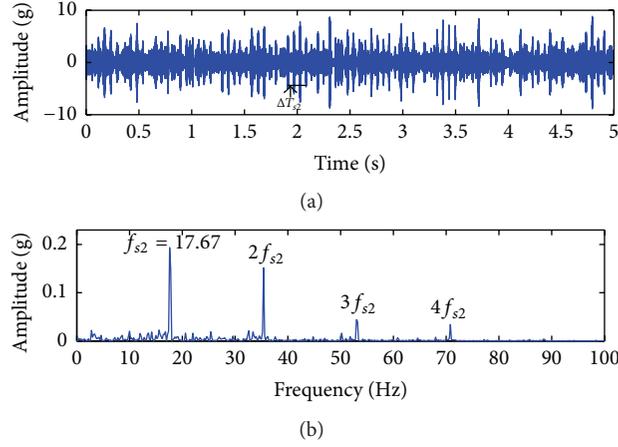


FIGURE 21: Spectral kurtosis for the signal sample in Figure 20.

FIGURE 22: (a) Filtered fault signal from the optimal band-pass filter in the case of the damaged sun gear  $s_2$  and (b) squared envelope spectrum of the filtered signal.

$$\begin{aligned}
 & \mathbf{K}_{s1\dot{p}i} \\
 &= k_{s1\dot{p}i} \begin{bmatrix} -\sin^2 \varphi_{s1\dot{p}i} & \sin \varphi_{s1\dot{p}i} \cos \varphi_{s1\dot{p}i} & -\sin \varphi_{s1\dot{p}i} \\ \sin \varphi_{s1\dot{p}i} \cos \varphi_{s1\dot{p}i} & -\cos^2 \varphi_{s1\dot{p}i} & \cos \varphi_{s1\dot{p}i} \\ \sin \varphi_{s1\dot{p}i} & -\cos \varphi_{s1\dot{p}i} & 1 \end{bmatrix}, \\
 & \mathbf{K}_{s2p} \\
 &= \sum_{i=1}^N k_{s2\bar{p}i} \begin{bmatrix} \sin^2 \varphi_{s2\bar{p}i} & -\sin \varphi_{s2\bar{p}i} \cos \varphi_{s2\bar{p}i} & -\sin \varphi_{s2\bar{p}i} \\ \text{sym} & \cos^2 \varphi_{s2\bar{p}i} & \cos \varphi_{s2\bar{p}i} \\ & & 1 \end{bmatrix}, \\
 & \mathbf{K}_{s2\bar{p}i} \\
 &= k_{s2\bar{p}i} \begin{bmatrix} -\sin^2 \varphi_{s2\bar{p}i} & \sin \varphi_{s2\bar{p}i} \cos \varphi_{s2\bar{p}i} & -\sin \varphi_{s2\bar{p}i} \\ \sin \varphi_{s2\bar{p}i} \cos \varphi_{s2\bar{p}i} & -\cos^2 \varphi_{s2\bar{p}i} & \cos \varphi_{s2\bar{p}i} \\ \sin \varphi_{s2\bar{p}i} & -\cos \varphi_{s2\bar{p}i} & 1 \end{bmatrix}, \\
 & \mathbf{K}_{\dot{p}i} \\
 &= k_{s1\dot{p}i} \begin{bmatrix} \sin^2 \varphi_{s1\dot{p}i} & -\sin \varphi_{s1\dot{p}i} \cos \varphi_{s1\dot{p}i} & \sin \varphi_{s1\dot{p}i} \\ \text{sym} & \cos^2 \varphi_{s1\dot{p}i} & -\cos \varphi_{s1\dot{p}i} \\ & & 1 \end{bmatrix} \\
 & \quad + k_{r\dot{p}i} \begin{bmatrix} \sin^2 \varphi_{r\dot{p}i} & -\sin \varphi_{r\dot{p}i} \cos \varphi_{r\dot{p}i} & -\sin \varphi_{r\dot{p}i} \\ \text{sym} & \cos^2 \varphi_{r\dot{p}i} & \cos \varphi_{r\dot{p}i} \\ & & 1 \end{bmatrix} \\
 & \quad + k_{b\dot{p}i} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\
 & \mathbf{K}_{\bar{p}i} \\
 &= k_{s2\bar{p}i} \begin{bmatrix} \sin^2 \varphi_{s2\bar{p}i} & -\sin \varphi_{s2\bar{p}i} \cos \varphi_{s2\bar{p}i} & \sin \varphi_{s2\bar{p}i} \\ \text{sym} & \cos^2 \varphi_{s2\bar{p}i} & -\cos \varphi_{s2\bar{p}i} \\ & & 1 \end{bmatrix} \\
 & \quad + k_{\bar{p}i\dot{p}i} \begin{bmatrix} \sin^2 \varphi_{\bar{p}i\dot{p}i} & -\sin \varphi_{\bar{p}i\dot{p}i} \cos \varphi_{\bar{p}i\dot{p}i} & -\sin \varphi_{\bar{p}i\dot{p}i} \\ \text{sym} & \cos^2 \varphi_{\bar{p}i\dot{p}i} & \cos \varphi_{\bar{p}i\dot{p}i} \\ & & 1 \end{bmatrix} \\
 & \quad + k_{b\bar{p}i} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},
 \end{aligned}$$

$$\begin{aligned}
& \mathbf{K}_{\widehat{p}i} \\
&= k_{r\widehat{p}i} \begin{bmatrix} \sin^2 \varphi_{r\widehat{p}i} & -\sin \varphi_{r\widehat{p}i} \cos \varphi_{r\widehat{p}i} & -\sin \varphi_{r\widehat{p}i} \\ & \cos^2 \varphi_{r\widehat{p}i} & \cos \varphi_{r\widehat{p}i} \\ \text{sym} & & 1 \end{bmatrix} \\
&+ k_{\widehat{p}i\widehat{p}i} \begin{bmatrix} \sin^2 \varphi_{\widehat{p}i\widehat{p}i} & -\sin \varphi_{\widehat{p}i\widehat{p}i} \cos \varphi_{\widehat{p}i\widehat{p}i} & \sin \varphi_{\widehat{p}i\widehat{p}i} \\ & \cos^2 \varphi_{\widehat{p}i\widehat{p}i} & -\cos \varphi_{\widehat{p}i\widehat{p}i} \\ \text{sym} & & 1 \end{bmatrix} \\
&+ k_{b\widehat{p}i} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\
& \mathbf{K}_{\widehat{p}i\widehat{p}i} \\
&= k_{\widehat{p}i\widehat{p}i} \begin{bmatrix} -\sin^2 \varphi_{\widehat{p}i\widehat{p}i} & \sin \varphi_{\widehat{p}i\widehat{p}i} \cos \varphi_{\widehat{p}i\widehat{p}i} & -\sin \varphi_{\widehat{p}i\widehat{p}i} \\ \sin \varphi_{\widehat{p}i\widehat{p}i} \cos \varphi_{\widehat{p}i\widehat{p}i} & -\cos^2 \varphi_{\widehat{p}i\widehat{p}i} & \cos \varphi_{\widehat{p}i\widehat{p}i} \\ \sin \varphi_{\widehat{p}i\widehat{p}i} & -\cos \varphi_{\widehat{p}i\widehat{p}i} & 1 \end{bmatrix}. \tag{A.1}
\end{aligned}$$

## Conflict of Interests

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

## Acknowledgments

This research is supported by the National High Technology Research and Development Program of China (no. 2013AA040204) and the National Key Basic Research Development Plan of China (no. 2011CB013403). The authors thank Dr. X. H. Liang for the help in validating the analytical mesh stiffness results. Comments and suggestions from anonymous referees and editors are quit valuable in the improvement of the quality of this paper.

## References

- [1] Z. P. Feng, M. Liang, and F. L. Chu, "Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 165–205, 2013.
- [2] X. F. 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] Z. L. Liu, J. Qu, M. J. Zuo, and H.-B. Xu, "Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel Fisher discriminant analysis," *International Journal of Advanced Manufacturing Technology*, vol. 67, no. 5, pp. 1217–1230, 2013.
- [4] Z. P. Feng and M. J. Zuo, "Vibration signal models for fault diagnosis of planetary gearboxes," *Journal of Sound and Vibration*, vol. 331, pp. 4919–4939, 2012.
- [5] Z. P. Feng, M. J. Zuo, J. Qu, T. Tian, and Z. L. Liu, "Joint amplitude and frequency demodulation analysis based on local mean decomposition for fault diagnosis of planetary gearboxes," *Mechanical Systems and Signal Processing*, vol. 40, no. 1, pp. 56–75, 2013.
- [6] M. Inalpolat and A. Kahraman, "A theoretical and experimental investigation of modulation sidebands of planetary gear sets," *Journal of Sound and Vibration*, vol. 323, no. 3–5, pp. 677–696, 2009.
- [7] M. Inalpolat and A. Kahraman, "A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors," *Journal of Sound and Vibration*, vol. 329, no. 4, pp. 371–393, 2010.
- [8] J. Lin and R. G. Parker, "Analytical characterization of the unique properties of planetary gear free vibration," *Journal of Vibration and Acoustics*, vol. 121, no. 3, pp. 316–321, 1999.
- [9] J. Lin and R. G. Parker, "Planetary gear parametric instability caused by mesh stiffness variation," *Journal of Sound and Vibration*, vol. 249, no. 1, pp. 129–145, 2002.
- [10] D. C. H. Yang and J. Y. Lin, "Hertzian damping, tooth friction and bending elasticity in gear impact dynamics," *Journal of Mechanisms, Transmissions, and Automation in Design*, vol. 109, no. 2, pp. 189–196, 1987.
- [11] X. H. Tian, *Dynamic Simulation for System Response of Gearbox including Localized Gear Faults*, University of Alberta, Edmonton, Canada, 2004.
- [12] X. Zhou, Y. M. Shao, Y. G. Lei, and M. J. Zuo, "Time-varying meshing stiffness calculation and vibration analysis for a 16DOF dynamic model with linear crack growth in a pinion," *Journal of Vibration and Acoustics*, vol. 134, no. 1, Article ID 011011, 2012.
- [13] Z. G. Chen and Y. M. Shao, "Dynamic simulation of spur gear with tooth root crack propagating along tooth width and crack depth," *Engineering Failure Analysis*, vol. 18, no. 8, pp. 2149–2164, 2011.
- [14] S. Y. Wu, *Dynamic Simulation and Estimation of Fault Growth*, University of Alberta Edmonton, Alberta, Canada, 2007.
- [15] S. Y. Wu, M. J. Zuo, and A. Parey, "Simulation of spur gear dynamics and estimation of fault growth," *Journal of Sound and Vibration*, vol. 317, no. 3–5, pp. 608–624, 2008.
- [16] Z. G. Tian, M. J. Zuo, and S. Y. Wu, "Crack propagation assessment for spur gears using model-based analysis and simulation," *Journal of Intelligent Manufacturing*, vol. 23, no. 2, pp. 239–253, 2012.
- [17] X. H. Liang, M. J. Zuo, and T. H. Patel, "Evaluating the time-varying mesh stiffness of a planetary gear set using the potential energy method," *Proceedings of the Institution of Mechanical Engineers C: Journal of Mechanical Engineering Science*, vol. 228, no. 3, pp. 535–547, 2014.
- [18] X. H. Liang, M. J. Zuo, and M. Pandey, "Analytically evaluating the influence of crack on the mesh stiffness of a planetary gear set," *Mechanism and Machine Theory*, vol. 76, pp. 20–38, 2014.
- [19] X. H. Liang, M. J. Zuo, and M. R. Hoseini, "Vibration signal modeling of a planetary gear set for tooth crack detection," *Engineering Failure Analysis*, vol. 48, pp. 185–200, 2015.
- [20] Z. G. Wan, H. R. Cao, Y. Y. Zi, W. P. He, and Z. J. He, "An improved time-varying mesh stiffness algorithm and dynamic modeling of gear-rotor system with tooth root crack," *Engineering Failure Analysis*, vol. 42, pp. 157–177, 2014.
- [21] Y. M. Shao and Z. G. Chen, "Dynamic features of planetary gear set with tooth plastic inclination deformation due to tooth root crack," *Nonlinear Dynamics*, vol. 74, no. 4, pp. 1253–1266, 2013.

- [22] R. G. Parker and J. Lin, "Mesh phasing relationships in planetary and epicyclic gears," *Journal of Mechanical Design—Transactions of the ASME*, vol. 126, no. 2, pp. 365–370, 2004.
- [23] T. Barszcz and R. B. Randall, "Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352–1365, 2009.

## Research Article

# Application of Artificial Neural Network for Damage Detection in Planetary Gearbox of Wind Turbine

**Marcin Strączkiewicz and Tomasz Barszcz**

*Department of Robotics and Mechatronics, Faculty of Mechanical Engineering and Robotics, AGH University of Science and Technology, 30 Mickiewicza Avenue, 30-059 Krakow, Poland*

Correspondence should be addressed to Marcin Strączkiewicz; [mstr@agh.edu.pl](mailto:mstr@agh.edu.pl)

Received 3 July 2015; Revised 28 August 2015; Accepted 1 September 2015

Academic Editor: Dong Wang

Copyright © 2016 M. Strączkiewicz and T. Barszcz. 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 the monitoring process of wind turbines the utmost attention should be given to gearboxes. This conclusion is derived from numerous summary papers. They reveal that, on the one hand, gearboxes are one of the most fault susceptible elements in the drive-train and, on the other, the most expensive to replace. Although state-of-the-art CMS can usually provide advanced signal processing tools for extraction of diagnostic information, there are still many installations, where the diagnosis is based simply on the averaged wideband features like root-mean-square (RMS) or peak-peak (PP). Furthermore, for machinery working in highly changing operational conditions, like wind turbines, those estimators are strongly fluctuating, and this fluctuation is not linearly correlated to operation parameters. Thus, the sudden increase of a particular feature does not necessarily have to indicate the development of fault. To overcome this obstacle, it is proposed to detect a fault development with Artificial Neural Network (ANN) and further observation of linear regression parameters calculated on the estimation error between healthy and unknown condition. The proposed reasoning is presented on the real life example of ring gear fault in wind turbine's planetary gearbox.

## 1. Introduction

In recent years, large power planetary gearboxes became very popular elements of wind turbine's drive-trains due to their ability to transmit a relatively high load from blades simultaneously providing a high gear ratio, which is required for efficient production of electrical energy by the generator.

On the other hand, as it was pointed out in a number of statistical summaries [1–4], gearboxes (with no special distinction for the planetary ones and others) are located on top of the list of most damage susceptible drive-train components in this group of rotating machinery. This is one of the reasons why wind turbine gearboxes, and planetary ones in particular, received the utmost attention of researchers and maintenance engineers. In consequence, this initiated the pursuit for providing the methodology that could detect the damage at the early stage and therefore limit the downtime of the entire machine. It should be mentioned that, due to the sharing of load between several meshes during operation of a planetary gearbox, usually a single malfunction results in total gear

damage requiring full gearbox replacement. According to [2], a 2.5 MW wind turbine planetary gearbox replacement costs over £400,000, which significantly influences a total wind turbine revenue. Figure 1 presents such high power gearbox with two parallel stages and single planetary stage.

One of the possibilities for damage assessment of planetary gearboxes is vibration analysis. Unlike other typical monitored industrial objects, vibration analysis of such planetary gearboxes is generally more challenging. This is due to few reasons. Firstly, wind turbines inherently operate under nonstationary conditions, including variable speed and load. Therefore, analysis of its data requires advanced signal selection algorithms [5, 6] as well as sophisticated signal resampling methods [7]. Secondly, in order to maximize even teeth wear, planetary gearboxes are designed with maximum length of a full cycle (i.e., exact repeatable position of planets and all teeth). Thus, faults related to planets may manifest themselves differently upon measurements not covering entire cycles. Tertiary, due to relatively low speed of wind turbine planetary gearbox elements consequent in relatively



FIGURE 1: Planetary gearbox.

low characteristic frequencies, accurate fault identification requires extremely long records, which are difficult to store and analyze. Moreover, usually low speed results in low fault induced energy, which in turn requires high sensitivity and high quality data acquisition. Finally, commercial wind turbine manufacturers do not allow any mechanical interference into planetary gearbox casing, limiting signal sources to vibration recorded from outside a gearbox.

The first attempts to the vibration analysis of gears were performed in the early 80s [8]. Over the years, a number of methods were developed; however many of them referred to analysis of spur gears [9–13] covering wide range of approaches like dynamic modelling [14, 15], wavelet transform [16], Wigner-Ville distribution [17], and Principal Component Analysis [18], just to name a few. The research interest in planetary gearboxes increased along with the computational and storage capabilities of modern CMS. This allowed providing methodology for detection of damage of a different nature, that is, spalling [19, 20] and pitting [21, 22], cracked tooth [23] and carrier [24], faulty planet gears [25, 26] and ring gears [27], and so forth. Zimroz and Bartkowiak [28] analyzed application of Principal Component Analysis and Canonical Discriminant Analysis with multidimensional matrix of variables for classification of planetary gearbox condition. This work was preceded by [29], where authors developed a method that takes advantage of highly varying operational parameters by relating gearbox's condition to susceptibility of diagnostic features to load. It has to be mentioned that the exhaustive list of references related to this topic can be found in the review papers like [30, 31].

Unfortunately, although many diagnostic methods were proposed, a lot of them suffer from shortcomings that prevent their industrial applications. In [32], the authors noticed a requirement for visual inspection of diagnostic results for comparative assessment between healthy and damaged condition. Jablonski and Barszcz [33] point that many of the approaches assume precise location of gears in time, which is impractical for the majority of commercial installation. It was also observed that for particular techniques the data analysis is performed without the inclusion of information about the unsteady conditions of work [28]. Finally, based on the experience of the authors of this paper, in many contemporary CMS, especially in SCADA-based systems, raw vibration signals are not available for the end user. This is due

to the limited data transfer, like for remote WTs, where data transmission is utilized via GPRS connection and prevents transmission of data packages containing highly sampled vibration signals of long duration. In such applications, the condition monitoring process is performed entirely on the basis of simple diagnostic features, like root-mean-square or the amplitude of vibration.

In such case, these indicators are observed over time (in the so-called *trends*) for the potential increase that might provide information of the occurrence of a malfunction (e.g., [34]). The additional obstacle is a nonlinear correlation between the level of diagnostic indicators and the corresponding operating conditions. Therefore the assessment on the behavior of the indicators is difficult [35, 36]. This demand for stationary or quasistationary conditions is rarely satisfied for quite large group of machines like wind turbines [37–39] or mining excavators [40].

The above issue may be overcome by introducing operational states of machine [41, 42]. In this approach, each trend is tracked in a precalculated range of operational conditions. Moreover, operational states are used for data classification during the acquisition procedure and consequently in data analysis process. Such modes are selected in order to most accurately describe a particular machine. Furthermore, it is desired to define a state for which calculated trends show low (or the lowest there is) level of variation [42]. Such stability is required for reliable justification of alarm threshold value. Otherwise, that is, for high variability, sudden increase of estimator would cause false positive alarms, misleading in the process of condition monitoring. Unfortunately, as mentioned in [42], the methodology for this approach is in the early stage of development.

In this paper, the authors follow the idea of the comparative study of simple vibration features in highly nonstationary operating conditions. The proposed reasoning is based on backpropagation Artificial Neural Network (ANN) and linear regression that fit the estimation error in the function of selected operation parameter. According to the literature review, so far such solution has not been proposed to the failure detection process of planetary gearboxes. Good accuracy of ANN for condition monitoring was shown for rolling element bearings [43–45], spur gears [46], centrifugal pumps [47], and power transformers [48]. In [49] the author employed ANN for estimation of remaining useful life of bearings. Bahmanyar and Karami [50] applied ANN to voltage signals while Crivelli et al. [51] verified its efficiency for acoustic emission. An interesting approach was proposed by Zhang and Wang [52] where ANN is used to model the behavior of generator's bearing, yet this method was applied to slowly varying parameter, that is, temperature.

The technique proposed in this paper takes into consideration fluctuation of vibration signal features under highly varying operating parameters of speed and load. Firstly, the ANN is trained at the normal (referential) behavior of the machine. Then, for each upcoming unclassified observation, the given value of a feature is compared against the value estimated using the taught net. In case of the regular operation, the error in assessment is relatively small, and it would

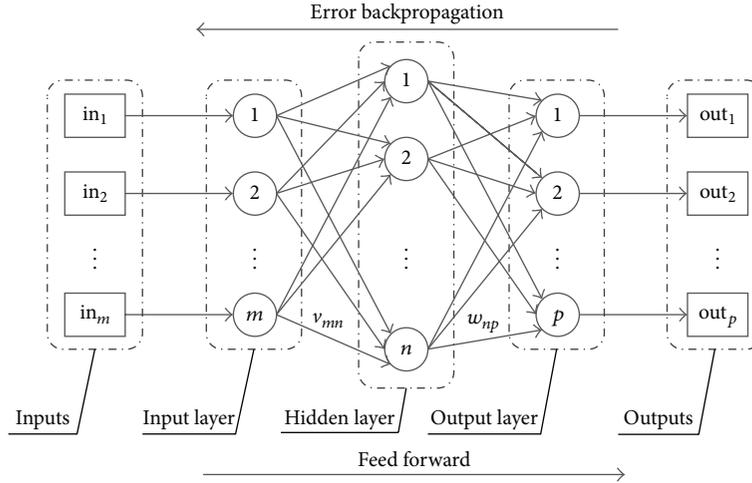


FIGURE 2: Scheme of Artificial Neural Network with backpropagation.

increase in case of arising malfunction. In order to limit the influence of changing conditions, the linear regression is employed and  $b$  parameter (also called the error term) is observed. The method is presented on the case study example of planetary gearbox stationary ring fault that was not detected using classical trend analysis.

The paper is organized as follows. Within the introductory part, the state of the art of planetary gearbox diagnostics is presented and discussed with its shortcomings related to implementation in SCADA-based systems. The authors propose to use Artificial Neural Network and discuss its hitherto applications. Next, the principles of Artificial Neural Networks are described. The proposed fault detection approach is presented in Section 3. In Section 4, the usage of the method is shown in data acquired from commercially used wind turbine. Finally, the results are discussed and paper is concluded.

## 2. Artificial Neural Networks

The concept of backpropagation ANN was firstly proposed in [53]. It is constructed as a multilayer structure (Figure 2), containing three essential layers: input, hidden, and output, whereas the hidden layer might consist of multiple layers. The inputs  $in_1, in_2, \dots, in_m$  information is passed to the unseen hidden layers with initial weights, generally described as  $v_{mn}$  which corresponds to weight given to information between input neuron  $m$  and hidden neuron  $n$ . In the process of a forward phase, the outputs  $out_1, out_2, \dots, out_m$  are calculated using different set of weights  $w_{np}$  between the hidden layer and the output layer. Then, the trained outputs are compared with the target ones and the error is established. In the following step, the error backpropagation is computed, which means that the weights are calibrated in order to minimize the estimation error for the entire set of training data. The last operation can be repeated in successive iterations called epochs, until the acceptable level of estimation error is established.

## 3. The Proposed Algorithm

As previously stated, the authors' intention is to present the method that takes into account process parameters, like oil temperature, wind, and rotational speed or load, and detect the increase of one or more of diagnostic features in the presence of fault development. In order to do that, the dependency on fluctuation of process parameters has to be reduced. It is possible due to the learning process of Artificial Neural Network that estimates a relationship between the input process parameters and the output diagnostic ones. The contribution of the paper is to use ANN as the model of the normal behavior of the vibration signal features, depending on the most important process variables, and later to use the trained model as the reference to the actual values of these features.

After establishing the network for the training set of inputs and outputs, the testing procedure can be performed. The net estimates values of diagnostic features based on given inputs. Next, this prediction is subtracted from the actual indication carried from CMS, resulting in estimation error. In fact, there are three reasons for the erroneous estimation.

The first one is the improper training of the neural network resulting in invalid output estimations of trained ANN for the unseen input. In this case, the estimation error changes with operational parameters, yet it stays at the similar level in time. In order to minimize this shortcoming, it is required to train the neural network using long time measurements that cover the entire range of operation parameters.

The second reason is related to the dynamics of the observed machine, that is, with unequivocal relation between operational parameters and diagnostic features. This happens when for a given combination of input parameters the output takes various values. In the training process of ANN such ambiguity is not possible to model since the taught net assigns a single value for each combination of inputs. It should be pointed out that, for the unchanged condition of machine, the estimation error related to imprecise training of ANN remains constant in time.

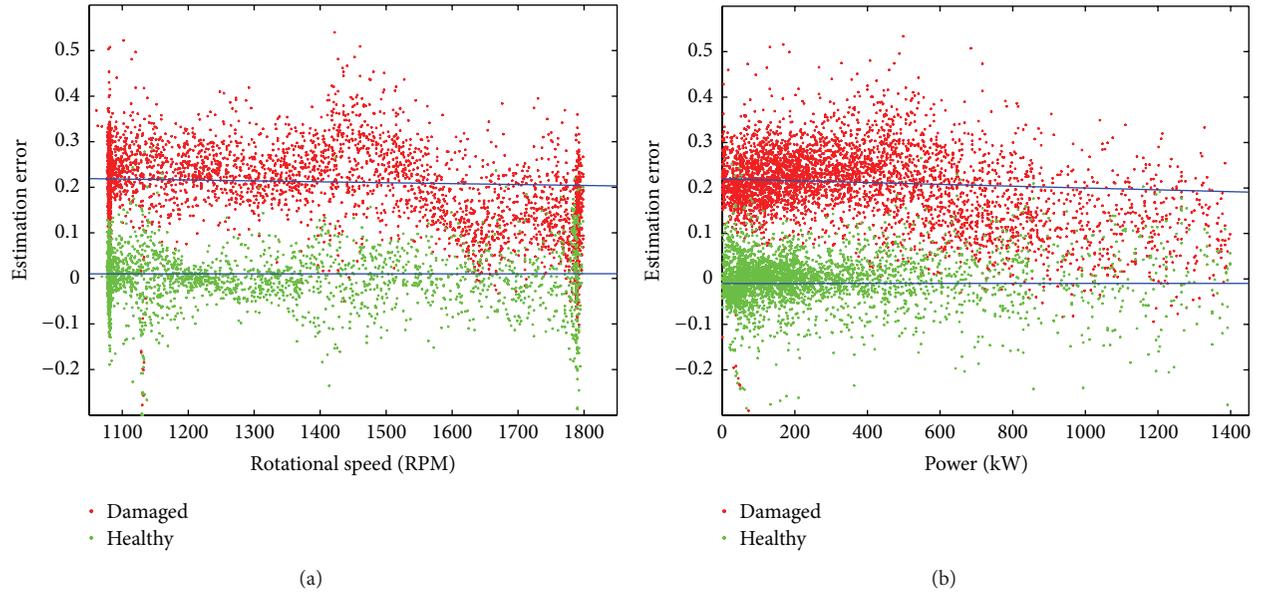


FIGURE 3: Exemplary regression analysis (blue) performed in two conditions: healthy (green) and damaged (red) plotted against operational parameters.

The third one appears when measured values do not correspond to those calculated in the referential period. This might indicate either sensor failure or, even more importantly, a damage development.

All of the aforementioned reasons might exist at the same time; therefore it is suggested to minimize its random influence and emphasize any deterministic phenomena. This can be done using the idea similar to the one presented in [29, 37]. Zimroz et al. proposed to observe diagnostic features against chosen operational parameter in several segments obtained from trend plot. For each segment, the linear regression function is fit to the cloud of observations. The obtained parameters  $a$  and  $b$  are used for decision making process.

In the approach proposed in this paper, the linear regression is used to fit the estimation error in the function of selected operation parameter and not to the raw diagnostic features. It is based on the conclusion that in bad condition of machine the values of diagnostic features, and simultaneously the estimation error, would increase regardless of the change of operation conditions. Thanks to use of ANN, the dependency of error on operational parameters is limited, so the diagnostic reasoning can be limited simply to parameter  $b$ , also called the error term. For good condition, this parameter remains at low level (close to zero), and for bad condition it is significantly increased (Figure 3). In this case, threshold value may be established using statistical methods as presented, for example, in [54, 55], with no concern on highly changing operating conditions. Additionally, in order to improve the time domain resolution and hence accelerate the diagnostic decision, the segmentation may be performed with overlap [37]. In the proposed reasoning parameter  $a$  is not taken into consideration since in general it is more susceptible to ambiguous association between inputs and

output of ANN, and it is related to character of machine's operation. As presented in Figure 3, the distribution of error might be uneven in the domain of operational parameter, so it influences fitted regression line (marked in blue). Therefore it is concluded that parameter  $a$  does not carry diagnostic information.

## 4. Case Study

*4.1. Object Description.* The proposed method is presented on the data acquired using vibration-based CMS installed on commercially utilized wind turbine (WT) located in western Germany. This 1.5 (MW) WT has a standard design of a drive-train; that is, gearbox and generator are in line in the nacelle, as presented in Figure 3. The main rotor is driven by three blades and supported by the main bearing (A1). The planetary gearbox (A2) has three planets, which are impelled by the planet carrier. The planets roll over the ring and transmit the torque to the sun gear, simultaneously increasing the rotational speed. The sun shaft passes the torque from the planetary gear to the two-stage parallel gear (A3 and A4). The parallel gears have three shafts: the slow one that clutches to the sun shaft, the intermediate one, and the fast shaft, which drives the generator. The generator produces AC current of a varying frequency. It is firstly converted into DC and then into AC current of a frequency equal to the grid one. Electric transformations are performed by the controller located at the base of the WT's tower. In general, such drive-train multiplies the rotational speed from about 25 (RPM) on the main rotor to about 1800 (RPM) at the generator.

In the typical vibration-based CMS, it takes six sensors to collect data adequate for the proper diagnostics of the entire drive-train. They are located as presented in Figure 4: on main bearing (A1), on planetary gearbox (A2), on two parallel

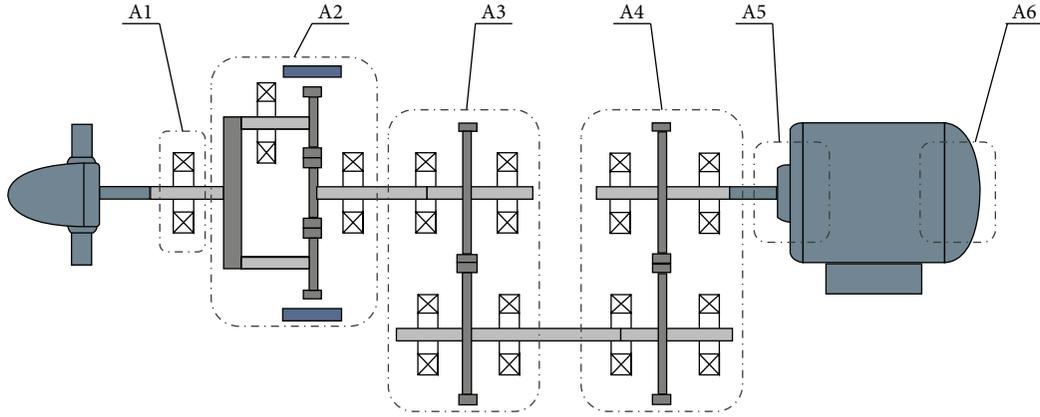


FIGURE 4: Kinematic scheme of the studied wind turbine.

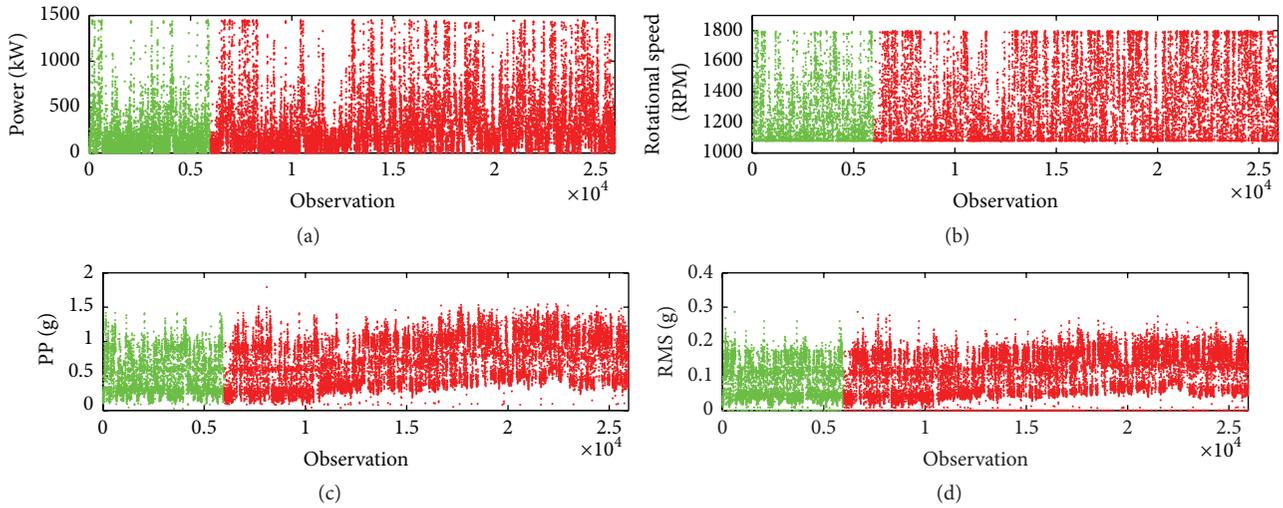


FIGURE 5: Observed parameters: power output (a), rotational speed (b), PP (c), and RMS of vibration signal from planetary gearbox (d). In green, data used for training; in red, data used for testing.

gearboxes (A3 and A4), and on the generator bearings (driven end (A5) and non-driven end (A6)).

As listed in [56], for machine with similar construction, nearly 150 indicators are extracted for the purpose of condition monitoring. The data is collected together with the corresponding operating parameters, namely, load and rotational speed.

In the described CMS, each parameter was acquired every 15 minutes from 10 (s) measurements. The vibration signals were collected with 25 kHz of sampling frequency. Process parameters were recorded simultaneously to vibration signal at rate of once per second, so that 10 measurements were collected for each vibration signal. In the next step, these values were averaged, so that, as the output, the end user of the system received one value per observation for each of the parameters. It should be mentioned that, before the procedure of feature extraction, the validation of signals was performed according to the guideline proposed in [5]. It included partial filtration of data for low rotational speed

(below 1050 (RPM)) and periods where load was below 0 (kW), that is, when wind turbine did not generate power.

In the presented case study, the wind turbine was observed for about 13 months—from the time of CMS installation until its stoppage for gearbox replacement. In total, about 38500 observations were recorded out of which about 26000 satisfied the validation requirement mentioned above. In the meantime, the extended wear of ring gear was developed. This phenomenon was not detected by CMS, yet it was found during periodical service measurements around observation 22500. In Figure 5, one can observe frequent changes of the operational state, although slight increase of wideband features was visible. The increase is only a small percent, and so typical approach for setting a threshold will not work, since it is based on the empirical distribution function [54].

The authors would like to discuss the approach based on ANN and linear regression that could reveal fault development in this gearbox. The data used for the proposed

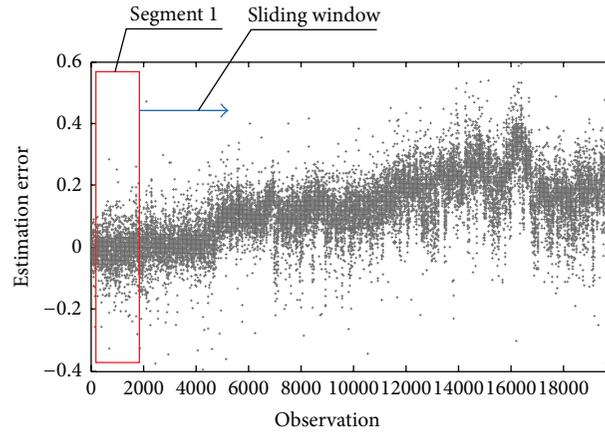


FIGURE 6: Estimation error with marked segment 1 (red) and sliding window (blue).

algorithm include two process parameters, namely, rotational speed of the generator's shaft and generator's output power, understood as load. There is a large number of features, which are calculated from raw vibration signals. Most popular ones are narrowband energies representing frequencies equal to gear mesh, shaft harmonics, and so forth. There are also broadband features, which are less susceptible to noise and windowing effects and are considered more robust ones. This is the main reason why it has been normalized and is used as an industry standard [57]. Two diagnostic features were selected, namely, root-mean-square (RMS) and peak-peak (PP).

**4.2. Construction of ANN.** One of the most important problems in condition monitoring of machinery in highly changing operational conditions is a nonlinear influence of such conditions, in particular rotational speed and load, on the level of vibration and thus on the fault indicators. It is often observed that the increase in amplitude related to changing conditions is much higher than the one created by a machinery fault, especially in its early stage. In the authors' opinion, one of the possible methods to reduce this problem is the use of ANN. In this approach, the network is trained using data collected in the referential (healthy) period of machine's operation and then used for estimation of values in the testing period, when the condition of machine is unknown. In the beginning of the investigation, a number of combinations for the ANN structure were tested. This resulted in conclusion that the least erroneous estimation is obtained using the 2-10-20-2 combination of nodes with mean square error equal to 0.0144 for PP and 0.0025 for RMS. The structure described as 2-10-20-2 means that two nodes were used as inputs (i.e., rotational speed and load); there were two hidden layers consisting of 10 and 20 nodes, respectively. Finally, the RMS and PP indicators were employed as the output (target) values.

**4.2.1. Training.** The proposed network was trained using about 3 months of observation (Figure 5), which amounted

to about 6000 samples for each parameter. As mentioned previously, the rotational speed and load (power output) were treated as input parameters for training of the network, since it can be stated that it mostly influences the diagnostic parameters [29]. The RMS and PP were used as outputs.

The training on the given structure took place in 500 repetitions (epochs), during which the initial weights were corrected for minimizing the output error of prediction in the process of backpropagation (Figure 2). Finally the established net was used for the remaining data for possible damage detection.

**4.2.2. Testing.** The testing set covered about 20000 observations for each of the selected features, as presented in Figure 5. The input data were used by the established net of RMS and PP as outputs in order to achieve estimations of these parameters in the substantial time of the unknown operation. In the next step, the approximations were compared with the values calculated from the actual vibration signals recorded by CMS. Thanks to this procedure, the error of the estimation was obtained.

**4.2.3. Linear Regression Fitting.** In the next step, the data was divided into overlapping segments (overlap was equal to 50%), each containing 2000 observations (Figure 6).

In Figures 7 and 8 one may observe distribution of estimation error against two operating parameters for PP and RMS, respectively. It can be seen that errors increase regardless of speed and load, yet they do not separate entirely from the initial segment. This is due to the small increase of diagnostic features in the initially observed trends and it prevents using classical methodology based on thresholds.

**4.3. Fault Recognition.** In the next step, for each segment linear regression was fit using least squares approach and parameters  $a$  and  $b$  were obtained and presented in Tables 1 and 2. Additionally, it is plotted in Figures 9 and 10. It can be seen that parameter  $a$  does not change significantly regardless

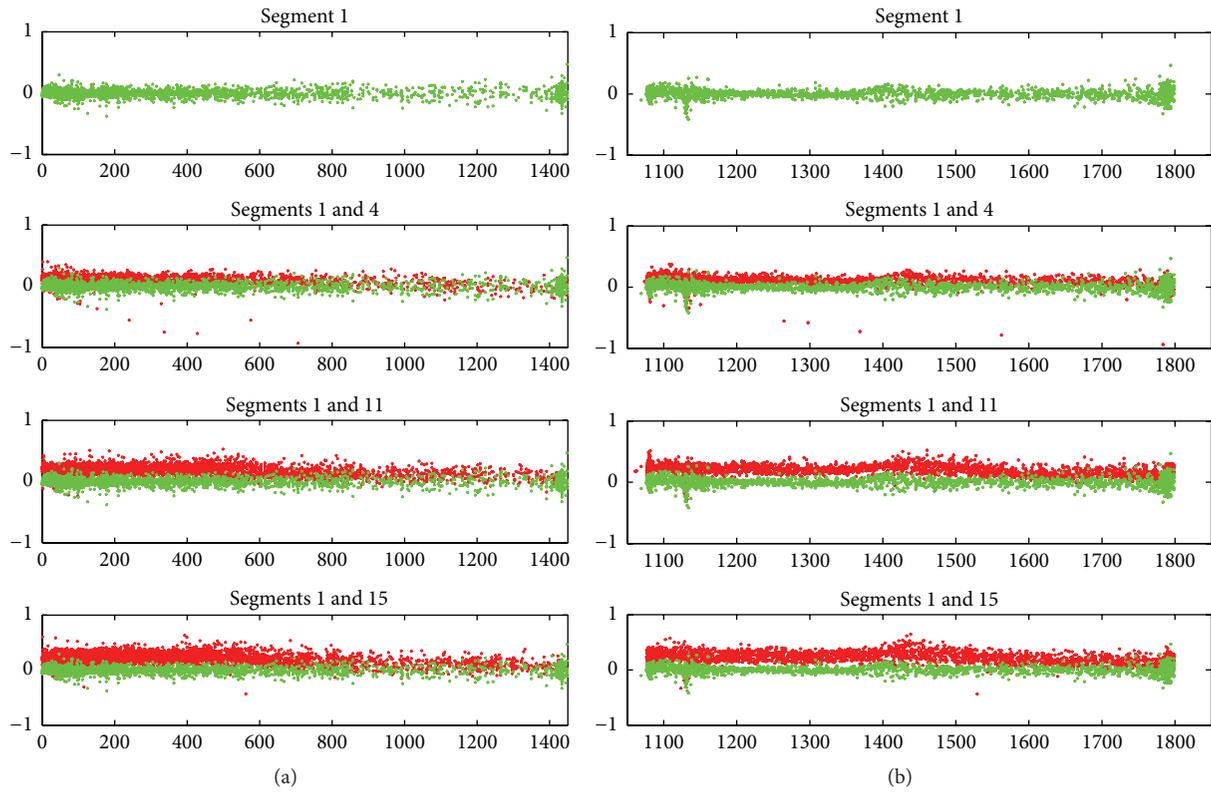


FIGURE 7: Regression of PP estimation error on segment 1 (green) and selected following segments (red) in function of power (a) and rotational speed (b).

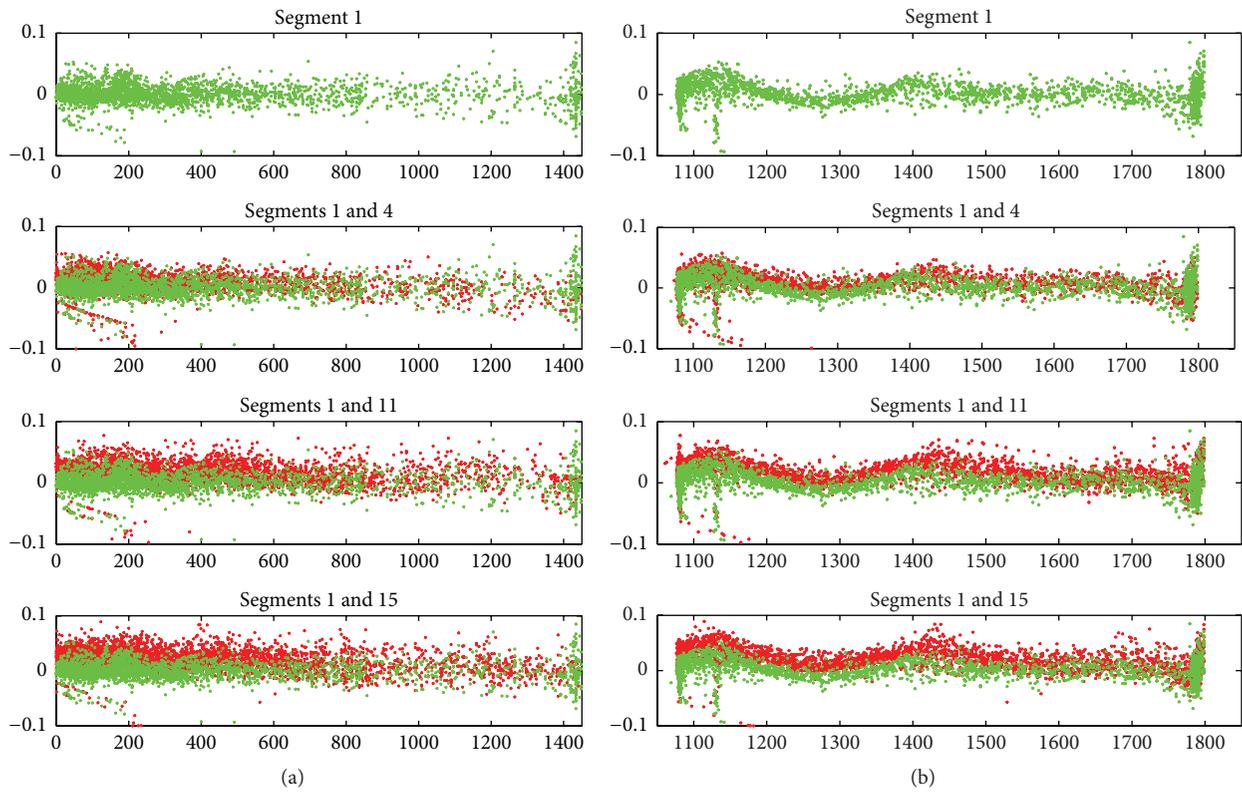


FIGURE 8: Regression of RMS estimation error on segment 1 (green) and selected following segments (red) in function of power (a) and rotational speed (b).

TABLE 1: Regression parameters  $a$  and  $b$  for PP feature.

Number of segment	Regression as a function of load		Regression as a function of speed	
	$a$	$b$	$a$	$b$
1	$-1.5766e-05$	$-0.0003$	$-1.6381e-05$	$0.0029$
2	$-1.5483e-05$	$0.0430$	$-1.7524e-05$	$0.0239$
3	$4.2233e-06$	$0.0697$	$5.8440e-06$	$0.0898$
4	$1.2853e-05$	$0.1115$	$1.8195e-05$	$0.1817$
5	$1.1192e-05$	$0.1173$	$1.7271e-05$	$0.1861$
6	$1.0557e-05$	$0.1326$	$1.5953e-05$	$0.2387$
7	$-1.2469e-07$	$0.1394$	$2.9331e-06$	$0.2648$
8	$-3.1138e-05$	$0.1403$	$-3.7741e-05$	$0.2510$
9	$-6.1005e-05$	$0.1554$	$-6.4024e-05$	$0.2474$
10	$-7.9284e-05$	$0.1652$	$-9.2298e-05$	$0.2422$
11	$-7.8454e-05$	$0.1922$	$-9.3476e-05$	$0.2892$
12	$-6.9781e-05$	$0.2298$	$-8.2785e-05$	$0.3566$
13	$-7.3974e-05$	$0.2496$	$-9.4870e-05$	$0.3802$
14	$-8.8399e-05$	$0.2664$	$-1.1258e-04$	$0.4148$
15	$-9.5407e-05$	$0.2748$	$-1.2264e-04$	$0.4091$
16	$-1.1561e-04$	$0.2629$	$-1.5154e-04$	$0.4069$
17	$-1.1373e-04$	$0.2471$	$-1.3672e-04$	$0.4122$
18	$-1.3014e-04$	$0.2261$	$-1.2909e-04$	$0.3680$

TABLE 2: Regression parameters  $a$  and  $b$  for RMS feature.

Number of segment	Regression as a function of load		Regression as a function of speed	
	$a$	$b$	$a$	$b$
1	$4.3312e-07$	$0.0008$	$6.8457e-08$	$0.0044$
2	$2.4075e-06$	$0.0057$	$1.1940e-06$	$0.0109$
3	$9.2405e-06$	$0.0081$	$8.9710e-06$	$0.0172$
4	$1.0180e-05$	$0.0115$	$9.6781e-06$	$0.0311$
5	$7.9734e-06$	$0.0117$	$6.5334e-06$	$0.0348$
6	$9.1124e-06$	$0.0130$	$7.0332e-06$	$0.0427$
7	$2.1228e-06$	$0.0130$	$-1.8971e-07$	$0.0452$
8	$-5.3487e-06$	$0.0122$	$-8.3286e-06$	$0.0433$
9	$-9.1654e-06$	$0.0132$	$-1.4587e-05$	$0.0396$
10	$-1.0978e-05$	$0.0140$	$-1.8990e-05$	$0.0355$
11	$-1.0032e-05$	$0.0173$	$-1.8843e-05$	$0.0368$
12	$-6.7979e-06$	$0.0223$	$-1.5419e-05$	$0.0442$
13	$-5.6093e-06$	$0.0247$	$-1.2386e-05$	$0.0451$
14	$-8.2130e-06$	$0.0267$	$-1.4679e-05$	$0.0502$
15	$-5.8187e-06$	$0.0287$	$-1.2387e-05$	$0.0555$
16	$-9.0436e-06$	$0.0276$	$-1.7583e-05$	$0.0545$
17	$-8.4208e-06$	$0.0261$	$-1.6874e-05$	$0.0563$
18	$-6.8415e-06$	$0.0237$	$-1.5494e-05$	$0.0516$

of the diagnostic indicator and operation parameter used for regression analysis. In each case, the oscillation around zero indicates even level of estimation error.

On the other hand, for parameter  $b$  the difference between initial and deteriorating condition of ring gear is significant. It is clear that the increase is higher using this parameter when

compared to trend of diagnostic feature alone. Moreover it is now possible to establish the control limit.

It is important to mention that the observed increase is not related to selected operating parameter, yet it is higher when observed in speed domain. This might be due to the wind turbine's design, which aims at establishing rotational

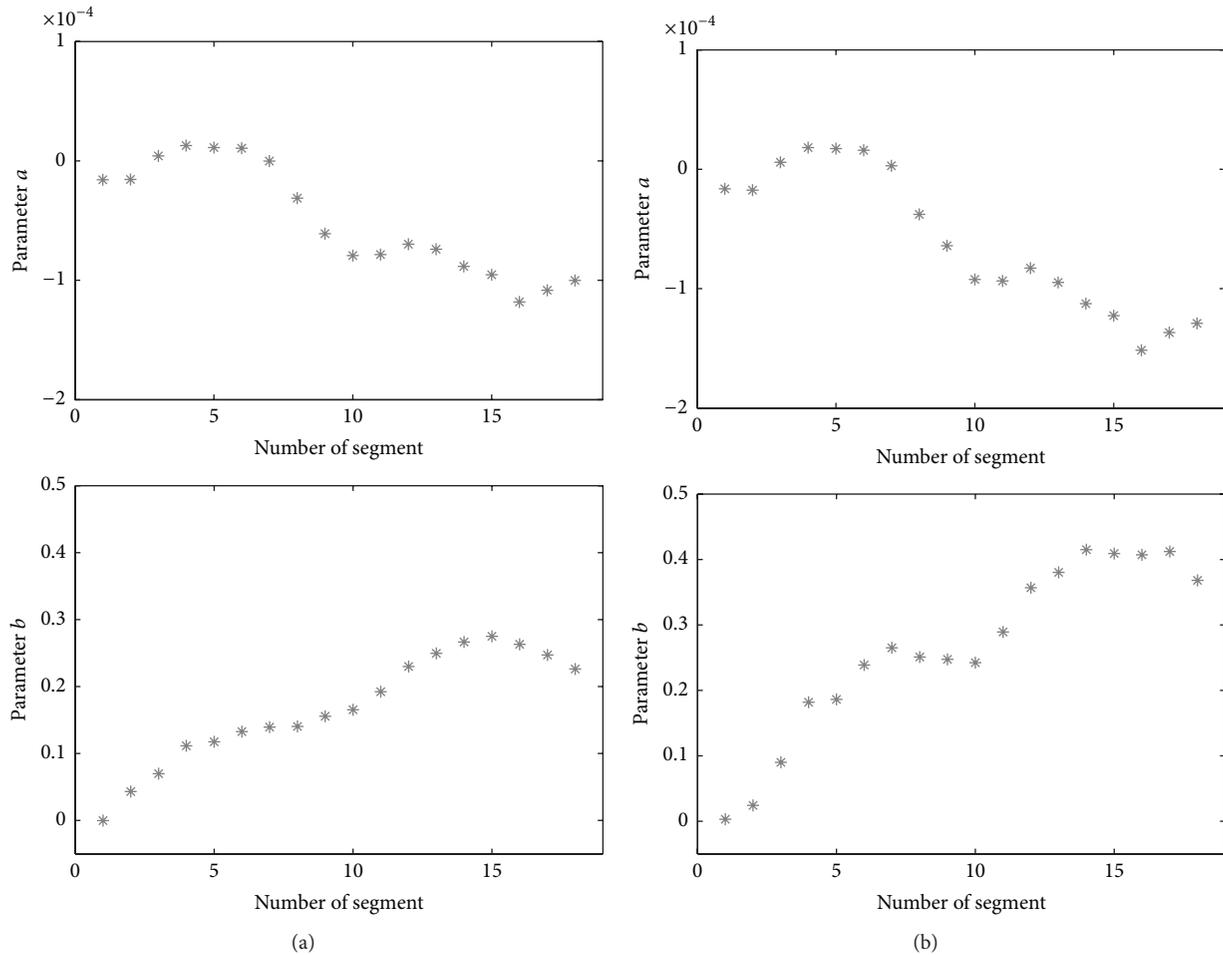


FIGURE 9: Fluctuation of parameters  $a$  and  $b$  obtained for consecutive segments of estimation error on PP over power (a) and rotational speed (b).

speed at certain levels (typically low and high). The arising error in estimation cumulates in few clouds around these speeds and influence the regression analysis. In the authors' opinion, this should not be considered as a drawback of a technique since it emphasizes arising malfunction.

## 5. Conclusions

Within the research paper, the application of backpropagation Artificial Neural Networks was investigated for detection of fault in wind turbine's planetary gearbox. ANN modelled the selected features depending on operational parameters and linear regression was later used to distinguish between the healthy and damaged state.

The analysis was focused on two commonly used diagnostic features, namely, root-mean-square (RMS) and peak-peak (PP). In the presented case study, it was shown that these indicators did not reveal sufficient symptoms of failure occurrence. To the contrary, the presented methodology exposed early stage of damage several months before gearbox replacement. This was possible by trending the erroneous estimation of diagnostic features and further fitting of linear regression in function of operation parameters. Thanks to

the ability of ANN to model nonlinear dependencies, it was possible to reduce difficulties related to constantly and highly varying operating parameters. Parameter  $b$  of linear regression provided information about general increase of error for the entire range of operating parameter and hence provided information of the development of damage in gearbox.

The additional benefit of the proposed approach is that it can be used with various input parameters and fault indicators. The method does not require preceding filtration of data, so there is no need for operational state setup. Clear distinction between initial and deteriorating condition makes it possible to establish a control limit and therefore automatize the diagnostic process. In the authors' opinion, it shows great potential for application in condition monitoring of various rotating machinery, such as wind turbines (as presented), mining excavators, or gas turbines.

## Conflict of Interests

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

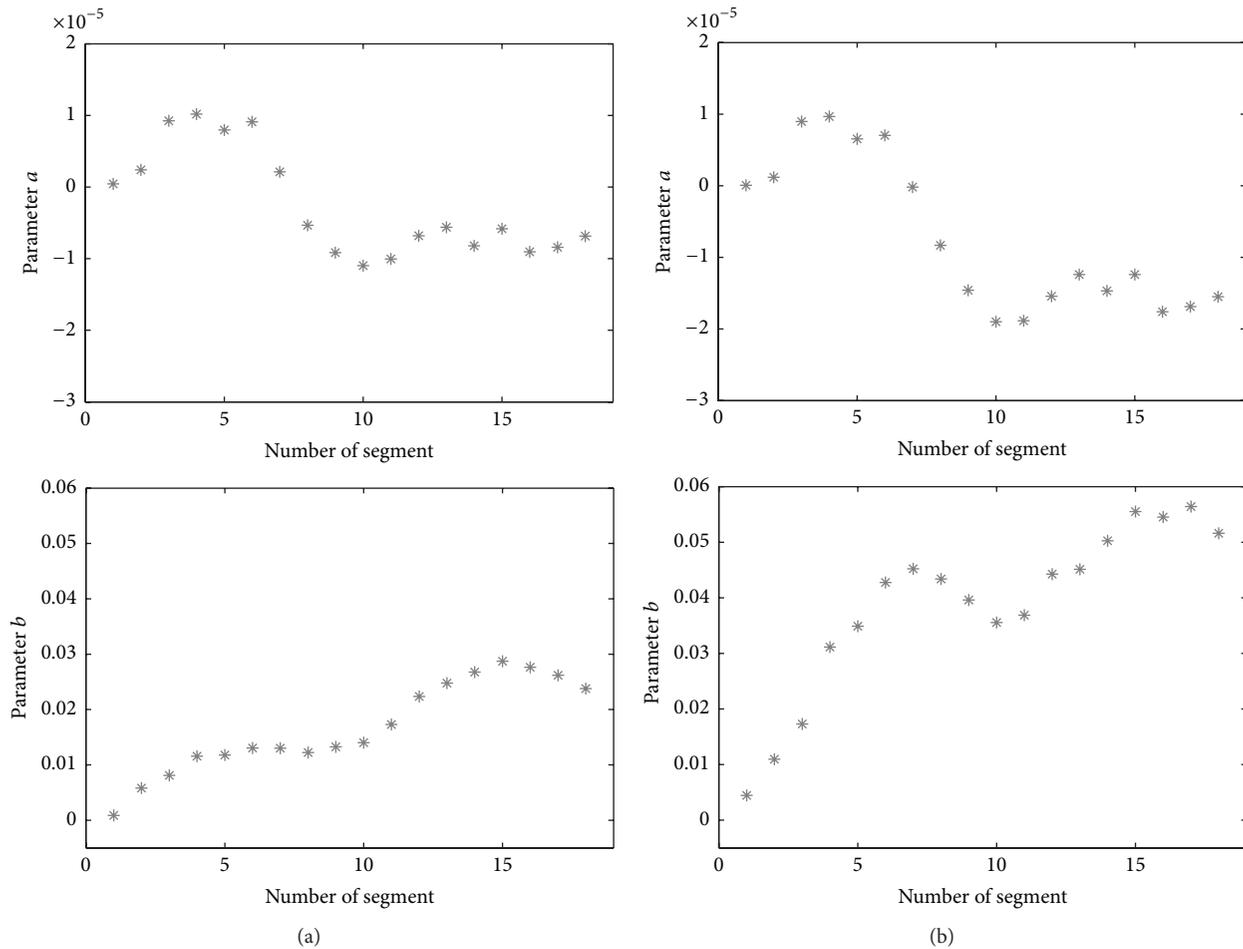


FIGURE 10: Fluctuation of parameters  $a$  and  $b$  obtained for consecutive segments of estimation error on RMS over power (a) and rotational speed (b).

## Acknowledgment

The authors would like to acknowledge the financial support of AGH University of Science and Technology within the program of Dean's Scholarships (no. 15.11.130 987).

## References

- [1] G. W. Van Bussel and M. B. Zaaijer, "Reliability, availability and maintenance aspects of large-scale offshore wind farms," in *Proceedings of the International Conference on Marine Renewable Energy*, pp. 119–126, Newcastle, UK, 2001.
- [2] D. McMillan and G. W. Ault, "Quantification of condition monitoring benefit for offshore wind turbines," *Wind Engineering*, vol. 31, no. 4, pp. 267–285, 2007.
- [3] J. Ribrant and L. M. Bertling, "Survey of failures in wind power systems with focus on Swedish wind power plants during 1997–2005," *IEEE Transactions on Energy Conversion*, vol. 22, no. 1, pp. 167–173, 2007.
- [4] S. Faulstich, B. Hahn, and P. J. Tavner, "Wind turbine downtime and its importance for offshore deployment," *Wind Energy*, vol. 14, no. 3, pp. 327–337, 2011.
- [5] A. Jablonski and T. Barszcz, "Validation of vibration measurements for heavy duty machinery diagnostics," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 248–263, 2013.
- [6] G. Song, Z.-Y. Li, P. Bellemain, N. Martin, and C. Mailhes, "Astrion data validation of non-stationary wind turbine signals," in *Proceedings of the 12th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, Oxford, UK, June 2015.
- [7] L. F. Villa, A. Reñones, J. R. Perán, and L. J. de Miguel, "Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation," *Mechanical Systems and Signal Processing*, vol. 25, no. 6, pp. 2157–2168, 2011.
- [8] R. B. Randall, "A new method of modeling gear faults," *ASME Journal of Mechanical Design*, vol. 104, no. 2, pp. 259–267, 1982.
- [9] P. D. McFadden, "Determining the location of a fatigue crack in a gear from the phase of the change in the meshing vibration," *Mechanical Systems and Signal Processing*, vol. 2, no. 4, pp. 403–409, 1988.
- [10] J. D. Smith, *Gear Noise and Vibration*, Marcel Dekker, New York, NY, USA, 2nd edition, 2003.
- [11] J. Maćzak, "Local meshing plane as a source of diagnostic information for monitoring the evolution of gear faults," in *Engineering Asset Lifecycle Management: Proceedings of the 4th World Congress on Engineering Asset Management (WCEAM)*

- 2009), 28–30 September 2009, pp. 661–670, Springer, London, UK, 2010.
- [12] A. Belsak and J. Flasker, “Method for detecting fatigue crack in gears,” *Theoretical and Applied Fracture Mechanics*, vol. 46, no. 2, pp. 105–113, 2006.
  - [13] W. J. Staszewski and G. R. Tomlinson, “Application of the wavelet transform to fault detection in a spur gear,” *Mechanical Systems and Signal Processing*, vol. 8, no. 3, pp. 289–307, 1994.
  - [14] I. Howard, S. Jia, and J. Wang, “The dynamic modelling of a spur gear in mesh including friction and a crack,” *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 831–853, 2001.
  - [15] W. Bartelmus, “Gearbox dynamic modelling,” *Journal of Theoretical and Applied Mechanics*, vol. 39, no. 4, pp. 989–999, 2001.
  - [16] W. J. Wang and P. D. McFadden, “Application of wavelets to gearbox vibration signals for fault detection,” *Journal of Sound and Vibration*, vol. 192, no. 5, pp. 927–939, 1996.
  - [17] W. J. Staszewski, K. Worden, and G. R. Tomlinson, “Time-frequency analysis in gearbox fault detection using the Wigner-Ville distribution and pattern recognition,” *Mechanical Systems and Signal Processing*, vol. 11, no. 5, pp. 673–692, 1997.
  - [18] R. Zimroz and A. Bartkowiak, “Investigation on spectral structure of gearbox vibration signals by principal component analysis for condition monitoring purposes,” *Journal of Physics: Conference Series*, vol. 305, Article ID 012075, pp. 1–11, 2011.
  - [19] P. D. Samuel and D. J. Pines, “A review of vibration-based techniques for helicopter transmission diagnostics,” *Journal of Sound and Vibration*, vol. 282, no. 1-2, pp. 475–508, 2005.
  - [20] P. D. Samuel and D. J. Pines, “Constrained adaptive lifting and the CAL4 metric for helicopter transmission diagnostics,” *Journal of Sound and Vibration*, vol. 319, no. 1-2, pp. 698–718, 2009.
  - [21] X. Zhao, M. J. Zuo, and Z. Liu, “Diagnosis of pitting damage levels of planet gears based on ordinal ranking,” in *Proceedings of the 2011 IEEE Conference on Prognostics and Health Management*, pp. 1–8, Denver, Colo, USA, June 2011.
  - [22] Z. Cheng, N. Hu, F. Gu, and G. Qin, “Pitting damage levels estimation for planetary gear sets based on model simulation and grey relational analysis,” *Transactions of the Canadian Society for Mechanical Engineering*, vol. 35, no. 3, pp. 403–417, 2011.
  - [23] Z. Chen, Z. Zhu, and Y. Shao, “Fault feature analysis of planetary gear system with tooth root crack and flexible ring gear rim,” *Engineering Failure Analysis*, vol. 49, pp. 92–103, 2015.
  - [24] D. M. Blunt and J. A. Keller, “Detection of a fatigue crack in a UH-60A planet gear carrier using vibration analysis,” *Mechanical Systems and Signal Processing*, vol. 20, no. 8, pp. 2095–2111, 2006.
  - [25] X. Zhao, M. J. Zuo, Z. Liu, and M. R. Hoseini, “Diagnosis of artificially created surface damage levels of planet gear teeth using ordinal ranking,” *Measurement*, vol. 46, no. 1, pp. 132–144, 2013.
  - [26] J. P. T. Koch and C. M. Vicuña, “Dynamic and phenomenological vibration models for failure prediction on planet gears of planetary gearboxes,” *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 36, no. 3, pp. 533–545, 2014.
  - [27] T. Barszcz and R. B. Randall, “Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine,” *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352–1365, 2009.
  - [28] R. Zimroz and A. Bartkowiak, “Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions,” *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 237–247, 2013.
  - [29] W. Bartelmus and R. Zimroz, “A new feature for monitoring the condition of gearboxes in non-stationary operating conditions,” *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1528–1534, 2009.
  - [30] Y. Lei, J. Lin, M. J. Zuo, and Z. He, “Condition monitoring and fault diagnosis of planetary gearboxes: a review,” *Measurement*, vol. 48, no. 1, pp. 292–305, 2014.
  - [31] 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.
  - [32] Y. Lei, D. Kong, J. Lin, and M. J. Zuo, “Fault detection of planetary gearboxes using new diagnostic parameters,” *Measurement Science and Technology*, vol. 23, no. 5, Article ID 055605, 2012.
  - [33] A. Jablonski and T. Barszcz, “Instantaneous circular pitch cyclic power (ICPCP)—a tool for diagnosis of planetary gearboxes,” *Key Engineering Materials*, vol. 518, pp. 168–173, 2012.
  - [34] D. Astolfi, F. Castellani, and L. Terzi, “Fault prevention and diagnosis through SCADA temperature data analysis of an onshore wind farm,” *Diagnostyka*, vol. 15, no. 2, pp. 71–78, 2014.
  - [35] J. R. Stack, T. G. Habetler, and R. G. Harley, “Effects of machine speed on the development and detection of rolling element bearing faults,” *IEEE Power Electronics Letters*, vol. 1, no. 1, pp. 19–21, 2003.
  - [36] W. Yang, R. Court, and J. Jiang, “Wind turbine condition monitoring by the approach of SCADA data analysis,” *Renewable Energy*, vol. 53, pp. 365–376, 2013.
  - [37] R. Zimroz, W. Bartelmus, T. Barszcz, and J. Urbanek, “Wind turbine main bearing diagnosis—a proposal of data processing and decision making procedure under non stationary load condition,” *Key Engineering Materials*, vol. 518, pp. 437–444, 2012.
  - [38] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. Song, “Condition monitoring and fault detection of wind turbines and related algorithms: a review,” *Renewable and Sustainable Energy Reviews*, vol. 13, no. 1, pp. 1–39, 2009.
  - [39] J. Urbanek, M. Strączkiewicz, and T. Barszcz, “Joint power-speed representation of vibration features: application to wind turbine planetary gearbox,” in *Proceedings of the 3rd International conference on Condition Monitoring of Machinery in Non-Stationary Operations*, Ferrara, Italy, May 2013.
  - [40] G. Żak, J. Obuchowski, A. Wylomańska, and R. Zimroz, “Application of ARMA modelling and alpha-stable distribution for local damage detection in bearings,” *Diagnostyka*, vol. 15, no. 3, pp. 3–10, 2014.
  - [41] M. A. Timusk, M. G. Lipsett, J. McBain, and C. K. Mechefske, “Automated operating mode classification for online monitoring systems,” *Journal of Vibration and Acoustics*, vol. 131, pp. 131–141, 2009.
  - [42] M. Strączkiewicz, P. Wiciak, A. Jabłoński, and T. Barszcz, “Machinery in highly changing operations: on designation of operational states,” in *Proceedings of the 12th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, pp. 1–12, Oxford, UK, June 2015.
  - [43] A. Hajnayeb, S. E. Khadem, and M. H. Moradi, “Design and implementation of an automatic condition-monitoring expert system for ball-bearing fault detection,” *Industrial Lubrication and Tribology*, vol. 60, no. 2, pp. 93–100, 2008.

- [44] B. Samanta and K. R. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features," *Mechanical Systems and Signal Processing*, vol. 17, no. 2, pp. 317–328, 2003.
- [45] M. Cocconcelli, R. Rubini, R. Zimroz, and W. Bartlemus, "Diagnostics of ball bearings in varying-speed motors by means of artificial neural network," in *Proceedings of the 8th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, pp. 760–771, Cardiff, UK, June 2011.
- [46] J. Rafiee, F. Arvani, A. Harifi, and M. H. Sadeghi, "Intelligent condition monitoring of a gearbox using artificial neural network," *Mechanical Systems and Signal Processing*, vol. 21, no. 4, pp. 1746–1754, 2007.
- [47] A. Azadeh, M. Saberi, A. Kazem, V. Ebrahimipour, A. Nourmohammadzadeh, and Z. Saberi, "A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyperparameters optimization," *Applied Soft Computing Journal*, vol. 13, no. 3, pp. 1478–1485, 2013.
- [48] D. V. S. S. S. Sarma and G. N. S. Kalyani, "ANN approach for condition monitoring of power transformers using DGA," in *Proceedings of the IEEE Region 10 Conference (TENCON '04)*, vol. 3, pp. 444–447, IEEE, November 2004.
- [49] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," *Journal of Intelligent Manufacturing*, vol. 23, no. 2, pp. 227–237, 2012.
- [50] A. R. Bahmanyar and A. Karami, "Power system voltage stability monitoring using artificial neural networks with a reduced set of inputs," *International Journal of Electrical Power and Energy Systems*, vol. 58, pp. 246–256, 2014.
- [51] D. Crivelli, M. Guagliano, and A. Monici, "Development of an artificial neural network processing technique for the analysis of damage evolution in pultruded composites with acoustic emission," *Composites Part B: Engineering*, vol. 56, pp. 948–959, 2014.
- [52] Z. Zhang and K. Wang, "Wind turbine fault detection based on SCADA data analysis using ANN," *Advances in Manufacturing*, vol. 2, no. 1, pp. 70–78, 2014.
- [53] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, pp. 318–362, MIT Press, Cambridge, Mass, USA, 1986.
- [54] A. Jablonski, T. Barszcz, M. Bielecka, and P. Breuhaus, "Modeling of probability distribution functions for automatic threshold calculation in condition monitoring systems," *Measurement*, vol. 46, no. 1, pp. 727–738, 2013.
- [55] C. Cempel, "Limit value in the practice of machine vibration diagnostics," *Mechanical Systems and Signal Processing*, vol. 4, no. 6, pp. 483–493, 1990.
- [56] T. Barszcz and M. Straczkiewicz, "Novel intuitive hierarchical structures for condition monitoring system of wind turbines," *Diagnostyka*, vol. 14, no. 3, pp. 53–60, 2013.
- [57] ISO, "Mechanical vibration—evaluation of machine vibration by measurements on non-rotating parts—part I: general guidelines," ISO 10816, International Organization for Standardization, Geneva, Switzerland, 1995.

## Research Article

# Application of Reassigned Wavelet Scalogram in Wind Turbine Planetary Gearbox Fault Diagnosis under Nonstationary Conditions

**Xiaowang Chen and Zhipeng Feng**

*School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China*

Correspondence should be addressed to Zhipeng Feng; [fengzpz@ustb.edu.cn](mailto:fengzpz@ustb.edu.cn)

Received 27 June 2015; Accepted 31 August 2015

Academic Editor: Chuan Li

Copyright © 2016 X. Chen and Z. Feng. 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.

Wind turbine planetary gearboxes often run under nonstationary conditions due to volatile wind conditions, thus resulting in nonstationary vibration signals. Time-frequency analysis gives insight into the structure of an arbitrary nonstationary signal in joint time-frequency domain, but conventional time-frequency representations suffer from either time-frequency smearing or cross-term interferences. Reassigned wavelet scalogram has merits of fine time-frequency resolution and cross-term free nature but has very limited applications in machinery fault diagnosis. In this paper, we use reassigned wavelet scalogram to extract fault feature from wind turbine planetary gearbox vibration signals. Both experimental and in situ vibration signals are used to evaluate the effectiveness of reassigned wavelet scalogram in fault diagnosis of wind turbine planetary gearbox. For experimental evaluation, the gear characteristic instantaneous frequency curves on time-frequency plane are clearly pinpointed in both local and distributed sun gear fault cases. For in situ evaluation, the periodical impulses due to planet gear fault are also clearly identified. The results verify the feasibility and effectiveness of reassigned wavelet scalogram in planetary gearbox fault diagnosis under nonstationary conditions.

## 1. Introduction

Wind turbines are playing an increasingly significant role in energy strategy. However, harsh working conditions, for example, wind gust, dust, and unpredictable heavy load, make the power transmission system prone to fault, which may lead to catastrophic breakdown or even productivity and economic losses. Planetary gearbox is one of the key components of wind turbine powertrain, for its merits of high gear transmission ratio and large load bearing capacity in a compact structure. Therefore, planetary gearbox fault diagnosis has been an important topic and has drawn increasing attentions recently [1–3].

Planetary gearbox exhibits quite unique dynamic behavior due to its special transmission structure and kinetics. McFadden [4] recognized that the varying phase angle of the planet gear vibration is responsible for spectra asymmetry. Inalpolat and Kahraman [5, 6] investigated the mechanisms of sideband harmonics in the vicinity of meshing frequency

and its integer multiples using a general model and further developed a dynamic model to predict modulation sidebands due to manufacturing errors. Feng and Zuo [7] proposed planetary gearbox vibration signal models and derived equations to calculate both local and distributed fault frequencies. These researches show that planetary gearbox vibration signals feature complexity and modulation and therefore have quite different spectral characteristics from that of fixed-shaft gearboxes.

In order to extract potential fault feature from planetary gearbox vibration signals, researchers proposed different methods. McNames [8] used continuous-time Fourier series (CTFS) analysis to invest the frequency characteristics. Chen et al. [9] applied ensemble multiwavelet transform with adaptive multiwavelet basis function to extract weak fault feature of a planetary gearbox. Lei et al. [10] proposed an adaptive stochastic resonance method to extract weak fault feature characteristics from noisy vibration signals of planetary gearbox. Considering the spectral complexity of

planetary gearbox vibration signals, Feng et al. [11] presented a joint amplitude and frequency demodulation analysis based on Teager energy operator and ensemble empirical mode decomposition (EEMD) to diagnose planetary gearbox fault. Barszcz and Randall [12] employed the spectral kurtosis to detect gear tooth crack in a wind turbine planetary gearbox. Sun et al. [13] customized the multiwavelets based on the redundant symmetric lifting scheme to detect the incipient pitting faults in planetary gearbox. Li et al. [14] extracted the fault characteristics frequencies of planetary gearbox based on empirical mode decomposition and adaptive multiscale morphological gradient filter.

Most of above works were based on an assumption of signal stationarity, that is, constant running speed and load. However, in real-world wind turbine applications, volatile wind conditions result in nonstationary vibration signals. For example, the running speed varies with respect to the unpredictable wind power, and transient strong wind may lead to a shock. Under nonstationary conditions, planetary gearbox vibration signals always feature intricate frequency component structure [15, 16], manifesting as time-varying amplitude modulation (AM) and frequency modulation (FM). Due to the nonstationary nature, conventional Fourier transform most likely fails to thoroughly reveal the true frequency structure. To the best of our knowledge, the literature reported on planetary gearbox fault diagnosis under nonstationary conditions has been very limited [17]. Bartelmus and Zimroz [18] proposed an indicator that reflects the linear dependence between the meshing frequency amplitude and the operating condition for condition monitoring of planetary gearboxes. Chaari et al. [19] developed a mathematical model to understand the dynamic behavior of planetary gear under variable load condition. Yang and Zhang [20] studied the running characteristics of wind turbine planetary gearbox considering time-varying speed and load.

Time-frequency analysis (TFA) is capable of revealing time-frequency structure despite time variation of frequency components and thus is suitable for analyzing nonstationary signals [21]. Linear time-frequency analysis methods like short time Fourier transform (STFT) and continuous wavelet transform (CWT) essentially represent signals by weighted sum of a series of bases localized in both time and frequency domains. But their time-frequency resolution is governed by Heisenberg uncertainty principle and resolution in one domain can only be improved at expense of the other, resulting in time-frequency smearing [22]. Bilinear time-frequency distribution presents the signal energy on time-frequency plane [23, 24]. The Wigner-Ville distribution (WVD) is the basis of almost all bilinear TFA methods, with merit of finest time-frequency resolution. However, it is subjected to inevitable cross-term interferences when analyzing multicomponent signals. Even for monocomponent signals, it still suffers from inner interference if the IF curve is nonlinear. Although improved methods, smoothed pseudo Wigner-Ville distribution (SPWVD), for example, can effectively suppress the cross-term interferences, they in turn deteriorate time-frequency resolution.

Fortunately, reassigned wavelet scalogram has merits of excellent time-frequency resolution and good suppression of

cross-terms and thus can clearly exhibit the time-frequency structure of nonstationary signals. However, its applications in machinery fault diagnosis have been very limited, and only a few published works focus on this topic. Peng et al. [25] studied the characteristics of rotor-stator rubbing impacts and applied both scalogram and reassigned scalogram for impact detection. Ma et al. [26] summarized the symptoms of rotor-stator rubbing fault based on reassigned scalogram.

In this paper we further extend the application of reassigned wavelet scalogram to planetary gearbox fault diagnosis under nonstationary conditions [27] and validate its effectiveness in extracting gear fault features. The remainder of this paper is structured as follows. In Section 2, the detailed algorithm of scalogram and reassigned scalogram is introduced. In Section 3, a numerical simulated nonstationary signal encompassing a time-varying frequency and periodical impulses is analyzed to illustrate the performance of reassigned scalogram. In Sections 4 and 5, the effectiveness of reassigned scalogram is further validated by analysis of lab experimental planetary gearbox vibration signals in both local and distributed sun gear fault cases, as well as in situ wind turbine planetary gearbox vibration signals measured from a wind farm. Finally, conclusions are drawn in Section 6.

## 2. Reassigned Wavelet Scalogram

*2.1. Continuous Wavelet Transform.* Wavelet transform (WT) has advantage of better time-frequency resolution over short time Fourier transform (STFT). Rather than using a constant window function for whole time-frequency plane analysis, it employs different window functions for analyzing different frequency bands of the signal. Therefore, WT is popular among linear TFA methods.

For an arbitrary finite-energy signal  $x(t) \in L^2(\mathbb{R})$ , its CWT is defined as

$$\begin{aligned} W_x(a, b; \psi) &= \langle x(t), \psi_{a,b}(t) \rangle \\ &= a^{-1/2} \int_{-\infty}^{\infty} x(t) \psi_{a,b}^*(t) dt, \end{aligned} \quad (1)$$

where asterisk  $*$  stands for complex conjugate,  $a > 0$ ,  $a$  and  $b$ , respectively, represent the scale and time offset, and  $\psi_{a,b}(t)$  is the family of wavelets generated by dilation and translation from the mother wavelet  $\psi(t)$ :

$$\psi_{a,b}(t) = a^{-1/2} \psi\left(\frac{t-b}{a}\right). \quad (2)$$

Morlet wavelet is most commonly used in CWT, and it is defined as

$$\psi(t) = \pi^{-1/4} \left( e^{-i\omega_0 t} - e^{-\omega_0^2/2} \right) e^{-t^2/2}. \quad (3)$$

The wavelet scalogram is defined as

$$S_x(a, b; \psi) = |W_x(a, b; \psi)|^2. \quad (4)$$

*2.2. Reassigned Wavelet Scalogram.* Although wavelet scalogram can reveal the constituent frequency components of

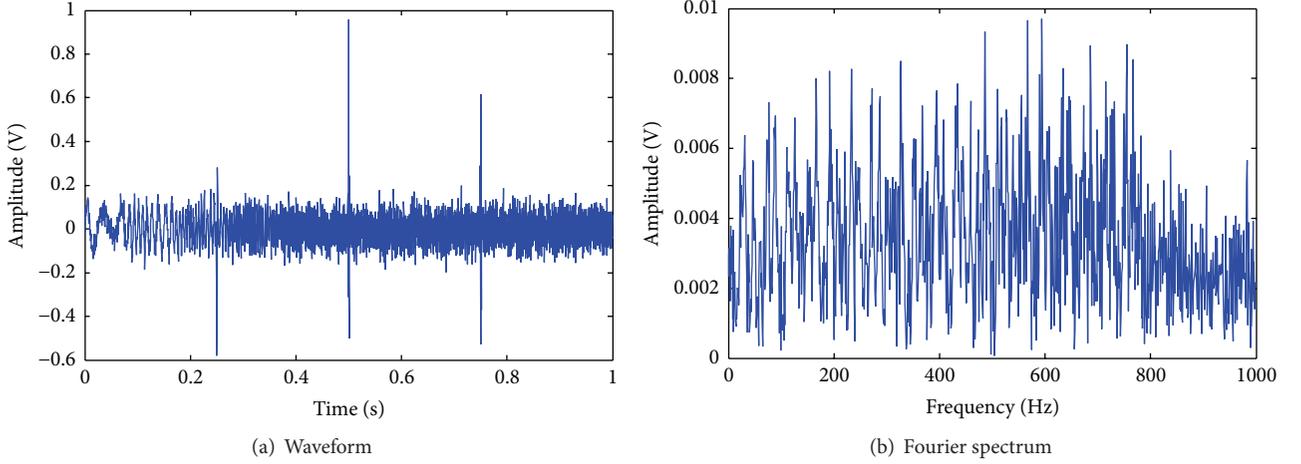


FIGURE 1: Simulated signal.

a signal and show how the energy of each component varies with time and frequency, its time-frequency resolution is still subject to the Heisenberg-Gabor inequality; that is, fine time localization and frequency resolution can hardly be obtained simultaneously. For example, it has fine time localization but coarse frequency resolution in high-frequency regions, while it has coarse time localization but fine frequency resolution in low-frequency regions. Therefore, for nonstationary signals with instantaneous frequency curves changing in a wide frequency region, the time-frequency blurring phenomenon may arise, thus hindering an accurate fault feature extraction.

In order to improve the readability of conventional wavelet scalograms, Auger and Flandrin [28] proposed the reassignment method. The wavelet scalogram  $S_x(a, b; \psi)$  does not represent the signal energy at the location  $(a, b)$  on the time-scale/frequency plane but the mean energy in the vicinity of location  $(a, b)$ . Therefore, it is reasonable to reallocate the signal energy  $S_x(a, b; \psi)$  to the center of gravity in the vicinity of location  $(a, b)$ . The new coordinates become

$$\begin{aligned} \frac{\omega_0}{a'(b, a)} &= \frac{\omega_0}{a} + \text{Im} \left\{ \frac{\text{CWT}_x(a, b; \hat{\psi}) \text{CWT}_x^*(a, b; \psi)}{2\pi a |\text{CWT}_x(a, b; \psi)|^2} \right\}, \\ b'(a, b) &= b - \text{Re} \left\{ a \frac{\text{CWT}_x(a, b; \psi') \text{CWT}_x^*(a, b; \psi)}{|\text{CWT}_x(a, b; \psi)|^2} \right\}, \end{aligned} \quad (5)$$

where

$$\begin{aligned} \psi'(t) &= t\psi(t), \\ \hat{\psi}(t) &= \frac{d\psi}{dt}(t). \end{aligned} \quad (6)$$

Thus the reassigned wavelet scalogram is cast as

$$\begin{aligned} RS_x(a', b') &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(\frac{\hat{a}}{a}\right)^2 S_x(a, b) \delta[\hat{b} - b'(a, b)] \\ &\cdot \delta[\hat{a} - a'(a, b)] da db. \end{aligned} \quad (7)$$

### 3. Simulation Evaluation

In this section, we examine the performance of reassigned wavelet scalogram in identifying time-varying frequency components and transient impulses characteristic of typical gear fault under nonstationary conditions. A numerical simulated signal is generated:

$$\begin{aligned} s(t) &= A \sin \left( 2\pi \left( \int_{\text{chirp}} 800t + \int_{\text{FM}} 50 \cos(2\pi 10t) \right) \right) \\ &+ \underset{\text{impulse}}{\delta(t)} + n(t), \end{aligned} \quad (8)$$

where  $t = 0:0.0005:1$  s,  $A = 0.1$  is the amplitude of time-varying frequency component, and  $\delta(t)$  stands for impulses at time instant of 0.25, 0.5, and 0.75 s. To mimic background noise, a Gaussian white noise  $n(t)$  at signal to noise ratio (SNR) of 8 dB is added. This signal encompasses three nonstationary conditions, that is, linear frequency modulation (chirp), frequency fluctuations (FM), and periodical impulses. Its waveform and Fourier spectrum are plotted in Figures 1(a) and 1(b).

From the signal waveform, Figure 1(a), we can clearly see the time instant of the three impulses but cannot identify the period of time-varying frequency components. Neither can we do from the Fourier spectrum, Figure 1(b), due to the nonstationarity of the simulated signal.

Figure 2 shows the time-frequency analysis results of the signal. In the wavelet scalogram, Figure 2(a), the major signal components and their time variation are roughly

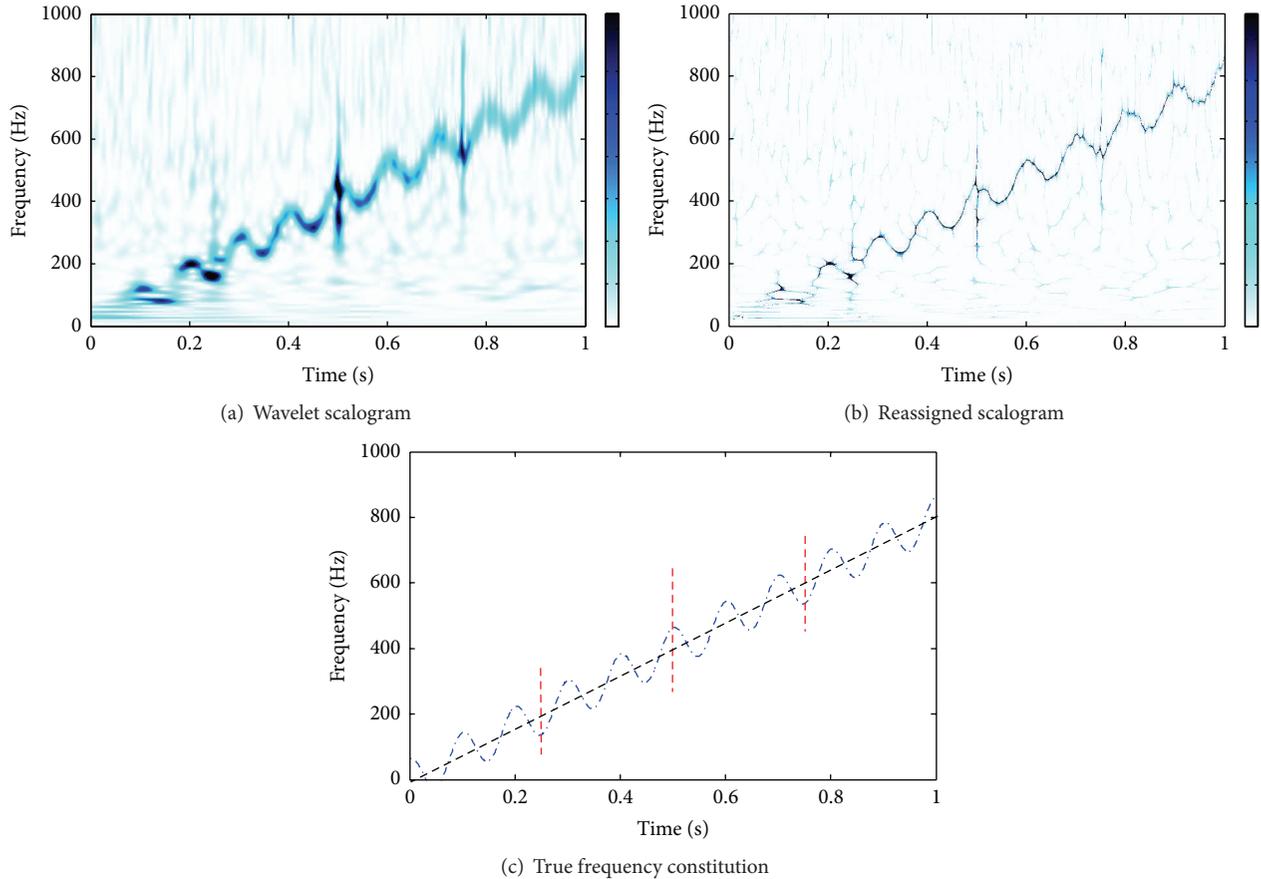


FIGURE 2: TFRs of the simulated signal.

shown: a nearly sinusoidal fluctuating frequency rides on a linear trend, together with three impulses. However, time-frequency blurring occurs because the instantaneous frequency of the time-varying frequency component is nonlinear. This results in a lower time-frequency resolution and may hinder an accurate time-frequency localization of each signal component.

In the reassigned scalogram, Figure 2(b), both the time-varying frequency component and the three impulses are clearly displayed, owing to the much better time-frequency resolution. More importantly, the identified instantaneous frequency curve of the time-varying component and the time-frequency distribution of the three impulses match well with the true time-frequency constitution shown in Figure 2(c). This result demonstrates the capability of reassigned scalogram in identifying both time-varying frequency component and transient impulses. Therefore, reassigned scalogram has the potential to diagnose planetary gearbox fault under nonstationary conditions.

#### 4. Experimental Evaluations

Experiments were carried out on a wind turbine drivetrain test rig in University of Ottawa lab, to examine the effectiveness of reassigned wavelet scalogram on real-world planetary gearbox fault diagnosis.

TABLE 1: Fixed-shaft gearbox configuration.

Gear	Number of gear teeth	
	Stage 1	Stage 2
Input	32	—
Intermediate	80	40
Output	—	72

*4.1. Experimental Settings.* In order to simulate the working condition of real-world planetary gearbox in wind farms, a wind turbine drivetrain composed of a two-stage planetary gearbox and a two-stage fixed-shaft gearbox is set up, as shown in Figure 3. The power flow path is as follows: drive motor  $\rightarrow$  fixed-shaft gearbox  $\rightarrow$  planetary gearbox  $\rightarrow$  main shaft  $\rightarrow$  magnetic powder brake. The configuration parameters of the two gearboxes are listed in Tables 1 and 2, respectively.

Accelerometers are mounted on the top of planetary gearbox casing to measure planetary gearbox vibration signal at a sampling frequency of 20,000 Hz. A load of 16.3 N·m is applied at the output end of planetary gearbox stage 2 by loading device. In consideration of the wind turbine blade rated speed of 25 rpm, that is, 0.417 Hz, we collect the vibration data during specific speed-down processes.

TABLE 2: Planetary gearbox configuration.

Gear	Number of gear teeth	
	Stage 1	Stage 2
Ring	100	100
Planet	40 (4)	36 (4)
Sun	20	28

Note: the number of planet gears is in the parenthesis.

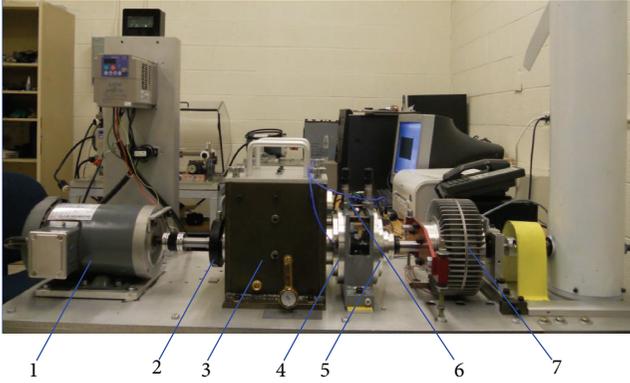


FIGURE 3: WT drivetrain test rig: (1) motor, (2) tachometer, (3) fixed-shaft gearbox, (4) planetary gearbox stage 1, (5) planetary gearbox stage 2, (6) accelerometers, and (7) loading device.

