

Mechanisms and Machine Science

Hao Zhang · Guojin Feng ·
Hongjun Wang · Fengshou Gu ·
Jyoti K. Sinha *Editors*

Proceedings of IncoME-VI and TEPEN 2021


Performance Engineering and
Maintenance Engineering

 Springer

Mechanisms and Machine Science

Volume 117

Series Editor

Marco Ceccarelli , Department of Industrial Engineering, University of Rome Tor Vergata, Roma, Italy

Advisory Editors


Sunil K. Agrawal, Department of Mechanical Engineering, Columbia University, New York, USA

Burkhard Corves, RWTH Aachen University, Aachen, Germany

Victor Glazunov, Mechanical Engineering Research Institute, Moscow, Russia

Alfonso Hernández, University of the Basque Country, Bilbao, Spain

Tian Huang, Tianjin University, Tianjin, China

Juan Carlos Jauregui Correa , Universidad Autonoma de Queretaro, Queretaro, Mexico

Yukio Takeda, Tokyo Institute of Technology, Tokyo, Japan

This book series establishes a well-defined forum for monographs, edited Books, and proceedings on mechanical engineering with particular emphasis on MMS (Mechanism and Machine Science). The final goal is the publication of research that shows the development of mechanical engineering and particularly MMS in all technical aspects, even in very recent assessments. Published works share an approach by which technical details and formulation are discussed, and discuss modern formalisms with the aim to circulate research and technical achievements for use in professional, research, academic, and teaching activities.

This technical approach is an essential characteristic of the series. By discussing technical details and formulations in terms of modern formalisms, the possibility is created not only to show technical developments but also to explain achievements for technical teaching and research activity today and for the future.

The book series is intended to collect technical views on developments of the broad field of MMS in a unique frame that can be seen in its totality as an Encyclopaedia of MMS but with the additional purpose of archiving and teaching MMS achievements. Therefore, the book series will be of use not only for researchers and teachers in Mechanical Engineering but also for professionals and students for their formation and future work.

The series is promoted under the auspices of International Federation for the Promotion of Mechanism and Machine Science (IFTOMM).

Prospective authors and editors can contact Mr. Pierpaolo Riva (publishing editor, Springer) at: pierpaolo.riva@springer.com

Indexed by SCOPUS and Google Scholar.

More information about this series at <https://link.springer.com/bookseries/8779>

Hao Zhang · Guojin Feng · Hongjun Wang ·
Fengshou Gu · Jyoti K. Sinha
Editors

Proceedings of IncoME-VI and TEPEN 2021

Performance Engineering and Maintenance
Engineering

Editors

Hao Zhang
School of Mechanical Engineering
Hebei University of Technology
Tianjin, China

Guojin Feng
School of Mechanical Engineering
Hebei University of Technology
Tianjin, China

Hongjun Wang
School of Mechanical and Electrical
Engineering
Beijing Information Science
and Technology University
Beijing, China

Fengshou Gu
University of Huddersfield
Huddersfield, UK

Jyoti K. Sinha
University of Manchester
Manchester, UK

ISSN 2211-0984

ISSN 2211-0992 (electronic)

Mechanisms and Machine Science

ISBN 978-3-030-99074-9

ISBN 978-3-030-99075-6 (eBook)

<https://doi.org/10.1007/978-3-030-99075-6>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Hebei University of Technology and WorldTech Intelligence, Tianjin, China, organised The Efficiency and Performance Engineering Network 2021 (TEPEN 2021) and Sixth International conference on Maintenance Engineering (IncoME-VI) on 20th–23rd October 2021.

The goal of TEPEN 2021 and IncoME-VI is to provide a common platform by which professionals, engineers, practitioners and researchers working in the field of condition monitoring, plant maintenance and reliability can share their experiences. The scope of the conference covered a broad area with multidisciplinary interests in the fields of plant maintenance, asset management, reliability, condition monitoring and related areas, ranging from fundamental research to real-world applications.

In this conference, participations and contributions were involved in both theoretical research and practical applications of all aspects of fault detection, diagnostics, prognostics in both the operational and manufacturing processes. In the course of this event, eight keynote speeches and parallel technical sessions were delivered in accordance with key following topics:

- Vibro-acoustics Monitoring
- Asset Management
- Condition-based Maintenance
- Condition Monitoring and Reengineering
- eMaintenance, Mobile Technology
- Health, Safety and Environment
- Sensors and Instrumentation
- Life Cycle Cost Optimisation
- Machine Health Monitoring
- Machine Lube Oil Analysis and Monitoring
- Artificial intelligence, Machine Learning
- Plant Outage
- Maintenance Auditing
- Prognostics and Health Management
- Maintenance Organisation

- Maintenance Performance Measurement
- Non-Destructive Testing
- Manufacturing Process Monitoring
- Reliability, Maintainability and Risk
- Signal and Image Processing Methods

Despite the challenging circumstances of year 2021, this book consists nonetheless of 87 peer-reviewed papers. The book offers the state of the art of advances in asset management and condition monitoring and also serves as an excellent reference work for academic and industrial scientists and graduate students, working in asset management, condition monitoring and related areas.

The editors would like to acknowledge and thank the following people for help in book initiation, preparation and completion:

- Ning Hu (China)
- Fulei Chu (China)
- Andrew D. Ball (UK)
- Rongfeng Deng (China)
- Yongjian Ji (China)
- Zuolu Wang (UK)
- Xiaoxia Liang (China)
- Haiyang Li (China)
- Yuandong Xu (UK)
- Xiaoli Tang (UK)
- Yinghui Liu (China)
- Peng Li (China)
- Shaoning Tian (China)
- Yang Chen (China)
- Xiuquan Sun (UK)
- Miaoshuo Li (UK)

Thank you.

Tianjin, China
 Tianjin, China
 Beijing, China
 Huddersfield, UK
 Manchester, UK

Dr. Hao Zhang
 Dr. Guojin Feng
 Prof. Hongjun Wang
 Prof. Fengshou Gu
 Prof. Jyoti K. Sinha

Contents

An Optical Gyroscope Based Technique for Calibrating Angular-Measuring Instrument	1
Chenpeng Cui, Yuanwei Jiu, Chun Wang, and Fengshou Gu	
Data Analysis for Predictive Maintenance Using Time Series and Deep Learning Models—A Case Study in a Pulp Paper Industry	11
Balduino Mateus, Mateus Mendes, José Torres Farinha, Alexandre Batista Martins, and António Marques Cardoso	
Reinforcement Learning Fault Diagnosis Method Based on Less Tag Data	27
Kuo Xin, Jianguo Wang, and Wenxing Zhang	
Optimization Design and Simulation Analysis of Miniature Boring Machine Based on ADAMS	41
Zhou Yue, Cao Yu, Lu Zhen-hua, Wei Qi-wen, Zhao Xue-mei, Wang Ye-zhen, Sun Jia-xing, Liu Ying, and Zhong Shan	
The Stability and Vibration Characteristic Optimization of the Pressure Shell of a Buoyancy Regulator of an Underwater Vehicle	49
Yonghui Cao, Chiye Yang, Jing Liu, Yu Xie, Shumin Ma, and Yong Cao	
Emulational and Experimental Research on a Sugarcane Field Excitation Device	57
Hanning Mo, Chen Qiu, Shangping Li, Guiqing He, Bang Zeng, and Daiyun Yang	
Experimental Research on Influence Factors of the Sugarcane Ratoon Cutting Quality Under Vibration Conditions	69
Chen Qiu, Hanning Mo, Shangping Li, Guiqing He, Bang Zeng, and Daiyun Yang	

Research on the Influence of Changes in Particle Sizes on Simulation Results in the Simulation Test Based on the Discrete Element Method	81
Xuhong Tan, Jingwei Gao, Cheng Hu, Xiaobo Song, Min Zhang, and Peng Zheng	
Wind Turbine Condition Monitoring Based on SCADA Data Co-integration Analysis	97
Chao Zhang, Guanghan Zhao, and Yue Wu	
Improvement and Application of YOLOv3 for Smartphone Glass Cover Defect Detection	105
Yuan Cheng, Jigang Wu, Jun Shaov, and Deqiang Yang	
Optimization and Design of Efficiency and Quality of a Company Based on Value Stream Analysis	119
Guo Jidong, Qiu Zixuan, Huang Zehao, Wu Jiaqi, Zheng Jianxin, Tan Runjia, Lai Lijuan, and Zhou Dawei	
Review of Using Operational Modal Analysis for Condition Monitoring	139
Fulong Liu, Wei Chen, Yutao Men, Xiaotao Zhang, Yuchao Sun, Jun Li, and Guoan Yang	
Vibration Analysis of the Rudder Drive System of an Underwater Glider	147
Liming Guo, Jing Liu, Guang Pan, Baowei Song, Yonghui Cao, Yong Cao, Yujun Liu, and Hengtai Ni	
Online Method for Assessment and Tracking of Wear in Kaplan Turbine Runner Blades Operating Mechanism	155
Oscar García Peyrano, Daniel Vaccaro, Rodrigo Mayer, and Matías Marticorena	
Bearing Fault Diagnosis Based on Improved Residual Network	167
Haofei Du, Chao Zhang, and Jianjun Li	
Study on Optimization and Improvement of Production Line of H Product	185
Guo Jidong, Liang Yuyan, Ma Zenan, Qiu Zijian, Mo Yuwei, Li Limin, and Zhou Dawei	
Towards Data Driven Dynamical System Discovery for Condition Monitoring a Reciprocating Compressor Example	199
Ann Smith and W. T. Lee	
Rolling Bearing Remaining Useful Life Prediction Based on LSTM-Transformer Algorithm	207
Xinglu Tang, Hui Xi, Qianqian Chen, and Tian Ran Lin	

Research and Application of Order Analysis Technology Without Tachometer Under Variable Speed Condition 217
 Ruibo Yang and Jianguo Wang

Simulation Analysis of Tooth Surface Wear Considering Axis Parallelism Error 227
 Ruiliang Zhang and Yandong Shi

A TFG-CNN Fault Diagnosis Method for Rolling Bearing 237
 Hui Zhang, Shuying Li, and Yunpeng Cao

A Gas Turbine Gas Path Digital Twin Modeling Method 251
 Junqi Luan, Yun Peng Cao, Shuying Li, and Ran Ao

Small Sample MKFCNN-LSTM Transfer Learning Fault Diagnosis Method 265
 Yonglun Guo, Guoxin Wu, and Xiuli Liu

Prediction of Sensor Values in Paper Pulp Industry Using Neural Networks 281
 João Antunes Rodrigues, José Torres Farinha, António Marques Cardoso, Mateus Mendes, and Ricardo Mateus

YOLOV4-Based Wind Turbine Blade Crack Defect Detection 293
 Xin Yan, Guoxin Wu, and Yunbo Zuo

State-of-Art of Metal Debris Detection in Online Oil Monitoring 307
 Dingxin Yang and Xiaorong Liu

Regression Prediction of Performance Parameters in Ship Propulsion Equipment Simulation Model Based on One-Dimensional Convolutional Neural Network 315
 Liangyuan Huang and Guoji Shen

Research on Numerical Simulation Study of Solidification Heat Transfer Coefficient of Extra-Wide Slab 329
 Xinxia Qi, Qi Jia, and Wenxiong Wu

Numerical Study of Motor Electrical Signature for Condition Monitoring of Gear Tooth Breakage in a Motor-Gear System 341
 Funso Otuyemi, Xiuquan Sun, Fengshou Gu, and Andrew D. Ball

Analysing the Fault Behavior of a Complex Mechanical System for Diagnosis: A Bond Graph-Based Approach 359
 Jinxin Wang, Shenglei Zhao, Xiuzhen Ma, and Fengshou Gu

Dynamic Responses of Clearance Induced Impacts in Big End Bearing Condition Monitoring of Diesel Engines 373
 Solomon Okhionkpanwonyi, Fengshou Gu, and Andrew D. Ball

Impeller Wear Diagnosis in Centrifugal Pumps Under Different Flow Rate Based on Acoustic Signal Analysis	385
Alsadak Daraz, Fengshou Gu, and Andrew D. Ball	
Fault Diagnosis for Gas Turbine Rotor Using MOMEDA-VNCMD	403
Yingjie Cui, Hongjun Wang, and Xinghe Wang	
Classification and Recognition Method of Bearing Fault Based on SDP-CNN	417
Wang Xing-he, Wang Hong-jun, Cui Ying-jie, and Liu Ze-rui	
Spindle Health Assessment Based on Rotor Perception	427
Zhuangzhuang Zhang, Hongjun Wang, Jishou Xing, Fengshou Gu, and Xinghe Wang	
Fault Diagnosis of Rolling Bearing Based on Wavelet Packet Decomposition and SVM-LMNN Algorithm	439
Zhengbo Wang, Hongjun Wang, and Yingjie Cui	
Fine-Tuning and Efficient VGG16 Transfer Learning Fault Diagnosis Method for Rolling Bearing	453
Jinglei Su and Hongjun Wang	
An Investigation of Unsupervised Data-Driven Models for Internal Combustion Engine Condition Monitoring	463
Xiaoxia Liang, Chao Fu, Xiuquan Sun, Fang Duan, David Mba, Fengshou Gu, and Andrew D. Ball	
Online Pipe Leakage Detection Using the Vibration-Based Wireless Sensing System	477
Xiaoli Tang, Yu Jia, Guojin Feng, Yuandong Xu, Fengshou Gu, and Andrew D. Ball	
Meta-Learning Guided Few-Shot Learning Method for Gearbox Fault Diagnosis Under Limited Data Conditions	491
Ming Zhang, Duo Wang, and Yuchun Xu	
Investigation into LSTM Deep Learning for Induction Motor Fault Diagnosis	505
Xiaoyu Zhao, Ibrahim Alqatawneh, Mark Lane, Haiyang Li, Yongrui Qin, Fengshou Gu, and Andrew D. Ball	
Degradation Trend Construction of Aircraft Engine Using Complex Network Model	519
Yongsheng Huang, Yongbo Li, Khandaker Noman, and Shun Wang	
Research on the Influence of Mesh Stiffness of Fixed Gearbox with Chipping Fault	529
Jiacong Zhang, Yongbo Li, Shun Wang, and Khandaker Noman	

Damage Diagnosis of Railway Vehicle Car Body Based on Strain Modes 543
Hui Cao, Gangjun Li, and Fengshou Gu

A Chip Defect Detection System Based on Machine Vision 555
Xindan Qiao, Ting Chen, Wanjing Zhuang, and Jinyi Wu

A Bayesian Probabilistic Score Matrix Based Collaborative Filtering Recommendation System for Rolling Bearing Fault Identification 569
Yinghang He, Guangbin Wang, Fengshou Gu, and Andrew D. Ball

On-Line Monitoring of the Dimensional Error in Turning of a Slender Shaft 583
Pengyu Lu, Kaibo Lu, Yipei Liu, Bing Li, Xin Wang, Meixia Tian, and Fengshou Gu

Research on Feature Extraction and Recognition of Dongba Hieroglyphs 595
Hao Huang, Guoxin Wu, and Xiaoli Xu

Study on Feature Extraction of Gearbox Vibration Signal for Wind Turbines 607
Jinang Guo, Guoxin Wu, Xiwei Zhao, Hao Huang, and Xiaoli Xu

Condition Monitoring of a Reciprocating Air Compressor Using Vibro-Acoustic Measurements 615
Debanjan Mondal, Fengshou Gu, and Andrew D. Ball

Application of Combined Normalized Least Mean Square and Ensemble Empirical Mode Decomposition Denoising Method in Fault Diagnosis of Rolling Bearings 629
Changsheng Xi, Jie Yang, Dong Zhen, Xiaohao Liao, Wei Hu, and Fengshou Gu

Rolling Bearing Fault Diagnosis Based on Weighted Variational Mode Decomposition and Cyclic Spectrum Slice Energy 643
Dongkai Li, Xiaolang Liu, Yue You, Dong Zhen, Wei Hu, Kuihua Lu, and Fengshou Gu

Nonlinear Dynamic Analysis of a Planetary Gear System with Sun Gear Fault 655
Yinghui Liu, Zhanqun Shi, Dong Zhen, Xiaolang Liu, Wei Hu, and Fengshou Gu

A Fault Diagnosis Method for Rolling Bearings Based on Improved EEMD and Resonance Demodulation Analysis 669
Wei Zhang, Xiang Tian, Guohai Liu, and Hui Liu

A Study on the Contact Characteristics of a Planetary Centrifugal Vari-Speed Drive 683
 Jin Li, Jing Liu, Chaojie Zhong, Wujun Zou, and Ruikun Pang

The Fatigue Failure Prediction of a Vari-Speed Drive with Different Rollers 695
 Jing Liu, Ximing Zhang, Jingtao Shang, Jin Li, and Shizhao Ding

State of Health Estimation of Lithium-Ion Batteries from Charging Data: A Machine Learning Method 707
 Zuolu Wang, Guojin Feng, Dong Zhen, Fengshou Gu, and Andrew D. Ball

Harmonic Response Analysis of a Dual-Rotor System with Mass Unbalance 721
 Yubin Yue and Hongjun Wang

A Novel Method for Stacking Optimization of Aeroengine Multi-stage Rotors Based on 3D Deviation Prediction Model 733
 Jia Kang, Jun He, Zhisheng Peng, Haizhou Huang, and Shixi Yang

A Sensor Fault Identification Method Based on Adaptive Particle Swarm Optimization Support Vector Machine 749
 Xuezen Cheng, Dafei Wang, Chuannuo Xu, and Jiming Li

Correlation Analysis of Sensor Fault Based on Fuzzy Petri Net and Apriori Algorithm 761
 Chuannuo Xu, Shenglei Zhao, Haitao Hao, Yandong Zhang, Jiming Li, and Xuezen Cheng

Review on Simulation and Optimization of Vehicle Ride Comfort Based on Suspension Model 773
 Tang Jianghu, Xiong Qing, Zhu Yingmou, and He Zhuoyu

The Application of ADAMS Software to Vehicle Handling Stability: A Review 785
 Li Yixuan, Xiong Qing, Zhu Yingmou, and He Zhuoyu

Analysis and Decompose of Nine Degrees of Freedom Motion Simulator Relative Positional Precision 797
 Bo Li, Huadong He, Yinjun Lian, Xia Wu, Tongling Fu, Weiling Zhao, and Huibo Zhang

A Study on Vibration Response in the Baseplate of a Delta 3D Printer for Condition Monitoring 811
 Xinfeng Zou, Zhen Li, Fengshou Gu, and Andrew D. Ball

Analysis of Metamaterials-Based Acoustic Sensing Enhancement 823
 Shiqing Huang, Yubin Lin, Lichang Gu, Rongfeng Deng, Fengshou Gu, and Andrew D. Ball