The drive motor speed decreases from 60 Hz to 50 Hz, and the corresponding blade speed varies continuously from 0.486 Hz to 0.405 Hz, within the typical rated speed range of wind turbine blades in real world. The motor speed is meanwhile recorded at a sampling frequency of 20 Hz.

To simulate gear faults, two types of gear damage were introduced to sun gears of planetary gearbox stage 1 and stage 2, respectively, while other gears are healthy. Stage 1 sun gear has wear on every tooth and stage 2 sun gear has an individual chipped tooth, as illustrated in Figures 4(a) and 4(b). Therefore three sets of experiments are carried out as baseline, wear, and chipping.

Given the drive motor speed  $f_d(t)$ , we can calculate the characteristic frequencies of both stages of planetary gearbox according to the aforementioned gearbox configuration, as listed in Table 3. For example, at the motor speed  $f_d(t) = 60$  Hz, the sun gear related characteristic frequencies are calculated as  $f_{m1}(t) = 222.2$  Hz,  $f_{s1}^{(r)}(t) = 13.3$  Hz, and  $f_{s1}(t) = 44.4$  Hz for stage 1 and  $f_{m2}(t) = 48.6$  Hz,  $f_{s2}^{(r)}(t) = 2.2$  Hz, and  $f_{s2}(t) = 6.9$  Hz for stage 2.

**4.2. Baseline Signal Analysis.** Firstly, we analyze the baseline signal for comparison study with the gear fault cases. In baseline case all gears are healthy. Figures 5(a) and 5(b) exhibit the waveform and Fourier spectrum of the baseline vibration signal, and Figure 5(c) shows the corresponding speed. The maximum meshing frequency is 222.2 Hz for stage 1 and 48.6 Hz for stage 2. Since gear fault usually results in sidebands around meshing frequency and its harmonics, we focus on two frequency ranges of 0–400 Hz and 0–80 Hz

TABLE 3: Planetary gearbox characteristic frequencies.

Frequency	Stage 1	Stage 2
Meshing $f_m(t)$	$(100/27) f_d(t)$	$(175/216) f_d(t)$
Sun rotating $f_s^{(r)}(t)$	$(2/9) f_d(t)$	$(1/27) f_d(t)$
Planet carrier rotating $f_c(t)$	$(1/27) f_d(t)$	$(7/864) f_d(t)$
Sun fault $f_s(t)$	$(20/27) f_d(t)$	$(175/1512) f_d(t)$
Planet fault $f_p(t)$	$(5/54) f_d(t)$	$(175/7776) f_d(t)$
Ring fault $f_r(t)$	$(4/27) f_d(t)$	$(175/5400) f_d(t)$

which cover 3/2 times stage 1 and stage 2 meshing frequencies, respectively.

Figures 6(a) and 6(b) show the reassigned wavelet scalogram of the baseline signal. It clearly displays the constituent frequency components and their variations along with time. The value of each characteristic frequency at any time can be calculated according to the equations in Table 3; thus the instantaneous frequency curves in the time-frequency analysis results can be identified. In Figure 6(a), the gear meshing frequency minus the sun gear fault frequency  $f_{m1}(t) - f_{s1}(t)$  and the gear meshing frequency minus four times the sun rotating frequency  $f_{m1}(t) - 4f_{s1}^{(r)}(t)$  are presented, but they are not dominant in comparison with the other components almost uniformly distributed in higher frequency region [200, 400] Hz. Even for healthy gearboxes, manufacturing errors and minor defects are inevitable. Hence, the presence of these frequency components does not imply the sun gear fault. In Figure 6(b), only a few weak components are visible, showing that the gearbox is healthy, which is consistent with the experimental setting.

**4.3. Detection of Sun Gear Wear.** In this section we replaced the healthy stage 1 sun gear with a worn one while other settings are unchanged and detect the sun gear wear by comparing worn sun gear signal with baseline signal in joint time and frequency domain. The waveform, Fourier spectrum, and corresponding motor speed are plotted in Figures 7(a)–7(c). Since gear damage locates on stage 1 sun gear, in this case we focus on 0–400 Hz to look for fault symptom around stage 1 meshing frequency.

Figure 8 presents the reassigned wavelet scalogram of faulty sun signal. Compared to the baseline signal case in Figure 6(a), the gear meshing frequency minus the sun fault characteristic frequency  $f_{m1}(t) - f_{s1}(t)$  and the gear meshing frequency minus four times the sun rotating frequency  $f_{m1}(t) - 4f_{s1}^{(r)}(t)$  still remain. Besides, more characteristic frequencies appear, including the gear meshing frequency  $f_{m1}(t)$ , the gear meshing frequency minus 7/4 times the sun fault characteristic frequency  $f_{m1}(t) - (7/4)f_{s1}(t)$ , and the gear meshing frequency plus three times the sun fault characteristic frequency  $f_{m1}(t) + 3f_{s1}(t)$ . All these frequency components have a pronounced magnitude, and they are all related to the sun gear. This finding shows that the 1st stage sun gear is faulty, in consistency with the experimental setting.

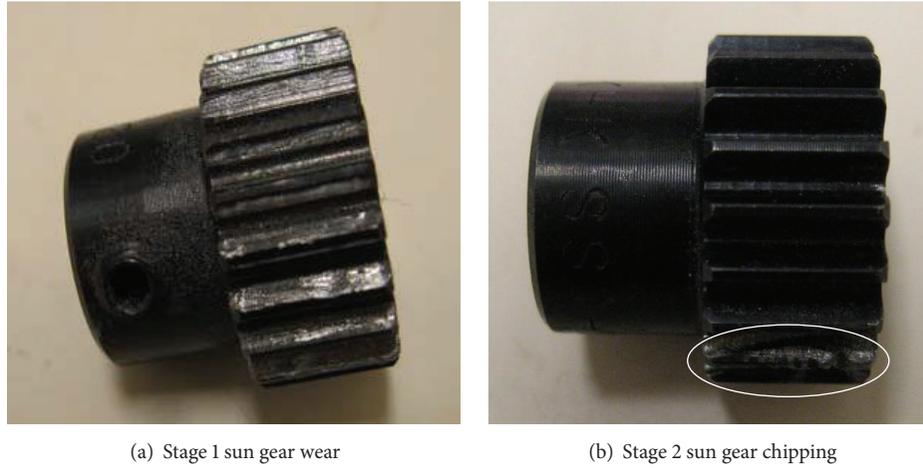


FIGURE 4: Sun gear damage.

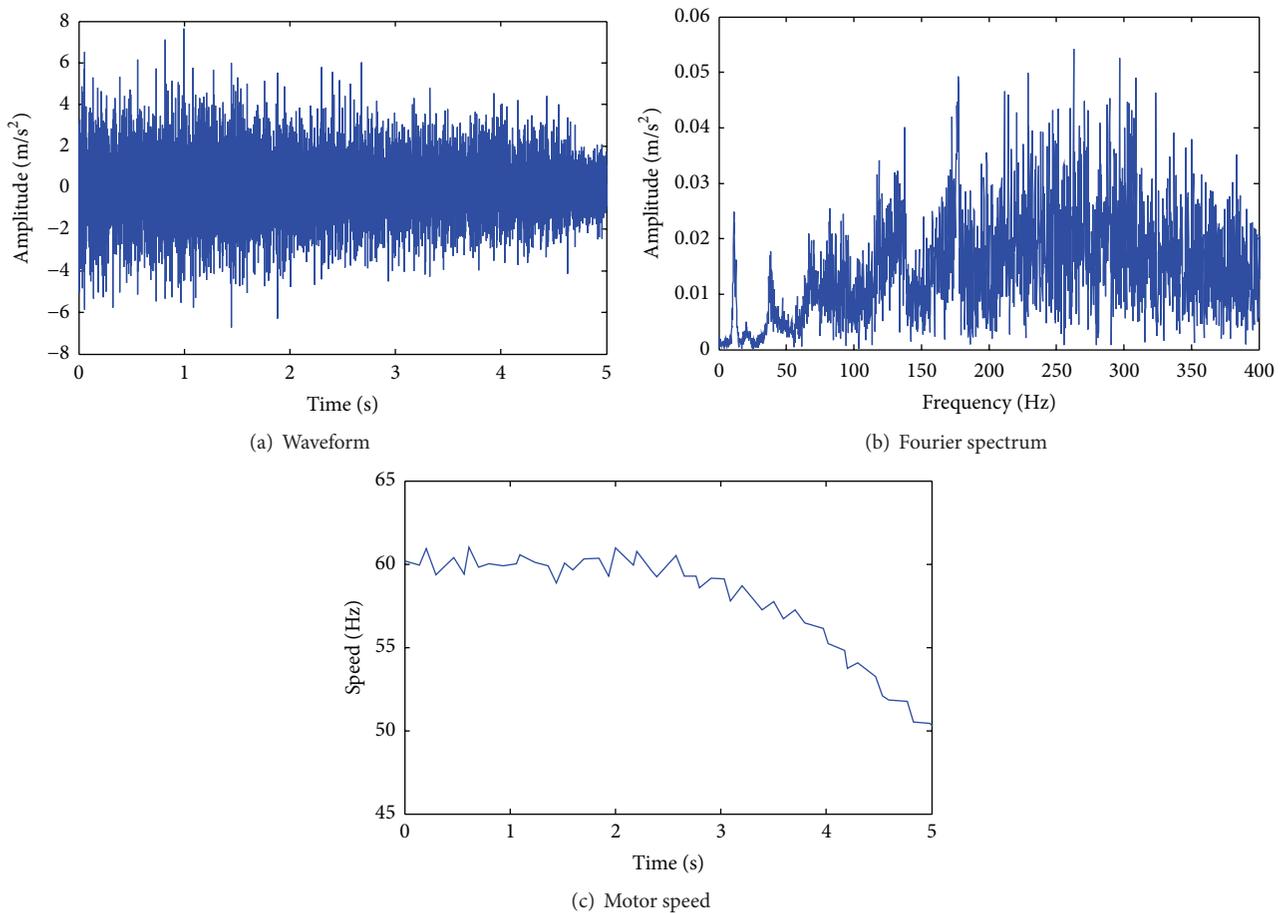


FIGURE 5: Baseline signal.

*4.4. Detection of Sun Gear Chipping.* Another type of gear fault, that is, gear chipping, is brought in by replacing the healthy stage 2 sun gear with a chipped one while other settings remain the same with baseline case. Still we detect the fault by comparing chipped sun gear signal with baseline

signal in joint time and frequency domain, but another frequency range of 0–80 Hz is focused on since the maximum stage 2 meshing frequency is 48.6 Hz. The waveform, Fourier spectrum, and corresponding motor speed of chipped sun gear signal are presented in Figures 9(a)–9(c).

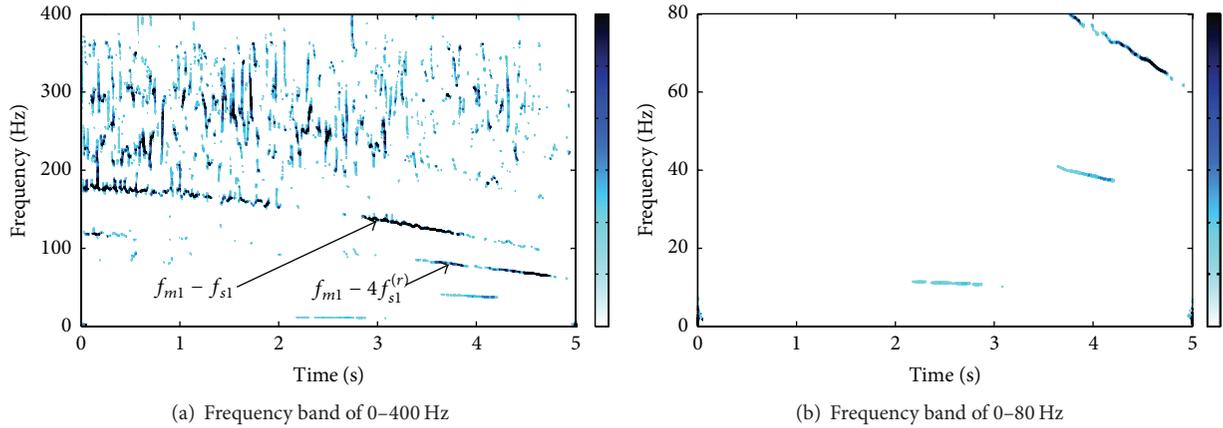


FIGURE 6: Reassigned scalogram of baseline signal.

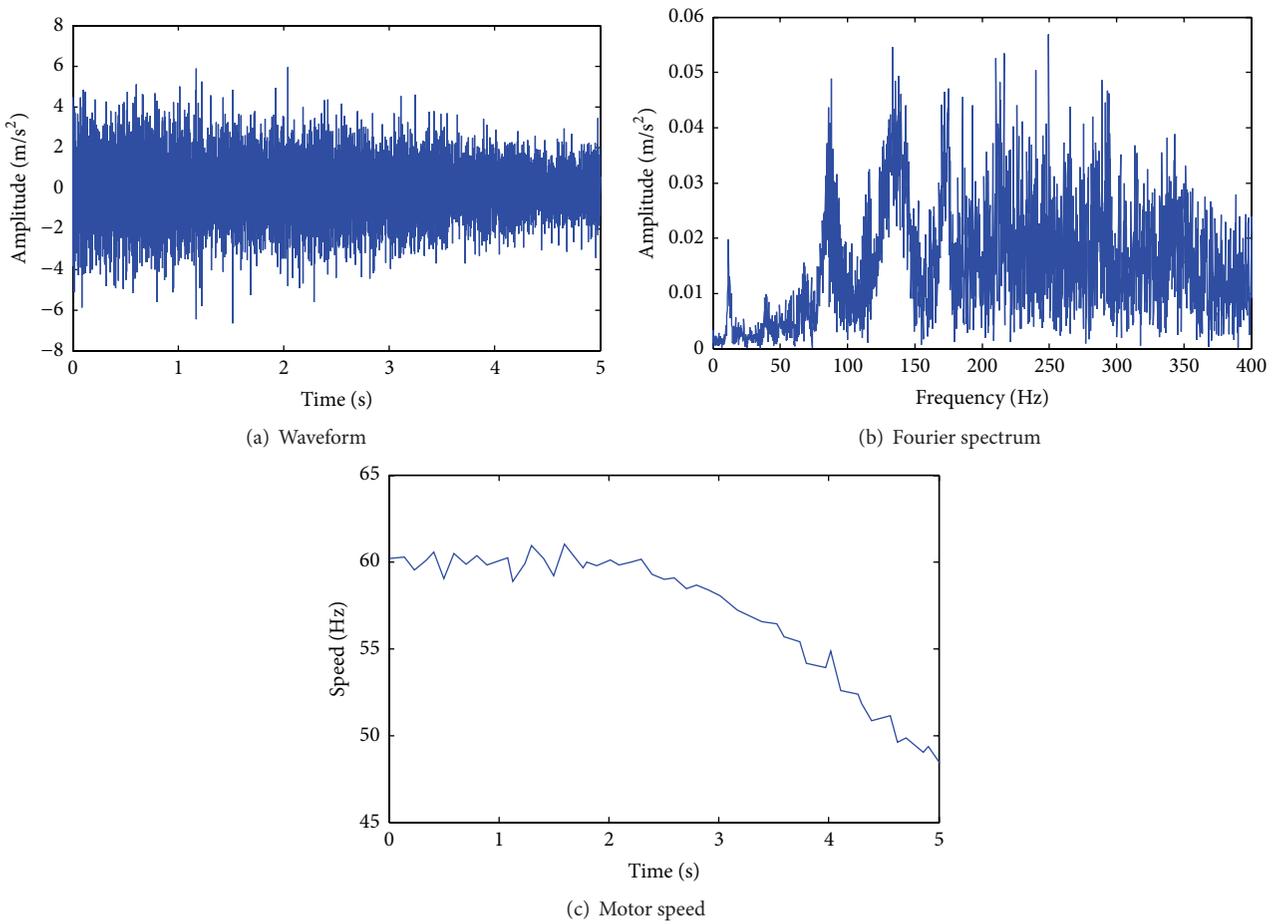


FIGURE 7: Sun gear wear signal.

In the reassigned scalogram, Figure 10, except for the motor rotating frequency component  $f_d$ , two new characteristic frequency components are also evident. One corresponds to the meshing frequency minus the third harmonic of the sum of the sun gear fault characteristic frequency and the sun gear rotating frequency of the 2nd stage,  $f_{m2} - 3[f_{s2}^{(r)} + f_{s2}]$ . The other is very close to motor rotating frequency, but

by virtue of the fine time-frequency resolution, it is identified to be the sum of the meshing frequency and the second harmonic of sun gear fault characteristic frequency of the 2nd stage,  $f_{m2} + 2f_{s2}$  (as labeled with a black dashed line). These two extra frequency components correlate to the 2nd stage sun gear fault characteristic frequency and have prominent magnitude, thus implying the 2nd stage sun gear fault.

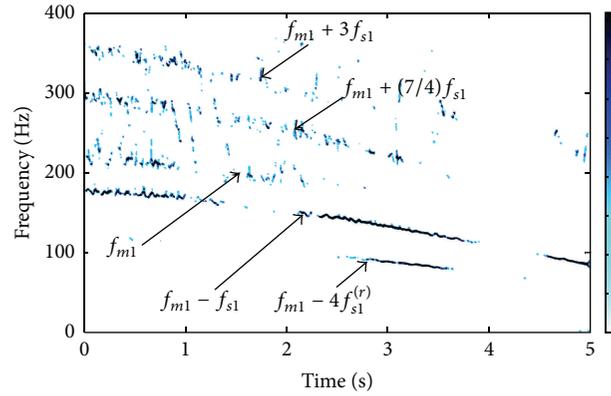


FIGURE 8: Reassigned scalogram of sun gear wear signal.

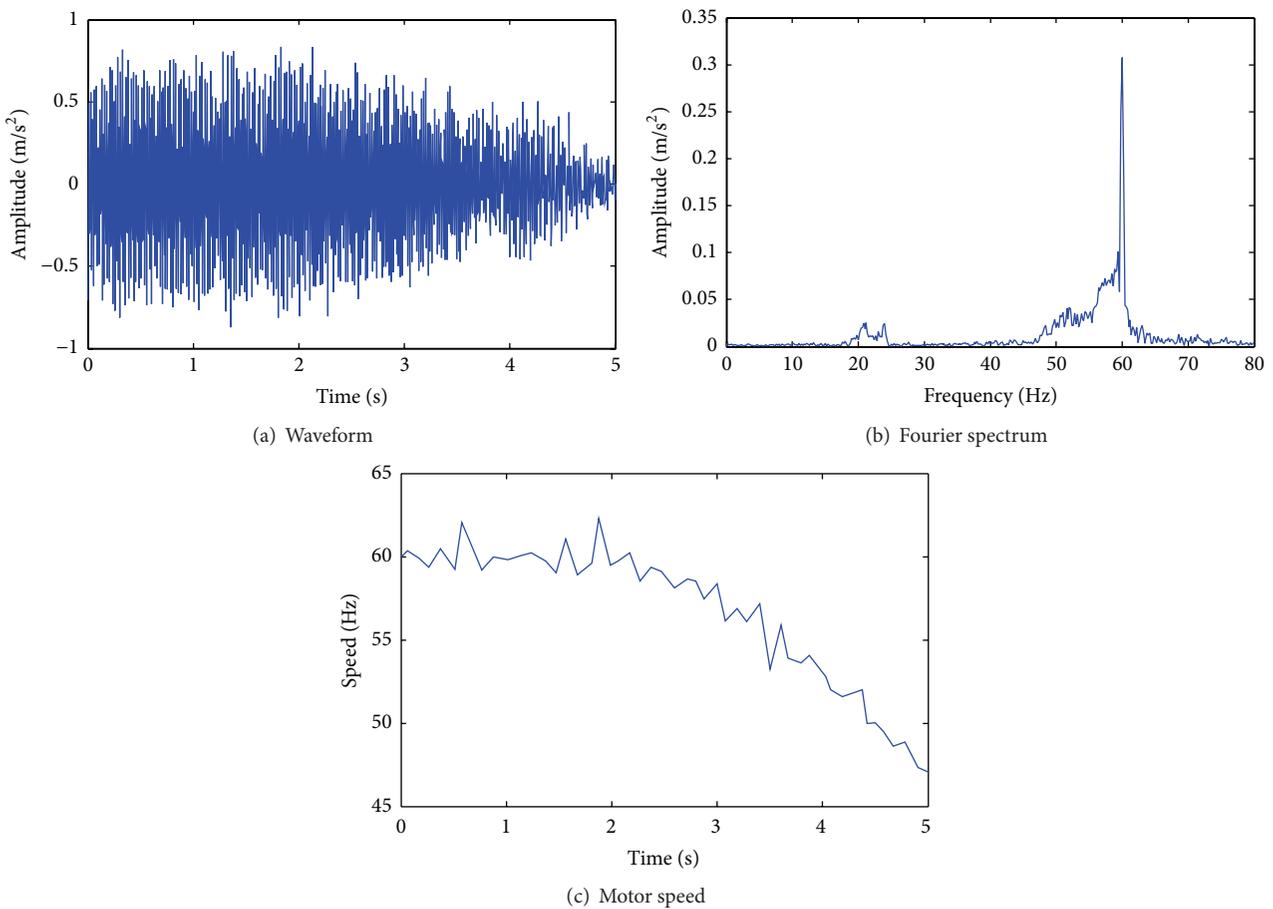


FIGURE 9: Chipped sun gear signal.

The sun gear fault usually exhibits weak symptom in the vibration signal, because the sun gear is relatively far from the accelerometer mounted on the gearbox casing. But the time-varying sun gear fault characteristic frequencies are still identified, owing to the fine time-frequency resolution and capability in suppressing cross-term interferences of reassigned scalogram. This shows its effectiveness in extracting planetary gearbox fault features under nonstationary conditions.

## 5. In Situ Signal Analysis

In this section, we analyze the in situ collected signals of a wind turbine planetary gearbox in a wind farm, to further validate the reassigned scalogram for engineering applications.

*5.1. Gearbox Configuration.* The wind turbine drivetrain in a wind farm to be analyzed is composed of a one-stage

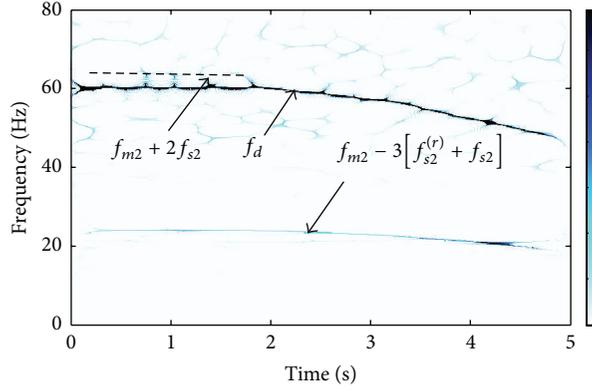


FIGURE 10: Reassigned scalogram of sun gear chipping signal.

TABLE 4: Planetary gearbox configuration parameters.

Gear	Number of gear teeth
Ring	96
Planet	37 (3)
Sun	21

Note: the number of planet gears is in the parenthesis.

TABLE 5: Fixed-shaft gearbox configuration parameters.

Gear	Number of gear teeth	
	Input stage	Output stage
Input gear	71	—
Intermediate shaft gear	19	89
Output gear	—	30

planetary gearbox and a two-stage fixed-shaft gearbox. The wind power flow path is as follows: blades → main shaft → planetary gearbox → fixed-shaft gearbox → generator. Considering the demand of wind farm information protection, photos of the wind turbine drivetrain are not attached in the paper. Tables 4 and 5 list the configuration parameters of the fixed-shaft gearbox and the planetary gearbox, respectively.

The rated rotating speed of the wind turbine blades is 25 rpm, that is, 0.416 Hz. At the rated speed, the characteristic frequencies of the drivetrain gearboxes are calculated as listed in Table 6. They will be used to estimate the frequency components of vibration signals, for the wind turbine speed usually fluctuates around the rated speed.

The vibration signals from an accelerometer mounted on the planetary gearbox casing of both a normal and a faulty wind turbine are collected. Since the gear fault always results in abnormal sidebands or impulses centered around gear meshing frequency and its harmonics, a frequency band of 0–70 Hz, which covers over 3/2 times the planetary gearbox meshing frequency (42.03 Hz), will be focused on.

## 5.2. Signal Analysis

**5.2.1. Normal Case.** The waveform and Fourier spectrum of the normal wind turbine planetary gearbox vibration

signal are plotted in Figures 11(a) and 11(b). Nevertheless, details of nonstationary characteristics are still veiled due to information loss. Figure 12 shows the reassigned wavelet scalogram of the normal wind turbine planetary gearbox vibration signal. The generator rotating frequency 27 Hz and the planetary gearbox meshing frequency 42 Hz are dominant. The second harmonic of generator rotating frequency 54 Hz is also visible, but in a relatively low magnitude. All these frequencies are time-varying due to the unsteady speed of wind turbine blades caused by the nonstationary wind power. Thus it is difficult to identify frequency components via traditional Fourier transform based spectral analysis.

**5.2.2. Faulty Planet Gear Case.** In the faulty case, similarly the waveform and Fourier spectrum of faulty planetary gearbox vibration signal are presented in Figures 13(a) and 13(b). Due to the unsteady motor speed and nonstationary conditions, it is difficult to identify frequency components via traditional spectral analysis. Figure 14 shows the reassigned scalogram of the faulty wind turbine planetary gearbox vibration signal. It has a time-frequency structure far different from that in the normal case. Only the planetary gearbox meshing frequency is dominant. In addition, some big impulses appear almost periodically, as labeled by evenly spaced vertical lines with capitals A–G in Figure 14(a). The interval between adjacent impulses corresponds approximately to the planet carrier rotating frequency 0.44 Hz. Furthermore, if we zoom in Figure 14(a) and focus on the interval between two consecutive primary impulses, for example, B and C, some small almost periodic impulses also appear, as indicated by lower-case letters a–f in Figure 14(b). The interval between these adjacent small impulses corresponds approximately to the planet gear fault characteristic frequency, that is, 2.27 Hz. These findings imply the planet gear fault, in accordance with the check result after opening the planetary gearbox.

The planet gear fault symptom is complicated, because the planet gear not only spins around its own axis but also revolves around the sun gear with the planet carrier, and meanwhile it meshes with both the sun gear and the ring gear. But the gear fault induced impulses are still identified, illustrating the effectiveness of reassigned scalogram in extracting complicated fault features under nonstationary conditions.

TABLE 6: Characteristic frequencies of drivetrain gearbox.

Mesh	Planetary			Fixed-shaft stage 1			Fixed-shaft stage 2		
	Ring	Planet	Sun	Mesh	Drive	Driven	Mesh	Drive	Driven
42.03	0.44	2.27	2.01	173.18	2.44	9.11	811.20	9.11	27.04

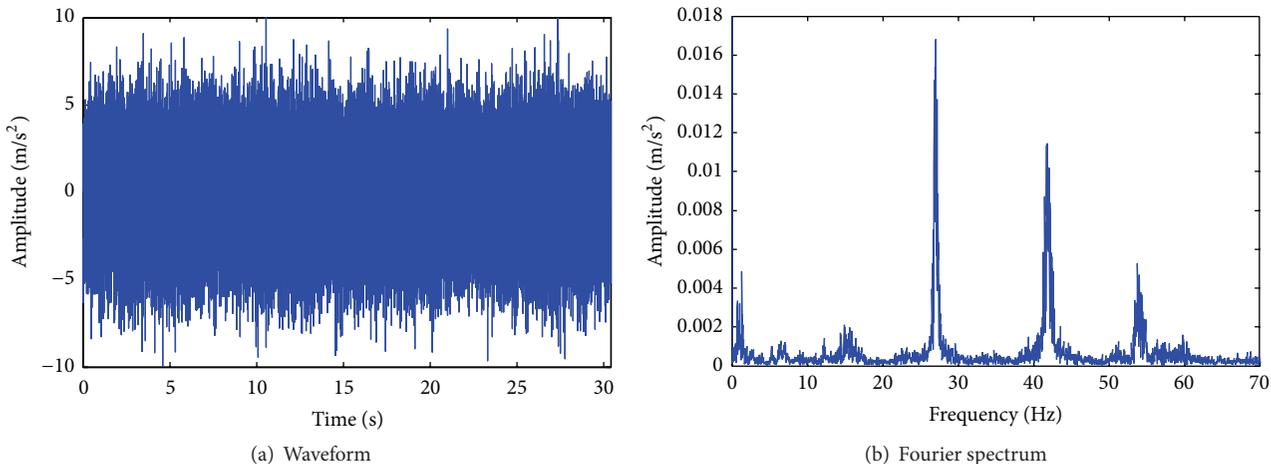


FIGURE 11: Normal signal.

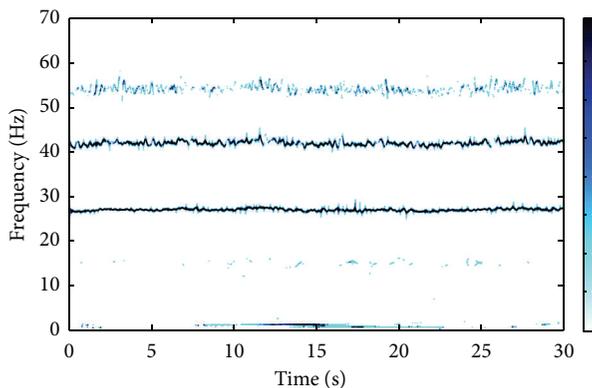


FIGURE 12: Reassigned scalogram of normal signal.

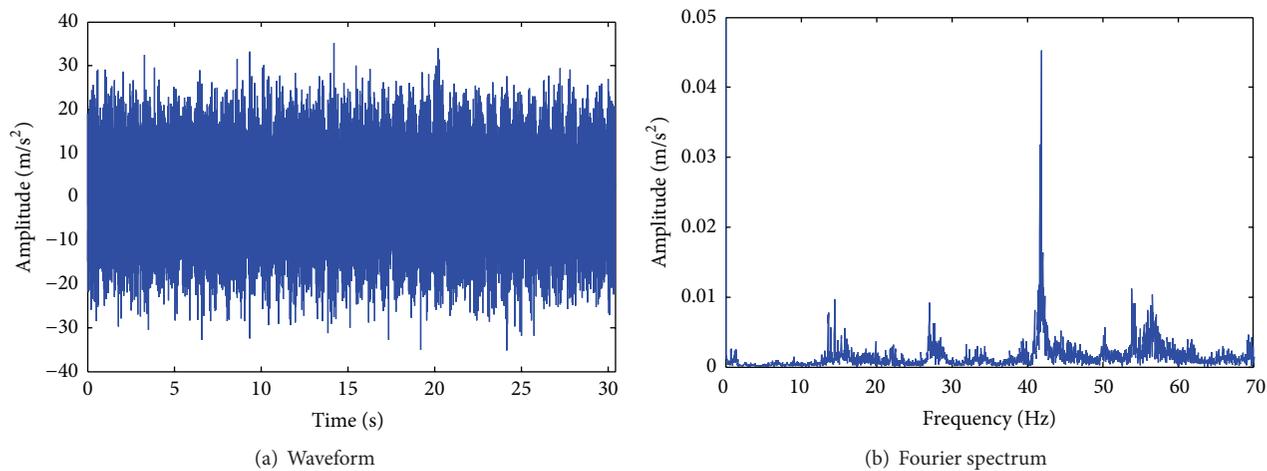


FIGURE 13: Faulty signal.

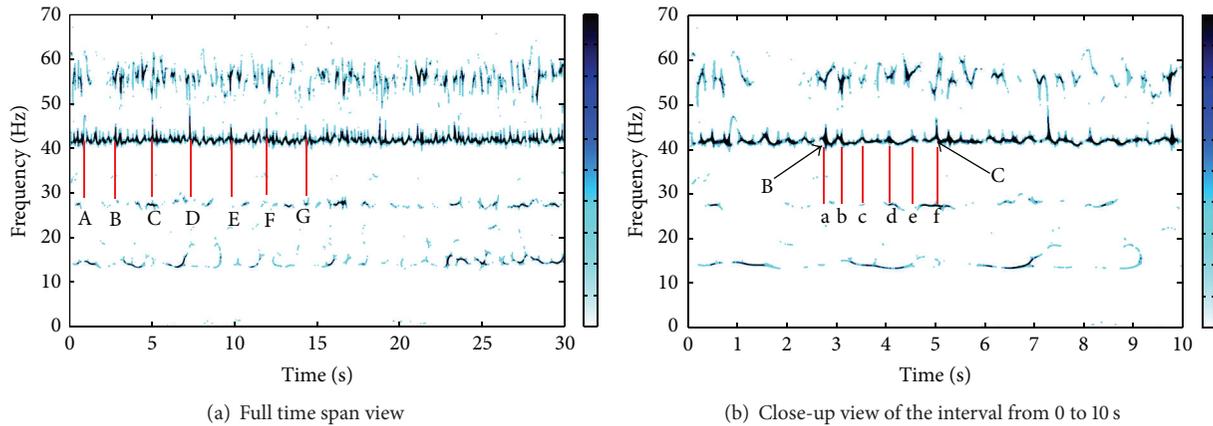


FIGURE 14: Reassigned scalogram of faulty signal.

## 6. Conclusions

The reassigned wavelet scalogram has excellent merits of fine time-frequency resolution and capability in suppressing cross-term interferences and thus is suited to extract the time-frequency features of nonstationary signals. We have validated the effectiveness of this method for planetary gearbox fault diagnosis under nonstationary conditions, using both lab experimental signals and in situ measured datasets. Both the time-varying gear characteristic frequencies and gear fault induced impulses are identified, and thereby the localized and distributed sun gear fault, as well as naturally developed planet gear fault, are diagnosed, even though they have weak and complicated fault symptoms under nonstationary conditions.

## Conflict of Interests

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

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (11272047 and 51475038), the Program for New Century Excellent Talents in University (NCET-12-0775), and the Natural Sciences and Engineering Research Council of Canada.

## References

- [1] P. D. Samuel and D. J. Pines, "A review of vibration-based techniques for helicopter transmission diagnostics," *Journal of Sound and Vibration*, vol. 282, no. 1-2, pp. 475-508, 2005.
- [2] Y. Guo and R. G. Parker, "Analytical determination of mesh phase relations in general compound planetary gears," *Mechanism and Machine Theory*, vol. 46, no. 12, pp. 1869-1887, 2011.
- [3] Y. Lei, J. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: a review," *Measurement*, vol. 48, no. 1, pp. 292-305, 2014.
- [4] P. D. McFadden, "Window functions for the calculation of the time domain averages of the vibration of the individual planet gears and sun gear in an epicyclic gearbox," *Journal of Vibration and Acoustics, Transactions of the ASME*, vol. 116, no. 2, pp. 179-187, 1994.
- [5] M. Inalpolat and A. Kahraman, "A theoretical and experimental investigation of modulation sidebands of planetary gear sets," *Journal of Sound and Vibration*, vol. 323, no. 3-5, pp. 677-696, 2009.
- [6] M. Inalpolat and A. Kahraman, "A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors," *Journal of Sound and Vibration*, vol. 329, no. 4, pp. 371-393, 2010.
- [7] Z. Feng and M. J. Zuo, "Vibration signal models for fault diagnosis of planetary gearboxes," *Journal of Sound and Vibration*, vol. 331, no. 22, pp. 4919-4939, 2012.
- [8] J. McNames, "Fourier series analysis of epicyclic gearbox vibration," *Journal of Vibration and Acoustics*, vol. 124, no. 1, pp. 150-160, 2002.
- [9] J. Chen, C. Zhang, X. Zhang, Y. Zi, S. He, and Z. Yang, "Planetary gearbox condition monitoring of ship-based satellite communication antennas using ensemble multiwavelet analysis method," *Mechanical Systems and Signal Processing*, vol. 54-55, pp. 277-292, 2015.
- [10] Y. Lei, D. Han, J. Lin, and Z. He, "Planetary gearbox fault diagnosis using an adaptive stochastic resonance method," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 113-124, 2013.
- [11] Z. Feng, M. Liang, Y. Zhang, and S. Hou, "Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy separation," *Renewable Energy*, vol. 47, pp. 112-126, 2012.
- [12] T. Barszcz and R. B. Randall, "Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352-1365, 2009.
- [13] H. Sun, Y. Zi, Z. He, J. Yuan, X. Wang, and L. Chen, "Customized multiwavelets for planetary gearbox fault detection based on vibration sensor signals," *Sensors*, vol. 13, no. 1, pp. 1183-1209, 2013.
- [14] H. Li, J. Zhao, and W. Song, "An offline fault diagnosis method for planetary gearbox based on empirical mode decomposition

- and adaptive multi-scale morphological gradient filter,” *Journal of Vibroengineering*, vol. 17, pp. 705–719, 2015.
- [15] Z. Feng, F. Chu, and M. J. Zuo, “Time-frequency analysis of time-varying modulated signals based on improved energy separation by iterative generalized demodulation,” *Journal of Sound and Vibration*, vol. 330, no. 6, pp. 1225–1243, 2011.
- [16] Z. Feng, X. Chen, M. Liang, and F. Ma, “Time-frequency demodulation analysis based on iterative generalized demodulation for fault diagnosis of planetary gearbox under nonstationary conditions,” *Mechanical Systems and Signal Processing*, vol. 62–63, pp. 54–74, 2015.
- [17] Z. Feng and M. Liang, “Fault diagnosis of wind turbine planetary gearbox under nonstationary conditions via adaptive optimal kernel time-frequency analysis,” *Renewable Energy*, vol. 66, pp. 468–477, 2014.
- [18] W. Bartelmus and R. Zimroz, “Vibration condition monitoring of planetary gearbox under varying external load,” *Mechanical Systems and Signal Processing*, vol. 23, no. 1, pp. 246–257, 2009.
- [19] F. Chaari, M. S. Abbes, F. V. Rueda, A. F. del Rincon, and M. Haddar, “Analysis of planetary gear transmission in nonstationary operations,” *Frontiers of Mechanical Engineering*, vol. 8, no. 1, pp. 88–94, 2013.
- [20] J. Yang and L. Zhang, “Dynamic response and dynamic load of wind turbine planetary gear transmission system under changing excitation,” *Applied Mechanics and Materials*, vol. 121–126, pp. 2671–2675, 2011.
- [21] Z. Feng, M. Liang, and F. Chu, “Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples,” *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 165–205, 2013.
- [22] O. Rioul and P. Flandrin, “Time-scale energy distributions: a general class extending wavelet transforms,” *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1746–1757, 1992.
- [23] F. Hlawatsch and G. F. Boudreaux-Bartels, “Linear and quadratic time-frequency signal representations,” *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.
- [24] L. Cohen, “Time-frequency distributions: a review,” *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.
- [25] Z. K. Peng, F. L. Chu, and P. W. Tse, “Detection of the rubbing-caused impacts for rotor-stator fault diagnosis using reassigned scalogram,” *Mechanical Systems and Signal Processing*, vol. 19, no. 2, pp. 391–409, 2005.
- [26] H. Ma, T. Yu, Q. Han, Y. Zhang, B. Wen, and C. Xuelian, “Time-frequency features of two types of coupled rub-impact faults in rotor systems,” *Journal of Sound and Vibration*, vol. 321, no. 3–5, pp. 1109–1128, 2009.
- [27] X. Chen, Z. Feng, and M. Liang, “Fault feature extraction of planetary gearboxes under nonstationary conditions based on reassigned wavelet scalogram,” in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference*, pp. 294–299, Pisa, Italy, May 2015.
- [28] F. Auger and P. Flandrin, “Improving the readability of time-frequency and time-scale representations by the reassignment method,” *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995.

## Research Article

# Research on the Sparse Representation for Gearbox Compound Fault Features Using Wavelet Bases

Chunyan Luo,<sup>1</sup> Changqing Shen,<sup>2</sup> Wei Fan,<sup>1</sup> Gaigai Cai,<sup>1</sup>  
Weiguo Huang,<sup>1</sup> and Zhongkui Zhu<sup>1</sup>

<sup>1</sup>School of Urban Rail Transportation, Soochow University, Suzhou 215137, China

<sup>2</sup>School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215021, China

Correspondence should be addressed to Changqing Shen; cqshen@suda.edu.cn

Received 29 June 2015; Accepted 12 August 2015

Academic Editor: Dong Wang

Copyright © 2015 Chunyan Luo 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 research on gearbox fault diagnosis has been gaining increasing attention in recent years, especially on single fault diagnosis. In engineering practices, there is always more than one fault in the gearbox, which is demonstrated as compound fault. Hence, it is equally important for gearbox compound fault diagnosis. Both bearing and gear faults in the gearbox tend to result in different kinds of transient impulse responses in the captured signal and thus it is necessary to propose a potential approach for compound fault diagnosis. Sparse representation is one of the effective methods for feature extraction from strong background noise. Therefore, sparse representation under wavelet bases for compound fault features extraction is developed in this paper. With the proposed method, the different transient features of both bearing and gear can be separated and extracted. Both the simulated study and the practical application in the gearbox with compound fault verify the effectiveness of the proposed method.

## 1. Introduction

As a fundamental mechanical component for transmitting power, gearbox has been widely used in modern industry. Because of its complicated structure, hostile working conditions, and other reasons, gearbox is usually easily damaged and breaks down. Therefore, it is of great significance to develop proper condition monitoring and fault diagnosis method for gearbox in order to prevent the unexpected machine fault during operation and even casualties [1].

When a fault is occurring in a bearing or a gear, both of which are vital components for gearbox, periodic transient impulses appear in its generated vibration signals. Researches have shown that the transients in the captured signal always comprise the important information of fault feature from the defective component [2]. Therefore, gearbox fault feature extraction can be transformed into extracting the transients in the generated signal.

During the past two decades, various advanced signal processing methods have been proposed for effective fault feature extraction in rotary machines. Time-frequency

analysis [3, 4], whose analysis of a signal is performed in both time and frequency domains, is developed for nonstationary signal processing. As a typical method of time-frequency analysis, Wigner-Ville Distribution (WVD) has proven its effectiveness in mechanical fault diagnosis [5, 6]. Wavelet transforms, which decompose an original signal into different scales with varying frequency bandwidths [7, 8], are also used to extract fault-related information of rotary machines [9, 10]. Empirical mode decomposition (EMD) [11], a self-adaptive signal processing technique, decomposes the nonlinear and nonstationary signal into a set of intrinsic mode functions (IMF). It has also been introduced to fault diagnosis of rotary machines [12, 13].

However, methods mentioned above are mostly used on the occasion that there is single fault in the machinery. A number of engineering practices have shown that there is usually more than one fault in a gearbox, which is demonstrated as compound fault [14]. When different faults exit simultaneously, vibration signal excited by several faults always contains different transient impulse responses, which makes it quite complex and difficult to identify each fault

from the observed signal in terms of traditional methods. Hence, some novel techniques for compound fault diagnosis have been developed gradually. The blind source separation (BSS) technique [15] can separate several original signals that cannot be observed directly from the superposed signals and has been used in extracting machinery faults features from different rotary components [16, 17]. However, the analyzed compound signals in BSS are usually derived from different channels through several transducers, which may bring some inconvenience in installing transducers during engineering application. Morphological component analysis (MCA) [18] was also developed for compound signal decomposition based on the morphological diversity of each component [19, 20]. However, MCA requests that the vibrations generated by each faulty component are totally irrelevant, which may lower the separation quality. Additionally, some intelligent methods based on models are proposed in recent years for compound faults signals separation [21, 22], but due to its difficulty in acquiring the appropriate data, these intelligent methods have not been widely promoted.

Meanwhile, there has been a growing interest in the study of sparse representation of signals recently [23, 24]. With an overcomplete dictionary which contains prototype signal-atoms, signals can be described as a sparse linear combination of these atoms [25, 26]. Till now, sparse representation has already been used in single fault feature extraction, and its excellent extraction property has been proven [27, 28]. According to the characteristics of gear fault vibration signal, Cai et al. [27] proposed the sparsity-enabled signal decomposition using tunable  $Q$ -factor wavelet transform and successfully extracted the fault feature of gear. In [28], Fan et al. constructed a sparse dictionary corresponding to the signal characteristic and combined Majorization-Minimization (MM) algorithm to extract the gear transient impulse responses sparsely.

Based on the engineering practical requirement, further research on the gearbox compound fault diagnosis should be conducted. On the basis of failure ratio investigation among the components in gearbox, the main failure components tend to be bearing and gear. However, there are few methods for extracting gearbox compound fault features of both bearing and gear faults. Hence, it is of much importance to pay attention to gearbox compound fault diagnosis when both bearing and gear faults occur. Due to its excellent self-adaptability, concise expression, and other merits, sparse representation is introduced in this paper for separating and extracting the gearbox compound fault features. Once there are both bearing and gear faults in a gearbox, there will always be two different kinds of transient impulse responses in the sampled signal. Considering the diversity in waveforms of each fault, different optimal wavelet bases thus can be constructed by correlation filtering [29, 30]. Then the constrained optimization algorithm with obtained special basis is incorporated to get a series of sparse coefficients which represent the specific fault. After representing the fault in sparse coefficients with the suitable constrained optimization algorithm one by one, the impulse time and the period parameter of each fault in the gearbox can be detected from the sparse coefficients properly.

The rest of this paper is outlined as follows. In Section 2, the basic theoretical background concerning the proposed method is introduced. A simulated study is given to verify the effectiveness of the proposed method in Section 3. In Section 4, the proposed method is applied to the gearbox compound fault features extraction to further verify its effectiveness. At last, Section 5 gives the conclusions.

## 2. Theoretical Framework

*2.1. Sparse Representation Theory.* Signal sparse representation is to represent the signal with as fewer nonzero values as possible in an overcomplete dictionary, in order to simplify the procedure of signal processing. The following part gives a concrete description on sparse representation theory.

Assume that a set  $\mathbf{A} = \{\mathbf{a}_r\}_{r \in \Gamma}$  contains  $M$  elements and  $N$  linearly independent vectors with  $N \ll M$ ; thus, the set  $\mathbf{A}$  is an overcomplete dictionary or basis. Each column of the matrix  $\mathbf{A}$  is a signal called an atom. Considering the signal sampled from the transducer contains noise, the observed signal  $\mathbf{y}(t)$  can be modeled as

$$\mathbf{y}(t) = \mathbf{x}(t) + n(t), \quad (1)$$

where  $\mathbf{y}(t)$  is the observed noisy signal,  $\mathbf{x}(t)$  is the true signal without noise, and  $n(t)$  is the noise. The true signal  $\mathbf{x}(t)$  can be represented sparsely with the overcomplete basis  $\mathbf{A}$  and also can be described as a linear combination of certain atoms of  $\mathbf{A}$ . Therefore, the representation of  $\mathbf{x}$  can be expressed as  $\mathbf{x} = \mathbf{A}\mathbf{c}$ , where  $\mathbf{c}$  is the vector of representation coefficients which also represents the transient. The occurrence of an impulse in the cyclic signal generates a value in the sparse coefficient vector  $\mathbf{c}$  accordingly. Therefore, when the cyclic impulses occur in the signal, cyclic values occur in the sparse coefficient vector  $\mathbf{c}$  correspondingly, and other values in the sparse coefficient vector will be zeros theoretically. Thus, the representation coefficient vector  $\mathbf{c}$  has sparsity, and then the cyclic transient components as well as the impulse time of the signal can be extracted from the sparse coefficients. As a result, the estimation modal in (1) turns into

$$\mathbf{y}(t) = \mathbf{A}\mathbf{c} + n(t), \quad (2)$$

where  $\mathbf{A}$  is an  $N \times M$  matrix, with  $N < M$ ;  $\mathbf{c}$  is a length- $M$  vector. The more similar the basis  $\mathbf{A}$  and the signal  $\mathbf{x}$  are, the sparser the vector  $\mathbf{c}$  will be. Based on the sparse representation theory, to obtain the sparse representation of signal  $\mathbf{x}$  under overcomplete basis  $\mathbf{A}$ , the optimization function can be constructed as

$$\begin{aligned} \min_{\mathbf{c}} \quad & \|\mathbf{c}\|_0 \\ \text{s.t.} \quad & \|\mathbf{y} - \mathbf{A}\mathbf{c}\|_2^2 \leq \varepsilon, \end{aligned} \quad (3)$$

where  $\|\mathbf{c}\|_0$  is the  $l_0$ -norm of vector  $\mathbf{c}$ , counting the nonzero values of vector  $\mathbf{c}$ . It is well known that the problem in (3) of estimating  $\mathbf{c}$  is a nondeterministic polynomial (NP)

problem, which is difficult to make out. Usually, problem (3) is transformed into

$$\begin{aligned} \min_{\mathbf{c}} \quad & \|\mathbf{c}\|_1 \\ \text{s.t.} \quad & \|\mathbf{y} - \mathbf{A}\mathbf{c}\|_2^2 \leq \varepsilon, \end{aligned} \quad (4)$$

where  $\|\mathbf{c}\|_1$  is the  $l_1$ -norm of vector  $\mathbf{c}$ , defined as  $\|\mathbf{c}\|_1 = \sum_{m=1}^M |\mathbf{c}(m)|$ .

**2.2. Optimization Algorithm for Sparse Representation.** At present, sparse algorithms including greedy pursuit algorithms and convex relaxation algorithms have been proposed to resolve the problem in (4). Greedy algorithms mainly contain Matching Pursuit (MP) algorithm [31] and Orthogonal Matching Pursuit (OMP) [32] which is developed by MP algorithm and so forth. Basis Pursuit (BP) algorithm, as one of the convex relaxation algorithms, is proposed by Chen et al. in [33]. It is different from the greedy methods by replacing the  $l_0$ -norm by  $l_1$ -norm to transform the combinatorial optimization problem into convex optimization problem. BPD algorithm, developed based on BP algorithm, having the advantage of eliminating noise inference, is introduced to solve the problem in (4):

$$\mathbf{J}(\mathbf{c}) = \arg \min_{\mathbf{c}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1, \quad (5)$$

where  $\|\mathbf{c}\|_2$  is the  $l_2$ -norm of vector  $\mathbf{c}$ , defined as  $\|\mathbf{c}\|_2^2 = \sum_{m=1}^M |\mathbf{c}(m)|^2$ , and  $\lambda$  is the regularization parameter. After the minimization of objective function in (5), a sparse representation vector  $\mathbf{c}$  can be obtained. To minimize  $\mathbf{J}(\mathbf{c})$ , an iterative algorithm must be introduced. The traditional gradient descent methods always converge slowly, such as iterative shrinkage/thresholding algorithm (ISTA) [34] and fast IST algorithm (FISTA) [35]. In order to improve the speed of convergence, Afonso et al. proposed a novel technique named the split augmented Lagrangian shrinkage algorithm (SALSA) [36], which is faster than the earlier methods. The algorithm updates the vector  $\mathbf{c}$  during each iteration so as to minimize the objective function  $\mathbf{J}(\mathbf{c})$  until the optimal solution  $\hat{\mathbf{c}}$  is gained.

Considering the unconstrained optimization problem in which the objective function is the sum of two functions, (5) can be written as

$$\min_{\mathbf{c}} \{f_1(\mathbf{c}) + f_2(\mathbf{c})\}, \quad (6)$$

where  $f_1(\mathbf{c}) = (1/2)\|\mathbf{y} - \mathbf{A}\mathbf{c}\|_2^2$ ,  $f_2(\mathbf{c}) = \lambda\|\mathbf{c}\|_1$ . Then variable splitting is introduced to create a new variable denoted by  $\mathbf{u}$ , to serve as the augment of  $f_1$ , under the constraint that  $\mathbf{u} = \mathbf{c}$ . This leads to the constrained problem:

$$\begin{aligned} \min_{\mathbf{u}, \mathbf{c}} \quad & \{f_1(\mathbf{u}) + f_2(\mathbf{c})\} \\ \text{s.t.} \quad & \mathbf{u} = \mathbf{c} \end{aligned} \quad (7)$$

which is obviously equivalent to the unconstrained problem in (6). Then, use the following definitions:  $E(\mathbf{z}) = f_1(\mathbf{u}) +$

$f_2(\mathbf{c})$ ,  $\mathbf{z} = [\mathbf{u}; \mathbf{c}]$ ,  $\mathbf{b} = \mathbf{0}$ ,  $\mathbf{H} = [\mathbf{I} \quad -\mathbf{I}]$ . With these definitions, (7) can be transformed into

$$\begin{aligned} \min_{\mathbf{z}} \quad & E(\mathbf{z}) \\ \text{s.t.} \quad & \mathbf{H}\mathbf{z} - \mathbf{b} = \mathbf{0}. \end{aligned} \quad (8)$$

The augmented Lagrangian function for this problem is defined as

$$L(\mathbf{z}, \lambda, \mu) = E(\mathbf{z}) + \lambda^T (\mathbf{H}\mathbf{z} - \mathbf{b}) + \frac{\mu}{2} \|\mathbf{H}\mathbf{z} - \mathbf{b}\|_2^2, \quad (9)$$

where  $\lambda$  is a vector of Lagrange multipliers and  $\mu \geq 0$  is the penalty parameter. The augmented Lagrangian method (ALM) is used to minimize the objective function  $L(\mathbf{z}, \lambda, \mu)$ , and the following results can be obtained:

$$\begin{aligned} \mathbf{z}^{(k+1)} &= \arg \min_{\mathbf{z}} \left\{ E(\mathbf{z}) + \frac{\mu}{2} \|\mathbf{H}\mathbf{z} - \mathbf{d}^{(k)}\|_2^2 \right\}, \\ \mathbf{d}^{(k+1)} &= \mathbf{d}^{(k)} - (\mathbf{H}\mathbf{z}^{(k+1)} - \mathbf{b}), \end{aligned} \quad (10)$$

where  $k$  is the iteration counter. Considering the concrete forms of the function  $E(\mathbf{z})$ , matrix  $\mathbf{H}$ , and the vector  $\mathbf{b}$ , novel results can be written as

$$\begin{aligned} \mathbf{u}^{(k+1)} &= \arg \min_{\mathbf{u}} \left\{ f_1(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{u} - \mathbf{c}^{(k)} - \mathbf{d}^{(k)}\|_2^2 \right\} \\ &= \arg \min_{\mathbf{u}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \frac{\mu}{2} \|\mathbf{u} - \mathbf{c}^{(k)} - \mathbf{d}^{(k)}\|_2^2 \right\}, \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbf{c}^{(k+1)} &= \arg \min_{\mathbf{c}} \left\{ f_2(\mathbf{c}) + \frac{\mu}{2} \|\mathbf{u}^{(k+1)} - \mathbf{c} - \mathbf{d}^{(k)}\|_2^2 \right\} \\ &= \arg \min_{\mathbf{c}} \left\{ \lambda \|\mathbf{c}\|_1 + \frac{\mu}{2} \|\mathbf{u}^{(k+1)} - \mathbf{c} - \mathbf{d}^{(k)}\|_2^2 \right\}, \end{aligned} \quad (12)$$

$$\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} - (\mathbf{u}^{(k+1)} - \mathbf{c}^{(k+1)}). \quad (13)$$

Equation (11) is a strictly convex quadratic function to be minimized, which leads to the solution  $\mathbf{u}^{(k+1)}$  directly, and the soft threshold facilitates the minimization of (12), after which the iteration procedure of SALSA can be listed as

$$\begin{aligned} \mathbf{u}^{(k+1)} &= (\mathbf{A}^H \mathbf{A} + \mu \mathbf{I})^{-1} (\mathbf{A}^H \mathbf{y} + \mu (\mathbf{c}^k + \mathbf{d}^k)), \\ \mathbf{c}^{(k+1)} &= \text{soft} \left( \mathbf{u}^{(k+1)} - \mathbf{d}^k, \frac{\lambda}{\mu} \right), \\ \mathbf{d}^{(k+1)} &= \mathbf{d}^{(k)} - \mathbf{u}^{(k+1)} + \mathbf{c}^{(k+1)}. \end{aligned} \quad (14)$$

By running the iterative numerical algorithm SALSA, the optimal sparse solution  $\hat{\mathbf{c}}$  can be found eventually. With the sparse solution  $\hat{\mathbf{c}}$ , which means most elements of the vector  $\hat{\mathbf{c}}$  are closer to zero, the reconstructed  $\hat{\mathbf{x}}$  can be represented as  $\hat{\mathbf{x}} = \mathbf{A}\hat{\mathbf{c}}$ . In the sparse vector  $\hat{\mathbf{c}}$ , there are successive periodic nonzero coefficients which present the transient responses in the original signal. Thus, the fault period can be calculated from the envelope spectrum analysis of the reconstructed signal, after which the fault feature is eventually extracted.

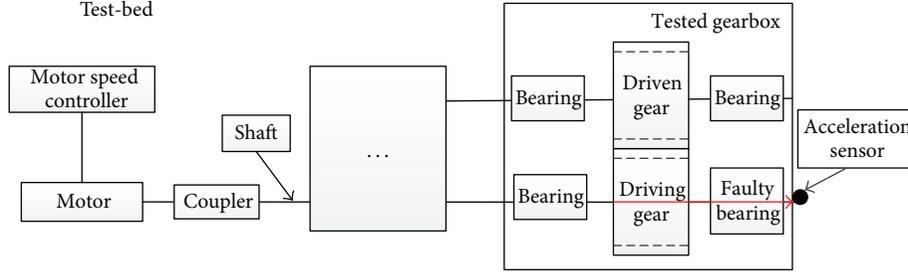


FIGURE 1: Diagram of test-bed (the red arrow illustrates the propagation path of the compound fault signal).

**2.3. Selection of the Optimal Wavelet Bases by Correlation Filtering.** After choosing SALSA as the sparse optimization algorithm, the selection of wavelet bases turns out to be another key work. In order to represent the original signal as sparse as possible, the basis should be as relevant to the signal as possible. Considering the characteristics of vibration signal of the faulty bearing, the Laplace wavelet, which is in shape similar to the signal transients caused by bearing localized fault, is selected to construct the wavelet basis during the extraction of bearing fault feature. The Laplace wavelet is defined as

$$\begin{aligned} A_1(t, \tau) &= \psi(f, \zeta, \tau, t) = \psi_\gamma(t) \\ &= B_1 e^{(-\zeta/\sqrt{1-\zeta^2})2\pi f(t-\tau)} \sin 2\pi f(t-\tau), \end{aligned} \quad (15)$$

where the parameter vector  $\gamma = (f, \zeta, \tau)$  determines the wavelet properties. These parameters  $(f, \zeta, \tau)$  denote frequency  $f \in \mathbf{R}^+$ , damping ratio  $\zeta \in [0, 1) \subset \mathbf{R}^+$ , and time index  $\tau \in \mathbf{R}$ , respectively. The parameter  $B_1$  is used to normalize the wavelet function.

The parameters  $f$ ,  $\zeta$ , and  $\tau$  belong to the subsets of  $\mathbf{F}$ ,  $\mathbf{Z}$ , and  $\mathbf{T}$ , respectively. With different parameters, the Laplace wavelet dictionary can be constructed as

$$\begin{aligned} \Psi &= \{\psi_\gamma(t) : \gamma \in \mathbf{F} \times \mathbf{Z} \times \mathbf{T}\} \\ &= \{\psi(f, \zeta, \tau, t) : f \in \mathbf{F}, \zeta \in \mathbf{Z}, \tau \in \mathbf{T}\}. \end{aligned} \quad (16)$$

With the constructed Laplace wavelet basis, correlation filtering is introduced to identify the optimal set of parameters  $(\bar{f}, \bar{\zeta}, \bar{\tau})$ . Correlation is used to determine the similarity between the wavelet basis and the original signal and measured by inner product operation. The correlation function  $c_\gamma$  is defined to calculate the correlation degree between the basis  $\psi_\gamma(t)$  and the original signal  $\mathbf{x}(t)$ :

$$c_\gamma = \cos \theta = \frac{|\langle \psi_\gamma(t), \mathbf{x}(t) \rangle|}{\|\psi_\gamma(t)\|^2 \|\mathbf{x}(t)\|^2}, \quad (17)$$

where  $\theta$  is the angle between  $\psi_\gamma(t)$  and  $\mathbf{x}(t)$ . The smaller the angle is, the more similar the basis  $\psi_\gamma(t)$  and the original signal  $\mathbf{x}(t)$  will be. Therefore, the optimal wavelet atom with optimal parameters  $(\bar{f}, \bar{\zeta}, \bar{\tau})$  can be obtained by maximizing the correlation function  $c_\gamma$  at each time value from the

constructed Laplace wavelet dictionary. Peaks of  $c_\gamma$  for a given time value  $\tau$  can be represented as

$$k_r(\tau) = \max_{f \in \mathbf{F}, \zeta \in \mathbf{Z}} c_\gamma = c(\bar{f}, \bar{\zeta}, \tau) \quad (18)$$

and the time index parameter  $\bar{\tau}$  can be calculated by maximizing the coefficient  $k_r(\tau)$ . With correlation filtering, the optimal parameters  $(\bar{f}, \bar{\zeta}, \bar{\tau})$  can be found effectively; then the optimal wavelet atom with these parameters can be constructed.

Meanwhile, the Morlet wavelet, which is in shape similar to the vibration signal transients caused by gear localized fault, is used to construct the wavelet basis during the extraction of gear fault feature. The Morlet wavelet is defined as

$$\begin{aligned} A_2(t, \tau) &= \psi(f, \zeta, \tau, t) = \psi_\gamma(t) \\ &= B_2 e^{(-\zeta/\sqrt{1-\zeta^2})[2\pi f(t-\tau)]^2} \cos 2\pi f(t-\tau), \end{aligned} \quad (19)$$

where the parameter vector  $\gamma = (f, \zeta, \tau)$  also determines the wavelet properties. These parameters  $(f, \zeta, \tau)$  denote frequency  $f \in \mathbf{R}^+$ , damping ratio  $\zeta \in [0, 1) \subset \mathbf{R}^+$ , and time index  $\tau \in \mathbf{R}$ , respectively, and the optimal set of the characteristic parameters also can be found by correlation filtering. The parameter  $B_2$  is used to normalize the wavelet function.

#### 2.4. Separation and Extraction for Gearbox Compound Fault.

However, the sequence of the compound fault features extraction should be determined first. The influence of propagation path of signals in the gearbox is taken into consideration to handle the problem. The vibration signals, generated from the gearbox, always contain not only the normal vibrations but also the fault vibrations. In theory, these signals transmit arbitrarily in the gearbox, but as a whole there is an overall propagation path of signals: gear-spline-shaft-bearing-casing [37]. When there are faults in both gear and bearing, the signal sampled by the sensor always contains different kinds of transient impulse responses. Because the sensor is placed on the casing, which is closer to the bearing according to the propagation path of the fault vibration signal as shown in Figure 1, the energy of bearing fault feature is thus higher than that of gear fault feature in the captured compound signal. As a result, the feature of bearing with higher energy is extracted

first in order to reduce the interference during the extraction of gear fault feature.

Since the choice for bases  $\mathbf{A}_1$  and  $\mathbf{A}_2$  has been made, by incorporating them into the algorithm SALSA during the extraction procedure of each fault feature, two sparse vectors  $\hat{\mathbf{c}}_1$  and  $\hat{\mathbf{c}}_2$  can be obtained one by one. That is, due to the characteristics of the vibration signal of the defective bearing, the Laplace wavelet basis, which is matched with the original compound fault signal by correlation filtering, is firstly incorporated into the iterative algorithm SALSA. Then, a sparse vector  $\hat{\mathbf{c}}_1$ , which represents the bearing fault feature sparsely, and the reconstructed signal  $\hat{\mathbf{x}}_1$  of bearing fault can be obtained. After the sparse representation of bearing fault feature, the amplitude of each transient impulse is represented by the sparse vector  $\hat{\mathbf{c}}_1$ . In order to estimate the real amplitude of the bearing fault transients, a constrained optimization strategy is proposed to estimate the amplitude of the single fault component by introducing the parameter  $k$ . The spectrum of the residual fault signal  $\mathbf{x} - k\hat{\mathbf{x}}_1$  is denoted by  $F_1(f)$

$$\begin{aligned} & \min_k \{F_1(f)\} \\ & \text{subject to } k > 0, \\ & f = f_{z1}, \end{aligned} \quad (20)$$

where  $\mathbf{x}$  is the original signal,  $\hat{\mathbf{x}}_1$  is the reconstructed signal of bearing,  $f_{z1}$  is the peak frequency of  $\hat{\mathbf{x}}_1$ , and  $k$  is a positive parameter. When  $F_1(f)$  is minimized eventually subjected to its constraints, it indicates that the bearing fault component in the residual fault signal has been removed to the largest extent. By solving problem in (20), an optimal value  $k_{\text{opt1}}$  is acquired, and the estimated bearing fault signal can be obtained by the function  $\mathbf{x}_1 = k_{\text{opt1}}\hat{\mathbf{x}}_1$ .

Then, after removing the bearing fault signal, the residual signal should only contain gear fault transient responses and underground noise. To extract the fault feature of gear, Morlet wavelet basis, matched with the obtained residual signal by correlation filtering, is incorporated into the iterative algorithm SALSA. After that, the second sparse vector  $\hat{\mathbf{c}}_2$  representing the gear transients is generated. Similarly, the reconstructed signal  $\hat{\mathbf{x}}_2$  of gear can be gained by  $\hat{\mathbf{x}}_2 = \mathbf{A}_2\hat{\mathbf{c}}_2$ . Similar to the bearing fault feature extraction, it is also necessary to estimate the real gear fault signal by solving another constrained optimization problem:

$$\begin{aligned} & \min_k \{F_2(f)\} \\ & \text{subject to } k > 0, \\ & f = f_{z2}, \end{aligned} \quad (21)$$

where the spectrum of the residual fault signal  $\mathbf{x}_{\text{res}} - k\hat{\mathbf{x}}_2$  is denoted by  $F_2(f)$ ,  $\mathbf{x}_{\text{res}}$  is the residual signal after removing the bearing fault component, and  $\hat{\mathbf{x}}_2$  is the preliminary reconstructed fault signal of gear. After solving problem in (21), another optimal value  $k_{\text{opt2}}$  is acquired naturally. The estimated gear fault signal thus can be obtained by the function  $\mathbf{x}_2 = k_{\text{opt2}}\hat{\mathbf{x}}_2$ .

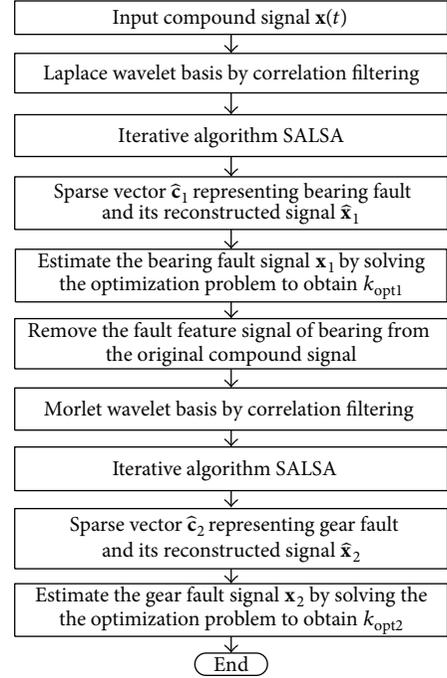


FIGURE 2: Procedure of the proposed compound fault transients extraction method.

To summarize, the procedure of the proposed method in this paper to separate and extract gearbox compound fault features using wavelet bases is presented in Figure 2.

### 3. Simulated Study

To verify the effectiveness of the proposed method, a simulated compound fault signal processing is performed for different features extraction. Considering the characteristics of the compound fault vibration signal in a gearbox, the signal is constructed as

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{x}_1(t) + \mathbf{x}_2(t) + A_n n(t), \\ \mathbf{x}_1(t) &= \sum_k \left[ B_1 e^{(-\zeta_1/\sqrt{1-\zeta_1^2})2\pi f_1(t-\tau_1-kT_1)} \right. \\ & \quad \left. \times \sin 2\pi f_1(t-\tau_1-kT_1) \right], \\ \mathbf{x}_2(t) &= \sum_k \left[ B_2 e^{(-\zeta_2/\sqrt{1-\zeta_2^2})[2\pi f_2(t-\tau_2-kT_2)]^2} \right. \\ & \quad \left. \times \cos 2\pi f_2(t-\tau_2-kT_2) \right], \end{aligned} \quad (22)$$

where  $\mathbf{x}_1(t)$  is a period cyclic impulse responses signal to simulate the bearing fault in Figure 3(a). The values of parameters are given. The frequency is  $f_1 = 3500$  Hz, the damping ratio is  $\zeta_1 = 0.080$ , the time index is  $\tau_1 = 0.1$  s, the cyclic period is  $T_1 = 0.007$  s, and the normalized parameter is  $B_1 = 1$ .  $\mathbf{x}_2(t)$  is also a period cyclic signal to simulate the vibration signal of faulty gear in Figure 3(b). Concrete

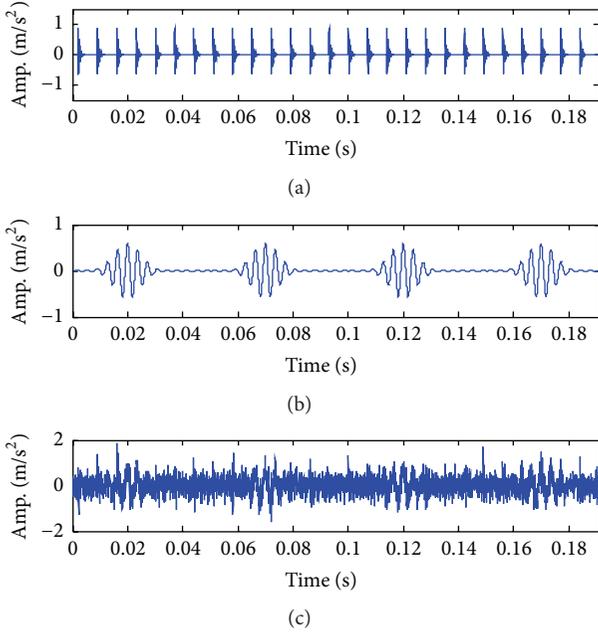


FIGURE 3: Simulated signals: (a) the simulated signal for faulty bearing, (b) the simulated signal for faulty gear, and (c) noisy compound fault signal.

values of the parameters are listed:  $f_2 = 275$  Hz,  $\zeta_2 = 0.0074$ ,  $\tau_2 = 0.02$  s,  $T_2 = 0.05$  s, and  $B_2 = 0.6$ . The signal  $n(t)$  is white Gaussian noise, which is weighted by  $A_n = 0.3$ . The sampling frequency is 25.6 KHz and the sampling number is 5000. Figure 3(c) gives the waveforms of the noisy compound signal.

To separate and extract each fault feature from the noisy compound signal, the proposed sparse representation under wavelet bases is applied. According to the procedure in Figure 2, the first step is to obtain the optimal Laplace wavelet basis matched with the original noisy signal by correlation filtering, which is shown in Figure 4(a). Then incorporate the matched Laplace wavelet basis into the iterative processes of the algorithm SALSA, after which the sparse coefficients representing the transient feature of the faulty bearing can be obtained, as shown in Figure 4(b). The reconstructed signal of bearing is shown in Figure 4(c), and Figure 4(d) gives the envelope spectrum analysis of reconstructed signal. In Figure 4(b), there are successive periodic nonzero values in the sparse vector, which represent the bearing fault transient in the original signal. In Figure 4(d), the fault characteristic frequency of bearing is obtained as 141.1 Hz, almost consistent with the theoretical value ( $f_0 = 1/T_1 = 142.9$  Hz). Figure 4(e) shows the estimated bearing fault signal with  $k_{\text{opt1}} = 1.716$ .

Remove the estimated bearing fault signal from the original noisy compound signal; we can get a residual signal as shown in Figure 5. Similar to the bearing, the first step is to obtain the optimal Morlet wavelet basis matched with the residual signal by correlation filtering, which is illustrated in Figure 6(a). Then, incorporate the Morlet wavelet basis into the iterative processes of SALSA, sparse coefficients representing the gear fault feature can be obtained in Figure 6(b),

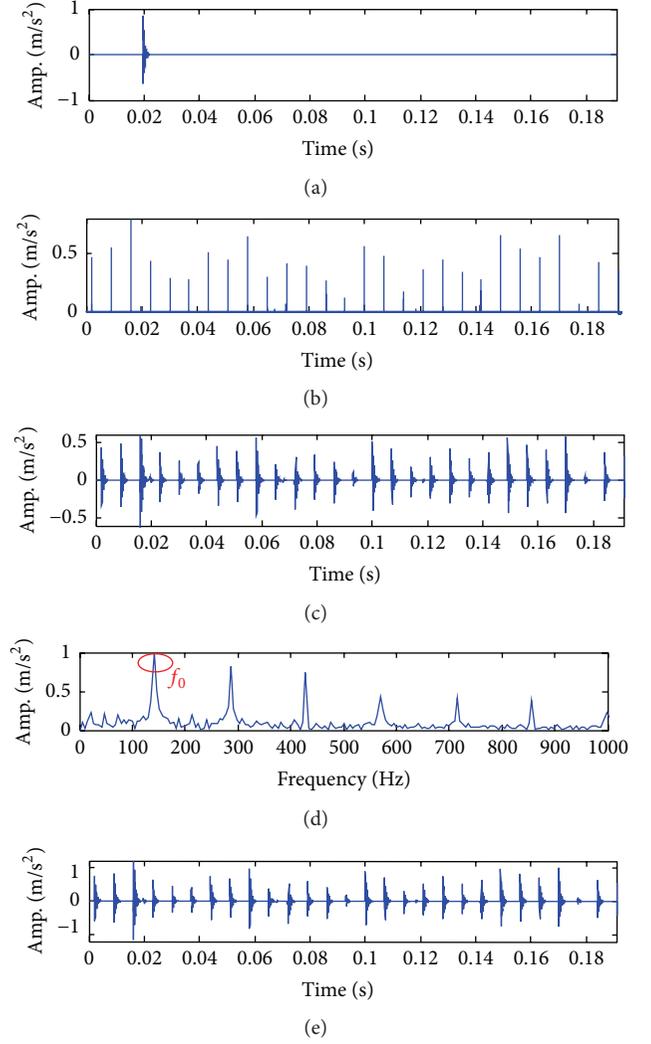


FIGURE 4: Results of bearing fault signal: (a) optimal Laplace basis, (b) sparse coefficients, (c) reconstructed signal, (d) the envelope spectrum analysis of reconstructed signal, and (e) the estimated bearing fault signal.

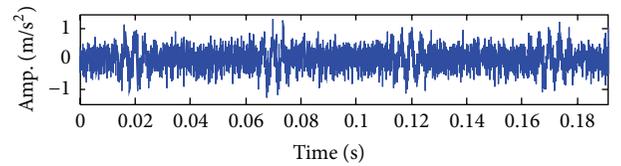


FIGURE 5: The residual signal after removing the estimated bearing fault signal.

and Figure 6(c) gives the reconstructed signal of defective gear. The envelope spectrum analysis of the reconstructed signal is given in Figure 6(d), where we can get the fault characteristic frequency as 20.9 Hz, nearly equal to the simulated value ( $f_0 = 1/T_2 = 20$  Hz). At last, Figure 6(e) shows the estimated gear fault signal with  $k_{\text{opt2}} = 11.563$ .

By analyzing the results of the simulated study, we can conclude that there are both bearing and gear faults in the

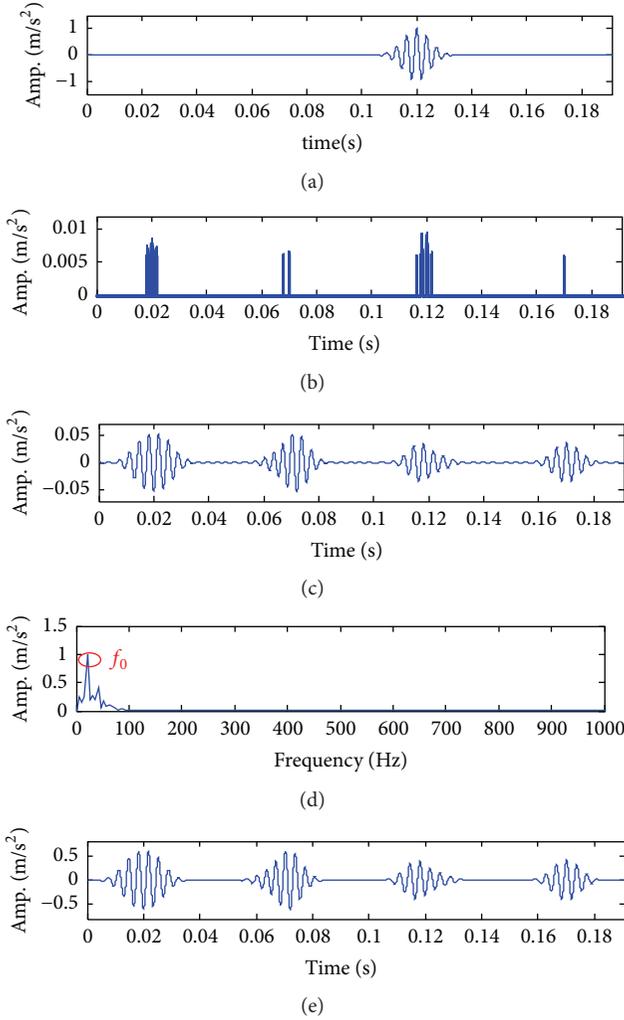


FIGURE 6: Results of gear fault signal: (a) optimal Morlet basis, (b) sparse coefficients, (c) reconstructed signal, (d) the envelope spectrum analysis of reconstructed signal, and (e) the estimated gear fault signal.

simulated compound fault signal. The effectiveness of the proposed method has been proven preliminarily.

Taking the noise inference into consideration, the noise intensity  $A_n$  is increased gradually in order to analyze the robustness of the proposed method. Figures 7–10 show the extraction results when  $A_n$  is selected as 0.4, 0.5, and 0.6, respectively. As shown in these figures, bearing and gear faults can still be separated and extracted from the compound signal accurately. Based on these above analyses, it can be concluded that the proposed method has the capability to suppress the noise inference until the noise intensity is increased to a higher value.

#### 4. Application to Gearbox Compound Fault Features Extraction

To further verify the effectiveness of the proposed method in practical engineering application, defective gearbox data

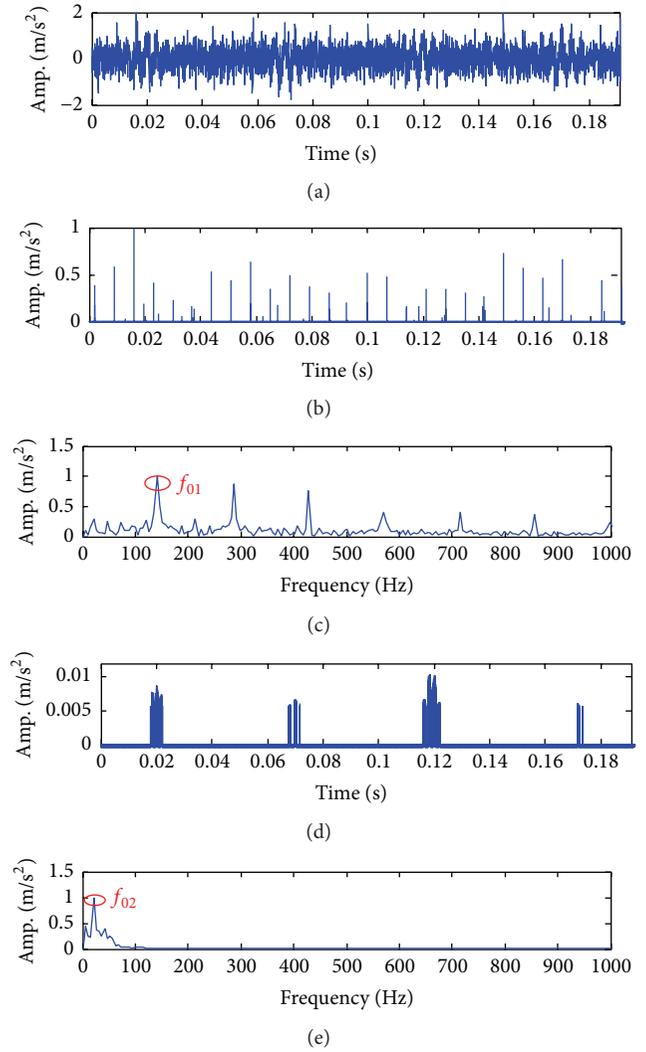


FIGURE 7: Extraction results of the compound signal with  $A_n = 0.4$ : (a) the noisy signal, (b) sparse coefficients of bearing fault component, (c) the envelope spectrum analysis of reconstructed bearing fault signal, (d) sparse coefficients of gear fault component, and (e) the envelope spectrum analysis of reconstructed gear fault signal.

is analyzed. The research object is a single stage transmission gearbox in a test-bed, as illustrated in Figure 10. The faulty gear is a helical gear, whose working parameters are listed in Table 1. The bearing model in the experiment is 30625, taper roller bearing, and its geometric parameters are listed in Table 2. Based on the known parameters, the theoretical fault frequency of the bearing outer race can be calculated as 176.18 Hz.

Aimed at getting gearbox compound fault data, a crack width of 0.4 mm is set in the outer race of the bearing using the linear cutting technique to simulate the localized fault of a bearing, and half a tooth is cut in the driving gear by electric sparkle technique to simulate the localized deflection of a gear as shown in Figure 11. Additionally, to reduce the influence of propagation path, the sensor is placed on the bearing end

TABLE 1: Working parameters of gears in the tested gearbox.

Gear	Number of teeth	Rotating frequency (Hz)	Rotating period (s)	Meshing frequency (Hz)
Driving gear	34	24.67	0.041	839
Driven gear	42	19.98	0.05	

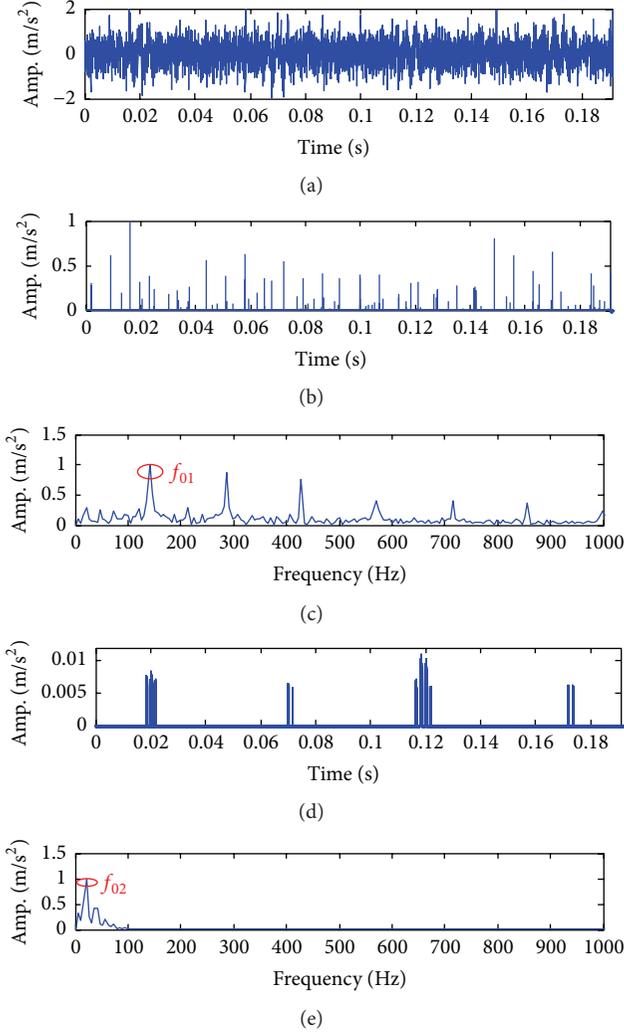


FIGURE 8: Extraction results of the compound signal with  $A_n = 0.5$ : (a) the noisy signal, (b) sparse coefficients of bearing fault component, (c) the envelope spectrum analysis of reconstructed bearing fault signal, (d) sparse coefficients of gear fault component, and (e) the envelope spectrum analysis of reconstructed gear fault signal.

TABLE 2: Geometry of the tested bearing.

Inside diameter (mm)	Outside diameter (mm)	Ball diameter (mm)	Number of rolling elements	Contact angle (°)
30	62	8	17	14

cover close to the faulty bearing. During the experiment, the rotating speed of the motor is 1496 r/min, and the sampling frequency is set at 25.6 KHz.

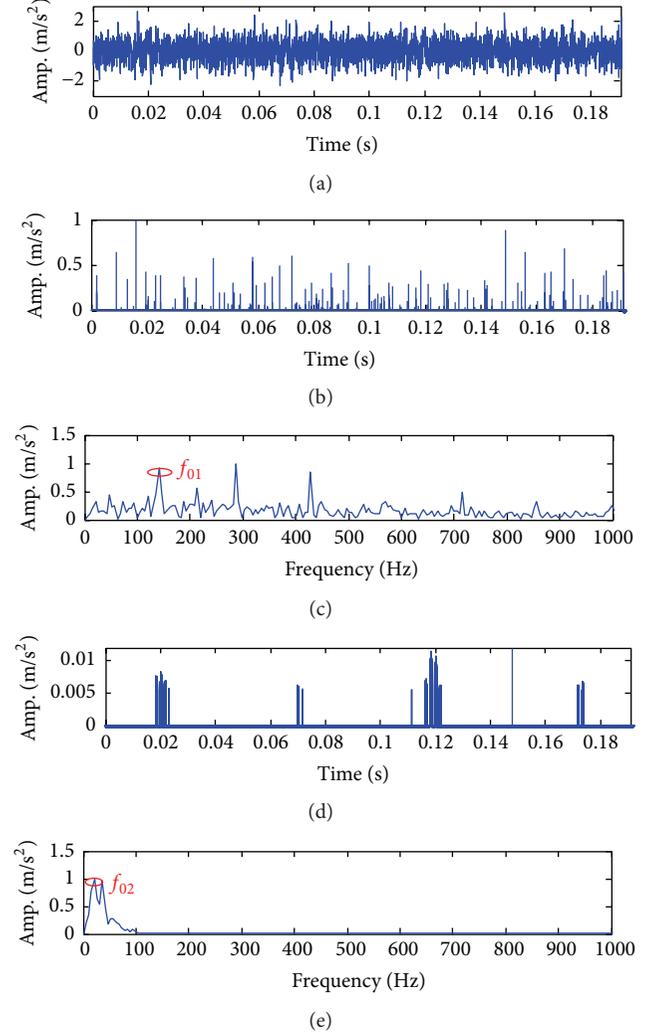


FIGURE 9: Extraction results of the compound signal with  $A_n = 0.5$ : (a) the noisy signal, (b) sparse coefficients of bearing fault component, (c) the envelope spectrum analysis of reconstructed bearing fault signal, (d) sparse coefficients of gear fault component, and (e) the envelope spectrum analysis of reconstructed gear fault signal.

The measured vibration signal with compound fault is shown in Figure 12(a). From Figure 12(a), the characteristics of each fault cannot be identified clearly. Therefore, the proposed method in this paper is employed to extract different transient features from the noisy signal. Figure 12(b) gives its frequency spectrum. Figure 12(c) is the envelope spectrum analysis of the original signal. In Figure 12(c), there exist different frequency components. Thus, the location of the fault cannot be identified exactly in the gearbox.

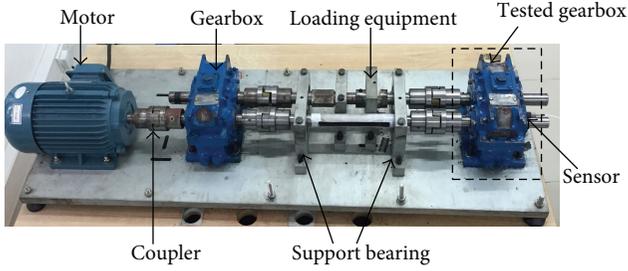


FIGURE 10: Experimental gearbox in a test-bed.

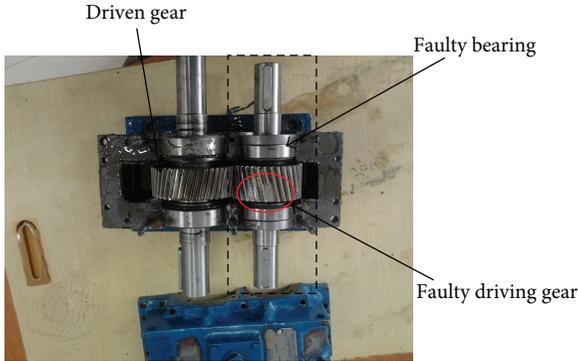


FIGURE 11: Fault components.

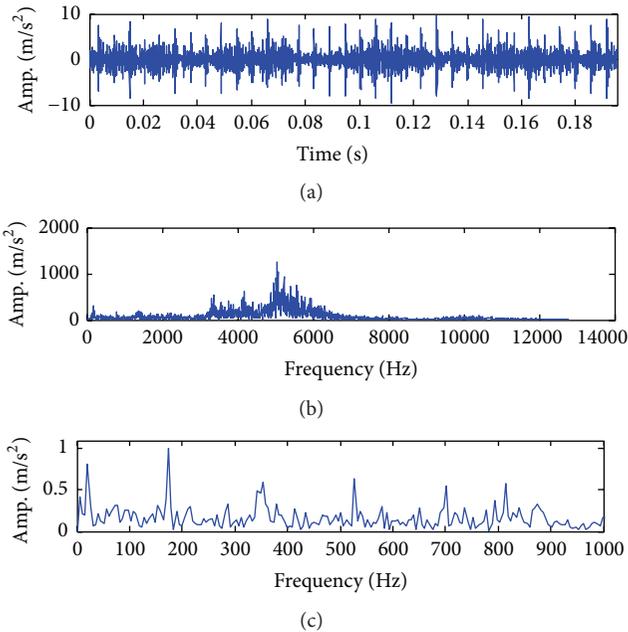


FIGURE 12: The measured signal and its spectral analysis: (a) the measured signal with compound fault, (b) its frequency spectrum, and (c) its envelope spectrum analysis.

According to the procedure of the proposed method in this paper, the first step is to obtain the optimal Laplace wavelet basis matched with the original measured signal by correlation filtering. Figure 13(a) gives the result of matched Laplace wavelet basis. Then, the matched Laplace wavelet basis is incorporated into the iterative algorithm SALSA.

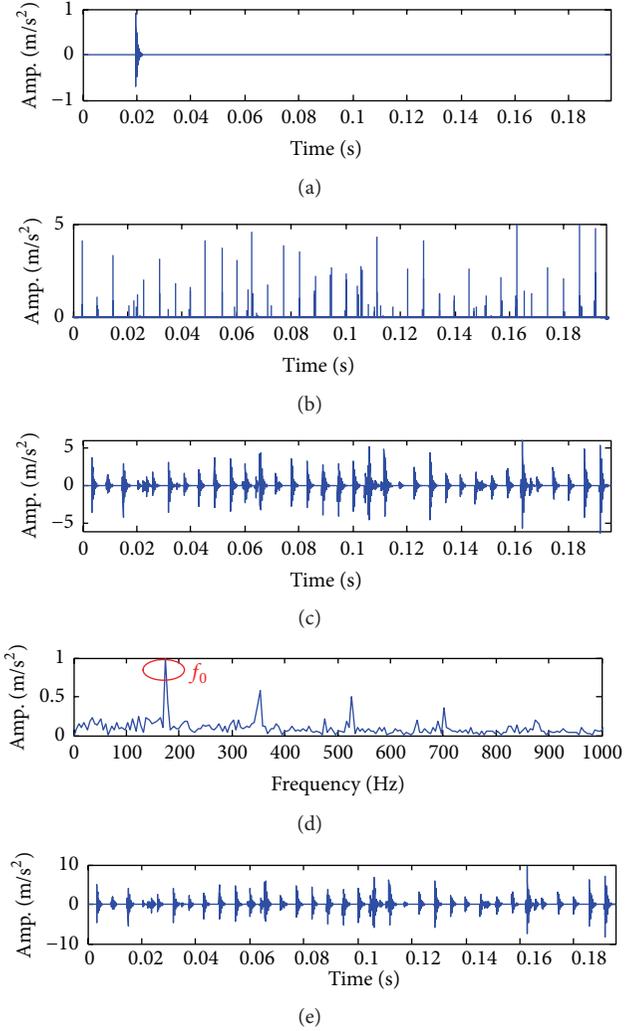


FIGURE 13: Results of bearing fault signal: (a) optimal Laplace basis, (b) sparse coefficients, (c) reconstructed signal, (d) the envelope spectrum analysis of the reconstructed signal, and (e) the estimated bearing fault signal.

Figure 13(b) shows the sparse coefficients of bearing. In Figure 13(b), there are successive nonzeros in the sparse vector, which represent the transient of bearing outer race fault. Figure 13(c) is the reconstructed signal of faulty bearing. Figure 13(d) gives the envelope spectrum analysis of the reconstructed signal, where we can conclude that the feature frequency of the bearing outer race is 174.1 Hz, and it is consistent with the theoretical value 176.18 Hz. Hence, it indicates that there exists a localized fault in the outer race of bearing indeed. Finally, Figure 13(e) shows the estimated bearing fault signal with  $k_{opt1} = 1.322$ .

After removing the estimated bearing fault signal from the original signal, we can obtain the residual signal in Figure 14. Then the next steps of the proposed method are conducted and the extraction results of the defective gear fault feature are shown in Figure 15. Figure 15(a) gives the optimal matched Morlet wavelet basis. Figure 15(b) gives the sparse coefficients. In Figure 15(b), there are successive nonzeros in

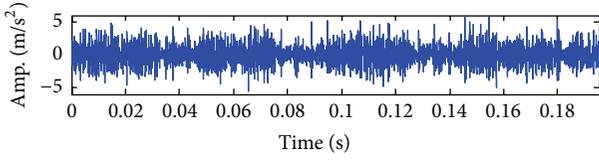


FIGURE 14: The residual signal after removing the bearing fault signal.

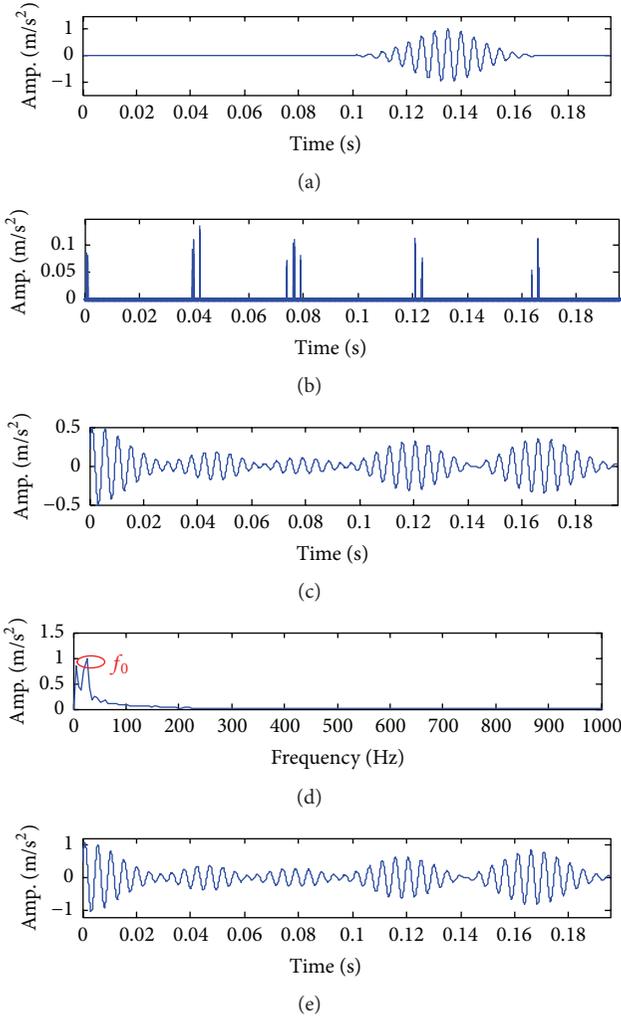


FIGURE 15: Results of gear fault signal: (a) optimal Morlet basis, (b) sparse coefficients, (c) reconstructed signal, (d) the envelope spectrum analysis of the reconstructed signal, and (e) the estimated gear fault signal.

the sparse vector, which represent the transients of gear fault. Figure 15(c) shows the reconstructed signal of the defective gear. Figure 15(d) is the envelope spectrum analysis of the reconstructed signal, from which the feature frequency can be obtained as 25.6 Hz, nearly consistent with the theoretical value 24.67 Hz. Therefore, it indicates that there is a localized fault in the gear indeed. Finally, Figure 15(e) shows the estimated gear fault signal with  $k_{\text{opt}2} = 2.193$ .

## 5. Conclusions

This paper proposes a novel method to represent the gearbox compound fault features sparsely using different wavelet bases so as to separate the different faulty components from the compound signal. Based on the sparse representation theory, the proposed method introduces the numerical iterative algorithm SALSA under Laplace wavelet basis and Morlet wavelet basis, respectively, to solve the BPD problem, after which two sparse vectors can be obtained one by one. One vector represents the transient feature of faulty bearing and the other represents the transient feature of defective gear. As a result, the proposed method converts the gearbox compound fault features into a series of sparse coefficients, which facilitates gearbox fault diagnosis. Both the simulated study and the application to the sampled gearbox compound fault data verify that the proposed method can separate and extract the compound fault features of the gearbox effectively.

## Conflict of Interests

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

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 51375322 and no. 51405321), the National Natural Science Foundation of Jiangsu Province (no. BK20140339), and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (no. 4KJB460023). The authors also would like to appreciate three anonymous reviewers for their constructive comments and suggestions for improving this paper.

## References

- [1] X. Wei and H. Pan, "Review of the gearbox fault diagnosis technology," *Journal of Test and Measurement Technology*, vol. 20, no. 4, pp. 368–376, 2006.
- [2] D. Wang, P. W. Tse, and Y. L. Tse, "A morphogram with the optimal selection of parameters used in morphological analysis for enhancing the ability in bearing fault diagnosis," *Measurement Science and Technology*, vol. 23, no. 6, Article ID 065001, 2012.
- [3] S. Wang, X. Chen, G. Cai, B. Chen, X. Li, and Z. He, "Matching demodulation transform and synchrosqueezing in time-frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 69–84, 2014.
- [4] L. Xiang, G. Tang, and A. Hu, "Vibration signal's time-frequency analysis and comparison for a rotating machinery," *Journal of Vibration and Shock*, vol. 29, no. 2, pp. 42–45, 2010.
- [5] S. Rajagopalan, J. A. Restrepo, J. M. Aller, T. G. Habetler, and R. G. Harley, "Wigner-Ville distributions for detection of rotor faults in brushless DC (BLDC) motors operating under non-stationary conditions," in *Proceedings of the 5th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDMPED '05)*, pp. 1–7, IEEE, Vienna, Austria, September 2005.

- [6] M. Amirmazlaghani and H. Amindavar, "Statistical modeling and denoising Wigner-Ville distribution," *Digital Signal Processing*, vol. 23, no. 2, pp. 506–513, 2013.
- [7] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.
- [8] C. Shen, D. Wang, F. Kong, and P. W. Tse, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier," *Measurement*, vol. 46, no. 4, pp. 1551–1564, 2013.
- [9] D. Wang, C. Shen, and P. W. Tse, "A novel adaptive wavelet stripping algorithm for extracting the transients caused by bearing localized faults," *Journal of Sound and Vibration*, vol. 332, no. 25, pp. 6871–6890, 2013.
- [10] W. J. Wang and P. D. McFadden, "Application of wavelets to gear box vibration signals for fault detection," *Journal of Sound and Vibration*, vol. 192, no. 5, pp. 927–939, 1996.
- [11] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [12] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 108–126, 2013.
- [13] L. Zhao, W. Yu, and R. Yan, "Rolling bearing fault diagnosis based on CEEMD and time series modeling," *Mathematical Problems in Engineering*, vol. 2014, Article ID 101867, 13 pages, 2014.
- [14] A. Ming, F. Chu, and W. Zhang, "Compound fault features separation of rolling element bearing based on the wavelet decomposition and spectrum auto-correlation," *Journal of Mechanical Engineering*, vol. 49, no. 3, pp. 80–87, 2013.
- [15] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–862, 2001.
- [16] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, and L. Li, "Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis," *Measurement*, vol. 46, no. 1, pp. 259–271, 2013.
- [17] P. W. Tse, J. Y. Zhang, and X. J. Wang, "Blind source separation and blind equalization algorithms for mechanical signal separation and identification," *Journal of Vibration and Control*, vol. 12, no. 4, pp. 395–423, 2006.
- [18] J. L. Starck, Y. Moudden, J. Bobin, M. Elad, and D. L. Donoho, "Morphological component analysis," in *Wavelets XI*, vol. 5914 of *Proceedings of SPIE*, pp. 1–15, San Diego, Calif, USA, July 2005.
- [19] J. Yang, H.-Q. Zheng, Z.-Z. Guan, and Y.-G. Wang, "Compound fault diagnosis for gearbox based on kernel morphological component analysis," *Journal of Vibration and Shock*, vol. 31, no. 10, pp. 97–101, 2012.
- [20] H. Li, H.-Q. Zheng, and L.-W. Tang, "Bearing multi-fault diagnosis based on improved morphological component analysis," *Journal of Vibration and Shock*, vol. 31, no. 12, pp. 135–140, 2012.
- [21] V. Purushotham, S. Narayanan, and S. A. N. Prasad, "Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition," *NDT & E International*, vol. 38, no. 8, pp. 654–664, 2005.
- [22] D. Wang, K. Tsui, P. W. Tse, and M. J. Zuo, "Principal components of superhigh-dimensional statistical features and support vector machine for improving identification accuracies of different gear crack levels under different working conditions," *Shock and Vibration*, vol. 2015, Article ID 420168, 14 pages, 2015.
- [23] I. Selesnick, "Introduction to sparsity in signal processing," *Connexions*, 2012, <http://cnx.org/contents/79bacf0d-b81d-4253-987a-f017996e4858@3/Introduction-to-Sparsity-in-Si>.
- [24] K. Huang and S. Aviyente, "Sparse representation for signal classification," *Advances in Neural Information Processing Systems*, vol. 19, pp. 609–616, 2006.
- [25] T. Poggio and F. Girosi, "A sparse representation for function approximation," *Neural Computation*, vol. 10, no. 6, pp. 1445–1454, 1998.
- [26] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computation*, vol. 15, no. 2, pp. 349–396, 2003.
- [27] G. Cai, X. Chen, and Z. He, "Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox," *Mechanical Systems and Signal Processing*, vol. 41, no. 1–2, pp. 34–53, 2013.
- [28] W. Fan, G. Cai, W. Huang, L. Shang, and Z. Zhu, "Sparse representation of transients based on wavelet basis and majorization-minimization algorithm for machinery fault diagnosis," *Mathematical Problems in Engineering*, vol. 2014, Article ID 696051, 11 pages, 2014.
- [29] S. Wang, J. Xu, and Z. Zhu, "LSM-based transient parameter identification and its application in feature extraction of bearing fault," *Journal of Mechanical Engineering*, vol. 48, no. 7, pp. 68–76, 2012.
- [30] S.-B. Wang, Z.-K. Zhu, and A.-Z. Wang, "Gearbox fault feature detection based on adaptive parameter identification with Morlet wavelet," in *Proceedings of the 8th International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR '10)*, pp. 409–414, IEEE, Qingdao, China, July 2010.
- [31] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [32] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Asilomar Conference on Signals, Systems & Computers*, pp. 40–44, November 1993.
- [33] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [34] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [35] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [36] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.
- [37] W. Li, *Vibration source detection system of gearbox based on vibration signal propagation path [Ph.D. thesis]*, Hangzhou Dianzi University, 2013.

## Research Article

# Cyclostationary Analysis for Gearbox and Bearing Fault Diagnosis

Zhipeng Feng<sup>1</sup> and Fulei Chu<sup>2</sup>

<sup>1</sup>*School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China*

<sup>2</sup>*Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Zhipeng Feng; [fengzp@ustb.edu.cn](mailto:fengzp@ustb.edu.cn)

Received 5 May 2015; Revised 24 July 2015; Accepted 27 July 2015

Academic Editor: Dong Wang

Copyright © 2015 Z. Feng and F. Chu. 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.

Gearbox and rolling element bearing vibration signals feature modulation, thus being cyclostationary. Therefore, the cyclic correlation and cyclic spectrum are suited to analyze their modulation characteristics and thereby extract gearbox and bearing fault symptoms. In order to thoroughly understand the cyclostationarity of gearbox and bearing vibrations, the explicit expressions of cyclic correlation and cyclic spectrum for amplitude modulation and frequency modulation (AM-FM) signals are derived, and their properties are summarized. The theoretical derivations are illustrated and validated by gearbox and bearing experimental signal analyses. The modulation characteristics caused by gearbox and bearing faults are extracted. In faulty gearbox and bearing cases, more peaks appear in cyclic correlation slice of 0 lag and cyclic spectrum, than in healthy cases. The gear and bearing faults are detected by checking the presence or monitoring the magnitude change of peaks in cyclic correlation and cyclic spectrum and are located according to the peak cyclic frequency locations or sideband frequency spacing.

## 1. Introduction

Gearboxes and rolling element bearings are critical mechanical components and widely used in many types of machinery [1–3]. Gear and bearing faults will result in deficiency of transmission or even shut-down of the entire machinery. Therefore, gearbox and bearing fault diagnosis play an important role.

The vibration signals of gearboxes and rolling element bearings are usually cyclostationary, since their statistics (in terms of ensemble average) change periodically with time due to their rotation. Hence, cyclostationary analysis is suitable to process gearbox and bearing vibration signals. Dalpiaz et al. [4] made a comparison study between various vibration signal analysis methods (including cepstrum, time-synchronous average, wavelet transform, and cyclostationary analysis) for gear localized fault detection, and they found spectral correlation density is effective in monitoring gear crack development. Sidahmed et al. [5, 6] showed that time-synchronous

average can be considered as a first order cyclostationarity and spectral correlation as a second order cyclostationarity, found that gear vibration signals have second order cyclostationarity, and early detected gear tooth spalling using spectral correlation analysis. Zhu et al. [7] investigated the effectiveness of cyclostationarity from the first order to the third order, that is, spectrum of time-synchronous average, cyclic spectrum and cyclic bispectrum, in gearbox condition monitoring. Bi et al. [8] proposed to extract the amplitude modulation and phase modulation information from gear vibration signals using slice spectral correlation density, so as to detect gear defects. Li and Qu [9] deduced the cyclic correlation and cyclic spectrum of amplitude modulation signals and applied them to rolling element bearing fault diagnosis. Recently, Antoni et al. [10–15] conducted a series of researches on cyclostationary signal analysis and applied it to fault diagnosis of rotating machinery. To reduce the computational complexity of cyclic energy indicator based on cyclic spectral density, Wang and Shen [16] proposed

an equivalent cyclic energy indicator for rolling element bearing degradation evaluation. These researches illustrate the effectiveness of cyclostationary analysis in gearbox and bearing fault diagnosis. However, the explicit relationship between the cyclostationary features of vibration signals and the gearbox and bearing dynamic nature still needs further investigation, in order for thoroughly understanding the vibration signal characteristics and thereby effectively diagnosing fault.

Gearbox and rolling element bearing vibration signals usually feature amplitude modulation and frequency modulation (AM-FM), and the modulation characteristics contain their health status information [17–21]. Cyclic correlation and cyclic spectrum are effective in extracting modulation features from amplitude modulation (AM), frequency modulation (FM), and AM-FM signals. Feng and his collaborators [22–24] derived the expressions of cyclic correlation and cyclic spectrum for gear AM-FM vibration signals and proposed indicators based on cyclic correlation and cyclic spectrum for detection and assessment of gearbox fault. Nevertheless, the carrier frequency of rolling element bearing vibration signals (resonance frequency) is completely different from that of gear vibration signals (gear meshing frequency and its harmonics). Therefore, it is important to investigate the cyclic correlation and cyclic spectrum of bearing vibration signals in depth, considering both the AM and the FM effects due to bearing fault. Meanwhile, how to explain the cyclostationary features displayed by the cyclic correlation and cyclic spectrum and to map the modulation characteristics to gear and bearing fault are still important issues for application of cyclostationary analysis in gearbox and bearing fault diagnosis. In this paper, we derive the explicit expressions of cyclic correlation and cyclic spectrum for general AM-FM signals, summarize their properties, and further extend the theoretical derivations to modulation analysis of both gear and rolling element bearing vibration signals, thus enabling cyclostationary analysis to detect and locate both gearbox and bearing fault.

## 2. Cyclic Correlation

**2.1. Definition.** The statistics of cyclostationary signals have periodicity or multiperiodicity with respect to time evolution. Cyclic statistics are suitable to process such signals. Among those, second order cyclic statistics, that is, cyclic correlation and cyclic spectrum, are effective in extracting the modulation features of cyclostationary signals.

For a signal  $x(t)$ , the cyclic autocorrelation function is defined as [25]

$$R_x^\alpha(\tau) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right)^* x\left(t - \frac{\tau}{2}\right) \exp(-j2\pi\alpha t) dt, \quad (1)$$

where  $\tau$  is time lag and  $\alpha$  is cyclic frequency.

**2.2. Cyclic Correlation of AM-FM Signal.** During the constant speed running of gearboxes and rolling element bearings, the existence of fault, machining defect, and assembling error often leads to periodical changes in vibration signals. For gearboxes, such periodical changes modulate both the amplitude envelope and instantaneous frequency of gear meshing vibration. For bearings, such repeated changes excite resonance periodically. The excited resonance vanishes rapidly due to damping before next resonance comes, resulting in AM feature. Meanwhile, in one repeating cycle, the resonance exists in early period and the instantaneous frequency equals approximately the resonance frequency, while in later period, the resonance vanishes due to damping and the instantaneous frequency becomes 0. That means the instantaneous frequency changes periodically, resulting in frequency modulation (FM). Hence the vibration signals of both gearboxes and rolling element bearings can be modeled as an AM-FM process [17, 21] plus a random noise

$$x(t) = \sum_{k=0}^K a_k(t) \cos[2\pi f_c t + b_k(t) + \theta_k] + \varepsilon(t), \quad (2)$$

where

$$a_k(t) = \sum_{n=0}^N (1 + A_{kn}) \cos(2\pi n f_m t + \phi_{kn}), \quad (3)$$

$$b_k(t) = \sum_{l=0}^L B_{kl} \sin(2\pi l f_m t + \varphi_{kl}) \quad (4)$$

are the AM and FM functions, respectively,  $A > 0$  and  $B > 0$  are the magnitude of AM and FM respectively,  $f_c$  is the carrier frequency (for gearboxes, it is the gear meshing frequency or its  $k$ th harmonics; for rolling element bearings, it is the resonance frequency of bearing system),  $f_m$  is the modulating frequency equal to the gear or bearing fault characteristic frequency,  $\theta$ ,  $\phi$ , and  $\varphi$  are the initial phase of AM and FM, respectively, and  $\varepsilon(t)$  is a white Gaussian noise due to random background interferences.

Without loss of generality, consider only the fundamental frequency of the AM and FM terms; then (2) becomes

$$x(t) = [1 + A \cos(2\pi f_m t + \phi)] \cdot \cos[2\pi f_c t + B \sin(2\pi f_m t + \varphi) + \theta] + \varepsilon(t). \quad (5)$$

According to the identity [26]

$$\exp[jz \sin(\beta)] = \sum_{k=-\infty}^{\infty} J_k(z) \exp(jk\beta), \quad (6)$$

where  $J_k(z)$  is Bessel function of the first kind with integer order  $k$  and argument  $z$ , the FM term in (5) can be expanded as a Bessel series, and then (5) becomes

$$x(t) = [1 + A \cos(2\pi f_m t + \phi)] \cdot \sum_{n=-\infty}^{\infty} J_n(B) \cos(2\pi f_c t + 2\pi n f_m t + n\varphi + \theta) + \varepsilon(t). \quad (7)$$

For such a signal expressed as (7), the time-varying feature of its autocorrelation function is mainly determined by the AM-FM part, since the autocorrelation function of a white Gaussian noise  $R_\varepsilon(\tau) = \delta(\tau)$  (i.e., its Fourier transform has peak at 0 only, and therefore the noise  $\varepsilon(t)$  does not affect identification of modulating frequency via cyclic correlation and cyclic spectrum analysis). In addition, it is the AM and

FM effects on the carrier signal that leads to the cyclostationarity of bearing and gearbox vibration signals, and we rely on detection of the modulating frequency of such AM and FM effects to diagnose bearing and gearbox fault. Therefore, we neglect the noise  $\varepsilon(t)$  and focus on the deterministic AM-FM part only in the following analysis. Then the cyclic autocorrelation function of (7) can be derived as [22]

$$R_x^\alpha(\tau) = \begin{cases} \frac{1}{2} \left[ h_L(0, 0) + \frac{1}{4} A^2 h_L(-2f_m, 0) + \frac{1}{4} A^2 h_L(2f_m, 0) \right], & \alpha = \pm(n - n') f_m, \\ \frac{1}{4} A [h_L(\mp f_m, \pm\phi) + h_L(\pm f_m, \pm\phi)], & \alpha = \pm(n - n' \pm 1) f_m, \\ \frac{1}{8} A^2 h_L(0, \pm 2\phi), & \alpha = \pm(n - n' \pm 2) f_m \end{cases} \quad (8a)$$

in lower cyclic frequency domain, and

$$R_x^\alpha(\tau) = \begin{cases} \frac{1}{2} \left[ h_H(0, 0) + \frac{1}{4} A^2 h_H(-2f_m, 0) + \frac{1}{4} A^2 h_H(2f_m, 0) \right], & \alpha = \pm[2f_c + (n + n') f_m], \\ \frac{1}{4} A [h_H(\mp f_m, \pm\phi) + h_H(\pm f_m, \pm\phi)], & \alpha = \pm[2f_c + (n + n' \pm 1) f_m], \\ \frac{1}{8} A^2 h_H(0, \pm 2\phi), & \alpha = \pm[2f_c + (n + n' \pm 2) f_m] \end{cases} \quad (8b)$$

in higher cyclic frequency domain, where the intermediate functions

$$h_L(\eta, \psi) = \sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} J_n(B) J_{n'}(B) \quad (9a)$$

$$\cdot \exp(\pm j \{ \pi [2f_c + (n + n') f_m + \eta] \tau + (n - n') \varphi + \psi \}),$$

$$h_H(\eta, \psi) = \sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} J_n(B) J_{n'}(B) \quad (9b)$$

$$\cdot \exp(\pm j \{ \pi [(n - n') f_m + \eta] \tau + (n + n') \varphi + 2\theta + \psi \}).$$

Set the time lag to 0, yielding the slice of cyclic autocorrelation function

$$R_x^\alpha(0) = \begin{cases} \left( \frac{1}{2} + \frac{1}{4} A^2 \right) h_{L0}(0), & \alpha = \pm(n - n') f_m, \\ \frac{1}{2} A h_{L0}(\pm\phi), & \alpha = \pm(n - n' \pm 1) f_m, \\ \frac{1}{8} A^2 h_{L0}(\pm 2\phi), & \alpha = \pm(n - n' \pm 2) f_m, \end{cases} \quad (10a)$$

$$R_x^\alpha(0) = \begin{cases} \frac{1}{2} \left( 1 + \frac{1}{2} A^2 \right) h_{H0}(0), & \alpha = \pm[2f_c + (n + n') f_m], \\ \frac{1}{2} A h_{H0}(\pm\phi), & \alpha = \pm[2f_c + (n + n' \pm 1) f_m], \\ \frac{1}{8} A^2 h_{H0}(\pm 2\phi), & \alpha = \pm[2f_c + (n + n' \pm 2) f_m], \end{cases} \quad (10b)$$

in lower and higher cyclic frequency domains, respectively, where the intermediate functions

$$h_{L0}(\psi) = \sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} J_n(B) J_{n'}(B) \quad (11a)$$

$$\cdot \exp\{\pm j [(n - n') \varphi + \psi]\},$$

$$h_{H0}(\psi) = \sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} J_n(B) J_{n'}(B) \quad (11b)$$

$$\cdot \exp\{\pm j [(n + n') \varphi + 2\theta + \psi]\}.$$

According to (10a) and (10b) and (11a) and (11b), the slice of cyclic autocorrelation function has two clusters of cyclic frequencies: one cluster concentrates around the cyclic frequency of 0 Hz separated by the modulating frequency  $f_m$ , and the other cluster spreads around twice the carrier

frequency  $2f_c$  with a spacing equal to the modulating frequency  $f_m$ .

In lower cyclic frequency domain, the cyclic frequency locations of present peaks are dependent on the difference of the two Bessel function orders  $n-n'$ . For any peak at a specific cyclic frequency,  $n-n' = \text{constant}$ . This leads to constant complex exponentials in (11a). Furthermore, according to the addition theorem of Bessel functions [26]

$$\sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} J_n(z) J_{n'}(z) = \begin{cases} 1, & n = n', \\ 0, & n \neq n', \end{cases} \quad (12)$$

peaks appear in the slice of cyclic autocorrelation function in lower cyclic frequency domain, if and only if the orders of the two Bessel functions are equal to each other; that is,  $n = n'$ . Therefore, the cyclic autocorrelation slice in lower cyclic frequency domain can be further simplified as

$$R_x^\alpha(0) = \begin{cases} \frac{1}{2} + \frac{1}{4}A^2, & \alpha = 0, \\ \frac{1}{2}A \exp(\pm j\phi), & \alpha = \pm f_m, \\ \frac{1}{8}A^2 \exp(\pm j2\phi), & \alpha = \pm 2f_m. \end{cases} \quad (13)$$

According to (13), in lower cyclic frequency domain of the cyclic autocorrelation slice, peaks appear at the cyclic frequencies of 0 Hz, the modulating frequency  $f_m$ , and its twice  $2f_m$ . If higher order harmonics of the modulating

frequency are taken into account, then peaks also appear at the modulating frequency harmonics  $kf_m$ .

The addition theorem of Bessel functions does not apply to the slice of cyclic autocorrelation function in higher cyclic frequency domain. The cyclic frequency locations of present peaks are dependent on the sum of the two Bessel function orders  $n+n'$ . For any specific peak,  $n+n' = \text{constant}$ . This does not mean  $n-n' = \text{constant}$ . According to (10b) and (11b), in higher cyclic frequency domain of the cyclic autocorrelation slice, sidebands appear at both sides of twice the carrier frequency  $2f_c$ , with a spacing equal to the modulating frequency  $f_m$ .

According to the above derivations, we can detect gearbox and bearing fault by monitoring the presence or magnitude change of sidebands around the cyclic frequency of 0 or twice the carrier frequency, that is, twice the meshing frequency for gearboxes and twice the resonance frequency for bearings, in the 0 lag slice of cyclic autocorrelation function. We can further locate the gearbox and bearing fault by matching the sideband spacing with the fault characteristic frequencies.

### 3. Cyclic Spectrum

**3.1. Definition.** For a signal  $x(t)$ , the cyclic spectral density is defined as the Fourier transform of the cyclic correlation function [25]

$$S_x^\alpha(f) = \int_{-\infty}^{\infty} R_x^\alpha(\tau) \exp(-j2\pi f\tau) d\tau, \quad (14)$$

where  $f$  is frequency.

**3.2. Cyclic Spectrum of AM-FM Signal.** Without loss of generality, we still consider the simplified gearbox and bearing vibration signal model, (5), by focusing on the fundamental frequency of the AM and FM terms. Its cyclic spectral density can be derived as [23, 24]

$$S_x^\alpha(f) = \begin{cases} \frac{1}{2}H(0), & \alpha = (n-n')f_m, f = f_c + \frac{1}{2}(n+n')f_m, \\ \frac{1}{8}A^2H(0), & \alpha = (n-n')f_m, f = f_c + \frac{1}{2}(n+n' \mp 2)f_m, \\ \frac{1}{4}AH(\pm\phi), & \alpha = (n-n' \pm 1)f_m, f = f_c + \frac{1}{2}(n+n' \mp 1)f_m, \\ \frac{1}{8}A^2H(2\phi), & \alpha = (n-n' \pm 2)f_m, f = f_c + \frac{1}{2}(n+n')f_m, \end{cases} \quad (15a)$$

$$S_x^\alpha(f) = \begin{cases} \frac{1}{2}H(2\theta), & \alpha = 2f_c + (n+n')f_m, f = \frac{1}{2}(n-n')f_m, \\ \frac{1}{8}A^2H(2\theta), & \alpha = 2f_c + (n+n')f_m, f = \frac{1}{2}(n-n' \mp 2)f_m, \\ \frac{1}{4}AH(2\theta \pm \phi), & \alpha = 2f_c + (n+n' \pm 1)f_m, f = \frac{1}{2}(n-n' \mp 1)f_m, \\ \frac{1}{8}A^2H(2\theta \pm 2\phi), & \alpha = 2f_c + (n+n' \pm 2)f_m, f = \frac{1}{2}(n-n')f_m, \end{cases} \quad (15b)$$

in lower and higher cyclic frequency domains, respectively, where the intermediate function

$$H(\psi) = \sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} J_n(B) J_{n'}(B) \exp(\pm j\psi). \quad (16)$$

According to (15a) and (15b), peaks appear at specific locations only on the cyclic frequency–frequency plane. Their cyclic frequency locations in lower cyclic frequency domain, as well as their frequency locations in higher cyclic frequency domain, are dependent on the difference of the two Bessel function orders  $n-n'$ . For any specific peaks,  $n-n' = \text{constant}$ . Then, the addition theorem of Bessel functions [26], (12), applies to the cyclic spectral density, (15a) and (15b). Hence, on the cyclic frequency–frequency plane of the cyclic spectral density, peaks appear if and only if the orders of the two Bessel

functions are equal to each other; that is,  $n = n'$ . Then cyclic spectral density can be simplified as

$$S_x^\alpha(f) = J_n^2(B) \begin{cases} \frac{1}{2}, & \alpha = 0, f = f_c + nf_m, \\ \frac{1}{8}A^2, & \alpha = 0, f = f_c + (n \mp 1) f_m, \\ \frac{1}{4}A \exp(\pm j\phi), & \alpha = \pm f_m, f = f_c + \left(n \mp \frac{1}{2}\right) f_m, \\ \frac{1}{8}A^2 \exp(\pm j2\phi), & \alpha = \pm 2f_m, f = f_c + nf_m, \end{cases} \quad (17a)$$

in lower cyclic frequency domain and

$$S_x^\alpha(f) = J_n^2(B) \begin{cases} \frac{1}{2} \exp(\pm j2\theta), & \alpha = 2f_c + 2nf_m, f = 0, \\ \frac{1}{8}A^2 \exp(\pm j2\theta), & \alpha = 2f_c + 2nf_m, f = \mp f_m, \\ \frac{1}{4}A \exp[\pm j(2\theta \pm \phi)], & \alpha = 2f_c + (2n \pm 1) f_m, f = \mp \frac{1}{2} f_m, \\ \frac{1}{8}A^2 \exp[\pm j(2\theta \pm 2\phi)], & \alpha = 2f_c + (2n \pm 2) f_m, f = 0, \end{cases} \quad (17b)$$

in higher cyclic frequency domain.

According to (17a) and (17b), for any peak on the cyclic frequency–frequency plane, its frequency location in lower

cyclic frequency domain, as well as its cyclic frequency location in higher cyclic frequency domain, is dependent on the Bessel function order  $n$ . Meanwhile, the peak magnitude only involves a few Bessel functions. Thus (17a) and (17b) can be further simplified as

$$S_x^\alpha(f) = \begin{cases} \frac{1}{2}J_n^2(B) + \frac{1}{8}A^2 [J_{n-1}^2(B) + J_{n+1}^2(B)], & \alpha = 0, f = f_c + nf_m, \\ \frac{1}{4}A [J_n^2(B) + J_{n+1}^2(B)] \exp(\pm j\phi), & \alpha = \pm f_m, f = f_c + \left(n + \frac{1}{2}\right) f_m, \\ \frac{1}{8}A^2 J_n^2(B) \exp(\pm j2\phi), & \alpha = \pm 2f_m, f = f_c + nf_m, \end{cases} \quad (18a)$$

$$S_x^\alpha(f) = \begin{cases} \frac{1}{2}J_n^2(B) + \frac{1}{8}A^2 [J_{n-1}^2(B) + J_{n+1}^2(B)] \exp(\pm j2\theta), & \alpha = 2f_c + 2nf_m, f = 0, \\ \frac{1}{4}A [J_n^2(B) + J_{n+1}^2(B)] \exp[\pm j(2\theta \pm \phi)], & \alpha = 2f_c + (2n + 1) f_m, f = \mp \frac{1}{2} f_m, \\ \frac{1}{8}A^2 J_n^2(B) \exp[\pm j(2\theta \pm 2\phi)], & \alpha = 2f_c + 2nf_m, f = \mp f_m, \end{cases} \quad (18b)$$

in lower and higher cyclic frequency domains, respectively.

According to (18a), in lower cyclic frequency domain, peaks appear at the cyclic frequencies of 0 Hz, modulating frequency  $f_m$ , and its twice  $2f_m$ . If higher order harmonics of modulating frequency are taken into account, then peaks also appear at the cyclic frequencies of the modulating frequency harmonics  $kf_m$ . Along the frequency axis, these peaks center

around the carrier frequency  $f_c$ , with a spacing equal to the modulating frequency  $f_m$ .

According to (18b), in higher cyclic frequency domain, peaks appear at the frequencies of 0 Hz, modulating frequency  $f_m$ , and its half  $1/2f_m$ . If higher order harmonics of modulating frequency are taken into account, then peaks also appear at the modulating frequency harmonics  $kf_m$  and

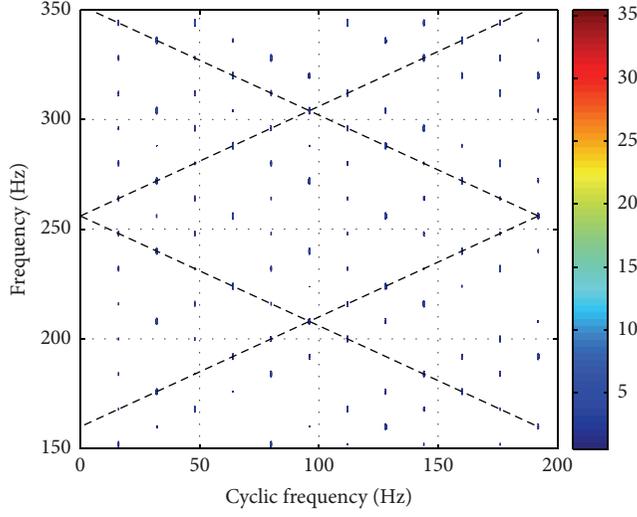


FIGURE 1: Cyclic spectrum distribution characteristics of an AM-FM signal.

their half  $k/2f_m$ . Along the cyclic frequency axis, these peaks center around twice the carrier frequency  $2f_c$ , with a spacing equal to twice the modulating frequency  $2f_m$ .

Observing the peak locations in (18a) and (18b), the peaks in cyclic spectrum distribute along the lines

$$f = \pm \frac{\alpha}{2} \pm (f_c \pm nf_m). \quad (19)$$

These lines form diamonds on the cyclic frequency–frequency plane, as illustrated by Figure 1, which is the cyclic spectrum of an AM-FM signal with a carrier frequency of 256 Hz and a modulating frequency of 16 Hz.

According to the above derivations, we can detect gearbox and bearing fault by monitoring the presence or magnitude change of peaks in the cyclic spectrum. For example, in lower cyclic frequency band, focus on the points corresponding to the cyclic frequency locations of modulating frequency  $f_m$  and its multiples  $kf_m$  and to the frequency locations of the carrier frequency plus the modulating frequency multiples  $f_c + nf_m$ . We can further locate the gearbox and bearing fault by matching the cyclic frequency spacing of peaks with the fault characteristic frequencies.

## 4. Gearbox Signal Analysis

**4.1. Specification of Experiment.** Figure 2 shows the experimental system. The gearbox has one gear pair. Table 1 lists the configuration and running condition of the gearbox. Two statuses of the gearbox are simulated. Under the healthy status, the gear pair is perfect. While under the faulty status, one of the pinion teeth is spalled, whereas the gear is perfect. During the experiment, accelerometer signals are collected at a sampling frequency of 20000 Hz and each record lasts for 3 s (i.e., 60000 samples) which covers more than 51 and 48 revolutions for the drive pinion and the driven

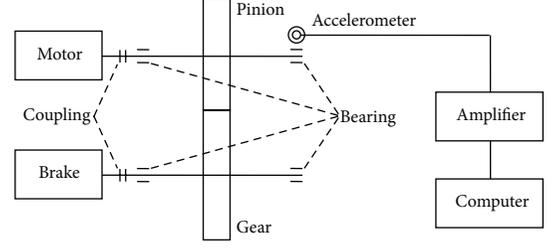


FIGURE 2: Gearbox experimental setup.

TABLE 1: Gearbox configuration and running condition.

	Drive pinion	Driven gear
Number of teeth	20	21
Rotating frequency (Hz)	17.285	16.462
Meshing frequency (Hz)		345.7
Brake torque (Nm)	—	200

gear, respectively, and is long enough to reveal the cyclostationarity.

**4.2. Signal Analysis.** Figure 3 shows the waveform, Fourier spectrum, cyclic correlation 0 lag slice, and cyclic spectrum (in lower cyclic frequency band) of the healthy gearbox. In the following analysis, the peak frequencies do not correspond exactly to those in Table 1. This is reasonable, because the actual speed is inevitably different from the set one in real experiments. In the cyclic correlation slice, Figure 3(c), most of the present peaks correspond to the driven gear rotating frequency harmonics. Although a few peaks appear at the pinion and gear characteristic frequencies and their harmonics, their magnitudes are not strong. In the cyclic spectrum (the colorbar on the right shows the magnitude, the same setting in the following cyclic spectra), Figure 3(d), along the frequency axis, the present peaks center around the gear meshing frequency 345.7 Hz, and its harmonics, and along the cyclic axis, they appear at the pinion and gear characteristic frequencies and their harmonics. The presence of these peaks is reasonable; since gear manufacturing errors and minor defects are inevitable, they will result in these peaks.

Figure 4 shows the waveform, Fourier spectrum, cyclic correlation 0 lag slice, and cyclic spectrum (in lower cyclic frequency band) of the faulty gearbox. In the cyclic correlation slice, Figure 4(c), peaks appear at the pinion and gear characteristic frequencies and their harmonics. Except the first two peaks, nearly all the other peaks appear at the cyclic frequency locations of the drive pinion rotating frequency harmonics, and they are higher than those in that of the healthy gearbox. In the cyclic spectrum, Figure 4(d), more and stronger peaks appear than in that of the healthy gearbox. Figure 4(e) shows the zoomed-in cyclic spectrum, for revealing the peak distribution details. Along the cyclic frequency axis, peaks appear along lines corresponding to the pinion

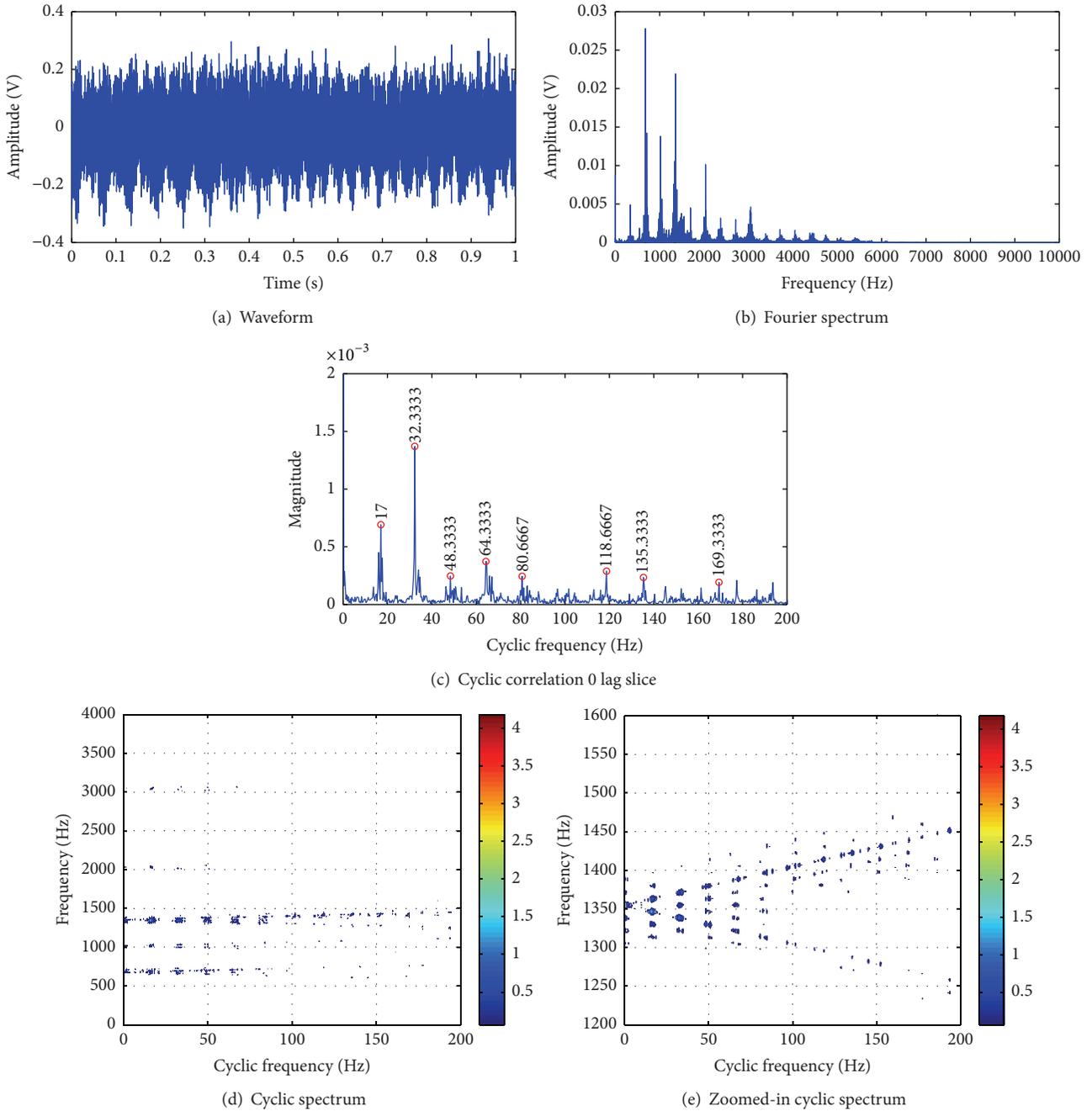


FIGURE 3: Healthy gearbox signal.

characteristic frequency and its multiples. While along the frequency axis, sidebands appear around 1400 Hz about four times the gear meshing frequency, with a spacing equal to the pinion characteristic frequency. Based on the theoretical derivations in Sections 2 and 3, the cyclic frequency locations of peaks in cyclic correlation slice and cyclic spectrum, as well as the sideband spacing along frequency axis in cyclic spectrum, relate to the modulating frequency of AM-FM signals. These features imply that more harmonics of the drive pinion rotating frequency get involved in modulating the gear

meshing vibration and that the drive pinion has a stronger modulation effect than the driven gear, indicating the pinion fault. These findings are consistent with the actual condition of the faulty gearbox.

## 5. Bearing Signal Analysis

5.1. Specification of Experiment. Figure 5 shows the experimental setup of the test [27]. A shaft is supported by four

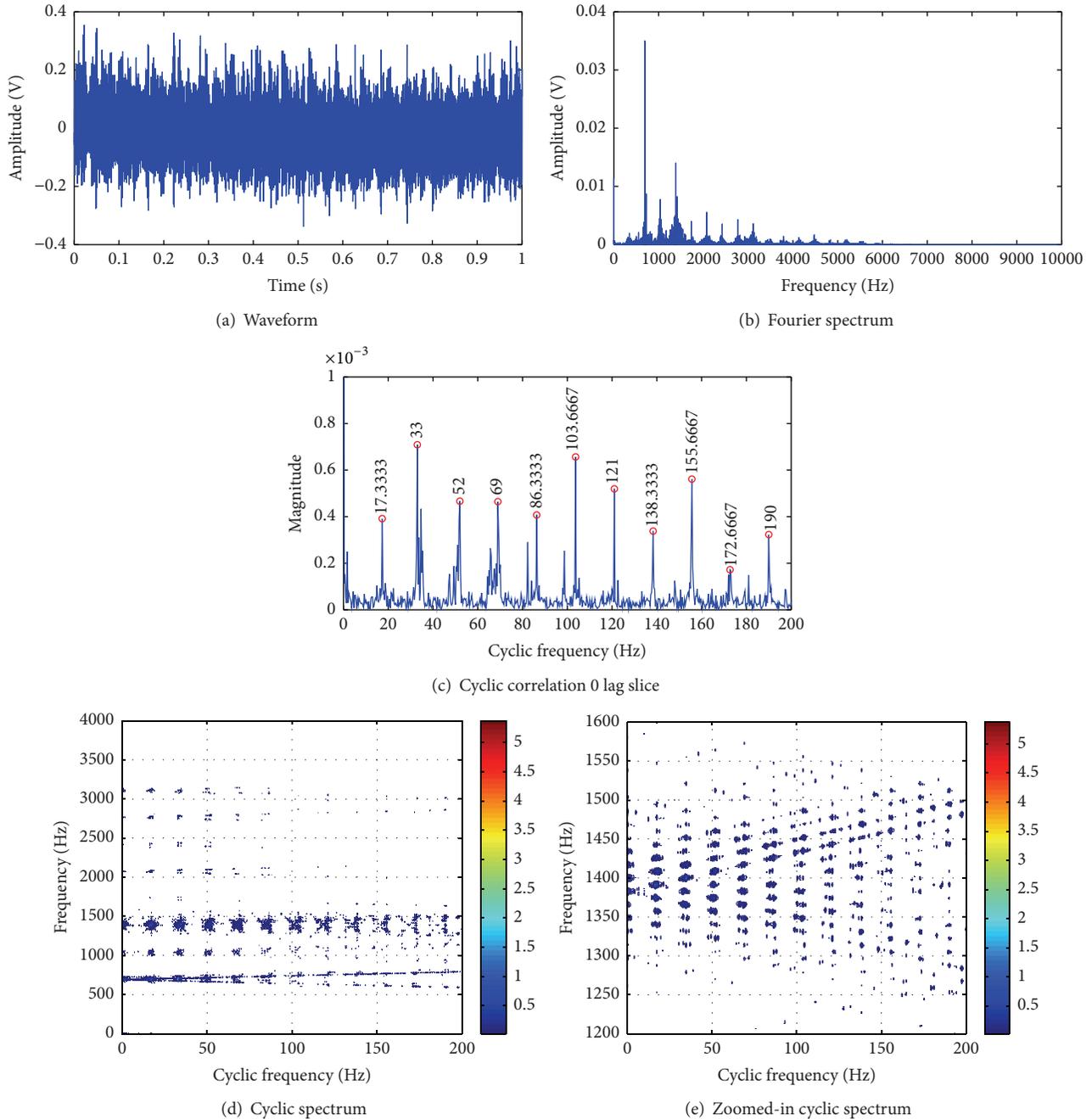


FIGURE 4: Faulty gearbox signal.

Rexnord ZA-2115 double row rolling element bearings and is driven by an AC motor through rub belts at a speed of 2000 rpm. A radial load of 6000 lbs is applied to the shaft and bearings by a spring mechanism. The bearing test was kept running for eight days until damage occurs to the outer race of bearing 1. In the normal case, all the four bearings are healthy. While in the faulty case, the outer race of bearing 1 has damage, as shown in Figure 6. Accelerometers are mounted on the bearing housings, and the vibration signals

are collected at a sampling frequency of 20480 Hz, and 20480 data points are recorded for both the healthy and faulty cases. The main parameters of the four bearings are listed in Table 2. The characteristic frequency of each element fault is calculated and listed in Table 3.

**5.2. Signal Analysis.** Figure 7 shows the waveform, Fourier spectrum, cyclic correlation 0 lag slice, and cyclic spectrum

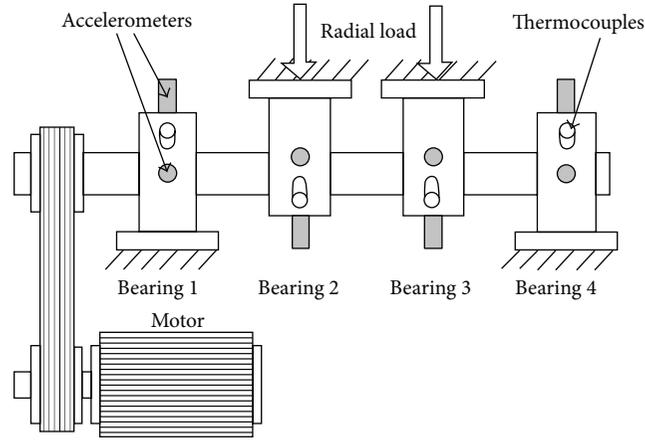


FIGURE 5: Bearing experimental setup [27].



FIGURE 6: Outer race damage of bearing 1 [27].

TABLE 2: Parameters of bearing Rexnord ZA-2115.

Number of rollers	Roller diameter	Pitch diameter	Contact angle
16	0.331 [in]	2.815 [in]	15.17 [°]

TABLE 3: Characteristic frequency of bearing Rexnord ZA-2115 [Hz].

Outer race	Inner race	Roller
236.404	296.930	139.917 279.833

(in lower cyclic frequency band) of the normal case. In the cyclic correlation slice, Figure 7(c), a few small peaks appear, but they do not correspond to any bearing component fault characteristic frequency or its multiples. Although some peaks appear in the cyclic spectrum, Figure 7(d), they do not distribute along lines corresponding to the cyclic frequencies of any bearing component fault characteristic frequency or its multiples. These features imply that no significant modulation on the resonance vibration and the bearing is healthy.

Figure 8 shows the waveform, Fourier spectrum, cyclic correlation 0 lag slice, and cyclic spectrum (in lower cyclic frequency band) of the faulty case. In the close-up view of the Fourier spectrum, some sidebands appear, but the frequency spacing is not identically equal to the outer race fault characteristic frequency. In the cyclic correlation slice, Figure 8(d), more peaks appear and their magnitudes are higher than in the normal case. Moreover, their cyclic frequency locations correspond to the outer race fault characteristic frequency and multiples up to the 25th order. In the cyclic spectrum, Figure 8(e), many peaks are present. More importantly, they distribute along lines associated with the outer race fault characteristic frequency and multiples. At any cyclic frequency of the outer race fault characteristic frequency or its multiple, these peaks form sidebands along the frequency axis, with a spacing equal to the outer race fault characteristic frequency. For example, in the zoomed-in cyclic spectrum, Figure 8(f), the four peaks labeled by capitals A, B, C, and D distribute along a line at the cyclic frequency of 710 Hz (about three times the outer race fault characteristic frequency), and they have a spacing of 236 Hz (approximately equal to the outer race fault characteristic frequency) along the frequency axis. These findings imply that the outer race characteristics

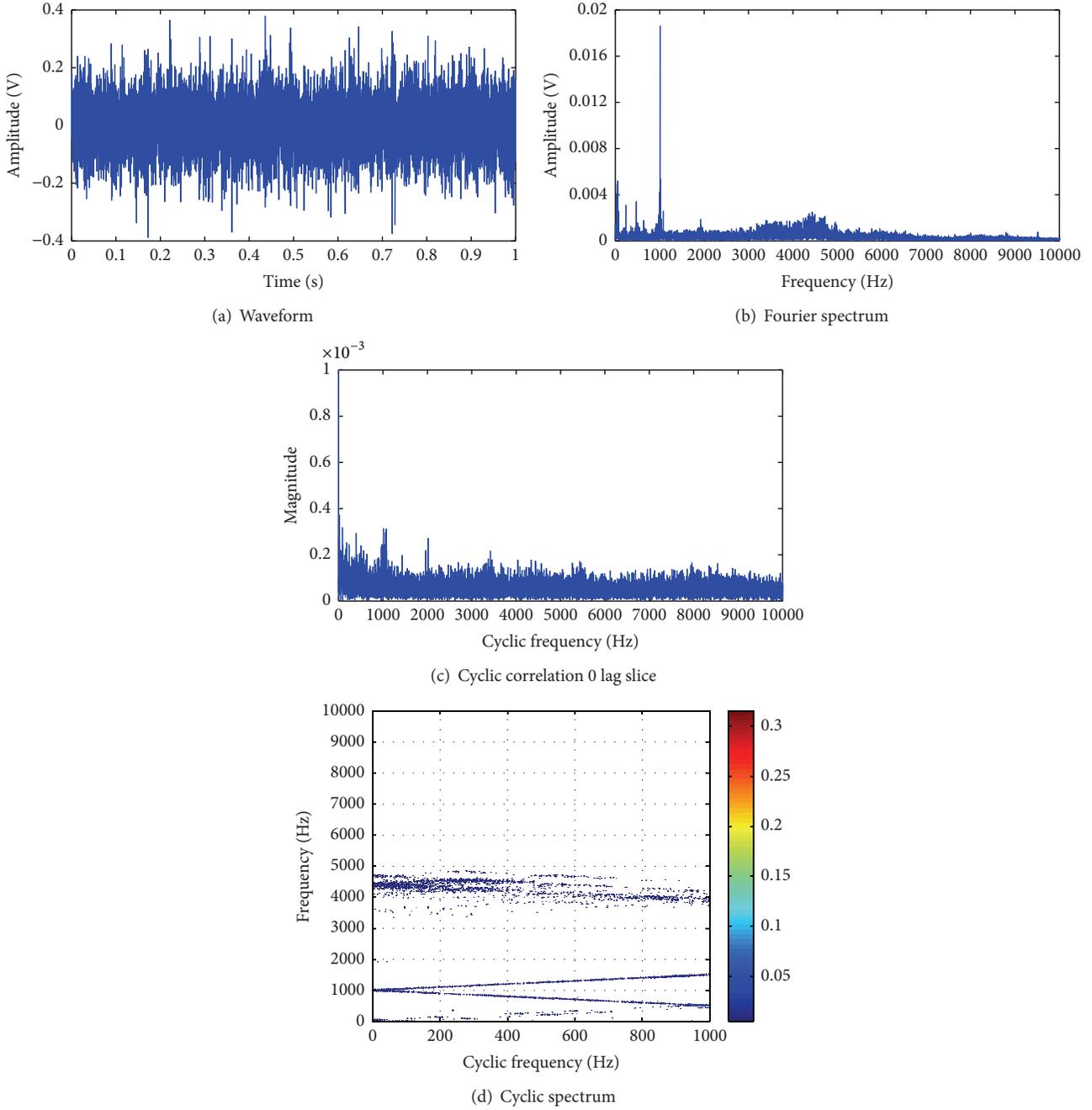


FIGURE 7: Normal bearing signal.

frequency has strong modulation effect on the resonance, indicating fault existence on the bearing outer race, in accordance with the actual settings.

**6. Discussion and Conclusions**

Since gearbox and rolling element bearing vibration signals are characterized by AM-FM feature, their Fourier spectra have complex sideband structure due to the convolution between the Fourier spectra of AM and FM parts as well as the

infinite Bessel series expansion of an FM term. Although gearbox and bearing fault can be identified via conventional Fourier spectrum, it involves complex sideband analysis and relies on the sideband spacing to extract the fault characteristic frequency.

Considering the cyclostationarity of gearbox and bearing vibration signals due to modulation and the suitability of cyclostationary analysis to extract modulation features of such signals, the explicit expressions of cyclic correlation and cyclic spectrum of general AM-FM signals are derived and

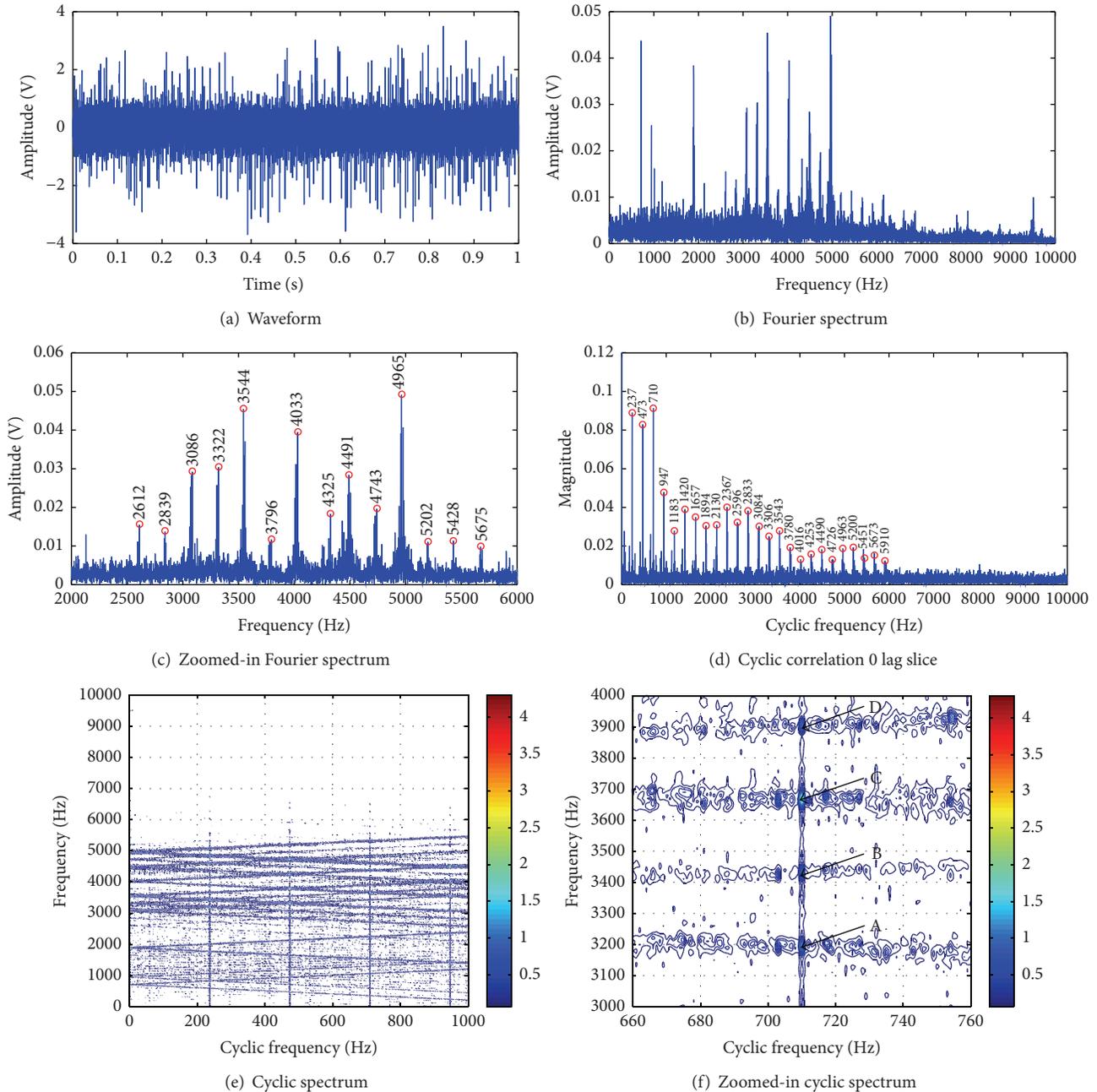


FIGURE 8: Faulty bearing signal.

validated via gearbox and bearing lab experimental signal analysis. According to the theoretical derivations, the cyclic frequency locations of peaks in cyclic correlation and cyclic spectrum directly correspond to the modulating frequency (i.e., the gear or bearing fault characteristic frequency) and harmonics. Therefore, cyclostationary analysis offers an effective approach to gearbox and bearing fault feature extraction.

**Conflict of Interests**

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

**Acknowledgments**

This work is supported by National Natural Science Foundation of China (11272047, 51475038), Program for New Century Excellent Talents in University (NCET-12-0775), and CETIM (Centre des Etudes Techniques des Industries Mécaniques de Senlis), France. The authors would like to thank the reviewers for their valuable comments and suggestions.

**References**

[1] R. Shao, W. Hu, and J. Li, “Multi-fault feature extraction and diagnosis of gear transmission system using time-frequency

- analysis and wavelet threshold de-noising based on EMD," *Shock and Vibration*, vol. 20, no. 4, pp. 763–780, 2013.
- [2] S. Luo, J. Cheng, and H. Ao, "Application of LCD-SVD technique and CRO-SVM method to fault diagnosis for roller bearing," *Shock and Vibration*, vol. 2015, Article ID 847802, 8 pages, 2015.
- [3] R. Yan, M. Shan, J. Cui, and Y. Wu, "Mutual information-assisted wavelet function selection for enhanced rolling bearing fault diagnosis," *Shock and Vibration*, In press.
- [4] G. Dalpiaz, A. Rivola, and R. Rubini, "Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears," *Mechanical Systems and Signal Processing*, vol. 14, no. 3, pp. 387–412, 2000.
- [5] C. Capdessus, M. Sidahmed, and J. L. Lacoume, "Cyclostationary processes: application in gear faults early diagnosis," *Mechanical Systems and Signal Processing*, vol. 14, no. 3, pp. 371–385, 2000.
- [6] L. Bouillaut and M. Sidahmed, "Cyclostationary approach and bilinear approach: comparison, applications to early diagnosis for helicopter gearbox and classification method based on hocs," *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 923–943, 2001.
- [7] Z. K. Zhu, Z. H. Feng, and F. R. Kong, "Cyclostationarity analysis for gearbox condition monitoring: approaches and effectiveness," *Mechanical Systems and Signal Processing*, vol. 19, no. 3, pp. 467–482, 2005.
- [8] G. Bi, J. Chen, F. C. Zhou, and J. He, "Application of slice spectral correlation density to gear defect detection," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 220, no. 9, pp. 1385–1392, 2006.
- [9] L. Li and L. Qu, "Cyclic statistics in rolling bearing diagnosis," *Journal of Sound and Vibration*, vol. 267, no. 2, pp. 253–265, 2003.
- [10] 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.
- [11] J. Antoni, "Cyclic spectral analysis in practice," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 597–630, 2007.
- [12] J. Antoni, "Cyclic spectral analysis of rolling-element bearing signals: facts and fictions," *Journal of Sound and Vibration*, vol. 304, no. 3–5, pp. 497–529, 2007.
- [13] A. Raad, J. Antoni, and M. Sidahmed, "Indicators of cyclostationarity: theory and application to gear fault monitoring," *Mechanical Systems and Signal Processing*, vol. 22, no. 3, pp. 574–587, 2008.
- [14] J. Antoni, "Cyclostationarity by examples," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 987–1036, 2009.
- [15] J. Urbanek, T. Barszcz, and J. Antoni, "Time-frequency approach to extraction of selected second-order cyclostationary vibration components for varying operational conditions," *Measurement*, vol. 46, no. 4, pp. 1454–1463, 2013.
- [16] D. Wang and C. Shen, "An equivalent cyclic energy indicator for bearing performance degradation assessment," *Journal of Vibration and Control*, 2014.
- [17] R. B. Randall, "A new method of modeling gear faults," *Journal of Mechanical Design*, vol. 104, no. 2, pp. 259–267, 1982.
- [18] 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.
- [19] P. D. McFadden and J. D. Smith, "Model for the vibration produced by a single point defect in a rolling element bearing," *Journal of Sound and Vibration*, vol. 96, no. 1, pp. 69–82, 1984.
- [20] P. D. McFadden and J. D. Smith, "The vibration produced by multiple point defects in a rolling element bearing," *Journal of Sound and Vibration*, vol. 98, no. 2, pp. 263–273, 1985.
- [21] M. Liang and I. S. 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.
- [22] Z. Feng, M. J. Zuo, R. Hao, and F. Chu, "Gear crack assessment based on cyclic correlation analysis," in *Proceedings of the 8th International Conference on Reliability, Maintainability and Safety (ICRMS '09)*, pp. 1071–1076, Chengdu, China, July 2009.
- [23] Z. Feng, R. Hao, F. Chu, M. J. Zuo, and M. El Badaoui, "Application of cyclic spectral analysis to gear damage assessment," in *Proceedings of the Prognostics and Health Management Conference*, p. 3058, Macau, China, January 2010.
- [24] Z. Feng and M. J. Zuo, "Gearbox diagnosis based on cyclic spectral analysis," in *Proceedings of the 3rd Annual IEEE Prognostics and System Health Management Conference (PHM '12)*, pp. 1–5, Beijing, China, May 2012.
- [25] W. A. Gardner, *Cyclostationarity in Communications and Signal Processing*, IEEE Press, New York, NY, USA, 1994.
- [26] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, NY, USA, 1972.
- [27] 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

# Feature Extraction of Faulty Rolling Element Bearing under Variable Rotational Speed and Gear Interferences Conditions

Dezun Zhao, Jianyong Li, and Weidong Cheng

*School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Weidong Cheng; wdcheng@bjtu.edu.cn

Received 6 May 2015; Revised 15 July 2015; Accepted 26 July 2015

Academic Editor: Wahyu Caesarendra

Copyright © 2015 Dezun Zhao 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 the field of rolling element bearing fault diagnosis, variable rotational speed and gear noise are main obstacles. Even though some effective algorithms have been proposed to solve the problems, their process is complicated and they may not work well without auxiliary equipment. So we proposed a method of faulty bearing feature extraction based on Instantaneous Dominant Meshing Multiply (IDMM) and Empirical Mode Decomposition (EMD). The new method mainly consists of three parts. Firstly, IDMM is extracted from time-frequency representation of original signal by peak searching algorithm, which can be used to substitute the bearing rotational frequency. Secondly, resampled signal is obtained by an IDMM-based resampling algorithm; then it is decomposed into a number of Intrinsic Mode Functions (IMFs) based on the EMD algorithm. Calculate kurtosis values of IMFs and an appropriate IMF with biggest kurtosis value is selected. Thirdly, the selected IMF is analyzed with envelope demodulation method which can describe the fault type of bearing. The effectiveness of the proposed method has been demonstrated by both simulated and experimental mixed signals which contain bearing and gear vibration signal.

## 1. Introduction

There are two main obstacles for rolling bearing diagnosis; they are variable rotational speed and gear vibration noise [1]. On one hand, time-varying speed will directly influence frequency and amplitude of rolling element bearing vibration signal. In terms of frequency, the fault impulse will not recur at a constant interval; hence demodulated resonance technique and any other methods based on the premise of constant speed are no longer applicable. Furthermore, the amplitude of the bearing vibration signal changes with the rotational speed in time domain. On the other hand, the amplitude of gear meshing noise is higher than bearing in rotating machinery. Even though the gear does not have fault, impulses of faulty bearing will be contaminated by gear noise, and the extraction of resonance frequency excited by bearing fault also will be influenced.

Many efforts have focused on the study of time-varying rotational speed and gear vibration noise. Order tracking [2, 3] is one of the most effective methods to deal with variable speed signal. The essence of the order tracking is resampling

the original vibration signal at a constant angle increment which converts the nonstationary signal (in time domain) into the stationary one (in angular domain) to remove the effects of speed fluctuation. Order tracking algorithm can be divided into three types: hardware-based order tracking, computed order tracking [2–4], and order tracking based on instantaneous frequency estimation [5, 6]. Compared with hardware-based order tracking and computed order tracking, the order tracking method based on instantaneous frequency estimation can overcome the limits of auxiliary equipment and costs. Hence it has become a new focus for many researchers, and its principle is extracting rotational frequency from original vibration signal then angular domain resampling. However, because of gear noise interference, it is difficult to directly extract the bearing rotational frequency from the mixed vibration signal. Traditional methods to eliminate the interference of the gear noise mainly include Time Synchronous Average (TSA) [7], Linear Prediction [8], Adaptive Noise Cancellation (ANC) [9], Self-Adaptive Noise Cancellation (SANC) [10], Discrete/Random Separation (DRS) [11], and Edited Cepstrum [12]. These algorithms

require rotational speed to keep stable except the ANC algorithm. ANC algorithm requires auxiliary equipment to obtain reference signals of gear, and the installation position of the auxiliary equipment is very important.

For faulty rolling element bearing feature extraction under time-varying rotational speed and gear noise, Borghesani et al. [13] proposed a method. This method consists of three steps: removing the effects of speed fluctuation based on the order tracking algorithm, eliminating gear noise by traditional denoising method in the gear angular domain, returning the denoised signal to time domain based on the reverse order tracking and angular resampling by bearing rotational frequency, and, at last, getting envelope order spectrum to determine bearing fault type. On the other hand, Wang [14] presented an autoregressive model-based method to diagnose faulty bearing under time-varying speed and gear noise conditions. The methods proposed by Borghesani and Wang effectively solve the problem of bearing fault diagnosis under time-varying rotational speed and gear noise, but these methods need speed information of bearing and gear which is obtained by auxiliary equipment. Hence it may not work in some conditions that have restrictions of installation space and costs. Afterwards, Wang [15, 16] proposed a new concept, named Instantaneous Dominant Meshing Multiply (IDMM), which can be directly extracted from mixed signal. Based on IDMM, a function is constructed as the reference signal for the ANC algorithm which is used to remove gear noise. Signal resampling based on instantaneous fault characteristic frequency (IFCF) which is extracted from time-frequency representation of remaining bearing signal component, then, obtains fault characteristic order (FCO) spectrum to determine bearing fault type. Although the effectiveness of the improved ANC algorithm has been validated, the process is very complex, because we need to try many times to obtain optimal parameters of the ANC algorithm, and the process must rely on FCO spectrum. Even though we have optimal parameters of the ANC algorithm, the multiples of the gear meshing frequency except for the IDMM cannot be removed. On the other hand, IFCF has errors which will influence bearing fault diagnosis.

In view of above analysis, a new method based on IDMM and EMD is proposed. The proposed method does not need auxiliary equipment to measure speed information, avoids complex parameters selecting, and simplifies the process of rolling element bearing fault diagnosis. It has three main steps: (a) extracting IDMM trend from time-frequency representation of mixed signal and signal resampling via the IDMM-based resampling algorithm, (b) obtaining IMFs by the EMD algorithm and selecting suitable IMF by kurtosis values, and (c) obtaining envelope order spectrum to diagnose rolling element bearing.

## 2. Theoretical Background

**2.1. IDMM Trend.** In the previous studies, gear noise has been generally classified as narrowband periodic signal. Due to this characteristic, we can easily find gear meshing frequency and its multiples. The amplitude of the dominant meshing

multiple has fixed proportion with gear rotational frequency, and it can be represented as

$$f'_g = c \cdot f_g = c \cdot \frac{z_i n_i}{60}, \quad (1)$$

where  $c$  is a number of the collection  $j$  and  $j$  ( $j = 1, \dots, G$ ) is the meshing multiple number; the value of  $c$  is different in different working conditions; in the most general case, it is 2 or 3;  $f_g$  is the gear meshing frequency;  $z_i$  is the number of gear teeth;  $n_i$  is the rotational speed of gear whose unit is r/min.

If gear rotational frequency changes with time, Instantaneous Dominant Meshing Multiple will change in proportion to it. The IDMM is defined as an instantaneous meshing frequency harmonic that has the highest amplitude among all harmonics of the instantaneous meshing frequency [15], and the IDMM trend is a collection of IDMM. IDMM can be easily extracted from time-frequency representation of the mixed signal, because it has the highest amplitude among all harmonics of the meshing frequency and the rolling element bearing signal component has relatively weaker one. Short-Time Fourier Transform (STFT) is used to compute time-frequency representation of mixed signal in this paper, and the IDMM is calculated as follows:

$$\text{STFT}(t, w) = \int_{-\infty}^{+\infty} g(u) f(t+u) e^{-jwu} du \quad (2)$$

$$\text{IDMM}_i = \arg \max \{ \text{STFT}_i \} \quad (i = 1, 2, \dots, n),$$

where  $t$  is the time;  $w$  is the frequency;  $g(u)$  is a Gaussian window function;  $\text{STFT}_i$  represents instantaneous frequency spectrum;  $i$  is the sequence number of instantaneous frequency spectrum whose value is from 1 to  $n$ ; the number of instantaneous frequency spectrum is determined by the length of Gaussian window function;  $\arg \max$  function is used to find frequency domain abscissa of  $\text{STFT}_i$  when  $\text{STFT}_i$  has the largest value.

**2.2. EMD Algorithm.** EMD is a self-adaptive analysis method proposed by Huang et al. [17] in 1998. It is based on the local characteristic time scale of signal and decomposes the signal into a number of narrowband components, called IMFs, that satisfies two conditions:

- (a) In the whole data set, the number of extrema and the number of zero-crossings must either be equal or differ at most by one.
- (b) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The process of EMD is shown as follows:

- (1) Obtain the local maxima and minima of the signal  $x(t)$ . Interpolate, using natural cubic splines, along the points of local maxima and minima in order to form an upper and a lower envelope.
- (2) The mean of the two envelopes is designated as  $m_1$ .  $h_1$ , the difference between  $x(t)$  and  $m_1$ , is calculated as

$$h_1 = x(t) - m_1. \quad (3)$$

If  $h_1$  satisfies all the requirements of IMF, it is the first IMF of  $x(t)$ .

- (3) If not,  $h_1$  is treated as the original signal and processed by previous steps. The mean of the two envelopes is designated by  $m_{11}$ . Let  $h_{11} = h_1 - m_{11}$ . If  $h_{11}$  is not an IMF, we can repeat the cycle until  $h_{1k}$  is an IMF; that is,

$$h_{1k} = h_{1\{k-1\}} - m_{1k}. \quad (4)$$

Then we define  $c_1 = h_{1k}$ ; it is the first IMF of the original signal.

- (4) Separate  $c_1$  from  $x(t)$  as follows:

$$r_1 = x(t) - c_1. \quad (5)$$

The residue,  $r_1$ , is treated as the original signal and subjected to the same cyclic process as described above. Repeat cycle  $n$  times and get  $n$  IMFs as follows:

$$r_2 = r_1 - c_2 \cdots r_n = r_{n-1} - c_n. \quad (6)$$

Stop cyclic process until  $r_n$  becomes a monotonic function from which no more IMF can be extracted. Via EMD decomposition, original signal can be represented as

$$x(t) = \sum_{i=1}^n c_i + r_n, \quad (7)$$

where  $r_n$  is a residue, which represents the signal's central tendency.

**2.3. IDMM-Based Resampling Algorithm.** We obtain resampled signal by the IDMM-based resampling algorithm. Traditional resampling algorithm is firstly to fit speed curve, secondly to calculate function relationship between the angle of reference shaft  $\theta$  and the time  $t$ , and thirdly to obtain resampled signal via interpolation algorithm. However, calculated burden of the traditional algorithm is too heavy, and its efficiency is quite low, as it needs to solve a number of quadratic equations. We use the IDMM-based resampling algorithm to resample the signal. The essence of the algorithm is to divide signal into a number of parts, larger speed parts with larger sampling rate and lower speed parts with lower one, and the ratio of sampling rate of two different time segments is equal to the ratio of rotational frequency. The process of IDMM-based resampling algorithm is shown as follows:

- (1) Divide mixed signal into  $n$  ( $n$  is 200 in this paper) parts, named  $x_1, x_2, \dots, x_n$ , and every part of signal has a rotational frequency:  $IDMM_1, IDMM_2, \dots, IDMM_n$ .
- (2) Define the initial sample rate, namely,

$$f_{s,base} = Fs, \quad (8)$$

where  $Fs$  is sampling rate of the original signal.

The initial IDMM is calculated as

$$IDMM_{base} = \min(IDMM_1, IDMM_2, \dots, IDMM_n). \quad (9)$$

- (3) Based on the ratios of rotational frequency of  $x_i$ , reset sampling rate of  $x_i$  is determined as

$$f_{s,i} = \frac{(f_{s,base} \times IDMM_i)}{IDMM_{base}}. \quad (10)$$

- (4) Obtain resampled signal based on the computed sampling rates,  $f_{s,i}$ . Firstly, determine the new sampling point based on new sampling rate,  $f_{s,i}$ ; secondly, based on original signal  $x_i$ , determine resampled signal  $x'_i$  by polynomial interpolation; lastly, the  $x'_i$ 's are arranged by the value of  $i$ , and the resampled signal is shown as  $x'_1, x'_2, \dots, x'_n$ .

### 3. Feature Extraction of Faulty Bearing Based on IDMM and EMD

Instantaneous Dominant Meshing Multiply (IDMM) is easily extracted from time-frequency representation of the mixed signal by peak searching algorithm. In addition, IDMM trend is proportional to bearing rotational frequency. Hence we proposed resampling the signal based on IDMM trend. EMD is a self-adaptive algorithm, which decomposes signal by local characteristic time scale of the signal. Furthermore, IMFs are distributed from high to low in turn according to their frequency. Bearing signal is high frequency resonance signal, and gear meshing signal generally has lower frequency. Hence we use the EMD algorithm to remove interference of gear noise. In the proposed method, the IMF with the biggest Kurtosis value is selected for further analysis, which means that it includes much more fault information than other IMFs. The Kurtosis value of a random variable  $x$  is defined as

$$K(x) = \frac{E(x - \mu)^4}{\sigma^4}, \quad (11)$$

where  $E(x)$  represents mathematical expectation of a random variable  $x$  and  $\sigma$  and  $\mu$  are standard deviation and mean value of  $x$ , respectively.

The steps of the proposed method can be given as follows. A detailed flowchart is presented in Figure 1.

- (1) Extract IDMM trend from time-frequency representation (TFR) of mixed signal by peak searching algorithm. Then obtain resampled signal by the IDMM-based resampling algorithm.
- (2) Decompose signal by EMD algorithm and obtain IMFs. Compute Kurtosis value of IMFs based on (11) and select an appropriate IMF.
- (3) Obtain envelope order spectrum of the selected IMF to determine bearing fault type.

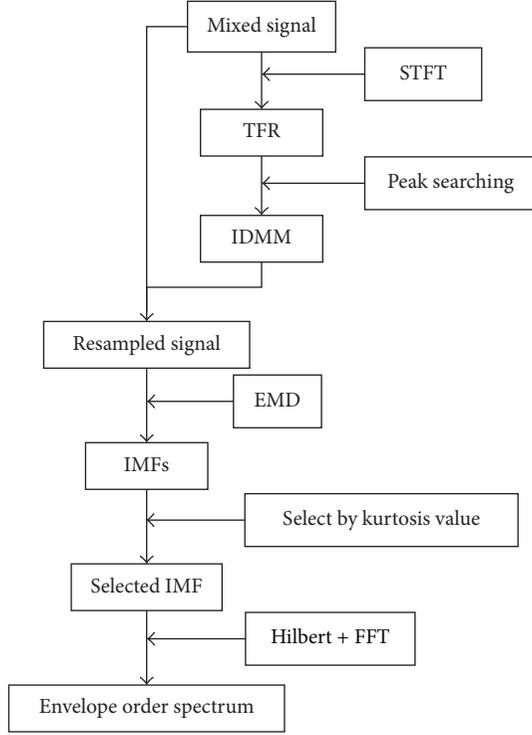


FIGURE 1: Flowchart of rolling element bearing faulty feature extraction based on IDMM and EMD.

## 4. Simulation and Experimental Validation

**4.1. Simulation Study.** In this subsection, a synthetic speed-up signal, which includes faulty rolling element bearing, gear, and the white Gaussian noise, is constructed to demonstrate the effectiveness of the proposed method.

The model representing the faulty rolling element bearing signal under speed-up,  $x_{\text{bearing}}$ , is given by

$$x_{\text{bearing}}(t) = \sum_{m=1}^N A_m e^{-\beta(t-t_m)} \sin(w_r(t-t_m)) u(t-t_m), \quad (12)$$

where  $A_m$  is the amplitude of the  $m$ th fault impulse,  $\beta$  is the structural damping frequency,  $w_r$  is the fault exiting resonance frequency,  $u(t)$  is a unit step function, and  $t_m$  is the occurrence time of the  $m$ th impulse and it can be calculated as

$$t_1 = (1 + \mu) \cdot \frac{1/(f(t_0))}{n} \quad (13)$$

$$t_m = (1 + \mu) \cdot \frac{1/(f(t_{m-1}))}{n} \quad m = 2, 3, \dots, N,$$

where  $u$ ,  $u = 0.01 \sim 0.02$ , is the errors caused by the random slippage of the rolling elements;  $f(t)$  is the rotational frequency function of rolling element bearing and  $f(t) = 2.5t + 5$ ;  $t_0 = 0$  and  $n$  is the number of fault impacts per shaft revolution.

TABLE 1: Parameters of mixed bearing and gear simulation model under time-varying rotational speed.

Parameters	Value
Length of the signal	8 s
Length of $f(t)$	5 Hz–25 Hz
Length of $f_g(t)$	6 Hz–30 Hz
Sampling rate $f_s$	20000 Hz
Fault characteristic coefficient	3
Bearing resonance frequency $w_r$	5000 Hz
Number of teeth $L$	18
Rotating frequency ratio $\tau$	1.2

The model representing the gear signal under speed-up can be written as

$$x_{\text{gear}}(t) = \sum_{j=1}^G X_j(t) \cos(2\pi j L F_g(t)), \quad (14)$$

where  $j$  ( $1, 2, \dots, G$ ) is the meshing multiple number and  $X_j(t)$  is the amplitude of  $j$ th harmonic frequency.  $L$  is the tooth number.  $F_g(t)$  is integral function of the gear rotational frequency,  $f_g(t)$ .  $\tau$  is the ratio of gear rotational frequency to bearing; namely,  $f_g(t) = \tau f(t)$ . The amplitudes of gear meshing frequency  $X_1 = 0.4$  and amplitudes of its 2nd and 3rd multiples are set as 5 and 0.6, which means that the 2nd multiple is the dominant meshing multiple. The other parameters in the model are listed in Table 1.

The simulation mixed signal can be represented as

$$x(t) = x_{\text{bearing}}(t) + x_{\text{gear}}(t) + \text{noise}(t). \quad (15)$$

Four seconds of the simulation signal is selected for analysis, and the selected simulation signal is shown in Figure 2(a). The amplitude of the mixed signal goes up in proportion to the rotational frequency and the fault impulses are not obvious. Time-frequency representation of the mixed signal is shown in Figure 2(b). From Figure 2(b), the resonance frequency can be found at around 5000 Hz. The three linearly rising lines in the lower part of Figure 2(b) are the three multiples of the gear meshing frequency, and 2nd meshing frequency is quiet obvious.

IDMM trend is extracted using peak searching algorithm from time-frequency representation of the mixed simulation signal, and it is shown in blue solid line in Figure 3, together with the preset gear meshing frequency trend (red solid line) and its 2nd multiple (red dash line). The extracted IDMM trend and the 2nd multiple of the preset gear meshing frequency are almost identical. Hence we can conclude that IDMM trend can be easily extracted. Because IDMM trend has invariable proportion (the proportion coefficient is 43.2, which is calculated by (1) and parameters of the simulation model) to bearing rotational frequency, it can be used to substitute bearing rotational frequency.

We obtain the resampled signal by the IDMM-based resampling algorithm. The angular resampled signal is shown in Figure 4(a). Figure 4(b) is the time-frequency representation of the angular resampled signal. The EMD algorithm is

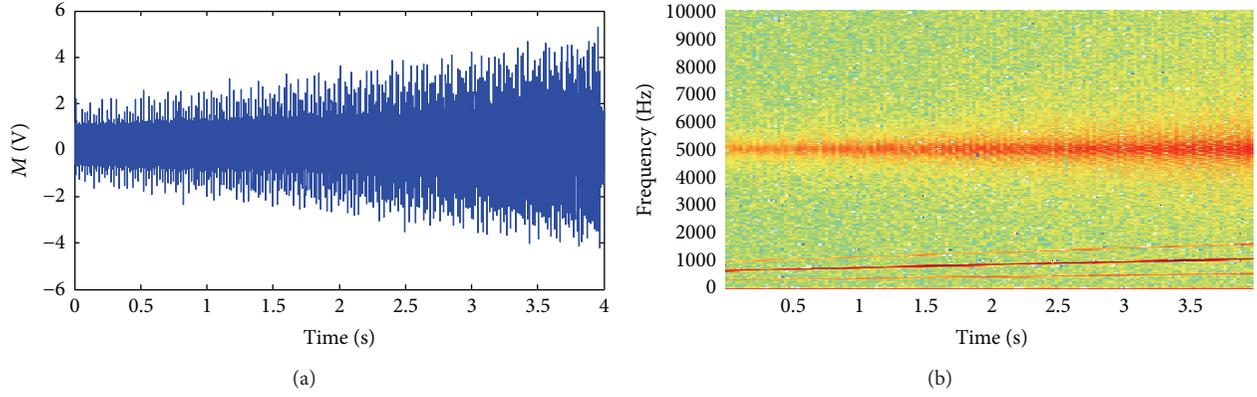


FIGURE 2: (a) Mixed simulation signal in time domain. (b) Time-frequency representation of the mixed signal.

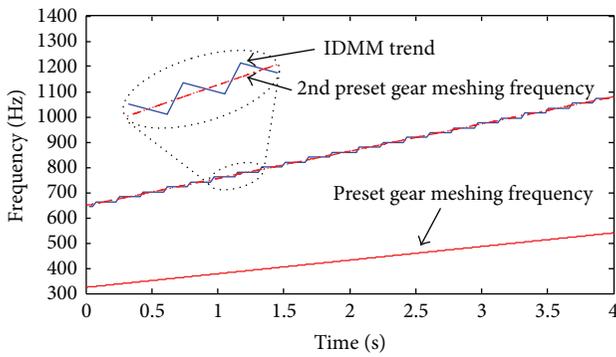


FIGURE 3: Extracted IDMM trend and preset gear meshing frequency.

TABLE 2: Kurtosis values of IMFs.

IMF	1	2	3	4	5	6
Kurtosis value	16.91	1.84	6.37	5.18	1.76	1.45

used to process the resampled signal, and the decomposed first six IMFs are shown in Figure 5. We respectively calculate Kurtosis values of IMFs and Kurtosis values of IMFs are listed in Table 2. IMF1 is selected for further analysis as the biggest kurtosis value.

Figure 6 is time-frequency representation of the IMF1. Compared to Figure 4(b), the results revealed that the three multiples of the gear meshing frequency in the mixed signal have been removed by EMD algorithm. Apply envelope demodulation method to process the IMF1, and the corresponding envelope order spectrum is shown in Figure 7(a). In this picture, fault characteristic order (FCO) and its 2nd and 3rd harmonics are easily identified whose values are 0.0696, 0.139, and 0.2086, respectively. Bearing Rotational Order (BRO) and its 2nd harmonic are also easily identified whose values are 0.0232 and 0.0462. The ratio of the FCO and the BRO is 3, which is equal to the preset fault characteristic coefficient. Hence the fault type of rolling element bearing can be identified.

TABLE 3: Comparison of the envelope order spectrum.

Harmonic order	Amplitudes of the harmonics		Changed ratio
	Resampled signal	Denoising signal	
BRO	0.035	0.036	2.9%
2BRO	0.019	0.027	42.1%
FCO	0.116	0.247	113%
2FCO	0.119	0.251	111%
3FCO	0.100	0.186	86%
4FCO	0.0787	0.125	58.8%

For the purpose of comparison, the resampled signal is directly analyzed with envelope demodulation method, and the corresponding envelope order spectrum is shown in Figure 7(b), from which the gear interference peaks are clearly revealed. Figures 7(a) and 7(b) indicated that the proposed method can effectively remove interference of gear meshing frequency. Furthermore, the amplitude of each harmonic is obviously increased. The changed ratio of the FCO amplitude between the resampled signal and the denoising signal is 113%; other amplitudes of harmonics (see column 2 and 3 of Table 3) and changed ratios (see column 4 of Table 3) are given in Table 3, respectively.

**4.2. Experimental Analyses.** The effectiveness of the proposed method is further verified by the mixed signal, which is measured on the vibration test rig. The layout of the test rig is illustrated in Figure 8. The gearbox is driven by an electrical motor, and the speed of the motor is controlled by a speed controller. A tachometer is fixed on the output shaft to measure speed information. The sampling rate is 24,000 Hz. The signal is collected by an acceleration sensor which is mounted on the top surface of the faulty bearing, YE6231 acquisition card, and its corresponding software. Gears do not have fault, and their reduction ratio is 3/11. A simulated outer race faulty bearing, 6206, is assembled in the forepart of the input shaft. The parameters of faulty rolling element bearing and fault characteristic coefficient of outer race (FCC<sub>o</sub>) are shown in Table 4.

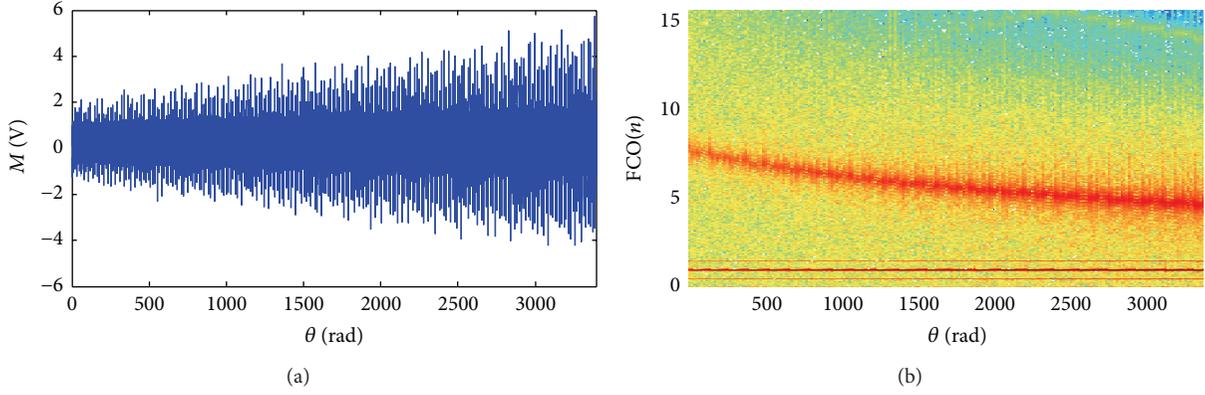


FIGURE 4: (a) Angular resampled signal. (b) Time-frequency representation of angular resampled signal.

TABLE 4: Parameters of bearing in experiment.