A Novel Cylindrical Mechanical Metastructure for Drone Vibration Isolation 831
 Yubin Lin, Shiqing Huang, Lichang Gu, Rongfeng Deng, Solomon Okhionkpanwonyi, Qingbo He, Fengshou Gu, and Andrew D. Ball

Design and Simulation of Broadband Piezoelectric Energy Harvester with Multi-Cantilever 841
 Weiqiang Mo, Shiqing Huang, and Na Liu

A Mobile Pipeline Leak Monitoring Robot Based on Power Spectrum Correlation Analysis and Sound Pressure Location 853
 Weijie Tang, Rongfeng Deng, Baoshan Huang, Fengshou Gu, and Andrew D. Ball

A Review of Acoustic Emission Monitoring on Additive Manufacturing 867
 Zhen Li, Xinfeng Zou, Fanbiao Bao, Fengshou Gu, and Andrew D. Ball

Modelling and Vibration Signal Analysis for Condition Monitoring of Industrial Robots 879
 Huanqing Han, Dawei Shi, Lichang Gu, Nasha Wei, and Fengshou Gu

Study on Fault Mode of Hybrid Electric Vehicle 893
 Guibo Liao, Fanbiao Bao, and Baoshan Huang

A New Two-Dimensional Condition-Based Maintenance Model by Using Copulas 907
 Hanyang Wang, Ming Luo, and Fengshou Gu

Rotary Valve to Improve the Problem of Big End and Needle Glue Overflow in Contact Dispensing Process 921
 Gaolian Huang, Shifei Zhang, Zhiguo Liu, Gaobo Xiao, Chucheng Chen, and Fanbiao Bao

The Positioning Accuracy Measurement of the Dispenser and Compensation Method 931
 Xiong Huang, Gaobo Xiao, Zhiguo Liu, Shifei Zhang, Qiwen Wu, and Fanbiao Bao

X/Y/Z High-Speed and High-Precision Operation Platform Design 945
 Jun Guo, Gaobo Xiao, Zhen Xing, Zhiguo Liu, Shifei Zhang, and Yongqi Wu

Differentiable Architecture Searched Network with Tree-Structured Parzen Estimators for Rotating Machinery Fault Diagnosis 959
 Jingkang Liang, Yixiao Liao, and Weihua Li

A Review of Fault Diagnosis Methods for Marine Electric Propulsion System 971
Dongqin Li, Rongfeng Deng, Zhexiang Zou, Baoshan Huang, and Fengshou Gu

Research on the Influence of Crack Parameters on the Vibration Characteristics of Gas Turbine Compressor Blades 985
Weiwen Yu, Shixi Yang, Hongwei Chi, Zhisheng Peng, and Jun He

Research on Vibration Characteristics of Last Stage Blade Based on Blade Tip-Timing Technology 999
Xinyu Hu, Daming Zhuang, Jun He, Haizhou Huang, and Shixi Yang

A Simulation Study of an Energy Harvester Operating on a Vertical Rotor System 1013
LiChang Gu, Yubin Lin, Rongfeng Deng, Dawei Shi, Wang Wei, Zhixia Wang, Qishan Chen, Fengshou Gu, and Andrew D. Ball

Image-Based 3D Shape Estimation of Wind Turbine from Multiple Views 1031
Minghao Huang, Mingrui Zhao, Yan Bai, Renjie Gao, Rongfeng Deng, and Hui Zhang

Health Status Assessment of Marine Diesel Engine Based on Testability Model 1045
Ru Xiao, Guojun Qin, Zeyun Zhou, and Min Wang

Modelling the Dynamics of a CNC Spindle for Tool Condition Identification Based on On-Rotor Sensing 1057
Chun Li, Dawei Shi, Bing Li, Hongjun Wang, Guojin Feng, Fengshou Gu, and Andrew D. Ball

Real-Time Condition Monitoring and Health Assessment of Equipment Power Transmission Device Based on Wireless Sensor Network 1073
Cheng Zhe, Jiang Wei, Hu Niaoqing, Zhang Hao, and Zhen Dong

Research on Fault Detection Method for Special Equipment Under the Condition of Sample Missing 1081
Lei Wei, Zhe Cheng, Niaoqing Hu, Junsheng Cheng, and Guoji Shen

Research on Dynamic Impact Force Calculation for Spline Coupling Teeth and Its Suppression 1093
Wenchao Pan, Hai Lan, Zhiyong Han, Lantao Yang, Liming Wang, and Yimin Shao

An Optical Gyroscope Based Technique for Calibrating Angular-Measuring Instrument



Chenpeng Cui, Yuanwei Jiu, Chun Wang, and Fengshou Gu

Abstract This paper analyzes the error model of inductosyn install in an angular-measuring instrument, and then build up a test error system using the ultra-high precision Ring Laser Gyroscope to get the error of inductosyn. Compare the traditional Fourier Function calibrations and the linear interpolation calibrations methods; this paper presents a bit memory based method to calibrate the error of inductosyn in electrical resolution. As a result, after compensation, the inductosyn error is ± 0.8 arcsecond, RMS of the error is less than 0.3 arcsecond.

Keywords Inductosyn · Error model · Ring laser gyroscope · Calibration

1 Introduction

The accuracy requirement of angular measurement is stringent in variety application occasions, like space image detection, high-precision turntable, scanning measurement. Not only the static angle accuracy is necessary, but also the dynamic angle measurement ability is needed. The traditional angle sensors are angle encoder, optical gyroscope, resolver, inductosyn, etc. Encoder (photoelectric rotary encoder) assemble demand is strict, eccentricity and the tilt in the installation process must under 3–5 μm at shaft and plane, optical gyroscope due to shift with time, could not be used in the motion control in space environment. Resolver accuracy can't fulfill the arcsecond requirement due to volume and weight. Inductosyn can get high precision at same time the installation requirement is not accurate, below 30–50 μm at shaft and plane. In order to achieve sub-arcsecond accuracy, there are two way to reduce errors of inductosyn instrument. One is to improve inductosyn manufacturing craft

C. Cui (✉) · Y. Jiu · C. Wang

Department of Mechanism and Control Technology, Beijing Institute of Space Mechanics and Electricity, Beijing 100094, China
e-mail: cuichenpengbj@foxmail.com

F. Gu

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

and raising angle solver circuit designs ability. The other one is to calibrate the error of the exiting angular-measuring instrument with compensation function or method [1]. In order to improve inductosyn accuracy, the error of inductosyn should be measured. Due to the dynamic angle standard device and traceability technology [2], which based on Ring Laser Gyroscope, the error table can be obtained. The RLG is measured by the national small angle standard and the high precision turntable. The index and interpolation measuring deviation were distributed within ± 0.4 arcsecond. The angular measuring repeatability was less than 0.2 arcsecond [3, 4].

RLG accuracy is much higher than the inductosyn based angular sensor, but it is shifting with time, 12 arcsecond per minute, as its principle shows, we must standardize the zero before using it to eliminating the Earth rotation effect. Beside the shorts above, the space environment like vacuum, radiation, terminal deflection restricted the applications of RLG in orbit. In contrast, the inductosyn is not sensitive to space environments with high stability and reliability. On the other hand, we can still using RLG on ground to testing the inductosyn error table in a short time (~ 5 s) according to the requirements of use.

Using the RLG equipment to get the error of inductosyn, the paper developed a bit based look up table software algorithm to improve the inductosyn precision. Compare to the function compensation algorithm, this method can make the inductosyn angle instrument to sub-arcsecond, error RMS is less than 0.3 arcsecond.

2 Common Mathematical Error Model of the Single-Axis Induction Synchronizer

The induction synchronizer also known as inductosyn, can be used as an angular measuring instrument. It is a magic-electrical angle measurement, can achieve high accuracy with 180 pair poles or more compare to resolver. It can be used in ultra-high precise position servo systems, especially in space crafts. This paper using inductosyn sensor as an angular instrument in a scanning mirror mechanism, and the mechanism is a part of a remote satellite. In order to satisfy the stringent angular measurement accuracy requirement, the error of inductosyn should be calibrated on ground.

From [1, 5–7], the error of inductosyn mainly consists of two parts, one is zero position errors and subdivision errors. The manufacturing and the eccentricity and tilt in the installation process cause zero position errors. The amplitude of induced voltage and harmonic voltage, along with electrical reasons cause subdivision errors.

Components and derivation of inductosyn error can be show in Table 1.

From [5–10], the commonly zero error mathematical model of inductosyn could show as “Eq. (1)”.

$$y(x) = a_0 + a_1 \cdot \sin(x) + a_2 \cdot \cos(x) \quad (1)$$

From [1], the subdivision errors and zero error can be modeled based on FFT.

Table 1 Error components of inductosyn

Cause of error	Manifestations	Harmonic order
Non-Ideal voltage (Captive voltage drop, edge voltage, interference)	Zero position error	p, 3p, 5p ...
3th, 4th harmonic voltage	Subdivision Errors performs as fourth sine function	4p
The errors of the amplitude of the induced voltage in sine phase and cosine phase of stator	Subdivision Errors performs as secondary sine function	2p, 2p ± 1
The eccentricity and the tilt in the installation process	Zero Position Errors perform as primary cosine function in 360 degree (mechanical angle)	p
The errors of the amplitude of the induced potential of the sine phase and cosine phase of stator or the rotor conductor	Zero Position Errors perform as frequency K random form	K and a series harmonic voltage

p is the number of pairs of poles

FFT technique can transform the error datas. Measure a limited length set of discrete data $y(n)$ in a fixed sampling period is the basic idea of FFT. We use FFT to get their expression spectrum coefficients as the N data points are measured on sampling interval.

$$Y(k) = \sum_{n=0}^{N-1} y(n) \cdot e^{-j \cdot k \cdot \frac{2\pi}{N} \cdot n}, 0 \ll k \ll N - 1 \quad (2)$$

Then calculate Fourier inverse transformation of the spectrum coefficients, the linear combination mathematical expressions could be gotten.

$$y(n) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{j \cdot k \cdot \frac{2\pi}{N} \cdot n} = \sum_{n=0}^{N-1} x(n) \cdot e^{j \cdot k \cdot \frac{2\pi}{N} \cdot n}, 0 \ll k \ll N - 1 \quad (3)$$

According to the Eqs. (2) and (3), we can get

$$a_k = \sum_{n=0}^{N-1} x(n) \cdot e^{j \cdot k \cdot \frac{2\pi}{N} \cdot n}, 0 \ll k \ll N - 1 \quad (4)$$

Then analyze amplitude frequency characteristic by taking the measured datas into the Eq. (4). As a result we get the inductosyn error model according to the actual datas and the amplitude frequency characteristic.

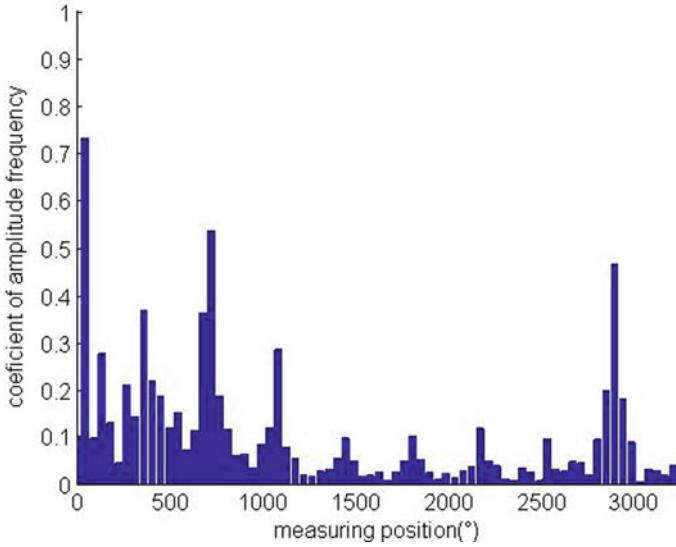


Fig. 1 Amplitude-frequency characteristic

$$y(x) = a_0 + a_1 \cdot \sin(p_1 \cdot x) + a_2 \cdot \cos(p_2 \cdot x) + a_3 \cdot \sin(p_3 \cdot x) + a_4 \cdot \cos(p_4 \cdot x) + \dots + a_{n-1} \cdot \sin(p_{n-1} \cdot x) + a_n \cdot \cos(p_n \cdot x) \quad (5)$$

Therein,

p_n —amplitude frequency characteristic parameters

a_n —system model parameters

$$x = \frac{2\pi i}{n}, i = 0, 1, 2 \dots N - 1$$

Form part 4, we get the error with position of the inductosyn instrument, the FFT result see Fig. 1. The main errors are 45° , 360° , 720° , 2880° .

The axis is shifting with position as the instrument is holding by two pivots, the error curve of each 2° present no repeatability. The error curve shift as the eccentricity and the tilt changed with position, see Fig. 2.

3 Comparison of Different Calibration Methods

From part 4, the error of inductosyn before compensation can be achieved, then we use the “cftool” of Matlab to compare the different methods of compensation angular error. The residual RMS error is 1.033 arcsecond in 1th Fourier equation fit, 0.7484 arcsecond in 4th Fourier equation, 0.5985 arcsecond in 8th Fourier equation.

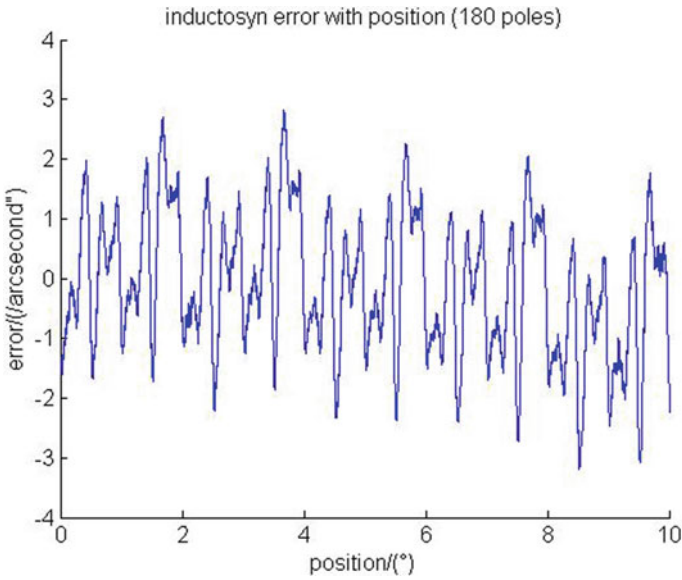


Fig. 2 Inductosyn error with position (180 poles)

The inductosyn be used is JGX/360-WND, which has 180 poles, mechanical angle 2° represent electrical angle 360° . See Figs. 3, 4 and 5.

The residual error is 0 uses the linear method, see (Fig. 6). The linear interpolation method is simple, intuitive and easy to implement in embedded control system.

In the test, the inductosyn was used to measure the single-axis scan mirror motion relative to the pivot base, the RLG mounted on the pivot were used to measure the

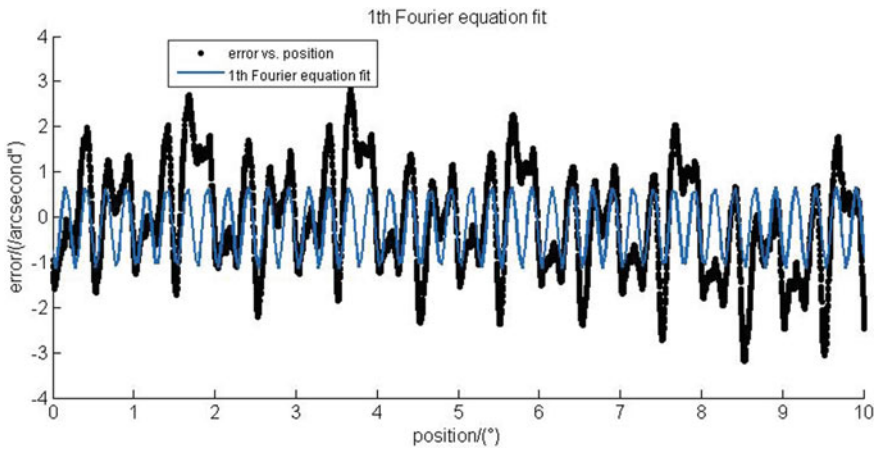


Fig. 3 1th Fourier equation fit

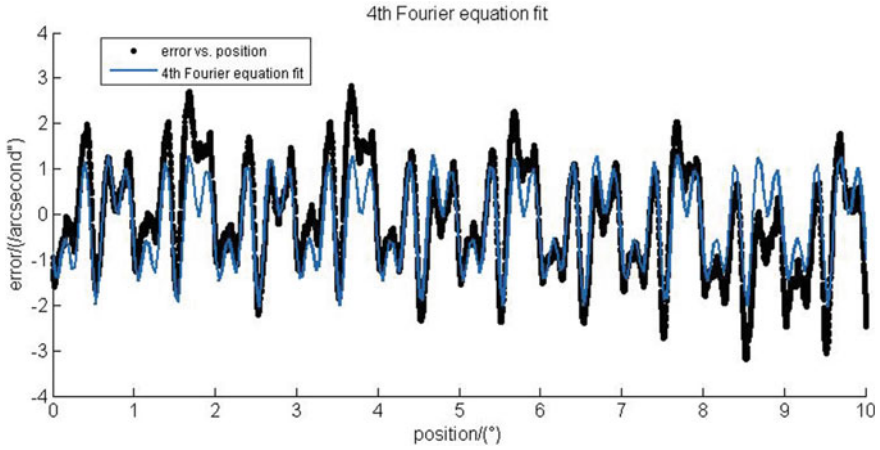


Fig. 4 4th Fourier equation fit

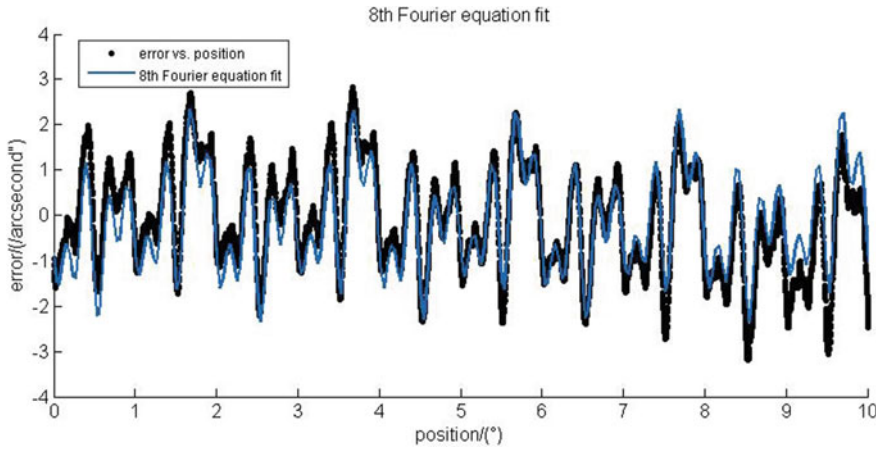


Fig. 5 8th Fourier equation fit

absolute motion of the scan mirror. After compensating for the pivot base motion (the motion due to Earth rotation), the difference between the gyro readouts and the inductosyn readouts representing the inductosyn error are computed and stored. As the datas are equal interval collection and storage, the calibration data of sequence i is (x_i, y_i) . So the calibration datas between i and $i + 1$ can be obtained by linear interpolation method. Consider the inductosyn out data of one angle is x , the model of interpolation is

$$x_c = x - \left(y_i + \frac{x - x_i}{\Delta x} \right) * (y_{i+1} - y_i), x_i < x < x_{i+1} \quad (6)$$

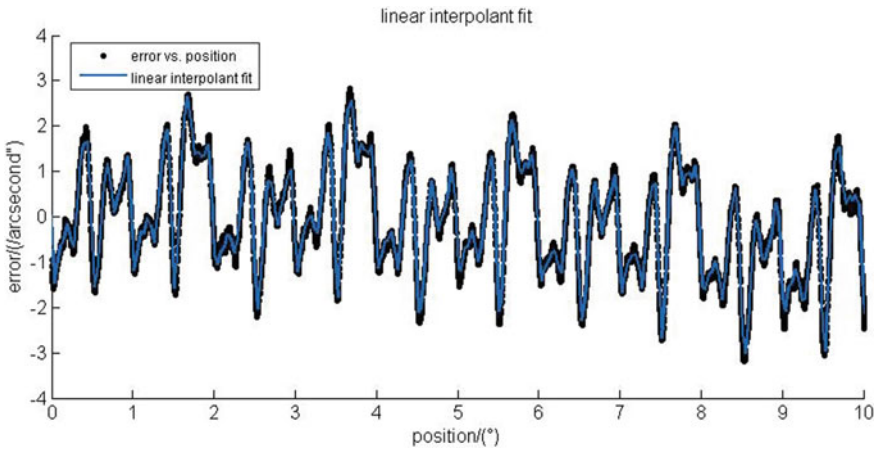


Fig. 6 Linear interpolation fit results

x_c is calibration value, Δx is the spacing of test datas, it determined by the resolution of the readout circuit and the calibration storage memory chip. In this operation, the resolution of inductosyn readout circuit is 2° divide by 2^{16} , 0.1098 arcsecond. As a result, we define this kind of calibrate method as bit memory based method, which means the spacing of datas is a bit of memory chip and equal the resolution of the inductosyn readout circuit.