Bearing type	Number of balls $n$	Ball diameter $d$ /mm	Pitch diameter $D$ /mm	Contact angle $\alpha$	FCC <sub>o</sub>
6206	9	9.52	51	0	3.66

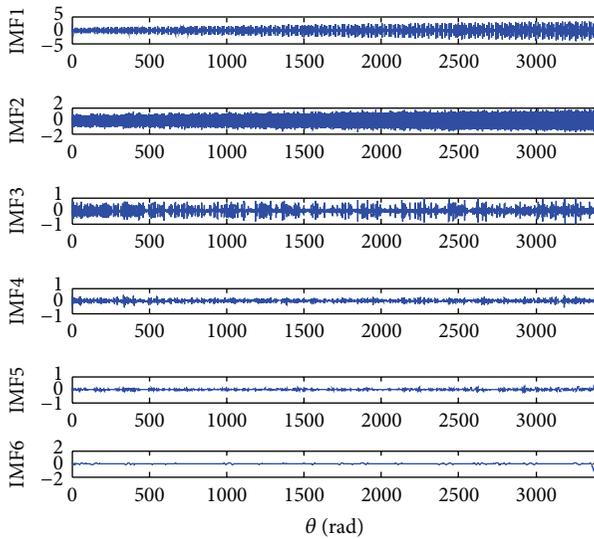


FIGURE 5: Decomposition result of the resampled signal.

The measured signal in time domain is shown in Figure 9(a), and the length of the signal is 1 second. Figure 9(b) is time-frequency representation of the measured signal. From the picture we can easily find the IDMM trend. The IDMM trend is extracted from Figure 9(b) using the peak searching algorithm. Figure 10 displays the IDMM trend (blue solid line) and the measured gear meshing frequency (red solid line) and its 3rd multiple trend (red dash line) for comparison. The measured gear meshing frequency is calculated by the measured speed information and (1). As shown in Figure 10, the IDMM trend and the 3rd multiple of gear meshing frequency almost overlap with each other. The proportional relationship between the IDMM trend and the gear meshing frequency is invariable, which is calculated by

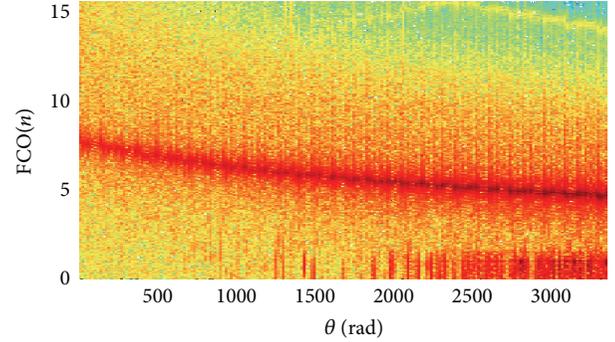


FIGURE 6: Time-frequency representation of IMF1.

the parameters of the gearbox. Therefore, the IDMM trend can be reliably used for signal resampling.

Resampled signal is obtained by an IDMM-based resampling algorithm, and the EMD algorithm is used to decompose the resampled signal. The IMFs are shown in Figure 11, respectively. Kurtosis values of IMFs are 25.34, 1.98, 4.18, 4.6, 3.31, and 1.12, respectively. Because IMF1 has the biggest Kurtosis value, we select IMF1 for analysis. The envelope order spectrum of IMF1 is shown in Figure 12(a). As shown in Figure 12(a), the FCO and its harmonics are obvious; we can also easily find the BRO and its 2nd harmonic. Values of the FCO and the BRO are 0.0786 and 0.0213, respectively. The calculated fault characteristic coefficient is 3.62, and the actual value is 3.66; hence we can determine that the fault is on the outer race of the rolling element bearing.

We directly analyze resampled signal by envelope demodulation method, and the envelope order spectrum is shown in Figure 12(b). There are obvious interference peaks, and the BRO cannot be found in the picture. Compared to Figure 12(b), the BRO and its 2nd harmonic are easily recognized in the envelope order spectrum of IMF1. In addition, the FCO and the BRO have higher amplitude. For easy

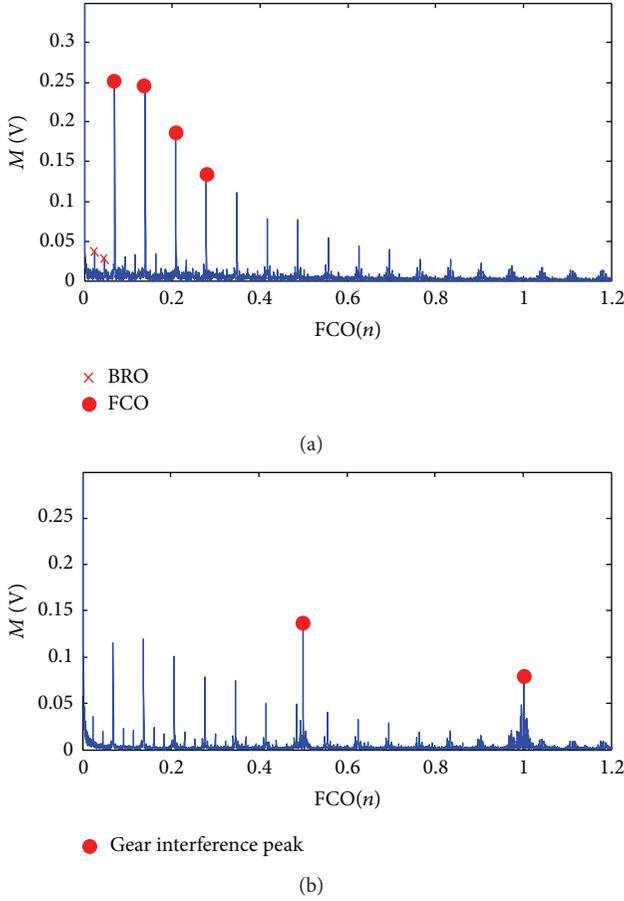


FIGURE 7: (a) Envelope order spectrum of IMF1. (b) Envelope order spectrum of resampled signal.

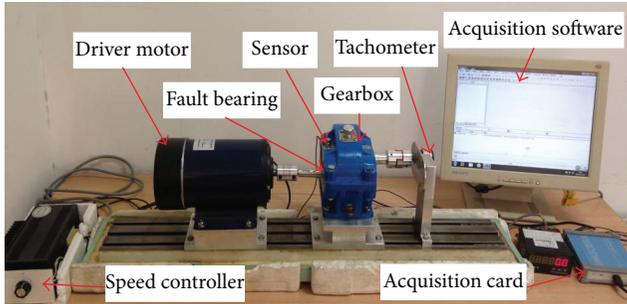


FIGURE 8: Layout of the test rig.

comparison, amplitude of each harmonic and the changed ratios of harmonics between the resampled signal and the denoising signal are listed in Table 5.

## 5. Conclusions

A novel method of rolling element bearing fault feature extraction based on IDMM and EMD under variable rotational speed and gear interferences conditions is proposed.

TABLE 5: Comparison of the envelope order spectrum.

Harmonic order	Amplitudes of the harmonics		Changed ratio
	Resampled signal	Denoising signal	
BRO		0.017	
2BRO		0.034	
FCO	0.065	0.092	41.5%
2FCO	0.054	0.072	33.3%
3FCO	0.045	0.064	42.2%
4FCO	0.035	0.049	40%

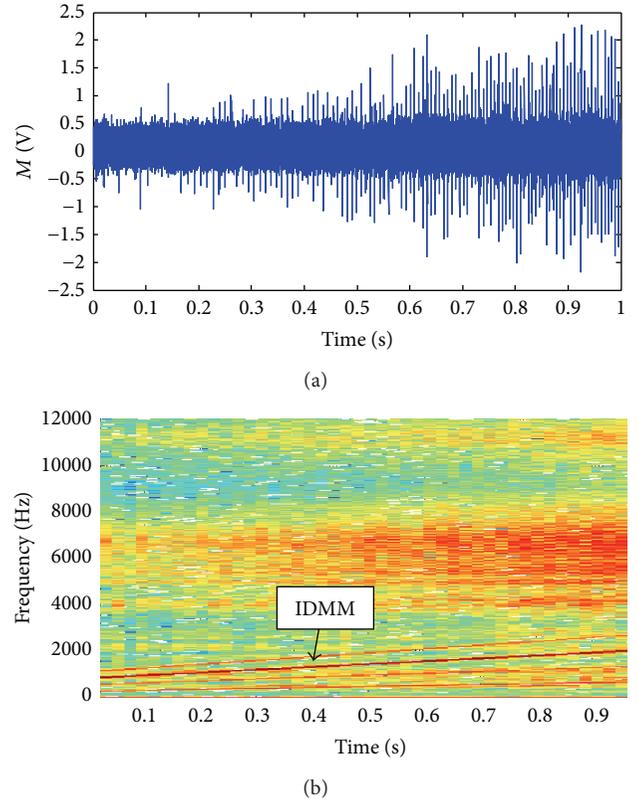


FIGURE 9: (a) Measured signal in time domain. (b) Time-frequency representation of the measured signal.

The major innovative features of the proposed method are reflected by the following:

- IDMM trend is used to substitute bearing rotational frequency, and the proposed method does not need auxiliary equipment to measure speed information.
- The EMD algorithm is applied to remove gear noise, and the results are satisfying.
- The proposed method avoids selecting complex parameters and simplifies the process of rolling element bearing diagnosis under time-varying rotational speed and gear interferences conditions.

Simulation and experiment analysis have demonstrated that the proposed method is effective in rolling element bearing fault diagnosis under time-varying rotational speed and

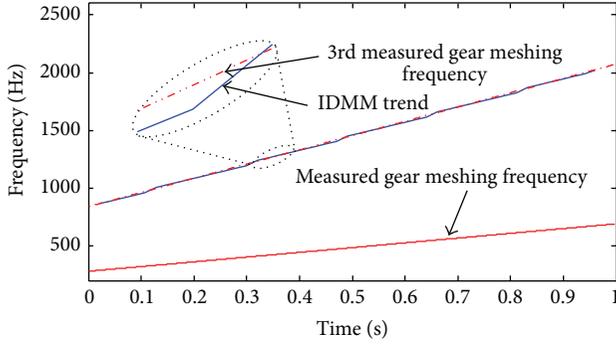


FIGURE 10: Extracted IDMM trend and measured gear meshing frequency.

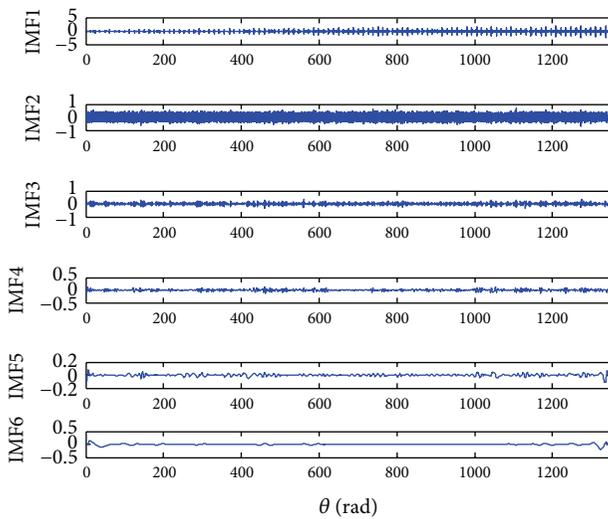


FIGURE 11: Decomposition result of the resampled signal.

gear noise. In addition, compared with the envelope order spectrum of the resampled signal, the BRO and the FCO in the envelope order spectrum obtained by the proposed method have higher amplitude.

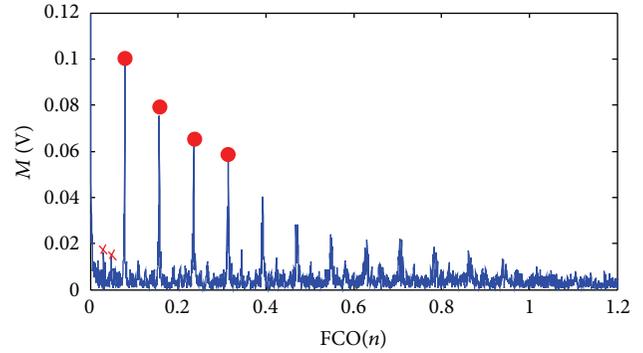
It should be pointed out that the proposed method is only suitable for some mechanism whose bearing rotational frequency is proportional to gear rotational frequency, such as gearbox.

### Conflict of Interests

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

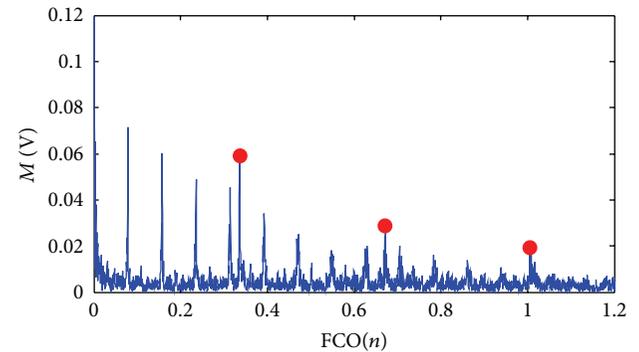
### Acknowledgment

This study is supported by the National Natural Science Foundation of China under Grant no. 51275030. The support is greatly appreciated.



× BRO  
● FCO

(a)



● Gear interference peak

(b)

FIGURE 12: (a) Envelope order spectrum of IMF1. (b) Envelope order spectrum of resampled signal.

### References

- [1] T.-Y. Wang, J.-Y. Li, and W.-D. Cheng, "Fault diagnosis of rolling bearing under a variable rotational speed and gear vibration noise based on revised ANC algorithm and FCO spectrum," *Journal of Vibration and Shock*, vol. 33, no. 18, pp. 7–13, 2014.
- [2] K. R. Fyfe and E. D. S. Munck, "Analysis of computed order tracking," *Mechanical Systems and Signal Processing*, vol. 11, no. 2, pp. 187–202, 1997.
- [3] W. Cheng, R. X. Gao, J. Wang, T. Wang, W. Wen, and J. Li, "Envelope deformation in computed order tracking and error in order analysis," *Mechanical Systems and Signal Processing*, vol. 48, no. 1-2, pp. 92–102, 2014.
- [4] P. N. Saavedra and C. G. Rodriguez, "Accurate assessment of computed order tracking," *Shock and Vibration*, vol. 13, no. 1, pp. 13–32, 2006.
- [5] J. Urbanek, T. Barszcz, and J. Antoni, "A two-step procedure for estimation of instantaneous rotational speed with large fluctuations," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 96–102, 2013.
- [6] G. Yu, Q. Shuren, T. Baoping, and et al, "Order tracking of rotating machinery based on instantaneous frequency estimation," *Chinese Journal of Mechanical Engineering*, vol. 39, no. 3, pp. 32–36, 2003.

- [7] P. D. McFadden and M. M. Toozhy, "Application of synchronous averaging to vibration monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 14, no. 6, pp. 891–906, 2000.
- [8] 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.
- [9] Y. Wang R, *Bearing fault detection and oil debris monitoring by adaptive noise cancellation [M.S. thesis]*, University of Ottawa, Ottawa, Canada, 2008.
- [10] D. Ho and R. B. Randall, "Effects of time delay, order of FIR filter and convergence factor on self-adaptive noise cancellation," in *Proceedings of the 5th International Congress on Sound and Vibration*, Adelaide, Australia, 1997.
- [11] P. Borghesani, P. Pennacchi, R. B. Randall, and R. Ricci, "Order tracking for discrete-random separation in variable speed conditions," *Mechanical Systems and Signal Processing*, vol. 30, pp. 1–22, 2012.
- [12] R. B. Randall and N. Sawalhi, "A new method for separating discrete components from a signal," *The Sound and Vibration*, vol. 45, no. 5, pp. 6–9, 2011.
- [13] P. Borghesani, R. Ricci, S. Chatterton, and P. Pennacchi, "A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 23–35, 2013.
- [14] W. Wang, "Autoregressive model-based diagnostics for gears and bearings," *Insight: Non-Destructive Testing and Condition Monitoring*, vol. 50, no. 8, pp. 414–418, 2008.
- [15] T. Wang, M. Liang, J. Li, W. Cheng, and C. Li, "Bearing fault diagnosis under unknown variable speed via gear noise cancellation and rotational order sideband identification," *Mechanical System and Signal Processing*, vol. 62–63, pp. 30–53, 2015.
- [16] T. Wang, M. Liang, J. Li, and W. Cheng, "Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis," *Mechanical Systems and Signal Processing*, vol. 45, no. 1, pp. 139–153, 2014.
- [17] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London Series A*, vol. 454, no. 1971, pp. 903–995, 1998.

## Research Article

# Two General Architectures for Intelligent Machine Performance Degradation Assessment

Yanwei Xu,<sup>1,2</sup> Aijun Xu,<sup>1</sup> and Tancheng Xie<sup>1</sup>

<sup>1</sup>*School of Mechatronics Engineering, Henan University of Science and Technology, Luoyang 471003, China*

<sup>2</sup>*School of Mechanical Engineering, Tianjin University, Tianjin 300072, China*

Correspondence should be addressed to Tancheng Xie; [xietc@mail.haust.edu.cn](mailto:xietc@mail.haust.edu.cn)

Received 5 June 2015; Revised 1 July 2015; Accepted 2 July 2015

Academic Editor: Dong Wang

Copyright © 2015 Yanwei 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.

Markov model is of good ability to infer random events whose likelihood depends on previous events. Based on this theory, hidden Markov model serves as an extension of Markov model to present an event from observations rather than states in Markov model. Moreover, due to successful applications in speech recognition, it attracts much attention in machine fault diagnosis. This paper presents two architectures for machine performance degradation assessment, which can be used to minimize machine downtime, reduce economic loss, and improve productivity. The major difference between the two architectures is whether historical data are available to build hidden Markov models. In case studies, bearing data as well as available historical data are used to demonstrate the effectiveness of the first architecture. Then, whole life gearbox data without historical data are employed to demonstrate the effectiveness of the second architecture. The results obtained from two case studies show that the presented architectures have good abilities for machine performance degradation assessment.

## 1. Introduction

Machine undergoes deterioration over time. In order to ensure its reliability, maintainability, and safety, timely condition monitoring and fault diagnosis are required. Temporal domain, frequency domain, and time-frequency domain based signal processing methods [1–3] are popular to detect fault related signatures for machine fault diagnosis. However, these methods need expertise. So, intelligent condition monitoring and fault diagnosis methods become attracting for identification of different machine health conditions [4].

Hidden Markov model (HMM) is a statistical model and it is successfully applied in speech recognition [5]. In machine condition monitoring and fault diagnosis, compared with Markov model, states used in HMM usually have no actual physic meaning and could not be directly observable. Nevertheless, state sequences in HMM can be estimated by a series of observation sequences, which may be features extracted from collected vibration signals. Bunks et al. [6] indicated that HMM is a very promising tool for constructing algorithms for condition based maintenance. Then, they

investigated case studies related to Westland helicopter gearbox data set to demonstrate the potential benefit of HMM. Ocak and Loparo [7] proposed a HMM based fault diagnosis scheme by selecting a HMM with the highest probability. For bearing wear assessment, Ocak et al. [8] used a HMM to build a normal bearing model and observed any deviations from a normal bearing health condition for bearing performance degradation assessment. Ertunc et al. [9] applied a HMM to track tool wear conditions during drilling operations. Miao et al. [10] used empirical mode decomposition to analyze gearbox vibration signals. Then, a HMM was used for gearbox performance degradation assessment. Wang et al. [11] used two HMMs for early gear fault diagnosis and its performance degradation assessment, respectively.

In this paper, we present two architectures for machine performance degradation assessment by considering two different kinds of situations, with or without historical data. Machine performance degradation assessment aims to evaluate the deviation of the current health condition from the normal health condition and different fault levels. When historical data are available, they can be used to train

different HMMs. Therefore, different health conditions and fault levels can be identified by finding the maximum log likelihood of the HMMs. Based on these HMMs, a slight fault can be initially identified. Then, based on this kind of fault pattern, different fault levels can be identified through multiple HMMs. On the other hand, in the case of unavailable historical data, an abnormal condition can only be inferred from any abnormal deviation of HMM from a normal condition. But it should be noted that the deviation can only be used to describe machine performance degradation trend rather than different performance degradation levels. Nevertheless, in some cases, some critical components, such as gears, may be directly replaced with new ones once an abnormal health condition is detected.

The rest of this paper is given as follows. Section 2 presents two architectures for machine performance degradation assessment under the situations with or without available historical data. In Section 3, two case studies are conducted to demonstrate the effectiveness of the presented architectures and to illustrate how they work. Conclusions are drawn at last.

## 2. Two Architectures for Intelligent Machine Performance Degradation Assessment

In this section, by considering two situations with or without historical data, two architectures are built by hidden Markov models for machine performance degradation assessment. One is an underlying state stochastic process, which can only be estimated by another observation stochastic process. A HMM can be expressed by a compact notation  $\lambda = \{A, B, \pi\}$  and is introduced by the following [5]:

- (a)  $N$  is the number of hidden states in the hidden Markov model. Denote individual states by  $S = \{S_1, S_2, \dots, S_N\}$  and the state at time  $t$  by  $q_t$ .
- (b)  $M$  is the number of distinct observation symbols per state. The observation symbols correspond to the physical output of the system being modeled, such as features extracted from vibration signals. Denote the individual symbols by  $V = \{V_1, V_2, \dots, V_M\}$ .
- (c) The state transition probability distribution  $A = \{a_{ij}\}$ , where

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i] \quad 1 \leq i, j \leq N. \quad (1)$$

- (d) The observation symbol probability distribution in state  $j$ ,  $B = \{b_j(k)\}$ , where

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j] \quad 1 \leq j \leq N, 1 \leq k \leq M. \quad (2)$$

If the observation is modeled as continuous, a continuous probability density function must be specified for each state. In general, a weighted sum of several Gaussian distributions is employed for continuous probability density function.

- (e) The initial state distribution  $\{\pi = \pi_i\}$ , where

$$\pi_i = P[q_1 = S_i] \quad 1 \leq i \leq N. \quad (3)$$

When HMM is applied, three problems should be solved. Firstly, how to evaluate a set of observations for a given hidden Markov model? This evaluation problem can be solved by Forward-Backward algorithm [5]. Secondly, how to find the optimal state sequence for a given set of observations? This decoding problem can be solved by Viterbi algorithm [5]. At last, how to find the optimal parameters of HMM for a given set of observations? The optimal parameters of HMM can be found by Baum-Welch algorithm [5].

*2.1. An Architecture for Machine Performance Degradation Assessment with Historical Data.* When there are historical data available, each machine health condition can be represented by a HMM. So, it is easy to distinguish different initial faults by selecting maximum log likelihood because HMM with maximum log likelihood indicates the current health condition of the machine. Here, taking different initial bearing faults, for example, bearings are prone to suffer from an inner race defect, an outer race defect, and a roller race defect. So, four HMMs are built to distinguish different bearing health conditions as shown in Figure 1(a). Once an abnormal health condition is detected, different fault levels can be estimated through Figure 1(b). In Figure 1(b), historical data with different fault levels are used to build HMMs. Therefore, HMM with maximum log likelihood can be used to indicate the current fault level. Machine performance degradation can be clearly assessed by the above procedure.

*2.2. An Architecture for Machine Performance Degradation Assessment without Historical Data.* When there are no fault historical data available, it is difficult to know which fault occurs firstly. But we can assess any log likelihood deviation from a normal condition as shown in Figure 2. In other words, only one HMM can be trained by normal data. Besides, only degradation trend rather than different fault levels can be assessed to judge how far the current health condition deviates from a normal condition. If a proper threshold is set, an abnormal condition can be detected and maintenance decision-making can be made. The proper threshold is usually set by expertise and users' requirement. The loop termination in Figures 1 and 2 is determined by users' requirements. It means that performance degradation levels are decided by users.

## 3. Case Studies

*3.1. A Case Study with Historical Data.* In this section, vibration data collected from a motor were used for our analyses. The sampling frequency was set to 12 kHz [12]. Single point faults including a fault diameter of 0.007 inches, 0.014 inches, 0.021 inches, and 0.028 inches were, respectively, introduced to normal bearings using electrodischarge machining. For observation sequence extraction, discrete wavelet packet transform was employed to obtain wavelet packet node energy [8]. But here each wavelet packet node energy was scaled by a log algorithm for simplifying parameter estimation. Moreover, in order to enhance training sequence length, each original signal was divided into  $L$  equal length

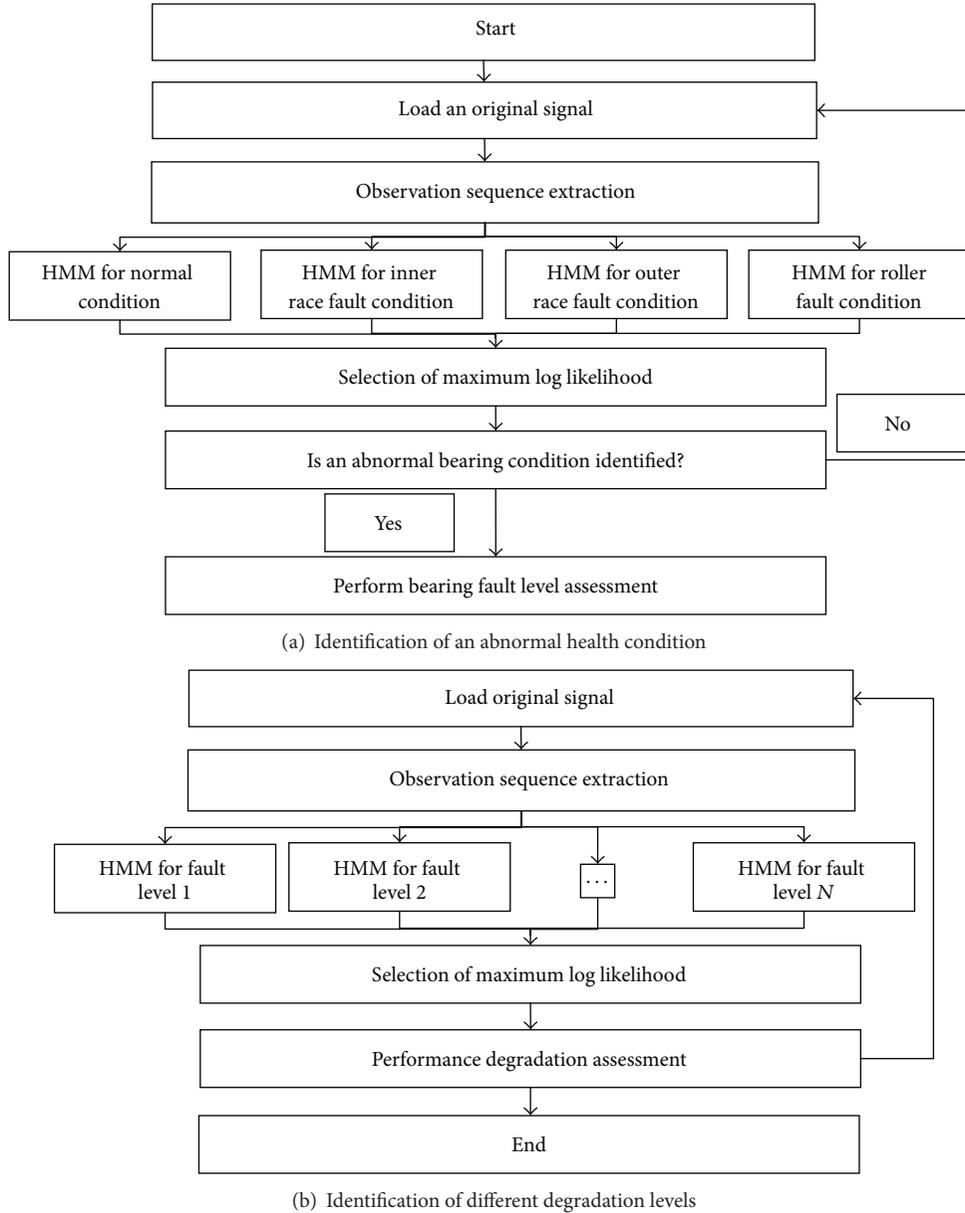


FIGURE 1: The presented flowchart for bearing performance degradation assessment in the case of available historical data.

subsignals before the use of discrete wavelet transform. In this paper,  $L$  was set to 12 and bearing vibration signals with a length of 10 seconds were used. So, each subsignal contains 10000 samples. Daubechies-9 wavelet was chosen and wavelet packet decomposition level was 4. The major reason for selection of Daubechies-9 is that it has been proven to be effective in diagnosing localized bearing faults. In all, the size of the observation feature matrix is  $12 \times 16$ . Assume that hidden states are governed by a homogeneous Markov chain of order 1. HMM used here is a 5-state model with a diagonal covariance matrix containing 3 Gaussian mixtures. Then, firstly, in order to distinguish different bearing faults, four hidden Markov models are trained by normal bearing data, inner race fault data with a fault diameter of 0.007

inches, outer race fault data with a fault diameter of 0.007 inches, and roller race fault data with a fault diameter of 0.007 inches. Moreover, they are denoted by  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , and  $\lambda_4$ . Then, inner race fault data with a fault diameter of 0.007 inches is used for validation data. The diagnostic result is shown in Figure 3(a), where it is obvious that  $\lambda_2$  has the maximum log likelihood among all HMMs  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , and  $\lambda_4$ . The result demonstrates that the HMMs are effective in identification of different initial defects.

Further, in order to assess bearing degradation, inner race fault with fault diameters of 0.007 inches, 0.014 inches, 0.021 inches, and 0.028 inches is used for further investigation. So, the size of the observation feature matrix is  $48 \times 16$ . Here, four HMMs are trained by inner race fault data with different fault

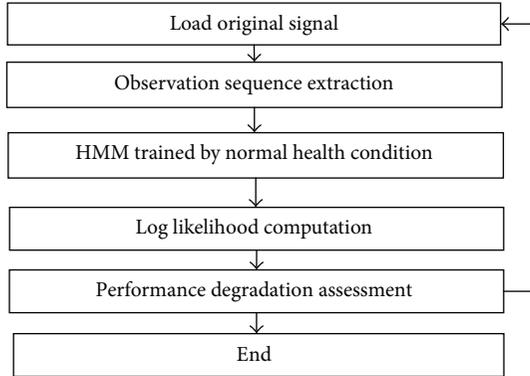


FIGURE 2: The presented flowchart for machine performance degradation assessment in the case of unavailable historical data.

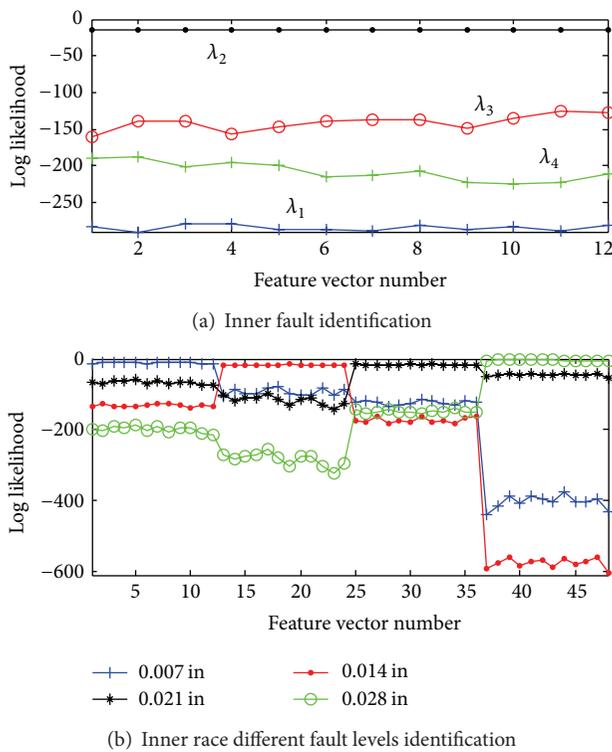


FIGURE 3: Bearing performance degradation assessment using presented architecture 1.

levels. The results are shown in Figure 3(b) using architecture 1. It is clear to see that the bearing with different inner fault levels can be well assessed by selecting the maximum log likelihood at different bearing inner race degradation stages.

**3.2. A Case Study without Historical Data.** In this section, a set of vibration signals collected from a run to failure gearbox were used to validate the second architecture in the case of unavailable historical data. The sampling rate was set to 20 kHz. The gear in a normal health condition ran to failure. A total of 148 data files were collected. Finally, the gear was subject to distributed faults. For observation sequence extraction, the same procedure was conducted as

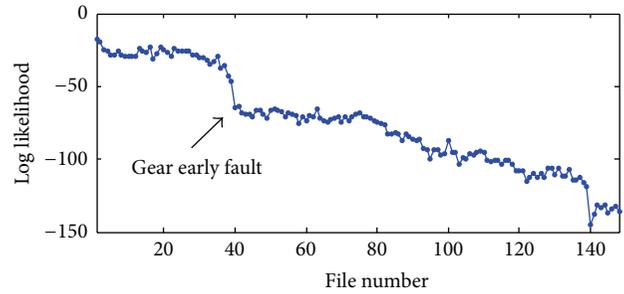


FIGURE 4: Gear performance degradation assessment using presented architecture 2.

mentioned step in Section 3.1.  $L$  was set to 10 and gearbox vibration signals with a length of 10 seconds are used. So, each subsignal contains 20000 samples. Daubechies-9 wavelet was chosen and wavelet packet decomposition level was 4. In all, the size of the observation feature matrix is  $10 \times 16$  at each inspection. There is only one hidden Markov model trained by a normal condition. Then, the observation feature matrices at other inspection times serve as the input to the HMM trained by the normal health condition. The gear performance degradation trend can be observed by any deviations from the normal log likelihood. For slightly normal wear, the trend keeps decreasing. For an abnormal health condition, there is a sudden change indicated in Figure 4, which can be regarded as early gear faults because of its abnormality. Without historical data, only gear degradation trend can be observed and specific gear degradation levels are unknown.

## 4. Conclusions

In this paper, two architectures were presented for machine performance degradation assessment, by considering two kinds of situations, with or without historical data. For available historical data, different HMMs could be built to respond to different health conditions, including different faults and their associated fault levels. So, different initial fault patterns could be identified firstly. Then, their associated fault levels could be further assessed. On the other hand, for unavailable historical data, only normal data could be employed to construct one HMM that represents a normal condition. Any deviations from the normal condition could be observed. If a large sudden change occurs, it indicates the occurrence of the abnormal condition. However, it should be pointed out that the HMM without historical fault data can only be used to describe the performance degradation trend rather than specific performance degradation levels.

## Conflict of Interests

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

## Acknowledgments

The authors are grateful for the financial support of the National Natural Science Foundation of China (51305127) and the scientific research key project fund of the Education Department of Henan Province of China (14A460018). The authors would like to thank Professor K. A. Loparo for providing the bearing data.

## References

- [1] X. Zhang, J. Kang, L. Xiao, J. Zhao, and H. Teng, "A new improved Kurtogram and its application to bearing fault diagnosis," *Shock and Vibration*, vol. 2015, Article ID 385412, 22 pages, 2015.
- [2] D. Wang, P. W. Tse, and K. L. Tsui, "An enhanced Kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 176–199, 2013.
- [3] D. Wang, P. W. Tse, and Y. L. Tse, "A morphogram with the optimal selection of parameters used in morphological analysis for enhancing the ability in bearing fault diagnosis," *Measurement Science and Technology*, vol. 23, no. 6, Article ID 065001, 2012.
- [4] C. Shen, D. Wang, F. Kong, and P. W. Tse, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier," *Measurement*, vol. 46, no. 4, pp. 1551–1564, 2013.
- [5] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [6] C. Bunks, D. McCarthy, and T. Al-Ani, "Condition-based maintenance of machines using hidden Markov models," *Mechanical Systems and Signal Processing*, vol. 14, no. 4, pp. 597–612, 2000.
- [7] H. Ocak and K. A. Loparo, "HMM-based fault detection and diagnosis scheme for rolling element bearings," *Journal of Vibration and Acoustics*, vol. 127, no. 4, pp. 299–306, 2005.
- [8] 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.
- [9] H. M. Ertunc, K. A. Loparo, and H. Ocak, "Tool wear condition monitoring in drilling operations using hidden Markov models," *International Journal of Machine Tools and Manufacture*, vol. 41, no. 9, pp. 1363–1384, 2001.
- [10] Q. Miao, D. Wang, and M. Pecht, "A probabilistic description scheme for rotating machinery health evaluation," *Journal of Mechanical Science and Technology*, vol. 24, no. 12, pp. 2421–2430, 2010.
- [11] D. Wang, Q. Miao, Q. Zhou, and G. Zhou, "An intelligent prognostic system for gear performance degradation assessment and remaining useful life estimation," *Transactions of The ASME—Journal of Vibration and Acoustics*, vol. 137, no. 2, Article ID 021004, 2015.
- [12] CWRU, Bearing Data Center, seeded fault test data, <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.

## Research Article

# In Situ Measurement of Seeking Speed and Seeking Induced Head-Disk Interface Instability in Hard Disk Drives

Yu Wang,<sup>1,2</sup> Xiongfei Wei,<sup>3</sup> Yanyang Zi,<sup>1</sup> and Kwok-Leung Tsui<sup>2</sup>

<sup>1</sup>*School of Mechanical Engineering, State Key Laboratory for Manufacturing and Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China*

<sup>2</sup>*Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong*

<sup>3</sup>*SAE Technologies Development (Dongguan) Co. Ltd., Dongguan 523087, China*

Correspondence should be addressed to Yu Wang; [ywang95@mail.xjtu.edu.cn](mailto:ywang95@mail.xjtu.edu.cn)

Received 29 April 2015; Accepted 28 June 2015

Academic Editor: Chuan Li

Copyright © 2015 Yu 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.

This paper investigated the instability of head-disk interface caused by the voice coil motor (VCM) end crashing the crash stop during the seeking of magnetic head. To make the whole process of that clear, an in situ measurement method based on maximum likelihood estimation and extended Kalman filter for seeking speed at component level was developed first and was then calibrated by a high speed camera. Given a crash between VCM end and crash stop that may be a consequence of the continuous increasing seeking speed, the seeking speed was carefully controlled by using our developed method to find a critical value that may induce vigorous head-disk interface instability. Acoustic emission sensor and laser Doppler vibrometer were used to capture the transient dynamic behaviors of magnetic head when the crash is happening. Damage analysis and mode identification were carried out to reveal the relationship between the damage patterns on disk surface and head dynamics. The results of this study are helpful to optimize the track seeking profile during the HDD operation, as well as the design of components such as head and head arm.

## 1. Introduction

The areal density of the hard disk drive (HDD) is expected to reach multiseveral Tb/in<sup>2</sup> in next few years [1, 2]. At such high density, the data transfer needs to further speed up due to the competition pressure from the solid-state drive (SSD) and the requirement of users. The access time or response time of a rotating drive that measures the time of HDD components' response before the drive transferring data is mainly related to the mechanical nature of the rotating disks and moving heads. However, due to the limitations of mechanical components in HDDs, low-speed motion has become one of the chief obstacles for further increasing the speed of data transfer [3, 4]. Track following and seeking is a key step in the loop of data transfer in HDDs, and thus increasing seeking speed is raised as a critical task in HDDs.

Generally, head seeking is implemented by inputting a digital current into the voice coil motor (VCM) to drive the head arm assembly move. To provide a specific speed, an available method is to add a constant current to VCM for

a prescribed amount of time. However, this method may not achieve the precise speed control due to the disturbances, such as friction of the pivot bearings and the resistance force of data flex cable [5, 6]. An alternative method taking advantage of back electromotive force (EMF) was proposed to create a negative acceleration in order to produce the VCM velocity which is widely adopted in current HDDs [7–9]. The mechanism of this method is that a back EMF representing the velocity of the VCM will be generated when the power is cut off. The seeking speed can be attained by adjusting the back EMF.

For a safety consideration, two crash stops were designed to limit the rotation of the VCM to prevent the head from moving off of the disk or hitting the spindle during head seeking [10]. No matter which manner used for VCM control, when the seeking speed is too high, it has a risk that the end of VCM may crash the crash stop. Such a crash may induce the HDD instability such as head slap on disk and head/disk damage. Although this topic is essential, there is rarely report on that in literature. It is therefore necessary to have a better

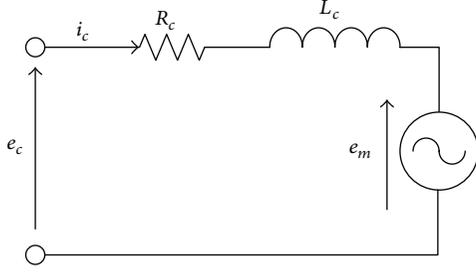


FIGURE 1: Simplified electronic circuit of a typical VCM.

understanding for the whole process of that, especially at component level, which is generally followed by system integration and thus is highly related to the component design and the yielding of the final products.

Given the above considerations, this paper investigated the instability of HDI caused by the VCM crashing the crash stop during head seeking at the component level. To investigate the effect of seeking speed on anti-impact performance of head-disk interface, a maximum likelihood estimation (MLE) and extended Kalman filter (EKF), MLE-EKF, method was developed to estimate and control the speed flexibly. The EKF method is an improved method of the Kalman filter that is incapable of handling the nonlinear problems [11, 12]. Then, different speeds were carefully controlled to find a critical value that may induce HDI instability. Acoustic emission (AE) sensor and laser Doppler vibrometer (LDV) were used to capture the transient dynamic behaviors of head when the crash is happening. Damage analysis and mode identification were carried out to reveal the relationship between the damage patterns on disk surface and head dynamics.

## 2. Design and Calibration of Seeking Speed

**2.1. Mathematic Model.** Before implementing the experiment, the seeking speed of head should be well controlled and calibrated. To this end, a drive-based tester was used to implement the seeking test. Head seeking controlling method by the back EMF-based control technology applies the back EMF that is generated by coil movement in the magnetic field to drive the VCM to move. A simplified electronic circuit is shown in Figure 1, and the electrical dynamics of the VCM can be described as [13]

$$L_c \frac{di_c(t)}{dt} + R_c i_c(t) + e_m(t) = e_c(t), \quad (1)$$

where  $L_c$  is the inductance of coil,  $R_c$  is the resistance of coil,  $e_c$  is applied voltage, and  $e_m$  is the induced back EMF in the coil due to the permanent magnets. According to Lenz's law, the running of VCM generates a back EMF proportional to rotational velocity. Therefore, the back EMF can be calculated by  $e_m = K_e \omega$ , and  $K_e$  is the back EMF constant.

According to the mechanical dynamic model of the VCM, the current can be transferred to the rotation speed by [14, 15]

$$K_t i_c(t) - T_f - K_c \cdot x = J \dot{\omega}(t), \quad (2)$$

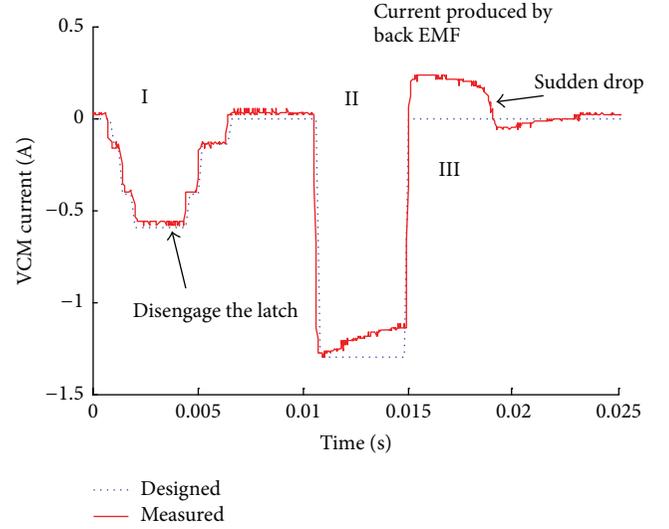


FIGURE 2: The measured VCM current profiles during seeking.

where  $K_t$  is the torque constant,  $i_c$  is the coil current,  $J$  is the inertia of head arm,  $T_f$  is the torque generated by pivot friction force,  $K_c$  is the stiffness of flex cable,  $x$  is the displacement, and  $\omega$  is the angular velocity. The values of  $K_t$  and  $J$  used in this experiment are 0.026 Nm/A and  $0.1309 \times 10^{-5} \text{ kgm}^2$ , respectively. By integrating  $\dot{\omega}$ , the rotation speed of arm can be obtained. Given the arm length  $l$ , the head seeking speed can be calculated. It is therefore necessary to know the implemented current and estimate the key parameters including  $K_t$ ,  $T_f$ , and  $J$  before identifying the seeking speed.

A programmable power source was used to control the current of VCM so as to control the seeking speed. Figure 2 shows the designed waveform of VCM current and the measured current profile by an oscilloscope. The measured current profile consists of three phases. Phase I provided the power for VCM to disengage the latch. Phase II generated the seeking speed. Phase III was produced by the back EMF. It can be observed at phase III that a sudden drop occurred at 0.018 s, implying that a crash happened. A more detailed explanation was given in Section 3. When the current turned off, the rotational speed of VCM can be directly determined by back EMF. Using this property, the unknown parameters  $T_f$  and  $K_c$  can be estimated accordingly.

**2.2. Parameter Estimation.** To estimate the parameters in (2), it is necessary to reconsider the mechanics of VCM. When the VCM current is cut off, (2) becomes

$$T_f + K_c \cdot x = -J \dot{\omega}(t), \quad (3)$$

assuming that the displacement of flex cable has a linear relationship with the duration

$$x = a + bt. \quad (4)$$

By integrating  $\dot{\omega}(t)$ , the angular speed of VCM can be calculated by

$$\hat{\omega} = \omega_0 - \frac{1}{J} \left[ (T_f + K_c a) t - \frac{1}{2} K_c b t^2 \right]. \quad (5)$$

For the sake of simplicity, (5) can be simplified by combining the parameters, that is, letting  $C_1 = (1/J)[T_f + K_c a]$ ,  $C_2 = K_c b/J$ , and (5) become

$$\hat{\omega} = \omega_0 - C_1 t - \frac{1}{2} C_2 t^2. \quad (6)$$

Due to the disturbance, the measurement always accompanied with measurement error. The measurement error can be regarded as an independent and identically distributed normal random variable. Given the measurement error  $\varepsilon \sim N(0, \sigma^2)$ , the angular speed can be written as

$$\omega = \hat{\omega} + \varepsilon. \quad (7)$$

The angular speed follows  $\omega \sim N(\hat{\omega}, \sigma^2)$ . To estimate the parameters  $C_1$ ,  $C_2$ , and  $\sigma$ , the maximum likelihood estimation method was used, which then turns to the logarithmic likelihood function of the angular speed:

$$\log L = \sum_{i=1}^m \left( \frac{n_i}{2} (-\ln(2\pi) - 2 \ln \sigma) - \frac{\sum_{k=1}^{n_i} (\omega_i - \omega_0 + C_1 t + (1/2) C_2 t^2)^2}{2\sigma^2} \right). \quad (8)$$

By maximizing the log-likelihood function, the parameters  $C_1$ ,  $C_2$ , and  $\sigma$  can be obtained.

Three VCM current waveforms were used to build the likelihood function, as illustrated in Figure 3. By maximizing the log-likelihood function (8), the estimated parameters are  $C_1 = 245.9$ ,  $C_2 = 0$ , and  $\sigma = 0.54$ . The seeking speed can be calculated by integrating  $\dot{\omega}$  (when VCM current was supplied) or by  $e_m/K_e$  (when VCM current was cut off).

**2.3. Real-Time Speed Estimation.** We treated these estimated parameters as the initial parameters for an EKF; a real-time VCM speed can be obtained through the prediction and update steps in the EKF. We assume the state of the EKF is  $x = [C_1 \ C_2 \ \omega_0]$ . According to the mechanical dynamic model of the VCM, the state and measurement function of the EKF can be expressed as follows.

State model is

$$x_k = f(x_{k-1|k-1}, k-1) + q_{k-1},$$

$$f(x_{k-1|k-1}, k-1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C_{1,k-1} \\ C_{2,k-1} \\ \omega_{0,k-1} \end{bmatrix}, \quad (9)$$

where  $q_{k-1} = [v_{C_1} \ v_{C_2} \ v_{\omega_0}]$ ,  $v_{C_1} \sim N(0, \sigma_{C_1})$ ,  $v_{C_2} \sim N(0, \sigma_{C_2})$ , and  $v_{\omega_0} \sim N(0, \sigma_{\omega_0})$ .

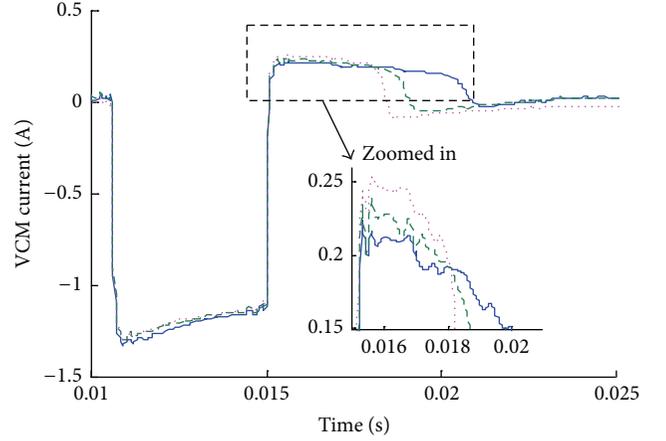


FIGURE 3: Three VCM current waveforms to generate different seeking speed.

Measurement model is

$$\omega_k = h(x_k, k) + v,$$

$$h(x_k, k) = \omega_{0,k-1} - C_{1,k} t - \frac{1}{2} C_{2,k} t^2, \quad (10)$$

where  $v \sim N(0, \sigma_y)$ .

Based on this state-space function, the first order EKF can be used to smooth the raw data. The steps for the first order EKF are as follows [12].

Predict

$$x_{k|k-1} = f(x_{k-1|k-1}, k-1),$$

$$P_{k|k-1} = F_x(x_{k-1|k-1}, k-1) P_{k-1|k-1} F_x^T(x_{k-1|k-1}, k-1) + Q_{k-1}, \quad (11)$$

where  $F_x$  is defined to be  $\partial f / \partial x$ .

Update

$$v_{k|k} = \omega_k - h(x_k, k),$$

$$S_{k|k} = H_x(x_{k-1|k-1}, k-1) P_{k-1|k-1} H_x^T(x_{k-1|k-1}, k-1) + R_k, \quad (12)$$

$$K_k = P_{k|k-1} H_x^T(x_{k-1|k-1}, k-1) S_{k|k}^{-1},$$

$$x_{k|k} = x_{k|k-1} + K_k v_{k|k},$$

$$P_{k|k} = P_{k|k-1} - K_k S_{k|k} K_k^T,$$

where  $H_x$  is defined to be  $\partial h / \partial x$ .

To illustrate the effectiveness of the proposed method, the EMF induced speed was estimated by EKF. As shown in Figure 4, the estimated speed follows the measured speed well

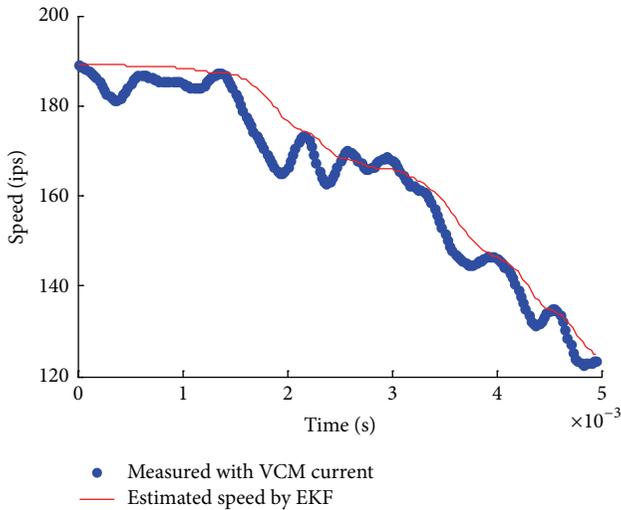


FIGURE 4: Real-time estimation of the head speed by EKF.

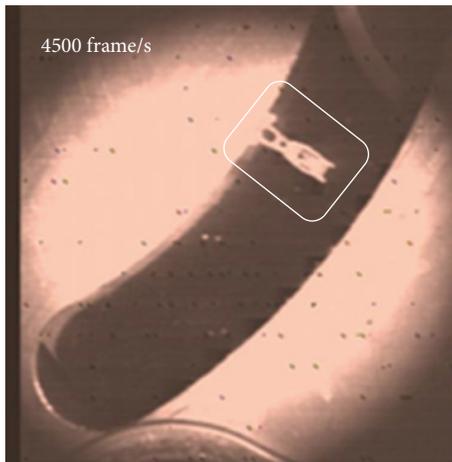


FIGURE 5: The pattern captured by high speed camera.

as the maximum relative error is 8%. This indicates that our proposed speed estimation method is valid for estimating the head speed.

**2.4. Speed Calibration by High Speed Camera.** To verify the aforementioned calculated seeking speed, a high speed camera taking the snapshots for head arm with a speed of 4500 frames per second was used. The head arm pattern at any location, as shown in Figure 5, was continuously captured by the camera.

Using a National Instruments image identification software to measure the angular displacement  $\Delta\theta$  of the featured pattern of head arm at a small time interval  $\Delta t$  during seeking process, the speed can be calculated by  $\Delta\theta/\Delta t$ .

A correlation between our estimated speed and the measured speed by camera was conducted, as illustrated by Figure 6. As observed from this figure, a good correlation exists between the estimated and the measured results as

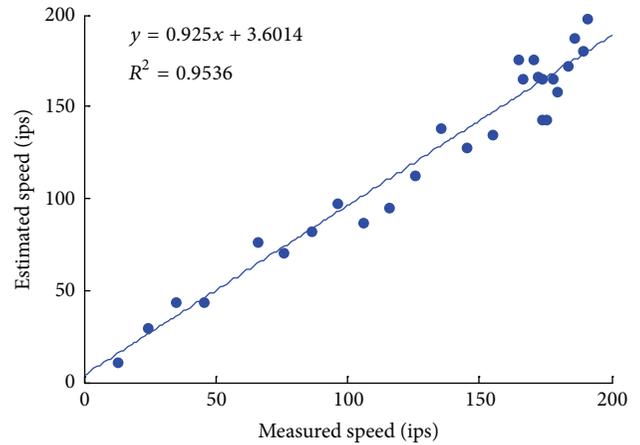


FIGURE 6: The seeking speed calibrated by high speed camera.

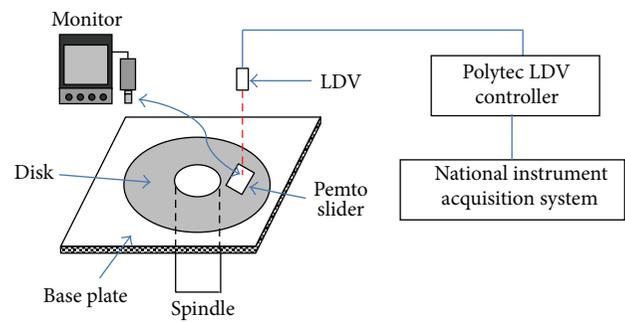


FIGURE 7: The schematic of the experiment setup.

$R^2$  is approximately 0.95. This means the seeking speed can be well quantified by our estimated speed.

### 3. Experimental Setup for Crashing Characterization

Based on back EMF control technology, a component-level evaluation system for seeking was developed. The schematic of the experimental configuration is shown in Figure 7. An acoustic emission sensor was mounted on the top head arm to monitor the arm dynamics. A laser Doppler vibrometer (LDV) was used to detect the head-disk contact and monitor the transient head dynamics when a crash is happening. In this experiment, a typical Pemto slider ( $1.235 \times 0.7 \times 0.23 \text{ mm}^3$ ) covered by a 2 nm diamond-like carbon (DLC) film was tested. The nominal flying attitude was 10 nm. The tested disk was a carbon coated 3.5-inch disk platter with a rotation speed of 5400 rpm. Signals were collected by an oscilloscope and a spectrum analyzer. The disk damage was inspected by an optical surface analyzer (OSA).

Figure 8 shows the signals monitored during the whole seeking process from inner diameter (ID) to outer diameter (OD) and then to ID. As observed from this figure, at phase III that the VCM current experienced, a vigorous vibration that occurred after the maximum back EMF was just generated. This vigorous vibration indicates a crash happening. Due to

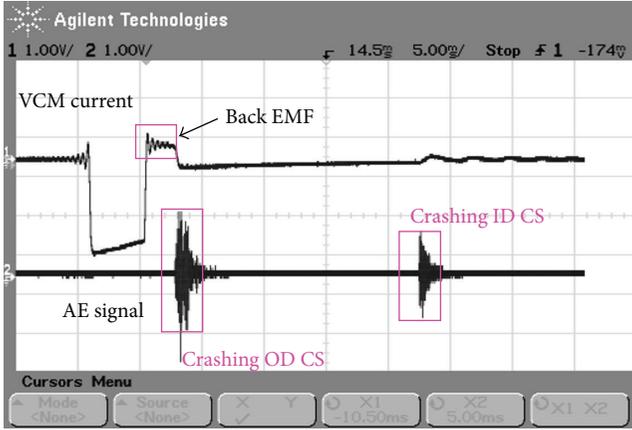


FIGURE 8: The monitored signals during seeking from inner diameter (ID) to outer diameter (OD), and then to ID. CS is the abbreviation of crash stop.

the crash, the seeking speed has a significant loss, leading to a sudden drop of the back EMF. By calculating the displacement that the head moved, this vigorous vibration can be identified as the crashing point between VCM end and OD crash stop. A very small response was observed at ID crash stop, indicating that the seeking speed has a significant loss after the crashing at OD crash stop.

#### 4. The Effect of Seeking Speed on HDD Instability

**4.1. Seeking Speed and Crashing Energy.** The crash between VCM end and crash stop may induce a vigorous vibration of head and arm and further result in HDD instability such as head slap on disk and head/disk damage depending on the crashing energy. Seeking speed is the main cause of the crashing energy. To reveal the relationship between speed and crashing energy, crashing energy characterized by AE signal under different seeking speeds was investigated, as illustrated by Figure 9. It can be observed that, with the increase of seeking speed during crashing, the crash-induced vibration increases quickly.

To describe the crashing energy, the root mean squares (RMS) of AE signals were calculated under different crashing speeds. A clear increasing trend can be observed in Figure 10. From around 160 ips, the energy has begun to increase sharply, which indicated that HDD may become unstable after this critical value. Further increasing the speed may lead to head/disk damage.

**4.2. Head Dynamics.** To monitor the head transient dynamics when crashing just happened, the laser spot of LDV was fixed on the position of OD crash stop to capture the vibration of head trailing edge center when head passed OD position. During the experiment, the laser will scan the trailing edge of head back when the head passed. Figure 11 shows the laser scanned area on the head back when head passed OD.

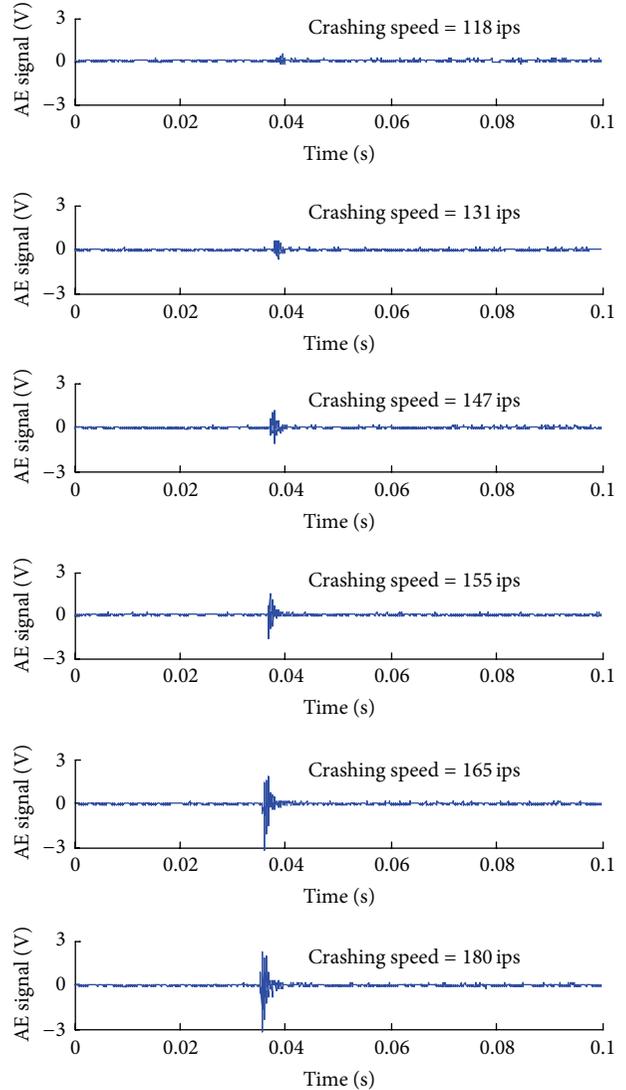


FIGURE 9: AE signals under different crashing speeds. ips is the abbreviation of inch per second.

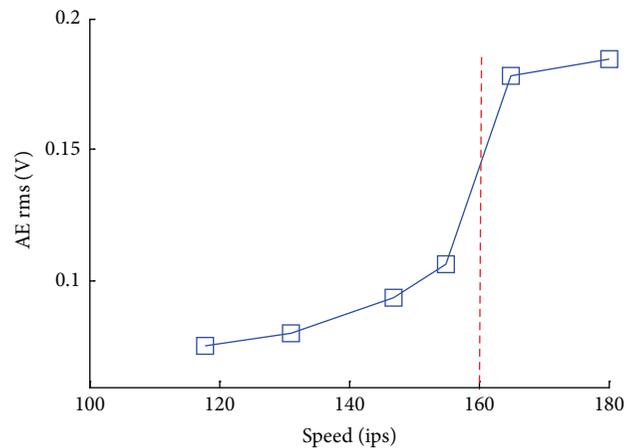


FIGURE 10: The energy generated by crashing under different seeking speeds.

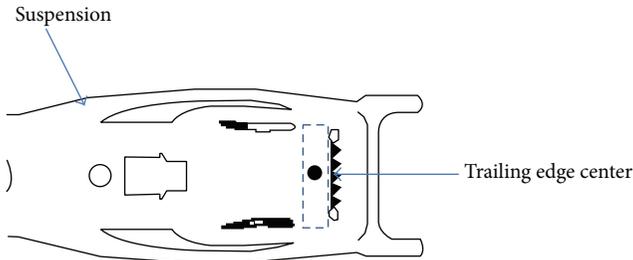


FIGURE 11: Laser scanned area on head when it passed to OD position.

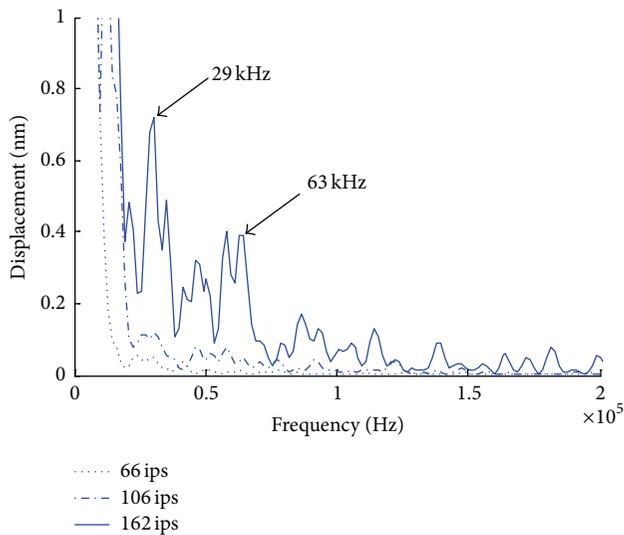


FIGURE 12: The frequency spectrum of LDV signal during crashing at OD crash stop.

A spectrum analyzer was used to acquire and analyze the laser signals with the cutoff of the frequency range of 10 kHz–200 kHz. Three different crashing speeds were conducted in order to reveal the relationship between the head dynamics and speed, as illustrated in Figure 12. As observed from this figure, when the crashing speed reached 162 ips, a large head vibration happened, while at 66 ips and 106 ips, there is nearly no significant response. This suggests that, at 162 ips, the air bearing may collapse and leads to HDI instability. The frequency spectrum of head response at 162 ips shows that the dominant modes are at 29 kHz and 63 kHz. The responses of these two modes at 162 ips are about 10 times bigger than that of the other two speeds.

To reveal the dynamic nature of the head during crashing, a finite element model of whole head arm was built to identify the mode shapes of the two modes. The air bearing was modeled by four springs whose stiffness was identified by a parameter identification code [16]. The finite element model and the constraints are shown in Figure 13. Both modal and harmonic analysis were carried out using the finite element model. The mode shape of head under a lateral excitation (a unit force on coil end) showed that a torsion mode of head gimbal and the rolling mode of the air

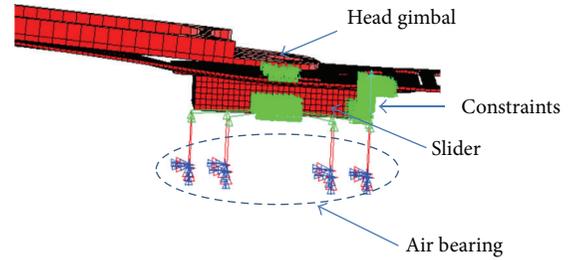


FIGURE 13: The finite element model of head arm.

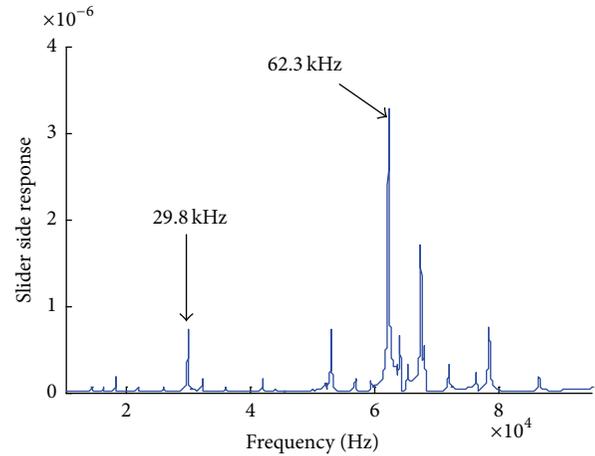


FIGURE 14: Head side response with a lateral excitation.

bearing dominated the head side response. The frequencies of these two modes (29.8 kHz and 62.3 kHz), as illustrated in Figure 14, are in accordance with the measured frequencies by LDV, respectively.

**4.3. Damage Inspection of Disk Surface.** After the experiment, the tested disks were inspected by OSA. Figure 15 shows the tested disk at crashing speed of 162 ips. As seen from this figure, some special patterns (damage) were observed, providing the evidence of air bearing collapse. The width of the patterns is about 0.7 mm, which is equivalent to the head width, indicating that the head rolling motion dominates the head vibration when the arm end is crashing crash stop. This result is in good accordance with the mode shape that was identified by finite element model.

## 5. Conclusion

This paper investigated the effect of seeking speed on HDD instability. The head seeking was controlled by a back EMF-based technology. A mathematical method based on MLE-EKF for deriving the seeking speed was developed. AE sensor and LDV were used to characterize the crashing energy and capture the head transient dynamics at the different speeds, respectively. With the help of finite element model and OSA inspection, the rolling mode of air bearing was found to dominate the head dynamics. This mode will lead to a serious damage of disk when a high seeking speed was adopted.

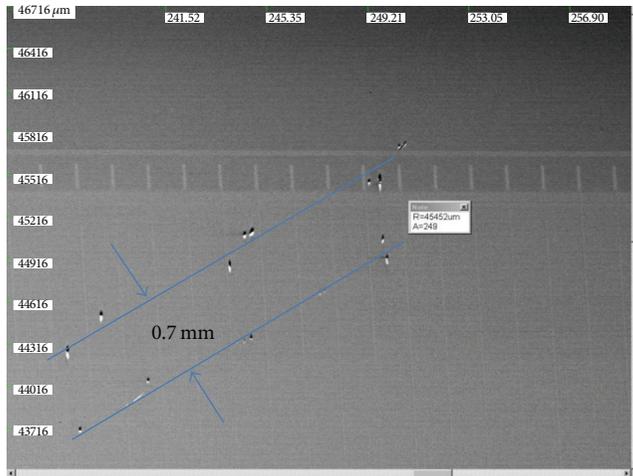


FIGURE 15: The OSA inspection result of disk after crashing test (crashing speed = 162 ips).

The results revealed in this paper will help to optimize the drive-level track seeking profile, as well as the design of components such as head and head arm.

### Conflict of Interests

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

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (51405380), China Postdoctoral Science Foundation (2014M560765), Hong Kong Innovation Technology Fund (GHX/003/12), and Hong Kong RGC GRF (121410).

### References

- [1] B. Marchon, T. Pitchford, Y.-T. Hsia, and S. Gangopadhyay, "The head-disk interface roadmap to an areal density of 4 Tbit/in<sup>2</sup>," *Advances in Tribology*, vol. 2013, Article ID 521086, 8 pages, 2013.
- [2] V. Venkataramanan, G. Feng, and B. Gaber, "Track following servo control for rotary piezoelectric motor based primary actuation in hard disk drives," *Microsystem Technologies*, vol. 18, no. 9-10, pp. 1751-1757, 2012.
- [3] B.-H. Chang and Y. Hori, "Trajectory design considering derivative of jerk for head-positioning of disk drive system with mechanical vibration," *IEEE/ASME Transactions on Mechatronics*, vol. 11, no. 3, pp. 273-279, 2006.
- [4] D. Wu, G. Guo, and Y. Wang, "Reset integral-derivative control for HDD servo systems," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 1, pp. 161-167, 2007.
- [5] K. Peng, B. M. Chen, G. Cheng, and T. H. Lee, "Modeling and compensation of nonlinearities and friction in a micro hard disk drive servo system with nonlinear feedback control," *IEEE Transactions on Control Systems Technology*, vol. 13, no. 5, pp. 708-721, 2005.
- [6] T. Yan and R. Lin, "Experimental modeling and compensation of pivot nonlinearity in hard disk drives," *IEEE Transactions on Magnetics*, vol. 39, no. 2, pp. 1064-1069, 2003.
- [7] B. J. Harmer, "Closed loop calibration of back EMF measurement," US20100182715 A1, 2010.
- [8] M. Yamashita, C. Ota, and N. Wasa, "HDD with VCM back EMF and voltage control circuitry," US patent 8497641 B2, 2013.
- [9] R. Oboe, F. Marcassa, and G. Maiocchi, "Hard disk drive with voltage-driven voice coil motor and model-based control," *IEEE Transactions on Magnetics*, vol. 41, no. 2, pp. 784-790, 2005.
- [10] K. Williams and T. Pham, "Rotating crash stop assembly for hard disk drives," US patent 5715119, 1998.
- [11] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, pp. 35-45, 1960.
- [12] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, vol. 141 of *Mathematics in Science and Engineering*, Academic Press, New York, NY, USA, 1979.
- [13] J. C. Gamazo-Real, E. Vázquez-Sánchez, and J. Gómez-Gil, "Position and speed control of brushless dc motors using sensorless techniques and application trends," *Sensors*, vol. 10, no. 7, pp. 6901-6947, 2010.
- [14] B. M. Chen, T. H. Lee, K. Peng, and V. Venkataramanan, *Hard Disk Drive Servo Systems*, Springer, London, UK, 2nd edition, 2006.
- [15] P.-D. Pfister and Y. Perriard, "Very-high-speed slotless permanent-magnet motors: analytical modeling, optimization, design, and torque measurement methods," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 1, pp. 296-303, 2010.
- [16] Q. H. Zeng and D. B. Bogy, "Stiffness and damping evaluation of air bearing sliders and new designs with high damping," *Journal of Tribology*, vol. 121, no. 2, pp. 341-347, 1999.

## Research Article

# A New Transmissibility Based Indicator of Local Variation in Structure and Its Application for Damage Detection

**X. Z. Li, Z. K. Peng, X. J. Dong, W. M. Zhang, and G. Meng**

*State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China*

Correspondence should be addressed to Z. K. Peng; [z.peng@sjtu.edu.cn](mailto:z.peng@sjtu.edu.cn)

Received 4 May 2015; Accepted 24 June 2015

Academic Editor: Dong Wang

Copyright © 2015 X. Z. 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.

The presence of damage in engineering structures could usually cause local variation in stiffness or damping, and therefore it is meaningful to detect the variation as early as possible for protecting the engineering structure from serious damage. In the study, a novel method is developed to detect and locate the local variation in stiffness and damping for structures based on transmissibility. Some important properties of transmissibility are first analytically revealed and, further, a variation detection indicator is proposed to locate the variation. The effectiveness of the proposed method is verified by numerical studies and the usefulness of it is demonstrated by application for detecting crack position in beam structures. The results show that the proposed new method has better performance than the three conventional transmissibility based methods when considering different frequency bands and noise.

## 1. Introduction

Researchers have developed many useful vibration-based techniques for detecting structural damage, and enormous literatures [1–3] have been published to detect, locate, and quantify the damage. Normally, existence of structural damage in an engineering system leads to modification of the vibration. The damage caused by stiffness variation could alter the dynamic properties of the structure, such as natural frequencies and mode shapes. Therefore, feature extraction through vibration testing of healthy or damaged structures is the main core of vibration-based methods. One of the major challenges in structural health monitoring is the detection and location of small-scale damage since it is local damage. The difficulty mainly relies on two reasons: (i) the modal properties of the structure are also influenced by environmental parameters, such as load and boundary condition, which may cause changes in the vibration characters of the same order of magnitude as damage; (ii) modal parameters are global properties in nature and are only sensitive to substantial amounts of damage.

Transmissibility functions (TFs) [4], defined as the ratios of response in frequency domain or energy between two different DOFs of structure, have been a meaningful theory to evaluate the structural condition [5]. One key property of

transmissibility functions is that it is local in nature. Zhang et al. [6] and Mottershead [7] demonstrated that transmissibility functions are determined solely by system zeros, which makes the transmissibility functions sensitive to the local variation. Structural damage affects both poles and zeros of system. While the poles are sensitive to the changes in structural health anywhere because it is a function of all the system parameters, the zeros are only sensitive to a certain subset of mass, damping, and stiffness that are localized in a region of the structure.

Taking advantage of this basic property, a large number of detection methods were proposed in the last 20 years [5]. In 1994, transmissibility functions were firstly proposed as potential features for damage detection by Chen et al. [8]. A neural network using TFs was developed to compute a novelty index for damage detection. After that, efforts have been made on the use of the transmissibility functions for local damage detection. In 1997, Schulz et al. [4] proposed a damage indicator using transmissibility of cross and autopower spectral density for the first time, which was used by many authors subsequently and was evolved into the spectra ratio of response directly later. The change of power transmissibility near damage was used to locate the crack. Then, Zhang et al. [6] used both magnitude and phase of transmissibility functions to improve the locating

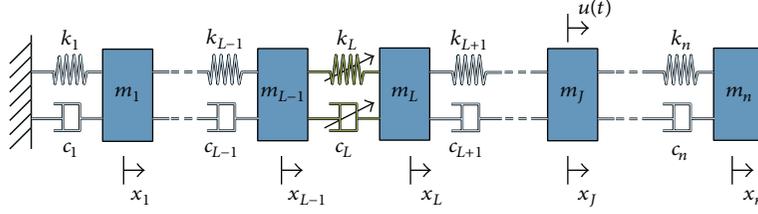


FIGURE 1: A one-dimensional MDOF system with local variation.

ability in 1999, and a concept of curvature TF was also developed to locate the damage. Experiment was made using cantilever beam to demonstrate the better ability than Schulz's method in the same frequency band. Johnson and Adams [9] extended the application of transmissibility to nonlinear system using quotient of transmissibility ratios in 2002. Structure with nonlinear damage can be also detected by the indicator in that literature.

Besides these methods proposed more than a decade ago, with the application of transmissibility in operational modal analysis [10, 11], some new indicators based on transmissibility were proposed for structural health monitoring in the last few years. In 2011, Maia et al. [12] extended the concept of Response Vector Assurance Criterion (RVAC) by replacing the FRF with transmissibility function in the indicator RVAC and damage quantification indicator (DRQ), which is the average value of RVAC along the frequency domain. The new method can detect and quantify the damage, but not location of damage. In 2013, Li and Hao [13] used the power spectral density transmissibility (PSDT) to detect the damage of shear connectors. PSDT matrices in the undamaged and damaged states or damaged states only were directly compared to identify the damage of shear connectors, while, in 2014, the same authors [14] proposed another method using transmissibility of frequency response function. By comparing the transmissibility vectors in undamaged and damaged states, the damage level of shear connectors can be identified. Then, in 2015, a method using response reconstruction with PSDT was published by them [15] again. In the same period, Zhou et al. proposed two new kinds of damage indicators using transmissibility in 2015. One kind of indicator [16] used the accumulation or modal assurance criterion of Transmissibility Coherence (TC) in frequency domain, which was defined as magnitude squared coherence. The detection results were compared with the FRF coherence and show a better performance. The other kind of damage detection approach [17] used the Mahalanobis distance of transmissibility as a damage feature. Mahalanobis distance can distinguish the damaged structural state condition from the undamaged one by considering the baseline data. While both kinds of method can quantify the damage severity, they cannot locate the damage.

The objective of the present research is to develop and test a novel method to detect local damage with transmissibility functions, which can be applied in linear and nonlinear MDOF systems. Properties of transmissibility functions in one-dimensional MDOF system with local variation are present, and a detection indicator is developed with the support of analytical derivation. Then, numerical simulations of linear, bilinear, and nonlinear systems are used to validate

the properties and indicator. The technique is applied for crack position detection in beam structure and is compared with the three popular indicators by Schulz et al. [4], Zhang et al. [6], and Johnson and Adams [9].

## 2. Transmissibility Functions of MDOF System

**2.1. MDOF System with Local Variation.** Considering a general dynamic structural system with multiple-input and multiple-output, the motion equation of that system with local variation can be described as

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) + \mathbf{F}_{LV}(\dot{\mathbf{x}}(t), \mathbf{x}(t)) = \mathbf{F}(t), \quad (1)$$

where  $\mathbf{M}$ ,  $\mathbf{C}$ , and  $\mathbf{K}$  represent the  $n \times n$  mass, damping, and stiffness matrices of the system,  $\mathbf{x}(t)$  is the  $n \times 1$  response vector,  $\mathbf{F}(t)$  is the  $n \times 1$  external force excitation, and  $n \times 1$  vector  $\mathbf{F}_{LV}(\dot{\mathbf{x}}(t), \mathbf{x}(t))$  models the additional forces induced by the local variation.

For the sake of simplicity, consider the one-dimensional MDOF structure in Figure 1, which has only one local variation component and is subjected to one input excitation. Since TFs describe the connection between two different points of the system, it is more convenient to study their properties in one-dimensional system.

As shown in Figure 1, the  $L$ th component has a local variation and a force excitation is acting on the  $J$ th mass. The local variation induces the local restoring force  $F_v$  between the  $(L - 1)$ th component and  $L$ th component, which not only affect the element nearby but also change the dynamic properties of the whole system.

(i) If  $F_v$  is a linear force, it can be described as

$$F_v = k_L^1 \Delta + c_L^1 \dot{\Delta}, \quad (2)$$

where  $k_L^1$  and  $c_L^1$  are linear stiffness and damping coefficient and  $\Delta = x_L - x_{L-1}$ ,  $\dot{\Delta} = \dot{x}_L - \dot{x}_{L-1}$ . Under this condition, the system is a linear system with linear local variation. And this kind of system might be caused by corrosion or loss in preload.

(ii) If  $F_v$  is a bilinear force, it can be described as

$$F_v = \begin{cases} k_L' \Delta + c_L' \dot{\Delta} & x_L(t) > 0 \\ 0 & \text{else,} \end{cases} \quad (3)$$

where  $k_L'$  and  $c_L'$  are the linear stiffness and damping coefficient which only connect with the  $L$ th component. Under this condition, the system is a system with a bilinear component. This condition could be induced by the dry friction or damage of viscous-elastic material.

(iii) If  $F_v$  is a nonlinear force, according to Weierstrass Approximation Theorem [18], it is assumed that the restoring force can be described by the polynomial functions, yielding

$$F_v = \sum_{i=1}^P k_L^i \Delta^i + \sum_{i=1}^P c_L^i \dot{\Delta}^i, \quad (4)$$

where  $P$  is the polynomial degree and  $k_L^i, c_L^i$  ( $i = 2, \dots, P$ ) are nonlinear stiffness and damping coefficient. Under this condition, the system becomes a nonlinear system, which might be caused by the crack or loosened fasteners.

So, vector  $F_{LV}$  in one-dimensional MDOF structure yields

$$F_{LV}(t) = \left( \begin{array}{c} \overbrace{0, \dots, 0}^{L-2}, F_v, -F_v, \overbrace{0, \dots, 0}^{n-L} \end{array} \right)'. \quad (5)$$

And, the other parameters of this model are shown below. Consider that

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_n \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} c_1 + c_2 & -c_2 & 0 & \cdots & 0 \\ -c_2 & c_2 + c_3 & -c_3 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -c_{n-1} & c_{n-1} + c_n & -c_n \\ 0 & \cdots & 0 & -c_n & c_n \end{bmatrix}, \quad (6)$$

$$\mathbf{K} = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & \cdots & 0 \\ -k_2 & k_2 + k_3 & -k_3 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -k_{n-1} & k_{n-1} + k_n & -k_n \\ 0 & \cdots & 0 & -k_n & k_n \end{bmatrix}$$

are the system mass, damping, and stiffness matrices, respectively. Consider that

$$\mathbf{x} = (x_1, \dots, x_n)',$$

$$\mathbf{F}(t) = \left( \begin{array}{c} \overbrace{0, \dots, 0}^{J-1}, u(t), \overbrace{0, \dots, 0}^{n-J} \end{array} \right)' \quad (7)$$

are the displacement and external force vector, respectively. It is assumed that  $J \geq L$ .

Since this model could contain linear, bilinear, or nonlinear component, it will be used to find a suitable method to detect the local variation in all kinds of structure. Here, we extend the concept of the traditional transmissibility into the nonlinear system directly and use the ratio of the frequency domain response to calculate TFs.

**2.2. Transmissibility Functions and Properties of the System.** A very powerful technique for solving the ordinary and partial differential equations describing the way in which certain quantities vary with time is that of the Laplace transform, which literally transforms these problems from time domain into complex frequency domain, from differential equations into an elementary algebraic expression [19]. Setting the initial conditions to zeros and applying the Laplace transform, the original motion equation yields

$$(\mathbf{M}s^2 + \mathbf{C}s + \mathbf{K}) \mathbf{x}(s) = \mathbf{F}_{LV}(s) + \mathbf{F}(s), \quad (8)$$

where  $\mathbf{x}(s)$ ,  $\mathbf{F}(s)$  are the complex frequency domain response and excitation vectors, respectively, and  $\mathbf{F}_{LV}(s)$  is the complex frequency domain additional force induced by the local variation.

Therefore, the complex frequency domain response  $\mathbf{x}(s)$  can be described as

$$\mathbf{x}(s) = \mathbf{H}(s) (\mathbf{F}_{LV}(s) + \mathbf{F}(s)), \quad (9)$$

where

$$\mathbf{H}(s) = (\mathbf{M}s^2 + \mathbf{C}s + \mathbf{K})^{-1}$$

$$= \begin{bmatrix} m_1 s^2 + (c_1 + c_2)s + k_1 + k_2 & -c_2 s - k_2 & 0 & \cdots & 0 \\ -c_2 s - k_2 & m_2 s^2 + (c_2 + c_3)s + k_2 + k_3 & -c_3 s - k_3 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -c_{n-1} s - k_{n-1} & m_{n-1} s^2 + (c_{n-1} + c_n)s + k_{n-1} + k_n & -c_n s - k_n \\ & & 0 & -c_n s - k_n & m_n s^2 + c_n s + k_n \end{bmatrix}^{-1} \quad (10)$$

is the transfer function matrix of the system.

From the transfer function matrix, it can be known that the response is coupled, which means that the frequency response of one component is connected with that of the component before and after. And, according to the definition of TFs, the transmissibility of the MDOF system in the complex frequency domain can be derived from (9).

For any two consecutive components, the TFs can be expressed as

$$T_{(i+1,i)}(s) = \frac{x_{i+1}(s)}{x_i(s)}, \quad (11)$$

where  $i = 1, \dots, n-1$ .

Thus, according to (11), when  $i = 1$ , the first row of (9) yields

$$\begin{aligned} (m_1 s^2 + (c_1 + c_2)s + (k_1 + k_2))x_1(s) \\ - (c_2 s + k_2)x_2(s) = 0. \end{aligned} \quad (12)$$

So

$$T_{(2,1)}(s) = \frac{m_1 s^2 + (c_1 + c_2)s + (k_1 + k_2)}{c_2 s + k_2}. \quad (13)$$

Subsequently, substituting (13) into the second row of (9) yields

$$T_{(3,2)}(s) = \frac{m_2 s^2 + (1 - 1/T_{(2,1)})(c_2 s + k_2) + c_3 s + k_3}{c_3 s + k_3}. \quad (14)$$

Iteratively using the above procedure until  $i = L-2$ , it can be known that, for  $2 \leq i \leq L-2$ , the TFs yield

$$\begin{aligned} T_{(i+1,i)}(s) \\ = \frac{m_i s^2 + (1 - 1/T_{(i,i-1)})(c_i s + k_i) + c_{i+1} s + k_{i+1}}{c_{i+1} s + k_{i+1}}. \end{aligned} \quad (15)$$

Then, for  $i = L-1$ , substituting  $T_{(L-1,L-2)}$  into the  $(L-1)$ th row of (9) yields

$$\begin{aligned} T_{(L,L-1)}(s) \\ = \frac{m_{L-1} s^2 + (1 - 1/T_{(L-1,L-2)})(c_{L-1} s + k_{L-1}) + c_L s + k_L}{c_L s + k_L} \left( 1 \right. \\ \left. + \frac{F_v(s)/x_{L-1}(s)}{m_{L-1} s^2 + (1 - 1/T_{(L-2,L-1)})(c_{L-1} s + k_{L-1}) + c_L s + k_L} \right). \end{aligned} \quad (16)$$

And, for  $i = L$

$$\begin{aligned} T_{(L+1,L)}(s) \\ = \frac{m_L s^2 + (1 - 1/T_{(L,L-1)})(c_L s + k_L) + c_{L+1} s + k_{L+1}}{c_{L+1} s + k_{L+1}} \left( 1 \right. \\ \left. - \frac{F_v(s)/x_L(s)}{m_L s^2 + (1 - 1/T_{(L,L-1)})(c_L s + k_L) + c_{L+1} s + k_{L+1}} \right). \end{aligned} \quad (17)$$

Since  $T_{(i+1,i)}$  is complicated for  $i = L-1, L$ , it is difficult to calculate and judge the TFs for the next components. However, a similar procedure can be conducted starting from the  $n$ th row of (9). So, for  $i = n$

$$T_{(n,n-1)}(s) = \frac{c_n s + k_n}{m_n s^2 + c_n s + k_n}. \quad (18)$$

Subsequently, for  $J+1 \leq i \leq n-1$

$$\begin{aligned} T_{(i,i-1)}(s) \\ = \frac{c_i s + k_i}{m_i s^2 + (1 - T_{(i+1,i)})(c_{i+1} s + k_{i+1}) + c_i s + k_i}. \end{aligned} \quad (19)$$

Then, when  $i = J$

$$\begin{aligned} T_{(J,J-1)}(s) \\ = \frac{c_J s + k_J}{m_J s^2 + (1 - T_{(J+1,J)})(c_{J+1} s + k_{J+1}) + c_J s + k_J} \left( 1 \right. \\ \left. + \frac{u(s)/x_{J-1}(s)}{c_J s + k_J} \right). \end{aligned} \quad (20)$$

Until now, the TFs in complex frequency domain have been calculated for this MDOF system. Then, we need to transfer these functions into frequency domain. Since the Fourier transform is a special case of the Laplace transform [19], the frequency domain of TFs yields

$$T_{(i+1,i)}(j\omega) = T_{(i+1,i)}(s)|_{s=j\omega}. \quad (21)$$

The equality holds up if and only if the convergence region of the Laplace transform contains the imaginary axis. After TFs in frequency domain are got from (21), the influence of the local variation could be studied by sensitivity.

According to the TFs above, the whole system is divided into four kinds of components.

(i) For  $1 \leq i \leq L-2$  and  $J \leq i \leq n-1$

$$\frac{\partial T_{(i+1,i)}(j\omega)}{\partial F_v(j\omega)} = 0, \quad (22)$$

$$\frac{\partial T_{(i+1,i)}(j\omega)}{\partial u(j\omega)} = 0.$$

(ii) For  $i = L-1, L$

$$\frac{\partial T_{(i+1,i)}(j\omega)}{\partial F_v(j\omega)} \neq 0, \quad (23)$$

$$\frac{\partial T_{(i+1,i)}(j\omega)}{\partial u(j\omega)} = 0.$$

(iii) For  $i = J-1$

$$\frac{\partial T_{(i+1,i)}(j\omega)}{\partial F_v(j\omega)} = 0, \quad (24)$$

$$\frac{\partial T_{(i+1,i)}(j\omega)}{\partial u(j\omega)} \neq 0.$$

(iv) For  $L < i < J - 1$

$$\begin{aligned} \frac{\partial T_{(i+1,i)}(j\omega)}{\partial F_v(j\omega)} &\neq 0, \\ \frac{\partial T_{(i+1,i)}(j\omega)}{\partial u(j\omega)} &\neq 0. \end{aligned} \quad (25)$$

It is difficult to calculate the sensitivity of TFs to  $F_v(j\omega)$ ,  $u(j\omega)$  for  $L < i < J - 1$  directly. But, from the derivation of the TFs at  $i = L - 1, L, J$ , it can be known that transmissibility should be influenced by  $F_v(j\omega)$  and  $u(j\omega)$ .

One conclusion can be got from the sensitivity study directly, which is that the TFs are influenced by the local variation as well as the location of excitation. This conclusion agrees with the research before [9]. With further study, the TFs in case (i) are irrelevant to the stiffness and damping coefficients  $k_L^i$ ,  $c_L^i$  and excitation  $u(j\omega)$ . For instance, the transmissibility function at  $i = 1$  is always equal to  $(-m_1\omega^2 + (c_1 + c_2)j\omega + (k_1 + k_2))/(c_2j\omega + k_2)$ , no matter that  $k_L^i$  is nonzero or not. While the TFs in cases (ii) and (iii) are influenced directly by the local variation and excitation, case (iv) is affected mediately.

For  $i = L - 1$ , if there is no local variation, TFs yield

$$\begin{aligned} \bar{T}_{(L,L-1)}(j\omega) \\ = \frac{-m_{L-1}\omega^2 + (1 - 1/T_{(L-1,L-2)})(c_{L-1}j\omega + k_{L-1}) + c_Lj\omega + k_L}{c_Lj\omega + k_L}. \end{aligned} \quad (26)$$

Set  $\alpha = (F_v(j\omega)/x_{L-1}(j\omega))/(-m_{L-1}\omega^2 + (1 - 1/T_{(L-2,L-1)})(c_{L-1}j\omega + k_{L-1}) + c_Lj\omega + k_L)$ , and then

$$\frac{T_{(L,L-1)}(j\omega)}{\bar{T}_{(L,L-1)}(j\omega)} = (1 + \alpha). \quad (27)$$

So, (27) shows the difference of the TFs before and after the local variation. The difference grows with the increase of  $\alpha$ . And  $\alpha$  changes with different frequency. When frequency is near the natural frequency,  $x_{L-1}(j\omega)$  will be much bigger than that of the frequency far away from the natural frequency. If the local variation and excitation force are constant, then  $\alpha$  near the natural frequency will be smaller than that of the other frequency bands. That means the difference between the TFs before and after the local variation will be smaller than that of the other frequency bands.

While for  $i = J$ , if there is no excitation, TFs yield

$$\begin{aligned} \bar{T}_{(J,J-1)}(j\omega) \\ = \frac{c_Jj\omega + k_J}{-m_J\omega^2 + (1 - T_{(J+1,J)})(c_{n+1}j\omega + k_{n+1}) + c_Jj\omega + k_J}. \end{aligned} \quad (28)$$

Set  $\beta = (u(j\omega)/x_{J-1}(j\omega))/(c_Jj\omega + k_J)$ , and then

$$\frac{T_{(J,J-1)}(j\omega)}{\bar{T}_{(J,J-1)}(j\omega)} = (1 + \beta). \quad (29)$$

Similar to (27), (29) shows the influence of the external excitation to the transmissibility. If  $u(j\omega)$  is constant,  $\beta$

becomes small when the frequency is near the natural frequency. That means the influence of the external excitation is much weak in the frequency bands near the natural frequency.

Thus, from the discussion above, two properties of TFs of the MDOF structure can be got.

*Property 1.* In theory,  $T_{(i+1,i)}(j\omega)$  is irrelevant to the stiffness and damping coefficients  $k_L^i$ ,  $c_L^i$  ( $i = 1, \dots, P$ ) and external excitation  $u(j\omega)$  for components in the range of  $1 \leq i \leq L - 2$  and  $J \leq i \leq n - 1$ , while, for  $L - 1 \leq i \leq J - 1$ ,  $T_{(i+1,i)}(j\omega)$  is influenced by these parameters.

*Property 2.* To increase the sensitivity of TFs to the local variation, frequency bands should be selected narrowing the range of natural frequency, while, to decrease the dependence of TFs on external excitation, frequency bands should include the range of natural frequency.

Property 1 can be used to detect the local variation by TFs, because it gives the discrimination between the positions with and without local variation. And Property 2 validates the conclusion again that it is difficult to decide the frequency bands of the TFs until the test date got from the practice.

*2.3. Detection Indicator Based on Transmissibility.* From the two properties, it is revealed that the difference between the transmissibilities  $T_{(i+1,i)}(j\omega)$  and  $\bar{T}_{(i+1,i)}(j\omega)$  is zero for the interval without the restoring force and external excitation in theory, while, for the other components, it is nonzero. Therefore, an indicator can be developed to detect the local variation using the transmissibility difference between the intact structure and local variation structure:

$$\begin{aligned} TD_i = \sum_{\omega_0}^{\omega_N} w_i(j\omega) |T_{(i+1,i)}(j\omega) - \bar{T}_{(i+1,i)}(j\omega)| \\ (i = 1, \dots, n - 1), \end{aligned} \quad (30)$$

where  $\omega \in (\omega_0, \omega_N)$  is the frequency band and

$$\begin{aligned} w_i(j\omega) \\ = \frac{x_{i+1}(j\omega) + x_i(j\omega) + \bar{x}_{i+1}(j\omega) + \bar{x}_i(j\omega)}{\sum_{\omega_0}^{\omega_N} (x_{i+1}(j\omega) + x_i(j\omega) + \bar{x}_{i+1}(j\omega) + \bar{x}_i(j\omega))} \end{aligned} \quad (31)$$

is the factor, which is used to increase the weight of resonance. According to the two properties, frequency band including the resonance could reduce the influence of excitation. Hence, in order to increase the sensitivity of the indicators to damage, not the excitation, the weighting factor is added to the differences of TFs before and after local variation. This indicator is a linear sum along all the spectral lines of the functions, and the frequency band can be chosen differently for each case.

### 3. Simulation Studies on Different Kinds of MDOF Systems

Since the restoring force  $F_v$  could be linear, bilinear, or nonlinear, the indicator should be useful in all kinds of system. In

TABLE I: Comparison of the natural frequencies before and after variation.

Modes	1	2	3	4	5	6	7	8
$\bar{\omega}_n$	5.536	16.420	26.744	36.158	44.341	51.013	55.948	58.978
$\omega_n$	5.276	16.251	24.903	36.066	41.918	49.793	54.993	57.502
$ \omega_n - \bar{\omega}_n /\bar{\omega}_n$ (%)	4.697	1.029	6.884	0.254	5.464	2.392	1.707	2.503

this section, numerical simulation will be studied on MDOF systems with linear, bilinear, or nonlinear component to validate the properties and detection indicator developed in the previous section.

*3.1. Detection Indicators of System with Linear Local Variation.* A linear local variation system is the case when the initially linear-elastic structure remains linear-elastic after local variation. The changes in dynamic properties could be induced by changes in the geometry or the material properties of the structure, but the structural response can still be modeled using linear equations of motion. Lots of methods have been proposed to detect and identify the local variation or damage in such system. For instance, in [20], a nondestructive method using measurements of the structural natural frequency was described to detect, locate, and quantify damage of an aluminum plate. And change in the flexibility matrix was used to detect and locate damage on a wide-flange steel beam in [21], while, in this simulation, the ability of detection indicator proposed in Section 2 will be demonstrated in linear system.

A linear local variation system with 8 degrees of freedom is built using the one-dimensional MDOF model in Figure 1. The linear variation locates in the 4th component, so  $L = 4$ . An impulse excitation is applied on the 6th mass, so  $J = 6$ . It is assumed that the damping is proportional to the stiffness; for example,  $\mathbf{C} = \mu\mathbf{K}$ . The values of the system parameters are taken as

$$\begin{aligned}
 m_1 &= \dots = m_8 = 1, \\
 k_1 &= \dots = k_8 = 3.5531 \times 10^4, \\
 \mu &= 0.001, \\
 u(t) &= \begin{cases} 2 & t = 0 \\ 0 & t \neq 0, \end{cases} \\
 k_L^1 &= -0.4k_4.
 \end{aligned} \tag{32}$$

The fourth-order Runge-Kutta method is used to calculate the system responses. And the spectra of the responses can be got through the FFT. Then, TFs can be determined using the frequency responses.

Table 1 shows the first 8 natural frequencies of the system before and after the variation, which indicates that the small local variation is difficult to detect from the change of natural frequencies. The frequency band of the transmissibility is determined by including the first 5 natural frequencies and is set as 0–50 Hz.

The TFs of the 8 points are shown in Figure 2. And their differences in frequency band of 0–50 Hz are also calculated. For  $i = 1, 2$ ,  $T_{(2,1)}$ ,  $T_{(3,2)}$  are very close to  $\bar{T}_{(2,1)}$ ,  $\bar{T}_{(3,2)}$ , respectively, except for few frequency points near the 6th natural frequency. And  $T_{(6,7)}$ ,  $T_{(7,8)}$  and  $\bar{T}_{(6,7)}$ ,  $\bar{T}_{(7,8)}$  are similar to them, while, for  $i = 3, 4, 5$ , there are much bigger differences between the TFs before and after the variation. Thus, the results agree with the properties in the last section that TFs are influenced by the local variation and excitation. In addition, the differences are not always very big in the whole frequency band even for the TFs influenced by the local variation and input directly. That means not all the frequencies are sensitive to this kind of local variation. On the other hand, TFs away from the local variation and input are not all exactly equal to zero. That might be caused by the error from the resonance and antiresonance.

Then, indicators of this linear system are calculated and the results are shown in Figure 3. In this system, stiffness and damping variation locate in the 4th component, which connect the 3rd and 4th component. And  $TD_3$ , the transmissibility indicator between the 3rd and 4th components, is the maximum of all the indicators. So, the position of the local variation is detected by the indicators correctly. Since the external force is applied on the 6th component, the 3rd, 4th, and 5th indicators are bigger than the others. So, this is consistent with Property 1 that the transmissibility between the local variation and external force is influenced, while others are not. So, for this system, local variation is indicated correctly.

*3.2. Detection Indicators of System with Bilinear Component.* A local bilinear variation system is the case when the initially linear-elastic structure changes to nonlinear structure because of bilinear component. One example of this kind of system is the formation of a fatigue crack that subsequently opens and closes under the normal operating vibration environment. Rivola and White [22] provided a possibility of using the bispectral analysis technique to detect damage in structures and conducted simulation and experiment test on a beam with a fatigue crack. In [23], the authors found that the numerically generated steady state response of the system betrays the presence of the bilinear spring. So, this phenomenon was used to determine crack location, depth, and opening load in a cracked beam. In [24], a solution based on the use of two square wave functions to model stiffness change was proposed for bilinear oscillators under low-frequency excitation, and the predicted spectral pattern and the magnitude of each harmonic component for a simply supported beam showed the existence of crack. Since

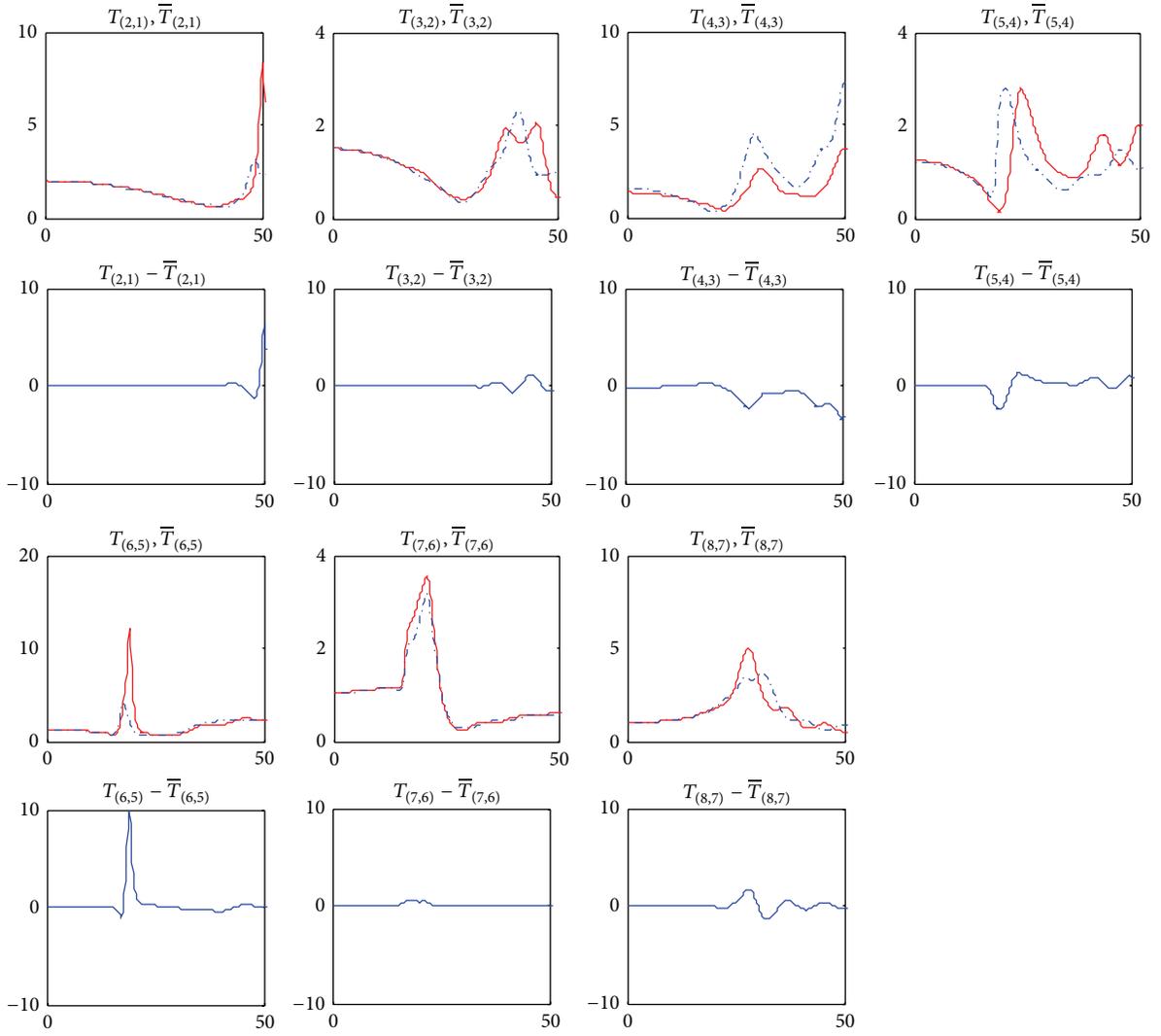


FIGURE 2: TFs of the system with linear local variation before and after the variation and their difference versus frequency (Hz): blue— $T_{(i+1,i)}$ , red— $\bar{T}_{(i+1,i)}$ .

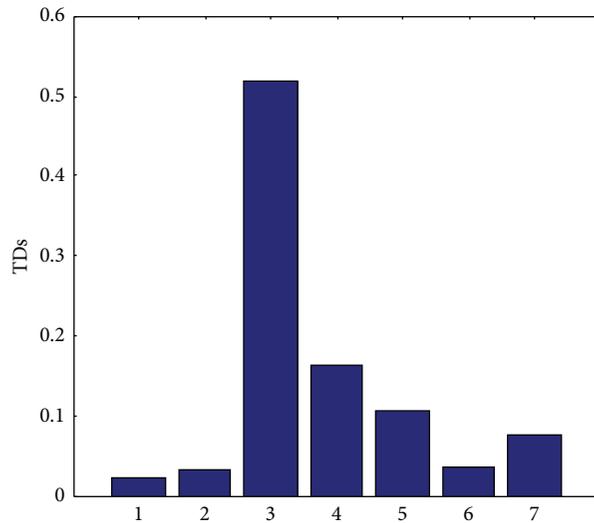


FIGURE 3: Detection indicators of the system with linear local variation.

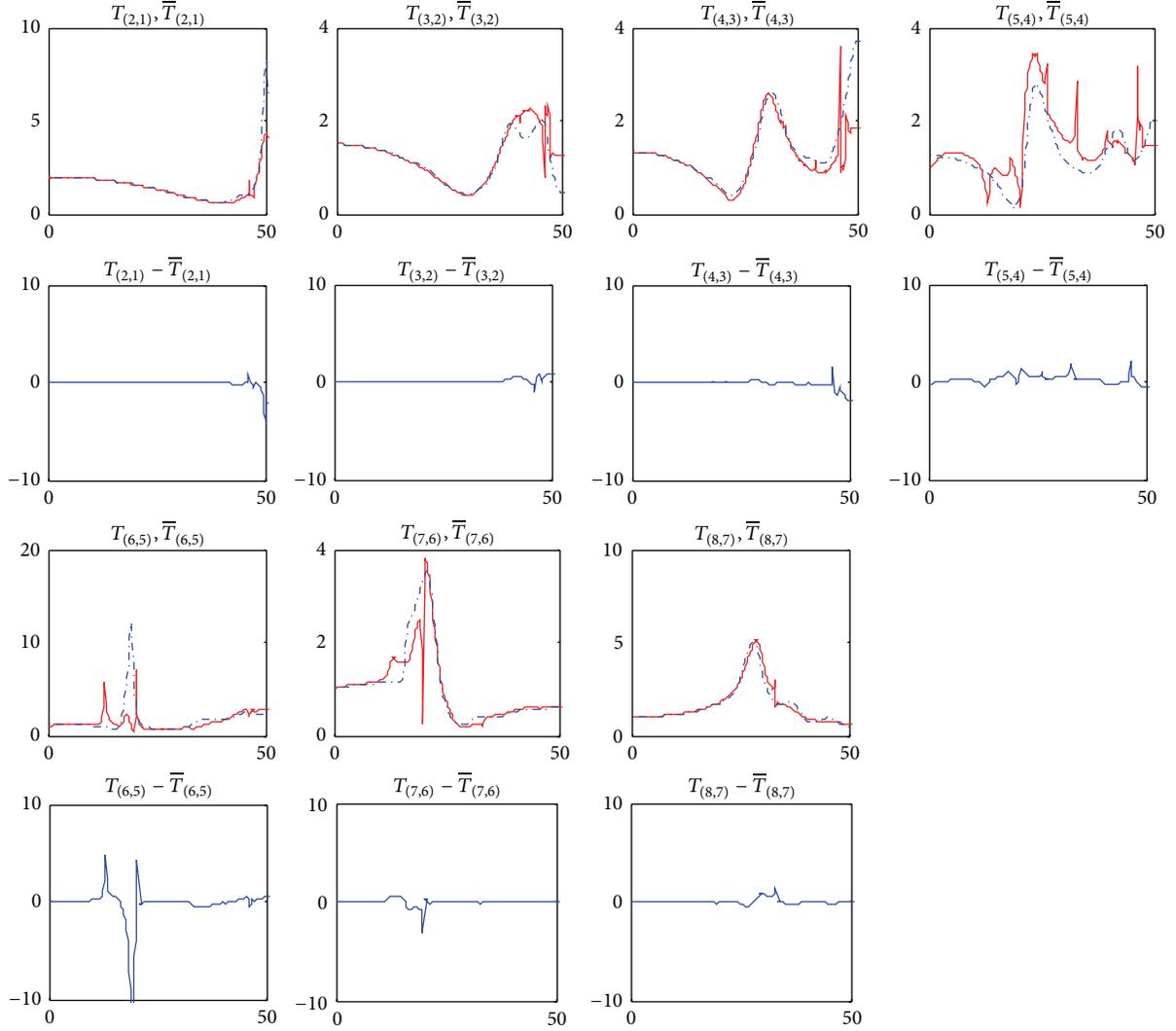


FIGURE 4: TFs of the system with bilinear component before and after the variation and their difference versus frequency (Hz): blue— $T_{(i+1,i)}$ , red— $T_{(i+1,i)}$ .

the bilinear system is a useful system in the crack detection, it will be used to validate the indicator of TFs here.

In this numerical simulation, the 4th component of one 8-DOF system contains bilinear stiffness; that is,  $L = 4$ . And an impulse excitation is applied on the 6th mass; that is,  $J = 6$ . The values of the system parameters are the same as the linear system in Section 3, except the additional stiffness  $k'$  attached to the 4th component alone. The value of  $k'$  is taken as

$$k' = \begin{cases} 0.6k_4 & x_4(t) > 0 \\ 0 & \text{else,} \end{cases} \quad (33)$$

where  $k_4$  and  $x_4(t)$  are linear stiffness and time response of the 4th component, respectively.

Then, the responses are calculated by the fourth-order Runge-Kutta method and response spectrums can be got.

Figure 4 shows the TFs of the system with bilinear component. It can be seen that, for  $i = 1, 2, 3, 6, 7$ ,  $T_{(i+1,i)}$  are close to  $\bar{T}_{(i+1,i)}$ , while, for  $i = 4, 5$ , there is much difference between them. Since the bilinear variation locates in the 4th component solely,  $T_{(5,4)}$  is affected seriously. And  $T_{(6,5)}$  is also influenced by location of the input force.

Then, detection indicators are calculated by the response spectrum of the 8 points. In Figure 5,  $TD_4$  and  $TD_5$  are much bigger than the other indicators. And  $TD_4$  is a little bigger than  $TD_5$ , so the local bilinear variation can be detected. In this simulation, it can be seen that the influence of the input force is strong compared with the local variation. So, location of the input force should be determined carefully.

**3.3. Detection Indicators of System with Nonlinear Component.** A nonlinear local variation system is the case when

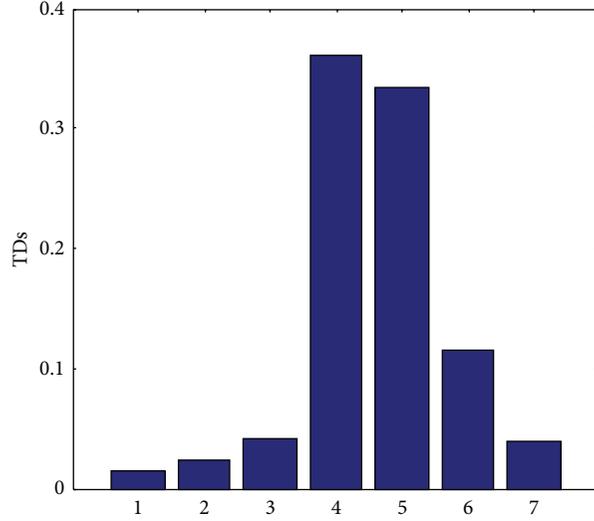


FIGURE 5: Detection indicators of the system with bilinear component.

the initially linear-elastic structure behaves in a nonlinear manner after the local nonlinear variation has been introduced. Its examples include loose connection, crack, and nonlinear material. Since the nonlinear variation is much complicated, some new nondestructive methods were proposed in the last few years. Ultrasonic wave is one of these methods. In [25], nonlinear Rayleigh surface wave was used to evaluate the fatigue damage induced by plastic deformation in A36 steel. And, in [26–28], another kind of ultrasonic wave, nonlinear Lamb wave, was used to detect and locate the nonlinear damage caused by material nonlinearity. The method even predicted the remaining lifetime of the structures. These two methods of the ultrasonic wave are normally based on the fact that the nonlinear wave could be induced by the damage. Compared with the traditional methods, these methods are much more sensitive to the damage, but strict requirements for the experiment equipment are needed. In this section, simulation will show the application of TFs to the nonlinear damage. This method is much more convenient compared to ultrasonic wave methods and is more sensitive to nonlinear damage compared with the traditional methods.

A nonlinear system is developed in this section. The nonlinear variation locates in the 4th component and we apply the impulse force on the 6th mass. So,  $L = 4$  and  $J = 6$ . The nonlinear stiffness and damping are set as

$$\begin{aligned}
 k_L^2 &= 0.4k_4^2, \\
 k_L^3 &= 1.4k_4^3, \\
 c_L^2 &= 0.001 \times 0.4k_4^2, \\
 c_L^3 &= 0.001 \times 1.4k_4^3.
 \end{aligned} \tag{34}$$

The other parameters of the nonlinear system are taken as the same values as the linear system. Then, the same procedure is applied on this nonlinear system and the results are shown in Figures 6 and 7.

In Figure 6, the results show great accordance to the properties proposed in Section 2. The local variation and position of input influence the transmissibility nearby. Then, indicators are calculated in Figure 7. The maximum value  $TD_3$  shows the position of the nonlinear variation.

Through the three simulation cases in this section, properties of the transmissibility functions can be validated and the indicator shows a great performance in the local variation detection in linear, bilinear, and nonlinear systems. It also can be seen that location of external force has a strong influence on the value of the indicator.

#### 4. Application for Crack Position Detection in Beam Structures

This section is focused on the crack position detection in cantilever beam. The responses of the beam structures with and without damage are calculated using the finite element model. So, the detection indicators can be got from their differences.

**4.1. Model Description.** The local variation is implemented using open crack. There are a number of methods to model crack in structures [29–33]. A simplified model of open crack by Sinha et al. [34] is adopted in this paper. In this model, for the area close to the crack, the stiffness is given by

$$EI(\xi) = \begin{cases} EI_0 - E(I_0 - I_{cj}) \frac{(\xi - \xi_{j1})}{(\xi_j - \xi_{j1})} & \xi_{j1} \leq \xi \leq \xi_j \\ EI_0 - E(I_0 - I_{cj}) \frac{(\xi_{j2} - \xi)}{(\xi_{j2} - \xi_j)} & \xi_j \leq \xi \leq \xi_{j2}, \end{cases} \tag{35}$$

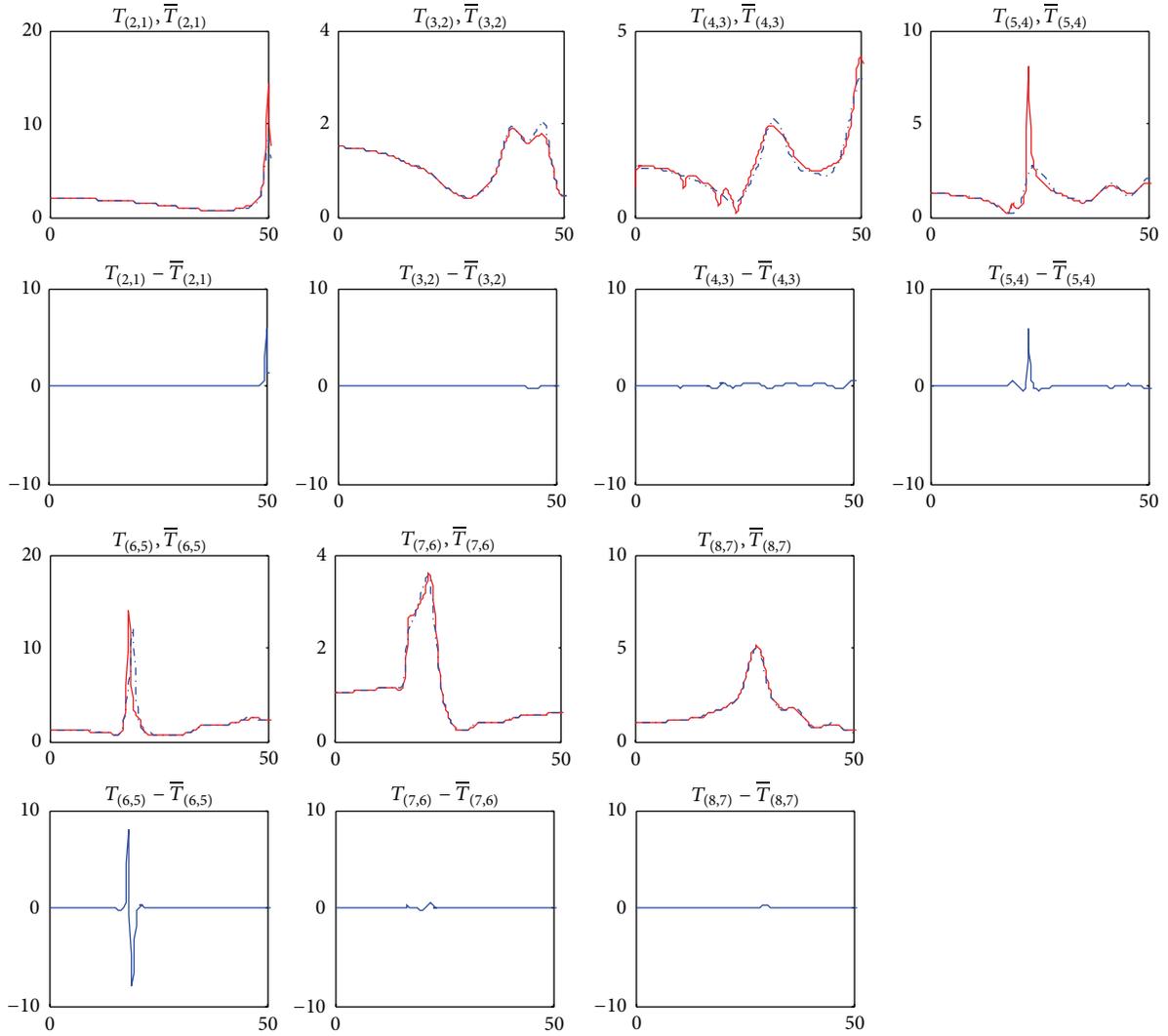


FIGURE 6: TFs of the system with nonlinear component before and after the variation and their difference versus frequency (Hz): blue— $T_{(i+1,i)}$ , red— $\bar{T}_{(i+1,i)}$ .

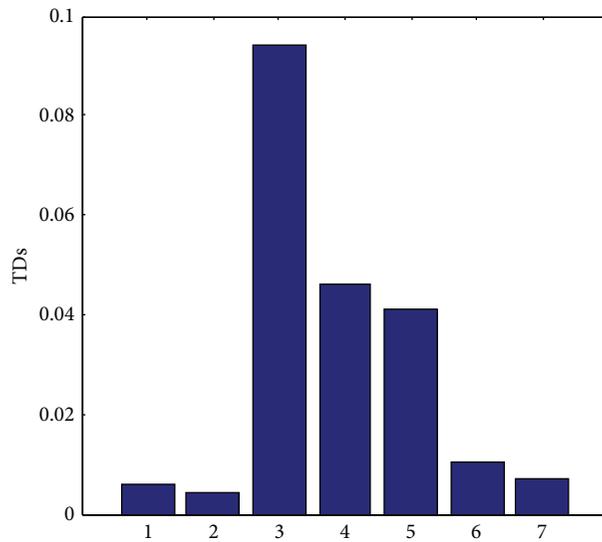


FIGURE 7: Detection indicators of the system with nonlinear component.

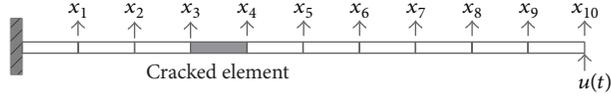
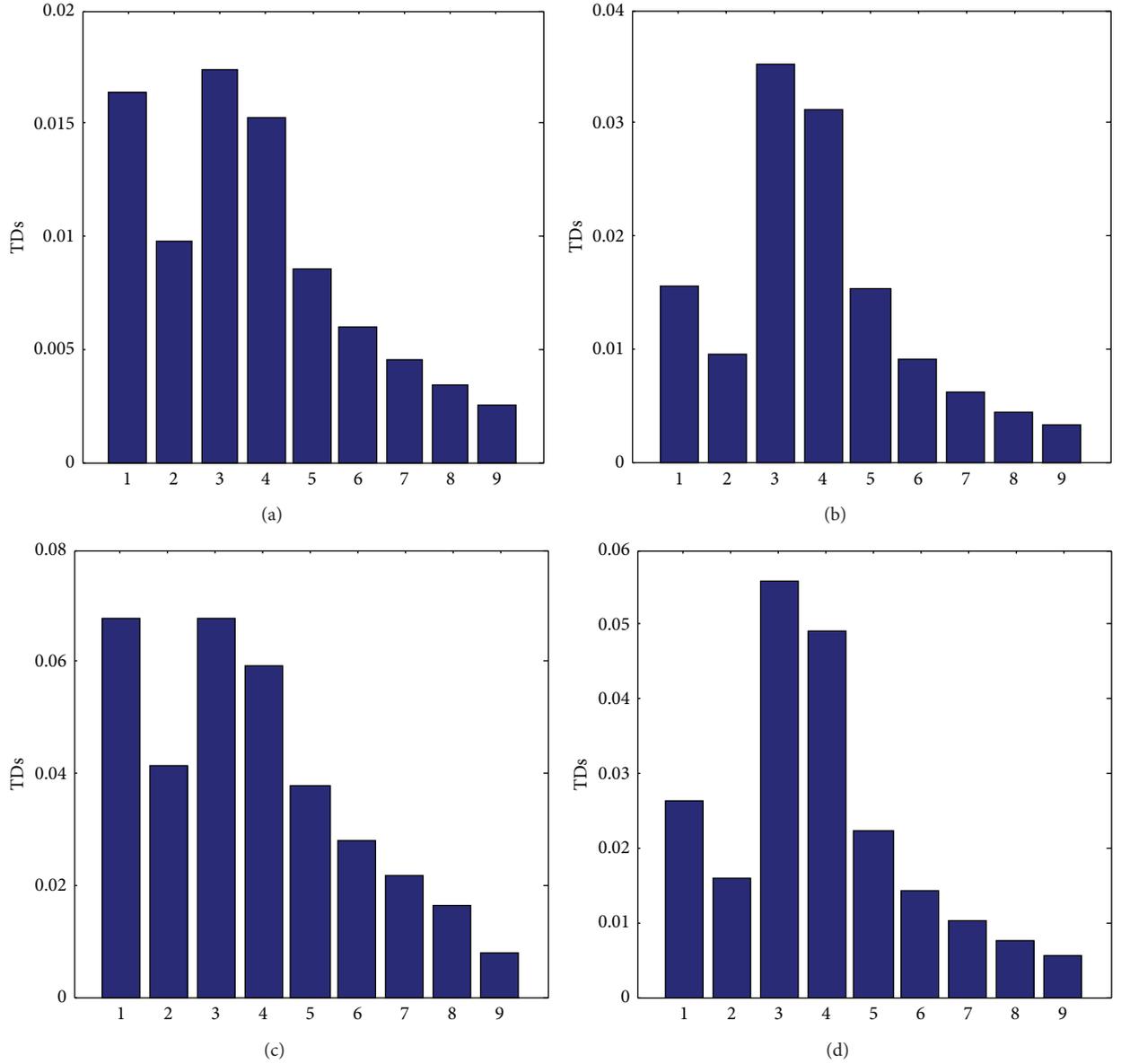


FIGURE 8: Finite element model of a cantilever beam.

FIGURE 9: Detection indicators of the cantilever beam for 0–50 Hz with different crack depth ratios: (a)–(d) are the indicators with  $\gamma$  from 0.1 to 0.4, respectively.

where  $\xi_j$  is the location of the  $j$ th crack within the element and  $\xi_{j1} = \xi_j - l_c$  and  $\xi_{j2} = \xi_j + l_c$  are the positions on either side of the crack where the stiffness reduction begins.  $I_0 = wd^3/12$  and  $I_{cj} = w(d - d_{cj})^3/12$  are the second moment of the areas of the undamaged beam and at the  $j$ th crack.  $d_{cj}$  is the crack depth of beam with dimensions  $L \times w \times d$ .

The advantage of this simplified model is that the stiffness reduction is local and the stiffness matrix may be written as an explicit function of the crack location and depth. The location of the crack is set at the center of one element, and  $\gamma = d_{cj}/d$  is set as the ratio of crack depth to depth of the beam. Hence, the stiffness matrix is the function of  $\gamma$ .

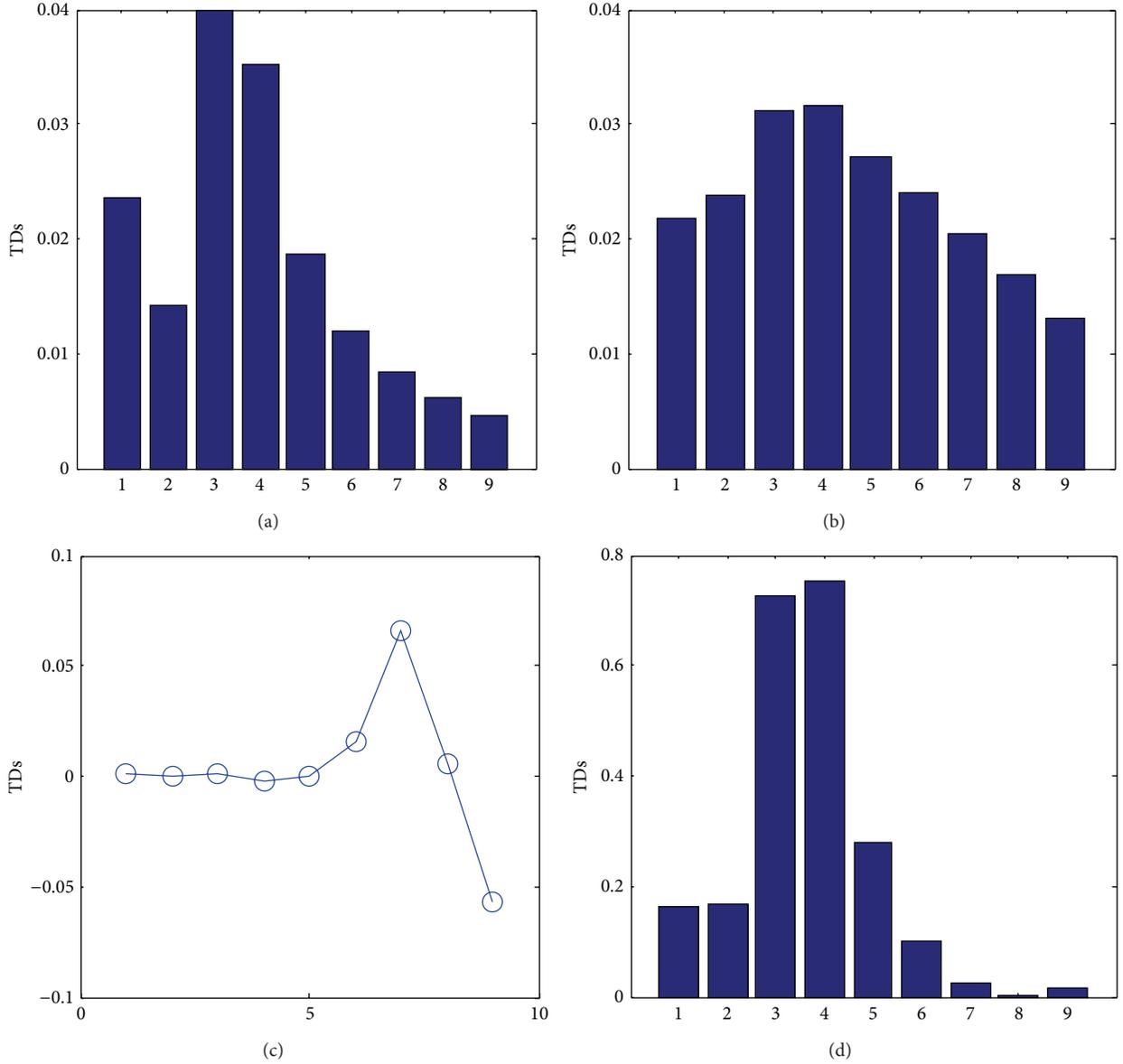


FIGURE 10: Detection indicators of the cantilever beam when  $\gamma = 0.2$  with frequency band of 0–10 Hz: (a) this paper, (b) Schulz et al. [4], (c) magnitude value of Zhang et al. [6], and (d) Johnson and Adams [9].

Figure 8 shows the finite element model of the beam used in this work. The dimensions of the beam are  $1.00 \times 0.05 \times 0.025 \text{ m}^3$  ( $L \times w \times d$ ). It has a mass density of  $2600 \text{ kg/m}^3$ , an elastic modulus of  $69.79 \text{ MPa}$ , and a damping ratio of  $0.001$ . The whole beam is divided into 10 elements. An open crack happens in the 4th element.

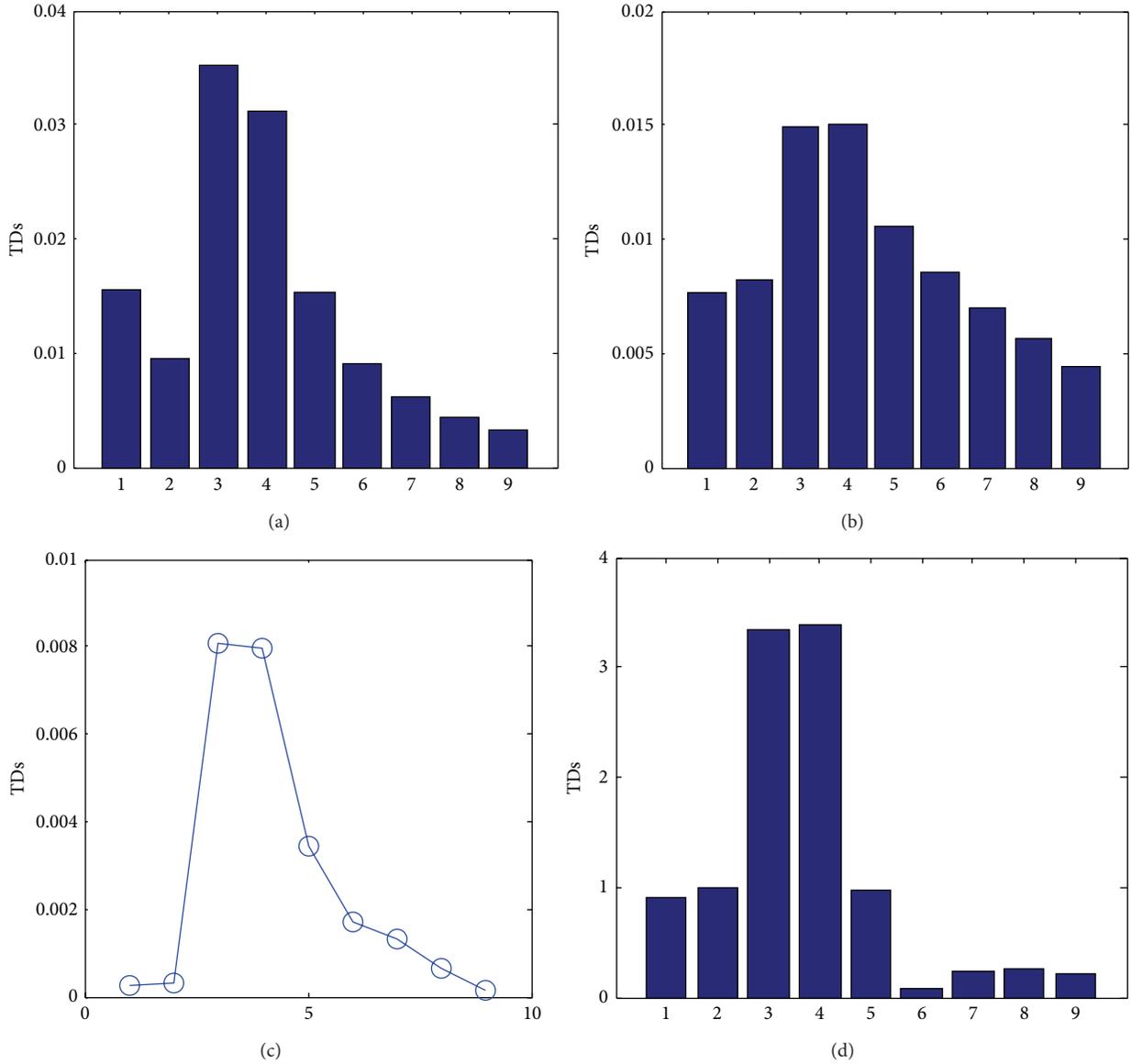
The first three natural frequencies of intact beam and cracked beam are shown in Table 2. Through the comparison, it shows that the natural frequencies change slightly with the increase of crack depth ratio  $\gamma$ , where the maximum decrease ratio is only  $1.699\%$  for the first natural frequency. That means it is difficult to detect the crack from the natural frequencies directly.

According to the natural frequencies of the beam, a frequency band of 0–50 Hz is selected for the detection tests in the first, and 20 N impulse force is applied on the right end of the beam to excite the structure. The fourth-order Runge-Kutta method is used to calculate the response of the beam. Then, the spectrum can be calculated and TFs can be got.

**4.2. Results and Comparison.** Figure 9 gives the result of the detection indicators of the cracked cantilever beam in four different damaged levels for frequency band of 0–50 Hz without noise. Since the responses are measured from the second point of the beam,  $\text{TD}_3$  represents the indicator of the 4th element, which is the maximum value of all indicators in

TABLE 2: Comparison of natural frequencies of intact and damaged cantilever beam with different crack depth ratios.

$\gamma$	First mode			Second mode			Third mode		
	Intact	Damaged	Rel. (%)	Intact	Damaged	Rel. (%)	Intact	Damaged	Rel. (%)
0.1	4.147	4.127	0.484	11.613	11.521	0.791	22.773	22.766	0.031
0.2	4.147	4.108	0.941	11.613	11.437	1.518	22.773	22.760	0.057
0.3	4.147	4.091	1.352	11.613	11.363	2.153	22.773	22.756	0.078
0.4	4.147	4.076	1.699	11.613	11.302	2.677	22.773	22.752	0.095

FIGURE 11: Detection indicators of the cantilever beam when  $\gamma = 0.2$  with frequency band of 0–50 Hz: (a) this paper, (b) Schulz et al. [4], (c) magnitude value of Zhang et al. [6], and (d) Johnson and Adams [9].

each case. Hence, crack location can be detected and located correctly, while, for  $\gamma$  equal to 0.1 and 0.3,  $TD_1$  is very big, which means that the detection indicators are influenced by the boundary conditions.

Since the damage indicators based on transmissibility functions are sensitive to the frequency bands and noise [5],

a comparison will be made to study the advantages of the proposed method in this paper. The indicator proposed in this paper is compared with the indicators defined by Schulz et al. [4], Zhang et al. [6], and Johnson and Adams [9] in the same damage conditions with different frequency bands or different noise levels.

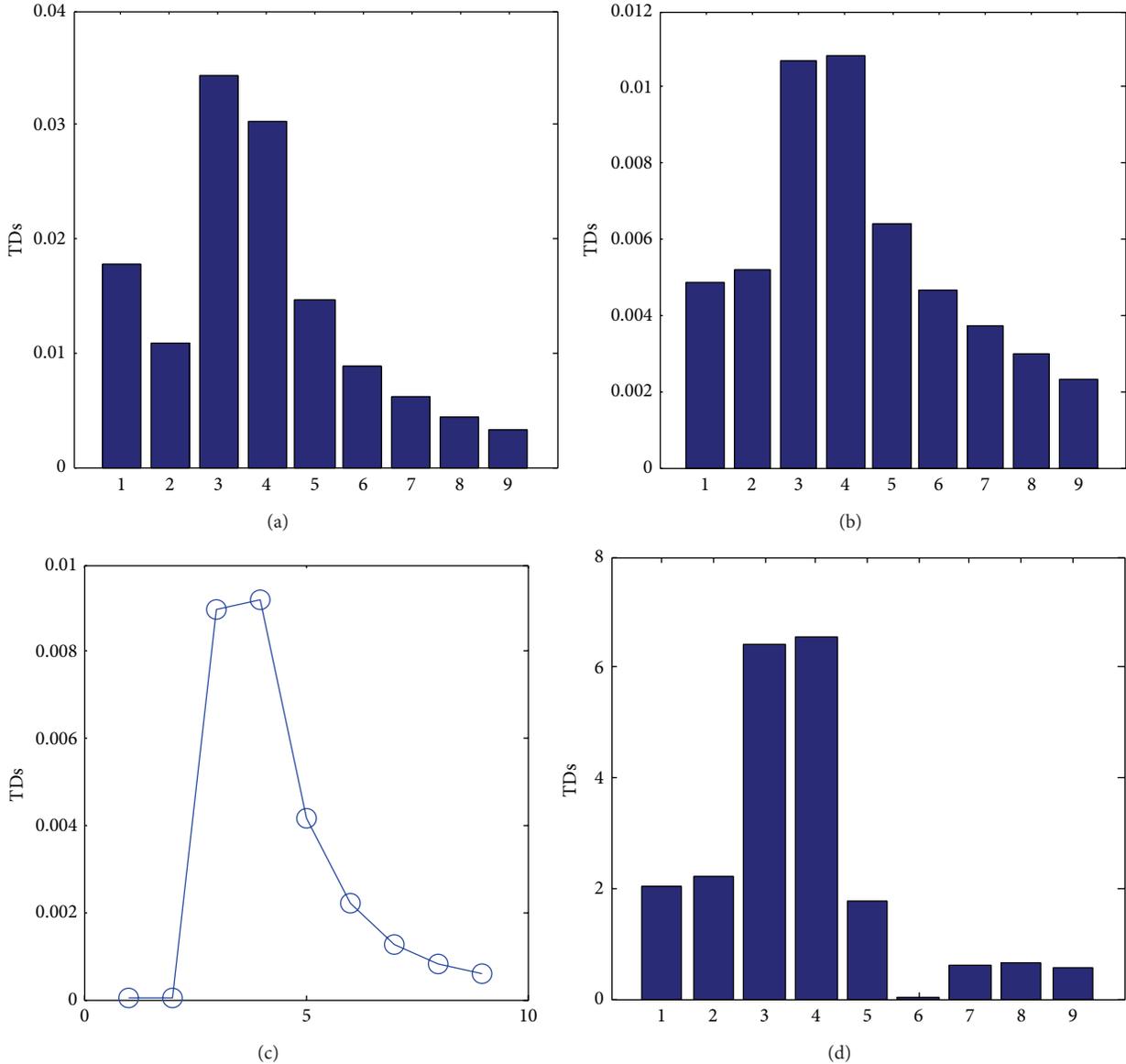


FIGURE 12: Detection indicators of the cantilever beam when  $\gamma = 0.2$  with frequency band of 0–100 Hz: (a) this paper, (b) Schulz et al. [4], (c) magnitude value of Zhang et al. [6], and (d) Johnson and Adams [9].

4.2.1. *Effect of Frequency Bands.* The effect of frequency bands is studied in this application. Considering the natural frequency in Table 2, and three different frequency bands, 0–10 Hz, 0–50 Hz, and 0–100 Hz, are used to compute the indicators. The crack depth ratio  $\gamma$  is 0.2, and stiffness reduction of the beam element is no more than 8.5%.

Figure 10 shows the indicators of the cantilever beam for frequency band of 0–10 Hz. It can be seen that the method proposed by this paper locates the crack position correctly. For indicators proposed by Zhang et al. [6], the magnitude value gives the complete wrong location. While the conventional method by Schulz et al. [4] and quotients of transmissibility ratios by Johnson and Adams [9] also misdetect the crack for this frequency band, the indicator of the fourth element is maximum, which is near the crack position.

Figures 11 and 12 show the detection indicators for frequency bands of 0–50 Hz and 0–100 Hz, separately. Besides the indicators proposed by this paper, the other three methods still cannot locate the crack all the time. But all these indicators give the crack position very close to the right location. For all these three frequency bands, the method proposed by this paper locates the crack correctly, which means that this method is more robust to the frequency band compared to the other three methods.

4.2.2. *Effect of Noise.* Besides the frequency band, measurement inaccuracy is another influence factor to the transmissibility based detection indicators. In this section, Gaussian noise is added on response directly, and the sensitivity to noise of the method is studied. Once again, scenario with

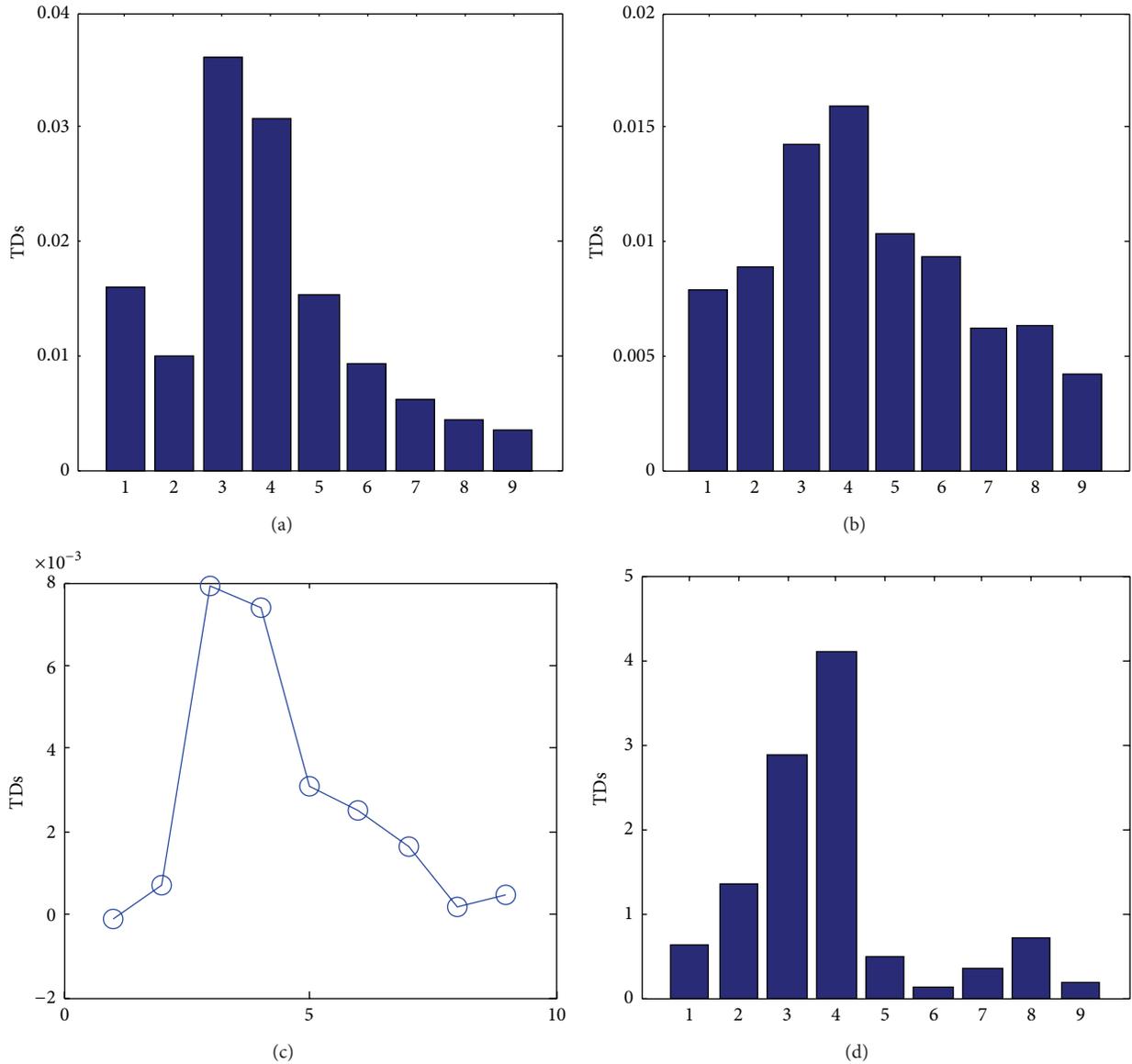


FIGURE 13: Detection indicators of the cantilever beam when  $\gamma = 0.2$  for frequency band of 0–50 Hz with 60 dB: (a) this paper, (b) Schulz et al. [4], (c) magnitude value of Zhang et al. [6], and (d) Johnson and Adams [9].

crack depth ratio of 0.2 is selected. According to the study of frequency band, 0–50 Hz is set as the suitable frequency band.

Three levels of noise, 60 dB, 40 dB, and 20 dB, are taken into consideration. When the noise is low, detection indicators change slightly in Figure 13 except for method proposed by Johnson and Adams [9]. The indicators of methods proposed by this paper and Zhang et al. [6] all give the right crack position, and the conventional method proposed by Schulz et al. [4] locates the fourth element, which is close to the crack. And when the noise grows in Figures 14 and 15, only the indicators of methods proposed by this paper and Zhang et al. [6] can give the right position or position near

the crack. The indicators by Schulz et al. [4] and Johnson and Adams [9] are completely disordered, and the crack is totally misdirected. Even when the noise is 20 dB, the method proposed by this paper still can locate the crack correctly. Therefore, the proposed method in this paper can decrease the influence of noise using weight function compared with the other three methods.

## 5. Conclusions

In this paper, the effects of the local variation on the TFs are studied, and the results show that TFs between the position of local variation and input change obviously, while the others

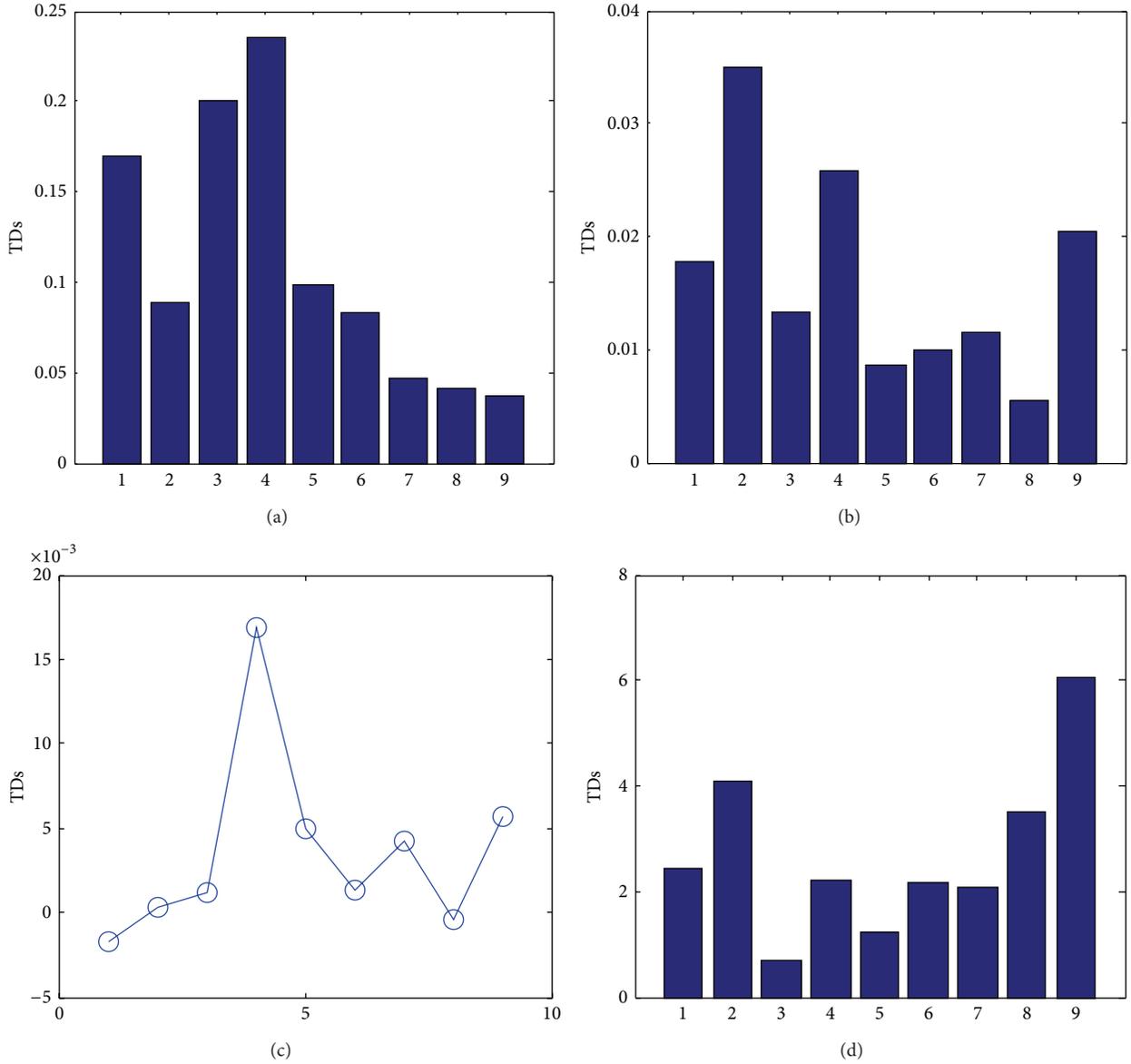


FIGURE 14: Detection indicators of the cantilever beam when  $\gamma = 0.2$  for frequency band of 0–50 Hz with 40 dB: (a) this paper, (b) Schulz et al. [4], (c) magnitude value of Zhang et al. [6], and (d) Johnson and Adams [9].

remain constant. Based on these properties, the differences between the transmissibility functions before and after the variation are used to develop a new detection indicator.

Through the simulation of linear, bilinear, and nonlinear one-dimensional MDOF system, the indicator was validated. And the detection indicators were applied in the crack detection in beams structures. From the study of the properties and the application of TFs, it was found that TFs are influenced by many factors, such as the local restoring force, location of external force, boundary, and frequency band. Even for a defined system, it is not easy to determine proper frequency bands to avoid the influence of the subordinate factors. But this paper indeed shows the advantage of the new indicator

in local variation detection through the comparison with the other three indicators based on transmissibility functions. The new indicator is more robust to the frequency bands and noise.

Based on the study above, the authors will focus on reducing the influence of these factors on the transmissibility in one-dimensional and multidimensional MDOF system in the future.

**Conflict of Interests**

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

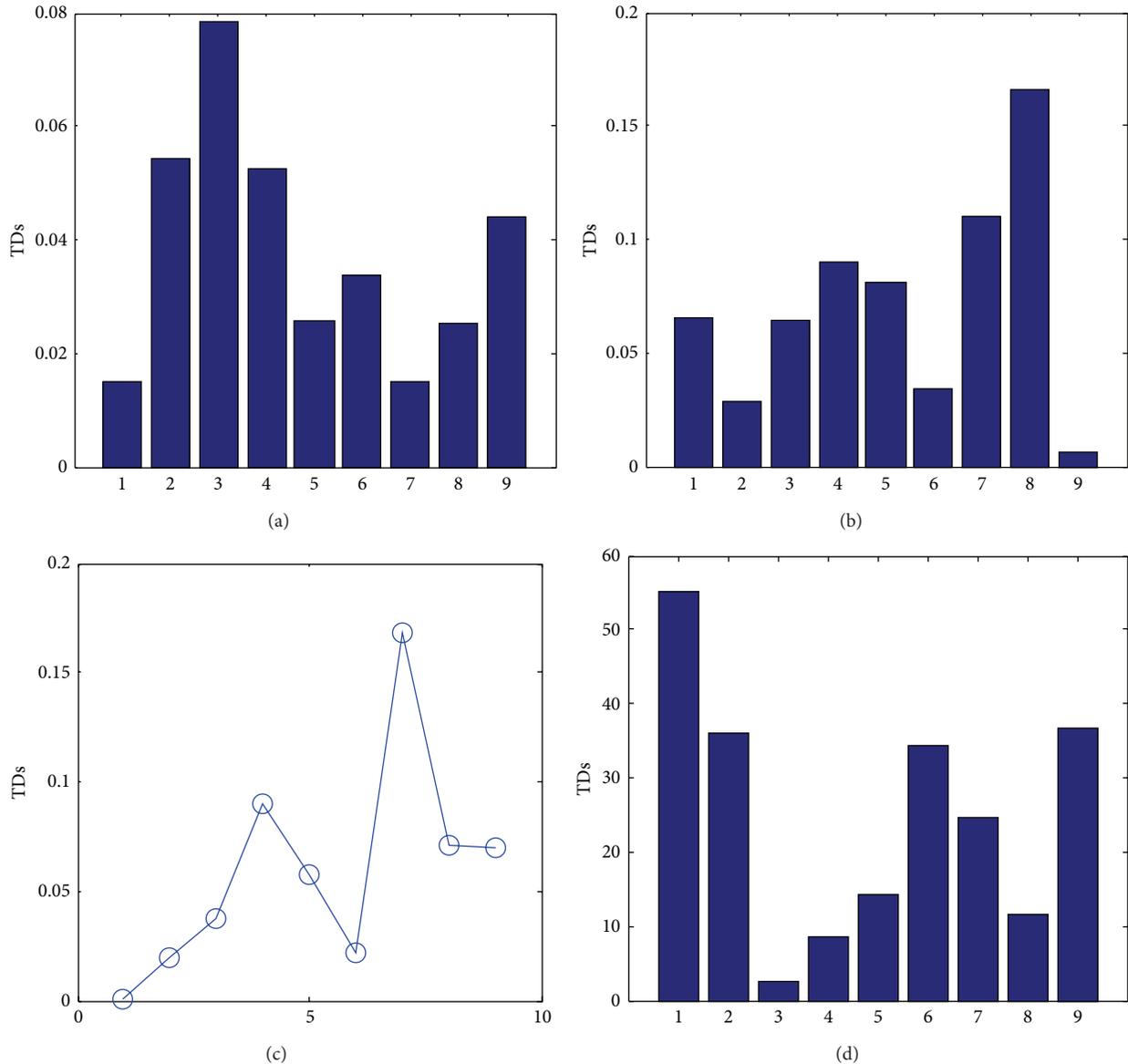


FIGURE 15: Detection indicators of the cantilever beam when  $\gamma = 0.2$  for frequency band of 0–50 Hz with 20 dB: (a) this paper, (b) Schulz et al. [4], (c) magnitude value of Zhang et al. [6], and (d) Johnson and Adams [9].

## Acknowledgment

The authors gratefully acknowledge that this paper was supported by the NSFC for Distinguished Young Scholars (11125209) and the NSFC (11472170, 51421092).

## References

- [1] W. Fan and P. Qiao, "Vibration-based damage identification methods: a review and comparative study," *Structural Health Monitoring*, vol. 10, no. 1, pp. 83–111, 2011.
- [2] 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 ID 122001, 2008.
- [3] Z. Q. Su, L. Ye, and Y. Lu, "Guided Lamb waves for identification of damage in composite structures: a review," *Journal of Sound and Vibration*, vol. 295, no. 3–5, pp. 753–780, 2006.
- [4] M. J. Schulz, A. S. Naser, P. F. Pai, M. S. Linville, and J. Chung, "Detecting structural damage using transmittance functions," in *Proceedings of the 15th International Modal Analysis Conference (IMAC '97)*, pp. 638–644, 1997.
- [5] S. Chesné and A. Deraemaeker, "Damage localization using transmissibility functions: a critical review," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 569–584, 2013.
- [6] H. Zhang, M. J. Schulz, A. Naser, F. Ferguson, and P. F. Pai, "Structural health monitoring using transmittance functions," *Mechanical Systems and Signal Processing*, vol. 13, no. 5, pp. 765–787, 1999.

- [7] J. E. Mottershead, "On the zeros of structural frequency response functions and their sensitivities," *Mechanical Systems and Signal Processing*, vol. 12, no. 5, pp. 591–597, 1998.
- [8] Q. Chen, Y. W. Chan, and K. Worden, "Structural fault detection using neural networks trained on transmissibility functions," in *Proceedings of the International Conference on Vibration Engineering*, pp. 446–456, Beijing, China, 1994.
- [9] T. J. Johnson and D. E. Adams, "Transmissibility as a differential indicator of structural damage," *Journal of Vibration & Acoustics*, vol. 124, no. 4, pp. 634–641, 2002.
- [10] C. Devriendt and P. Guillaume, "Identification of modal parameters from transmissibility measurements," *Journal of Sound and Vibration*, vol. 314, no. 1-2, pp. 343–356, 2008.
- [11] W.-J. Yan and W.-X. Ren, "Operational modal parameter identification from power spectrum density transmissibility," *Computer-Aided Civil & Infrastructure Engineering*, vol. 27, no. 3, pp. 202–217, 2012.
- [12] N. M. M. Maia, R. A. B. Almeida, A. P. V. Urgueira, and R. P. C. Sampaio, "Damage detection and quantification using transmissibility," *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2475–2483, 2011.
- [13] J. Li and H. Hao, "Damage detection of shear connectors based on power spectral density transmissibility," *Key Engineering Materials*, vol. 569-570, pp. 1241–1248, 2013.
- [14] J. Li, H. Hao, Y. Xia, and H.-P. Zhu, "Damage detection of shear connectors in bridge structures with transmissibility in frequency domain," *International Journal of Structural Stability & Dynamics*, vol. 14, no. 2, Article ID 1350061, pp. 634–640, 2014.
- [15] J. Li, H. Hao, and J. V. Lo, "Structural damage identification with power spectral density transmissibility: numerical and experimental studies," *Smart Structures & Systems*, vol. 15, no. 1, pp. 15–40, 2015.
- [16] Y. L. Zhou, E. Figueiredo, N. Maia, and R. Perera, "Damage detection and quantification using transmissibility coherence analysis," *Shock and Vibration*, vol. 2015, Article ID 290714, 16 pages, 2015.
- [17] Y. L. Zhou, E. Figueiredo, N. Maia, R. Sampaio, and R. Perera, "Damage detection in structures using a transmissibility-based Mahalanobis distance," *Structural Control & Health Monitoring*, 2015.
- [18] H. Jeffreys and B. Swirles, *Methods of Mathematical Physics*, Cambridge University Press, Cambridge, UK, 1999.
- [19] J. L. Schiff, *The Laplace Transform: Theory and Applications*, Springer, New York, NY, USA, 1999.
- [20] P. Cawley and R. D. Adams, "The location of defects in structures from measurements of natural frequencies," *Journal of Strain Analysis for Engineering Design*, vol. 14, no. 2, pp. 49–57, 1979.
- [21] A. K. Pandey and M. Biswas, "Damage detection in structures using changes in flexibility," *Journal of Sound & Vibration*, vol. 169, no. 1, pp. 3–17, 1994.
- [22] A. Rivola and P. R. White, "Bispectral analysis of the bilinear oscillator with application to the detection of fatigue cracks," *Journal of Sound & Vibration*, vol. 216, no. 5, pp. 889–910, 1998.
- [23] J. N. Sundermeyer and R. L. Weaver, "On crack identification and characterization in a beam by non-linear vibration analysis," *Journal of Sound and Vibration*, vol. 183, no. 5, pp. 857–871, 1995.
- [24] Y. C. Chu and M.-H. H. Shen, "Analysis of forced bilinear oscillators and the application to cracked beam dynamics," *AIAA Journal*, vol. 30, no. 10, pp. 2512–2519, 1992.
- [25] S. V. Walker, J.-Y. Kim, J. Qu, and L. J. Jacobs, "Fatigue damage evaluation in A36 steel using nonlinear Rayleigh surface waves," *NDT & E International*, vol. 48, pp. 10–15, 2012.
- [26] C. Bermes, J.-Y. Kim, J. Qu, and L. J. Jacobs, "Experimental characterization of material nonlinearity using Lamb waves," *Applied Physics Letters*, vol. 90, no. 2, Article ID 021901, 3 pages, 2007.
- [27] S. S. Kessler, S. M. Spearing, and C. Soutis, "Damage detection in composite materials using Lamb wave methods," *Smart Materials & Structures*, vol. 11, no. 2, pp. 269–278, 2002.
- [28] W. Li, Y. Cho, T. Ju, H. Choi, N. Kim, and I. Park, "Evaluation of material degradation of composite laminates using nonlinear Lamb wave," in *Nondestructive Testing of Materials and Structures*, pp. 593–598, Springer, Amsterdam, The Netherlands, 2013.
- [29] M. I. Friswell and J. E. T. Penny, "Crack modeling for structural health monitoring," *Structural Health Monitoring*, vol. 1, no. 2, pp. 139–148, 2002.
- [30] A. Khorram, M. Rezaeian, and F. Bakhtiari-Nejad, "Multiple cracks detection in a beam subjected to a moving load using wavelet analysis combined with factorial design," *European Journal of Mechanics—A/Solids*, vol. 40, no. 1, pp. 97–113, 2013.
- [31] G. M. Owolabi, A. S. J. Swamidass, and R. Seshadri, "Crack detection in beams using changes in frequencies and amplitudes of frequency response functions," *Journal of Sound & Vibration*, vol. 265, no. 1, pp. 1–22, 2003.
- [32] A. O. Ayhan, "Simulation of three-dimensional fatigue crack propagation using enriched finite elements," *Computers & Structures*, vol. 89, no. 9-10, pp. 801–812, 2011.
- [33] M. Götting and J. Rösler, "A finite element creep crack growth model for waspaloy," *Computers & Structures*, vol. 85, no. 3-4, pp. 225–232, 2007.
- [34] J. K. Sinha, M. I. Friswell, and S. Edwards, "Simplified models for the location of cracks in beam structures using measured vibration data," *Journal of Sound & Vibration*, vol. 251, no. 1, pp. 13–38, 2002.

## Research Article

# Gearbox Fault Identification and Classification with Convolutional Neural Networks

ZhiQiang Chen,<sup>1</sup> Chuan Li,<sup>1,2</sup> and René-Vinicio Sanchez<sup>3</sup>

<sup>1</sup>Chongqing Engineering Laboratory for Detection, Control and Integrated System, School of Computer Science and Information Engineering, Chongqing Technology and Business University, Chongqing 400067, China

<sup>2</sup>Research Center of System Healthy Maintenance, Chongqing Technology and Business University, Chongqing 400067, China

<sup>3</sup>Department of Mechanical Engineering, Universidad Politécnica Salesiana, Cuenca, Ecuador

Correspondence should be addressed to Chuan Li; [chuanli@21cn.com](mailto:chuanli@21cn.com)

Received 11 March 2015; Revised 20 April 2015; Accepted 24 April 2015

Academic Editor: Dong Wang

Copyright © 2015 ZhiQiang Chen 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.

Vibration signals of gearbox are sensitive to the existence of the fault. Based on vibration signals, this paper presents an implementation of deep learning algorithm convolutional neural network (CNN) used for fault identification and classification in gearboxes. Different combinations of condition patterns based on some basic fault conditions are considered. 20 test cases with different combinations of condition patterns are used, where each test case includes 12 combinations of different basic condition patterns. Vibration signals are preprocessed using statistical measures from the time domain signal such as standard deviation, skewness, and kurtosis. In the frequency domain, the spectrum obtained with FFT is divided into multiple bands, and the root mean square (RMS) value is calculated for each one so the energy maintains its shape at the spectrum peaks. The achieved accuracy indicates that the proposed approach is highly reliable and applicable in fault diagnosis of industrial reciprocating machinery. Comparing with peer algorithms, the present method exhibits the best performance in the gearbox fault diagnosis.

## 1. Introduction

Gearboxes play crucial roles in the mechanical transmission systems, are used to transmit power between shafts, and are expected to work 24 hours a day in the production system. Any failures with the gearboxes may introduce unwanted downtime, expensive repair, and even human casualties. Therefore it is essential to detect and diagnose faults in the initial stage [1–4]. As an effective component for the condition-based maintenance, the fault diagnosis has gained much attention for the safe operations of the gearboxes [5, 6].

Machine fault identification can be done with different methodologies such as vibration signature analysis, lubricant signature analysis, noise signature analysis, and temperature monitoring. The gearbox conditions can be reflected by such measurements as vibratory, acoustic, thermal, electrical, and oil-based signals [7–12]. Of the above the diagnostic by vibrations is the most employed for the reason that every machine is considered to have a normal spectrum until there is a fault, where the spectrum changes [13, 14]. The vibration signals

have been proven effective to reflect the healthy condition of the rotating machinery. In the vibration-based gearbox fault diagnostics, Wang et al. [15] proposed the application of local mean decomposition of the vibration signal to diagnose a low-speed helical gearbox. Hong et al. [16] investigated the vibration measurements for the planetary gearbox fault detection. The vibration characteristics in both the time and the frequency domains were analyzed by Lei et al. [17] for the diagnostics of the planetary gearboxes.

Various studies exist, of algorithms for detection and diagnostics of faults in gearboxes; among these are support vector machines and artificial neural network. A support vector machines based envelope spectrum was proposed by Guo et al. [18] to classify three health conditions of the planetary gearboxes. An intelligent diagnosis model based on wavelet support vector machine (SVM) and immune genetic algorithm (IGA) was proposed for the gearbox fault diagnosis [19]. The IGA was developed to determine the optimal parameters for the wavelet SVM with the highest accuracy and generalization ability. Tayarani-Bathaie et al. [20]

suggested a dynamic neural network to diagnose the gas turbine fault. The artificial neural network combining with empirical mode decomposition was applied for automatic bearing fault diagnosis based on vibration signals [21]. Among all the typical classifiers, the support vector classification (SVC) family (i.e., the standard SVC and its variants) attracted much attention due to their extraordinary classification performance. According to the researches, the SVM family received good results in comparison with the peer classifiers.

Recently, deep learning received great success in the classification field. The deep learning gained better classification performance owing to its “deeper” representations for the faulty features. Up to now, different deep learning networks such as deep belief network [22], deep Boltzmann machines (DBMs) [23], deep autoencoder [24], and convolutional neural networks [25] have been introduced, but few been used for the fault diagnosis cases. Tran et al. [26] introduced the application of the deep belief networks to diagnose reciprocating compressor valves. Tamilselvan and Wang [27] employed the deep belief learning based health state classification for iris dataset, wine dataset, Wisconsin breast cancer diagnosis dataset, and *Escherichia coli* dataset. The limited reports used the deep learning structure for the fault diagnosis, with commonly one modality feature.

This paper presents a study for the application of the convolutional neural network in the identification and classification of the gearboxes fault. Convolutional neural network (CNN) is a type of feed-forward artificial neural network. Its individual neurons are tiled in such a way that they respond to overlapping regions in the visual field [28]. CNN and its variations are widely used models for image and video recognition [29, 30]. In this work, it is used as a classifier for the gearbox faults diagnosis.

The most successful methods of vibration-based fault diagnoses are composed of two main steps: extracting the sensitive features and classifying the condition patterns. In the vibration-based fault diagnosis, the most commonly used features have been generated from the temporal [31], spectral [32], wavelet [33], and other representations of the signals. Different representations can be regarded as different observations on the vibration signals [34]. In this work, statistical measurements such as standard deviation, skewness, and kurtosis are computed from the acquired time domain data. In the frequency domain, the spectrum obtained with a FFT is divided into multiple bands. The root mean square value is calculated for each band so the energy maintains its shape at the spectrum peaks. Vectors of the features of the preprocessed signal are formed, which are used as input parameters for the CNN. It is important to point out that the testing is performed under five different rotation frequencies and for each one four different load conditions are applied, which simulates the most likely scenario within an industrial application.

The rest of this paper is structured as follows. The CNN model and method of extracting statistical features are introduced in Section 2; Section 3 explains the mechanical conditions for the experiment; Section 4 presents the implementation of classifier based on the CNN model; and

Section 5 shows the obtained results and their evaluation. Finally some conclusions are drawn.

## 2. Methodologies

In this section, we first present the representations of the convolutional neural network. And then the approach of extracting the sensitive features is introduced, where some classical statistical parameters are calculated from the time and the frequency.

*2.1. Deep Learning with Convolutional Neural Network.* Convolutional neural network was inspired by the visual system's structure [35] and in particular by the models of it proposed by [36]. The first computational models are based on local connectivity between neurons and on hierarchically organized transformations of the image in Fukushima's neocognitron [37]. LeCun and collaborators, following up on this idea, designed and trained convolutional networks using the error gradient, where state-of-the-art performance was obtained [38, 39] on several pattern recognition tasks. Modern understanding of the physiology of the visual system is consistent with the processing style found in convolutional networks in the literature [40]. To this day, pattern recognition systems based on convolutional neural networks are among the best performing systems [41]. This has been shown clearly for handwritten character recognition [38], which has served as a machine learning benchmark for many years.

A typical convolutional neural network [38] is organized in layers of two types: convolutional layers and subsampling layers. Each layer has a topographic structure.

At each location of each layer, there are a number of different neurons. Each has its set of input weights that is associated with neurons in a rectangular patch in the previous layer. The same set of weights, but a different input rectangular patch, is associated with neurons at different locations.

Figure 1 presents the architecture of typical convolutional neural networks, in which the early analysis consists of alternating convolution and subsampling operations, while the last stage of the architecture consists of a generic multilayer network: the last few layers (closest to the outputs) will be fully connected 1-dimensional layers. CNNs work on the 2-dimensional data, so called maps, directly, unlike normal neural networks which would concatenate these into vectors. Typically convolutional layers are interspersed with subsampling layers to reduce computation time and to gradually build up further spatial and configural invariance. A small subsampling factor is desirable in order to maintain specificity at the same time.

Convolutional layers move forward with deriving the back propagation updates in a network, which compose feature maps by convolving kernels over feature maps in layers below them. At a convolution layer, the previous layer's feature maps are convolved with learnable kernels and put through the activation function to form the output feature map. Each output map may combine convolutions with

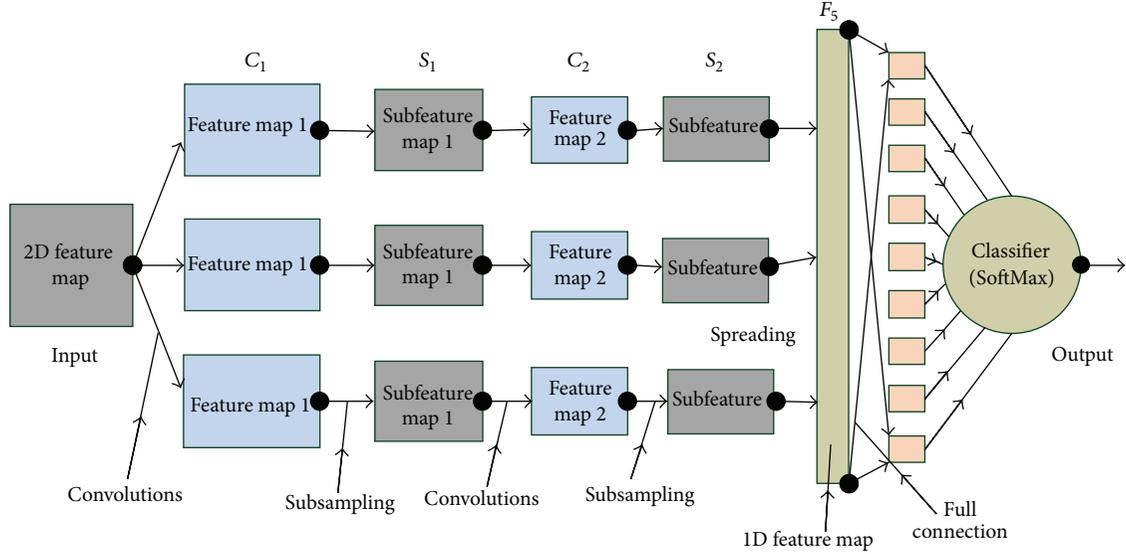


FIGURE 1: Architecture of convolutional neural networks.

multiple input maps. In general, it is calculated as follows [41]:

$$\mathbf{x}_j^l = f \left( \sum_{i \in M_j} \mathbf{x}_i^{l-1} * \mathbf{k}_{ij}^l + b_j^l \right), \quad (1)$$

where  $M_j$  represents a selection of input feature maps;  $l$  is the  $l$ th layer in a network,  $\mathbf{k}$  is a matrix of  $S \times S$ ; here,  $S$  is the size of convolutional kernels;  $f$  is a nonlinearity active function, typically hyperbolic tangent or sigmoid function. Each output map is given an additive bias; for a particular output map, the input maps will be convolved with distinct kernels  $\mathbf{k}$ . That is to say, if output map  $m$  and map  $n$  both sum over input map  $i$ , then the kernels applied to map  $i$  are different for output maps  $m$  and  $n$ .

A subsampling layer produces downsampled versions of the input maps. If there are  $N$  input maps, then, there will be exactly  $N$  output maps, although the output maps will be smaller. More formally [41],

$$\mathbf{x}_j^l = f \left( \beta_j^l \text{down}(\mathbf{x}_j^{l-1}) + b_j^l \right), \quad (2)$$

where  $\text{down}(\cdot)$  represents a subsampling function. Typically this function will sum over each distinct  $n$ -by- $n$  block in the input feature map so that the output feature map is  $n$  times smaller along both spatial dimensions. Each output map is given its own multiplicative bias  $\beta$  and an additive bias  $b$ .

To discriminate between  $T$  classes a fully connected output layer with  $T$  neurons is added. The output layer takes as input the concatenated feature maps of the layer below it, denoting the feature vector,  $f_v$ ,

$$O = f(b_o + w_o f_v), \quad (3)$$

where  $b_o$  is a bias vector and  $w_o$  is a weight matrix.

$\mathbf{k}_{ij}^l$ ,  $\beta_j^l$ ,  $b_o$ , and  $w_o$  of the model are learnable parameters. Learning is done using gradient descent which can be implemented efficiently using a convolutional implementation of

the backpropagation algorithm as shown in [41]. It should be clear that because kernels are applied over entire input maps, there are many more connections in the model than weights; that is, the weights are shared. This makes learning deep models easier, as compared to normal feedforward-backpropagation neural nets, as there are fewer parameters, and the error gradients goes to zero slower because each weight has greater influence on the final output.

## 2.2. Statistical Features of the Gearbox Vibration Signals.

The gearbox condition can be reflected by the information included in different features in frequency and time domain. From the set of signals obtained from the measurements of the vibrations at different speeds and loads, the features in frequency and time domain are obtained. From the group of graphs the values that can be used as input parameters for the CNN are selected. Sixty percent of the samples set are used for the training of the CNN, and forty percent are used for testing.

### 2.2.1. Time Data Statistical Features.

Usually, statistical parameters are good indices for extracting the condition information. In this research, statistical measurements such as standard deviation, skewness, and kurtosis for each node are used. Standard deviation, skewness, and kurtosis are computed from the acquired time domain data; the formulas used for this are shown in Table 1, where  $E(x)$  is the expected value of  $x$ . Correction bias is used for the evaluation of skewness and kurtosis. The standard deviation, skewness, and kurtosis evaluated on each of the vibration signals are used for training and testing of the CNN. The evaluation of these is done using standard MATLAB functions.

### 2.2.2. Fast Fourier Transform Banded RMS Value.

Figure 2 shows the vibration signal spectrum obtained during the test under the following condition patterns: gear  $Z_4$  with face

TABLE 1: Formula for the evaluation of statistical values.

Feature	Definition
Mean	$\bar{x} = \frac{1}{N} \sum_{n=1}^N x(n)$
Standard deviation	$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N (x(n) - \bar{x})^2}$
Skewness	$S = \frac{E(x - \bar{x})^3}{\sigma^3}$
Kurtosis	$K = \frac{1}{N} \sum_{n=1}^N \frac{(x(n) - \bar{x})^4}{\sigma^4}$

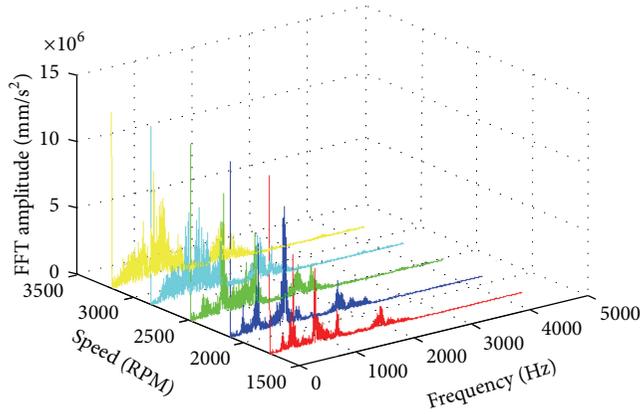


FIGURE 2: Frequency spectrum in function of the speed, under the following condition patterns: 375 W load,  $Z_1$  with face wear 0.4 [mm],  $Z_2$  with face wear 0.5 [mm],  $B_2$  with 2 pitting on inner ring, and  $B_3$  with 2 pitting on outer ring.

wear 0.4 [mm], gear  $Z_3$  with face wear 0.5 [mm], bear  $B_3$  with 2 pits on inner ring, and bear  $B_2$  with 2 pits on outer ring for 5 different rotation speeds, and load at 375 W. Figure 3 shows frequency spectrum under five combinations of different condition patterns. The time domain signal was multiplied by a Hanning window to obtain the FFT spectrum, in which a shift in the frequency and an increment in the amplitude in function of the speed increment are noticeable during the test. The different spectrum graphs showed that the amplitude of each component increases in a proportional manner to the load variation. Also in the spectra some accentuations and attenuations were observed on certain spectral component, which suggests dependency of the fault features with respect to the amount of load applied.

With the objective of reducing the amount of input data to the CNN the spectrum was split in multiple bands, because with this number of bands the root mean square (RMS) values keep track of the energy in the spectrum peaks [42], where the RMS value is evaluated with (4), and  $N$  is the number of samples of each frequency band. Consider

$$\text{FFT}_{\text{rms}} = \sum_{n=1}^N \text{FFT}(n). \quad (4)$$

Vectors of the features of the preprocessed signal are formed as input parameters for the CNN as follows:  $N_{\text{RMS}}$

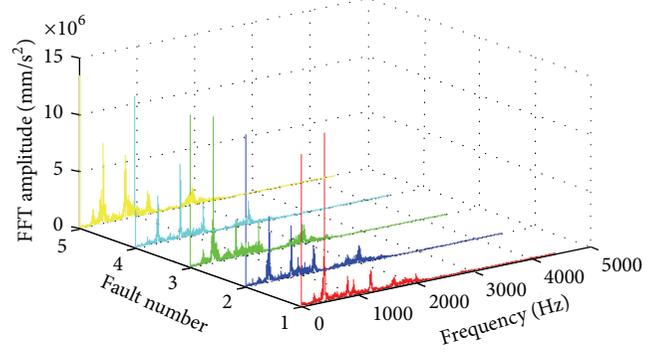


FIGURE 3: Frequency spectrum under five combinations of different condition patterns.

TABLE 2: Test's conditions.

Characteristic ( $C_1$ )	Value
Sample frequency	44100 [Hz] (16 bits)
Sampled time	10 [s]
Power	1000 [W]
Minimum speed	700 [RPM]
Maximum speed	1600 [RPM]
Minimum load	250 [W]
Maximum load	750 [W]
Speeds	1760, 2120, 2480, 2840, 3200 [mm/s]
Loads	375, 500, 625, 750 [W]
Number of loads per test	10
Type of accelerometer	Uniaxial
Trademark	ACS
Model	ACS 3411LN
Sensibility	330 [mV/g]

RMS values, standard deviation, skewness, kurtosis, rotation frequency, and applied load measurements. In this work, the frequency range is 0 to 22050 Hz and the size of the data vector in the frequency is 18000 samples. The spectrum is divided into  $N_{\text{RMS}}$  frequency bands,  $N_{\text{RMS}} = 251$ .

### 3. Experimental Setup

To validate the effectiveness of the proposed method, we carried out the experiments on a gearbox fault experimental platform. Figure 4 indicates the internal configuration of the gearbox and positions for accelerometers. There are 3 shafts and 4 gears composing a two-stage transmission of the gearbox. An input gear ( $Z_4 = 27$ , modulus = 2, and  $\Phi$  of pressure = 20) was installed on the input shaft. Two intermediate gears ( $Z_2 = Z_3 = 53$ ) were installed on an intermediate shaft for the transmission between the input gear and the output gear ( $Z_1 = 80$ , installed on the output shaft). The faulty components used in the experiments included gears  $Z_1$ ,  $Z_2$ ,  $Z_3$ , and  $Z_4$  and bearings  $B_1$ ,  $B_2$ ,  $B_3$ , and  $B_4$  as labeled in Figure 4(a). Test's conditions are described in the Table 2. The vibration signal is obtained from

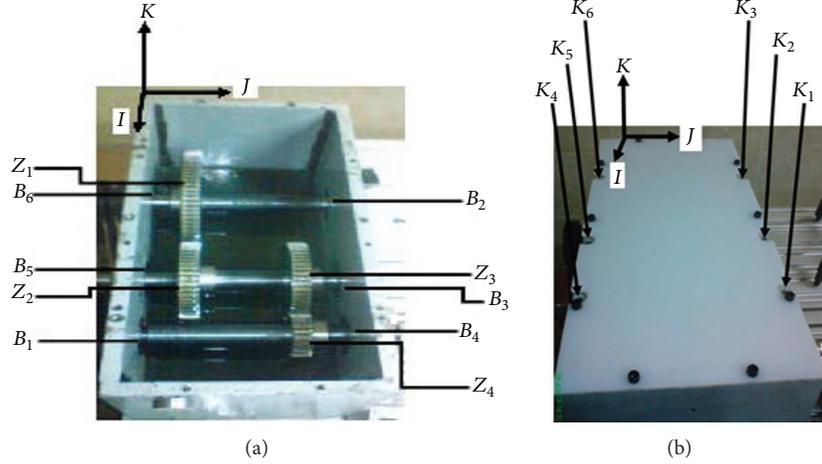


FIGURE 4: (a) The internal configuration of the gearbox; (b) positions for accelerometers.

the measurements of a vertically allocated accelerometer in the gearbox case. Tables 3 and 4 present the description of each fault condition of each component of the gearbox used in the experiment. We call them basic condition pattern. In our experiment, a test case includes several basic condition patterns, which is a combination of multiple component faults. For example, the test case A shown in Table 5 includes the following information of faults:

- Gear  $Z_4$ : gear with pitting on teeth.
- Gear  $Z_3$ : gear with face wear 0.5 [mm].
- Bear  $B_2$ : bearing with 4 pits on outer ring.
- Bear  $B_3$ : bearing with 2 pits on outer ring.
- Gears  $Z_1$  and  $Z_2$  and bears  $B_1$  and  $B_4$ : normal.

To evaluate the performance of the proposed method for gearbox fault diagnosis, first, we constructed 12 condition patterns as listed in Table 5. Each pattern with 4 different load conditions and 5 different input speeds was applied during the experiments. For each pattern, load and speed condition, we repeated the tests for 5 times. In each time of the test, the vibratory signals were collected with 24 durations each of which covered 0.4096 sec.