Then we put the interpolation data in the memory, use look-up table method in the practice angle compensation.

4 Testing of Inductosyn Angle Measuring Errors

From the principle of optical gyroscope angle measuring instrument, using the angular measuring deviation mechanism and effectiveness of the scale factor, bias, and frequency stability etc. to testing the gyroscope. As a result, the Ring Laser Gyroscope (RLG) index error, segmentation errors (interpolation measuring deviation) were distributed within ± 0.24 arcsecond. The angular measuring repeatability was less than 0.1 arcsecond [3, 5, 11]. Then use RLG (MG033 of AVIC Xi'an Flight Automatic Control Research Institute) as an angular measuring instrument to measure the dynamical error of inductosyn is useful, the installation requirements are not strict. Compare to the move-less measurement instrument, such as Photo-electric autocollimator (± 1000 arcsecond), can measure much more large scale.

The basic idea of testing the inductosyn errors was using an ultra-high precision Ring Laser Gyroscope (RLG) to build an inductosyn errors test system. The rotor and stator of inductosyn were installed in parallel on the rotor and stator of the pivot bearing instrument, as shown in Figs. 7 and 8.

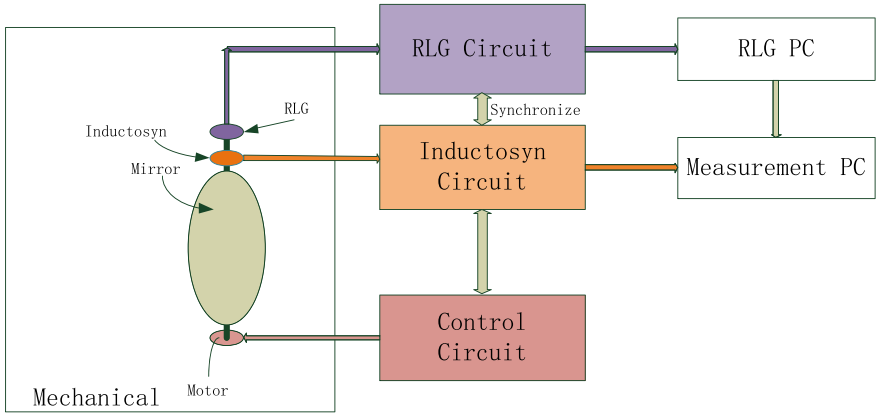


Fig. 7 Inductosyn errors test system

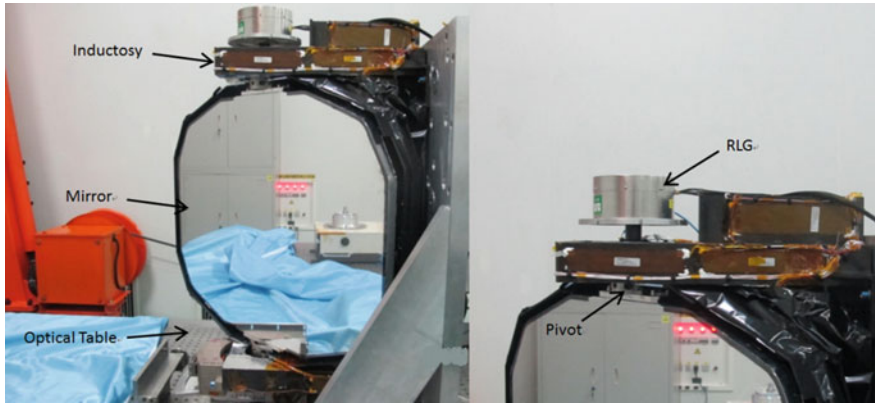


Fig. 8 Inductosyn instrument and ultra-high precision RLG

Figure 9 shows the lab of inductosyn error test. The inductosyn circuit consist amplifier and resolver to digit converter. The 360 electrical degree of the inductosyn is 2 mechanical degrees; the circuit converts the 2 mechanical degrees to a 16 bit digital signal, 1 bit represent 0.1099 arcsecond. The inductosyn circuit send synchronize signals to RLG circuit to get the angle of the mechanism at the same time. After storage the angle of inductosyn measurement PC and RLG PC, subtract the two can get error. Then using linear interpolation method, store the error table in resolution of 0.1099 arcsecond per bit in the memory of inductosyn circuit. After calibration, the error of inductosyn Re (Fig. 10), ± 0.8 arcsecond. RMS of the error is less than 0.3 arcsecond.

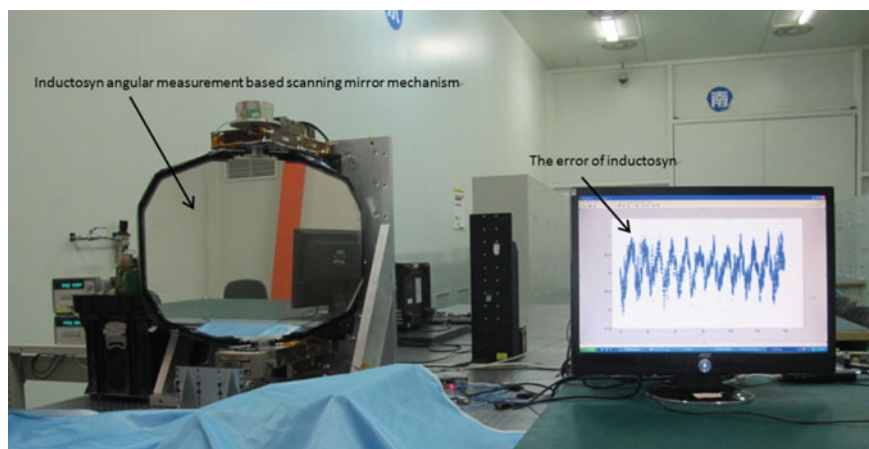


Fig. 9 Inductosyn error test lab

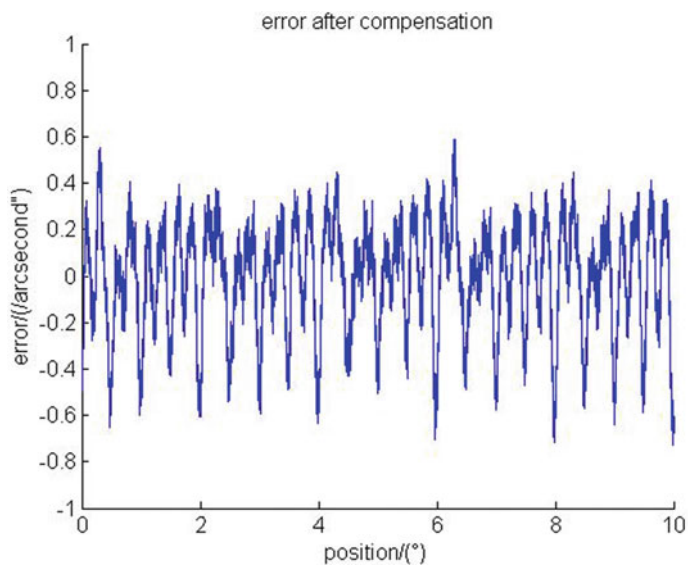


Fig. 10 Inductosyn error after compensation

5 Conclusion

Because of limitation in installation and electronics shortage, the accuracy of inductosyn is corrupted with errors. Using ultra-high precision RLG to measure the error of inductosyn, then we can get a high angular instrument after calibrate the error of inductosyn. Especially the mounting limitation of pivot axis, the repeatability of

each electrical pole is not similar. Using precision RLG gyroscope as calibration references can get the full scale dynamic position error.

The error table generated by RLG on ground may not satisfied usage on space operations as the error characteristics of inductosyn may vary due to terminal environments, mounting transmutation and component aging. In the future, the difference error calibration scheme of ground and space environments should be studied to fulfill the using range of angular compensation.

References

1. Liu, C., Zhu, M.: Error separation method of Inductosyn. In: Third International Conference on Instrumentation, Measurement, Computer, Communication and Control 2013, pp. 1233–1236. IEEE, Shenyang, China (2013)
2. Dan, Q.: Study on Dynamic Angle Standard and Traceability Technology, pp. 49–51. China Jiliang University, Master Dissertation (2017)
3. Yao, H., Zi, X., Dan, Q.: The metrological performance test of optical gyroscope angle measuring instrument. *ACTA Metrologica SINICA* **38**(12), 61–64 (2017)
4. Ring Laser Gyroscope MG033 testing report. CDJC2017-7740, National Institute of Metrology (2017)
5. Du, C., Yang, G.: Error analysis and compensation for Inductosyn-based position measuring system. In: Industry Applications Conference 2003, pp. 6–10. IEEE, China (2003)
6. Zou, A., Zhao, H., Ma, Y., Li, D.: Analysis calculation and testing of rotary Inductosyn angle measuring errors. In: The 33rd Chinese Control Conference 2014, pp. 28–30. IEEE, Nanjing, China(2014)
7. Wang, X., Wu, M., Zhu, G., Feng, J.: Detection and compensation of amplitude error and quadrature error for Inductosyn. In: The 2010 International Conference on Measuring Technology and Mechatronics Automation, pp. 1085–1088. IEEE, Changsha, China (2010)
8. Wang, X., Wu, M., Zhu, G., Feng, J.: The error analysis and harmonic wave correction of Inductosyn. In: 2nd International Asia Conference on Informatics in Control, Automation and Robotics 2010, pp. 358–361. IEEE, Wuhan, China (2010)
9. Bao, M.: Research on Technologies of Precise Angle Measurement Based on Inductosyn Transducer, pp. 8–28. Harbin Institute of Technology, Master Dissertation (2013)
10. Zhu, M.: Research on Back to Back Inductosyn Angle measuring System, pp. 4–6. Harbin Institute of Technology, Master Dissertation (2013)
11. Wu, Y.A.: On-orbit calibration of Inductosyn error. In: American Control Conference 1994, pp. 2887–2891. IEEE, Baltimore, MD, USA (1994)

Data Analysis for Predictive Maintenance Using Time Series and Deep Learning Models—A Case Study in a Pulp Paper Industry



Balduino Mateus, Mateus Mendes, José Torres Farinha,
Alexandre Batista Martins, and António Marques Cardoso

Abstract Predictive maintenance is fundamental for modern industries, in order to improve the physical assets availability, decision making and rationalize costs. That requires deployment of sensor networks, data storage and development of data treatment methods that can satisfy the quality required in the forecasting models. The present paper describes a case study where data collected in an industrial pulp paper press was pre-processed and used to predict future behavior, aiming to anticipate potential failures, optimize predictive maintenance and physical assets availability. The data were processed and analyzed, outliers identified and treated. Time series models were used to predict short-term future behavior. The results show that it is possible to predict future values up to ten days in advance with good accuracy.

Keywords Data analysis · Autoregressive models · ARIMA · Deep Learning · Time series forecasting · Predictive maintenance

B. Mateus (✉) · A. B. Martins

EIGeS—Research Centre in Industrial Engineering, Management and Sustainability, Lusófona University, Campo Grande, 376, 1749-024 Lisboa, Portugal
e-mail: balduino.mateus@ubi.pt

A. B. Martins

e-mail: p5922@ulusofona.pt

B. Mateus · A. B. Martins · A. M. Cardoso

CISE—Electromechatronic Systems Research Centre, University of Beira Interior, Calçada Fonte Do Lameiro, P-62001-001, Covilhã, Portugal

J. T. Farinha

CEMMPRE—Centre for Mechanical Engineering, Materials and Processes—CEMMPRE, 3030-788 Coimbra, Portugal
e-mail: torresfarinha@dem.uc.pt

M. Mendes · J. T. Farinha

Polytechnic of Coimbra—Polytechnic of Coimbra, ISEC, 3045-093 Coimbra, Portugal
e-mail: mmendes@isr.uc.pt

M. Mendes

Institute of Systems and Robotics, University of Coimbra, 3004-531 Coimbra, Portugal

1 Introduction

Life cycle optimization has been a concern for decades; it becomes clear that a physical asset with an adequate maintenance will have a longer life with a greater return for the organization [1]. Monitoring industrial equipment is essential to anticipate and avoid potential failures, which can endanger people and assets. Sensors are deployed and data are collected to facilitate and automate the process. The methods applied to treat and analyse the data are relevant for improving the fault detection performance, predictions and decision making. Data cleaning is one of the key challenges [2, 3], so that excess data or wrong data can be removed out of the analysis process. Using the data collected and properly treated, machine learning models can be trained, parameters can be calculated and obtained, so that actions, decision making, control, supervision and planning can be implemented to optimize manufacturing plants production processes [4, 5].

One of the biggest challenges is the elimination of duplicate data and noise. Gong et al. [6] propose a simple binary classifier to separate useful data from bad data with 99% accuracy. Veit et al. [7] propose an approach that consists of combining clean and noisy data, pre-training a network using a large noisy data set, and then fine-tune it with the clean data set. Plutowski and White in 1993 use a multi-layer feedforward neural network architecture to find patterns of bad quality data in a dataset [8].

Sensor data recorded along the time can be processed using time series methods. According to Zhang [9], the classical decomposition method of the time series is, for example, to decompose a seasonal time series into trend, seasonal, cyclical, and irregular components. After the components are known, the data can then be used to adjust or train suitable machine learning models.

Time series prediction models such as Autoregressive Integrated Moving Average (ARIMA) models, as well as Artificial Neural Networks (ANN) are frequently used and compared, with mixed conclusions about the superiority in forecasting performance [9, 10]. Mateus et al. [11] discuss the disadvantages that the Autoregressive Moving Average (ARMA) time series model presents to forecast when faced with oscillatory data (dummy variables). In the case of complex problems that have both linear and non-linear correlation structures, the combination of ARMA with ANN is an effective way to improve forecasting performance. Although ANN are essentially nonlinear models, they have a capacity of modelling linear processes as well [9].

Deep learning methods are capable of identifying the structure and patterns of data, such as non-linearity and problems of complexity in time series forecasting [10]. Backpropagation networks (BP) are good at solving a wide variety of problems, and are used in time series forecasting [12]. According Hecht-Nielsen (1989), the standard BP network using a subjective transfer function can learn any measurable function in a very precise manner when a sufficient number of hidden neurons are used to [13].

The paper is structured as follows: Section two gives an overview of prediction problems on maintenance and some problem reviews about the prediction model and

some solutions; In the third section it is presented Data Characterization and Pre-processing; In the fourth section a case study, to evaluate and validate the forecasting models, is presented; Finally, the conclusions of the study are made.

2 Related Work

Machine learning methods are increasingly popular in predictive maintenance. Jimenez et al. [14] showed that there exists potential in the development of predictive models for application in predictive maintenance. Rodrigues et al. [15] use neural networks and principal component analysis to assess diesel engine oil degradation and determine the optimal point for oil replacement.

Daniyan et al. [16] combine ANN with a dynamic time series model in diagnosing failures, to optimize maintenance intervention time in industrial equipment. Ayvaz and Alpay [17] propose a method to improve maintenance planning to minimize unexpected stops, through the combined use of Ensemble Empirical Mode Decomposition and Long Short-Term Memory. Huang et al. [18] apply Long Short-Term Memory (LSTM) neural network approaches to forecast real production data, obtaining satisfactory results, superior to conventional models.

Using deep networks to carry out stock market forecasting, Nti et al. [19] reach a fairly satisfactory result of forecasting. They concluded that the efficient fusion of information from different sample indicators offers greater precision than individual data. Liu et al. [20], using an elastic mesh algorithm and LSTM to calculate the remaining bearing life, demonstrate that this algorithm can achieve good stability in terms of problem prediction. Still et al. [21] used an LSTM network to predict the current situation of an engine—their model demonstrated good forecasting reliability.

3 Data Characterization and Pre-processing

3.1 Dataset and Framework

Data used in the present work are the result of monitoring an industrial paper press system. Six sensors are monitoring the functioning of the press, with a sampling period of 1 min. The variables monitored are: (1) Electrical Current Intensity; (2) Hydraulic Unit Oil Level; (3) VAT Pressure; (4) Rotation Velocity, (5) Temperature in the Hydraulic Unit; and (6) Torque.

The dataset contains the sensor readings from 1 February 2018 to 30 October 2020. There are 1,490,400 samples in the dataset. Figure 1 shows a plot of the values of all variables in the original dataset. This dataset was loaded in python and processed, using ScyPy libraries.

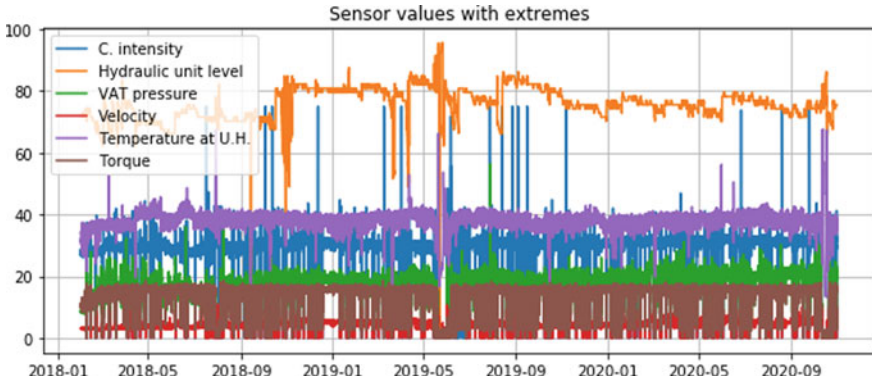


Fig. 1 Plot of the original dataset values. The variables are: C. intensity, Hydraulic unit level, VAT pressure, velocity, Temperature u. l., and Torque

3.2 Data Characterization and Identification of Discrepant Data

As Fig. 1 shows, there are some sensor readings which show extreme levels. The very large values may be reading errors or overload moments. The very low values may be when the press was stopped, malfunctioning, underused, or they may also be reading errors. Those extremely low or extremely high values provide information about abnormal functioning of the press. They may negatively impact the performance of the forecasting algorithms.

Table 1 shows some statistical values of the data, namely the mean, standard deviation, minimum and maximum values: Fig. 2 shows histograms of the variables' quartiles.

Figure 3 shows the amplitude of each sample concerning the lower and upper bounds for each variable. As the figure shows, the distribution of data is skewed for all variables (Fig. 4).

Table 1 Statistical parameters of the variables: C. intensity, Hydraulic unit level, VAT pressure, velocity, Temperature u. l., and Torque

	C. intensity	Hydraulic	Torque	VAT	Velocity	Temperature
Mean	30.26	75.90	15.28	18.25	4.59	38.22
Std	1.32	4.54	0.69	2.67	0.977	1.62
Min	26.34	62.93	13.59	9.67	1.27	33.19
25%	29.30	72.86	14.90	17.13	3.92	37.17
50%	30.46	75.53	15.43	18.72	4.57	38.33
75%	31.28	79.52	15.78	19.97	5.28	39.35
Max	34.26	88.97	17.09	26.17	7.87	43.10

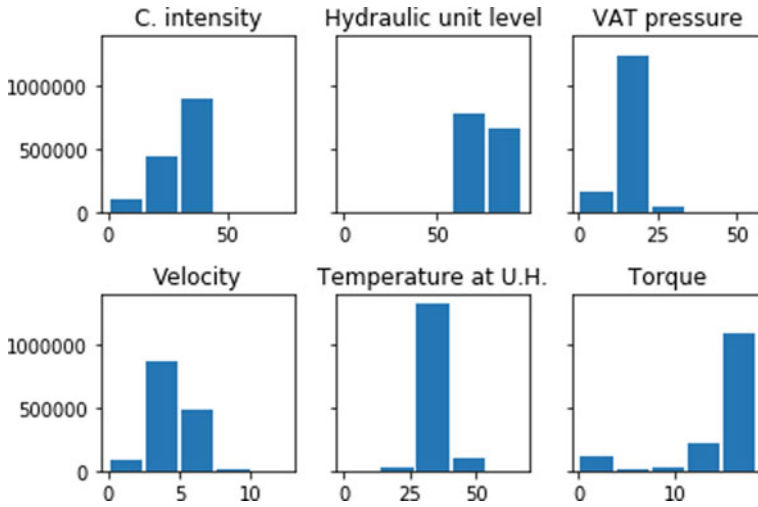


Fig. 2 Histogram of variables showing the number of samples per quartile

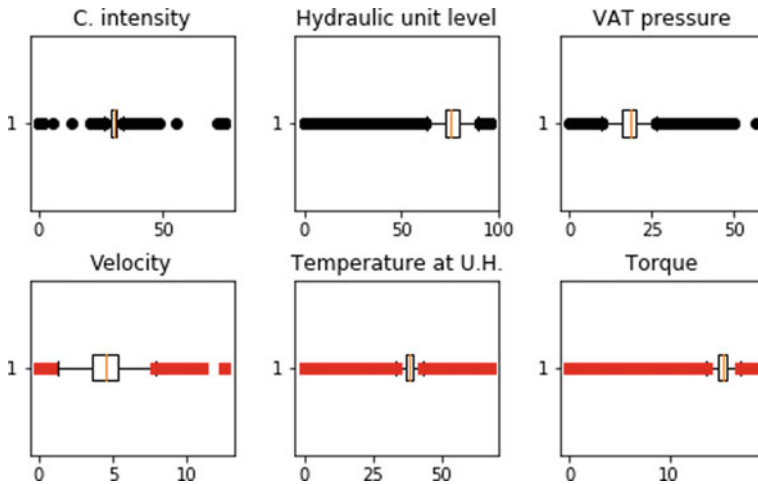


Fig. 3 Distribution of data points of all the sensors, with low and high extremes

In order to achieve best performance in training predictive machine learning models, discrepant data (Ning and You 2017) must be identified and possibly removed. The method used was the quartile approach, as explained in Formulae (1) to (5). In the formulae, $Q_{\frac{1}{4}}$ is the first quartile, $Q_{\frac{3}{4}}$ is the end quartile, n is sample number and IQ Interquartile Range.

$$Q_{\frac{1}{4}} = \frac{1}{4}(n + 1) \tag{1}$$

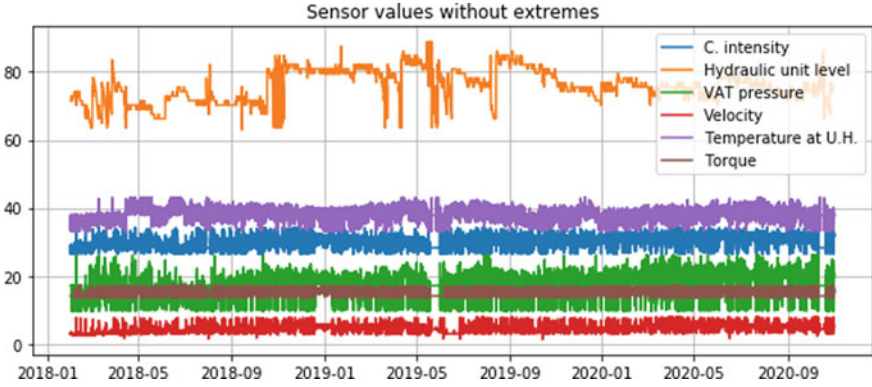


Fig. 4 Plot of the dataset variables without extreme values: current intensity, hydraulic unit level, VAT pressure, rotation velocity, temperature in the hydraulic unit, and torque

$$Q_{\frac{3}{4}} = \frac{3}{4}(n + 1) \quad (2)$$

$$IQ = Q_{\frac{1}{4}} - Q_{\frac{3}{4}} \quad (3)$$

$$\text{Down}_{\text{limit}} = Q_{\frac{1}{4}} - k.IQ. \quad (4)$$

$$\text{Up}_{\text{limit}} = Q_{\frac{3}{4}} + k.IQ \quad (5)$$

$\text{Down}_{\text{limit}}$ is the lower bound limit accepted for the variable, calculated by subtracting of the constant k multiplied IQ to $Q_{\frac{1}{4}}$. Up_{limit} is the upper bound limit accepted for the variable, calculated by adding the constant k multiplied IQ to $Q_{\frac{3}{4}}$, where k is the variation constant of the limits.

After the application of the quartile method described above, the discrepant samples are taken out of the dataset. Namely, samples which are not in the interval $\text{Down}_{\text{limit}}$ were removed. As shown in Figs. 4, 5 and 6.

3.3 Study of Correlations

Correlations between variables, as well as autocorrelations, are very important to have a better insight into the dependence of variables and determine which data models can be applied with higher probability of success. shows the matrix of correlations between variables. It is possible to verify some strong correlations between the variables Current intensity and Velocity, among others that are presented in the graphs below. Most of the correlations, however, are weak.

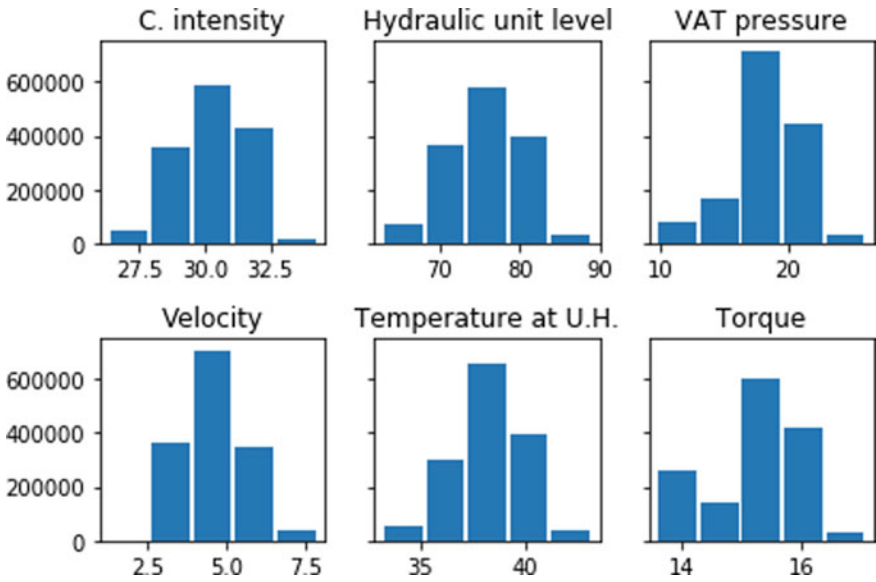


Fig. 5 Histogram of variables after removing discrepant data. The variables are Current intensity, hydraulic unit level, VAT pressure, rotation velocity, temperature in the hydraulic unit, and torque

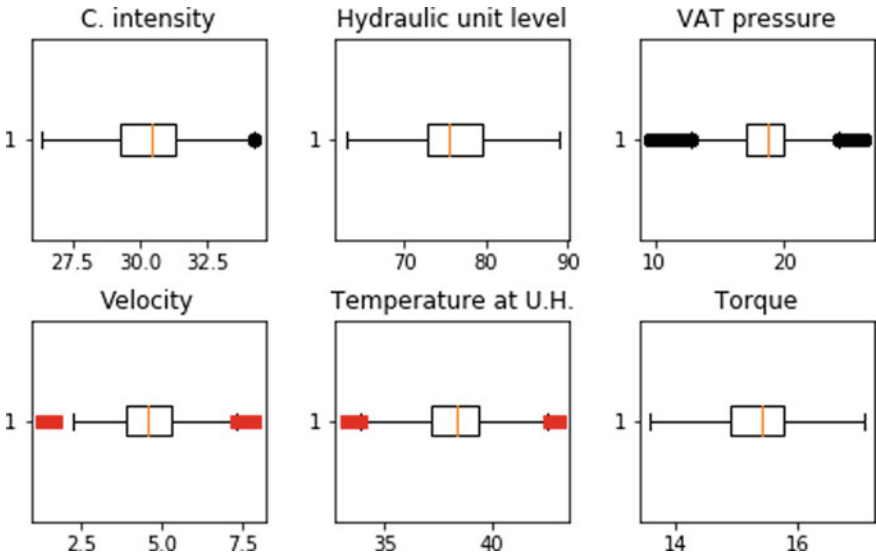


Fig. 6 Distribution of samples for all the sensors after removal of discrepant data. The variables are: current intensity, hydraulic unit level, VAT pressure, rotation velocity, temperature in the hydraulic unit, and torque

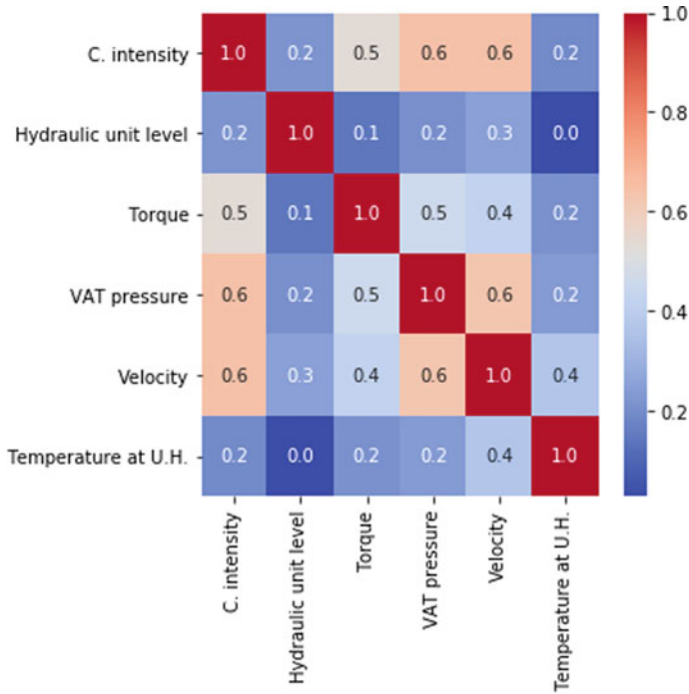


Fig. 7 Correlations among all variables

Figure 7, shows the autocorrelation of each variable. As the figure shows, the autocorrelations decay very quickly to less than 0.5. The charts show, therefore, that the correlation and autocorrelation of variables are very weak.

In Autocorrelation there is a decay in the period making the correlation increasingly lower. The graph in Fig. 8, shows only a correlation of up to 200 samples that also served for the test and for the forecast. Since the number of the samples is very high (1,490,400), there was a need to down sample the dataset, from a period of minutes to a period of days, in order to have a forecast in days. That was done by averaging the samples of each day using the python pandas function “df.resample(‘D’). Mean ()”.

4 Modelling Using Time Series

4.1 Autoregressive Model

As a first approach to predict future behaviour, an autoregressive model was applied. Autoregressive models are adequate to model variables that depend mostly on their previous behavior and a stochastic value, thus satisfying the following equation:

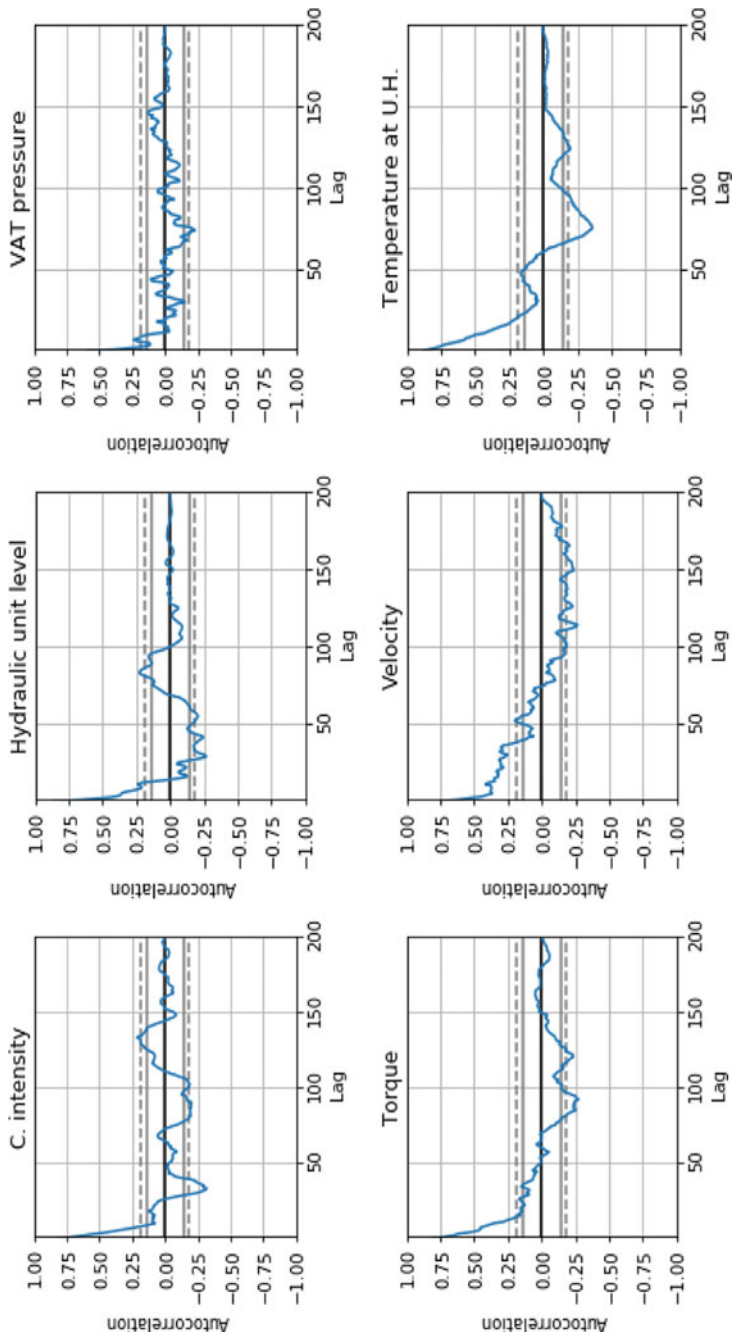


Fig. 8 Autocorrelation between samples of all variables, calculated for 200 days

$$AR = \vartheta_1 X_{t-1} + \vartheta_2 X_{t-2} + \vartheta_p X_{t-p} + \epsilon_t \quad (6)$$

where $\vartheta_1, \dots, \vartheta_p$ are real parameters and ϵ_t is a white noise process independent and identically distributed.

4.2 ARIMA and SARIMA Models

Some time series present a seasonal periodic component. A seasonal autoregressive model is characterized by the existence of a significant correlation between observations spaced by a multiple time interval [22].