#### 4. Implemented Classifier Based on CNN and Statistical Features

In this section, the implementation of classifier based on CNN will be introduced. Figure 5 shows the block diagram of the process followed in the processing of the signal. The CNN-based classifier includes parameters as follows:

- (1) The size of input feature map,  $S_{in}$ , depends on the feature representation of the preprocessed signal.
- (2) The number of alternating convolution and subsampling layers that decides the architecture of CNN is as follows. Two schemes are investigated: one is two convolutional layers and two subsampling layers; another is one convolutional layer and one subsampling layer.
- (3) The number of output feature maps of convolution layer,  $C_i$ , expresses  $i$ th convolution layer.  $n_{C_i}$  is as the number of output feature maps of  $C_i$ .
- (4) The scale of subsampling layer,  $S$ , which means the size of output feature map of subsampling layer, is  $1/S^2$  of that of the input feature map.  $S_i$  expresses that of  $i$ th subsampling layer; and  $S_{S_i}$  is as the scale of layer  $S_i$ .

TABLE 3: Nomenclature of gears fault.

Designator	Description
1	Normal
2	Gear with face wear 0.4 [mm]
3	Gear with face wear 0.5 [mm]
4	Gear with chafing on tooth 50%
5	Gear with chafing on tooth 100%
6	Gear with pitting on tooth depth 0.05 [mm], width 0.5 [mm], and large 0.05 [mm]
7	Gear with pitting on teeth
8	Gear with incipient fissure on 4 mm teeth to 25% of profundity and angle of 45°
9	Gear teeth breakage 20%
10	Gear teeth breakage 50%
11	Gear teeth breakage 100%

TABLE 4: Nomenclature of bears fault.

Designator	Description
1	Normal
2	Bearing with 2 pits on outer ring
3	Bearing with 4 pits on outer ring
4	Bearing with 2 pits on inner ring
5	Bearing with 4 pits on inner ring
6	Bearing with race on inner ring
7	Bearing with 2 pits on ball
8	Bearing with 2 pits on ball

TABLE 5: Condition patterns of the experiment.

Number of patterns	Basic faults							
	Gear faults				Bear faults			
	$Z_4$	$Z_3$	$Z_2$	$Z_1$	$B_4$	$B_3$	$B_2$	$B_1$
A	7	3	1	1	1	2	3	1
B	7	3	6	8	1	1	1	1
C	5	5	1	1	6	7	2	1
D	7	1	1	1	6	7	2	1
E	1	2	1	1	1	6	3	1
F	1	3	1	1	1	5	3	1
G	2	9	1	1	6	7	3	1
H	5	5	1	1	6	3	2	4
I	2	6	1	1	6	5	2	1
J	1	11	1	1	1	3	4	1
K	1	1	1	1	1	6	3	1
L	1	1	1	1	1	1	3	1

(5) For each input map convolve with corresponding kernel and add to output map; the convolutional kernel is usually a matrix of  $k \times k$ , where  $k$  is called convolutional kernel size.

To confirm the optimal architecture of CNN-based classifier for gearbox fault diagnosis, some parameter tunings are performed. Table 6 presents 11 schemes with different parameters of the CNN-based classifier. They are applied to a test case with the 12 patterns indicated in Table 5, using data with 12000 sample signals, where sixty percent of the samples set are used for the training of the CNN, and forty percent are used for testing. The classification rate and computation time (Intel Core i7-4710MQ CPU @2.50 GHz 2.50 GHz, Memory 8.00 GB) of each epoch training are recorded in Table 6. From Table 6, we can assume that the cases of  $16 \times 16$  input feature map are superior to those of  $28 \times 28$ . The cases with one convolutional layer and one subsampling layer are superior to those of two convolutional layers and two subsampling layers. #7~#11 cases have very good classification accuracy. #9~#11 cases have less computation times. So we select a configuration for the proposed CNN-based classifier as follows: one convolutional layer and one subsampling layer,  $S_{in} = 16 \times 16$ ,  $n_{C_1} = 8$ ,  $S_{S_1} = 2$ , and  $k = 5$ . The suggested architecture of the CNN-based classifier for gearbox fault diagnosis is described in Figure 6.

## 5. Experiment Evaluations

The training is done in first instance with the 12 patterns indicated in Table 5. The used data have 12000 sample signals, where sixty percent of the samples set are used for the training of the CNN, and forty percent are used for testing. For further tuning parameters, we consider the #9~#11 cases in Table 6 with different number of iteration epochs: 50, 100, 150, 200, 250, and 300, respectively. Table 7 indicates the classification rate for the first instance. As shown in Table 7, each parameters pair has excellent performance for the gearbox faults classification. The least classification rate

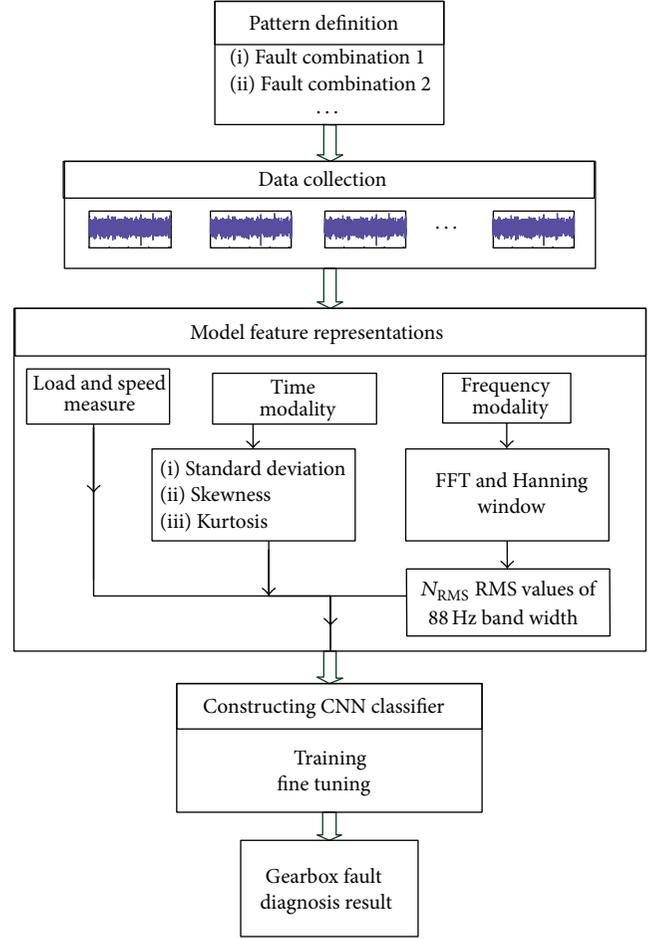


FIGURE 5: Training and testing process block diagram.

TABLE 6: Parameters tuning of the architecture of CNN.

Number	Architecture of CNN						Classification rate	Time (s/epoch)
	$S_{in}$	$n_{C_1}$	$S_{S_1}$	$n_{C_2}$	$S_{S_2}$	$k$		
#1	$28 \times 28$	6	2	12	2	5	86.73%	11.6 s
#2	$28 \times 28$	8	2	8	2	5	88.48%	12.8 s
#3	$28 \times 28$	12	2	12	2	5	92.50%	21.7 s
#4	$28 \times 28$	8	4	—	—	5	86.71%	8.00 s
#5	$16 \times 16$	6	2	12	2	5	90.23%	3.90 s
#6	$16 \times 16$	8	2	8	2	5	89.50%	3.80 s
#7	$16 \times 16$	6	2	6	1	5	95.71%	2.40 s
#8	$16 \times 16$	6	1	6	1	5	98.77%	4.50 s
#9	<b><math>16 \times 16</math></b>	<b>6</b>	<b>2</b>	—	—	<b>5</b>	<b>96.71%</b>	<b>1.04 s</b>
#10	<b><math>16 \times 16</math></b>	<b>8</b>	<b>2</b>	—	—	<b>5</b>	<b>98.35%</b>	<b>1.30 s</b>
#11	<b><math>16 \times 16</math></b>	<b>12</b>	<b>2</b>	—	—	<b>5</b>	<b>98.20%</b>	<b>2.02 s</b>

is 89.46% of the pair of  $n_{C_1} = 12$  and epochs = 50; the best one is 98.35% of the pair of  $n_{C_1} = 8$  and epochs = 200. In the following experiment,  $n_{C_1}$  and epochs are set to 8, 200, respectively.

Confusion matrix is an effective tool and is a visualization tool of the performance of a classification algorithm. Each

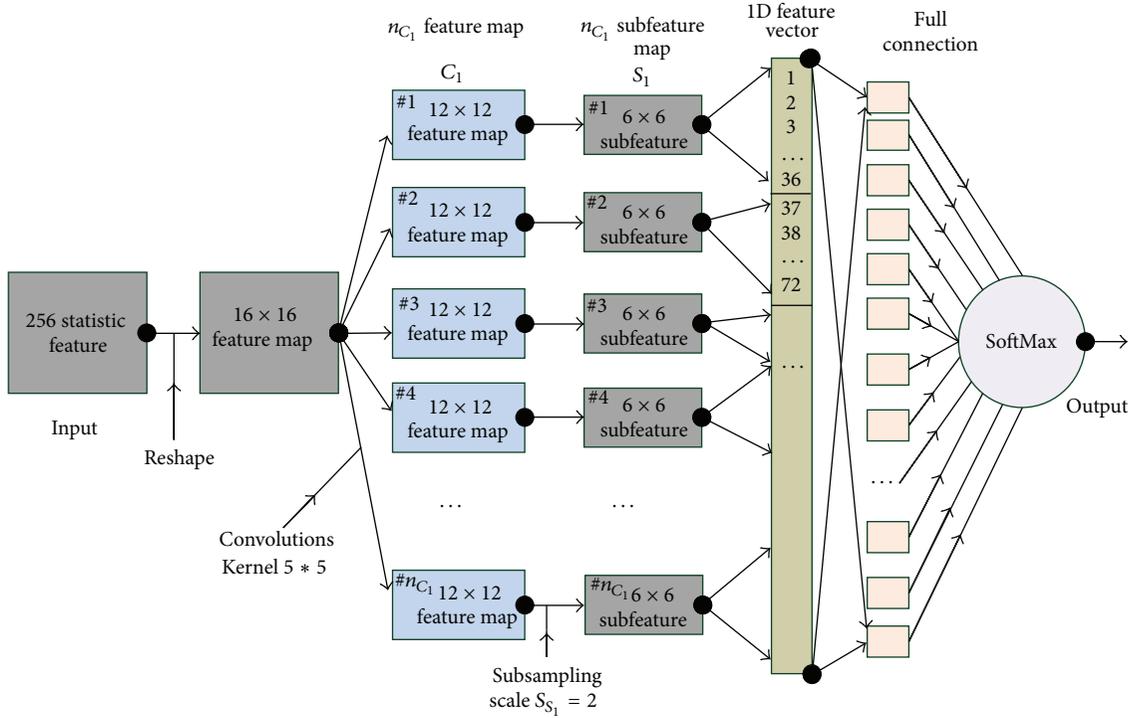


FIGURE 6: Implementation of classifier based on CNN and statistical features.

Output class	1	401	0	0	0	3	0	0	0	0	0	0	0	0	0	99.3%
	2	0	421	0	0	2	0	0	0	0	0	0	0	0	0	99.5%
	3	0	0	380	11	1	0	1	1	0	0	0	0	0	0	96.5%
	4	0	0	30	382	1	0	5	0	0	0	0	0	0	0	91.4%
	5	0	0	0	2	432	0	0	0	0	0	0	0	0	0	99.5%
	6	0	1	0	0	0	410	8	0	0	0	0	0	0	0	97.9%
	7	0	0	0	0	0	0	416	0	0	0	0	0	0	0	100%
	8	0	1	0	0	0	0	2	382	0	0	0	0	0	0	99.2%
	9	0	0	0	0	0	0	0	0	411	1	0	0	0	0	99.8%
	10	0	0	0	0	2	0	0	0	3	396	0	0	0	0	98.8%
	11	0	0	0	0	1	0	0	0	0	0	406	0	0	0	99.8%
	12	0	0	0	0	2	0	0	0	0	1	0	284	0	0	99.0%
		100%	99.5%	92.7%	96.7%	97.3%	100%	96.3%	99.7%	99.3%	99.5%	100%	100%	100%	98.4%	
		1	2	3	4	5	6	7	8	9	10	11	12			
		Target class														

FIGURE 7: Confusion matrix using CNN.

TABLE 7: Parameters tuning of CNN.

$n_{C1}$	Epochs					
	300	250	200	150	100	50
12	97.98%	98.0%	<b>98.02%</b>	96.71%	95.35%	89.46%
8	97.92%	98.19%	<b>98.35%</b>	97.98%	96.31%	91.19%
6	97.98%	97.27%	<b>96.71%</b>	96.71%	96.25%	93.1%

column of the confusion matrix represents the instances in a predicted class (output class), while each row represents

the instances in an actual class (target class). Figure 7 presents the confusion matrix using CNN model for 12 patterns indicated in Table 5. As shown in Figure 7, the global percentage of true positive classification of the 12 condition patterns of faults is 98.4% and the total error is 1.6%. The smallest percentage of true positive classifications is obtained for type 3; this is because this kind of conditions patterns with 6 basic faults. This is evident by observing the confusion matrix in which 30 times of type 4 are classified as type 3, noticing that mostly there is confusion between type 4 and type 3, in which they have 4 same basic faults. The percentages of true positive

1	320	8	2	36	12	0	0	14	12	0	0	0	79.2%
2	0	329	9	36	49	0	0	0	0	0	0	0	77.8%
3	0	0	222	150	18	0	4	0	0	0	0	0	56.4%
4	0	0	110	300	8	0	0	0	0	0	0	0	71.8%
5	0	0	35	85	313	0	1	0	0	0	0	0	72.1%
6	0	0	139	42	63	129	46	0	0	0	0	0	30.8%
7	0	0	162	82	25	7	140	0	0	0	0	0	33.7%
8	0	0	53	69	9	0	0	254	0	0	0	0	66.0%
9	0	0	0	0	0	0	0	0	412	0	0	0	100%
10	0	0	0	0	2	0	0	0	48	254	97	0	63.3%
11	0	0	0	0	1	0	0	0	9	0	397	0	97.5%
12	19	0	8	7	10	0	0	0	0	0	0	243	84.7%
	94.4%	97.6%	30%	37.2%	61.4%	94.9%	73.3%	94.8%	85.7%	100%	80.4%	100%	69.0%
	1	2	3	4	5	6	7	8	9	10	11	12	
	Target class												

FIGURE 8: Confusion matrix using SVM.

TABLE 8: Comparisons of classification rate with SVM using 20 test cases.

Number	1	2	3	4	5	6	7	8	9	10
CNN	91.4%	98.5%	97.1%	98.0%	98.4%	98.6%	97.2%	98.8%	97.4%	98.9%
SVM	73.8%	77.4%	65.9%	67.5%	65.5%	79.2%	69.1%	81.5%	66.8%	72.0%
Number	11	12	13	14	15	16	17	18	19	20
CNN	98.5%	98.7%	94.2%	97.7%	93.5%	98.7%	97.7%	94.7%	92.7%	97.4%
SVM	72.3%	66.4%	55.9%	64.7%	63.5%	67.0%	62.1%	61.3%	64.0%	60.9%
	Mean	Std.		Least			Most		Median	
CNN	96.8%	2.57%		91.4%			98.9%		97.7%	
SVM	67.8%	6.49%		55.9%			81.5%		66.8%	

classification of Type 1, Type 6, and Type 12 are all 100%. Confusion matrix in Figure 7 shows that the presenting CNN model has very high percentage of true positive classification.

To further validate the robustness of the present CNN model, a fault condition pattern library was constructed, which has 58 kinds of combinations based on the basis patterns described in Tables 3 and 4. 20 test cases are used to test the robustness of the present CNN method; each test case has 12 kinds of combinations that are randomly selected from the pattern library. The experiment results of 20 test cases using the CNN are shown in Table 8. With regard to the CNN method, its smallest classification rate is 91.4% of 1st test case, and the largest one reaches 98.9% of 10th test case. The mean, standard deviation, and median of classification rate using CNN are 96.8%, 2.57%, and 97.7%, respectively.

In addition, the CNN method was compared with “shallow” learning algorithms SVM. As for the SVM, one of the most important representatives in the “shallow” learning community, good classification results can be found for the gearbox fault diagnosis, which is similar with some existing researches (e.g., [43]). The algorithm SVM is applied using the LibSVM [44]. The parameters for SVM are chosen as  $C = 137.187$  and core (kernel) given by a radial basis  $K(u, v) = \exp(-\gamma|u - v|^2)$  function where  $\gamma = 1910.852$ .

These parameters were found through a search, aiming at the best model for the SVM. Figure 8 shows the confusion matrix using SVM method for 12 patterns indicated in Table 5. The experiment results of 20 test cases using the SVM method are shown in Table 8. Its global percentage of true positive classification of the 12 condition patterns of faults is only 69.0% and the total error is 31.0%. The smallest percentage of true positive classifications is 30% of type 3. The mean, standard deviation, and median of classification rate using SVM for the 20 test cases are 67.8%, 6.49%, and 66.8%, respectively. Comparing with the deep learning CNN method, the SVM method exhibits inferior performance for the gearbox fault diagnosis.

## 6. Conclusions

In this paper, a deep learning technique based CNN for the vibration measurements has been proposed to diagnose the fault patterns of the gearbox. The present CNN method identifies and classifies faults in gearbox by using the vibration signals measured with an accelerometer. Feature representations are selected as the input parameters of the CNN with a vector formed by RMS values, standard deviation, skewness, kurtosis, rotation frequency, and applied

load. For evaluating the proposed CNN method, the gearbox fault diagnosis experiments were carried out using different techniques. The results show that the present method has the outstanding performance of the gearbox fault diagnosis, comparing with peer methods. This type of classifiers could make a contribution to maintenance routines for industrial systems, towards lowering costs and guarantying a continuous production system, and, with the appropriate equipment, online diagnostics could be performed.

## Conflict of Interests

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

## Acknowledgments

This work is supported by Natural Science Foundation Project of CQ CSTC (nos. cstc2012jjA40041 and cstc2012jjA40059), Science Research Fund of Chongqing Technology and Business University (no. 2011-56-05), the National Natural Science Foundation of China (51375517), and the Project of Chongqing Innovation Team in University (KJTD201313).

## References

- [1] Y. Lei, M. J. Zuo, Z. He, and Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1419–1430, 2010.
- [2] Z. Yang, W. I. Hoi, and J. Zhong, "Gearbox fault diagnosis based on artificial neural network and genetic algorithms," in *Proceedings of the International Conference on System Science and Engineering (ICSSE '11)*, pp. 37–42, IEEE, Macao, China, June 2011.
- [3] A. Aherwar and M. S. Khalid, "Vibration analysis techniques for gearbox diagnostic: a review," *International Journal of Advanced Engineering Technology*, vol. 3, no. 2, 2012.
- [4] D. Wang, P. W. Tse, and K. L. Tsui, "An enhanced Kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 176–199, 2013.
- [5] D. Wang, P. W. Tse, W. Guo, and Q. Miao, "Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis," *Measurement Science and Technology*, vol. 22, no. 2, Article ID 025102, 2011.
- [6] C. Li and M. Liang, "Separation of the vibration-induced signal of oil debris for vibration monitoring," *Smart Materials & Structures*, vol. 20, no. 4, Article ID 045016, 2011.
- [7] S. Hou, M. Liang, Y. Zhang, and C. Li, "Vibration signal demodulation and bearing fault detection: a clustering-based segmentation method," *IMEche Proceedings Part C: Journal of Mechanical Engineering Science*, vol. 228, no. 11, pp. 1888–1899, 2014.
- [8] D. Wang, Q. Miao, and R. Kang, "Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition," *Journal of Sound and Vibration*, vol. 324, no. 3–5, pp. 1141–1157, 2009.
- [9] T. Toutountzakis, C. K. Tan, and D. Mba, "Application of acoustic emission to seeded gear fault detection," *NDT & E International*, vol. 38, no. 1, pp. 27–36, 2005.
- [10] J. R. Ottewill and M. Orkisz, "Condition monitoring of gearboxes using synchronously averaged electric motor signals," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 482–498, 2013.
- [11] A. M. D. Younus and B.-S. Yang, "Intelligent fault diagnosis of rotating machinery using infrared thermal image," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2082–2091, 2012.
- [12] C. Li and M. Liang, "Extraction of oil debris signature using integral enhanced empirical mode decomposition and correlated reconstruction," *Measurement Science and Technology*, vol. 22, no. 8, Article ID 085701, 2011.
- [13] P. Jayaswal, A. K. Wadhvani, and K. B. Mulchandani, "Machine fault signature analysis," *International Journal of Rotating Machinery*, vol. 2008, Article ID 583982, 10 pages, 2008.
- [14] 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.
- [15] Y. Wang, Z. He, J. Xiang, and Y. Zi, "Application of local mean decomposition to the surveillance and diagnostics of low-speed helical gearbox," *Mechanism and Machine Theory*, vol. 47, no. 1, pp. 62–73, 2012.
- [16] L. Hong, J. S. Dhupia, and S. Sheng, "An explanation of frequency features enabling detection of faults in equally-spaced planetary gearbox," *Mechanism and Machine Theory*, vol. 73, pp. 169–183, 2014.
- [17] Y. Lei, D. Kong, J. Lin, and M. J. Zuo, "Fault detection of planetary gearboxes using new diagnostic parameters," *Measurement Science and Technology*, vol. 23, no. 5, Article ID 055605, 2012.
- [18] L. Guo, J. Chen, and X. Li, "Rolling bearing fault classification based on envelope spectrum and support vector machine," *Journal of Vibration and Control*, vol. 15, no. 9, pp. 1349–1363, 2009.
- [19] F. Chen, B. Tang, and R. Chen, "A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm," *Measurement*, vol. 46, no. 1, pp. 220–232, 2013.
- [20] S. S. Tayarani-Bathaie, Z. N. Sadough Vanini, and K. Khorasani, "Dynamic neural network-based fault diagnosis of gas turbine engines," *Neurocomputing*, vol. 125, pp. 153–165, 2014.
- [21] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.
- [22] N.-N. Ji, J.-S. Zhang, and C.-X. Zhang, "A sparse-response deep belief network based on rate distortion theory," *Pattern Recognition*, vol. 47, no. 9, pp. 3179–3191, 2014.
- [23] B. Leng, X. Zhang, M. Yao, and Z. Xiong, "A 3D model recognition mechanism based on deep Boltzmann machines," *Neurocomputing*, vol. 151, pp. 593–602, 2015.
- [24] K. Katayama, M. Ando, and T. Horiguchi, "Model of MT and MST areas using an autoencoder," *Physica A: Statistical Mechanics and Its Applications*, vol. 322, pp. 531–545, 2003.
- [25] T. N. Sainath, B. Kingsbury, G. Saon et al., "Deep convolutional neural networks for multi-modality iso-intense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

- [26] V. T. Tran, F. Althobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4113–4122, 2014.
- [27] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering and System Safety*, vol. 115, pp. 124–135, 2013.
- [28] DeepLearning 0.1. LISA Lab, "Convolutional Neural Networks (LeNet)—DeepLearning 0.1 documentation," August 2013.
- [29] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," *Neural Networks*, vol. 16, no. 5-6, pp. 555–559, 2003.
- [30] Y. LeCun, "LeNet-5, convolutional neural networks," November 2013.
- [31] M. Amarnath and I. R. P. Krishna, "Local fault detection in helical gears via vibration and acoustic signals using EMD based statistical parameter analysis," *Measurement*, vol. 58, pp. 154–164, 2014.
- [32] K. K. McKee, G. L. Forbes, I. Mazhar, R. Entwistle, M. Hodkiewicz, and I. Howard, "A vibration cavitation sensitivity parameter based on spectral and statistical methods," *Expert Systems with Applications*, vol. 42, no. 1, pp. 67–78, 2015.
- [33] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.
- [34] C. Lim, M. Liang, and T. Wang, "Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals," *Mechanical Systems and Signal Processing*, 2015.
- [35] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–27, 2009.
- [36] 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.
- [37] K. Fukushima, "Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.
- [38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.
- [39] Y. LeCun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [40] T. Serre, G. Kreiman, M. Kouh, C. Cadieu, U. Knoblich, and T. Poggio, "A quantitative theory of immediate visual recognition," in *Computational Neuroscience: Theoretical Insights into Brain Function*, vol. 165 of *Progress in Brain Research*, pp. 33–56, Elsevier, 2007.
- [41] J. Bouvrie, "Notes on convolutional neural networks," Tech. Rep., MIT, 2006.
- [42] I. A. Abu-Mahfouz, "A comparative study of three artificial neural networks for the detection and classification of gear faults," *International Journal of General Systems*, vol. 34, no. 3, pp. 261–277, 2005.
- [43] D. L. de Souza, M. H. Granzotto, G. M. de Almeida, and L. C. Oliveira-Lopes, "Fault detection and diagnosis using support vector machines—a SVC and SVR comparison," *Journal of Safety Engineering*, vol. 3, no. 1, pp. 18–29, 2014.
- [44] C.-C. Chang and C.-J. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.

## Research Article

# Multifault Diagnosis of Rolling Element Bearings Using a Wavelet Kurtogram and Vector Median-Based Feature Analysis

**Phuong H. Nguyen and Jong-Myon Kim**

*School of Electrical, Electronics and Computer Engineering, University of Ulsan, Ulsan 680-749, Republic of Korea*

Correspondence should be addressed to Jong-Myon Kim; [jongmyon.kim@gmail.com](mailto:jongmyon.kim@gmail.com)

Received 3 July 2015; Accepted 18 August 2015

Academic Editor: Wahyu Caesarendra

Copyright © 2015 P. H. Nguyen and J.-M. Kim. 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 comprehensive multifault diagnosis methodology for incipient rolling element bearing failures. This is done by combining a wavelet packet transform- (WPT-) based kurtogram and a new vector median-based feature analysis technique. The proposed approach first extracts useful features that are characteristic of the bearing health condition from the time domain, frequency domain, and envelope power spectrum of incoming acoustic emission (AE) signals by using a WPT-based kurtogram. Then, an enhanced feature analysis approach based on the linear discriminant analysis (LDA) technique is used to select the most discriminant bearing fault features from the original feature set. These selected fault features are used by a Naïve Bayes (NB) classifier to classify the bearing fault conditions. The performance of the proposed methodology is tested and validated under various bearing fault conditions on an experimental test rig and compared with conventional state-of-the-art approaches. The proposed bearing fault diagnosis methodology yields average classification accuracies of 91.11%, 96.67%, 98.89%, 99.44%, and 98.61% at rotational speeds of 300, 350, 400, 450, and 500 rpm, respectively.

## 1. Introduction

For the past several decades, the development of reliable fault diagnosis systems to accurately detect and classify various bearing faults has been at the heart of research in the field of machine condition monitoring for preventive and predictive maintenance. These fault diagnosis systems aim to accurately detect bearing faults at the early stages of their development in order to prevent potential breakdown of industrial machines, thereby improving their reliability and reducing maintenance costs.

Several signal processing-based methods have been developed for fault diagnosis of rolling element bearings. These methods extract characteristic fault features from the inherently nonstationary signals obtained from faulty bearings. Antoni [1] presented the kurtogram, which has proven to be a powerful method for characterizing and extracting hidden nonstationary features in bearing fault signals [2–4]. The kurtogram method relies on spectral kurtosis, which was originally proposed by Dwyer [5], to detect the presence of nonstationary transients and accurately indicate their

location in the frequency bands of bearing signals. Recently, in attempts to enhance the performance of the kurtogram method for fault diagnosis, many researchers have integrated the kurtogram either with the short-time Fourier transform (STFT) [6] or with multirate filter banks (MRFB) [1]. These approaches, however, have yielded little improvement in extracting transient characteristics and have rendered kurtogram analysis more sensitive to irrelevant impulsive components [7]. These shortcomings have been addressed by introducing enhanced variations of the kurtogram, such as the one proposed by Shi et al. [8], which uses the complex Morlet wavelet transform. The complex Morlet wavelet transform, however, has its own limitations. These include its high computational complexity and its inability to effectively separate the high frequency components, which usually carry the characteristic fault information. In this paper, we propose a way to improve the kurtogram-based method by using the envelope power spectrum and the wavelet packet transform (WPT). The wavelet packet transform (WPT) is computationally more efficient and highly effective in frequency decomposition. The proposed kurtogram is performed by computing

kurtosis values from the envelope power spectrum of each subband signal at different levels of decomposition. The subband signal that yields the highest spectral kurtosis value is selected for extracting the features that are characteristic of the bearing fault conditions.

In addition to feature extraction techniques, feature analysis methods have also been widely studied in machine fault diagnosis in order to prevent the degradation of classification performance that is caused by redundant information in the feature space [9–13]. Dimensionality reduction techniques have been used to identify the features that effectively represent the high-dimensional data in a lower-dimensional space while retaining the intrinsic information of the bearing defects; these eventually lead to improvements in the classification performance of the fault diagnosis system. Among various dimensionality reduction methods, principle component analysis (PCA) [14], independent component analysis (ICA) [15], and linear discriminant analysis (LDA) [16] are the most popular methods that have been effectively utilized in machinery fault diagnosis. For instance, Widodo and Yang [11] applied the PCA technique to machine health prognostics to extract population characteristics from the available condition monitoring data. Guo et al. [12] investigated the performance of an ICA algorithm in separating the envelopes of vibration signals to extract the weak impulsive features of incipient faults in rolling element bearings. In their study looking into fault diagnosis in ball bearings, Harmouche et al. [13] demonstrated the use of an LDA technique to effectively discriminate the rather weak and otherwise difficult to detect/identify spectral features extracted from the envelope spectra of vibration signals. The LDA feature space yields better separation between different classes of bearing faults than conventional feature analysis methods, thereby resulting in improved classification performance. PCA and ICA are unsupervised methods that ignore the label information during dimensionality reduction, whereas LDA is a supervised method that utilizes the class label information to find an optimal low-dimensional representation of the original feature set while also preserving discriminant information between classes. As a result, LDA yields better classification results as compared to PCA and ICA [9, 13, 14]. This paper presents a new LDA-based feature analysis method that uses a vector median-based discriminant criterion (VMDC) to characterize the intraclass compactness and interclass separability of the feature space in order to select the most discriminative subset of bearing fault features.

This paper proposes a reliable multifault diagnosis scheme for rolling element bearings that combines an improved WPT-based kurtogram and the VMDC-based feature analysis methods. The contributions discussed in this paper can be summarized as follows:

- (1) An improved WPT-based kurtogram is proposed for extracting the most characteristic fault features from nonstationary acoustic emission (AE) signals. The kurtogram is performed by computing the spectral kurtosis values for the envelope power spectrum of each subband signal at different levels of decomposition. The subband signal that yields the highest

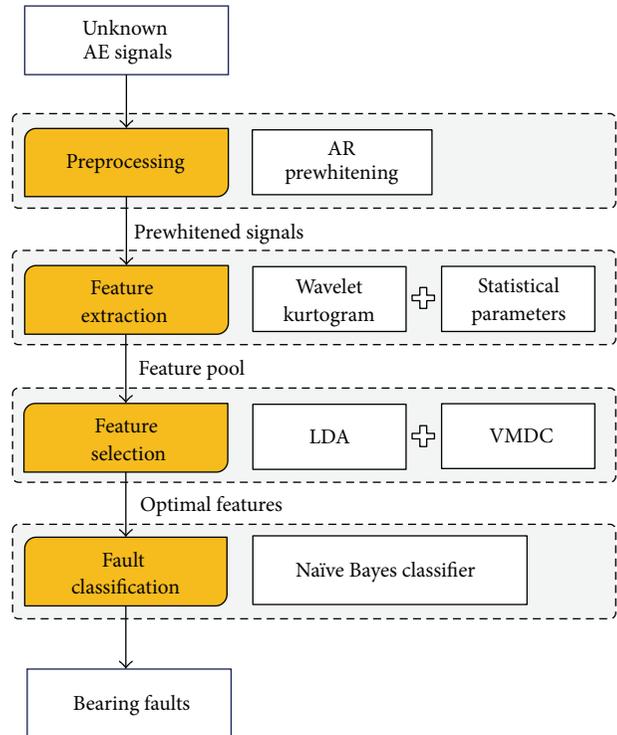


FIGURE 1: Flowchart of the proposed bearing fault diagnosis methodology.

spectral kurtosis value is selected for extracting the fault features.

- (2) An efficient VMDC-based discriminative fault feature analysis approach is proposed for selecting the most discriminative subset of fault features. A vector median-based discriminant criterion is presented to minimize the intraclass compactness and maximize the interclass separability of the feature space.
- (3) A comprehensive multiclass fault diagnosis methodology for rolling element bearings is presented, and its performance is validated using AE data for various single and compound bearing defects acquired under different simulated crack sizes and bearing rotational speeds.

The remainder of the paper is structured as follows. Section 2 describes the overall bearing fault diagnosis methodology, Section 3 introduces the self-designed fault simulator (including the data acquisition system and seeded bearing defects for collecting fault signals), and Section 4 shows the experimental results. Finally, Section 5 concludes this paper.

## 2. Proposed Bearing Fault Diagnosis Methodology

The proposed comprehensive bearing fault diagnosis scheme consists of four main processes including signal preprocessing, feature extraction, feature selection, and fault classification. Figure 1 illustrates the overall flowchart of the proposed methodology.

**2.1. Preprocessing.** The performance of wavelet kurtogram analysis can be improved if the variations between the peaks and flanks of the spectral energy are minimized as shown in Table 6; this can be achieved by preprocessing the AE signal before feature extraction [17]. In this study, the prewhitening technique is used to reduce the spectral energy variations in the incoming AE fault signal. To obtain the prewhitened signal, an autoregressive (AR) model  $y(n)$  is utilized, which is defined as follows:

$$y(n) = \sum_{i=1}^p a_i x(i+n) + e(n), \quad (1)$$

where  $x(n)$  is the incoming bearing signal,  $a_i$  are the AR coefficients,  $p$  denotes the order of the AR model, and  $e(n)$  is the residual signal representing a spectrum close to the white noise spectrum.

**2.2. Feature Extraction.** For the purpose of extracting the characteristic fault signatures from the AE signals, this paper presents a two-step feature extraction process that works as follows.

*Step 1.* In this step, the proposed WPT-based kurtogram method is performed on the prewhitened AE signal to determine its most informative subband. The flowchart of the proposed WPT-based kurtogram is illustrated in Figure 2, and its details are provided below.

First, the prewhitened bearing signal is decomposed into a series of subbands by performing the wavelet packet transform (WPT) with five decomposition levels ( $n = 5$ ) using a Daubechies 2 (or db2) filter. The five-level WPT decomposition results in a total of 63 subband signals. The envelope spectrum of each subband signal is then calculated using the fast Fourier transform (FFT) to identify the characteristic bearing fault frequencies. The envelope spectrum of an AE signal obtained from a defective bearing is usually flooded with the characteristic defect frequency and its harmonics. We utilize kurtosis to measure the degree of protrusion of the envelope spectrum and then characterize the hidden bearing defect signatures. In order to do this effectively, we first determine the bearing defect frequencies, which include the ball pass frequency of the bearing outer raceway (BPFO), the ball pass frequency of the bearing inner raceway (BPFI), the ball spin frequency of the bearing roller (BSF), and the first  $M$  harmonics ( $M = 3$  in this paper) of each of these frequencies. These defect frequencies are defined as follows:

$$\begin{aligned} \text{BPFO} &= \frac{N_r \cdot F_o}{2} \left( 1 - \frac{B_d}{P_d} \cos \alpha \right), \\ \text{BPFI} &= \frac{N_r \cdot F_o}{2} \left( 1 + \frac{B_d}{P_d} \cos \alpha \right), \\ \text{BSF} &= \frac{P_d \cdot F_o}{2 \cdot B_d} \left( 1 - \left( \frac{B_d}{P_d} \cos \alpha \right)^2 \right), \end{aligned} \quad (2)$$

where  $N_r$  denotes the number of cylindrical rollers,  $F_o$  is the shaft speed,  $\alpha$  is the contact angle, and  $B_d$  and  $P_d$  are

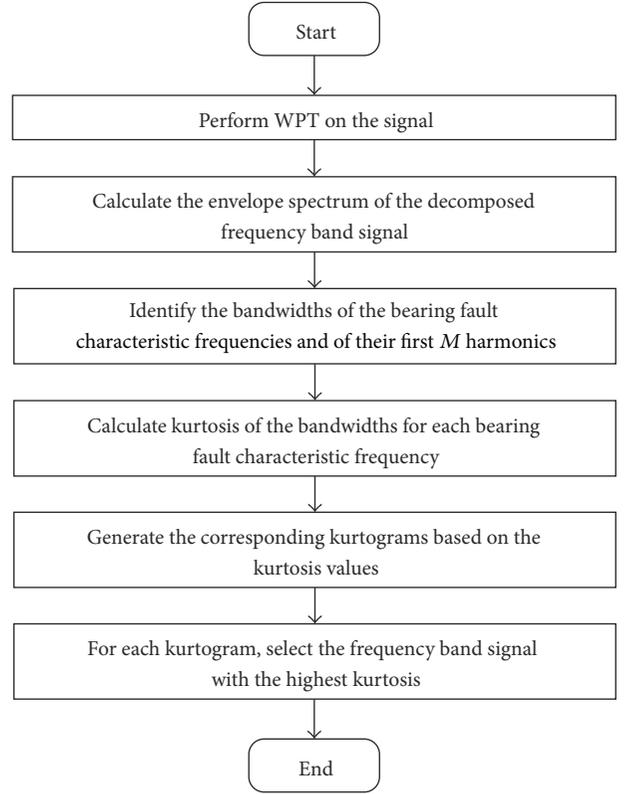


FIGURE 2: Flowchart of the wavelet kurtogram algorithm.

the roller and pitch diameters, respectively. Each defect-related bandwidth  $BW_i$  ranges from  $a_i$  to  $b_i$  ( $i = 1, 2, \dots, M$ ), which can be calculated as follows:

$$\begin{aligned} a_i &= \begin{cases} (i - \text{ER}) * \text{BPFO} & \text{if BPFO} \\ (i - \text{ER}) * \text{BPFI} - 2 * (1 + \text{ER}) * F_C & \text{if BPFI} \\ (i - \text{ER}) * \text{BSF} - 2 * (1 + \text{ER}) * F_C & \text{otherwise,} \end{cases} \\ b_i &= \begin{cases} (i + \text{ER}) * \text{BPFO} & \text{if BPFO} \\ (i + \text{ER}) * \text{BPFI} + 2 * (1 + \text{ER}) * F_C & \text{if BPFI} \\ (i + \text{ER}) * \text{BSF} + 2 * (1 + \text{ER}) * F_C & \text{otherwise.} \end{cases} \end{aligned} \quad (3)$$

Here, ER is the error rate of the bearing test rig and  $F_C$  is the cage frequency. Once the bearing fault bandwidths are identified, three different kurtosis values are calculated corresponding to the three bearing defect frequencies (i.e., BPFO, BPFI, and BSF). After obtaining the kurtosis values for the subband signals, the wavelet kurtogram is generated (Figure 3). Figure 3(a) shows the frequency-scale paving of the proposed wavelet kurtogram, while Figure 3(b) illustrates an example color kurtogram of the three bearing fault frequencies (i.e., BPFO, BPFI, and BSF), which represents the kurtosis values of all of the subbands.

Finally, for each kurtogram, the subband signal with the highest kurtosis value (i.e., the signal that has the most discriminative fault features) is selected for extracting the bearing fault signatures. In Figure 3(b), the selected subband

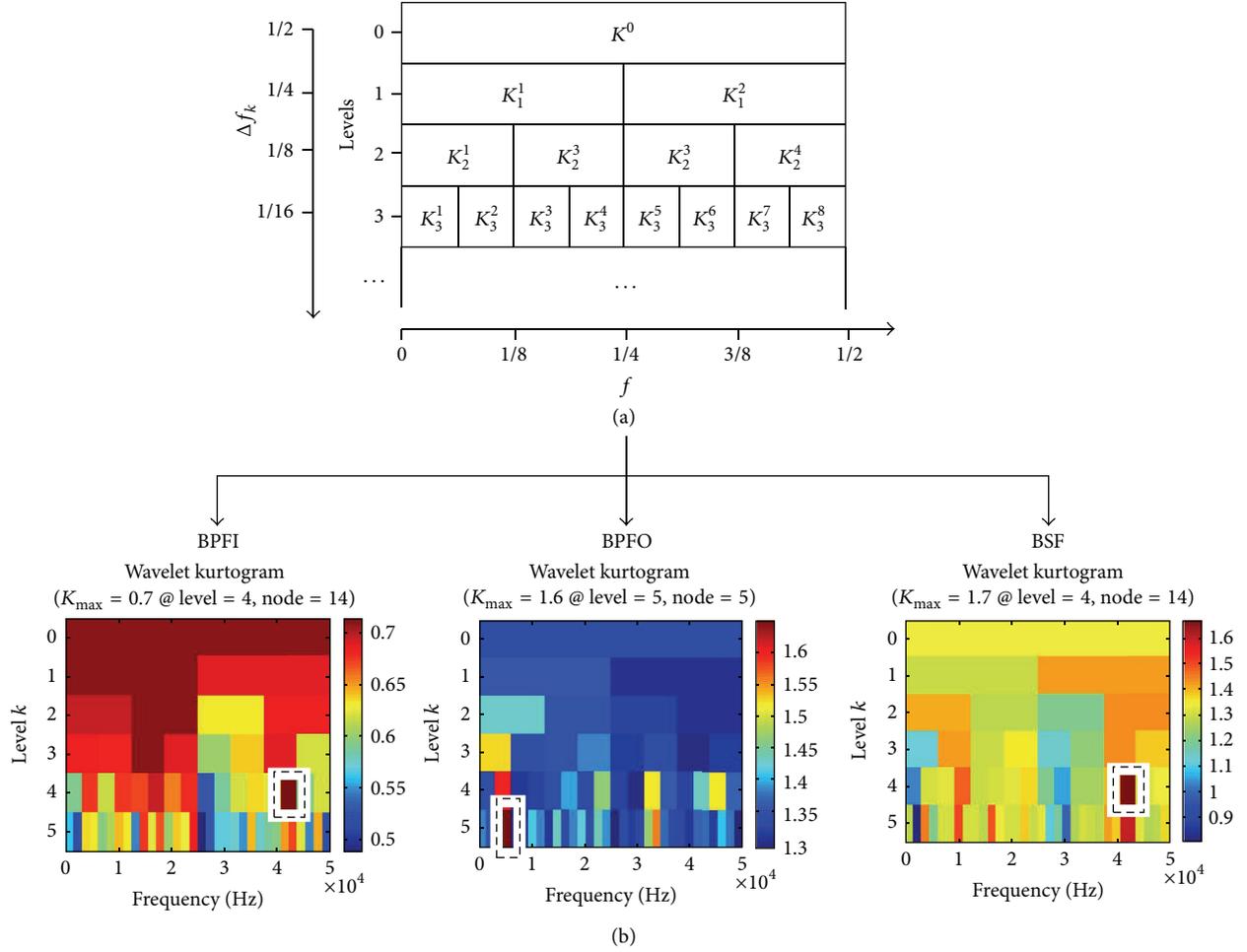


FIGURE 3: The proposed wavelet kurtogram. (a) The frequency-scale paving of the proposed wavelet kurtogram. (b) An example of a proposed wavelet kurtogram.

with the highest kurtosis value is highlighted using a dashed rectangle.

*Step 2.* In this step, 10 time domain ( $f_1, f_2, \dots, f_{10}$ ) and three frequency domain statistical features ( $f_{11}, f_{12}, f_{13}$ ) are extracted from each of the selected subband signals, as shown in Tables 1 and 2, respectively. Furthermore, we calculate three root mean square (RMS) features ( $f_{14}-f_{16}$ ) for the first three harmonics of the corresponding defect frequencies in the envelope spectrum of each of the selected subband signals, as described in Table 3. This results in a feature pool that consists of a total of  $N_c \times N_s \times N_f$  features, where  $N_c$  is the number of bearing fault classes,  $N_s$  is the number of AE signals obtained for each fault type, and  $N_f$  is the total number of features. In this case,  $N_f$  is  $10 + 3 + 3 = 16$  and includes 10 time domain statistical features, three frequency domain statistical features, and three envelope spectrum RMS features (calculated using the extracted wavelet kurtogram) for each of the three bearing defect frequencies (i.e., BPMFO, BPMFI, and BSF).

*2.3. Feature Selection.* In this paper, we propose a feature analysis method based on the linear discriminant analysis

(LDA) technique plus a vector median-based discriminant criterion (VMDC) to find the most discriminative subset of the extracted fault features for accurate fault diagnosis. LDA is a supervised dimensionality reduction algorithm that utilizes the class label information to find an optimal linear transformation matrix that projects the original high-dimensional feature space onto the lower-dimensional representation while also preserving the information that can help discriminate among different classes. Let  $X = \{x_1, x_2, \dots, x_N\} \in R^{D \times N}$  be a dataset, where  $D$  is the original dimensionality and  $N$  is the number of samples in  $X$ . Each sample  $x_i$  belongs to a class  $c_i = \{1, 2, \dots, C\}$ . Let  $N_i$  be the number of samples in class  $c_i$  and let  $N$  be the total number of samples in all of the classes. Then, the intraclass scatter matrix  $S_w$  and the interclass scatter matrix  $S_b$  can be evaluated with (4) and (5), respectively:

$$S_w = \sum_{i=1}^C \sum_{x_k \in C_i} (x_k - \mu_i)^T (x_k - \mu_i), \quad (4)$$

$$S_b = \sum_{i=1}^C (\mu_i - \mu)^T (\mu_i - \mu). \quad (5)$$

TABLE 1: Time domain statistical feature parameters.

Parameter	Definition
Root mean square ( $f_1$ )	$\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
Root mean square of the amplitude ( $f_2$ )	$\left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i }\right)^2$
Kurtosis ( $f_3$ )	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma}\right)^4$
Skewness ( $f_4$ )	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma}\right)^3$
Peak-to-peak ( $f_5$ )	$\max(x_i) - \min(x_i)$
Crest factor ( $f_6$ )	$\frac{\max( x_i )}{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}$
Impulse factor ( $f_7$ )	$\frac{(1/N) \sum_{i=1}^N  x_i }{\max( x_i )}$
Margin factor ( $f_8$ )	$\frac{\max( x_i )}{\left((1/N) \sum_{i=1}^N \sqrt{ x_i }\right)^2}$
Shape factor ( $f_9$ )	$\frac{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}{(1/N) \sum_{i=1}^N  x_i }$
Kurtosis factor ( $f_{10}$ )	$\frac{(1/N) \sum_{i=1}^N ((x_i - \bar{x})/\sigma)^4}{\left((1/N) \sum_{i=1}^N x_i^2\right)^2}$

TABLE 2: Frequency domain statistical feature parameters.

Parameter	Definition
Frequency center ( $f_{11}$ )	$\frac{1}{N} \sum_{i=1}^N f_i$
Root mean square frequency ( $f_{12}$ )	$\sqrt{\frac{1}{N} \sum_{i=1}^N f_i^2}$
Root variance frequency ( $f_{13}$ )	$\sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - X_{fc})^2}$

TABLE 3: Envelope spectrum statistical feature parameters.

Parameter	Definition
RMS frequency of the first harmonic ( $f_{14}$ )	$\sqrt{\frac{1}{(b_1 - a_1)} \sum_{i=a_1}^{b_1} f_i^2}$
RMS frequency of the second harmonic ( $f_{15}$ )	$\sqrt{\frac{1}{(b_2 - a_2)} \sum_{i=a_2}^{b_2} f_i^2}$
RMS frequency of the third harmonic ( $f_{16}$ )	$\sqrt{\frac{1}{(b_3 - a_3)} \sum_{i=a_3}^{b_3} f_i^2}$

Here,  $\mu_i = (1/N_i) \sum_{x_i \in c_i} x_i$  is the mean of the samples labelled as class  $c_i$  and  $\mu = (1/N) \sum_{i=1}^N x_i$  is the mean of all of the samples. The LDA projects the space of the original variables

onto a  $(C - 1)$ -dimensional space by maximizing the Fisher discriminant rule [18], which is defined as follows:

$$W^* = \arg \max \frac{W^T S_b W}{W^T S_w W}. \quad (6)$$

This discriminant rule maximizes the ratio of inter-class separability to intraclass compactness by selecting a reduced number of the most discriminative components only. However, LDA-based feature analysis approaches are limited in their ability to efficiently select the optimal number of discriminative components to gain the highest classification performance. In the existing literature, there is no clear consensus on the optimal number of LDA components that yield the highest classification performance. To address this problem, we present a robust method based on the vector median approach to obtain the optimal number of LDA components that can maximize the classification performance. In order to identify the vector median of each class, this study utilizes a cumulative distance criterion using the Euclidean distance. The vector median  $VM_i$  of class  $c_i$  is defined as the data point that yields the minimum accumulated distance and is determined as follows:

$$VM_i = \arg \min_{k \in c_i} \sum_{j \in c_i, j \neq k} \|x_k - x_j\|_2. \quad (7)$$

Once the vector medians of all classes are determined, the modified intraclass compactness and interclass separability values are calculated using (8) and (9), respectively:

$$S_w = \frac{1}{C} \sum_{i=1}^C \frac{1}{N_i} \sum_{j=1}^{N_i} \|VM_i - x_j\|_2, \quad (8)$$

$$S_b = \frac{1}{C} \sum_{i=1}^C \arg \min_{j \in C, j \neq i} \|VM_i - VM_j\|_2. \quad (9)$$

In (8), the intraclass compactness of a class is reflected by the distance from its vector median to other samples in that class. Alternatively, in (9), the interclass separability of a class from other classes is represented by the minimum distance of the vector median of that class to the vector medians of other classes. The effect of outliers in the feature space is minimized by averaging these distance measurements, which improves the discriminatory power of the proposed feature analysis method. The vector median-based discriminant criterion that is used to select the optimal number of useful LDA components is defined as follows:

$$K^* = \arg \max \frac{S_b}{S_w}. \quad (10)$$

**2.4. Fault Classification.** In this study, a Naïve Bayes classifier [19] is used to classify various single and compound bearing faults. This classifier is selected due to its simplicity and high classification performance at a relatively low computational cost [19, 20]. The Naïve Bayes classifier has complexity  $O(Nd)$ , where  $N$  is the total number of training samples and  $d$  is the number of conditional features. Other classification

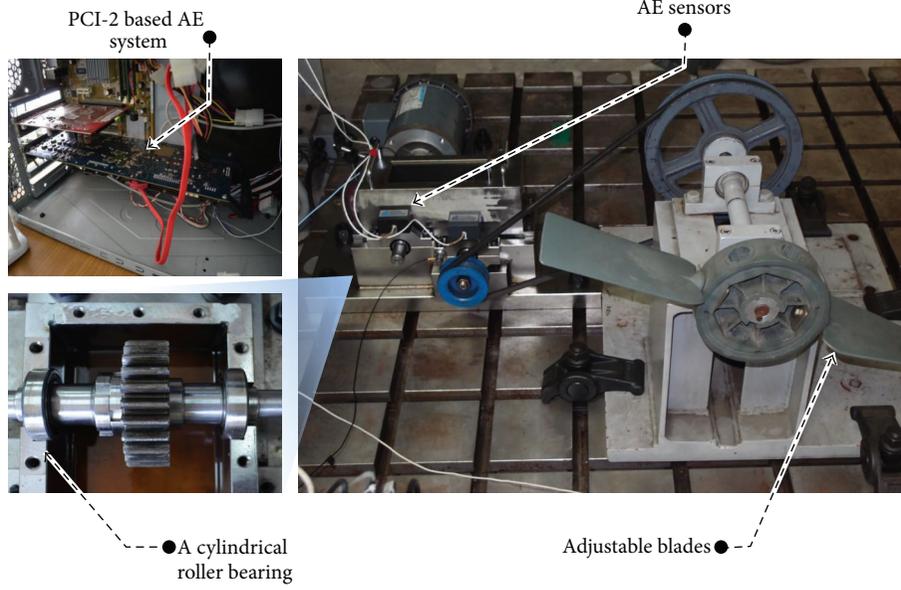


FIGURE 4: Experimental test rig for fault diagnosis of rolling element bearings [10].

algorithms, such as KNN, ANN, and SVM, have complexities  $O(N^2)$ ,  $O(N^2)$ , and  $O(N^3)$ , respectively [20]. Furthermore, the Naïve Bayes classifier is adaptive in nature and does not require any fixed parameters (unlike other classification algorithms).

In this study, in order to evaluate the generalized classification performance of the proposed comprehensive bearing fault diagnosis method,  $k$ -fold cross validation ( $k$ -cv) [21] is employed. In  $k$ -cv, the feature pool is randomly partitioned into  $k$  mutual subsets ( $k = 3$  in this study). Then, at the  $i$ th iteration of  $k$ -cv, one of the subsets is used as the testing dataset and the other  $(k - 1)$  subsets are employed to train the Naïve Bayes classifier. The classification performance is averaged over  $k$  trials of  $k$ -cv. The advantage of this technique is that no matter how we divide the data every data sample is in the testing set once and is included in the training set  $(k - 1)$  times. Thus, variance in the classification results is minimized. To validate the classification performance of the proposed method, this paper utilizes the average classification accuracy (ACA) and the true positive rate (TPR) as performance measures. These are defined in (11) and (12), respectively:

$$ACA = \frac{1}{k} \sum_{i=1}^k \frac{1}{N} \sum_{j=1}^C N_{TP}^{i,j} \times 100 (\%), \quad (11)$$

$$TPR_j = \frac{1}{k} \sum_{i=1}^k \frac{N_{TP}^{i,j}}{N_{TP}^{i,j} + N_{FN}^{i,j}} \times 100 (\%), \quad (12)$$

where  $k$  is the number of cross validation folds and  $N_{TP}^{i,j}$  and  $N_{FN}^{i,j}$  are the number of true positives and false negatives of class  $c_j$  resulting in the  $i$ th iteration of  $k$ -cv, respectively. In addition, the number of true positives is defined as the total number of samples in class  $c_j$  that are accurately classified as

class  $c_j$ . Alternatively, the number of false negatives indicates the number of samples in class  $c_j$  that are not classified as  $c_j$ .

### 3. Experimental Setup

In this study, we used an experimental test rig developed by Intelligence Dynamics Lab (Gyeongsang National University, Korea) to validate the performance of the proposed fault diagnosis methodology, as shown in Figure 4. An acoustic emission (AE) sensor (WS $\alpha$  type, from Physical Acoustic Cooperation) was employed to capture continuous AE signals. Table 3 describes the specifications of the data acquisition system used in this study.

The bearings used in this study were cylindrical rolling element type (FAG NJ206-3-TVP2). These bearings were engrained with different bearing defects, which are shown in Figure 5 and listed as follows: bearing-crack-on-outer-raceway (BCO), bearing-crack-on-inner-raceway (BCI), bearing-crack-on-roller (BCR), bearing-crack-on-outer-and-inner-raceways (BCOI), bearing-crack-on-outer-raceway-and-roller (BCOR), bearing-crack-on-inner-race-and-roller (BCIR), and bearing-crack-on-outer-and-inner-raceways-and-roller (BCOIR). A healthy or defect-free bearing (DFB) was used as a reference for baseline measurements. In total, eight types of bearing fault signals, including a normal defect-free bearing (as a baseline), were acquired under different rotational speeds (300, 350, 400, 450, and 500 rpm). A total of 360 AE fault signals, each 10 seconds in duration, were collected at a sampling rate of 250 KHz for each bearing defect in this study.

### 4. Experimental Results

**4.1. Data Preprocessing.** In this subsection, we investigate and justify the preprocessing of the AE fault signals to

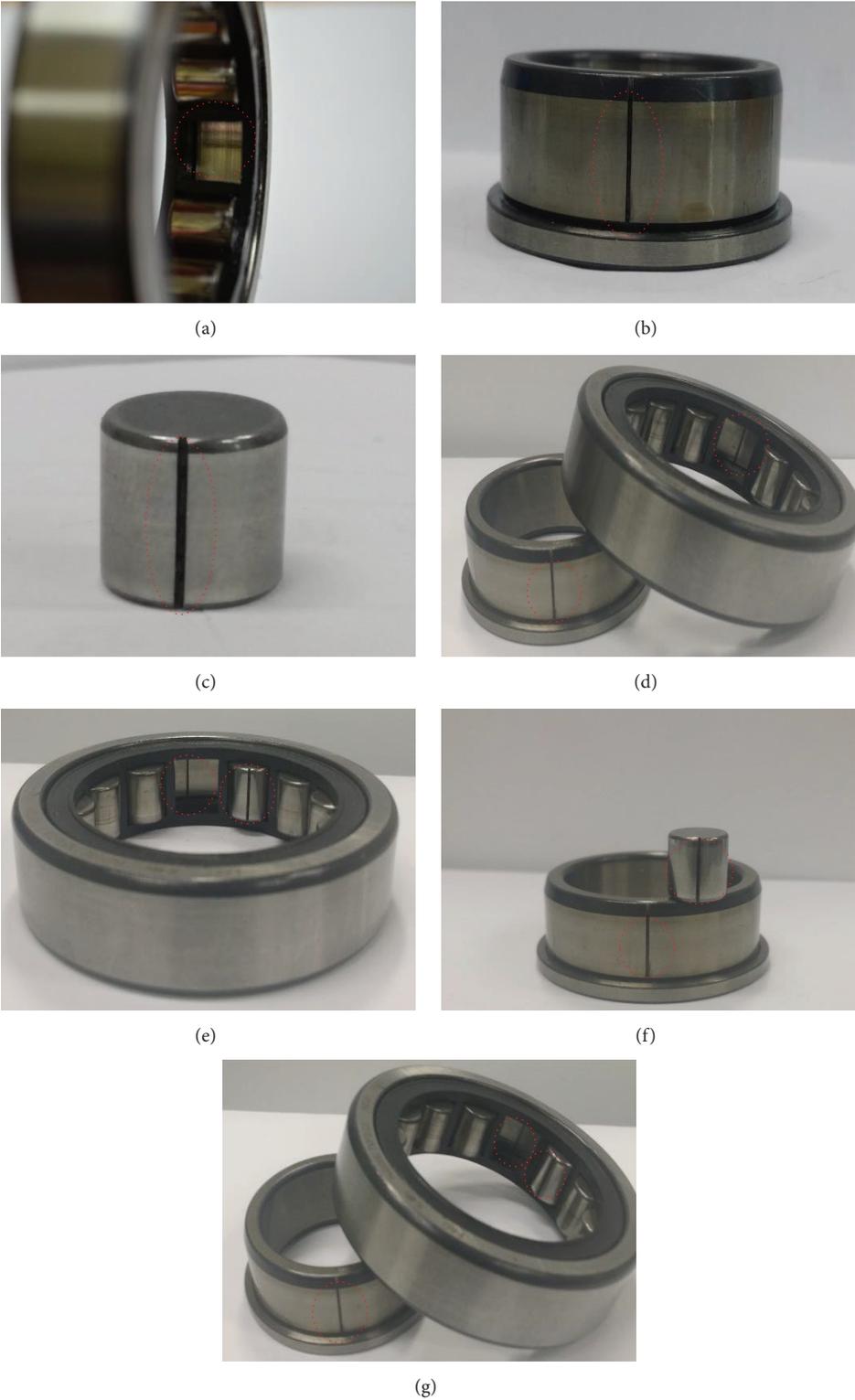


FIGURE 5: Single and multiple-combined bearing defects. (a) BCO, (b) BCI, (c) BCR, (d) BCOI, (e) BCOR, (f) BCIR, and (g) BCOIR.

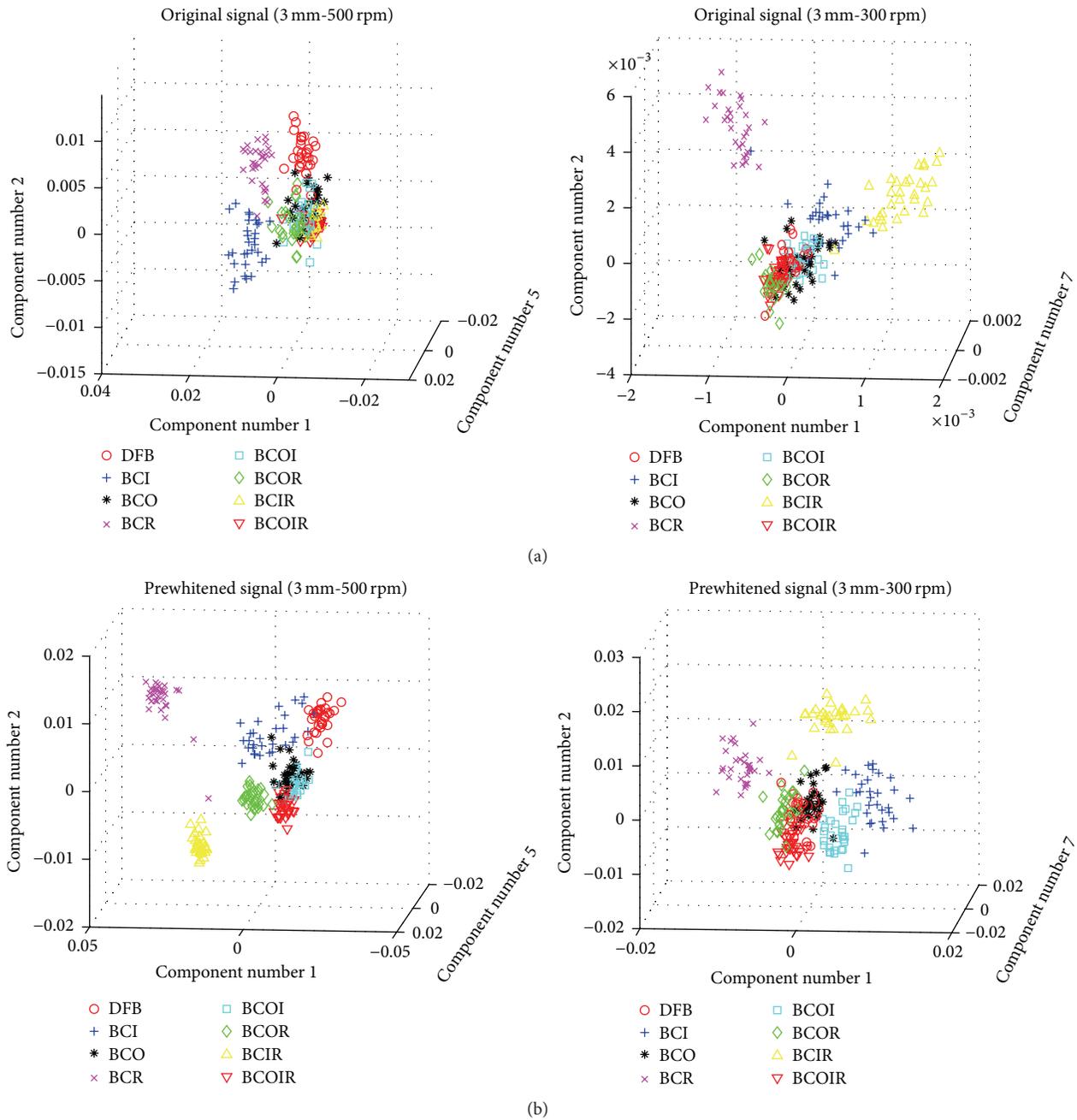


FIGURE 6: Feature distribution spaces with a bearing crack size of 3 mm. (a) Before prewhitening and (b) after prewhitening.

improve the classification performance of our methodology. As mentioned earlier, a total of eight single and compound bearing defects were investigated in this study. Figures 6 and 7 show the feature distribution of the optimal LDA discriminant components for four datasets. These datasets were acquired at two different rotational speeds (i.e., 500 rpm and 300 rpm) and for two different bearing crack sizes (i.e., 3 mm and 12 mm). From Figure 6, it is clear that the fault features are not well separated and discrimination between different classes is poor without preprocessing. For instance, the class clusters of the 3 mm crack length and 300 rpm rotational speed datasets are heavily overlapped,

which leads to higher error rate in classification. Conversely, preprocessing the original signals to obtain the prewhitened signals reduces overlapping between clusters and improves separation between their boundaries.

Likewise, Figure 7 shows that the boundaries between some class clusters of the original signals are blurred and unclear, whereas the prewhitened signals have clearer and more distinct boundaries. These results strongly suggest that prewhitened signals are able to strongly project the discriminative information that is hidden in the bearing signals. The variation between the peaks and flanks of the spectral energy in spectral kurtosis analysis is one reason why each class of

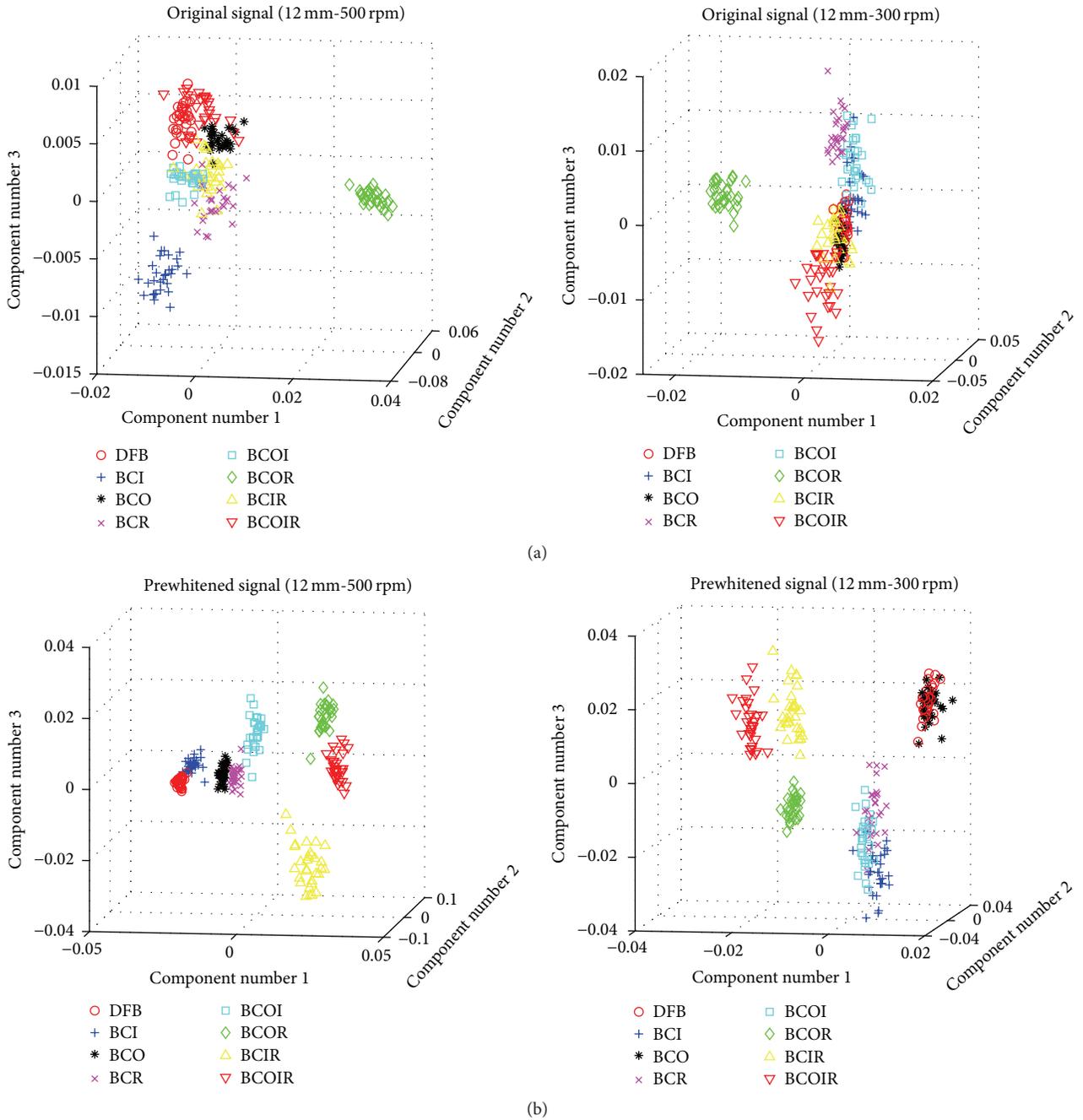


FIGURE 7: Feature distribution spaces with a bearing crack size of 12 mm. (a) Before prewhitening and (b) after prewhitening.

the original signals cannot be clustered accurately. Therefore, a prewhitening process is required to improve separation between different clusters, which makes classification easier and more accurate, even with a Naïve Bayes classifier.

**4.2. Proposed VMDC-Based Feature Analysis Method.** In order to validate the effectiveness of the proposed feature analysis method, this study compares the feature selection performance (in terms of the classification accuracy) of the proposed method and the original LDA for 10 different datasets, as shown in Figures 8 and 9. The vertical axis

represents the average classification accuracy calculated for the classification results of different fault types through  $k$  trials of  $k$ -cv, and the horizontal axis represents the number of LDA discriminant components that are selected. There is no general consensus about the appropriate number of LDA components that always guarantees the highest classification accuracy. As evident in Figures 8 and 9, the LDA method achieves its highest classification accuracy when the number of LDA discriminant components reaches seven. Nevertheless, there are some exceptions. For example, in Figure 8 (for the dataset of a 3 mm length crack and a rotational speed of

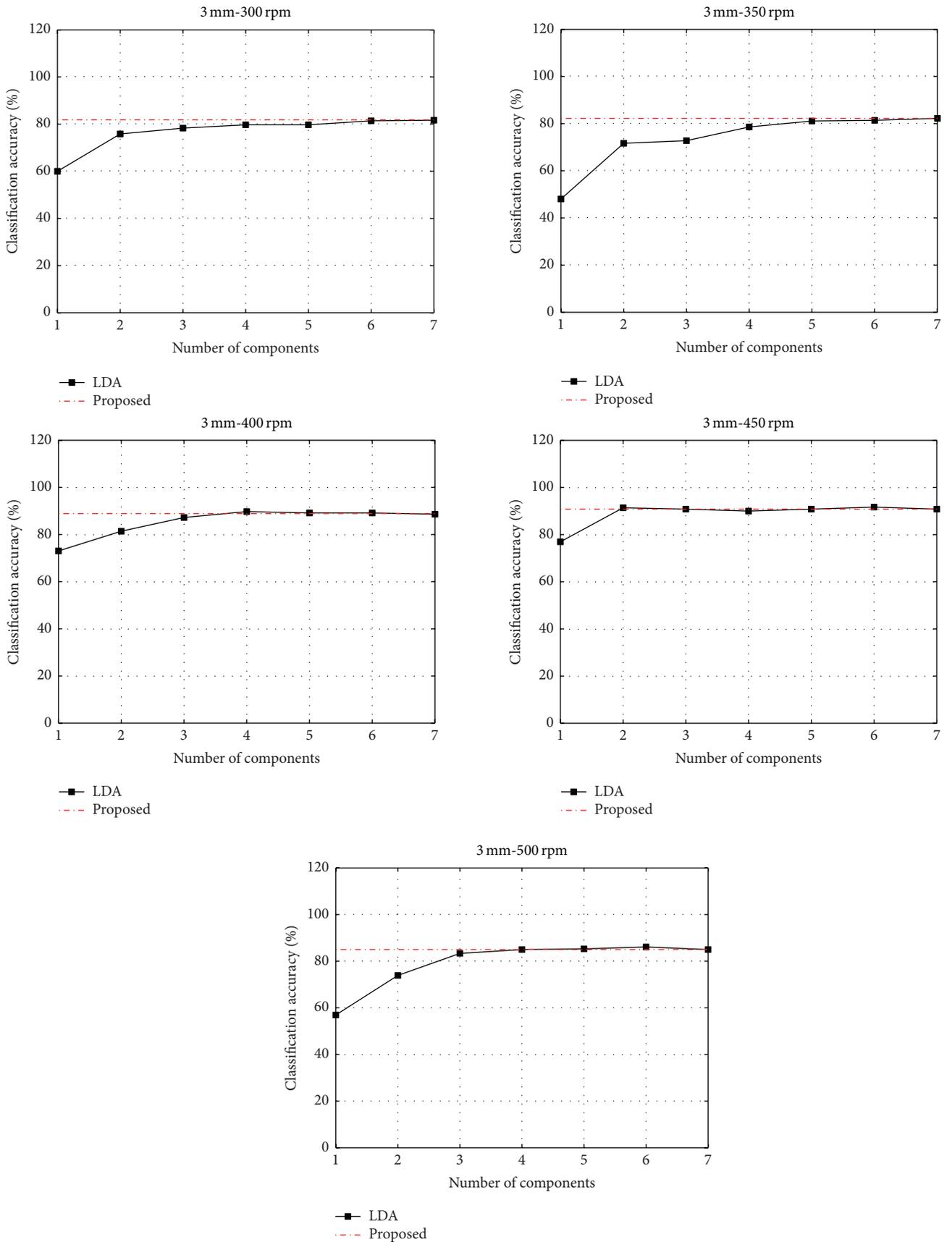


FIGURE 8: Performance of the proposed feature analysis at a bearing crack size of 3 mm.

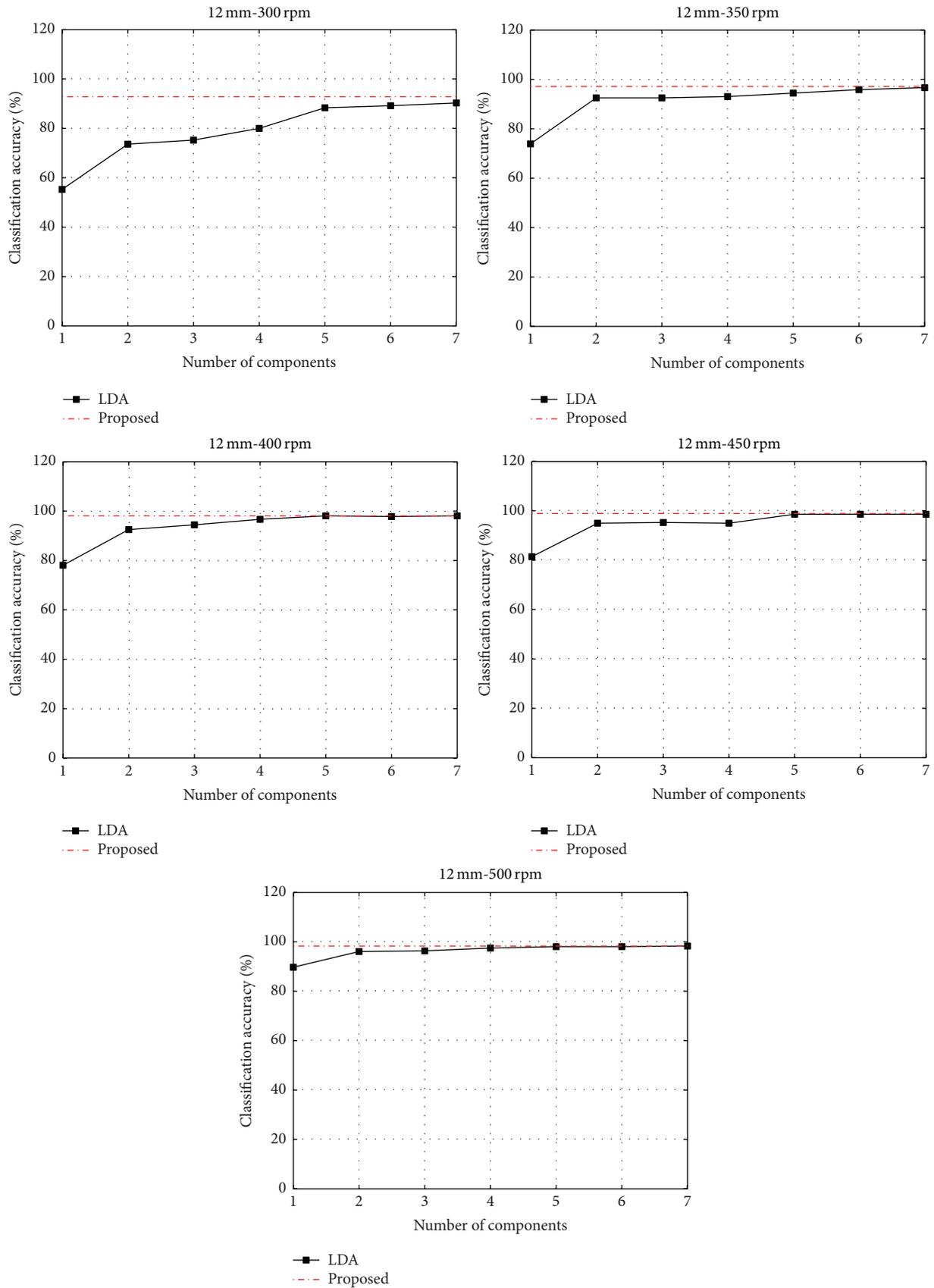


FIGURE 9: Performance of the proposed feature analysis at a bearing crack size of 12 mm.

TABLE 4: Specifications of the data acquisition system.

PCI-2 based AE system	(i) ADC: 18-bit 40 MS/s per channel maximum (ii) Frequency response: 1 kHz–3 MHz (at –3 dB points) (iii) Sample rate: 100 kS/s, 200 kS/s, 500 kS/s, 1 MS/s, 2 MS/s, 5 MS/s, 10 MS/s, 20 MS/s, and 40 MS/s are selectable
WS $\alpha$ sensor	(i) Peak sensitivity (V/ $\mu$ bar): –62 dB (ii) Operating frequency range: 100–900 kHz (iii) Resonant frequency: 650 kHz

kS/s: kSamples per second; MS/s: MSamples per second.

TABLE 5: Performance comparison between conventional feature analysis approaches and the proposed feature analysis method in terms of the average classification accuracy and the true positive rate at a bearing crack size of 3 mm (unit: %).

	Shaft speed	TPR per bearing fault condition under 3 mm crack size								ACA
		DFB	BCI	BCO	BCR	BCOI	BCOR	BCIR	BCOIR	
PCA	300 RPM	40.00	62.22	62.22	66.67	53.33	62.22	66.67	75.56	<b>61.11</b>
	350 RPM	57.78	55.56	28.89	91.11	44.44	68.89	31.11	55.56	<b>54.17</b>
	400 RPM	66.67	71.11	15.56	86.67	75.56	51.11	97.78	60.00	<b>65.56</b>
	450 RPM	86.67	86.67	26.67	91.11	33.33	75.56	95.56	80.00	<b>71.94</b>
	500 RPM	82.22	82.22	33.33	91.11	48.89	71.11	93.33	46.67	<b>68.61</b>
ICA	300 RPM	48.89	60.00	62.22	71.11	40.00	57.78	71.11	77.78	<b>61.11</b>
	350 RPM	42.22	51.11	13.33	86.67	35.56	73.33	57.78	40.00	<b>50.00</b>
	400 RPM	88.89	80.00	42.22	80.00	66.67	35.56	100.00	62.22	<b>69.44</b>
	450 RPM	88.89	82.22	28.89	88.89	53.33	82.22	95.56	73.33	<b>74.17</b>
	500 RPM	88.89	84.44	35.56	91.11	31.11	75.56	93.33	44.44	<b>68.06</b>
LDA	300 RPM	53.33	93.33	73.33	100.00	77.78	84.44	82.22	77.78	<b>80.28</b>
	350 RPM	82.22	91.11	71.11	86.67	77.78	86.67	80.00	88.89	<b>83.06</b>
	400 RPM	93.33	91.11	57.78	93.33	91.11	77.78	97.78	80.00	<b>85.28</b>
	450 RPM	88.89	91.11	77.78	100.00	86.67	95.56	100.00	71.11	<b>88.89</b>
	500 RPM	88.89	97.78	68.89	91.11	73.33	86.67	100.00	80.00	<b>85.83</b>
Proposed	300 RPM	53.33	80.00	62.22	100.00	75.56	80.00	82.22	80.00	<b>76.67</b>
	350 RPM	80.00	88.89	73.33	88.89	80.00	86.67	82.22	91.11	<b>83.89</b>
	400 RPM	91.11	91.11	60.00	93.33	93.33	82.22	100.00	82.22	<b>86.67</b>
	450 RPM	88.89	97.78	86.67	97.78	88.89	97.78	95.56	84.44	<b>92.22</b>
	500 RPM	88.89	97.78	64.44	91.11	75.56	91.11	100.00	80.00	<b>86.11</b>

450 rpm), the highest accuracy that can be achieved is 91.67% (with six LDA components); the highest accuracy obtained with seven LDA components is only 90.83%. Thus, the number of LDA components that can achieve the maximum classification accuracy must be determined empirically for each study. On the contrary, the proposed feature analysis approach always shows better average classification accuracies when compared with the original LDA algorithm. For instance, in Figure 9 (for the dataset of a 12 mm length crack and a rotational speed of 300 rpm), the proposed method yields a classification accuracy of up to 92.8%, clearly outperforming the LDA method with any number of components. This can be explained by the fact that our proposed method utilizes the vector-median discriminant criterion, which can adaptively select the optimal number of LDA components to maximize the ratio of interclass separability and intraclass compactness, thereby achieving the maximum discriminatory power in a low-dimensional feature space. This enables the proposed approach to always achieve the best classification performance. The analysis of

these results highlights the main contributions of this study in analyzing and selecting the optimal features from the original feature pool to achieve superior classification performance as compared to conventional feature analysis algorithms. Using the LDA method, the number of LDA components that yield the best classification results must be determined empirically on a dataset to dataset basis, whereas the proposed VMDC-based approach does this adaptively.

**4.3. Classification Performance.** In this section, we compare the classification performance of the proposed methodology with other state-of-the-art feature analysis approaches (i.e., PCA, ICA, and LDA). The optimal number of components for PCA, ICA, and LDA was determined by utilizing the training set of  $k$ -cv fold at each iteration and finding the number of components that yielded the highest accuracy for each of these methods. The final classification results, which have been calculated through  $k$  folds of  $k$ -cv, are summarized in Tables 4 and 5. We used two evaluation indexes: the true positive rate (TPR) and the average classification accuracy

TABLE 6: Performance comparison between conventional feature analysis approaches and the proposed feature analysis method in terms of the average classification accuracy and the true positive rate at a bearing crack size of 12 mm (unit: %).

	Shaft speed	TPR per bearing fault condition under 12 mm crack size								ACA
		DFB	BCI	BCO	BCR	BCOI	BCOR	BCIR	BCOIR	
PCA	300 RPM	84.44	75.56	64.44	66.67	86.67	95.56	71.11	75.56	<b>77.50</b>
	350 RPM	95.56	53.33	75.56	71.11	77.78	88.89	80.00	95.56	<b>79.72</b>
	400 RPM	93.33	73.33	86.67	82.22	75.56	91.11	95.56	97.78	<b>86.94</b>
	450 RPM	91.11	77.78	88.89	73.33	91.11	95.56	91.11	91.11	<b>87.50</b>
	500 RPM	88.89	95.56	91.11	86.67	95.56	93.33	91.11	93.33	<b>91.94</b>
ICA	300 RPM	51.11	48.89	86.67	66.67	71.11	93.33	71.11	64.44	<b>69.17</b>
	350 RPM	64.44	42.22	88.89	71.11	86.67	91.11	77.78	91.11	<b>76.67</b>
	400 RPM	88.89	62.22	88.89	75.56	75.56	88.89	91.11	97.78	<b>83.61</b>
	450 RPM	100.00	48.89	93.33	68.89	88.89	97.78	93.33	88.89	<b>85.00</b>
	500 RPM	91.11	88.89	91.11	84.44	97.78	93.33	88.89	95.56	<b>91.39</b>
LDA	300 RPM	77.78	88.89	82.22	93.33	95.56	100.00	97.78	88.89	<b>90.56</b>
	350 RPM	91.11	100.00	91.11	93.33	100.00	100.00	97.78	100.00	<b>96.67</b>
	400 RPM	95.56	100.00	97.78	95.56	100.00	100.00	97.78	97.78	<b>98.06</b>
	450 RPM	100.00	100.00	97.78	93.33	97.78	100.00	100.00	97.78	<b>98.33</b>
	500 RPM	91.11	100.00	100.00	97.78	100.00	97.78	100.00	100.00	<b>98.33</b>
Proposed	300 RPM	80.00	86.67	88.89	93.33	93.33	100.00	97.78	88.89	<b>91.11</b>
	350 RPM	91.11	100.00	88.89	93.33	100.00	100.00	100.00	100.00	<b>96.67</b>
	400 RPM	95.56	100.00	97.78	97.78	100.00	100.00	100.00	100.00	<b>98.89</b>
	450 RPM	100.00	100.00	100.00	97.78	97.78	100.00	100.00	100.00	<b>99.44</b>
	500 RPM	93.33	100.00	100.00	97.78	100.00	97.78	100.00	100.00	<b>98.61</b>

(ACA). From the figures in Tables 4 and 5, it is clear that the proposed feature analysis scheme consistently outperforms the conventional approaches of PCA, ICA, and LDA for all of the datasets that were analyzed in this study. For instance, Table 5 reveals that the TPR of the proposed method is around 100% for almost all bearing faults (except for the BCR fault condition, where it ranges between 93 and 97% but is still higher than that of PCA and ICA). In Table 4, it can be observed that the proposed method achieves good classification performance even though the feature clustering of the 3 mm length crack datasets was poor and not well separated as compared to the 12 mm length crack datasets. This comparatively high classification performance demonstrates that the proposed approach is more effective in feature analysis as compared to other conventional methods. Our approach is also able to yield maximum discrimination in a lower-dimensional feature space.

## 5. Conclusion

This paper proposed a comprehensive multifault diagnosis methodology based on AE analysis to detect multiple localized bearing faults of a rolling element bearing. The method comprises a prewhitening step, a feature extraction based on wavelet kurtogram, a useful feature selection with the proposed VMDC-based feature analysis, and a fault classification by employing the NB classifier. One of the main advantages of the proposed methodology is that the subband of the nonstationary AE signal, which carries the most sensitive information related to fault characteristic,

is adaptively identified without any prior knowledge of the bearing fault type by using an enhanced WPT-based kurtogram algorithm. In addition, the proposed VMDC-based discriminative feature analysis approach efficiently identifies optimal features where their discriminatory power is maximized. The proposed methodology was evaluated by an experimental setup using a healthy bearing and seven damaged ones under different rotational speeds and different simulated crack sizes. Experimental results indicated that the proposed multifault diagnosis methodology achieves the highest classification accuracy up to 91.11%, 96.67%, 98.89%, 99.44%, and 98.61% under bearing rotational speeds of 300 rpm, 350 rpm, 400 rpm, 450 rpm, and 500 rpm, respectively.

## Conflict of Interests

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

## Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (no. NRF-2013RIA2A2A05004566).

## References

- [1] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 108–124, 2007.

- [2] N. Sawalhi, R. B. Randall, and H. Endo, "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2616–2633, 2007.
- [3] T. Barszcz and R. B. Randall, "Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352–1365, 2009.
- [4] P. Bošković, J. Petrovičić, B. Musizza, and D. Juričić, "Detection of lubrication starved bearings in electrical motors by means of vibration analysis," *Tribology International*, vol. 43, no. 9, pp. 1683–1692, 2010.
- [5] R. F. Dwyer, "Detection of non-Gaussian signals by frequency domain kurtosis estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '83)*, pp. 607–610, Boston, Mass, USA, 1984.
- [6] J. Antoni and R. B. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 308–331, 2006.
- [7] B. Chen, Z. Zhang, Y. Zi, Z. He, and C. Sun, "Detecting of transient vibration signatures using an improved fast spatial-spectral ensemble kurtosis kurtogram and its applications to mechanical signature analysis of short duration data from rotating machinery," *Mechanical Systems and Signal Processing*, vol. 40, no. 1, pp. 1–37, 2013.
- [8] L. Shi, R. B. Randall, and J. Antoni, "Rolling element bearing fault detection using improved envelope analysis," in *Proceedings of the Eighth International Conference on Vibrations in Rotating Machinery*, pp. 301–311, Swansea, Wales, September 2004.
- [9] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and B.-K. Choi, "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2786–2797, 2015.
- [10] M. Kang, J. Kim, B.-K. Choi, and J.-M. Kim, "Envelop analysis with a genetic algorithm-based adaptive filter bank for bearing fault detection," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, article EL65, 6 pages, 2015.
- [11] A. Widodo and B.-S. Yang, "Machine health prognostics using survival probability and support vector machine," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8430–8437, 2011.
- [12] Y. Guo, J. Na, B. Li, and R.-F. Fung, "Envelope extraction based dimension reduction for independent component analysis in fault diagnosis of rolling element bearing," *Journal of Sound and Vibration*, vol. 333, no. 13, pp. 2983–2994, 2014.
- [13] J. Harmouche, C. Delpha, and D. Diallo, "Improved fault diagnosis of ball bearings based on the global spectrum of vibration signals," *IEEE Transactions on Energy Conversion*, vol. 30, no. 1, pp. 376–383, 2015.
- [14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Computer Science and Scientific Computing, Academic Press, Boston, Mass, USA, 1990.
- [15] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [16] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, NY, USA, 1992.
- [17] F. Cong, J. Chen, and G. Dong, "Spectral kurtosis based on AR model for fault diagnosis and condition monitoring of rolling bearing," *Journal of Mechanical Science and Technology*, vol. 26, no. 2, pp. 301–306, 2012.
- [18] K. Mardia, J. Kent, and J. Bibby, *Multivariate Analysis*, Academic Press, London, UK, 1979.
- [19] J. Vaidya, M. Kantarcioğlu, and C. Clifton, "Privacy-preserving Naïve Bayes classification," *The VLDB Journal*, vol. 17, no. 4, pp. 879–898, 2008.
- [20] Y.-L. He, R. Wang, S. Kwong, and X.-Z. Wang, "Bayesian classifiers based on probability density estimation and their applications to simultaneous fault diagnosis," *Information Sciences*, vol. 259, pp. 252–268, 2014.
- [21] J. Rodriguez, A. J. Perez, and J. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 569–575, 2010.

## Research Article

# Condition Monitoring and Fault Diagnosis for an Antifalling Safety Device

Guangxiang Yang<sup>1,2</sup> and Hua Liang<sup>1</sup>

<sup>1</sup>Chongqing Engineering Laboratory for Detection, Control and Integrated System, Chongqing Technology and Business University, Chongqing 400067, China

<sup>2</sup>Research Center of System Health Maintenance, Chongqing Technology and Business University, Chongqing 400067, China

Correspondence should be addressed to Guangxiang Yang; ygxmonkey@126.com

Received 7 April 2015; Accepted 26 May 2015

Academic Editor: Chuan Li

Copyright © 2015 G. Yang and H. Liang. 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.

There is a constant need for the safe operation and reliability of antifalling safety device (AFSD) of an elevator. This paper reports an experimental study on rotation speed and catching torque monitoring and fault diagnosis of an antifalling safety device in a construction elevator. Denoising the signal using wavelet transform is presented in this paper. Based on the denoising effects for several types of wavelets, the sym8 wavelet basis, which introduces the high order approximation and an adaptive threshold, is employed for denoising the signal. The experimental result shows a maximum data error reduction of 7.5% is obtained and SNRs (signal-to-noise ratio) of rotation speed and catching torque are improved for 3.9% and 6.4%, respectively.

## 1. Introduction

The construction hoist is one of the essential mechanical devices in the construction. With flourish development of construction industry in China, higher demands are also asked for the performance and reliability of the construction hoist. A construction elevator is widely used in tall building construction and some civil engineering fields such as bridge construction, underground construction, large chimney construction, and shipbuilding.

A construction hoist is equipped with an antifalling safety device to keep the unexpected falling and ensure the safe operation of the elevator. An antifalling safety device is the most important safety device in the pinion and rack builder's hoist, which can prevent the cage running over speed and effectively reduce the cage dropping accident. It is a kind of velocity controlled safety device which prevents the dropping of the elevator cage mounted on rack rails. When the speed of elevator cage reaches the triggering speed of antifalling safety device, the compression spring in the safety device is compressed and the braking force is increased by the friction of cone drum braking surface which is in contact with the brake band. With this braking force increasing, the elevator

cage is braked and stopped on the rack rails [1, 2]. The motor power and the circuits of controller of elevator are cut off as soon as the antifalling safety device is triggered to action. An antifalling safety device is one typical model of building hoist gear cone drum progressive antifalling safety device and is designed for prevent building hoist cages or counterweight from falling. It is made up of gear, centrifugal speed limiter device, cone drum, and brake device. An antifalling safety device is shown in Figure 1.