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is a general case of the models proposed by [23] Box And Jenkins at 1976, for the adjustment of stationary time series. However, when there is a seasonal component in the data, the model class is called SARIMA (p, d, q) (P, D, Q), given by:

$$MA = -\theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (7)$$

where $\theta_1, \dots, \theta_p$, are parameters of an order of structures, ϵ_t is white noise with zero mean.

$$AR_s = \Phi_1 X_{t-1s} + \Phi_2 X_{t-2s} + \dots + \Phi_p X_{t-ps} \quad (8)$$

$$MA_s = -\Theta_1 \epsilon_{t-1} - \Theta_2 \epsilon_{t-2} - \dots - \Theta_Q \epsilon_{t-Q} \quad (9)$$

$$\nabla^D \Delta^d X_t = AR - MA + AR_s + MA_s \quad (10)$$

where (p, d, q) refer to the model orders of the seasonal part: p is trend autoregression order, d is trend difference order and q is trend moving average order. (P, D, Q) is the same but with the Seasonal component. The parameters Φ_1, \dots, Φ_p , are the parameters referring to the seasonal autoregressive part and $\Theta_1, \dots, \Theta_Q$, are the parameters of moving averages, and i is an error that cannot be estimated from the model and D indicates the number of seasonal differences made in the series to park it. The calculation of the parameters of the models that best fits was made using the most frequent Akaike Information Criterion (AIC), which is defined by:

$$AIC = 2\log(L \cdot k) + 2(k) \quad (11)$$

where $L \cdot k$ is the maximized log-likelihood and k is the number of parameters in the model.

5 Experiments and Results

5.1 Results of the Autoregressive Model

The model was applied with a 20-day sliding window, thus corresponding to $1440 \times 10 = 14,400$ data samples and a forecast window with the same size, thus predicting the values for the next 10 days.

After eliminating the discrepant data samples, some irregularities in the samples, which may be momentary or prolonged damage, are still visible in Fig. 9. Nonetheless, the autoregressive model shows a good fit to all variables. The prediction errors between the forecasted values and the actual values are given in Table 2.

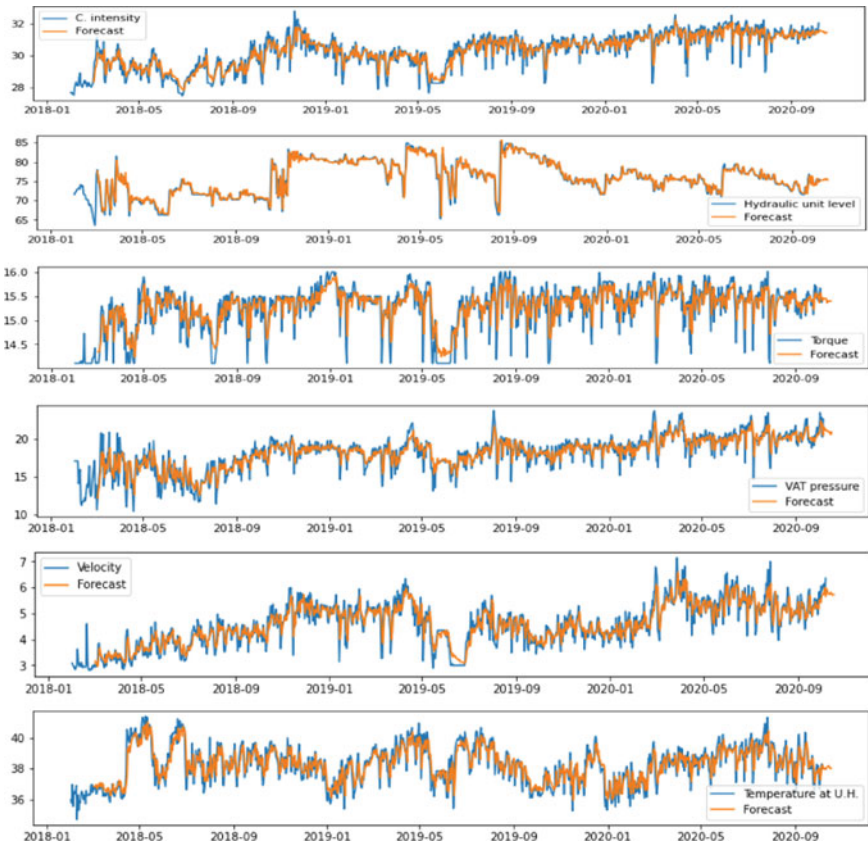


Fig. 9 Prediction of the six variables using a retrogressive method, with 20 days lag and 10 days predict window

Table 2 Summary of the MAE, MSE and MAPE errors for the autoregressive and SARIMA models tested

		C. intensity	Hydraulic unit level	Torque	VAT pressure	Velocity	Temp. at U.H
AR	MAE	3.60	3.06	0.93	2.97	1.15	0.36
	MSE	13.08	18.37	1.15	10.54	1.47	0.24
	MAPE	12.92	3.76	6.57	16.95	25.17	0.97
SARIMA	MAE	0.21	1.05	0.15	0.57	0.25	0.64
	MSE	0.07	1.32	0.03	0.59	0.11	0.66
	MAPE	0.68	1.40	0.95	2.83	4.50	1.71

5.2 Results of the SARIMA Model

The AIC was used to define the hyperparameters p , d , q , P , D and Q of the SARIMA model. The seasonal period was fixed at 12 for all-time series.

Figure 10 shows that the SARIMA model gives stable predictions, when using 20 days sliding window and parameters SARIMA (0, 1, 2) (1, 1, 2).

Table 2 shows the results of the forecasting errors in the period referring to the two models, AR and SARIMA. The table shows the Mean Average Error (MAE), the Mean Squared Error (MSE) and the Mean Average Percent Error (MAPE).

6 Discussion

Using the two models, (Auto-regressive and SARIMA), it was possible to verify that both offer acceptable prediction errors, with the data evaluated. The SARIMA model shows better performance than the AR model, what is expectable since it encompasses the three different components (autoregressive, moving averages and seasonal component). However, that implies a cost of an additional processing time. The SARIMA model takes approximately 15 times more computing time than the AR model. As the SARIMA model, its processing lasted 40 s and for the AR model 4 s.

For the regressive model prediction, there were no hyperparameters to optimize. However, to find the best model, it was necessary to evaluate several models and to choose the parameters that best fit the data, using the AIC information criterion. It can be concluded that there is a good capacity of these models to predict based on data that presents a moderate variation.

For short-term forecasting, the models are satisfactory, emphasizing the need to clean the discrepant data. According to new studies in this area, they show superiority in the growth of the use of Neural Networks for those objectives, namely Recurrent Neural Networks that have greater long-term and short-term forecasting efficiency due to their Long-Short Term Memory capacity [19, 20, 24]. That is planned as

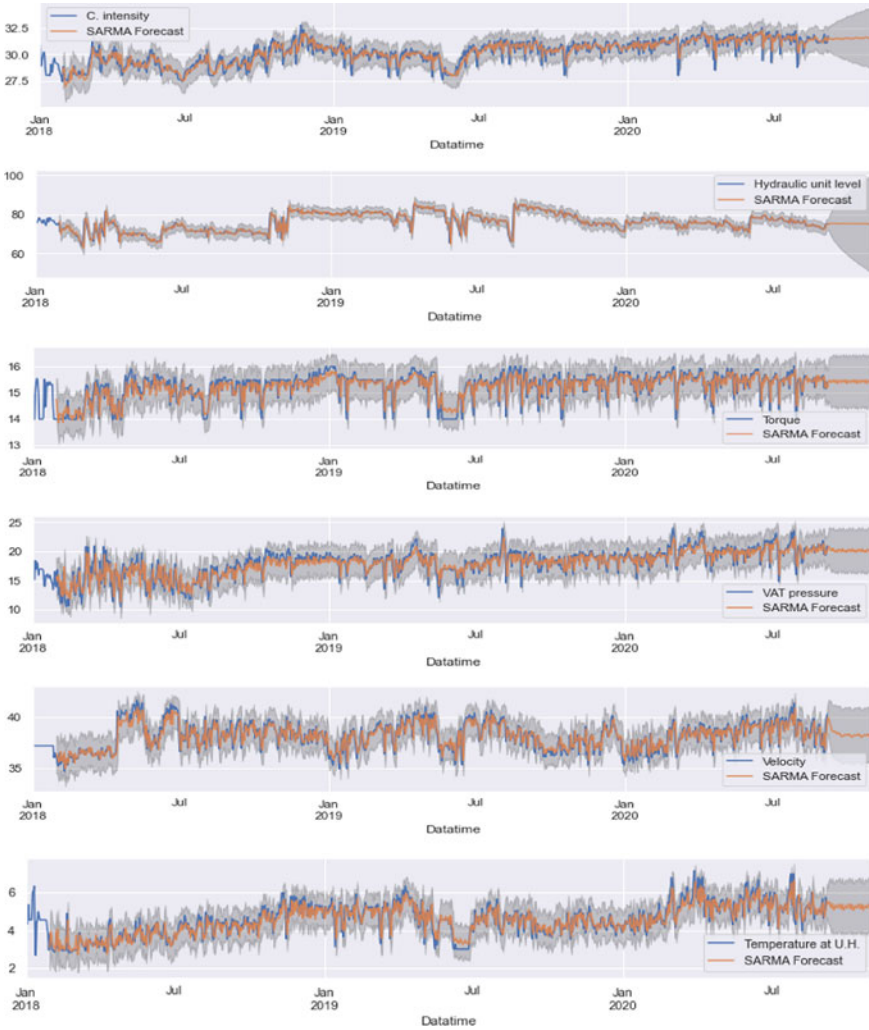


Fig. 10 Prediction of variables using the SARIMA (0, 1, 2) (1, 1, 2) method, with 20 days lag and 10 days predict window

future work in the present project, where deep neural models will be designed and optimized for prediction.

7 Conclusion

Sensor data is fundamental to monitor industrial equipment and processes. The present paper describes a case study where six variables were sampled during almost three years, with a period of one minute. The data were selected and cleaned of discrepant data samples, analysed and used to forecast future behaviour with time series models, namely Autoregressive and SARIMA. Data processing and experiments were carried out in Python using ScyPy libraries. The SARIMA model showed smaller errors in the test set, so it is more adequate for the data analysed.

Future work includes experiments with Neural Networks and larger forecast range predictions.

Acknowledgements The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 871284 project SSHARE and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-029494, and by National Funds through the FCT—Portuguese Foundation for Science and Technology, under Projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

References

1. Pais, E., Farinha, J.T., Cardoso, A.J.M., Raposo, H.: Optimizing the life cycle of physical assets—a review. *WSEAS Trans. Syst. Control* **15**, 417–430 (2020). <https://doi.org/10.37394/23203.2020.15.42>
2. Chaudhuri, S., Dayal, U.: An overview of data warehousing and OLAP technology. *ACM SIGMOD Rec.* **26**(1), 65–74 (1997). <https://doi.org/10.1145/248603.248616>
3. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM* **39**(11), 27–34 (1996). <https://doi.org/10.1145/240455.240464>
4. Prado, L.O., Ribeiro, P.F., Duque, C.A., Abdel Aleem, S.H.E.: Chapter 19—Modeling and processing of smart grids big data: study case of a university research building. In: Abdel Aleem, S.H.E., Abdelaziz, A.Y., Zobaa, A.F., Bansal, R. (Eds.) *Decision Making Applications in Modern Power Systems*, pp. 507–538. Academic Press (2020)
5. Martins, A.B., Torres Farinha, J., Marques Cardoso, A.: Calibration and certification of industrial sensors—a global review. *WSEAS Trans. Syst. Control* **15**, 394–416 (2020). <https://doi.org/10.37394/23203.2020.15.41>
6. Gong, Z., Wang, W., Ku, W.-S.: Adversarial and clean data are not twins. *ArXiv170404960 Cs*, Apr. 2017. Accessed: Mar. 02, 2021. [Online]. Available: <http://arxiv.org/abs/1704.04960>
7. Veit, A., Alldrin, N., Chechik, G., Krasin, I., Gupta, A., Belongie, S.: Learning From Noisy Large-Scale Datasets With Minimal Supervision, pp. 839–847, Accessed: Mar. 02, 2021. [Online] (2017). Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Veit_Learning_From_Noisy_CVPR_2017_paper.html
8. Plutowski, M., White, H.: Selecting concise training sets from clean data. *IEEE Trans. Neural Netw.* **4**(2), 305–318 (1993). <https://doi.org/10.1109/72.207618>
9. Zhang, Z.: Neural networks: further insights into error function, generalized weights and others. *Ann. Transl. Med.* **4**(16) (2016). <https://doi.org/10.21037/atm.2016.05.37>

10. Siami-Namini, S., Namin, A.S.: Forecasting economics and financial time series: ARIMA vs. LSTM, ArXiv180306386 Cs Q-Fin Stat, Mar. 2018, Accessed: Mar. 09, 2021. [Online]. Available: <http://arxiv.org/abs/1803.06386>
11. Mateus, B., Farinha, J.T., Cardoso, A.M.: Production Optimization Versus Asset Availability—A Review, vol. 15, p. 13 (2020). <https://doi.org/10.37394/23203.2020.15.33>
12. Kaastra, I., Boyd, M.: Designing a neural network for forecasting financial and economic time series. *Neurocomputing* **10**(3), 215–236 (1996). [https://doi.org/10.1016/0925-2312\(95\)00039-9](https://doi.org/10.1016/0925-2312(95)00039-9)
13. Hecht-Nielsen, R.: Neurocomputer applications. In: *Neural Computers*, pp. 445–453, Berlin. https://doi.org/10.1007/978-3-642-83740-1_45
14. Jimenez, V.J., Bouhmala, N., Gausdal, A.H.: Developing a predictive maintenance model for vessel machinery. *J. Ocean Eng. Sci.* **5**(4), 358–386 (2020). <https://doi.org/10.1016/j.joes.2020.03.003>
15. Rodrigues, J., Cost, I., Farinha, J.T., Mendes, M., Margalho, L.: Predicting motor oil condition using artificial neural networks and principal component analysis. *Eksploat. Niezawodn. - Maint. Reliab.* **22**(3), 440–448 (2020). <https://doi.org/10.17531/ein.2020.3.6>
16. Daniyan, I., Mpofu, K., Oyesola, M., Ramatsetse, B., Adeodu, A.: Artificial intelligence for predictive maintenance in the railcar learning factories. *Proc. Manuf.* **45**, 13–18 (2020). <https://doi.org/10.1016/j.promfg.2020.04.032>
17. Ayvaz, S., Alpay, K.: Predictive maintenance system for production lines in manufacturing: a machine learning approach using IoT data in real-time. *Expert Syst. Appl.* **173**, 114598 (2021). <https://doi.org/10.1016/j.eswa.2021.114598>
18. Huang, X., Zanni-Merk, C., Crémilleux, B.: Enhancing Deep Learning with semantics: an application to manufacturing time series analysis. *Proc. Comput. Sci.* **159**, 437–446 (2019). <https://doi.org/10.1016/j.procs.2019.09.198>
19. Nti, I.K., Adekoya, A.F., Weyori, B.A.: A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *J. Big Data* **8**(1) (2021). <https://doi.org/10.1186/s40537-020-00400-y>
20. Liu, M.-D., Ding, L., Bai, Y.-L.: Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. *Energy Convers. Manag.* **233**, 113917 (2021). <https://doi.org/10.1016/j.enconman.2021.113917>
21. Aydin, O., Guldamlasioglu, S.: Using LSTM Networks to Predict Engine Condition on Large Scale Data Processing Framework, pp. 281–285 (2017). <https://doi.org/10.1109/ICEEE2.2017.7935834>
22. Khandelwal, I., Adhikari, R., Verma, G.: Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. *Proc. Comput. Sci.* **48**, 173–179 (2015). <https://doi.org/10.1016/j.procs.2015.04.167>
23. Yip, H., Fan, H., Chiang, Y.: Predicting the maintenance cost of construction equipment: comparison between general regression neural network and Box-Jenkins time series models. *Autom. Constr.* **38**, 30–38 (2014). <https://doi.org/10.1016/j.autcon.2013.10.024>
24. Gui, Z., et al.: LSI-LSTM: an attention-aware LSTM for real-time driving destination prediction by considering location semantics and location importance of trajectory points. *Neurocomputing* **440**, 72–88 (2021). <https://doi.org/10.1016/j.neucom.2021.01.067>

Reinforcement Learning Fault Diagnosis Method Based on Less Tag Data



Kuo Xin, Jianguo Wang, and Wenxing Zhang

Abstract Vibration signals are often used in the fault diagnosis of rotating machinery. However, due to the influence of complex environment, environmental noise is often doped, and the diagnostic accuracy is reduced. The traditional deep self-encoder is used in the noise reduction process of rotating machinery fault diagnosis. The pooling model is poor and easy to lead to over-fitting problems, and deep learning training needs a large number of labeled data. Therefore, this paper proposes a reinforcement learning fault diagnosis method based on less label data. The random pooling is used to replace the pooling layer of the original convolutional self-encoder, and the exponential linear unit (ELU) is used to replace the original activation function to enhance the convolutional self-encoder. A large number of unlabeled samples are used for training, and then the deep reinforcement learning is used for network fine tuning. The experimental results of the sensor data collected by the fault diagnosis test bench show that the method used has a good improvement in denoising ability and feature extraction ability, and the recognition accuracy and stability are better than traditional convolutional autoencoder and traditional machine learning methods.

Keywords Reinforcement learning · Stochastic pooling · Convolutional auto-encoder · Fault diagnosis

1 Introduction

In the era of electromechanical big data, the use of advanced theories and methods to automatically mine information from the big data of mechanical equipment, replace diagnostic experts for feature extraction, conduct real-time monitoring and diagnosis of basic components such as bearings, and ensure the accuracy and efficiency of fault diagnosis and prediction has become a hot topic of current research.

K. Xin (✉) · J. Wang · W. Zhang
College of Mechanical Engineering, Inner Mongolia University of Science and Technology,
Baotou 014010, Inner Mongolia, China
e-mail: xinkuo1121@163.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_3

In the field of fault diagnosis, most machine learning algorithms are supervised learning [1, 2]. These methods require a large number of label data (known fault types) to extract features for fault classification, and label data requires strong professional knowledge and a large number of experiments. Yan et al. [3] used deep neural networks based on Denoising Sparse Auto-encoders to diagnose faults of bearings and motors. He et al. [4–6] used multi-channel signals to pre-train a set of source CNNs improved by random pooling and LReLU. A new framework of enhanced convolution neural network (ECNN) based on small-label infrared thermal image and convolution self-encoder (CAE) transmission. Zhou et al. [7] proposed a new deep neural network model, one-dimensional residual convolutional auto-encoder (1DRCAE). Although these methods have good performance in recognition accuracy, there are still the following problems: (1) Supervisory training methods often require a large number of label data and limit its applicability; (2) The pooling model has poor adaptability and low feature selection ability, and is easy to lead to over-fitting problems.

To solve the above problems, random pooling is used to replace the pooling layer of the original convolutional self-encoder, and the exponential linear unit (ELU) is used to replace the original activation function to enhance the convolutional self-encoder. A large number of unlabeled samples are used for training, and then the deep reinforcement learning is used for network fine tuning. The experimental results of the sensor data collected by the fault diagnosis test bench show that the method used has a good improvement in denoising ability and feature extraction ability, and the recognition accuracy and stability are better than traditional convolutional autoencoder and traditional machine learning methods.

2 Deep Reinforcement Learning

2.1 Markov Decision Process Modeling

Markov decision process is generally used to describe reinforcement learning tasks. The Markov decision process includes four tuples $\langle S, A, p, r \rangle$ that the model is easier. Markov decision process has Markov attributes: at the current time, the feedback of the environment only depends on the state and action of the previous time, and there is no correlation with the earlier time [8]. In the fault identification task in this paper, state s is one-dimensional fault signal; Action a is the category of mechanical fault: $0, 1, 2, \dots, Y - 1$, where Y is the number of fault categories; Whether the model identification results are consistent with the fault sample type is an important criterion for rewarding r . When the sample type is consistent with the identification results, r takes $+1$, otherwise r takes -1 . For the state transition probability p , although there is no correlation between the states, in order to avoid the over-fitting phenomenon during the network operation, the training data samples will be arbitrarily disrupted and the number of training data samples will be evenly distributed. The state transition

probability will become $p = 1/Y$, and s' can only be obtained by the probability p transfer of state s after action a is executed, which is not correlated with the state of the previous moment, and in line with the Markov property.

2.2 Strategy

The goal of reinforcement learning is to solve the optimal strategy of Markov decision process. The optimal strategy can be obtained by solving the optimal value function [9]. In short, the solution of reinforcement learning is finally evolved into the optimal Bellman equation. The mathematical model is:

$$V(s) = E[R_{t+1} + \gamma V(s_{t+1})] \quad (1)$$

where γ is a discount coefficient, which is a constant between $[0, 1]$. The existence of γ can ensure that the model is easier to obtain long-term cumulative reward.

The above equation shows the recursive relationship between the current state value function $V(s)$ and the future state value function $V(s_{t+1})$. Obviously the above is more concerned with the state, but to know the expected return obtained by using a policy π using action a under state s , there is a Q function:

$$Q(s, a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V(s') \quad (2)$$

where $P_{ss'}^a$ is the state transition probability of the current state s moving from action a to state s' .

$$V(s) = \sum_{a \in A} \pi(a|s) Q(s, a) \quad (3)$$

So the optimal Q function under the optimal strategy is:

$$Q^*(s, a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \max_{a'} Q^*(s', a') \quad (4)$$

In this task, our state transition probability $P_{ss'}^a$ is a constant value, which can be understood as a sampling method, and the following formula can be obtained:

$$\begin{aligned} Q(s, a) &\approx r_s^a + \gamma \max_{a'} Q(s', a') \\ &\approx (1 - \alpha) Q(s, a) + \alpha \left[r_s^a + \gamma \max_{a'} Q(s', a') \right] \end{aligned} \quad (5)$$

after processing, the above equation uses the exponential weighted average of α , and the above equation is the Bellman optimization equation.

In the fault diagnosis task, the number of state space set S is very large, and the limited space cannot effectively store the Q value of the state action pair. The strong function generalization ability is a major feature of the deep learning network, which can be used to fit the Q function, denoted as $Q(s, a; \theta)$, where θ is the unknown parameter of neural network. The Q function of the current state s when the probability P is transferred to s' after the execution of action a is:

$$y = r + \gamma \max_{a'} Q(s', a'; \theta) \quad (6)$$

where r denotes the reward received after state s executes action a .

According to the ε -greedy strategy, the action with the maximum current Q value is selected according to the probability ε , and the action is randomly selected according to the probability $1 - \varepsilon$ to obtain $Q(s, a)$. Therefore, the loss function is:

$$L = (y - Q(s, a; \theta))^2 \quad (7)$$

The above formula can make the Q function converge to the optimal Q value function under the minimum mean square error according to the ε -greedy strategy.

2.3 ε -Greedy Strategy

In reinforcement learning tasks, the exploration of the external environment is the main way to acquire knowledge, so the balance between exploration and utilization has become a problem to be solved in reinforcement learning. Too much exploration will affect the convergence speed of the model, and too much use will make the reinforcement learning model io local optimum. The commonly used ε -greed strategies are mostly fixed values. This paper uses a method that changes with the number of training steps i . Make the model in the early exploration and later use. The setting method of ε is:

$$\varepsilon = \max \left\{ \varepsilon_{\min}, 1 - \frac{1 - \varepsilon_{\min} \times i}{T} \right\} \quad (8)$$

where i is the current step number, T is the total step number.

3 The Proposed Method

3.1 Stochastic Pooling

Stochastic pooling is a recently developed technique to overcome the constraints of maximum and average pools [61]. In the random pool, each element is sampled according to its contribution, and the sampling probability can be calculated as follows:

$$p_{i,j} = \frac{a_{i,j}}{\sum_{(i,j) \in R_s} a_{i,j}} \quad (9)$$

In each pooling area, the composite elements of the output feature map can be weighted in an average probability form, defined as:

$$f_{i,j} = \sum_{(i,j) \in R_s} (p_{i,j} \cdot a_{i,j}) \quad (10)$$

By Eq. (9), it can be seen that stochastic pooling considers all the feature information and pays more attention to those strong features.

3.2 Exponential Linear Unit (ELU)

Selecting the appropriate activation function is of great significance for the design of neural network model. Both Sigmoid and Tanh have disadvantages of high computational cost and vanishing gradient. ReLU can well solve these two problems and become a priority for many deep learning models. However, when the input is negative, the ReLU neurons completely stop learning. Sigm and ReLU are currently the most widely used activation functions in CNNs. Sigm has the problem of gradient disappearance. When the input is negative, ReLU cannot continue training. Exponential linear unit (ELU) can effectively solve the problems existing in Sigm and ReLU, and its superiority is proved in some benchmark data. ELU is defined as:

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha(e^x - 1), & \text{if } x \leq 0 \end{cases} \quad (11)$$

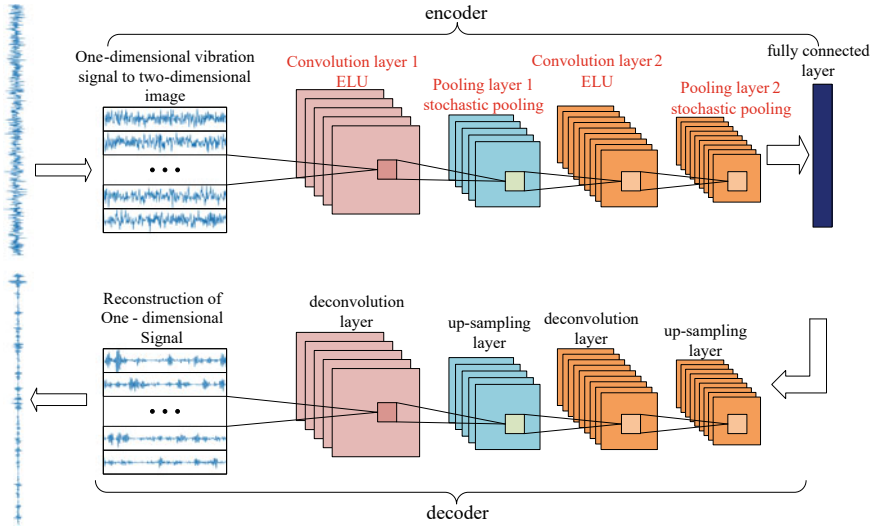


Fig. 1 Enhanced convolution self-encoder

3.3 Enhanced Deep Convolution Self-encoder

As shown in Fig. 1, the enhanced convolution autoencoder first converts one-dimensional vibration signal into two-dimensional gray image signal by using the signal stacking method [10], and the two-dimensional signal output by the encoder and decoder constitutes one-dimensional vibration signal. Among them, in order to better extract fault features and suppress noise, this paper uses new pooling methods and activation function random pooling and exponential linear unit (ELU) to replace the original pooling and activation methods in the encoder, so that it can more effectively extract features and suppress noise signals.

3.4 Fault Diagnosis Method of Enhanced Convolution Self-encoder and Reinforcement Learning

The fault diagnosis process of reinforcement learning based on small label data is shown in Fig. 2. The training process of the model is divided into two steps. The first step is to input a large number of unlabeled data into the enhanced convolutional autoencoder to reduce noise and dimension of the data and extract key fault features, and save the relevant weights of the encoder W, b ; In the second step, the encoder weights W and b are used as the initial weight parameters of reinforcement learning to fine-tune the network.

The specific process is as follows:

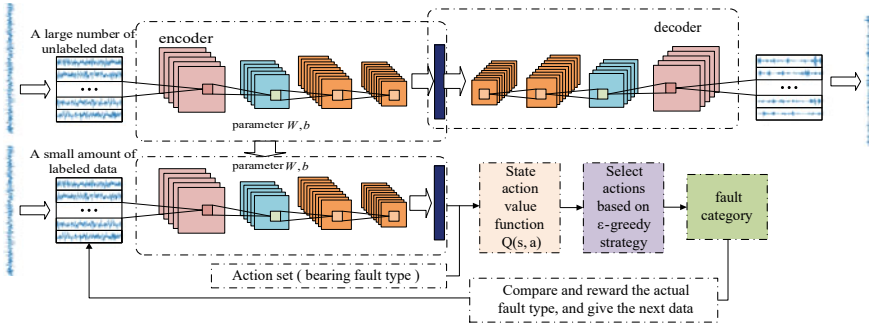


Fig. 2 Fault diagnosis model for enhanced convolutional autoencoder and reinforcement learning

- Step 1:** The collected fault test bench data are divided into training samples and test samples. Most of the training samples are unlabeled and the rest are labeled.
- Step 2:** The activation function and pooling mode of the original convolutional autoencoder network are replaced by ELU activation function and stochastic pooling.
- Step 3:** The label-free training samples are used to pre-train the enhanced convolutional self-encoder to obtain W and b .
- Step 4:** Initialization of reinforcement learning parameters with weights W, b and other parameters in trained encoders.
- Step 5:** The reinforcement learning network is trained by using fewer label samples to further adjust the weights and deviations.
- Step 6:** The test samples are used to test the diagnostic performance of the proposed method.

4 Experiment and Verification

In order to verify the effectiveness of the proposed method, the open source bearing data set of the University of Western Reserve in the United States was used for verification [11]. The data used are the data of the drive end, and the data contain four fault states: normal bearing, roller fault, inner ring fault and outer ring fault. Each fault state contains three fault degrees of 0.07, 0.14 and 0.21 mm, respectively. Therefore, the data set contains a total of 10 fault types, and each fault type contains four load data (1, 2, 3, 4hp). The corresponding rotational speeds are 1797, 1772, 1750 and 1730 r / min, respectively. It can be roughly considered as a constant speed, and the sampling frequency of vibration data is 12 kHz. In 10 fault types, each fault type contains 940 samples, a total of 9400 samples. Of these, 940 are test samples and others are training samples.

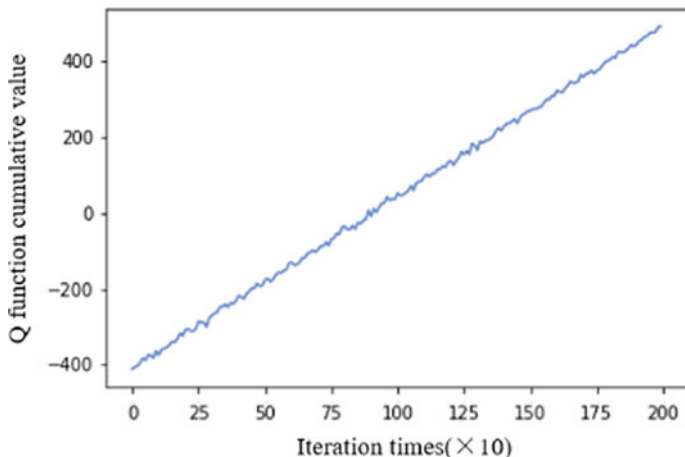


Fig. 3 Q function cumulative value change graph

4.1 Feasibility Validation of Reinforcement Learning

Using convolutional neural network [12] to fit the Q-value function, the network parameters are as follows:

The training parameters of the model are as follows: the number of training steps per round $M = 512$, the number of iterations 2000, the learning rate 0.5, and the probability $\varepsilon = \max\{\varepsilon_{\min}, 1 - \frac{1-\varepsilon_{\min} \times i}{1000}\}$, where i is the number of training steps and the fault samples are discrete, so the discount coefficient γ is 0.

In each round of training, the number of steps set in each round of this model is 512. Since the model does not learn anything at the beginning, there is a 1/10 probability to guess at the beginning, that is, the score of the first round should be between -512 and -456 points. The model has trained 2000 rounds, and the average value of reward and loss is calculated and plotted for each 10 rounds of 2000 rounds of training. Figure 3 shows the relationship between the cumulative value of Q function after model training and the number of iterations. It can be observed that the model is almost linearly increasing, and the final highest score reaches 488 points. The recognition accuracy of the model can reach 98.87%. Figure 4 shows that the loss value of the model after each round of training can be maintained at a relatively low state, indicating that the model has learned something in each round of training. The above experiments verify the feasibility of this method in fault diagnosis.

4.2 Verify the Effectiveness of Stochastic Pooling and ELU

In order to verify the enhancement effect of stochastic pooling and ELU in convolutional self-encoder, the max pooling + ELU, mean pooling + ELU, stochastic

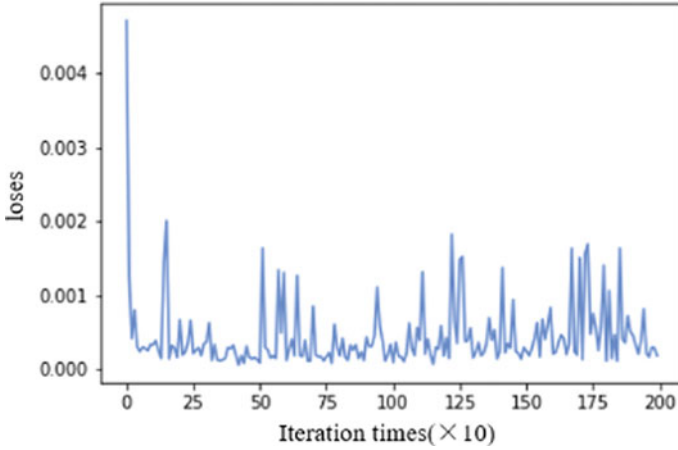


Fig. 4 Change of loss value

pooling + ELU and random pooling + rectifier linear function (ReLU) are used for verification. In order to better verify the enhancement effect of stochastic pooling and ELU, different noise levels are added to the original vibration signal. The probability distribution of noise conforms to the probability distribution of mean 0 and variance 1. The noise level is 10, 30, 50, 60, 70, 80, 85, 90, 93, and 95%. The parameters of each layer are shown in Table 1, where the enhanced convolutional autoencoder network learning rate is 0.01 and the α in ELU is 0.05.

In the experiment, the fixed activation function is used ELU, and the influence of different pooling methods on the reconstruction error of the encoder is compared. The results are shown in Fig. 5a. It can be seen from the figure that the reconstruction error of different pooling methods will increase with the increase of noise. However, the stochastic pooling used in this paper has better performance and lower reconstruction

Table 1 Fitting parameters of Q function network

Network layer	Description of each layer
Input data size	1024×1
Reshape	32×32
First layer convolution kernel size	$(5 \times 5) \times 15$
Activation functions of the first and third layers	ELU
Second layer stochastic pooling	$(2 \times 2) \times 15$
The third layer convolution kernel size	$(5 \times 5) \times 20$
The fourth layer stochastic pooling	$(2 \times 2) \times 20$
Fully connected layer	256

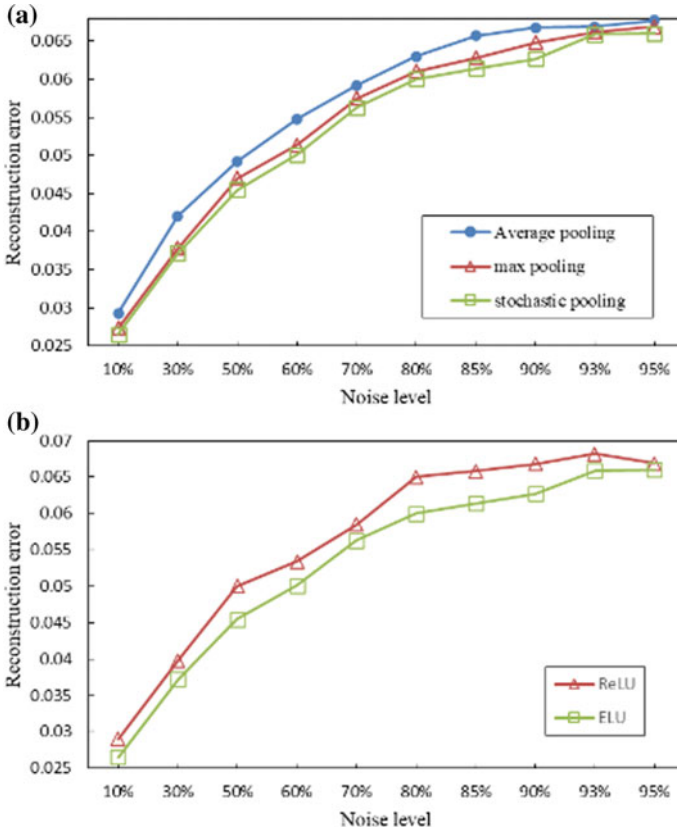


Fig. 5 **a** Influence of pooling method on reconstruction error and **b** Influence of activation function on reconstruction error

error. This is because the stochastic pooling not only considers the largest feature element in the signal, but also takes into account all the feature elements equally.

Similarly, in order to verify the enhancement effect of the activation function, stochastic pooling is used in the fixed pooling mode in the experiment, and the influence of different activation functions on the reconstruction error of the encoder is compared. The results are shown in Fig. 5b. It can be seen from the figure that the reconstruction error of different activation functions increases with the increase of noise, but the ELU used in this paper has better performance and lower reconstruction error.

The waveform comparison between the reconstructed output and the input signal is shown in Fig. 5. It can be seen that there is a relatively distinct difference between the reconstructed signals, indicating that the key characteristics of the signal reconstructed by the convolution self-encoder of the input signal become more significant, and the noise effectively eliminated. This shows that the enhanced convolutional

self-encoder has the ability to filter out the original signal noise, so as to effectively extract the key fault features (Fig. 6).