An antifalling safety device is composed of the following components: a pinion, a friction plate, a centrifugal block, an outer cone, a disc spring, a microswitch, a screw nut, and a spring.

When the hoist cage is speeding down, the rotation speed of pinion is increasing and the centrifugal block rotated with pinion at the same speed is thrown outwards and meshed in the outer cone. This makes the disc spring compressed and the positive pressure between the outer cone and the friction plate is increased gradually. Hence, the torque is produced. Catching torque is increased with the catching process to achieve gradual braking, until the cage is stopped. The construction elevator related specifications for an antifalling safety device in China including JG121-2000



FIGURE 1: An antifalling safety device.

building hoist, pinion and cone progressive type safety device, GB10055-2007 construction hoist safety regulations, and GB/T10054-2005 builder's hoist. These documents specify the operation, maintenance, and verification. For example, it is strictly pointed out in GB10055-2007 that the valid calibration deadline is one year; the tripping speed and catching distance of must meet the requirement of GB/T10054-2005.

According to the reports from a special laboratory of Liaoning province in China [3], there are 143 unqualified or failed antifalling safety devices in totally 1428 detected antifalling safety device in 2012. The unqualified rate reached 10% and in which 90% of the faults are due to rotation speed fault, the other is mainly the failure of braking distance and braking load. This fully demonstrates the importance and necessity of antifalling safety device of regular inspection.

Current research on the antifalling safety device mainly concentrated in the aspects of design, safety verification, and structural analysis. The condition monitoring and fault diagnosis scheme are rarely reported. Furthermore, the detected data used to calculate the catching distance and the maximum catching load of antifalling safety device is often contaminated or influenced by the electromagnetic interference. Therefore the accuracy and the reliability of data have to be improved.

In this paper, a condition monitoring system is developed and a fault diagnosis scheme is presented by using the wavelet transform thresholding denoising method. By reviewing the state-of-the-art in wavelet for rotating machine condition monitoring, the paper discusses specific needs of an antifalling safety device monitoring system addressed by wavelet, describes features of wavelet, and investigates the experiment study of wavelet in an antifalling safety device monitoring system. On the basis of comparison of denoising effect for several types of wavelets, the sym8 wavelet basis, which introduces the high order approximation, tuned scale factors, and an adaptive threshold, is employed for denoising the original catching torque and rotation speed signals. Experimental results show that the denoised data is smoother and the reliability of signal is improved. The sym8 wavelet with thresholding is proven to be better than the others. The average error of pulse peak data is reduced by 7.5% when the sym8 wavelet is applied for denoising the observed signal which proves the effectiveness and reliability of this method.

## 2. Related Works

An antifalling safety device in nature is a kind of rotating machinery. Rotating machine analysis has been conducted for many decades [4]. It is becoming popular for on-going production condition based monitoring system. Rotating machines operating in production is greatly benefited from accurate performance data. For an antifalling safety device, Bai and Zhang present the system design of detection platform [5]. The condition of an antifalling safety device is monitored and the data is obtained by sensors with the detection platform. Therefore the original data is important for calibration and computation of the catching distance and the catching load. Li et al. [6] propose an intelligent diagnosis method for condition diagnosis of rotating machinery by using wavelet transform and ant colony optimization, in order to detect faults and distinguish fault types at an early stage. Venugopal et al. [7] present a signal processing technique for machine performance monitoring which exploits fluctuations in phase angles of machine rotational frequency signals to determine their dynamic temporal coherence. Sun et al. [8] propose a fault diagnosis method for rotating machinery using ant colony optimization and possibility theory. Mueller and Pratt [9] present the theory and practical application for using radio frequency polarimetry in analyzing rotating machines. A neural network-based fault detection and isolation scheme is presented to detect and isolate faults in a highly nonlinear dynamics of an aircraft jet engine by Tayarani-Bathaie et al. [10]. Yang et al. [11] present a generalization of multidimensional linear regression to facilitate multisensor fault detection and signal reconstruction through the use of analytical optimization.

However, as a rotating machine, the antifalling safety device is more easily to be verified on condition that the original data are not noised. The most key problem is the confused detected result by the noised signal. These signals are used for calibration and parameter calculation such as catching distance and the catching load of a safety device. There are few researches of detected signals and data reported in the past. For the noised signal, an efficient technique for signal processing is the wavelet transform.

The wavelet transform, a signal processing technique based on a windowing approach of dilated "scaled" and shifted wavelets, is being applied to a broad range of engineering applications and other applications such as coding, data compression, structural detection and noise reduction, and spectroscopy signal processing [12–15]. Wavelet transform is also widely used in machinery diagnostics and measurement [16, 17].

Dziedziech et al. [18] present the time-variant frequency response function-based on the continuous wavelet transform for the analysis of time-variant systems. The focus is on the comparative study of various broadband input excitations. Measurement of groove race defect width in thrust bearing is presented by processing vibration signal using wavelet based technique by Singh and Kumar [19]. Wavelet features are computed using discrete wavelet transform (DWT) from the vibration signals of centrifugal pumps by Muralidharan and Sugumaran [20]. The seismic damage

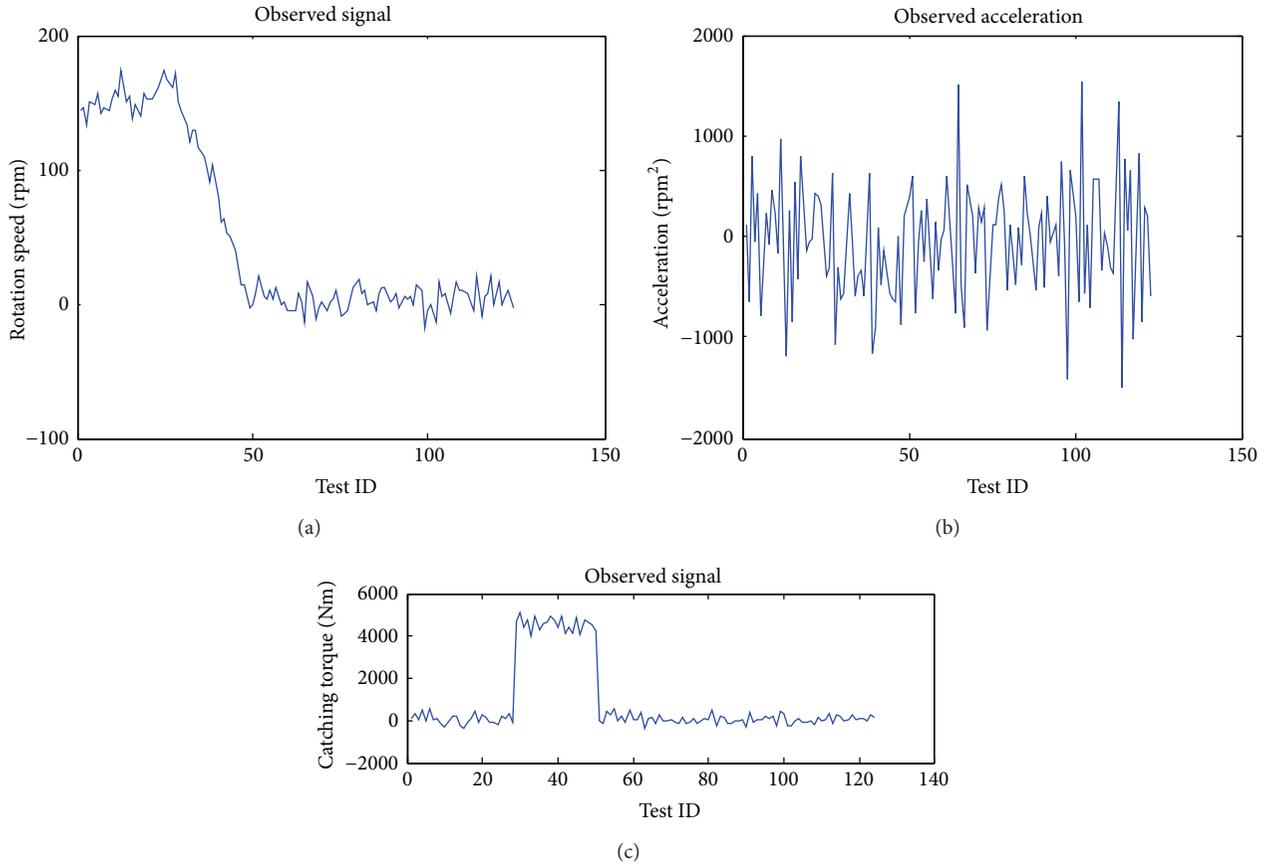


FIGURE 2: Observed signal of an antifalling safety device.

level of a structure is assessed using damage indices and the continuous wavelet transformation (CWT) technique is applied to compute the dissipated energy during ground shaking by Rajeev and Wijesundara [21]. Li et al. [22] review some important developments in TDLAS (tunable diode laser absorption spectroscopy), from its basic principles as a spectroscopic tool to the demonstration of gas absorption measurements, emphasizing signal enhancement and noise reduction techniques developed for improving current TDLAS performance. A recursive algorithm [23] is presented to estimate the variance of the noise and prove its convergence and investigate its mathematical properties. All of these works have greatly improved the research and application of wavelet transform.

In this paper, thresholding in wavelet domain is used for noise reduction of rotation speed and catching torque in antifalling safety device diagnosis. The proposed signal denoising method in this paper is based on the method of Johnstone and Donoho. It has been proven that this method has been successfully applied to a wide class of nonstationary signals. The signal denoising method is successfully applied in the measurement of safety device in this paper. The denoised signal is smoother and the noised data is reduced, which guarantee the accuracy and reliability of an antifalling safety device.

### 3. Fault Diagnosis with Wavelet Transform

Because of the environment interface, the signal of an antifalling safety device is more easily to be noised. The typical originals are shown in Figure 2.

In Figure 2, “test ID” means the data sequence number when the signal is sampled from sensors. These signals are used for calibration and parameter calculation such as catching distance and the catching load of a safety device. Therefore the accuracy and reliability of acquired data may have a significant impact on the detection result of safety device. For the noised signal, an efficient technique for signal processing is the wavelet transform.

A novel approach for noise reduction has been established by Donoho and Johnstone in the 1990s [24]. It employs thresholding in the wavelet transformation and can be shown to be asymptotically near optimal for signals corrupted by additive white Gaussian noise (AWGN). This method has a wide variety of related problems such as linear inverse problems [25], data compression, and statistical estimation [26]. It is also used in construction, structure health monitoring [27], and civil engineering [28–31]. Noise reduction of wavelet transform has been proved to work well for a wide class of 1D and 2D signals.

Given that finite samples  $f_i$  of signal  $f(t)$ , which is overlaid with a white Gaussian noisy signal, is presented by:

$$y_i = f_i + \sigma z_i; \quad i = 0, 1, 2, \dots, n-1, \quad (1)$$

where  $z_i$  is a white Gaussian noise and  $\sigma$  is the noise level.

For this continuous signal  $f(t)$ , its integral transformation is given by

$$W_f(\alpha, \tau) = \int_{\mathbb{R}} f(t) \overline{\psi_{\alpha\tau}(t)} dt = \langle f, \psi_{\alpha\tau}(t) \rangle. \quad (2)$$

Equation (1) is a continuous wavelet transform (CWT) of  $f(t)$ , in which

$$\psi_{\alpha\tau}(t) = \frac{1}{\sqrt{\alpha}} \psi\left(\frac{t-\tau}{\alpha}\right) \quad (3)$$

is the wavelet basis function which is a group of functions scaled or shifted by a mother wavelet  $\psi(t)$ , where  $\alpha$  and  $\tau$  are scaling and shifting parameters of wavelet function  $\psi(t)$ , respectively.

The discrete wavelet function is given as

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k); \quad j, k \in \mathbb{Z}, \quad (4)$$

where  $\psi(t)$  is the wavelet function and  $2^{-j/2}\psi(2^{-j}t - k)$  are scaled and shifted versions of  $\psi(t)$  based on the values of scaling coefficient  $j$  and shifting coefficient  $k$ , which is usually written as  $\psi_{j,k}(t)$ . The  $j$  and  $k$  coefficients take integer values for different scaling and shifted versions of  $\psi(t)$ .

In this case the continuous wavelet transform (CWT) of a continuous signal  $f(t)$  (1) is given as

$$D[W_f(j, k)] = 2^{-j/2} \int_{\mathbb{R}} f(t) \overline{\psi(2^{-j}t - k)} dt \quad (5)$$

and the discrete wavelet transform (DWT) of a discrete time sequence  $f(t)$  can be derived:

$$d_{j,k} = 2^{-j/2} \sum_t f(t) \psi(2^{-j}t - k). \quad (6)$$

Provided that  $f_k$  is the discrete signal of  $f(t)$ ,  $f_k = c_{0,k}$ , the orthogonal wavelet decomposition formula [32] of signal  $f(t)$  is computed as

$$\begin{aligned} c_{j,k} &= \sum_n c_{j-1,n} h_{n-2k}, \\ d_{j,k} &= \sum_n c_{j-1,n} g_{n-2k}, \end{aligned} \quad (7)$$

$$k = 0, 1, 2, \dots, N-1,$$

where  $c_{j,k}$  is the scale coefficient,  $d_{j,k}$  is wavelet coefficient,  $h, g$  is one pair of orthogonal mirror filters,  $j$  is resolution level, and  $N$  is the total number of data. The reconstruction of signal  $f(t)$  is the reverse process of decomposition. The reconstruction equation is computed as

$$c_{j-1,k} = \sum_n c_{j,n} h_{k-2n} + \sum_n d_{j,n} g_{k-2n}. \quad (8)$$

The Donoho denoising scheme can recover the original signal  $x_i$  from the noisy signal  $y_i$  by the following steps.

TABLE 1: The calibrated tripping speed (m/s).

Rated lifting speed of construction hoist: $v$	Calibrated tripping speed of safety device: $v1$
$v \leq 0.60$	$v1 \leq 1.00$
$0.60 < v \leq 1.33$	$v1 \leq v + 0.40$
$v > 1.33$	$v1 \leq 1.3v$

- (1) Computing the orthogonal wavelet transform of noisy signal, selecting the proper wavelet and wavelet resolution level  $j$ . Based on (7), the noisy signal is decomposed to resolution level  $j$  and the corresponding wavelet coefficient  $d_{j,k}$  can be obtained.
- (2) Computing the wavelet coefficient with soft thresholding and hard thresholding.
- (3) Inverse wavelet transform. The original signal is obtained by thresholding with wavelet.

The thresholding feature of wavelet denoising is proved a more efficient method than conventional denoising methods [33]. The procedure starts by choosing an appropriate level for decomposition and decomposing the dynamics signal up to this level. For each level of decomposition, a certain threshold is selected and an appropriate thresholding criterion is applied. The threshold criterion can be either hard or soft thresholding. It is proved in this paper that denoising signal method using soft heuristic thresholding at level 4 by sym8 wavelet is with an improvement of SNR (signal-to-noise ratio). Symlets wavelet [34] is a function system and obtained by improving Daubechies wavelet. Symlets wavelet is denoted as sym $N$  ( $N = 2, 3, \dots, 8$ ), where  $N$  is the order. Similar to Daubechies wavelet, Symlets wavelet does not have specified algebraic expression. The catching torque signal and rotation speed signal measured on the detection platform is a real-world signal. This signal is particularly interesting because of noise introduced when a defect developed in the monitoring equipment during the measurement. Wavelet is used to smooth out and to remove the noised data of the measured signal. The obtained noiseless signals of catching torque and rotation speed are then reconstructed in time domain for diagnosis and calibration of antifalling safety device.

The detected data of an antifalling safety device including rotation speed and catching torque are often contaminated and noised by the environment electromagnetic interference. The data are key parameters to compute the tripping speed and the catching distance of an antifalling safety device, which is the criteria of whether it is qualified or not. Therefore, the detected signals of rotation speed and catching torque are denoised by thresholding with wavelet transform in this paper.

The calibrated tripping speed and the catching distance of an antifalling safety device are specified in the Chinese national specification for construction hoist, which are shown in Tables 1 and 2.

The rotation speed  $s_i$  and the catching torque  $t_i$  are measured on the detection platform of an antifalling platform. The sample size of two measured parameters is 125 which

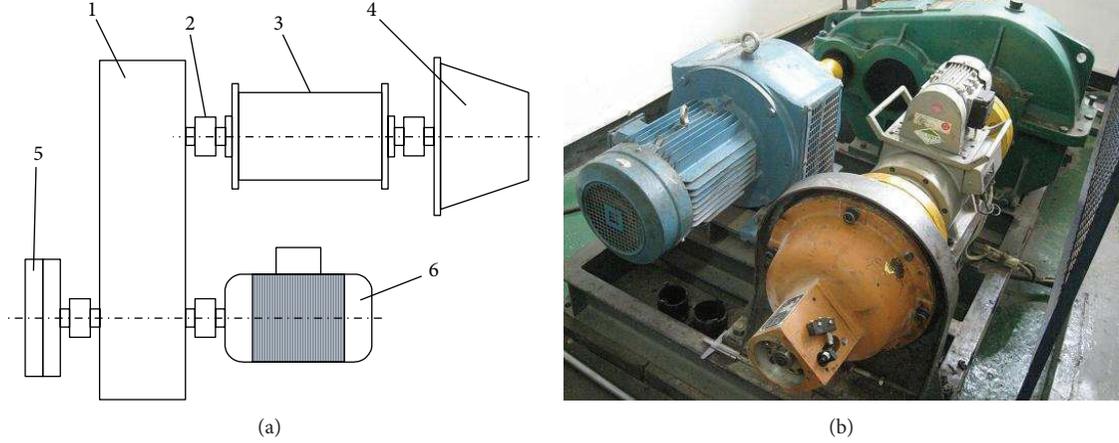


FIGURE 3: Experiment platform, where 1 is a gear box, 2 is a shaft coupling, 3 is sensor system of rotation speed and catching torque, 4 is an antifalling safety device, 5 is an electric motor, and 6 is an inertia load block.

TABLE 2: The catching distance.

Specified lifting speed of construction hoist: $v$ (m/s)	Specified catching distance of safety device (m)
$v \leq 0.65$	0.15–1.40
$0.65 < v \leq 1.00$	0.25–1.60
$1.00 < v \leq 1.33$	0.35–1.80
$v > 1.33$	0.55–2.00

are obtained in 5 seconds. The wavelet transform sample size is 125. The threshold value [35, 36] in the experiment is calculated by the following equation:

$$\delta_j = \frac{\sigma \sqrt{2 \log N}}{\ln(e + j - 1)}, \quad (9)$$

where  $j$  is wavelet scale coefficient,  $N$  is the sample size,  $e$  is a constant coefficient 2.718, and  $\sigma$  is the standard deviation of the noise signal.

The rotation speed computation equation is given by

$$v = \frac{(3.1415926 \times 8 \times 15 \times B)}{60000} \quad (10)$$

in which  $B$  is the measured rotation speed.

The catching distance is computed by the following formula:

$$\begin{aligned} \text{catching distance} &= \pi \times \text{gear pitch diameter} \\ &\times \frac{\text{compression of disc spring}}{\text{pitch}}, \end{aligned} \quad (11)$$

in which, the gear pitch diameter and pitch is a fixed value; therefore, the catching distance is simplified as

$$d = 0.1885 \times D, \quad (12)$$

in which  $D$  is the compression of disc spring.

TABLE 3: Parameters of SAJ40-1.2 antifalling safety device.

Parameters	Units	Value
Rated lifting speed	m/min	33
Rated loading capacity	kN	40
Rated tripping speed	m/s	1.2
Calibrated tripping speed	m/s	1
Pinion gear modulus/teeth	mm/tooth	8/15
Large nut pitch	mm	2.0

The catching load is computed by the following equation:

$$M = \frac{(2 \times C)}{(1.5 \times 8 \times 15)}, \quad (13)$$

where  $C$  is the measured catching torque.

## 4. Experiments

The condition monitoring of an antifalling safety device is conducted on a detection platform which is shown in Figure 3. The pinion and centrifugal governor of safety device are mounted on the platform. The rotation speed of the pinion is adjusted to the calibrated tripping speed when the centrifugal governor starts working.

The antifalling safety device monitored in this experiment is SAJ40-1.2 type, which is produced in Shanghai. The parameters of this type of safety device are shown in Table 3.

A large number of antifalling safety devices have to be diagnosed every day since there are many construction enterprises in Chongqing, China. The antifalling safety devices to be detected are shown in Figure 4.

The safety device is mounted on the detection platform and is driven by a stepless speed regulation controller. When the safety device began to work, the rotation speed and the catching torque are sampled and stored in the computer. The data curve graph of these two parameters is shown on the computer screen. In the process of measurement, the nut

TABLE 4: The denoised maximum rotation speed.

Observed signal	Denoised signal	Reduction (%)	Average reduction (%)
173.8	156.2	11.3	
168.7	158.7	6.3	
167.2	155.1	7.8	
168.5	152.7	10.3	
166.2	154.9	7.3	7.5
164.1	156.1	5.1	
168.1	157.7	7.0	
169.6	154.6	9.7	
159.2	151.3	5.2	
163.4	155.4	5.1	



FIGURE 4: The antifalling safety devices to be detected.

and the pin cannot be reset but the centrifugal governor must be reset after each detecting. This measurement process repeated for 3 times until the maximum detected catching torque reaches or exceeds the computed torque. After 3 measurements, the nut and the pin are reset manually.

In the experiment, the compressed amount of disc spring, which is the length between the top of the pin and the small end cap plane, is recorded when the maximum catching torque reaches the required value in Chinese construction specification. The compressed length is used for the catching distance computation.

The rotation speed  $s_i$  and the catching torque  $t_i$  are measured on the detection platform and are transmitted to computer with a RS485 communication port. When received the data, they are denoised by thresholding with wavelet transform. The observed catching torque and the denoised catching torque are shown in Figure 5.

From Figure 5, it can be observed that there are some noisy signals on the two sides of the valid catching torque. The hard thresholding and soft thresholding noise reduction methods are applied in the data handling. The denoised catching torque signals with hard thresholding and soft thresholding are smoother than the observed signal. The original signal is the true signal or the expected output of the antifalling safety device based on its characteristics.

The observed signal of rotation speed is shown in Figure 6. The denoised rotation speed signals with hard thresholding and soft thresholding are shown in this figure,

too. The rotation speed and acceleration of observed and denoised signals are shown in Figure 7.

According to the figures, it is shown that the wavelet transform denoising with thresholding is more suitable for the detecting. With the soft thresholding denoising, the maximum rotation speed of 173.8 rpm is reduced to 156.2 rpm, which is 11.3% error reduction. There are 10 groups of data collected in Table 4, which shows an average reduction of maximum rotation speed is 7.5%. The theoretical maximum rotation speed is a constant which is 156 rpm based on the datasheet of detection platform.

Rotation speed and catching torque signal denoising algorithms are evaluated based on signal-to-noise ratio (SNR) and Mean Square Error (MSE) for obtaining low noise and readily observable sampling data. For simplicity, the results of rotation speed and catching torque data are partially listed. SNR [37, 38] of observed data and the denoised data is calculated in this experiment.

The SNR is given by the following equation [39]:

$$\text{SNR} = 10 \lg \left[ \frac{\sum_{i=1}^N s_i^2}{\sum_{i=1}^N (s_i - g_i)^2} \right], \quad (14)$$

where  $s_i$  is the observed signal,  $g_i$  is the denoised signal, and  $N$  is the length of signal.

The MSE is estimated between the denoised signal and observed signal given by the following equation:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (s_i - g_i)^2, \quad (15)$$

where  $s_i$  is the observed signal,  $g_i$  is denoised signal, and  $N$  is the length of signal.

The detection platform aims to detect the rotation speed and the catching torque in which the falling of an elevator is simulated by a motor connected with safety device. The original rotation speed of the safety device is a motion with constant acceleration. The catching torque is theoretically a constant in each testing when the antifalling safety device is triggered to action.

The rotation speed is calculated by the following equation:

$$v_t = v_0 + \alpha t. \quad (16)$$

When an object moves with constant acceleration,  $\alpha$  is the instantaneous acceleration at any point which is a constant,  $v_0$  is the initial velocity, and  $v_t$  is the velocity at any point.

According to Figure 7, the rotation speed of safety device has two sections, that is, acceleration and deceleration. Each testing point is sampled every 0.04 seconds. The acceleration phase is composed of 28 testing points which is  $28 * 0.04 \text{ s} - 0.04 \text{ s} = 1.08 \text{ s}$ . The deceleration section is composed of 24 testing points which consume 0.92 seconds. The original rotation speed at any testing point in acceleration and

TABLE 5: The SNR results of rotation speed with hard thresholding and soft thresholding (dB).

	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10
Hard thresholding	26.51	26.53	26.95	26.89	27.02	26.97	26.11	26.43	26.64	26.12
Soft thresholding	26.35	27.10	27.04	27.17	27.13	26.91	27.01	26.88	27.10	26.73

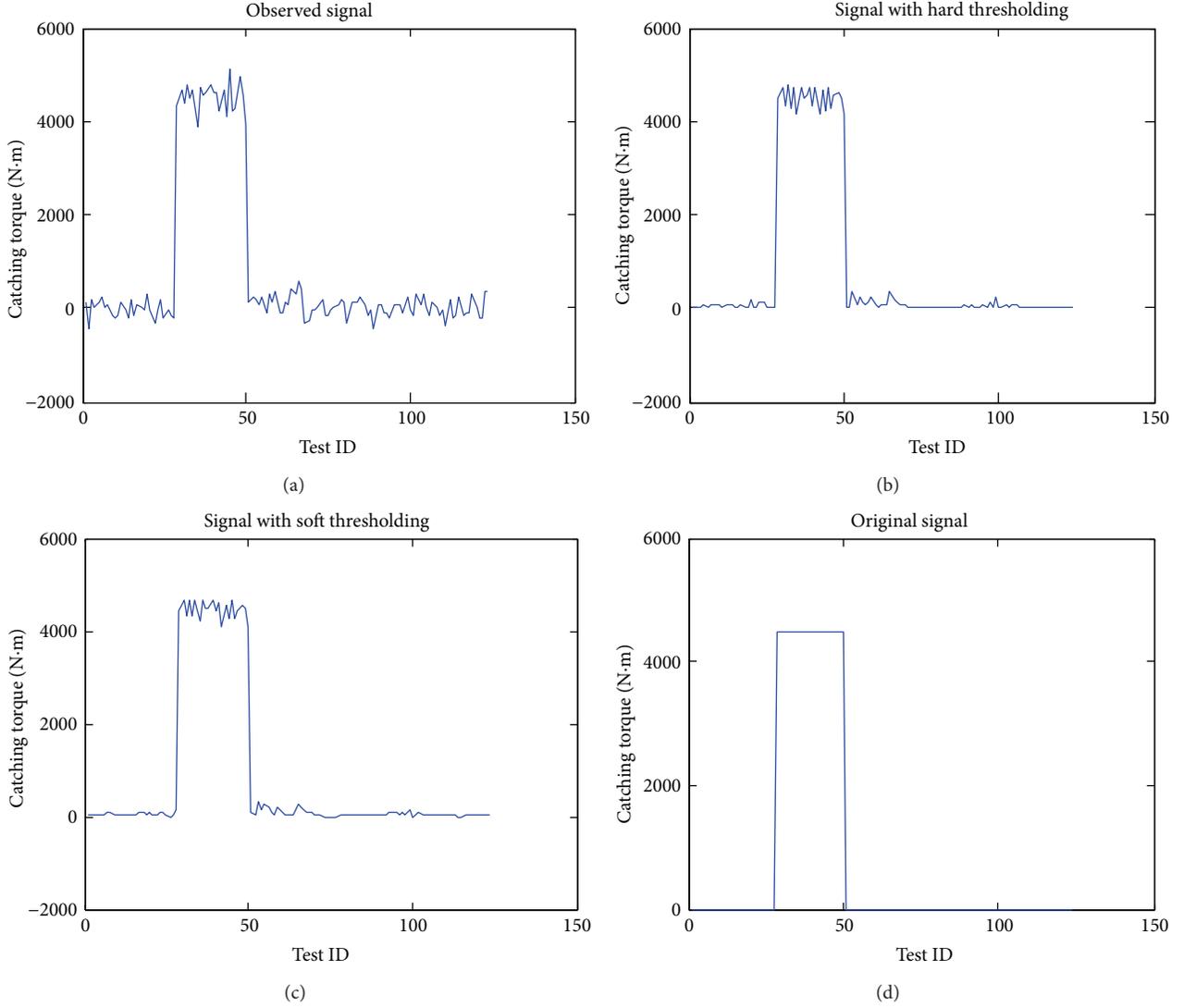


FIGURE 5: Comparisons of catching torque.

deceleration section is computed by the following equations, respectively:

$$v_{i1} = v_{01} + \alpha_1 i = 142.72 + (0.04 + i - 1) * \frac{737.78}{60},$$

$$(i = 1 \sim 29),$$

$$(17)$$

$$v_{i2} = v_{02} + \alpha_2 i = 156 - (i - 29) * \frac{10173.91}{60},$$

$$(i = 29 \sim 52),$$

where  $v_{01} = 142.72$  rpm,  $\alpha_1 = (156 - 142.72) * 60/1.08 = 737.78$  rpm<sup>2</sup>,  $v_{02} = 156$  rpm, and  $\alpha_2 = -(156) * 60/(2 - 1.08) = -10173.91$  rpm<sup>2</sup>.

The original rotation speed  $s_1$  is the assembly of  $v_{i1}$  and  $v_{i2}$ ; that is,

$$s_i = \{v_{i1}, v_{i2}\}. \quad (18)$$

The theoretical catching torque is computed by the following equation according to datasheet of antifalling safety device:

$$C_r = (1.5 \times 8 \times 15) \times \frac{M}{2}, \quad (19)$$

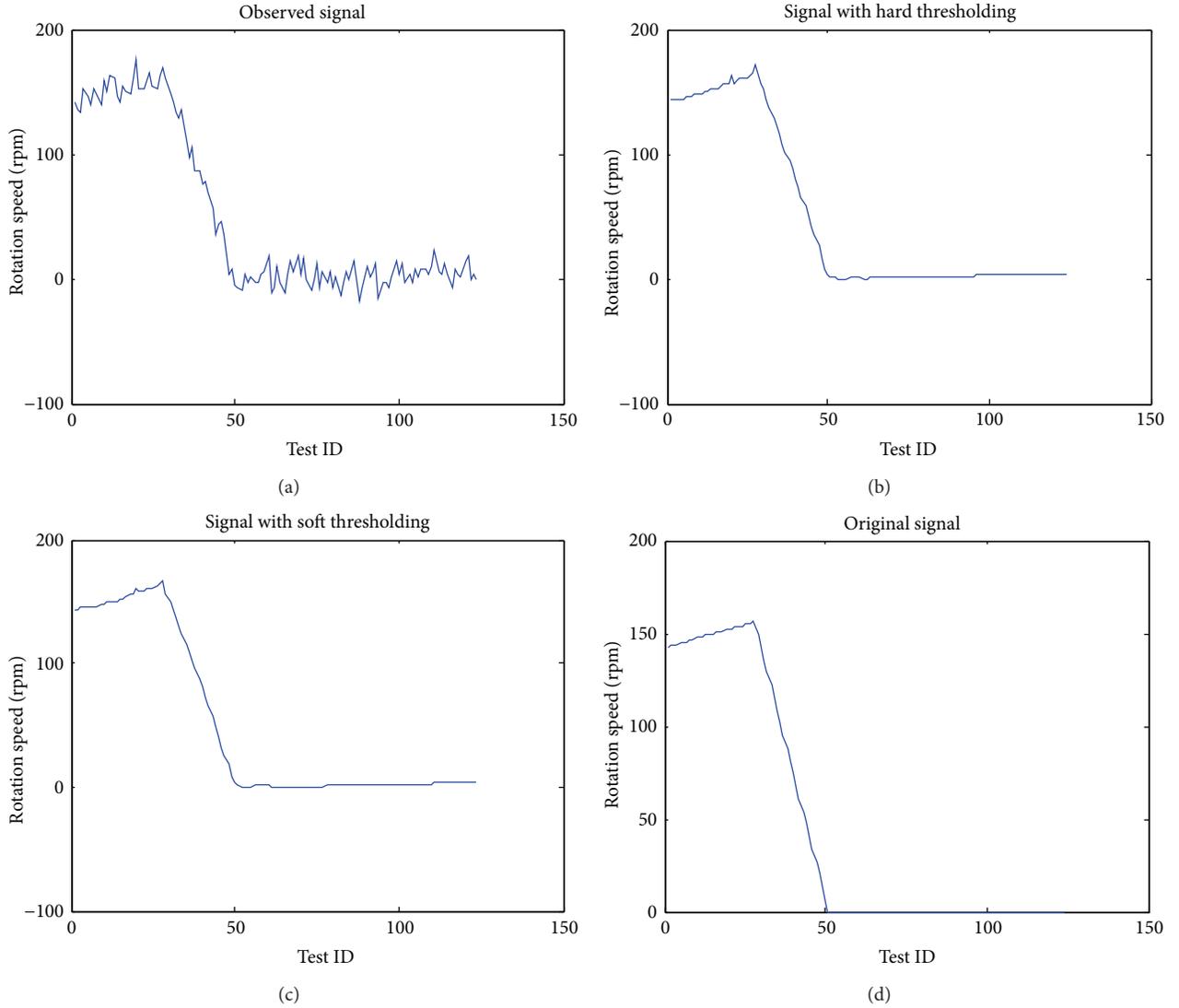


FIGURE 6: Comparisons of rotation speed.

TABLE 6: The SNR results of rotation speed with different wavelet scale coefficient (dB).

Wavelet scale coefficient	1	2	3	4	5	8	10
SNR (dB)	23.16	25.03	26.22	27.17	27.17	27.16	27.17

TABLE 7: The SNR and MSE results of rotation speed with different wavelet functions of  $\text{sym}N$ .

	$\text{sym}N$					
	3	4	5	6	7	8
SNR (dB)	26.62	26.71	26.83	27.10	27.12	27.17
MSE	0.5217	0.5072	0.4956	0.4898	0.4703	0.4699

where  $M$  is the rated catching load which is 44.8 kN; hence, the original catching torque  $C_r$  is 4032 N·m.

TABLE 8: The SNR and MSE of denoised signal and observed data.

	Denoised rotation speed	Observed rotation speed	Denoised catching torque	Observed catching torque
SNR (dB)	27.17	26.15	17.33	16.28
MSE	0.4699	0.6612	36.7545	41.9621

TABLE 9: Catching distance of antifalling safety device (m).

Number	Compression of disc spring			Catching distance (m)
	$a_1$ (mm)	$a_2$ (mm)	$ a_1 - a_2 $ (mm)	
1	23.1	26.8	3.7	0.697
2	26.8	29.4	2.6	0.490
3	29.4	31.5	2.1	0.396

TABLE 10: Detected catching torque and the maximum catching load.

Number	Rated catching load (kN)	Detected catching torque (kN·m)	Catching load measurement	
			Computed catching load (kN)	Maximum catching load (kN)
1	44.8	3.028	33.644	
2	44.8	3.870	43.000	50.322
3	44.8	4.529	50.322	

TABLE 11: The computed results with wavelet transform denoised data.

Number	Tripping rotation speed (rpm)	Catching torque (Nm)	Spring compression (mm)	Catching distance (m)	Catching load (kN)
101585	164.95	4819	3.6	0.679	53.544
	157.98	4425	2.3	0.434	49.167
101586	160.06	4461	1.8	0.339	49.567
	157.28	3839.2	5.1	0.96	42.661
101587	155.56	3537.8	5.5	1.037	39.309
	156.42	3811.4	5.1	0.961	42.349
101588	157.41	3793.6	5.2	0.980	42.151
	157.46	3834.2	5.1	0.961	42.602
101589	157.49	3383.4	5.7	1.074	37.593
	157	3443.8	5.6	1.056	38.264
101590	155.35	3544.9	5.5	1.037	39.388
	157.43	3815.6	5.1	0.96	42.396
101591	155.31	3659.9	5.4	1.02	40.666
	157.31	3581.6	5.4	1.02	39.796
101592	157.49	3363.8	5.8	1.09	37.376
	154.19	3478.7	5.5	1.04	38.652
101593	150.96	3383.6	5.7	1.07	37.60
	150.43	3589.1	5.4	1.02	39.88
101594	151.92	3622.9	5.4	1.02	40.25
	144.88	3413	5.6	1.06	37.92

A denoising signal method using soft heuristic thresholding at level 4 by sym8 wavelet is proposed in this paper, in which an improved threshold value is given by (9).

Two thresholding methods hard thresholding and soft thresholding are compared from level 1–10 by sym8 wavelet. The SNR results of denoised rotation speed of two thresholding option are shown in Table 5. It is shown in this table that the maximum SNR is obtained when using soft thresholding at level 4.

The SNR of rotation speed signal with different wavelet scale coefficient  $j$  ( $j = 1, 2, 3, 4, 5, 8, 10$ ) by soft thresholding is shown in Table 6. According to this table, when wavelet scale coefficient  $j$  is from 4 to 10, an improved SNR of about 27.17 dB is achieved.

It can be shown in Table 7 that sym8 is better than sym3 to sym7. This can be explained by the fact that the wavelet basis with higher order approximation can achieve better denoising results than that of lower order approximation.

On the other hand, the higher order approximation may cause the original signal vanished. Therefore the actual wavelet basis has to be ensured by different experiments.

The SNRs of rotation speed signal with different wavelet functions of sym $N$  ( $N = 3, 4, \dots, 8$ ) at level 4 are listed in Table 7.

The SNR results of the denoised signal with soft thresholding and the observed data are compared at level 4 by sym8 wavelet and the results are shown in Table 8.

From Tables 7 and 8, it is shown that the MSE using a soft thresholding at level 4 by sym8 wavelet is smaller than that by other sym $N$  and in all cases smaller than MSE using hard thresholding. The SNR using sym8 wavelet is the highest among different wavelet functions of sym $N$ . Moreover, SNR of denoised signal is higher than the detected signal. In summary, the performance of the denoising method of this paper is improved compared to that of hard thresholding. Additionally, the denoising method with soft thresholding at

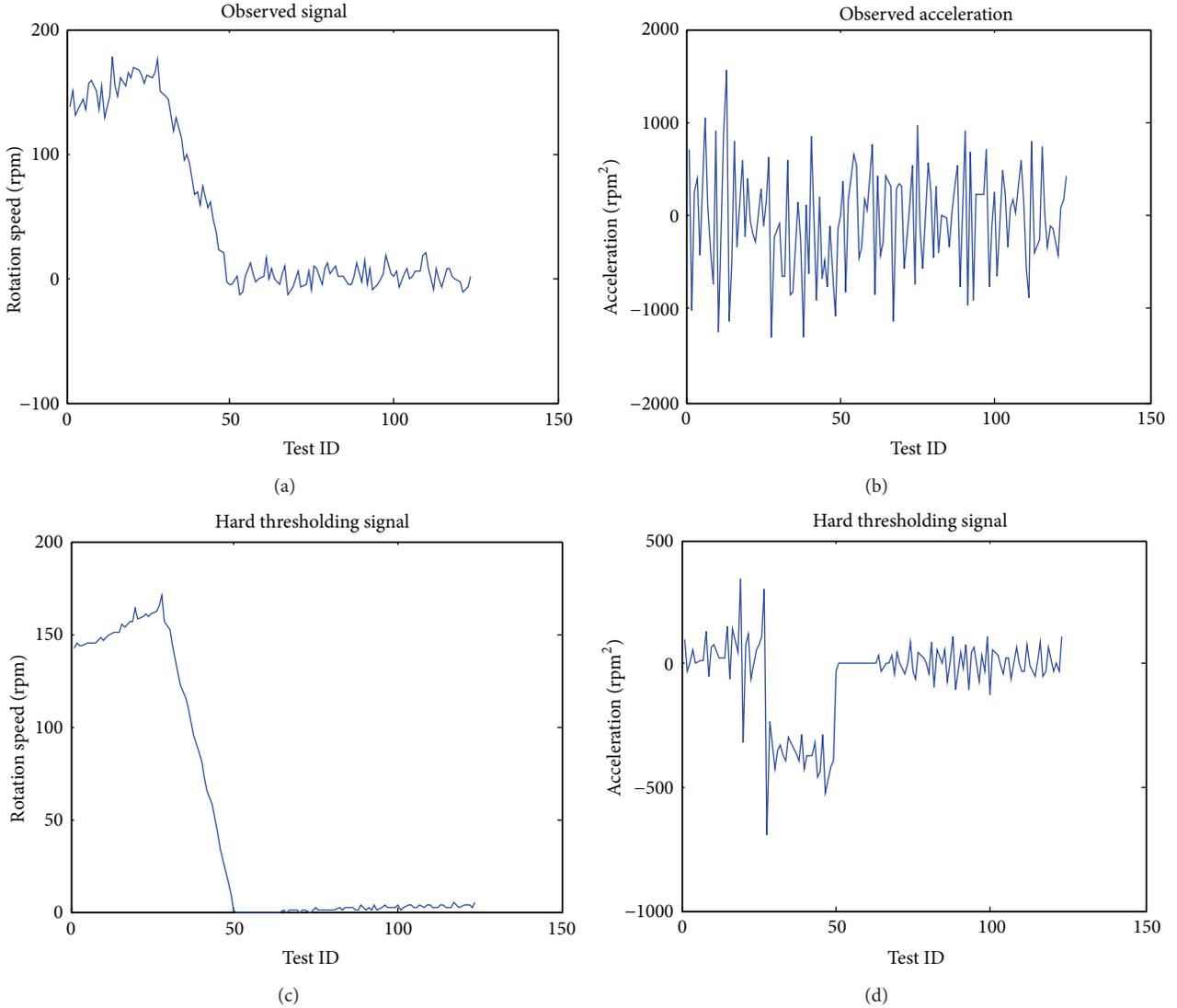


FIGURE 7: Comparisons of rotation speed and acceleration.

level 4 by sym8 wavelet achieves a better SNR and MSE than that at other level by sym3 to sym7. The SNR of denoised rotation speed and catching torque is increased from 26.15 dB to 27.17 dB which is 3.9% improvement. Similarly, 6.4% improvement of SNR of denoised catching torque is achieved.

With the wavelet transform, the data are smoothed and the errors are reduced. The figures show certain noise suppression, the denoised catching torque, and rotation speed have a relatively small defect and signal amplitude loss.

The computed catching distance is shown in Table 9.

It is specified in Chinese specification GB/T10054 that the maximum catching distance must be less than 2 m. Therefore, the diagnosed results in Table 9 show that the safety device is qualified.

The examples of the catching torque and the maximum catching load are shown in Table 10.

Ten groups of data of the original catching torque and the denoised catching torque with wavelet transform during

the industrial detection process are selected randomly and shown in Figure 8. From this figure, it can be observed that the computation results with observed signal distribute much loosely, which may confuse the operator and the customer. When the signal is denoised, the results are closer which is more reliable.

Ten groups of denoised data with wavelet transform for computation and diagnosis are shown in Table 11. In this table, column “number” represents the ID of an antifalling safety device. Each device is tested twice. Signals in column “tripping rotation speed” and column “catching torque” are denoised by wavelet transform. The data in “spring compression” is measured manually. Catching distance column is obtained with (12) and catching load with (13).

Based on the denoising signal method using thresholding with wavelet transform, the catching load of the antifalling safety device is more accurate and reliable than the observed signals.

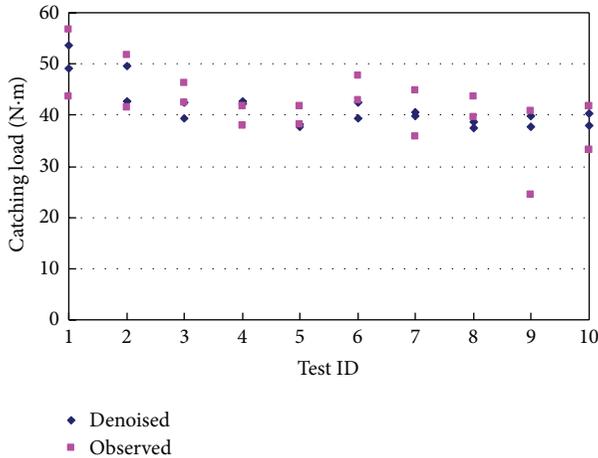


FIGURE 8: Catching load comparisons of observed signal and denoised signal.

## 5. Conclusions

Condition diagnosis of an antifalling safety device depends largely on the rotation speed and catching torque signals measured on the platform because the signals carry dynamic information about the machine state. However, the measured signal for fault diagnosis is difficult because the signal contains strong noise. Noise in the actual signal may lead to misrecognition and unreliability of useful information for the condition diagnosis. To solve these problems, we proposed a wavelet based condition diagnosis method for antifalling safety device. A noise reduction method employing hard and soft thresholding of DWT is proposed to denoise two important signals in the condition diagnosis. The wavelet transform with thresholding shows a better result than the measured rotation speed and catching torque signals. The experimental result shows that the signals are smoothed and the errors are reduced. An averagely maximum data error reduction of 7.5% is obtained. SNRs are improved by 3.9% and 6.4% for rotation speed and catching torque, respectively. The denoised signals improve the accuracy and reliability of diagnosis result for the antifalling safety device in construction elevator. The safety of elevator is ensured in high-rise building construction.

## Conflict of Interests

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

## Acknowledgments

This work is partially funded by Chongqing Technology and Business University through a Young Scholar seed Grant (Grant no. 1352015). The experiment environment and the antifalling safety device are provided by Chongqing Yingfengsheng Machinery and Equipment Co., Ltd. The authors thank for the efforts made by Professor Chen Shijiao,

Mr. Chen Yong, and Mr. Wang Hongyi during the experimental process.

## References

- [1] G. Tian, "Limit range of braking distance for anti-falling device with gear conedrum," *Construction Machinery*, vol. 2, pp. 25–28, 2000.
- [2] X.-J. Hao, "The analysis of auto detecting the brake of elevator using numerical arithmetic," *Microcomputer Information*, vol. 25, no. 5, pp. 310–312, 2009.
- [3] B. Zhang, T. Luan, and T. Ma, "On the importance of inspection of construction hoist anti falling safety device," *Value Engineering*, vol. 6, pp. 96–97, 2013.
- [4] P. J. Tavner, "Review of condition monitoring of rotating electrical machines," *IET Electric Power Applications*, vol. 2, no. 4, pp. 215–247, 2008.
- [5] J. Bai and J. Zhang, "Developing of anti-falling safety device detection system of construction hoist based on VB," *Gansu Science and Technology*, vol. 28, no. 7, pp. 53–55, 2012.
- [6] K. Li, P. Chen, and H. Wang, "Intelligent diagnosis method for rotating machinery using wavelet transform and ant colony optimization," *IEEE Sensors Journal*, vol. 12, no. 7, pp. 2474–2484, 2012.
- [7] S. Venugopal, R. A. Wagstaff, and J. P. Sharma, "Exploiting phase fluctuations to improve machine performance monitoring," *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 2, pp. 153–166, 2007.
- [8] H. Sun, K. Li, P. Chen, H. Wang, X. Ping, and Y. Cao, "A sequential fuzzy diagnosis method for rotating machinery using ant colony optimization and possibility theory," *Journal of Mechanical Science and Technology*, vol. 28, no. 4, pp. 1189–1201, 2014.
- [9] J. G. Mueller and T. G. Pratt, "A radio frequency polarimetric sensor for rotating machine analysis," *IEEE Sensors Journal*, vol. 13, no. 12, pp. 4866–4873, 2013.
- [10] S. S. Tayarani-Bathaie, Z. N. S. Vanini, and K. Khorasani, "Dynamic neural network-based fault diagnosis of gas turbine engines," *Neurocomputing*, vol. 125, pp. 153–165, 2014.
- [11] Z. Yang, B. W.-K. Ling, and C. Bingham, "Fault detection and signal reconstruction for increasing operational availability of industrial gas turbines," *Measurement*, vol. 46, no. 6, pp. 1938–1946, 2013.
- [12] Y. Kim, J. W. Chong, K. H. Chon, and J. Kim, "Wavelet-based AR-SVM for health monitoring of smart structures," *Smart Materials and Structures*, vol. 22, no. 1, Article ID 015003, pp. 1–12, 2013.
- [13] K. V. Nguyen, "Comparison studies of open and breathing crack detections of a beam-like bridge subjected to a moving vehicle," *Engineering Structures*, vol. 51, pp. 306–314, 2013.
- [14] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [15] J. Li, B. Yu, and H. Fischer, "Wavelet transform based on the optimal wavelet pairs for tunable diode laser absorption spectroscopy signal processing," *Applied Spectroscopy*, vol. 69, no. 4, pp. 496–506, 2015.
- [16] C. Li, M. Liang, and T. Wang, "Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals," *Mechanical Systems and Signal Processing*, vol. 64–65, pp. 132–148, 2015.

- [17] C. Li and M. Liang, "Enhancement of oil debris sensor capability by reliable debris signature extraction via wavelet domain target and interference signal tracking," *Measurement*, vol. 46, no. 4, pp. 1442–1453, 2013.
- [18] K. Dziejciech, W. J. Staszewski, and T. Uhl, "Wavelet-based frequency response function: comparative study of input excitation," *Shock and Vibration*, vol. 2014, Article ID 502762, 11 pages, 2014.
- [19] M. Singh and R. Kumar, "Thrust bearing groove race defect measurement by wavelet decomposition of pre-processed vibration signal," *Measurement*, vol. 46, no. 9, pp. 3508–3515, 2013.
- [20] V. Muralidharan and V. Sugumaran, "Rough set based rule learning and fuzzy classification of wavelet features for fault diagnosis of monoblock centrifugal pump," *Measurement*, vol. 46, no. 9, pp. 3057–3063, 2013.
- [21] P. Rajeev and K. K. Wijesundara, "Energy-based damage index for concentrically braced steel structure using continuous wavelet transform," *Journal of Constructional Steel Research*, vol. 103, pp. 241–250, 2014.
- [22] J. Li, B. Yu, W. Zhao, and W. Chen, "A review of signal enhancement and noise reduction techniques for tunable diode laser absorption spectroscopy," *Applied Spectroscopy Reviews*, vol. 49, no. 8, pp. 666–691, 2014.
- [23] A. Azzalini, M. Farge, and K. Schneider, "Nonlinear wavelet thresholding: a recursive method to determine the optimal denoising threshold," *Applied and Computational Harmonic Analysis*, vol. 18, no. 2, pp. 177–185, 2005.
- [24] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [25] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals. Densities, and spectra from indirect and noisy data," in *Proceedings of Symposia in Applied Mathematics*, vol. 47, pp. 173–205, American Mathematical Society, 1993.
- [26] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *Applied and Computational Harmonic Analysis*, vol. 1, no. 1, pp. 100–115, 1993.
- [27] V. A. Tran, S. T. Quek, and W. H. Duan, "Sensor validation in damage locating vector method for structural health monitoring," *International Journal of Structural Stability and Dynamics*, vol. 11, no. 1, pp. 149–180, 2011.
- [28] N. Wu and Q. Wang, "Experimental studies on damage detection of beam structures with wavelet transform," *International Journal of Engineering Science*, vol. 49, no. 3, pp. 253–261, 2011.
- [29] S.-T. Quek, Q. Wang, L. Zhang, and K.-K. Ang, "Sensitivity analysis of crack detection in beams by wavelet technique," *International Journal of Mechanical Sciences*, vol. 43, no. 12, pp. 2899–2910, 2001.
- [30] Q. Wang and N. Wu, "Detecting the delamination location of beam with a wavelet transform: an experimental study," *Smart Materials and Structures*, vol. 20, no. 1, Article ID 012002, 2011.
- [31] Q. Wang and X. Deng, "Damage detection with spatial wavelets," *International Journal of Solids and Structures*, vol. 36, no. 23, pp. 3443–3468, 1999.
- [32] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, 1992.
- [33] D. T. Nguyen and S. T. Quek, "Reliability of switched model-based controller for vessel dynamic positioning with switching under estimated motion frequency," *Journal of Offshore Mechanics and Arctic Engineering*, vol. 132, no. 2, pp. 1–9, 2010.
- [34] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pa, USA, 1994.
- [35] H. Wen and Z. Zhang, "Improvement of denoising of nonlinear wavelet transform threshold value method," *Bulletin of Surveying and Mapping*, vol. 3, pp. 318–321, 2006.
- [36] Z. Ruizhen and S. Guoxiang, "An improved method for white noise reduction based on wavelet transform," *Journal of Xidian University*, vol. 27, no. 5, pp. 619–622, 2000.
- [37] J. J. G. de la Rosa, I. Lloret, A. Moreno, C. G. Puntonet, and J. M. Górriz, "Wavelets and wavelet packets applied to detect and characterize transient alarm signals from termites," *Measurement*, vol. 39, no. 6, pp. 553–564, 2006.
- [38] M. E. Alexander, R. Baumgartner, A. R. Summers et al., "A wavelet-based method for improving signal-to-noise ratio and contrast in MR images," *Magnetic Resonance Imaging*, vol. 18, no. 2, pp. 169–180, 2000.
- [39] P. S. Gokhale, "ECG signal De-noising using discrete wavelet transform for removal of 50 Hz PLI noise," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 5, pp. 81–87, 2012.

## Research Article

# Rolling Element Bearing Fault Diagnosis Based on Multiscale General Fractal Features

Weigang Wen,<sup>1</sup> Zhaoyan Fan,<sup>2</sup> Donald Karg,<sup>2</sup> and Weidong Cheng<sup>1</sup>

<sup>1</sup>*School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing 100044, China*

<sup>2</sup>*Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269, USA*

Correspondence should be addressed to Weigang Wen; [wgwen@bjtu.edu.cn](mailto:wgwen@bjtu.edu.cn)

Received 30 March 2015; Accepted 16 July 2015

Academic Editor: Chuan Li

Copyright © 2015 Weigang Wen 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.

Nonlinear characteristics are ubiquitous in the vibration signals produced by rolling element bearings. Fractal dimensions are effective tools to illustrate nonlinearity. This paper proposes a new approach based on Multiscale General Fractal Dimensions (MGFDs) to realize fault diagnosis of rolling element bearings, which are robust to the effects of variation in operating conditions. The vibration signals of bearing are analyzed to extract the general fractal dimensions in multiscales, which are in turn utilized to construct a feature space to identify fault pattern. Finally, bearing faults are revealed by pattern recognition. Case studies are carried out to evaluate the validity and accuracy of the approach. It is verified that this approach is effective for fault diagnosis of rolling element bearings under various operating conditions via experiment and data analysis.

## 1. Introduction

Rolling element bearings are common mechanical parts which are subject to damage. Faulty bearings often cause machine failure and even contribute to the disaster in industry. Therefore, fault diagnosis of rolling element bearing is necessary in condition monitoring of machines. It is critical that fault pattern identification of bearing is performed to prevent machine breakdown and reduce economic loss in early period.

Vibration signal analysis has been used extensively in various bearing condition monitoring techniques and has become one of the most important methods applied for bearing fault diagnostics. The vibration signals generated by faults in bearings have been widely studied. Many studies have developed sound theoretical bases and approaches to diagnose bearing failure [1–3]. Much research focuses on obtaining fault information through time and frequency domain signal processing techniques. It is well known that the impact vibration produced by rolling elements in bearing excites resonances of the surrounding structures. But analysis of the vibration signal is complicated due to the stochastic movement of rolling elements. A method based on envelope spectrum analysis becomes a primary way for bearing fault

diagnosis which was stated systematically in [4]. Various faults can be diagnosed through fault characteristic frequencies according to the bearing structure parameters. However, when the bearing rotational speed varies over time, the characteristic frequencies cannot be obtained in frequency domain. Order tracking (OT), which may involve extra computation or auxiliary equipment such as speedometer and tachometer, is used to remove speed fluctuation. The method is complete and complicated, to which many studies are related [5, 6]. And this technique is successful for a wide range of cases.

The specific characteristics of rolling element bearing vibration signal are not periodic, especially under variable speed. It is inappropriate to diagnose fault of rolling element bearing only by adopting traditional diagnosis techniques. Time-frequency domain methods have been adopted to implement bearing fault diagnosis, such as Wigner-Ville distributions (WVD) [7], empirical mode decomposition (EMD) [8], and wavelet transform (WT) [9]. Many kinds of features are extracted to represent the characteristics of vibration signal in different domains, for example, statistics of Root Mean Square (RMS), kurtosis, crest factor, correlation coefficient and spectrum, and independent component

analysis (ICA) [10]. Finally, detection of bearing fault can be implemented by intelligent learning methods, such as perceptron, artificial neural network (ANN), or support vector machines (SVM) [11–13]. There are increasing research works that combine traditional time-frequency domain methods and intelligent learning methods for varied conditions of speed and load [14]. This ensemble of methods is developed as advanced hybrid intelligent fault diagnosis for rolling element bearings.

The vibration signals of bearings, especially with faults, often show mutation and nonlinearity [15]. As we know, the vibration signals that are excited by impacts of rolling element present nonstationary characteristics. Moreover, the nonstationary characteristics caused by faults in bearing are often mixed with nonlinear factors due to the complexity of the structure and operating conditions of rolling bearing, such as instantaneous variations in rolling ball movement, changing speed, and various loads. The traditional signal analysis methods based on linear system fail to extract nonlinear features in vibration signal. In order to analyze the nonlinear signals of bearings, a series of advanced techniques have been applied to extract fault features. It has been discovered that nonlinear analysis could provide a great alternative way to extract fault features out of vibration signals. Many nonlinear methods, such as chaos, fractal dimension, Lyapunov exponent, and approximate entropy, have been investigated [16–18]. The results have shown that nonlinear method is an effective way for rolling bearing fault diagnosis.

Fractal dimensions are widely used in nonlinearity analysis because it can quantitatively characterize nonlinear behavior. The correlation dimension is used to reveal the fault feature of rotating machinery in [19]. The combination of box-counting dimension, information dimension, and correlation dimension is applied to realize bearing fault diagnosis in [20]. Wavelet packet fractal technology is also utilized to diagnose rotating machinery in [21]. Even so, many fractal dimensions are sensitive to a flurry of factors besides bearing faults. It was proposed that the nonlinear feature of correlation dimension is related to the length of signal, the embedded dimension, the time delay, and so on in the bearing fault diagnosis experiment [22]. Moreover, in practical applications, how to decide the threshold of the nonlinear features is quite a problem. Multiscale fractal dimension can describe local nonlinear feature in different scales. To address these problems, multiscale nonlinear features were introduced for bearing fault diagnosis [23, 24]. This paper combines nonlinear analysis and intelligent diagnostics to implement bearing fault diagnosis. Specifically, a whole methodology based on Multiscale General Fractal Dimensions (MGFDs) of vibration signal and pattern recognition method is proposed. General fractal dimension is defined and utilized to reveal the approximation and detail of vibration signal in different scales. Then a feature space is constructed through MGFDs. Finally, intelligent pattern recognition method is utilized to implement classification of fault pattern in the feature space.

The rest of this paper is organized as follows. In the second section, the definition of general fractal dimension is introduced. The principle and methodology of multiscale

general fractal dimensions are addressed. In the third section, the vibration signals of rolling element bearings under different conditions are collected and analyzed. The experimental parameters are optimized according to the effectiveness of the methodology. The feasibility and reliability of the methodology for different bearing faults in various conditions are also proven in this section. In the fourth section, the conclusions are presented in closing.

## 2. Principle and Methodology

*2.1. Preliminaries on Fractal Dimension.* Theoretical fractals are infinitely self-similar, iterated locally and globally which are not easily described in traditional Euclidean geometric language. Fractals are not limited to nonlinear geometric patterns but can also describe processes in time. So fractal properties in the vibration time series can be suggested because of its nonstationary and nonlinear characteristics.

Fractal patterns are characterized by fractal dimension that is a ratio providing a statistical index of complexity. Fractal dimension can describe the changing of pattern with scale at which it is measured. But in reality, fractal characteristics only exist in a certain scale. Fractal dimensions in different scale can be estimated, respectively.

There are many types of definition of fractal dimension and several methods available to estimate fractal dimension, such as box-counting dimension, correlation dimension, and information dimension. The fractal dimension method is essentially a sequence of approximation associated with decreasing scale that is a geometric factor of simple figure forming the approximation [25]. Here, a general fractal dimension based on the time series cover is introduced to approximate the signal in time domain. This kind of fractal dimension concentrates on the changing pattern of time series.

*2.2. Principle of General Fractal Dimension.* In order to investigate fractal dimension of vibration signal in time domain, a two-dimensional graph can be made for vibration time series by way of the sample time as  $X$ -coordinate and the signal amplitude as  $Y$ -coordinate. According to the principle of fractal dimension, there is

$$M_\delta(S) \sim c\delta^D \quad \text{at } \delta \rightarrow 0, \quad (1)$$

where  $M_\delta(S)$  is approximation area of the sampled signal trajectory  $S$ ,  $\delta$  is the scale, and  $D$  is fractal dimension of the trajectory. Based on the principle of cover fractal dimension, the fractal dimension of sampled vibration signal can be calculated by minimal cover of series ichnography.

Suppose  $y = f(t)$  is signal function in domain of closed time interval  $[a, b]$ ; the domain is divided into  $m$  sections of  $\delta$ , the division is  $\omega_m = [a = t_0 < t_1 < \dots < t_m = b]$ , and  $\delta = (b - a)/m$ . To cover these  $m$  ( $m = 1, 2, \dots$ ) sections, every minimal cover of each section is a rectangle of which the length is  $\delta$  and the height is  $A_i(\delta)$  that is the difference between maximal value and minimal value of the signal in

section  $[t_{i-1}, t_i]$  ( $i = 1, 2, \dots, m$ ). The minimal cover of the time series of closed interval  $[a, b]$  is

$$M(\delta) = \sum_{i=1}^m M_i(\delta) = \sum_{i=1}^m A_i(\delta) \delta. \quad (2)$$

Define variance  $A(\delta)$  as

$$A(\delta) = \sum_{i=1}^m A_i(\delta). \quad (3)$$

According to the definition of time series fractal dimension, set  $D_\mu$  as minimal cover dimension:

$$M(\delta) = A(\delta) \delta \sim \delta^{D_\mu}. \quad (4)$$

As we know

$$\langle A(\delta) \rangle = \frac{1}{m} \sum_{i=1}^m A_i(\delta), \quad (5)$$

where  $\langle \rangle$  denotes the average within the time domain, because  $m * \delta$  is equal to the length of time series so  $m^{-1} \sim \delta$ , and there is

$$\langle A(\delta) \rangle \sim \delta^{D_\mu}. \quad (6)$$

With  $m$  increasing, the scale factor  $\delta$  is decreasing. An approximation of vibration time series in decreasing scale is made by this way. For the signal time series when  $\delta \rightarrow 0$

$$\langle |f(t+\delta) - f(t)| \rangle \sim \langle A(\delta) \rangle. \quad (7)$$

So

$$\langle |f(t+\delta) - f(t)| \rangle \sim \delta^{D_\mu}, \quad (8)$$

where  $D_\mu$  is a kind of general fractal dimension of the sampled time series. It denotes statistical property and fractal property of the time series.  $D_\mu$  can be calculated as

$$D_\mu = \lim_{\delta \rightarrow 0} \frac{\ln(\langle A(\delta) \rangle)}{\ln(\delta)}. \quad (9)$$

If same fractal characteristics exist in all the scales, a straight line in the graph of  $\ln(\langle A(\delta) \rangle) \sim \ln(\delta)$  can be fitted by least square method, and the slope of this line gives us an approximate estimation of fractal dimension. It can represent the change of vibration time series pattern. The time series minimum covering method can be completely independent from affine scaling of signal amplitude range. The general fractal dimension method is robust to variations in operating conditions.

**2.3. Methodology of Multiscale Fractal Dimensions.** For real-world signals with fractal structures, a single global fractal dimension at all scales is impossible. The practical fractal dimension of signal is also dependent on the used scale. Hence, a single noninteger number is not enough to represent entire complexity of a signal. In order to solve this

deficiency in characterization of the signal, multiscale fractal dimensions methodology is developed. Unlike global fractal dimension estimated by the slope of log-log curve, the multiscale fractal dimension scheme estimates local fractal dimensions along the scales [26].

For the scales  $\delta_1, \delta_2, \dots, \delta_m$  ranking from small to large, the local fractal dimension is estimated by calculating the slope of a line segment fitted by least squares over the adjacent scales in  $\ln(\langle A(\delta) \rangle) \sim \ln(\delta)$  plane. In this way, MGFs can describe a signal by a series of fractal dimensions along the scales. The computation process is listed as follows.

- (1) A section of accelerometer data are collected for computing. Here, a section of time series including  $L$  ( $L = 2^{13}$ ) sampled points is truncated for each evaluation in the experiments.
- (2) The value of  $\ln(\langle A(\delta) \rangle)$  and  $\ln(\delta)$  with increasing scale  $\delta$  is computed, where  $\delta = L/2^m$  ( $m = 1, 2, \dots, 13$ ).
- (3) The series of local fractal dimensions are estimated through the adjacent points on  $\ln(\langle A(\delta) \rangle) \sim \ln(\delta)$  plane. A series of MGFs are obtained by this way.
- (4) A fractal feature space is constructed through MGFs as the input of pattern recognition to identify fault patterns.

Here, the most popular intelligent methods of  $K$ -nearest neighbor classifier (KNNC), back-propagation neural networks (BPNNs), and least squares support vector machines (LS-SVMs) are selected as pattern recognition for training and testing [27–29]. Among these three classifiers, the KNNC algorithm predicts the test sample's category according to the  $K$  training samples which are the nearest neighbors to the test sample, and judge it to the category which has the largest category probability. BPNNs are constructed by three layers of neurons: input layer, hidden layer, and output layer. BPNNs are able to represent any functional relationship between input and output if there are enough neurons in the hidden layers. LS-SVMs are a class of kernel-based learning methods, which are a set of related supervised learning methods that analyze data and recognize patterns. In the experiments,  $K = 3$  is set for KNNC. The number of input neurons equals the dimension of feature space, the output neurons equal the type of bearing faults, and the hidden layer units are set to 3 in three levels of BPNNs. And the kernel function of LS-SVMs is 2nd order polynomial kernel as

$$K(x, x_i) = \left( 1 + \frac{x_i^T x}{c} \right)^d, \quad (10)$$

where polynomial degree  $d = 2$ .

By this methodology, MGFs are estimated on the signal amplitude-time plane. They describe quantitatively nonlinear features of vibration signal through minimum covers that represent transformation of the vibration signal contour in different scales. We do not need to take into account the influence of the embedded dimension, time delay, signal shift, and so forth. The only parameters that need to be considered



FIGURE 1: Bearing components with faults.

in computation are the length of signal and division of scales. Real-time calculation can be performed in the engineering field. So MGFs can reveal the approximate and detailed essence of vibration signal in different scales, which are seldom interfered with by external condition, for example, the variation of rotating speed and noise. The detailed verification of this methodology via experiments is described in the next section.

### 3. Experiments and Discussion

**3.1. Experiments of Different Bearing Faults.** A series of vibration signals of rolling element bearings with different faults were acquired from a rolling element bearing test rig. In the experiments, the rolling bearings were NSK-6000 deep groove ball bearings. A single point fault was introduced to the test bearings, respectively, by electrodischarge machining with fault diameters of 0.3 mm, 0.6 mm, and 1.0 mm. The rolling element bearing components with faults are shown in Figure 1. Four data sets of normal condition, ball fault, inner race fault, and outer race fault were sampled from the experimental system with a sampling frequency of 12 kHz. The motor rotating speed was set to 1500 rpm at first.

The vibration signals and the envelope power spectrums of different bearings are shown in Figure 2. The envelope spectrum of the normal bearing is relatively flat in (a). The inner race fault characteristic frequency of 154.5 Hz and the outer race fault characteristic frequency of 104 Hz can be distinguished easily from the envelope spectrum in (b) and

(c). The ball fault characteristic frequency of 137 Hz in (d) is not clear because of the more random motion of the balls.

The multiscale general fractal dimensions of these four types of bearings calculated by the presented MGFs method are shown in Figure 3. There are obvious distinctions among the four types of bearings along the scale  $\delta$ . The normal bearing displays the largest fractal dimension in small scales and decreases smoothly. The fractal dimensions of inner race faulty bearing and outer race faulty bearing rise in different middle scales. The fractal dimension of the ball faulty bearing fluctuates in large scales. The vibration signals of different types of bearings show diverse fractal dimensions in different scales which provide a way to diagnose the faults in bearing.

**3.2. Experiments under Variable Speed.** The vibration signal under increasing rotational speed from 500 rpm to 3500 rpm is shown in Figure 4. The characteristic frequency cannot be extracted from the envelope spectrum, but the shapes of multiscale general fractal dimensions do not change much with variation of shaft rotating speed as shown in Figure 5. MGFs of vibration signals sampled under variable rotating speeds, for example, varying from 500 rpm to 3000 rpm, and varying from 500 rpm to 3500 rpm, 1000 rpm, 1500 rpm, 2000 rpm, 2500 rpm, 3000 rpm, and 3500 rpm, are shown in Figure 5. The legend label of 500–3000 means outer race faulty bearing with speed varying from 500 rpm to 3000 rpm, and N500–3000 means normal bearing with speed varying from 500 rpm to 3000 rpm, and so on. The fold lines represent the multiscale fractal dimensions, among which the blue ones are normal bearings and the red ones are outer race faulty bearings. The vibration signals of bearings in the same state under different speeds show almost similar multiscale fractal dimensions in Figure 5.

**3.3. Experiments with White Noise.** When different levels of white noise are added to the vibration signal, MGFs of normal bearing and inner faulty bearing are shown in Figure 6. The legend label of IS0 means inner race faulty bearing signal without noise, IS2 means inner race faulty bearing signal with added white noise with SNR = 2, NS0 means normal bearing signal without noise, and NS2 means normal bearing signal with added white noise with SNR = 2, and so on. The fold lines represent the multiscale general fractal dimensions, among which the blue ones are the normal bearings signal with white noise and the red ones are the inner race faulty bearings signal with white noise. Figure 6 shows that MGFs are almost robust to the white noise with SNR ranging up to 2.

**3.4. Classification of Multiscale General Fractal Dimensions.** In this section, the ability of MGFs to distinguish different bearing conditions in varied speed is evaluated. The experiments of four types of bearings, for example, normal bearing, inner race faulty bearing, outer race faulty bearing, and ball faulty bearing under six different rotating speeds of 1000 rpm, 1500 rpm, 2000 rpm, 2500 rpm, 3000 rpm, and 3500 rpm, were implemented, respectively. Ten vibration signal samples containing 8192 points were selected from each experiment. Thus, there were sixty samples for each

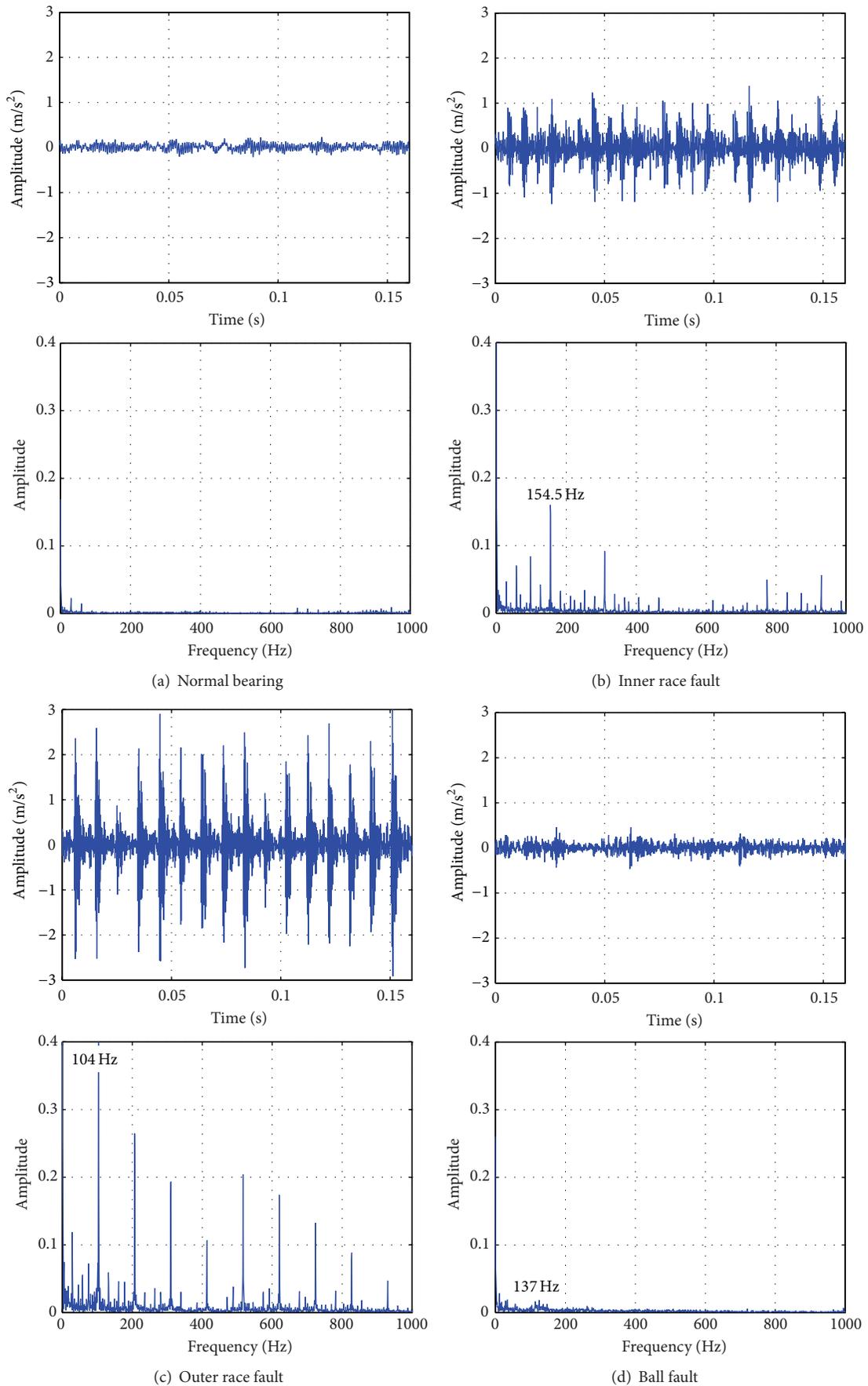


FIGURE 2: Vibration signal and envelope power spectrum of different bearings.

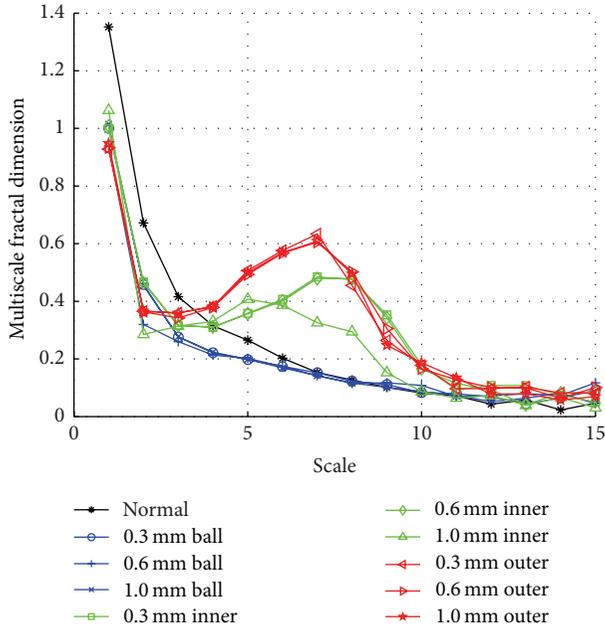


FIGURE 3: Multiscale general fractal dimensions of different bearings.

type of bearing. 240 samples were collected for training and testing of pattern recognition in total. For each type of bearing, thirty samples were randomly selected for training and remaining thirty samples are used for testing. Twelve scale general fractal dimensions were computed from each sample as the input feature vector for pattern recognition. Hence, a training dataset with dimension  $120 \times 12$  and a testing dataset with the same dimension were obtained. KNNC, BPNNs, and LS-SVMs were employed to classify the four types of bearings under variable speeds. In order to get a more accurate evaluation, the processes of random selection, training, and testing were repeated 100 times. Finally, average testing classification error rates were chosen as the evaluation standards.

For comparison, the methods based on wavelet packet (WP) and empirical mode decomposition (EMD) were used to classify the same datasets. In wavelet packet decomposition, the discrete Meyer wavelet was utilized to decompose the vibration signal into vectors of coefficients. The WP decomposition was applied up to wavelet packet level of  $J = 3$ . So  $2^J = 8$  vectors  $C_{J,k}(i)$  were produced in the  $J$ th level, where  $k = 0, 1, \dots, 2^J - 1$  and  $i = 1, \dots, I$ . Each vector contained approximately  $I = N_t/2^J$  coefficients that implicated the information in a specific frequency band of the signal. Here the statistical features, for example, kurtosis, skewness, and standard deviation, were calculated from each wavelet coefficient vector [30, 31]. Therefore, 24 features can be calculated for each sample as the input feature vector of pattern recognition based on statistical analysis and wavelet packet decomposition.

In empirical mode decomposition, each vibration signal can be decomposed into a number of Intrinsic Mode Functions (IMFs) [32, 33]. Thus, we can achieve a decomposition

TABLE 1: The classification error rates of pattern recognition.

	KNNC	BPNN	SVM
MGFDs	0.54%	4.34%	0.30%
Energies of EMD	1.67%	9.08%	8.04%
Statistics of WP	6.25%	7.62%	11.6%

of signal into  $n$ -empirical modes and a residue  $r_n$ . The coefficients of IMFs  $C_1, \dots, C_n$  include different frequency bands ranging from high to low. The frequency components contained in each frequency band are different and they change with the variation of original signal. Normalized energies of IMFs  $E_i/E_T$  constitute the input feature vectors for pattern recognition, where  $E_i$  ( $i = 1, \dots, n$ ) is the energy of each IMF and  $E_T$  is the total energy of the signal. Here, the EMD level was set as  $n = 14$ . So, 14 features could be calculated for each sample based on EMD. Finally the processes of random selection, training, and testing were repeated 100 times to evaluate effectiveness of these methods in the same way.

The average testing classification error rates of MGFDs, the normalized IMF energies of EMD, and the WP-statistic features using three classifiers on four bearing datasets under varied speed are given in Table 1. The average classification error rates in the table show clearly that MGFDs are always outperforming the other methods by using different classifiers on four bearing datasets. It reveals the capability of MGFDs to distinguish bearing fault under varied speed.

**3.5. Optimization of MGFDs.** As we discussed above, the result of MGFDs could be affected by several external factors, for example, bearing rotational speed, noise, selection of the scales of MGFDs, and the length of the samples. It has been verified that MGFDs were robust to varied speed and external white noise. Here the effect of scale and length of samples are evaluated.

The figures of MGFDs (Figures 3, 5, and 6) show that the distinctions among MGFDs at small scales are significant, while the distinctions become blurred when the scale becomes larger and all MGFDs tend to zero. This is due to the average minimum cover of signal remaining unchanged when the scale becomes large enough. In order to implement real-time computation and accuracy of this methodology, the size of scales and the length of samples can be determined according to the testing classification error rates of pattern recognition.

The number  $m$  of MGFDs from 1 to 12 corresponding to the scale  $\delta = L/2^m$  was chosen for assessment. For each choice of  $m$ , the processes of training and testing were repeated for 100 times. Then the average classification error rates were obtained for determining the size of scale and the length of samples. The results of average classification error rates are shown in Figure 7. It can be observed that when  $m \geq 10$ , the average classification error rates of three classifiers are under 1%. If the size of the scale is too large, the accuracy of classification could decline because of too much distracting information. Therefore, the size can be set

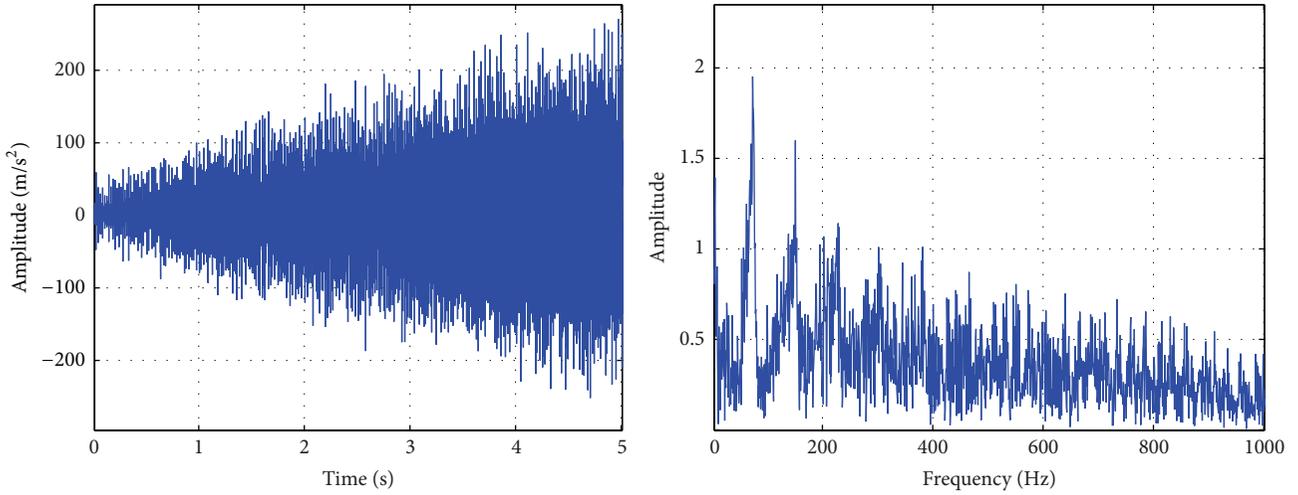


FIGURE 4: Vibration signal and envelope power spectrum under variable speed.

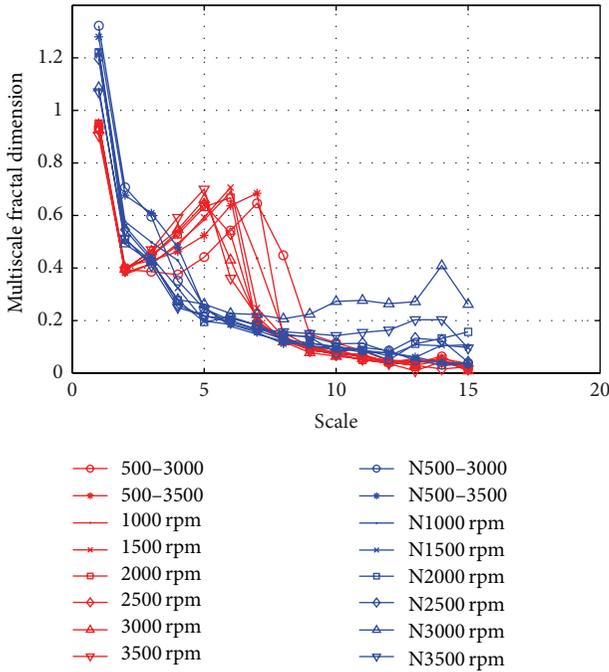


FIGURE 5: Multiscale general fractal dimensions under variable speed.

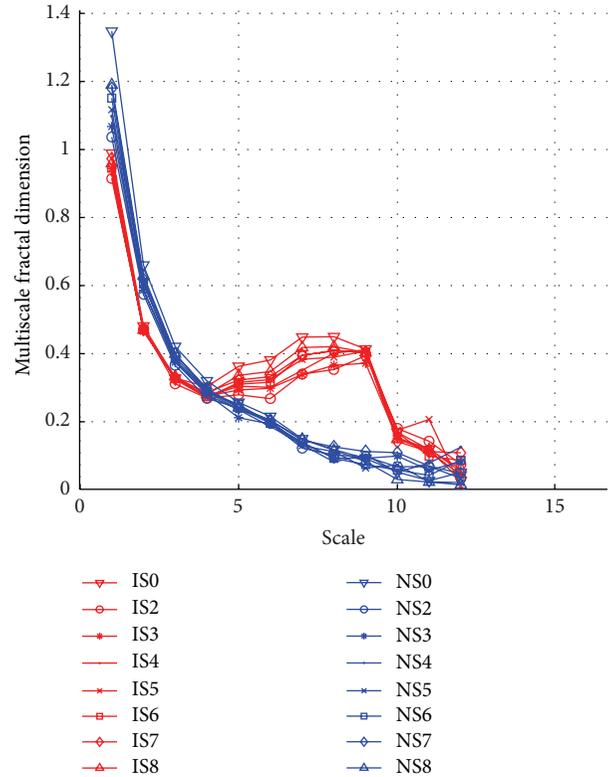


FIGURE 6: Multiscale general fractal dimensions with different SNR white noise.

to  $\delta = L/2^m$  ( $m = 1, 2, \dots, 10$ ) and the length of samples can be set to  $L = 2^{12}$ .

#### 4. Conclusions

In this paper, a novel rolling element bearing fault diagnosis approach based on multiscale fractal features of vibration signal was presented. This approach combined fractal theory and intelligent pattern recognition methods to implement bearing fault diagnosis. The methodology was proposed to

classify different types of rolling element bearings under different operating conditions.

Multiscale General Fractal Dimensions (MGFDs) were defined to illustrate the fractal feature of rolling element bearing vibration signal at first. The vibration signals demonstrated different fractal structures when different deficiencies developed in rolling element bearing. MGFDs could provide

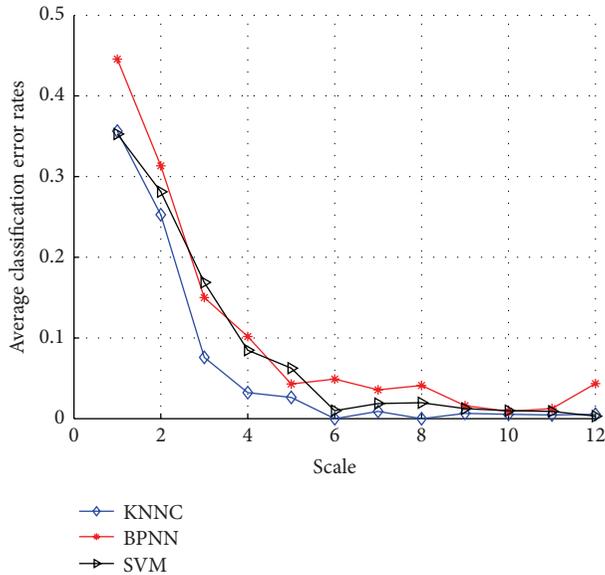


FIGURE 7: Diagram of average classification error rate versus scale.

a lot of discriminative information which revealed the change of vibration signal pattern with the scale. Through theoretical analysis and experimental verification, the approach could accurately classify the rolling element bearing states, for example, ball fault, inner race fault, outer race fault, and normal state under varied operating conditions. MGFs were robust to variation of the rotational speed and external white noise in the experiments.

The performance of MGFs and other methods, such as EMD and WPD, were evaluated by using three classifiers of KNNC, BPNNs, and LS-SVMs in comparative experiments. It has been observed that MGFs outperformed other approaches in varied operating conditions. Lastly, the size of scales and length of samples were optimized through testing classification error rates that ensured the reliability and effectiveness of this approach. It has been demonstrated that MGFs were capable of revealing fault of rolling element bearings accurately and could be generalized for fault diagnosis of other rotating machines in the future.

### Conflict of Interests

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

### Acknowledgments

This project is supported by the Fundamental Research Funds for the Central Universities (2011JBM093) and National Natural Science Foundation of China (51275030). Weigang Wen would like to express appreciation to Professor Robert X. Gao and the Electromechanical Systems Lab at the University of Connecticut. The authors would like to thank the anonymous peer reviewers for their valuable suggestions.

### References

- [1] R. M. Jones, "Enveloping for bearing analysis," *Sound & Vibration*, vol. 30, no. 2, pp. 10–15, 1996.
- [2] 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.
- [3] C. Bujoreanu, V. Horga, and B. Dragan, "Vibration analysis methods in bearing damage detection," *Applied Mechanics and Materials*, vol. 371, no. 1, pp. 622–626, 2013.
- [4] 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.
- [5] L. Guo, J. Chen, and X. Li, "Rolling bearing fault classification based on envelope spectrum and support vector machine," *Journal of Vibration and Control*, vol. 15, no. 9, pp. 1349–1363, 2009.
- [6] J. Luan, H. Kang, H. Zheng, Q. Cui, and J. Cao, "Application in fault diagnosis of bearing with order envelope spectrum analysis," *Journal of Vibration, Measurement and Diagnosis*, vol. 26, no. 9, pp. 215–217, 2006.
- [7] Y. Zhou, J. Chen, G. M. Dong, W. B. Xiao, and Z. Y. Wang, "Wigner-Ville distribution based on cyclic spectral density and the application in rolling element bearings diagnosis," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 225, no. 12, pp. 2831–2847, 2011.
- [8] H. Dong, K. Qi, X. Chen, Y. Zi, Z. He, and B. Li, "Sifting process of EMD and its application in rolling element bearing fault diagnosis," *Journal of Mechanical Science and Technology*, vol. 23, no. 8, pp. 2000–2007, 2009.
- [9] 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.
- [10] Y. Guo, J. Na, B. Li, and R.-F. Fung, "Envelope extraction based dimension reduction for independent component analysis in fault diagnosis of rolling element bearing," *Journal of Sound and Vibration*, vol. 333, no. 13, pp. 2983–2994, 2014.
- [11] D. Kateris, D. Moshou, X.-E. Pantazi, I. Gravalos, N. Sawalhi, and S. Loutridis, "A machine learning approach for the condition monitoring of rotating machinery," *Journal of Mechanical Science and Technology*, vol. 28, no. 1, pp. 61–71, 2014.
- [12] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, no. 3, pp. 16–27, 2015.
- [13] D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, and A. Alonso-Betanzos, "Automatic bearing fault diagnosis based on one-class m-SVM," *Computers and Industrial Engineering*, vol. 64, no. 1, pp. 357–365, 2013.
- [14] R. Zeng, L. Zhang, and Y. Xiao, "A method combining order tracking and fuzzy c-means for diesel engine fault detection and isolation," *Shock and Vibration*. In press.
- [15] J. Zheng, J. Cheng, and Y. Yang, "Multiscale permutation entropy based rolling bearing fault diagnosis," *Shock and Vibration*, vol. 2014, Article ID 154291, 8 pages, 2014.
- [16] W. Caesarendra, B. Kosasih, A. K. Tieu, and C. A. S. Moodie, "Application of the largest Lyapunov exponent algorithm for

- feature extraction in low speed slew bearing condition monitoring,” *Mechanical Systems and Signal Processing*, vol. 50-51, pp. 116–138, 2015.
- [17] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, and C.-C. Wang, “Bearing fault diagnosis based on multiscale permutation entropy and support vector machine,” *Entropy*, vol. 14, no. 8, pp. 1343–1356, 2012.
- [18] W. J. Wang, Z. T. Wu, and J. Chen, “Fault identification in rotating machinery using the correlation dimension and bispectra,” *Nonlinear Dynamics*, vol. 25, no. 4, pp. 383–393, 2001.
- [19] D. Logan and J. Mathew, “Using the correlation dimension for vibration fault diagnosis of rolling element bearings. I. Basic concepts,” *Mechanical Systems and Signal Processing*, vol. 10, no. 3, pp. 241–250, 1996.
- [20] J. Yang, Y. Zhang, and Y. Zhu, “Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension,” *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2012–2024, 2007.
- [21] C.-H. Chen, R.-J. Shyu, and C.-K. Ma, “A new fault diagnosis method of rotating machinery,” *Shock and Vibration*, vol. 15, no. 6, pp. 585–598, 2008.
- [22] D. B. Logan and J. Mathew, “Using the correlation dimension for vibration fault diagnosis of rolling element bearings—II. Selection of experimental parameters,” *Mechanical Systems and Signal Processing*, vol. 10, no. 3, pp. 251–264, 1996.
- [23] P.-L. Zhang, B. Li, S.-S. Mi, Y.-T. Zhang, and D.-S. Liu, “Bearing fault detection using multi-scale fractal dimensions based on morphological covers,” *Shock and Vibration*, vol. 19, no. 6, pp. 1373–1383, 2012.
- [24] G. Xiong, W. Huang, and L. Zhang, “Fault severity identification of rolling bearing based on multiscale entropy,” *Journal of Applied Sciences*, vol. 13, no. 13, pp. 2404–2408, 2013.
- [25] M. M. Dubovikov, N. V. Starchenko, and M. S. Dubovikov, “Dimension of the minimal cover and fractal analysis of time series,” *Physica A*, vol. 339, no. 3-4, pp. 591–608, 2004.
- [26] A. R. Backes and O. M. Bruno, “Shape classification using complex network and Multi-scale Fractal Dimension,” *Pattern Recognition Letters*, vol. 31, no. 1, pp. 44–51, 2010.
- [27] E. Bax, “Validation of  $k$ -nearest neighbor classifiers,” *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3225–3234, 2012.
- [28] F. Li, C. Wu, K. Wu, and J. Xu, “An improved back propagation neural network model and its application,” *Journal of Computers*, vol. 9, no. 8, pp. 1858–1862, 2014.
- [29] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [30] P. K. Kankar, S. C. Sharma, and S. P. Harsha, “Rolling element bearing fault diagnosis using wavelet transform,” *Neurocomputing*, vol. 74, no. 10, pp. 1638–1645, 2011.
- [31] N. G. Nikolaou and I. A. Antoniadis, “Rolling element bearing fault diagnosis using wavelet packets,” *NDT & E International*, vol. 35, no. 3, pp. 197–205, 2002.
- [32] Z. Shen, X. Chen, X. Zhang, and Z. He, “A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM,” *Measurement*, vol. 45, no. 1, pp. 30–40, 2012.
- [33] Y. Yang, D. Yu, and J. Cheng, “A roller bearing fault diagnosis method based on EMD energy entropy and ANN,” *Journal of Sound and Vibration*, vol. 294, no. 1, pp. 269–277, 2006.