In order to avoid the contingency brought by network training, each experiment was repeated 10 times. The test accuracy of each experiment was shown in Fig. 7. It can be seen from Fig. 7 that the enhanced convolutional self-coded neural network can improve the average recognition rate of fault, and the average recognition accuracy can reach more than 97%, which is higher than that of other pooling methods and activation methods.

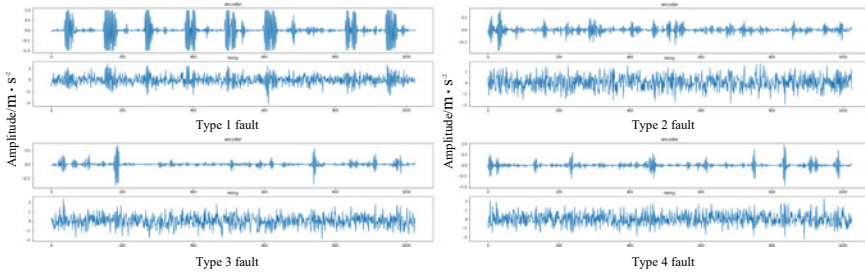


Fig. 6 Comparison between reconstructed signal and input signal of several types of data

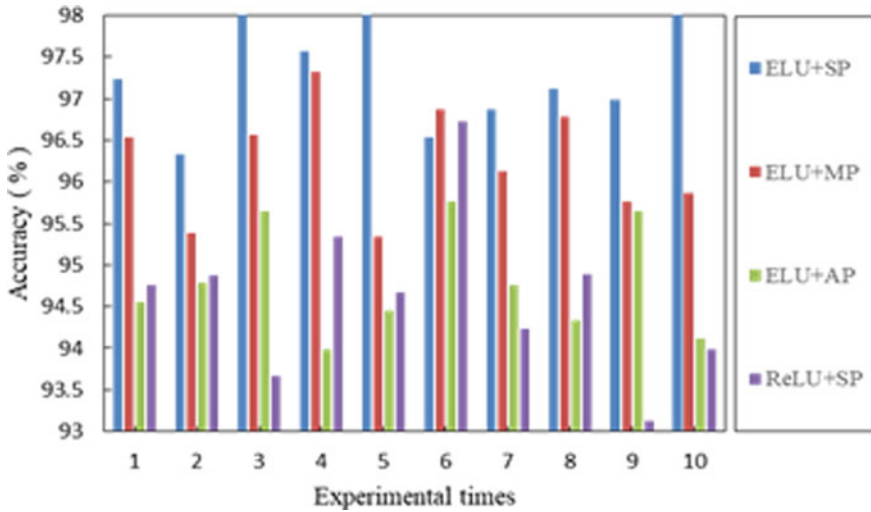


Fig. 7 Diagnostic accuracy of each experiment

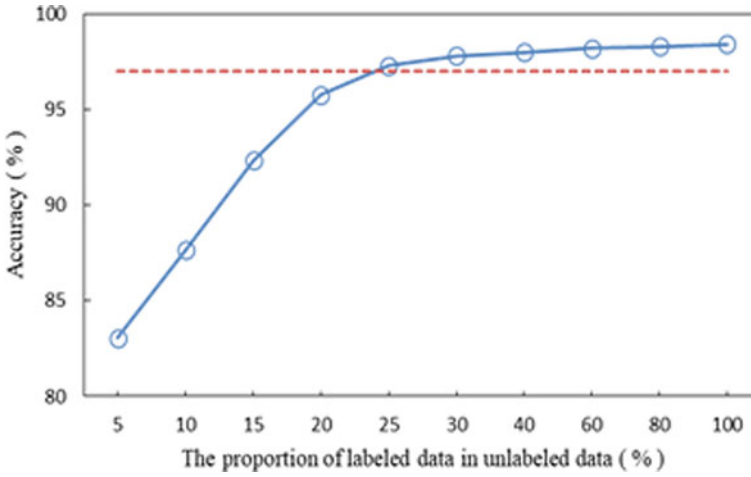


Fig. 8 Proportional diagnostic effect of labeled data and unlabeled data

4.3 Discussion on the Number of Label Samples

Compared with labeled samples, making a large number of unlabeled samples will be easier and faster. Therefore, it is necessary to explore the impact of the relationship between the number of labeled samples and the number of unlabeled samples on the recognition accuracy.

Here, the proportion of labeled samples in the number of unlabeled samples is 5%, 10%, 15%, 20%, 25%, 30%, 40%, 60%, 80% and 100%, respectively. It should be noted that the label-free samples for the training of the enhanced convolution self-encoder are always a fixed value, and only the number of label-free samples changes. The diagnostic results of the proposed method in different proportions are shown in Fig. 8. It can be seen that the average accuracy rate increases with the increase of the number of labeled samples. When labeled samples account for 25% of unlabeled samples, the accuracy rate can reach more than 97%.

5 Conclusion

This method effectively combines the enhanced convolution self-encoder with reinforcement learning, and can adaptively and unsupervised learn the abstract characteristics of the vibration signal of the sensor, so as to effectively diagnose the gearbox fault. The main conclusions are as follows. (1) The convolution self-encoder is enhanced by stochastic pooling and exponential linear unit (ELU), so that it has better noise reduction and feature extraction effect. (2) Compared with other algorithms, this method does not require a lot of label data, and has more practical value.

(3) The relationship between the number of labeled samples and unlabeled samples is discussed. Experiments show that the recognition effect can reach 97% when labeled samples account for 25% of the number of unlabeled samples.

Acknowledgements The authors are grateful to the support from the National Natural Science Foundation of China (No. 51865045).

References

1. Liu, G., Zhang, G.: Study on genetic neural network of the fault diagnosis based on wavelet packet. *J. Mech. Eng.* **36**(9), 110–112 (2000)
2. Li, W., Zhang, S.: Fault classification based on improved evidence theory and multiple neural network fusion. *J. Mech. Eng.* **46**(9), 93–99 (2010)
3. Sun, W., Shao, S., Yan, R.: Induction motor fault diagnosis based on deep neural network of sparse auto-encoder. *J. Mech. Eng.* **52**(09), 65–71 (2016)
4. He, Z., Shao, H., Zhong, X., Yang, Y., Cheng, J.: Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions. *Knowl.-Based Syst.* **207** (2020)
5. He, Z., Shao, H., Zhong, X., Yang, Y., Cheng, J.: An intelligent fault diagnosis method for rotor-bearing system using small labeled infrared thermal images and enhanced CNN transferred from CAE. *Adv. Eng. Informatics* **46** (2020)
6. He, Z., Shao, H., Zhong, X., Yang, Y., Cheng, J.: Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder. *Measurement* **152** (2020)
7. Zhou, X., Yu, J.: Gearbox fault diagnosis based on one-dimension residual convolutional auto-encoder. *J. Mech. Eng.* **56**(07), 96–108 (2020)
8. Mnih, V., Kavukcuoglu, K., Silver, D., et al.: Human-level control through deep reinforcement learning. *Nature* **518**(7540), 529–541 (2015)
9. Zhou, Z.: *Machine learning*, pp. 380–381. Tsinghua University Press, Beijing (2016)
10. Wen, L., Li, X., Gao, L., et al.: A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Industr. Electron.* **65**(7), 5990–5998 (2018)
11. Loux, L.: Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech. Syst. Signal Process.* **18**(5), 1077 (2004). [https://doi.org/10.1016/S0888-3270\(03\)00077-3](https://doi.org/10.1016/S0888-3270(03)00077-3)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Proceedings of International Conference on Neural Information Processing Systems*. Spain: Neural Information Processing Systems Foundation, pp. 1097–1105 (2012)

Optimization Design and Simulation Analysis of Miniature Boring Machine Based on ADAMS



Zhou Yue, Cao Yu, Lu Zhen-hua, Wei Qi-wen, Zhao Xue-mei,
Wang Ye-zhen, Sun Jia-xing, Liu Ying, and Zhong Shan

Abstract In order to improve the processing precision and surface quality of line-drawing and boring in ship manufacturing process, a virtual prototype of micro-boring machine was built on the basis of existing portable boring machine, aiming at the inconvenient boring operation and low processing precision of small and medium-sized ships at present. Then it carries on the optimization design to its transmission mechanism, uses ADAMS to carry on the kinematic simulation analysis to the virtual prototype; Finally, reasonable simulation data are obtained to ensure the performance, accuracy and efficiency of boring, and provide basis for the subsequent optimization design.

Keywords The ship · Micro boring machine · Three-dimensional modeling · ADAMS · Motion simulation

In 2019, the number of civil motor vehicles in China reached 121,440, and the total tonnage of ship's is increasing. At the same time, the shipbuilding industry is also developing [1]. The Xiji-ang River is an important part of the Pearl River system, known as the "golden waterway" and is now gradually becoming a gold corridor that actively moves closer to advanced productivity and undertakes the transfer of eastern industries [2], The construction of new land and sea corridors in the western region, the implementation of national strategies such as the Guangdong-Hong Kong-Macao Greater Bay Area, and the improvement of grade and the formation of the Pan-Pearl River Delta and Guangxi Beibu Gulf Economic Zone have made the rapid growth of the national economy and great demand for ships. Therefore, Guangxi shipbuilding industry has obtained good market opportunities necessary for development. With the rapid development of shipbuilding industry, the processing quality requirements for each process in the shipbuilding process are also increasing, and the upgrading of processing equipment is imminent [3]. In the shipbuilding process stage, the

Z. Yue · C. Yu (✉) · L. Zhen-hua · W. Qi-wen · Z. Xue-mei · W. Ye-zhen · S. Jia-xing · L. Ying · Z. Shan

School of Mechanical and Resource Engineering, Wuzhou University, Wuzhou 543002, China
e-mail: connor521@qq.com

accuracy of the wire drawing boring is very strict. The shafting installation is affected by many factors such as wire drawing, boring and so on. If the accuracy of the wire drawing boring is not too close, it will affect the speed of the ship's shafting installation and become the bottleneck of shipbuilding speed. The boring accuracy of the new ship shafting is required to be higher and higher [4], especially the stern shaft, which is an important part of the ship shafting. The working condition of ship stern shaft is equivalent to the safety and stability of ship operation [5]. Domestic small and medium-sized ship shafting mainly adopts the process of clamp alignment and manual hinged stroke; The large-scale ship shafting basically adopts the clamping site alignment, and then the f-flange hole is refined by the special key machine. This process can meet the technical requirements of the installation, but because the size of the flange hole is not interchangeable, the smoothness is not ideal; In addition, the high labor intensity and long processing cycle have a great impact on the integrity of the platform and the cycle of shipbuilding [6], which cannot meet the production needs of the shipbuilding industry [7]. Therefore, a miniature boring machine for ships with small size, light weight, simple operation, low cost and high working efficiency is developed and improved based on the actual problems of hole processing in china Shipbuilding Guijiang company ltd.

Aiming at the problems of low machining accuracy, large volume and inconvenient operation of traditional boring machine, some problems are solved by replacing parts. SolidWorks software is used to complete the virtual, and ADAMS software is used to simulate and analyze it. The feasibility of micro boring machine is tested and relevant data are obtained.

1 Design of Micro Boring Machine

1.1 Working Principle

The mechanism of the micro boring machine is mainly composed of a power head, a motor, a handle, a screw rod, a belt, a PLC (Programmable Logic Controller) control part, a support frame, and a fixture. Place the boring machine at the designated position and start it with power. Adjust the horizontal position of the boring cutter through the handle to move the machining center to the position to be processed; set the motor speed, spindle feed speed, feed rate and other values; servo motor and step The input motor is energized to run, respectively driving the main shaft (through a belt drive) and the screw rod to rotate. The precision planetary gear connects the main shaft and the boring bar. The boring bar is the benchmark for clamping the tool and transmits power to the tool [8]; The reciprocating motion and the boring tool rotates to realize the boring action.

1.2 Optimized Design

1.2.1 Optimal Choice of Prime Mover

The motors of traditional boring machines are mostly three-phase asynchronous AC motors plus frequency converters, and it is easy to malfunction due to unreasonable environment settings such as humidity [9]. In addition, the additional copper loss caused by the skin effect will cause the motor to generate additional heat, resulting in reduced efficiency and reduced output power [10]; and its mechanical characteristics are relatively hard, occupying a certain position in volume, and when running there may also be unstable vibrations. For this, a motor suitable for use on a ship should be used, and it can make up for the lack of mechanical rigidity and improve work efficiency. At the same time, the overall volume of the machine is reduced, the overload capacity is enhanced, and the working performance is more safe and reliable. It can be seen that the servo motor can do these tasks. Servo motors not only have the characteristics of small electromagnet time constant, high linearity, low starting voltage, etc. but also small size, fast action response, large overload capacity, wide speed adjustment range, large low-speed torque, small fluctuations, stable operation, and low noise, High efficiency and other advantages [11]. The motor obtains a lower processing speed through the frequency converter, and then combines with a planetary gear reducer with a medium transmission ratio. The combination of the three not only increases the output torque, but also meets the processing requirements under the premise of precise positioning of the servo motor. Adjust the speed between 100 ~ 1000 r/min.

1.2.2 Optimal Selection of Transmission Mechanism

In order to ensure high-precision output when the micro boring machine is working, the Mohs 2 power head is selected as the boring output power head, and it can be assembled and disassembled freely in assembly. When boring on a ship, the working environment is mostly at sea, and the impact of irregular waves will cause the ship to shake. In order to ensure the machining accuracy of the boring under the shaking environment, the ball screw is selected as the feeding mechanism. The transmission efficiency of ordinary sliding screws is 0.25–0.50, while the transmission efficiency of ball screws can reach 0.90. When driving the same large load, using the latter can obtain higher transmission efficiency, low energy consumption, high precision and smooth movement, ensuring the reliability and stability in the feeding transmission process, and realizing high-speed feed and Micro feed motion [12].

1.2.3 Optimized Design of Control System

Ship shaking is inevitable when working on the water. The system controlled by the boring machine needs to control the operation efficiency of the motor [13]. Here, the servo control system is used to accurately control the mechanical position, torque, and speed [14]. And the system has a frequency analysis function, which can detect the resonance point of the machine, which is convenient for system adjustment. While achieving high-precision output, it can also ensure the stability of the work and high-efficiency responsiveness. In addition, coupled with PLC to achieve closed-loop control, Using the advantages of PLC, such as strong anti-interference ability, low cost, fast speed, strong function and small size, the boring machine is easy to operate, can be controlled, save manpower and time [15].

2 The Mechanical Principle of the Micro Boring Machine

The boring machine is mainly composed of a feed mechanism and a transmission mechanism. Through the hand wheel control screw transmission mechanism. Then the plane guide rail is moved to the position of the machining center that is ready for positioning; Subsequently, the servo control system starts and controls the motor 1 and 3. The servo motor 3 transmits power through the pulley to drive the boring cutter to rotate for preparation. The feed mechanism controls the ball screw transmission mechanism through stepping motor 1, and drives the tool head through the feed slider to move it down to the displacement point and reset, so as to realize the boring movement. The specific plan layout of this device is shown in Fig. 1.

3 Adams Dynamic Simulation Analysis

The prototype model established by ADAMS kinematic analysis software is as follows. Through kinematic simulation analysis, the maximum machining stroke of micro boring machine is analyzed (Fig. 2).

Taking the micro boring machine model that has completed the three-dimensional modeling as the research object, the kinematic simulation analysis is carried out in ADAMS. The maximum stroke analysis of X, Y and Z axes is carried out, and the velocity and acceleration of Y axis are analyzed. The simulation time is set to 22 s, and steps are 500 steps.

The function of the maximum moving stroke of boring bar in X axis direction is:

$$step(time, 0, 0, 5, 60) + step(time, 5, 0, 10, -70)$$

The function of the maximum moving stroke of boring bar in Z axis direction is:

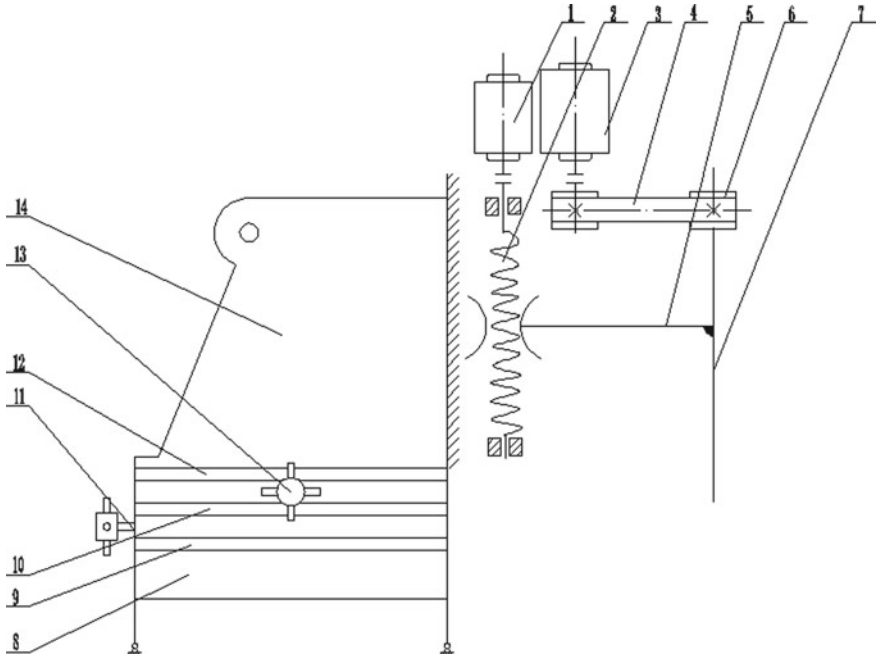


Fig. 1 Structure of micro boring machine. Drawing number identification: 1. Stepper motor; 2. Ball screw; 3. Servo motor; 4. Leather strap; 5. Fast feed; 6. Belt pulley; 7. Boring rod; 8. Base; 9. Bottom guide rail; 10. Middle guide rail; 11. Screw rod; 12. Upper guide rail; 13. Hand wheel; 14. Frame

$$step(time, 0, 0, 5, -22) + step(time, 5, 0, 10, 30)$$

The function of maximum moving stroke of boring bar in Y axis direction is:

$$step(time, 0, 0, 5, -5) + step(time, 5, 0, 20, 80)$$

The resulting data is exported as shown in Fig. 3.