## Research Article

# Screw Performance Degradation Assessment Based on Quantum Genetic Algorithm and Dynamic Fuzzy Neural Network

**Xiaochen Zhang, Hongli Gao, and Haifeng Huang**

*School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China*

Correspondence should be addressed to Hongli Gao; [ghl2248@hotmail.com](mailto:ghl2248@hotmail.com)

Received 16 April 2015; Revised 23 June 2015; Accepted 1 July 2015

Academic Editor: Wahyu Caesarendra

Copyright © 2015 Xiaochen Zhang 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.

To evaluate the performance of ball screw, screw performance degradation assessment technology based on quantum genetic algorithm (QGA) and dynamic fuzzy neural network (DFNN) is studied. The ball screw of the CINCINNATIV5-3000 machining center is treated as the study object. Two Kistler 8704B100M1 accelerometers and a Kistler 8765A250M5 three-way accelerometer are installed to monitor the degradation trend of screw performance. First, screw vibration signal features are extracted both in time domain and frequency domain. Then the feature vectors can be obtained by principal component analysis (PCA). Second, the initialization parameters of the DFNN are optimized by means of QGA. Finally, the feature vectors are inputted to DFNN for training and then get the screw performance degradation model. The experiment results show that the screw performance degradation model could effectively evaluate the performance of NC machine screw.

## 1. Introduction

NC machine tool is an important foundation for modern manufacturing, which directly impacts the economic development [1, 2]. With the continuous development of modern manufacturing technology, NC machine tool has been widely used in aerospace, automobile, shipbuilding, and other industries [3, 4]. In the manufacturing process, the continuous degradation of ball screw directly leads to the decrease of NC machine tool's static geometric accuracy and dynamic precision, which means that enterprises should increase their production and maintenance costs. Since ball screw's performance would change greatly with the variation of working conditions and processing environment, replacement of the ball screw regularly is neither scientific nor reasonable [5, 6]. Therefore, screw performance degradation needs real-time assessment in order to cut down enterprises' maintenance costs.

DFNN is a kind of information processing method combined with fuzzy set theory. Its essence is a dynamic mapping network with fuzzy input signal and weights [7].

During the learning process, network parameters and dimension will change according to the rules. Compared with the traditional neural network, DFNN is more suitable for describing dynamic system. However, it is difficult to choose the initialization parameters of the DFNN, which means DFNN will easily fall into local optimum [8, 9]. To the defect of DFNN, this paper adopts the QGA, which has the advantage of the high efficiency and avoiding local optimum, to select the best initialization parameters of the DFNN that can improve the performance of DFNN and increase the operation stability of system.

The purpose of this paper is to present a useful method for performance degradation assessment of ball screw based on QGA and DFNN. Meanwhile feature vectors selection method is proposed here. Particularly, we extract screw vibration signal features both in time domain and frequency domain. The dimensionalities of the input signal feature space are reduced with the help of PCA. Then the initialization parameters of the DFNN are optimized by means of QGA. Screw performance degradation model can be gotten after training with the feature vectors. Finally, we compare

the prediction accuracies among different kinds of neural networks to examine the effectiveness of the proposed method.

## 2. Screw Vibration Feature Vectors

**2.1. Screw Online Monitoring System.** Figure 1 shows the screw online monitoring system applied to CINCINNATIV5-3000 machining center. Two Kistler 8704B100M1 one-way accelerometers are installed at two bearing chocks to monitor the radial vibration of screw ends, while a Kistler 8765A250M5 three-way accelerometer is installed on the screw nut to monitor three-dimensional shaking of the screw nut. INV1870 is the signal conditioner connected with Advantech PCII710 data acquisition card. Industry computer can store the vibration data through the PCII710 data acquisition card, and the sampling frequency is 2.56 kHz. Ball screw works in horizontal installation state; meanwhile the installation method of the ball screw is one-end fixed while the other end floated. Ball screw reaches 45 mm in axis diameter, 12 mm in lead, and 762 mm in journey.

**2.2. Signal Analysis and Feature Extraction.** With the increase of service life, ball screw's performance will gradually reduce; meanwhile the vibration of the screw fixed end changes gradually. After more than five years of service life, the vibration increased significantly. Failure occurs at the seventh year of service life. Vibrations of screw fixed end under different service life are showed in Figure 2.

Considering the distribution characteristics of the ball screw vibration signal, the original features consist of the following parts. Time domain or frequency domain features (such as root mean square value, peak value, and gravity frequency), which are presently used to reflect the time domain or frequency domain, are extracted as a part of the original signal features. By using the wavelet analysis method, the vibration signal is decomposed into 5 levels selecting wavelet "db1," and distinct time-frequency features based on wavelet packet energy are obtained. Approximate entropy, a recently developed statistic theory in mechanical fault diagnosis, has also been applied to enrich the original features here.

The original features extracted by different methods, which have characteristics of high dimension and heavy computation, are not conducive to the online modeling and evaluation. Meanwhile, considering the certain correlation among the high dimensional original features, eliminating redundant information is one of the main focuses in the research of feature extraction.

PCA is a statistical method that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. The  $n$ -dimensional original features can be expressed as  $\mathbf{X} = [x_1, x_2, \dots, x_n]$ . By means of PCA, linearly uncorrelated feature vectors,  $\mathbf{Y} = [y_1, y_2, \dots, y_n]$ , can be obtained. The contribution of the  $i$ th component  $\eta_i$  can be expressed as follows:

$$\eta_i = \frac{\lambda_i}{\sum_{k=1}^n \lambda_k}, \quad (1)$$

where  $\lambda_i$  is the variance of the  $y_i$ .

Then the contributions of the first  $m$  principal component  $\eta'_m$  can be built as follows:

$$\eta'_m = \frac{\sum_{i=1}^m \lambda_i}{\sum_{k=1}^n \lambda_k}. \quad (2)$$

The first ten principal component contributions are plotted, as shown in Figure 3. The first three principal component contributions are more than 20%, and the accumulative contribution of the first five principal components is over 90%. Taking into account the real-time operation speed and operation precision, this paper takes the first five principal components as the feature vectors of the screw performance assessment system.

## 3. Screw Performance Degradation Assessment

**3.1. Dynamic Fuzzy Neural Network.** The DFNN has five layers, including input layer, output layer, and three hidden layers [10, 11]. Figure 4 shows the structure of DFNN. In Figure 4,  $x_1, x_2, \dots, x_r$  are the input variables,  $y$  is the system output,  $MF_{ij}$  is the  $j$ th membership function of the  $i$ th input variable,  $N_j$  is the  $j$ th normalized node, and  $w_j$  is the weight of the  $j$ th rule. The number of the system rules is  $u$ .

The main function of each layer can be explained as follows.

(1) Input layer: The variables  $x_1, x_2, \dots, x_r$  are input to the corresponding nodes, respectively.

(2) Membership function layer: Each node represents a membership function, which can be expressed as follows:

$$\mu_{ij}(x_i) = \exp \left[ -\frac{(x_i - c_{ij})^2}{\sigma_j^2} \right], \quad (3)$$

where  $\mu_{ij}$  ( $i = 1, 2, \dots, r; j = 1, 2, \dots, u$ ) is the  $j$ th membership function of the  $x_i$ .  $c_{ij}$  and  $\sigma_j$  represent the center and width of the  $\mu_{ij}$ , respectively.

(3)  $t$ -norm layer: In the layer, fuzzy rules are represented by nodes. The output of the  $j$ th node  $R_j$  is defined as

$$\phi_j = \exp \left[ -\frac{\sum_{i=1}^r (x_i - c_{ij})^2}{\sigma_j^2} \right] = \exp \left[ -\frac{\|\mathbf{X} - \mathbf{C}\|^2}{\sigma_j^2} \right], \quad (4)$$

$j = 1, 2, \dots, u.$

(4) Normalized layer: Normalization processing is made for the output of the  $t$ -norm layer in this layer. The output of  $j$ th node  $N_j$  can be expressed as follows:

$$\psi_j = \frac{\phi_j}{\sum_{k=1}^u \phi_k}, \quad j = 1, 2, \dots, u. \quad (5)$$

(5) Output layer: The system output which is the superposition of the input variables can be formulated as

$$y(\mathbf{X}) = \sum_{k=1}^u w_k \cdot \psi_k. \quad (6)$$

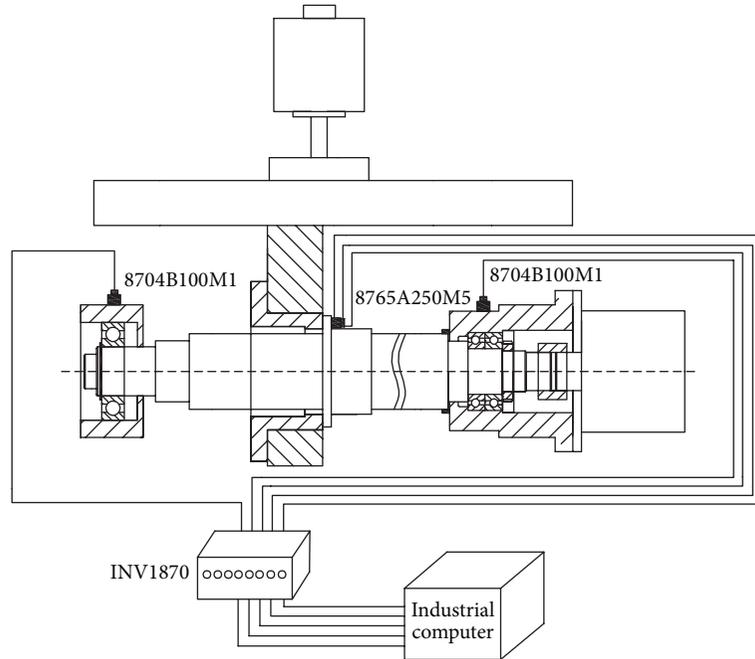


FIGURE 1: Screw online monitoring system.

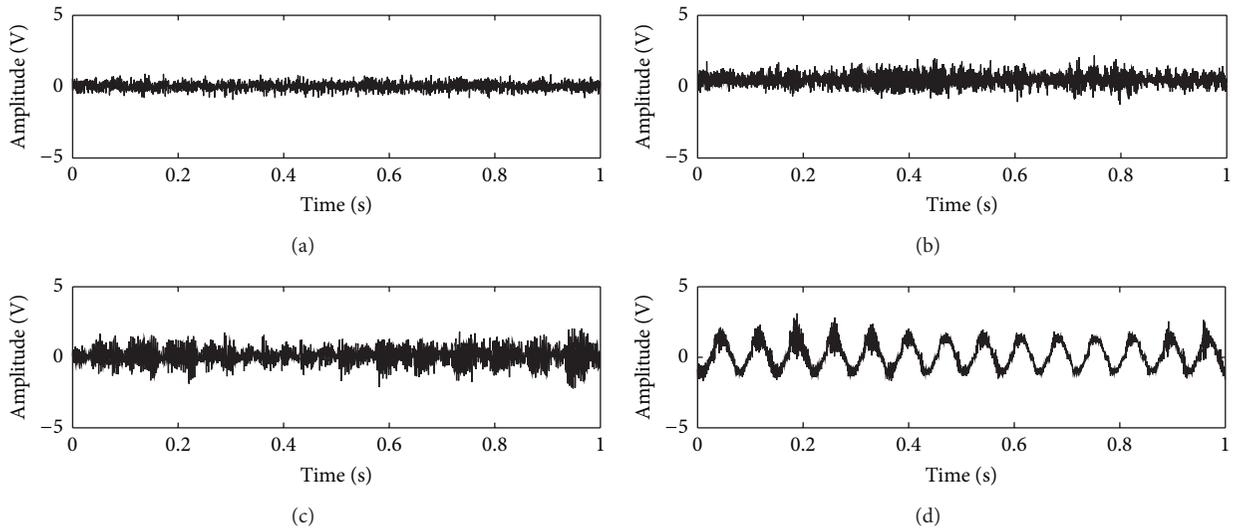


FIGURE 2: Vibrations of screw fixed end under different service life: (a) vibration of 0.5-year-service screw, (b) vibration of 2-year-service screw, (c) vibration of 5-year-service screw, and (d) vibration of 7-year-service screw.

Before generating the first rule, DFNN should set the initialization parameters of the network. Ten parameters of DFNN need to be initialized, including membership function width of the first rule  $\sigma_0$ , overlap factor of the radial based function  $k$ , width renewal factor  $k_w$ , rule threshold  $k_{err}$ , decay constant  $\gamma$ , convergence constant  $\beta$ , maximum debugging standard  $\varepsilon_{max}$ , minimum debugging standard  $\varepsilon_{min}$ , maximum output error  $e_{max}$ , and minimum output error  $e_{min}$ . The specific details of the ten initialization parameters are introduced in [12–15].

DFNN will easily fall into local optimum due to the random initialization of the network parameters. Therefore, this paper applies QGA to obtain the optimal initialization parameters of DFNN.

**3.2. Quantum Genetic Algorithm.** The main ideas of QGA can be expressed as follows: according to the parameter characteristics of DFNN, chromosomal genes with quantum bit coding system are constructed and the population that includes several chromosomes is generated. By adopting

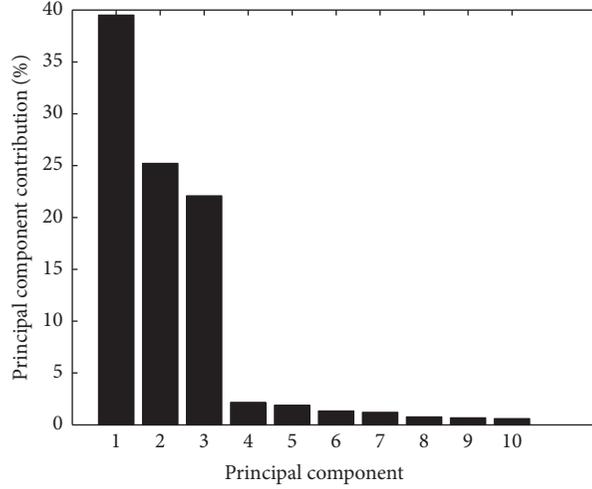


FIGURE 3: The first ten principal component contributions.

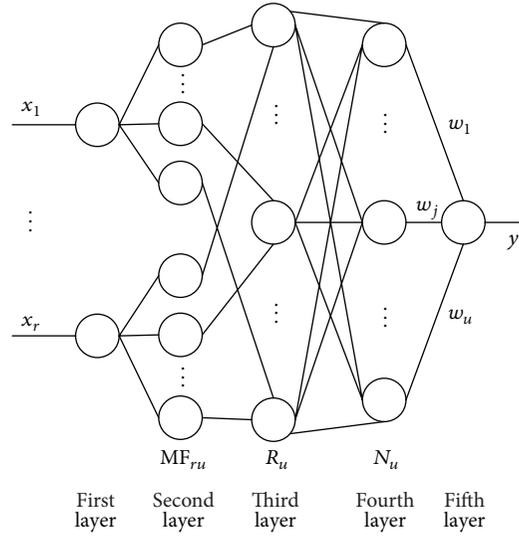


FIGURE 4: The structure of DFNN.

quantum cross, variation, and quantum rotation gate, the optimal initialization parameters of DFNN can be obtained. Figure 5 shows the algorithm flowchart of QGA.

**3.2.1. Population Initialization.** The population size of the initialized population  $\mathbf{Q}(t_0) = \{\mathbf{q}_1^{t_0}, \mathbf{q}_2^{t_0}, \dots, \mathbf{q}_N^{t_0}\}$  is  $N$ . By means of the quantum bit coding system, arbitrary chromosome  $\mathbf{q}_j^{t_0}$  of the initialized population  $\mathbf{Q}(t_0)$  can be expressed as follows:

$$\mathbf{q}_j^{t_0} = \begin{bmatrix} \alpha_{11}^{t_0} & \alpha_{12}^{t_0} & \dots & \alpha_{1k_1}^{t_0} & \alpha_{21}^{t_0} & \alpha_{22}^{t_0} & \dots & \alpha_{2k_2}^{t_0} & \dots & \alpha_{m1}^{t_0} & \alpha_{m2}^{t_0} & \dots & \alpha_{mk_m}^{t_0} \\ \beta_{11}^{t_0} & \beta_{12}^{t_0} & \dots & \beta_{1k_1}^{t_0} & \beta_{21}^{t_0} & \beta_{22}^{t_0} & \dots & \beta_{2k_2}^{t_0} & \dots & \beta_{m1}^{t_0} & \beta_{m2}^{t_0} & \dots & \beta_{mk_m}^{t_0} \end{bmatrix}, \quad (7)$$

where  $m$  is the number of the chromosomal genes and  $k_1, k_2, \dots, k_m$  represent the quantum bit number of each chromosomal gene, respectively.  $\mathbf{q}_j^{t_0}$  is the  $j$ th chromosome of the  $t_0$  generation. In this paper, chromosome  $\mathbf{q}_j^{t_0}$  includes ten genes, which are, respectively, corresponding to ten initialization parameters of DFNN. Initialization probability

amplitude  $[\alpha, \beta]^T$  is  $[1/\sqrt{2}, 1/\sqrt{2}]^T$  so that each chromosome expresses the same state. The fitness function is built as follows [16, 17]:

$$f = 1 - \frac{1}{s} \sum_{p=1}^s \left| \frac{y_p' - t_p}{t_p} \right|, \quad (8)$$

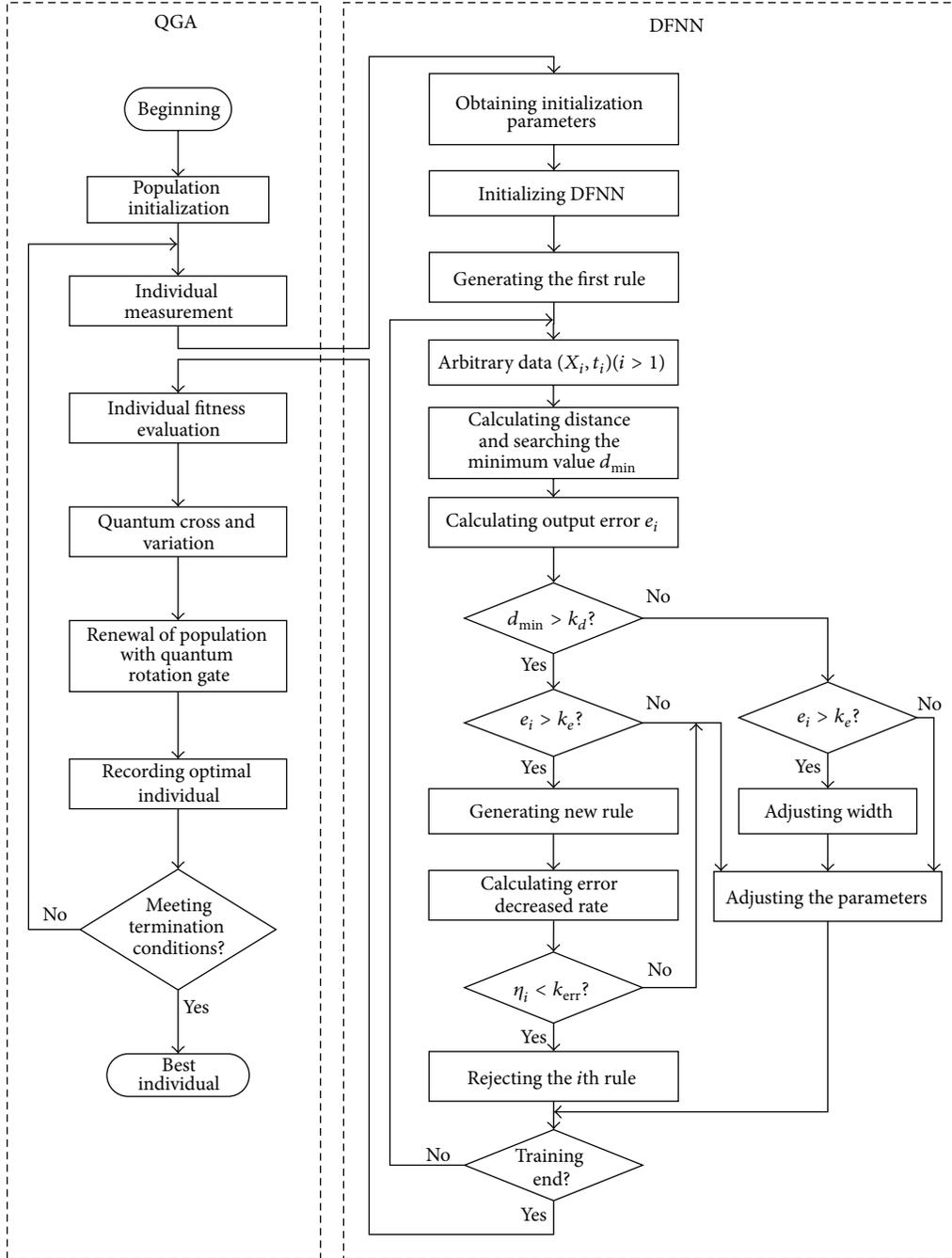


FIGURE 5: Algorithm flowchart of QGA.

where  $y'_p$  is the predictive value of the imitating prediction sample,  $t_p$  is the true value of the imitating prediction sample, and  $s$  is the number of the imitating prediction samples.

For DFNN training processing, the predictive values of the testing samples are prone to distortion though the predictive values of the training samples are very good. To avoid having a seriously distorted prediction, the fitness function and imitating prediction samples can be used. The fitness function takes into account the fitting degree of the

training samples and the portability of optimized DFNN. Therefore, the prediction accuracy of the optimized DFNN is guaranteed.

**3.2.2. Quantum Cross and Variation.** In order to avoid falling into population local optimum, quantum cross has been used here. With the help of quantum cross, new chromosomes are generated which means information exchange among chromosomes is realized. The cross processes are as follows.

TABLE 1: The adjustment strategy of the quantum rotating angle.

$x_i$	$b_i$	$f(x) > f(b)$	$\Delta\theta_i$	$s(\alpha_i, \beta_i)$			
				$\alpha_i\beta_i > 0$	$\alpha_i\beta_i < 0$	$\alpha_i = 0$	$\beta_i = 0$
0	0	TRUE/FALSE	0	0	0	0	0
0	1	FALSE	$\omega$	+1	-1	0	$\pm 1$
0	1	TRUE	$\omega$	-1	+1	$\pm 1$	0
1	0	FALSE	$\omega$	-1	+1	$\pm 1$	0
1	0	TRUE	$\omega$	+1	-1	0	$\pm 1$
1	1	TRUE/FALSE	0	0	0	0	0

- (1) Two chromosomes are randomly selected from the population, and whether cross operation should be considered is determined by cross probability.
- (2) If it is necessary to consider the cross operation, exchanging the random cross position information between two chromosomes is applied here.
- (3) Examining chromosome feasibility, cross operation is finished.

By using quantum variation, we can disturb the current evolution direction of the chromosome to avoid early maturity, so good global search capacity can be obtained.

**3.2.3. Quantum Rotation Gate.** Quantum gate is the actuator of the evolution process for QGA. The update process of quantum rotation gate is as follows:

$$\begin{bmatrix} \alpha' \\ \beta' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad (9)$$

where  $[\alpha \ \beta]^T$  and  $[\alpha' \ \beta']^T$  represent quantum bit probability amplitudes before and after updating, respectively.  $\theta$  is the quantum rotating angle. The adjustment strategy of the quantum rotating angle is shown in Table 1.

Where  $x_i$  and  $b_i$  are the  $i$ th bit of the current chromosome and the best chromosome, respectively,  $\Delta\theta_i$  is the adjusting angle step and  $s(\alpha_i, \beta_i)$  is the rotating angle direction. This paper adopts dynamic adjustment strategy based on expansion coefficient for quantum rotation gate [18–21]; rotating angle  $\omega$  is defined as follows:

$$\omega = \omega_{\min} + (\omega_{\max} - \omega_{\min}) \left[ 1 - \left( \frac{t_i}{t_{\max}} \right)^\varepsilon \right], \quad (10)$$

where  $\omega_{\min}$  and  $\omega_{\max}$  are the minimum value and maximum value of the  $\omega$ , respectively,  $t_i$  is the number of the current genetic generations,  $t_{\max}$  is the number of the maximum genetic generations, and  $\varepsilon$  is the expansion coefficient.

**3.3. Screw Performance Degradation Model.** Sensor signals are acquired by screw online monitoring system and high dimension features are extracted in both time domain and frequency domain. Then the feature vectors can be obtained by PCA. Feature vectors are sent to feature vectors library together with the real-time working conditions. Training samples and imitating prediction samples are randomly selected from the feature vectors library and used for training

DFNN and QGA. After training, testing samples selected from the feature vectors library are used to test the optimized and trained DFNN. If the prediction accuracy falls on the system error allowable range, this DFNN can be used as screw performance degradation model. Otherwise, the DFNN must be retrained by means of increasing samples or modifying network parameters until the prediction accuracy can be guaranteed. Flowchart of building screw performance degradation model is showed in Figure 6.

## 4. The Experiment Results

Screw vibration signals are not only determined by screw performance degradation degree, but also determined by working conditions. Feature vectors library which includes real-time working conditions is accumulated with the help of the screw online monitoring system. In this paper, one ball screw of CINCINNATIV5-3000 is used as the experiment object. Four class performance degradation samples of the ball screw are randomly selected from the feature vectors library together with their working conditions. The length of each screw performance sample is 10 seconds while the sampling frequency is 2.56 kHz. Four class screw performance samples include 0.5-year-service screw samples, 2-year-service screw samples, 5-year-service screw samples, and 7-year-service screw samples. Training samples include 400 samples (100 samples from each class), imitating prediction samples include 400 samples (100 samples from each class), and testing samples include 400 samples (100 samples from each class).

A mapping method used for describing the output of the screw performance degradation model is presented. In this method, different output intervals of the screw performance degradation model show different kinds of screw performance. Interval  $[0, 0.2)$  indicates good performance. Interval  $[0.2, 0.4)$  indicates the performance degraded slightly. Interval  $[0.4, 0.6)$  indicates middle degraded performance which can still ensure the machining accuracy. Interval  $[0.6, 0.8)$  presents serious degraded performance which means the screw is easy to go wrong. Interval  $[0.8, 1]$  means screw with failure. Combined with the field experiences, the model output of the 0.5-year-service screw samples, 2-year-service screw samples, 5-year-service screw samples, and 7-year-service screw samples is 0.1, 0.25, 0.7, and 0.9.

The QGA population scale and the number of the iterations should be based on comprehensive consideration of training samples and searching efficiency. The parameters of the quantum cross, variation, and rotation gate can be determined by experiences on the premise of fast searching optimization. The initialization parameters range of DFNN can be obtained with multiple algorithm running tests [22–25]. In this paper, parameters of the QGA are selected as follows [26, 27]: population scale is 40, the number of the iterations is 100, the initialization parameters interval of DFNN is  $[0, 3]$ , cross probability is 0.3 while variation probability is 0.1, maximum rotating angle  $\omega_{\max}$  is  $0.15\pi$  and minimum rotating angle  $\omega_{\min}$  is  $0.01\pi$ , and expansion coefficient  $\varepsilon$  is 2.

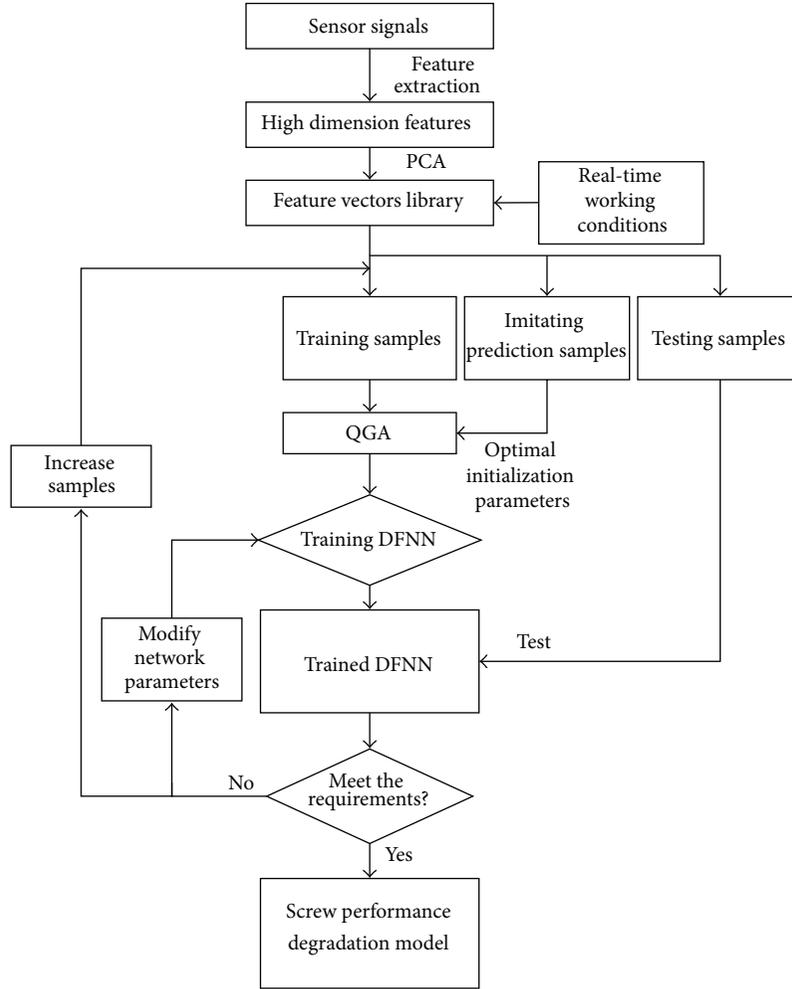


FIGURE 6: Flowchart of building screw performance degradation model.

TABLE 2: The prediction accuracies of two DFNN.

	$\sigma_0$	$k$	$k_w$	$k_{err}$	$\gamma$	$\beta$	$\epsilon_{max}$	$\epsilon_{min}$	$e_{max}$	$e_{min}$	Accuracy of training samples	Accuracy of testing samples
Unoptimized DFNN	2	1.5	2	0.1	0.9	0.9	2	0.1	2	0.1	0.7350	0.6975
Optimized DFNN	2.3896	1.0065	2.0226	0.0036	0.7432	0.1840	1.6578	0.3766	2.3861	0.5571	0.8225	0.8175

Table 2 shows the prediction accuracies of two DFNN. Compared with the unoptimized DFNN, the optimized DFNN has higher prediction accuracies in both training samples and testing samples. From the prediction accuracies of training samples and testing samples, it is clear that the optimized DFNN maintains a stable prediction accuracy.

Figures 7 and 8 show the confusion matrixes of unoptimized DFNN. It can be found that unoptimized DFNN could not distinguish 5-year-service screw very well, especially for testing samples. Figures 9 and 10 are the confusion matrixes of optimized DFNN. It can be seen that optimized DFNN shows better prediction accuracies. Figures 11 and 12 show the training and testing accumulated errors of two DFNN.

It is clear that optimized DFNN presents better prediction accuracies.

In order to validate the effect of the optimized DFNN, backpropagation (BP) neural network and radial basis function (RBF) neural network are trained and tested with the same samples. Table 3 shows the prediction accuracies of three networks. According to the results presented in Table 3, prediction accuracy of the optimized DFNN is better than that of BP neural network and RBF neural network.

Figures 13 and 14 are the confusion matrixes of BP neural network while Figures 15 and 16 are the confusion matrixes of RBF neural network. From Figure 13 to 16, it is found that 2-year-service screw and 5-year-service screw are

TABLE 3: The prediction accuracies of three networks.

	Optimized DFNN	BP neural network	RBF neural network
Accuracy of training samples	0.8225	0.6950	0.6925
Accuracy of testing samples	0.8175	0.6375	0.6450

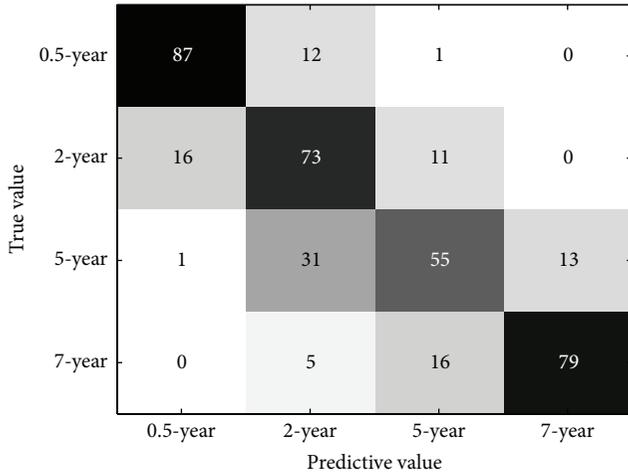


FIGURE 7: Unoptimized DFNN confusion matrix for training samples.

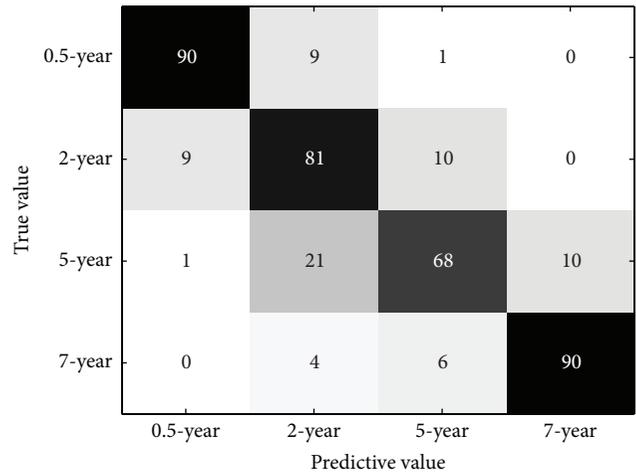


FIGURE 9: Optimized DFNN confusion matrix for training samples.

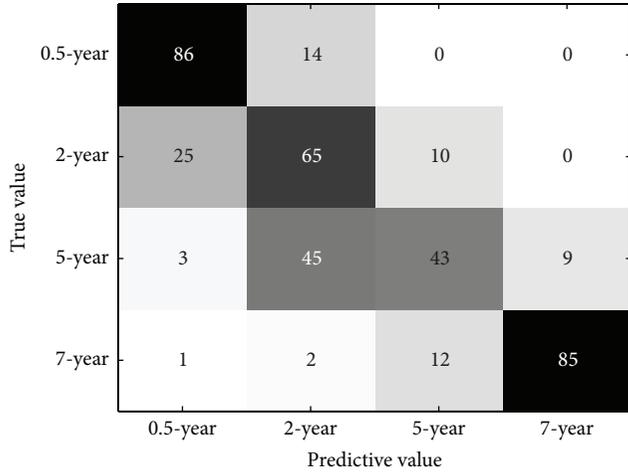


FIGURE 8: Unoptimized DFNN confusion matrix for testing samples.

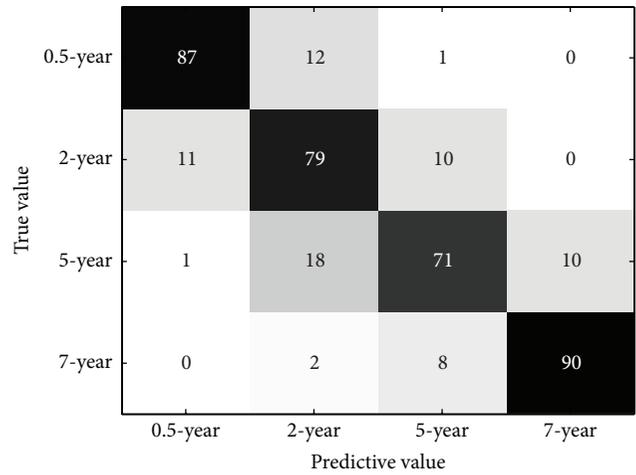


FIGURE 10: Optimized DFNN confusion matrix for testing samples.

difficult to distinguish both for BP neural network and for RBF neural network. It is known that screw performance degradation would obey the rule of bathtub curve; 2-year-service screw and 5-year-service screw are in random failure period. Therefore it is not easy to predict 2-year-service screw and 5-year-service screw. Compared with Figures 9 and 10, the optimized DFNN shows better prediction accuracies than BP neural network and RBF neural network.

Figures 17 and 18 show the training and testing accumulated errors of three networks. As seen in Figures 17 and

18, the accumulated errors of BP network and RBF network are larger than those of the optimized DFNN. As previously discussed, it can be concluded that the optimized DFNN can assess screw performance degradation effectively.

### 5. Conclusion

Screw performance degradation assessment based on QGA and DFNN is studied in this paper. The experiment results classified by the optimized DFNN show that our proposed

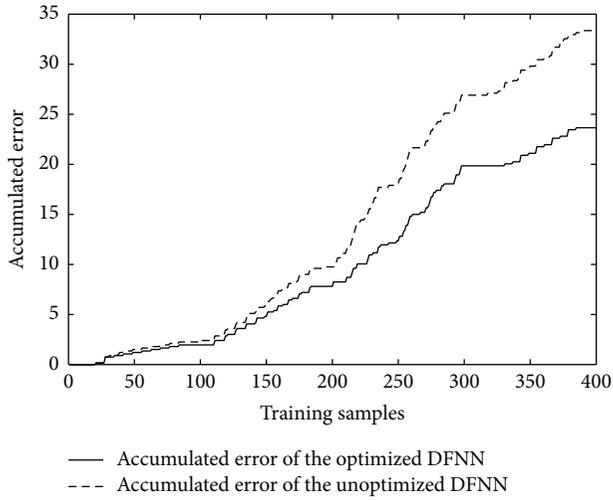


FIGURE 11: Training accumulated errors of two DFNN.

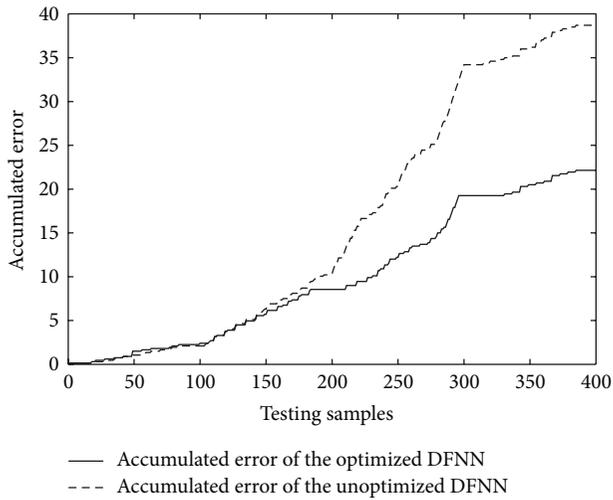


FIGURE 12: Testing accumulated errors of two DFNN.

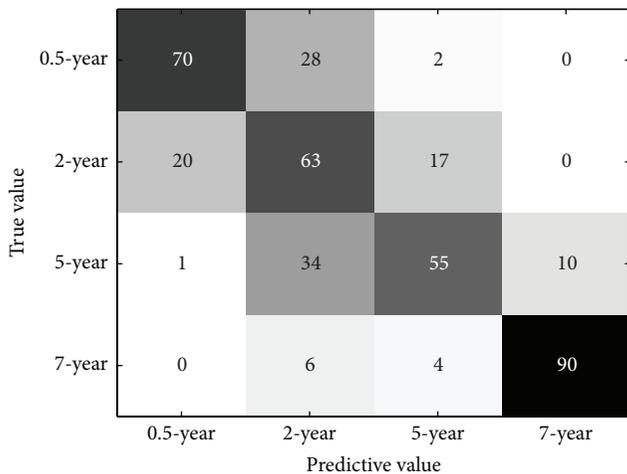


FIGURE 13: BP neural network confusion matrix for training samples.

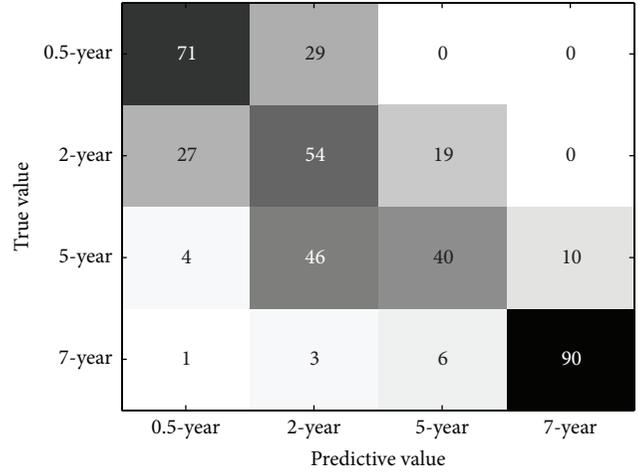


FIGURE 14: BP neural network confusion matrix for testing samples.

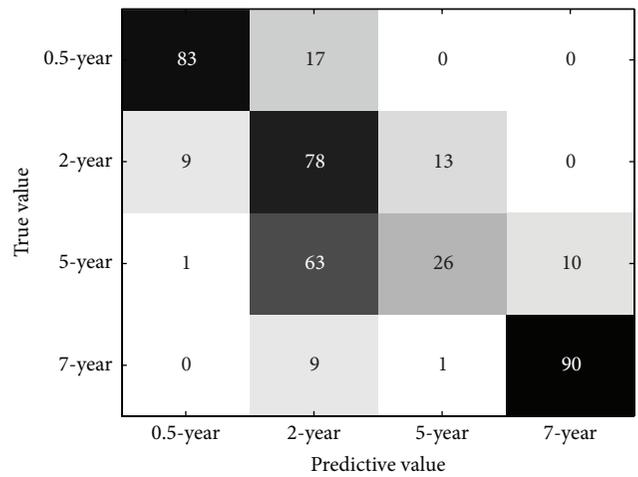


FIGURE 15: RBF neural network confusion matrix for training samples.

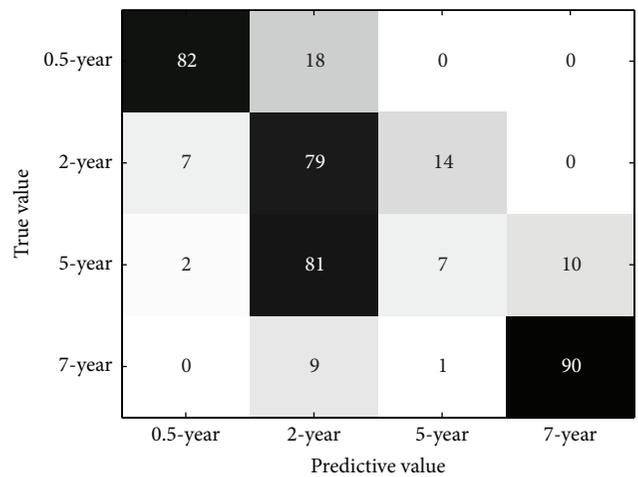


FIGURE 16: RBF neural network confusion matrix for testing samples.

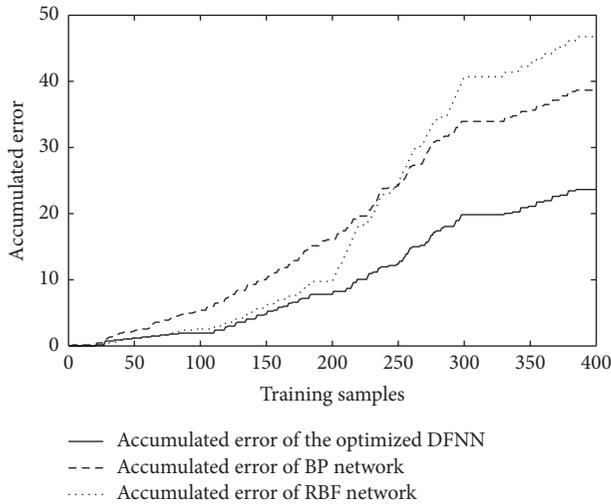


FIGURE 17: Training accumulated errors of three networks.

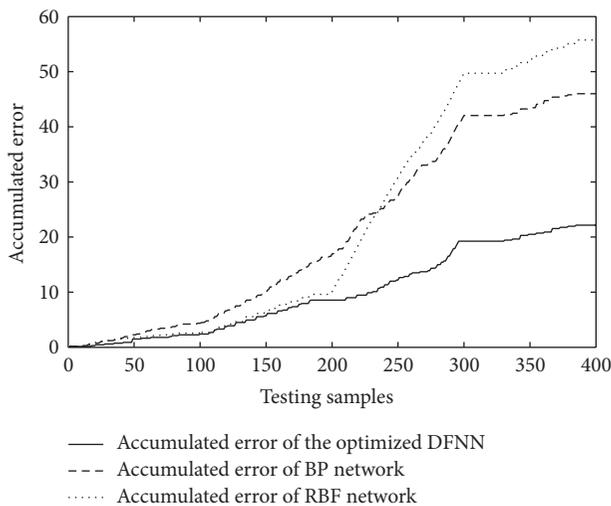


FIGURE 18: Testing accumulated errors of three networks.

method has the best performance compared to other three models.

- (1) Screw online monitoring system is applied to CINCINNATIV5-3000 machining center. Considering that the machining center always works on changeable working conditions, the real-time working conditions are also stored as an important part of the feature vectors library.
- (2) To solve the prediction stability problem of using DFNN model, an optimization algorithm based on QGA and imitating prediction samples is presented in this paper. Compared to the unoptimized DFNN, the optimized DFNN maintains a good prediction accuracy.
- (3) Screw feature vectors of different service life are applied to test the model performance in the experiment; from the discussion among three network

models, the conclusion can be reached that the optimized DFNN presents better prediction accuracy than BP and RBF network and is suitable for screw performance assessment.

## Conflict of Interests

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

## Acknowledgment

The authors would like to extend their thanks to the joint financial support by National Natural Science Fund of China (51275426).

## References

- [1] C.-H. Lee, M.-Y. Yang, C.-W. Oh, T.-W. Gim, and J.-Y. Ha, "An integrated prediction model including the cutting process for virtual product development of machine tools," *International Journal of Machine Tools and Manufacture*, vol. 90, pp. 29–43, 2015.
- [2] R. Pérez, A. Molina, and M. Ramírez-Cadena, "Development of an integrated approach to the design of reconfigurable micro/mesoscale cnc machine tools," *Journal of Manufacturing Science and Engineering*, vol. 136, no. 3, Article ID 031003, 10 pages, 2014.
- [3] A. Gok, C. Gologlu, and H. I. Demirci, "Cutting parameter and tool path style effects on cutting force and tool deflection in machining of convex and concave inclined surfaces," *International Journal of Advanced Manufacturing Technology*, vol. 69, no. 5–8, pp. 1063–1078, 2013.
- [4] A. Deshpande, "An empirical study to evaluate machine tool production readiness and performance," *International Journal of Advanced Manufacturing Technology*, vol. 64, no. 9–12, pp. 1285–1296, 2013.
- [5] P. C. Tsai, C. C. Cheng, and Y. C. Hwang, "Ball screw preload loss detection using ball pass frequency," *Mechanical Systems and Signal Processing*, vol. 48, no. 1-2, pp. 77–91, 2014.
- [6] Z. H. Li, K. G. Fan, J. G. Yang, and Y. Zhang, "Time-varying positioning error modeling and compensation for ball screw systems based on simulation and experimental analysis," *International Journal of Advanced Manufacturing Technology*, vol. 73, no. 5–8, pp. 773–782, 2014.
- [7] Y. Pan, M. J. Er, X. Li, H. Yu, and R. Gouriveau, "Machine health condition prediction via online dynamic fuzzy neural networks," *Engineering Applications of Artificial Intelligence*, vol. 35, no. 1, pp. 105–113, 2014.
- [8] J. Mateo, A. M. Torres, and M. A. García, "Dynamic fuzzy neural network based learning algorithms for ocular artefact reduction in EEG recordings," *Neural Processing Letters*, vol. 39, no. 1, pp. 45–67, 2014.
- [9] C.-F. Juang, "A TSK-type recurrent fuzzy network for dynamic systems processing by neural network and genetic algorithms," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 155–170, 2002.
- [10] C.-H. Lee and C.-C. Teng, "Identification and control of dynamic systems using recurrent fuzzy neural networks," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 4, pp. 349–366, 2000.

- [11] R.-J. Wai and P.-C. Chen, "Robust neural-fuzzy-network control for robot manipulator including actuator dynamics," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 4, pp. 1328–1349, 2006.
- [12] A. Subasi, "Automatic detection of epileptic seizure using dynamic fuzzy neural networks," *Expert Systems with Applications*, vol. 31, no. 2, pp. 320–328, 2006.
- [13] S. Q. Wu, M. J. Er, and Y. Gao, "A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 578–594, 2001.
- [14] H. Adeli and X. M. Jiang, "Dynamic fuzzy wavelet neural network model for structural system identification," *Journal of Structural Engineering*, vol. 132, no. 1, pp. 102–111, 2006.
- [15] C.-F. Juang and C.-D. Hsieh, "A locally recurrent fuzzy neural network with support vector regression for dynamic-system modeling," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 2, pp. 261–273, 2010.
- [16] C. G. Huang, G. Li, Z. X. Xu, and L. P. Chang, "Design of optimal digital lattice filter structures based on genetic algorithm," *Signal Processing*, vol. 92, no. 4, pp. 989–998, 2012.
- [17] L. Liu, L.-H. Guo, H. Xiao, J.-J. Wang, and G.-G. Wang, "Software reliability growth model based on SAA-DFNN," *Journal of Jilin University*, vol. 42, no. 5, pp. 1225–1230, 2012.
- [18] X. C. Zhang, H. L. Gao, H. F. Huang, L. Guo, and S. D. Xiao, "Optimization design of mathematical morphology filter based on quantum genetic algorithm," *Journal of Southwest Jiaotong University*, vol. 49, no. 3, pp. 462–469, 2014.
- [19] K.-H. Han and J.-H. Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 6, pp. 580–593, 2002.
- [20] J. Zhang, H. Z. Li, Z. G. Tang, Q. P. Lu, X. O. Zheng, and J. L. Zhou, "An improved quantum-inspired genetic algorithm for image multilevel thresholding segmentation," *Mathematical Problems in Engineering*, vol. 2014, Article ID 295402, 12 pages, 2014.
- [21] S. E.-O. Bahlous, M. Neifar, S. El-Borgi, and H. Smaoui, "Ambient vibration based damage diagnosis using statistical modal filtering and genetic algorithm: a bridge case study," *Shock and Vibration*, vol. 20, no. 1, pp. 181–188, 2013.
- [22] A. Montazeri and J. Poshtan, "Optimizing a multi-channel ANC system for broadband noise cancellation in a telephone kiosk using genetic algorithms," *Shock and Vibration*, vol. 16, no. 3, pp. 241–260, 2009.
- [23] A. Lin and J. Phillips, "Optimization of random diffraction gratings in thin-film solar cells using genetic algorithms," *Solar Energy Materials and Solar Cells*, vol. 92, no. 12, pp. 1689–1696, 2008.
- [24] J.-C. Lee, W.-M. Lin, G.-C. Liao, and T.-P. Tsao, "Quantum genetic algorithm for dynamic economic dispatch with valve-point effects and including wind power system," *International Journal of Electrical Power and Energy Systems*, vol. 33, no. 2, pp. 189–197, 2011.
- [25] M. J. Er, F. Liu, and M. B. Li, "Channel equalization using dynamic fuzzy neural networks," *International Journal of Fuzzy Systems*, vol. 11, no. 1, pp. 10–19, 2009.
- [26] A. SaiToh, R. Rahimi, and M. Nakahara, "A quantum genetic algorithm with quantum crossover and mutation operations," *Quantum Information Processing*, vol. 13, no. 3, pp. 737–755, 2014.
- [27] J. W. Gu, M. Z. Gu, C. W. Cao, and X. S. Gu, "A novel competitive co-evolutionary quantum genetic algorithm for stochastic job shop scheduling problem," *Computers & Operations Research*, vol. 37, no. 5, pp. 927–937, 2010.

## Research Article

# Planetary Gearbox Vibration Signal Characteristics Analysis and Fault Diagnosis

Qiang Miao<sup>1,2</sup> and Qinghua Zhou<sup>2</sup>

<sup>1</sup>State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, China

<sup>2</sup>School of Aeronautics and Astronautics, Sichuan University, Chengdu, Sichuan 610065, China

Correspondence should be addressed to Qiang Miao; mqiang@scu.edu.cn

Received 26 June 2015; Accepted 20 August 2015

Academic Editor: Chuan Li

Copyright © 2015 Q. Miao and Q. Zhou. 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.

Planetary gearboxes are widely used in helicopters, wind turbines, mining machinery, and so forth. The structure and motion type of a planetary gearbox are more complex in comparison with a fixed-shaft one, which makes condition monitoring and fault diagnosis of planetary gearbox a challenging issue in practical applications. In order to understand the fundamental nature of planetary gearbox vibration, this paper conducts an investigation on vibration characteristics of a single-stage planetary gearbox. Assuming that the gearbox and the sensor revolve inversely at the speed of planet carrier, the problem can be transformed into two easier parts: research on fixed-shaft gearbox signal model and research on influence of sensor spinning. Based on this assumption, a vibration signal model of planetary gearbox is obtained. Experimental data are used to validate the model.

## 1. Introduction

The development of equipment manufacturing industry raises urgent demand on high performance transmissions. Compared with traditional fixed-shaft gearing systems, a planetary gearbox has compact structure and light weight, while its load-carrying capacity, transmission precision, and efficiency are much higher. Therefore, planetary gearboxes have been widely used in helicopters, wind turbines, mining machinery, and so forth. However, the reliability of these high performance transmissions is a critical issue in practical applications, because the occurrence of gearbox failure may lead to catastrophic accidents and cause severe economic losses. Therefore, it is necessary to conduct planetary gearbox condition monitoring and fault diagnosis so as to reduce equipment operational cost and risk.

In general, a planetary gear train in a gearbox consists of a sun gear, a ring gear, a planet carrier, and several planet gears. Compared with a fixed-shaft gearing system, a planetary gear train has a composite rotating motion style. Planet gears revolve on their own shafts and along central sun gear. The composite motion caused by multi-gear-meshing results in time-variant vibration propagation paths, and this generates

distinctive nonstationarity in planetary gearbox dynamic response. The composite motion induced nonstationarity is coupled with nonstationarity caused by gear train failure, which makes it a challenging problem in planetary gearbox condition monitoring.

Gearbox dynamic response analysis is an essential step to reveal fault root causes and their response features. For example, Ericson and Parker [1] investigated the impact of torque changes on planetary gear natural frequencies, mode shapes, and damping parameters. Bartelmus et al. [2] conducted fault detection and diagnosis on fixed-shaft and planetary gearboxes under time-varying nonstationary operations and established gearbox dynamics models to understand the fault related phenomena through vibration signal analysis. Gu and Vexlex [3] proposed a lumped parameter model to analyze the influence of planet position errors in planet gears. It should be noted that a gearbox dynamics model should be built on the basis of certain simplifications, weakening its feasibility under very complicated working environments.

In planetary gearbox condition monitoring and diagnosis, an early research is to separate meshing vibration signals of planet gear and sun gear through the time-domain averaging (TSA) technique [4]. Based on this idea, Forrester

[5] proposed a signal filtering technique to realize the time-domain averaging for each individual planet gear. However, the TSA technique is established on the basis of constant rotating speed, while speed variation in practical applications limits signal separation accuracy. Recently, a lot of research has been conducted on planetary gearbox fault diagnosis. Orchard and Vachtsevanos [6] proposed a particle-filtering based framework for helicopter planetary gear train diagnosis and prognosis. Barszcz and Randall [7] presented a tooth crack detection method in wind turbine gearboxes using the spectral kurtosis. Lei et al. [8] proposed two diagnostic parameters based on the analysis of vibration features of planetary gearboxes. Feng and Zuo [9] investigated vibration signal models for planetary gearbox fault diagnosis. Reference [10] conducts a comprehensive survey on the state-of-the-art techniques in planetary gearbox fault diagnosis.

Engineering practices showed that vibration signal generated by different fault types and meshing components may demonstrate their respective waveform characteristics, and dynamic signal collected from equipment is an overall description of these vibration sources [11–14]. Due to the fact that a planetary gear train has complicated motion style, vibration induced by a planetary gear train differs a lot from a fixed-shaft gearing system. In order to understand the fundamental nature of planetary gear train vibration and develop reliable fault diagnosis methods, it is necessary to analyze vibration characteristics of planetary gearbox under different conditions.

In general, most of work on planetary gearbox fault diagnosis focuses on the fault feature extraction and fault recognition. The purpose of this paper is to investigate vibration characteristics of a planetary gear train, and a planetary gearing signal model was obtained by assuming that the planetary gearbox and the sensor revolve inversely at the speed of carrier. The results obtained in this research may provide a new way to understand the fault behavior and develop new fault diagnosis methods. Experimental data collected from a planetary gearbox test rig were used to validate the developed signal model in gearbox fault identification.

The rest of the paper is organized as follows. Section 2 briefly describes the assumptions that are made in this research. Section 3 gives theoretical calculation formula for planetary gear train meshing frequency and characteristic frequencies. Section 4 shows the development of vibration signal model based on the assumptions defined in Section 2. In Section 5, the planetary gearbox test rig is briefly introduced, and the vibration data collected from this test rig are used for model validation. Finally, conclusions are summarized in Section 6.

## 2. Assumptions in Planetary Gearbox Vibration Characteristics Analysis

In the process of gearbox running, the number of meshing teeth varies alternatively along with gear train rotation, which generates gear meshing vibration induced by periodic changes of tooth meshing stiffness. If a tooth fault exists in a planetary gearbox, it introduces periodic impulses or serious modulating phenomena in vibration signal, and the

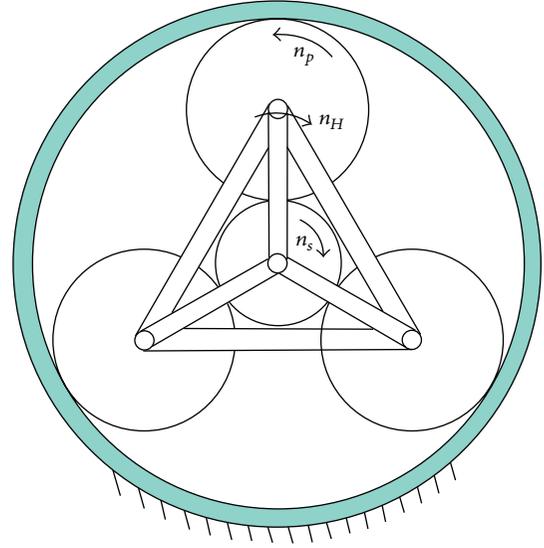


FIGURE 1: A schematic graph of planetary gear train.

corresponding characteristic frequency is associated with the rotating frequency of faulty gear relative to planet carrier [9].

Given a planetary gear train, usually the ring gear is fixed, while the carrier and the planet gears are bonded together, as shown in Figure 1. In this paper,  $r$ ,  $p$ ,  $s$ , and  $H$  represent ring gear, planet gear, sun gear, and planet carrier, respectively. In addition,  $N$  is the rotation speed,  $z$  is the number of gear teeth, and  $K$  is the number of planet gears.

In the computation of planet gear train transmission ratio, it is usually assumed that the whole gearset reversely rotates at the carrier speed, and the calculation of a planet gear train transmission ratio is transformed to a fixed-shaft gearing one. Similarly, in the analysis of planetary gearing vibration characteristics, it is assumed that the gearbox and the sensor rotate reversely with the speed of planet carrier. Then, the analysis of planetary gearbox vibration can be divided into two simple parts: the fixed-shaft gearset vibration analysis and the impact of sensor's spinning motion on fixed-shaft gearset vibration signal. Unless otherwise specified, the term "relative" in this paper means the relative motion to planet carrier.

## 3. Planetary Gear Train Meshing Frequency and Characteristic Frequencies

**3.1. Gear Meshing Frequency.** When a planet gear completes one relative revolution, it has the same number of meshing teeth with the sun gear and the ring gear. Therefore, the meshing frequency of sun gear to planet gear is equal to the meshing frequency of planet gear to ring gear. Given the planetary gear train shown in Figure 1, if the rotation frequency of sun gear's shaft is known as  $f_s$ , then the meshing frequency  $f_m$  can be calculated as

$$f_m = f_H \cdot z_r = \frac{z_r z_s}{z_r + z_s} \cdot f_s. \quad (1)$$

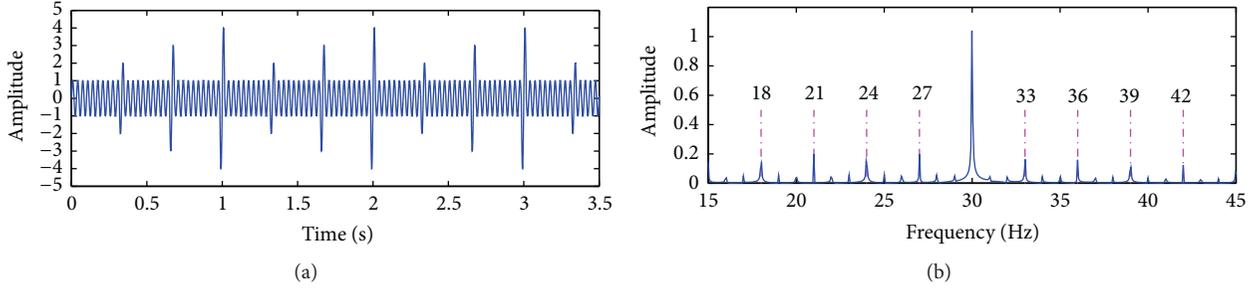


FIGURE 2: A simulation signal. (a) Time-domain waveform; (b) frequency spectrum.

### 3.2. Characteristic Frequency of Sun Gear

(1) *Characteristic Frequency of Sun Gear with Distributed Fault.* Given a distributed fault on a sun gear, the change of meshing condition is in one period when the sun gear completes one relative revolution. Then, the corresponding characteristic frequency of sun gear with distributed fault ( $f_{csd}$ ) is equal to the relative rotating frequency  $f_s^H$ :

$$f_{csd} = f_s^H = \frac{f_m}{z_s} = \frac{z_r}{z_r + z_s} \cdot f_s. \quad (2)$$

(2) *Characteristic Frequency of Sun Gear with Local Fault.* Assume that a local fault exists on a sun gear. When the local fault meshes with a planet gear, the vibration signal is modulated by an impulse signal. When the local fault exits the meshing area, the vibration signal returns to its normal condition. In one revolution of the sun gear, such modulating phenomenon happens  $K$  times. Assume that these modulation conditions are identical; the characteristic frequency of sun gear with local fault ( $f_{csl}$ ) is  $K$  times of the relative rotating frequency  $f_s^H$ :

$$f_{csl} = K \cdot f_s^H = K \cdot \frac{f_m}{z_s} = K \cdot \frac{z_r}{z_r + z_s} \cdot f_s. \quad (3)$$

In fact, due to the existence of manufacturing and assembly errors in planet gear, planet carrier, and ring gear, the modulating phenomena in vibration signal caused by sun gear fault area meshing may not be identical. This paper utilizes a simulation signal to further discuss this case.

Take the planetary gear train shown in Figure 1 as an example. Assume that the amplitude of meshing vibration signal produced by sun gear and planet gear is 1, the meshing frequency is 30 Hz, and the relative rotating frequency of sun gear is 1 Hz. When the sun gear has a local fault, the fault area meshes with 3 planet gears, and the amplitudes of resulting amplitude-modulation (AM) phenomenon produced by these meshing processes are 2, 3, and 4, respectively. For simplification purpose, the frequency modulation (FM) phenomenon is not considered, and it is assumed that the waveform of modulating signal is a square wave. Figure 2 shows the time-domain waveform and its frequency spectrum. It can be observed that the values of spectral lines at  $f_m \pm 3f_s^H \cdot i$  ( $i \in \mathbb{Z}$ ) are obviously large, while the other spectral lines are relatively small. Therefore, even if the manufacturing and assembly errors from planet gear, planet carrier,

and ring gear are considered, the characteristic frequency of sun gear with local fault can still be treated as  $K \cdot f_s^H$ , which is consistent with (3).

### 3.3. Characteristic Frequency of Planet Gear

(1) *Characteristic Frequency of Planet Gear with Distributed Fault.* Given a distributed fault on a planet gear, the change of meshing condition is in one period when the planet gear completes one relative revolution. Thus, the corresponding characteristic frequency of planet gear with distributed fault ( $f_{cpd}$ ) is equal to the relative rotating frequency of planet gear  $f_p^H$ :

$$f_{cpd} = f_p^H = \frac{f_m}{z_p} = \frac{z_r \cdot z_s}{z_p(z_r + z_s)} \cdot f_s. \quad (4)$$

(2) *Characteristic Frequency of Planet Gear with Local Fault.* Assume that a local fault exists on a single side of a planet gear tooth. When the local fault meshes with the sun gear or the ring gear, the vibration signal is modulated by an impulse signal. When the local fault exits the meshing area, the vibration signal returns to its normal condition. In one revolution of planet gear, such modulating situation happens once. Thus, the characteristic frequency of planet gear with single-sided local fault ( $f_{cpl}$ ) is equal to the relative rotating frequency  $f_p^H$ :

$$f_{cpl} = f_p^H = \frac{f_m}{z_p} = \frac{z_r \cdot z_s}{z_p(z_r + z_s)} \cdot f_s. \quad (5)$$

If a local fault happens on both sides of a planet gear tooth, the vibration signal modulation happens twice during its one period revolution. Although the amplitudes of these two modulations may be different, similar results can be obtained from the discussion of sun gear local fault characteristic frequency. Therefore, the characteristic frequency of planet gear with two-sided local fault is twice the relative rotating frequency  $f_p^H$ :

$$f_{cpl} = 2f_p^H = 2 \cdot \frac{f_m}{z_p} = 2 \cdot \frac{z_r \cdot z_s}{z_p(z_r + z_s)} \cdot f_s. \quad (6)$$

3.4. *Characteristic Frequency of Ring Gear.* Similar to the results from sun gear characteristic frequencies, the characteristic frequency of ring gear with distributed fault ( $f_{\text{crd}}$ ) and the characteristic frequency of ring gear with local fault ( $f_{\text{crl}}$ ) can be obtained, respectively:

$$f_{\text{crd}} = \frac{z_s}{z_r + z_s} \cdot f_s = f_H, \quad (7)$$

$$f_{\text{crl}} = K \cdot \frac{z_s}{z_r + z_s} \cdot f_s = K \cdot f_H. \quad (8)$$

## 4. Planetary Gear Train Vibration Signal Model

4.1. *Vibration Signal Model after Transforming to Fixed-Shaft Gearing System.* For a pair of gears under working condition, the number of meshing teeth changes alternatively during the meshing process. This results in the periodic change of tooth meshing stiffness and causes gear meshing vibration. Due to the unavoidable errors in manufacturing and assembly process, as well as faults on gears, the gear meshing condition changes frequently. The change is closely associated with the gear relative rotating period, and it generates amplitude modulating and frequency modulating phenomena in the original signal. Thus, the vibration signal caused by faulty gear meshing can be described by amplitude and frequency modulations:

$$x_o(t) = \sum_{k=0}^{\infty} a_k(t) \cos [2\pi k f_m t + b_k(t) + \theta_k]. \quad (9)$$

Here,  $a_k(t)$  and  $b_k(t)$  are the amplitude modulation and the frequency modulation functions, respectively:

$$a_k(t) = A_0 + \sum_{n=1}^{\infty} A_{kn} \cos (2\pi n f_c t + \varphi_{kn}), \quad (10)$$

$$b_k(t) = \sum_{l=1}^{\infty} B_{kl} \sin (2\pi l f_c t + \phi_{kl}). \quad (11)$$

Here,  $f_c$  is the characteristic frequency.  $A_0$ ,  $A_{kn}$ , and  $B_{kl}$  are constants, representing the amplitudes of vibration signal, AM, and FM, respectively.

For the vibration signal collected under gear local fault, its modulating signal is a periodic impulse. Since the periodic function can be approximately expressed by an infinite series of sine functions and cosine functions, the amplitude modulating function and the frequency modulating function caused by local fault can be expressed by (10) and (11).

For simplification purpose, we only analyze the fundamental frequencies in (9)~(11). That is, only the case of  $k = n = l = 1$  is considered. Then, (9) can be simplified as

$$x_o(t) = A_1 [1 + A \cos (2\pi f_c t + \varphi_1)] \cdot \cos [2\pi f_m t + B \sin (2\pi f_c t + \phi_1) + \theta_1]. \quad (12)$$

In addition, we have

$$e^{jz \sin \varphi} = \sum_{k=-\infty}^{+\infty} J_k(z) e^{jk\varphi}, \quad (13)$$

where  $J_k(z)$  is the  $k$ th-order first kind Bessel function of the variable  $z$ .

According to (13), (12) can be expanded as

$$x_o(t) = A_1 [1 + A \cos (2\pi f_c t + \varphi_1)] \cdot \sum_{k=-\infty}^{+\infty} J_k(B) \cos [2\pi (f_m + k f_c) t + k\phi_1 + \theta_1]. \quad (14)$$

Following the properties of trigonometric functions, we have

$$\cos \alpha \cdot \cos \beta = \frac{1}{2} [\cos (\alpha + \beta) + \cos (\alpha - \beta)]. \quad (15)$$

Then, (14) can further be expanded as

$$x_o(t) = \sum_{k=-\infty}^{+\infty} A_1 \cdot J_k(B) \left\{ \cos [2\pi (f_m + k f_c) t + \varphi_k] + \frac{A}{2} \cos [2\pi (f_m + k f_c + f_c) t + \phi_k] + \frac{A}{2} \cos [2\pi (f_m + k f_c - f_c) t + \theta_k] \right\}, \quad (16)$$

where  $f_c$  is the characteristic frequency,  $\varphi_k = k\phi_1 + \theta_1$ ,  $\phi_k = k\phi_1 + \varphi_1 + \theta_1$ , and  $\theta_k = -k\phi_1 + \varphi_1 - \theta_1$ . Since  $A_1$ ,  $J_k(B)$ , and  $A$  are constants, (16) can be further simplified by properties of trigonometric functions, and it can be obtained that

$$x_o(t) = \sum_{k=-\infty}^{+\infty} C_k \cos [2\pi (f_m + k f_c) t + \alpha_k]. \quad (17)$$

Here,  $C_k$  and  $\alpha_k$  are constants that are related to  $A_1$ ,  $A$ ,  $\{J_i(B)\}$ ,  $\{\varphi_i\}$ ,  $\{\phi_i\}$ , and  $\{\theta_i\}$ .

According to (17), a series of sidebands emerge around the meshing frequency ( $f_m$ ) of a faulty gear, and the interval between the neighbouring sidebands is equal to the characteristic frequency of faulty gear ( $f_c$ ). If we take into account the harmonics  $n f_m$  ( $n \in N^*$ ) of the meshing frequency, the sidebands emerge around the positions  $n f_m \pm k f_c$  ( $k \in N^*$ ).

4.2. *Influence of Sensor Spinning to Vibration Signal.* Given different vibration propagation paths, the impact of sensor spinning may not be the same. Assume that a fault exists on a sun gear. As shown in Figure 3, the vibration caused by the meshing process between sun gear and planet gear may be transmitted to the sensor through 3 paths.

In path 1, the vibration starts from the meshing point, passes through the planet gear, the ring gear, and the case of gearbox, and finally reaches the signal measurement point. In path 2, the vibration starts from the meshing point, passes through the planet gear with its bearing, the planet carrier with its shaft, the bearing, and the case of gearbox, and finally reaches the signal measurement point. In path 3, the vibration starts from the meshing point, passes through the sun gear, the shaft of sun gear with its bearing, and the case of gearbox,

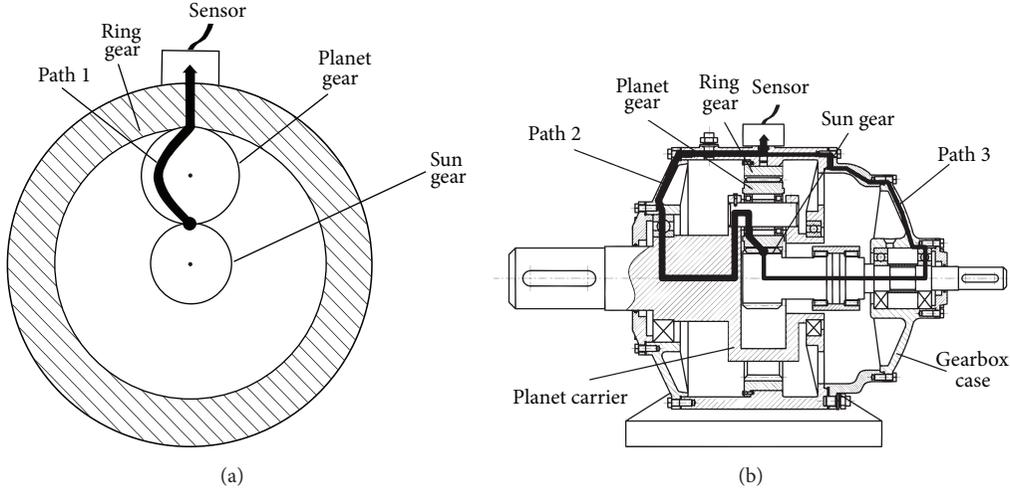


FIGURE 3: Propagation paths of signal. (a) Path 1; (b) path 2 and path 3.

and finally reaches the signal measurement point. There are more components in path 2 and path 3, which causes serious signal attenuation in comparison with path 1. Therefore, this paper only considers the signal transmitted through path 1.

Assume that the planet gear train is simplified as a fixed-shaft gear train, and a fault exists on a certain gear tooth. Due to the movement of the sensor, the relative distance between the fault area and the sensor continuously changes, given that at time  $t$  the distance between the sensor and the meshing point is the closest. With the rotating movement of the sensor around the center of planetary gear train, the sensor gradually leaves from the meshing point, and the amplitude of the vibration collected by the sensor becomes smaller. When the sensor moves to the farthest distance, the strength of vibration is the weakest one. Further movement of the sensor leads to smaller distance, and the vibration collected becomes stronger. In fact, during the spinning process of the sensor, the frequency of collected signal may have a small amount of variation. Once the sensor revolves one period around the gear train and returns to the position that has the closest distance with the meshing point, both the amplitude and the frequency of the collected signal have one period of change. Thus, the impact of sensor revolving around center of gear train can be described as amplitude modulating and frequency modulating processes.

**4.3. Vibration Signal Model of Planetary Gear Train.** On one hand, (17) describes the vibration signal model of the fixed-shaft gear train that is transformed from the planetary gear train. On the other hand, the impact of sensor spinning is to cause amplitude modulating and frequency modulating phenomena in vibration signal. For simplification purpose, only the modulating phenomena of the fundamental frequency are considered, and the vibration signal  $x(t)$  collected by sensor can be described as

$$x(t) = A_{H0} [1 + A_{H1} \cos(2\pi f_H t + \theta_H)] x_o(t). \quad (18)$$



FIGURE 4: Test rig of a single-stage planetary gearbox.

Here,  $x_o(t)$  is the original vibration signal, and its expression is shown by (17);  $f_H$  is the relative rotating frequency of sensor, and its value is equal to the rotating frequency of planet carrier;  $A_{H0}$  is the signal attenuation coefficient;  $A_{H1}$  is the amplitude modulating strength.

Substitute (17) into (18), and the following result can be obtained:

$$x(t) = \sum_{k=-\infty}^{+\infty} D_k \cos[2\pi(f_m + kf_c \pm f_H)t + \beta_k], \quad (19)$$

where  $D_k$  and  $\beta_k$  ( $k \in Z$ ) are two constants.

Accordingly, the impact of sensor spinning on the collected vibration signal can be described as the emergence of spectral lines at the positions of  $\pm f_H$  around sidebands  $f_m \pm kf_c$  ( $k \in Z$ ).

If we consider the harmonics of the meshing frequency  $f_m$ , the modulating effect, and the harmonics of  $f_H$ , the positions of sidebands are  $nf_m + kf_c + lf_H$  ( $n \in N^*$ ,  $l \in Z$ ). Here,  $f_c$  is the characteristic frequency of the corresponding gear fault.

It should be noted that after a period of gearbox running gears suffer certain degree of wear. In addition, errors may also be introduced during manufacturing and assembly processes. All these may cause some variations in the theoretical signal model.

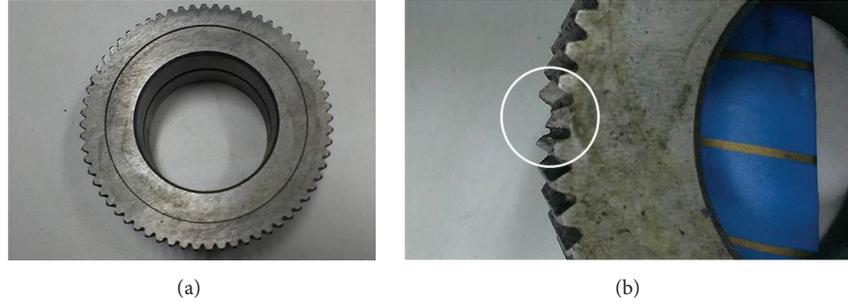


FIGURE 5: The normal planet gear and the artificially seeded faulty gear. (a) Normal planet gear; (b) artificially seeded faulty gear.

TABLE 1: Parameters of the planetary gearbox.

The number of teeth in the sun gear	The number of teeth in the planet gear	The number of teeth in the ring gear	The number of planet gears
13	64	146	3

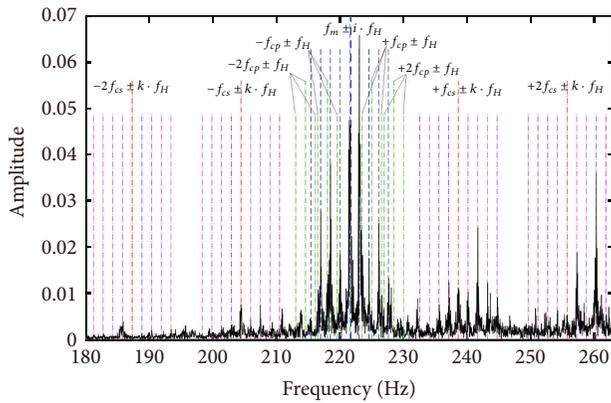


FIGURE 6: Frequency spectrum of vibration signal collected from the planetary gearbox.

## 5. Case Study

**5.1. Experimental Setup.** In this paper, an experiment is conducted on a single-stage planetary gearbox. The model type of gearbox is NGW11, and the gearbox transmission ratio is 12.23. Other parameters are provided in Table 1. The test rig includes a single-stage planetary gearbox, a drive motor, a magnetic powder brake, and a control and data acquisition system, which is shown in Figure 4. In the experiment, the IMI 603C1 sensor is mounted on the top of gearbox. The NI PCI-4472B data acquisition card is used for data collection, where the sampling frequency is 80 kHz, and the sampling time is 30 seconds. Figure 5 shows the normal planet gear and the faulty one, where the fault is artificially seeded.

**5.2. Data Analysis.** After a certain time of gearbox running, a piece of vibration signal was collected, and Figure 6 shows the spectrum of the signal. From Figure 6, the meshing frequency of 221.694 Hz can be observed, which is represented by the highest blue dashed line in the graph. The remaining blue

TABLE 2: Time-domain statistic metrics of the normal and faulty signals.

Metric	Normal signal	Faulty signal
Root Mean Square (RMS)	0.5729	0.8820
Peak Value	3.1431	14.3611
Crest Factor	5.4861	16.2823
Kurtosis	3.4298	17.6152

dashed lines correspond to the combination of modulating phenomena caused by the planet carrier rotation and the ring gear error. The green dash-dot lines represent the planet gear characteristic frequency and the sidebands of modulation phenomena caused by the planet carrier rotation. The red and purple dash-dot lines correspond to the sun gear characteristic frequency and the sidebands of modulation phenomena caused by the planet carrier rotation. Therefore, it can be drawn from Figure 6 that the actual sideband distribution in the signal spectrum of planetary gearbox is consistent with the vibration signal model presented in Section 4.3.

In order to validate the proposed vibration model for fault diagnosis, two pieces of vibration signals collected from the normal gearbox and the one with planet gear failure were used for comparison. Table 2 shows some time-domain statistic metrics calculated from these signals. Figure 7 shows the time-domain waveforms of the gearbox under normal condition and planet gear failure. Figure 8 shows the spectra of two signals around the first-order meshing frequency.

From Table 2, it can be seen that all the statistic metrics of the faulty vibration signal are higher than the ones of the normal signal. The increase of RMS indicates that the existence of fault causes stronger vibration. The Peak Value, Crest Factor, and Kurtosis of fault signal are several times larger than the ones calculated from the normal signal, which means there exist obvious impulses in the time-domain signal. In fact, a comparison of Figures 7(a) and 7(b) also proves the existence of impulses in gear failure condition.

In Figure 8, the dashed lines represent the meshing frequency (the position of the highest dashed line) and the sidebands of the modulating phenomena caused by planet carrier rotation and ring gear error (the positions of remaining dashed lines). The dash-dot lines correspond to the characteristic frequency of the planet gear (the position with the highest dash-dot line) and the sidebands of the

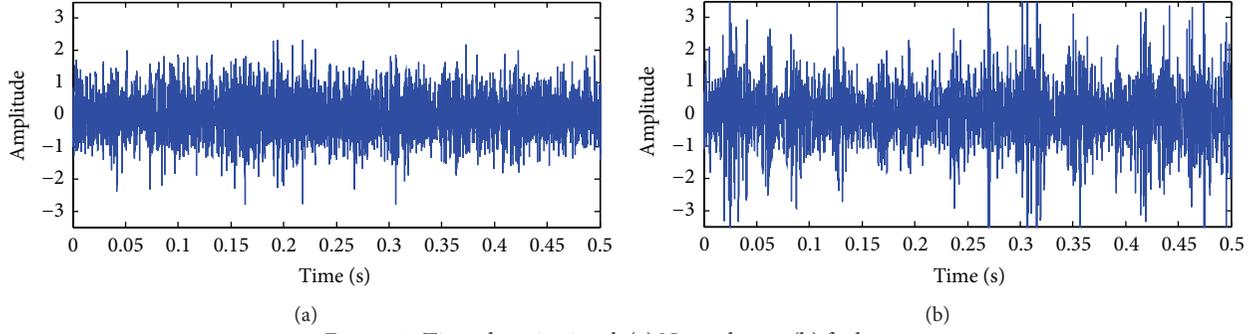


FIGURE 7: Time-domain signal. (a) Normal gear; (b) faulty gear.

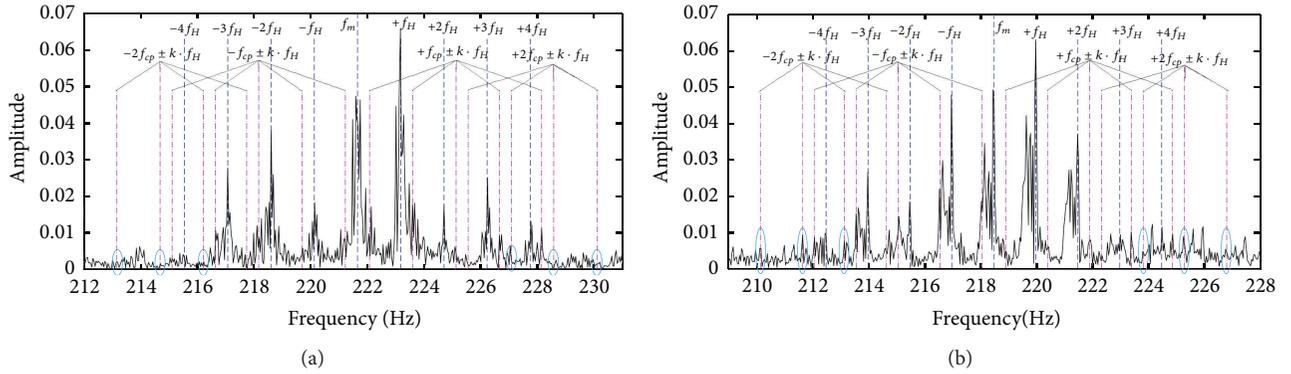


FIGURE 8: Frequency spectrum. (a) Normal gear; (b) faulty gear.

modulating phenomena caused by planet carrier rotation (the positions of remaining dash-dot lines). Through comparison between Figures 8(a) and 8(b), it can be observed that the amplitudes of the planet gear characteristic frequencies and their corresponding sidebands increase dramatically after planet gear failure. Specifically, obvious spectral lines can be found at the second harmonic of the characteristic frequency of planet gear with distributed fault (i.e., the characteristic frequency of planet gear with two-sided gear tooth fault).

Based on the aforementioned analysis results obtained from the time-domain statistic metrics, the time-domain waveform, and the emergence of sidebands, it can be concluded that there exists planet gear failure in the gearbox.

## 6. Conclusions

Planetary gearboxes have been widely used in transmission systems of helicopters, wind turbines, and engineering machinery. However, due to its complicated structure, traditional condition monitoring methods for fixed-shaft gearing system are not suitable for planetary gearbox. It is necessary to develop new methods for planetary gearbox condition and fault diagnosis.

The purpose of this research is to investigate vibration signal model of planetary gear train, which helps us to understand fault phenomena of planetary gearbox and develop reliable fault diagnosis methods. Assume that the whole gearbox reversely rotates around center of sun gear at a speed of planet gear rotation. Then, planetary gearbox fault diagnosis can be divided into two tasks: the fixed-shaft

gearbox fault diagnosis and the impact of sensor spinning on vibration signal. Based on the assumptions, the vibration signal model of faulty planetary gearbox is established, and the proposed model provides us with a way to understand different fault phenomena of planetary gearbox. A test rig of planetary gearbox was built, and the vibration data collected from the test rig validated the proposed model. Further, the proposed vibration signal model successfully identified planet gear fault in the gearbox, which may be useful in developing planetary gearbox fault diagnosis methods using the information extracted from vibration signal spectrum.

## Conflict of Interests

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

## Acknowledgments

This research was partially supported by the National Natural Science Foundation of China (Grant nos. 51275554 and 51405316) and the 2012 Open Project of State Key Laboratory of Mechanical Transmission, Chongqing University (Grant no. SKLMT-KFKT-201205). The authors also would like to thank Mr. Lu Yang for his help in the test rig setup in this research.

## References

- [1] T. M. Ericson and R. G. Parker, "Experimental measurement of the effects of torque on the dynamic behavior and system

- parameters of planetary gears,” *Mechanism and Machine Theory*, vol. 74, pp. 370–389, 2014.
- [2] W. Bartelmus, F. Chaari, R. Zimroz, and M. Haddar, “Modelling of gearbox dynamics under time-varying nonstationary load for distributed fault detection and diagnosis,” *European Journal of Mechanics—A/Solids*, vol. 29, no. 4, pp. 637–646, 2010.
- [3] X. Gu and P. Velex, “A dynamic model to study the influence of planet position errors in planetary gears,” *Journal of Sound and Vibration*, vol. 331, no. 20, pp. 4554–4574, 2012.
- [4] P. D. McFadden, “A technique for calculating the time domain averages of the vibration of the individual planet gears and the sun gear in an epicyclic gearbox,” *Journal of Sound and Vibration*, vol. 144, no. 1, pp. 163–172, 1991.
- [5] B. D. Forrester, “Method for the separation of epicyclic planet gear vibration signatures,” United States Patent, US6298725 B1, 2001.
- [6] M. E. Orchard and G. J. Vachtsevanos, “A particle-filtering approach for on-line fault diagnosis and failure prognosis,” *Transactions of the Institute of Measurement and Control*, vol. 31, no. 3-4, pp. 221–246, 2009.
- [7] T. Barszcz and R. B. Randall, “Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine,” *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352–1365, 2009.
- [8] Y. Lei, D. Kong, J. Lin, and M. J. Zuo, “Fault detection of planetary gearboxes using new diagnostic parameters,” *Measurement Science and Technology*, vol. 23, no. 5, Article ID 055605, 2012.
- [9] Z. P. Feng and M. J. Zuo, “Vibration signal models for fault diagnosis of planetary gearboxes,” *Journal of Sound and Vibration*, vol. 331, pp. 4919–4939, 2012.
- [10] Y. Lei, J. Lin, M. J. Zuo, and Z. He, “Condition monitoring and fault diagnosis of planetary gearboxes: a review,” *Measurement: Journal of the International Measurement Confederation*, vol. 48, no. 1, pp. 292–305, 2014.
- [11] D. Wang, Q. Miao, and R. Kang, “Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition,” *Journal of Sound and Vibration*, vol. 324, no. 3–5, pp. 1141–1157, 2009.
- [12] C. Li, M. Liang, and T. Wang, “Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals,” *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 132–148, 2015.
- [13] D. Wang, P. W. Tse, W. Guo, and Q. Miao, “Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis,” *Measurement Science and Technology*, vol. 22, no. 2, Article ID 025102, 2011.
- [14] D. Wang, P. W. Tse, and K. L. Tsui, “An enhanced Kurtogram method for fault diagnosis of rolling element bearings,” *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 176–199, 2013.