From Fig. 3 and the function, it can be seen that the X, Y, Z three axis curve in the fifth second to one side of the maximum stroke, X, Z axis in 5-10 s to the other side of the maximum stroke; The spindle Y axis travels to the other side in 5-20 s. The maximum travel value of X axis is 130 mm, the maximum travel value of Z axis is 52 mm, and the maximum travel value of the spindle (Y axis) is 85 mm. The travel value obtained by the three-axis motion simulation in ADAMS is compared with the displacement range value of the plane guide rail given by the three-dimensional modeling data, which is consistent with the design value.

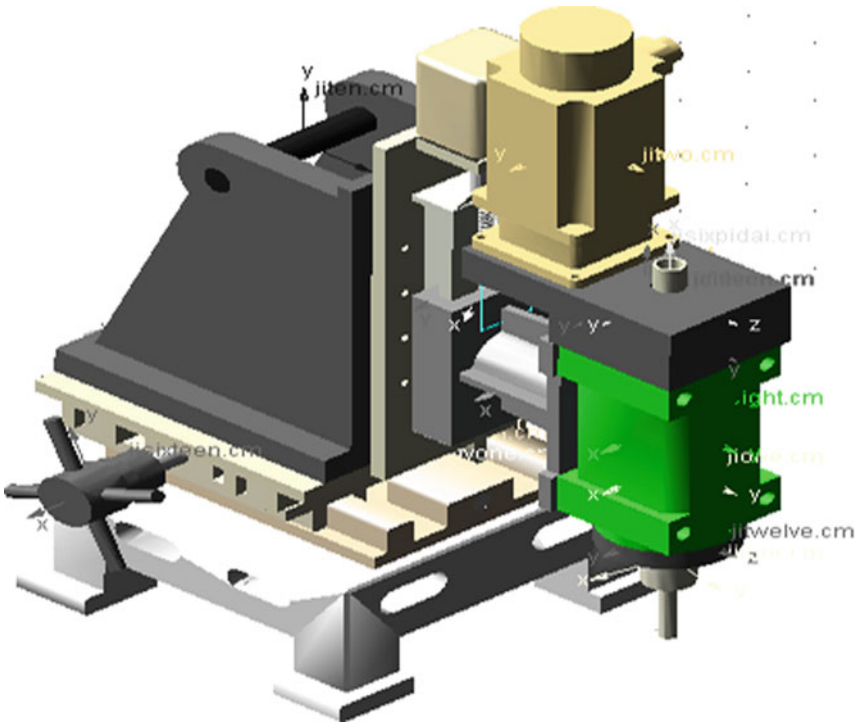


Fig. 2 ADAMS micro boring machine prototype model

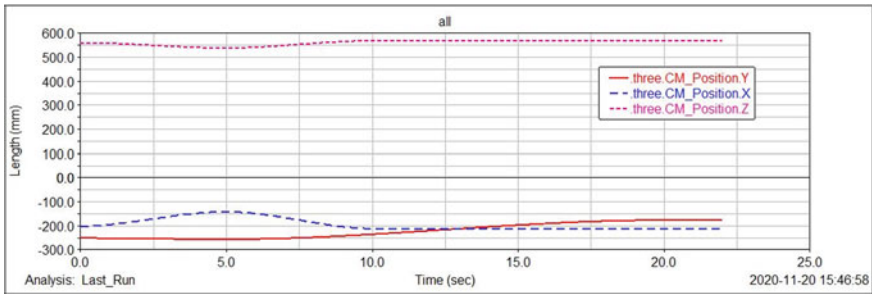


Fig. 3 Triaxial displacement travel characteristic curves of X, Y and Z axes

As shown in Fig. 4, the Y-axis curve in 0–22 s to go through a maximum stroke of the Y-axis at the same time the boring cutter uniformly accelerated motion, acceleration unchanged. Transmission is relatively stable in the ideal state, so there will be no speed curve oscillation phenomenon.

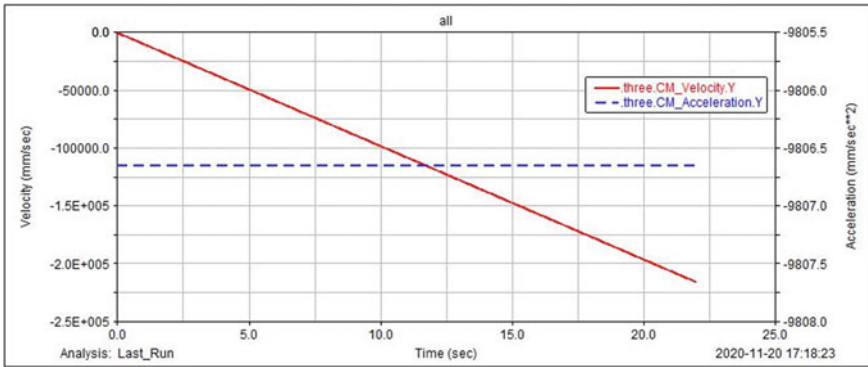


Fig. 4 Y-axis velocity and acceleration characteristic curve

4 Conclusion

Basde on modeling software SolidWorks and simulation software ADAMS, the whole process of research direction, design optimization, modeling and simulation of micro-boring machines is discussed, including the design of boring machine mechanism, 3D modeling process, ADAMS motion simulation analysis and other concrete steps. Using SolidWorks to build a three-dimensional model of each part of the micro boring machine and assemble it, using the parasolid.x_t file format to import the assembled sample geometry in SolidWorks into ADAMS and add the constraints, etc., the micro boring machine sample is finally formed.

Using ADAMS to carry out motion simulation analysis of the boring machine mechanism, the total stroke of the spindle (Y-axis) is consistent with the given design value, and the maximum displacement value of the spindle is 85 mm, i.e. the maximum boring depth of the bar is 85 mm, all of which are consistent with the initial design value. Through the simulation analysis of the displacement of X-axis and Z-axis, the travel range is 130 mm and 53 mm respectively. The rationality of the miniature boring machine is analyzed in detail through various performances, which provides a certain theoretical basis for the optimization design of the various parts of the miniature boring machine.

Fund Project College Student’s Innovation and Entrepreneurship Training Project (201911354136).

References

1. Song-zhen, L.I.: Study on the layout of railway-water intermodal transport of xijiang golden waterway. *China Railway* **10**, 32–35 (2013)

2. Editorial Department of the Journal. Xijiang Waterway under the Background of ‘ The Belt and Road ‘—Records of Guangxi Golden Waterway Construction. *Pearl River Waterway* **16**, 15–17 (2017)
3. Lin, W.: Research on the Development of Shipbuilding Industry in the Four ASEAN Countries. Xiamen University, Xiamen (2018)
4. Xue-shan, W.: Improvement of shafting boring process in shipbuilding. *Scientif. Technol. Innov.* **04**, 35–37 (2018)
5. Xi-chuan, W.: Application of high-speed arc spraying in ship stern shaft repair. *Scientif. Technol. Prospect* **26**(08), 133 (2016)
6. Fei-ming, D.U.: Portable Marine Boring Machine Design Based on Numerical Control Technology. Zhejiang Ocean University, Zhoushan (2013)
7. Cheng-gang, S.: Production planning and control of China ‘ s shipbuilding industry. *Sci. Technol. Wind* (**04**), 117–118 (2020)
8. Cheng, W., Shi-min, Z.: Research on boring technology of ship site. *Scientif. Technol. Innov. Appl.* **25**, 47 (2013)
9. Cong-yu, Y., Yi-bing, S.: Analysis of common faults and maintenance countermeasures of inverter used in three-phase asynchronous AC motor. *Sci. Technol. Innov. Report* **23**, 65 (2011)
10. Hui, D.: Inverter selection method and application. *Electron. world* **12**, 200–201 (2020)
11. Fei-ming, D., Heng, C., Yu-lian, Z.: Control system design of HC-600 portable marine CNC boring machine. *Mechan. Design Manuf. Eng.* **42**(06), 53–56 (2013)
12. Guo-ying, L.: The selection advantages and installation and use precautions of ball screw in CNC lathe modification. *Heilongjiang Sci. Technol. Inform.* **12**, 70 (2012)
13. Jun-hao, H.: Comprehensive comparison of stepper motor and AC servo motor performance. *Commun. World* **05**, 184–185 (2019)
14. Hong-sen, Y., Wei-ren, C.: On the output motion characteristics of variable input speed servo-controlled slider-crank mechanisms. *Mech. Mach. Theory* **35**(4), 541–561 (2000)
15. Ru-wan, W.: Application of PLC technology in automatic control of electrical engineering. *Managem. Technol. Small and Medium-Sized Enterprises (late issue)* **09**, 166–167 (2020)

Zhou Yue (1998.12), female, Zhaoqing, Guangdong Province, Wuzhou University, school of mechanical and tesource engineering, post: student, positional title: undergraduate, mainly engaged in machinery, Wuzhou College of Mechanical Engineering, No. 82 Fumin Sanlu, Wanxiu District, Wuzhou City, Guangxi, 543002, 13318115456.

Cao Yu (1986.3), male, Jilin Tonghuaren, Wuzhou University, school of mechanical and tesource engineering, position: dean of teaching-researching section, positional title: Senior engineer, master of materials science, mainly engaged in materials science and education research. Wuzhou University, No. 82 Fumin Third Road, Wanxiu District, Wuzhou City, Guangxi, College of Mechanical and Chemical Engineering, 543002, 17758656917.

The Stability and Vibration Characteristic Optimization of the Pressure Shell of a Buoyancy Regulator of an Underwater Vehicle



Yonghui Cao, Chiye Yang, Jing Liu, Yu Xie, Shumin Ma, and Yong Cao

Abstract Underwater vehicles (UV) with a deeper operation ability are the important research field in the marine industry. In order to obtain a better and safer operation performance, the strength, stability and vibration characteristics of the pressure shell of UV should be analyzed. In this paper, a finite element model of the pressure shell of buoyancy regulator is developed. The influences of the stiffener thickness, width, position, and shell thickness on the strength, deformation, and load factor of the pressure shell are studied. In addition, a lighter and safer shell structure is obtained by using the response surface optimization method. The simulation results show that the above factors have great influence on the shell characteristics, such as strength, stability and modal parameters. Moreover, a lighter pressure shell used in the UV can be helpful for providing a better possibility to carry more equipment.

Keywords Underwater vehicle · Pressure shell · Stability · Vibration characteristic

Y. Cao · C. Yang · J. Liu · Y. Xie · S. Ma · Y. Cao (✉)

School of Marine Science and Technology, Northwestern Polytechnical University, Xian 710072, China

e-mail: cao_yong@nwpu.edu.cn

Y. Cao

e-mail: caoyonghui@nwpu.edu.cn

C. Yang

e-mail: yangchiye@mail.nwpu.edu.cn

J. Liu

e-mail: jliu0922@nwpu.edu.cn

Y. Xie

e-mail: xieyu@mail.nwpu.edu.cn

S. Ma

e-mail: msm@mail.nwpu.edu.cn

1 Introduction

The piston-type buoyancy regulator is generally used in large-scale UVs such as gliders, which is the key device for adjusting buoyancy. As shown in Fig. 1, when the piston moves outwards, the drainage volume of the buoyancy regulator increases, and the buoyancy provided increases accordingly; on the contrary, the buoyancy provided decreases. When the piston moves to the outermost side, the compartment where the piston cylinder and the hydraulic block are located; and it will generate a larger cavity. Due to the effect of water pressure, this is the weakest status of the whole device. Since the inside is a cavity and the outside are under the water pressure, this section of the cavity can be regarded as a pressure shell.

In 2014, the pressure cabin of the “Nereus” submarine developed by the Woods Hole Institute of Oceanography in the United States [1] ruptured due to the insufficient strength, which caused the submarine to be lost in the Kermadec Trench in northeastern New Zealand. Therefore, it is necessary to calculate the strength of the pressure shell. In addition, it can be known from the critical length formula that this pressure-resistant shell is a thin wall short cylinder at the external pressure, and buckling may occur before the strength yields [2]. Therefore, it is also necessary to conduct the stability analysis. Finally, the moving parts such as the motors inside the device may cause the casing to vibrate during operation. In order to ensure the normal operation of the device, a better dynamic performance is also a required

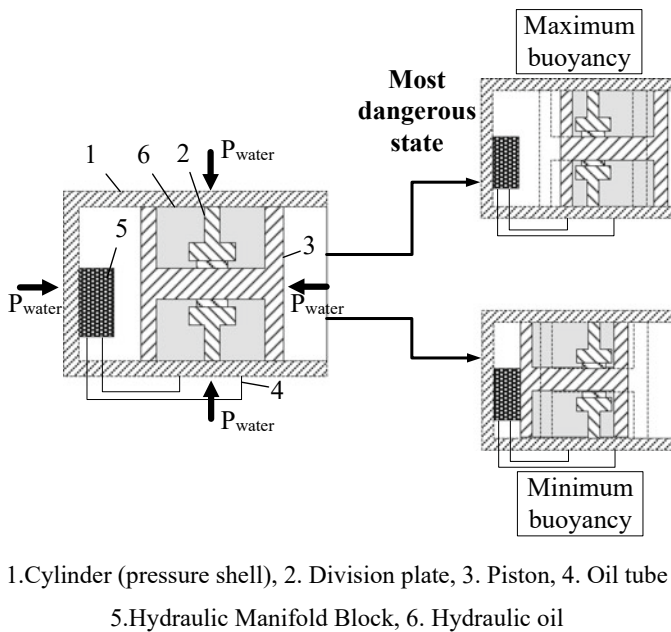


Fig. 1 Working principle of piston-type buoyancy regulator

feature. Vasilikis Daniel and Karamanos Spyros A examined the mechanical response of thin-walled cylinders surrounded by a rigid or deformable medium, subjected to uniform external pressure [3]. Rouzbeh Hashemian and Magdi Mohareb developed a general eigenvalue buckling solution for the buckling analysis of sandwich pipes with thick cores subjected to internal and external hydrostatic pressure [4]. Razakamiadana and Zidi studied buckling and postbuckling of concentric cylindrical tubes under external pressure by finite element method [5]. Isvandzibaei et al. presented the energy method for the vibration of thin-walled homogeneous isotropic and manifold layered isotropic cylindrical shells under uniform external lateral pressure [6]. Liu et al. presented an analytical procedure and closed-form vibration solutions with analytically determined coefficients for orthotropic circular cylindrical shells having classical boundary conditions [7]. This paper presents a finite element analysis to simulate the shell to study the strength, stability and vibration characteristics of the shell.

The organization structure of this article is as follows: Sect. 2 introduces the finite element model. Section 3 introduces the simulation results. The lightweight optimization of the shell is completed by the response surface optimization method. Finally, the modal analysis of the shell before and after the optimization verifies that the dynamic performance of the optimized shell is also better than that before the optimization. A conclusion and suggestions for future work will be presented in Sect. 4.

2 Finite Element Model

Aiming at the most dangerous state of the buoyancy adjusting device as the calculation condition, the external pressure shell composed of the inner side of the piston cylinder and the compartment where the hydraulic block is located is taken as the calculation object, and the work is simulated at a water depth of 2000 m. The original model can be simplified to a cylinder with a wall thickness of 12.5 mm as shown in Fig. 2. The wall thickness of the piston cylinder is H ; the number of stiffeners is n ; the height of the stiffener is h ; and the width of the stiffener is w .

For the cylindrical model, Shell181 has higher calculation efficiency and accuracy. Therefore, the calculations in this paper use Shell181 to construct the conceptual model. A quadrilateral dominant mesh is used, the mesh size is 3.0 mm; and the number of mesh is 57,749. Since the rigidity of the piston and the end cap is much greater than that of the cylinder; two ends can be regarded as the fixed constraints; and a pressure of 20 Mpa is applied to the outer wall of the cylinder to simulate a water depth of 2000 m. The mesh division and loading conditions are shown with six stiffeners as shown in Fig. 3.

The response surface optimization method is used to optimize the pressure shell. The variable range is as follows: The number of stiffeners is set to $2 \leq n \leq 6$, where n is an integer; the thickness of the stiffener is $5 \leq h \leq 25$; the width of the stiffener is $5 \leq w \leq 15$; the wall thickness of the piston cylinder is $5 \leq H \leq 30$, and h , w ,

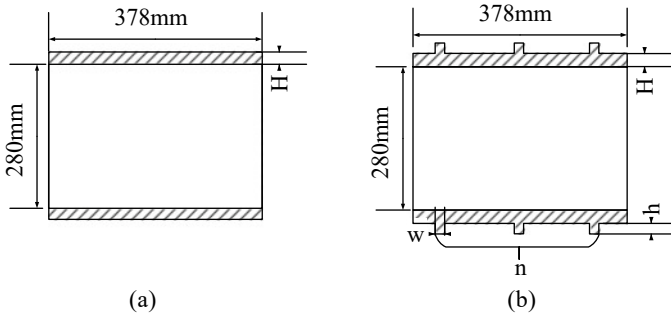


Fig. 2 Geometric model

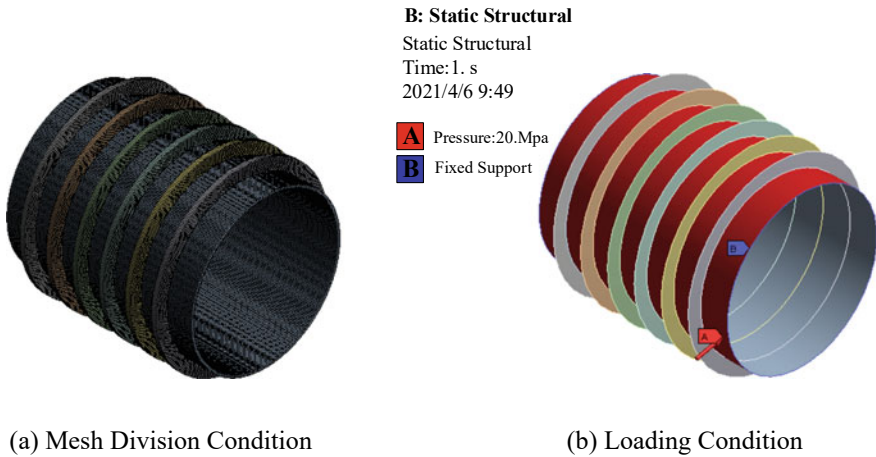


Fig. 3 Mesh division and loading conditions

and H are continuous values. The central composite design method (CCD) is used to produce experimental points for calculation. The genetic aggregation method is used to generate the response surface for the calculated experimental points.

3 Simulation Results

The sensitivity of each variable to the impact of the shell's strength, stiffness and total mass, and the results are shown in Fig. 4. The impact of the shell strength can be reflected by the maximum stress. The impact on the stiffness can be reflected by the load factor. It can be seen from the results that the stiffener has little effect on the strength of the shell, far less than the impact of the thickness of the shell on the strength; with the increase of the stiffeners, the influence of the thickness of stiffeners

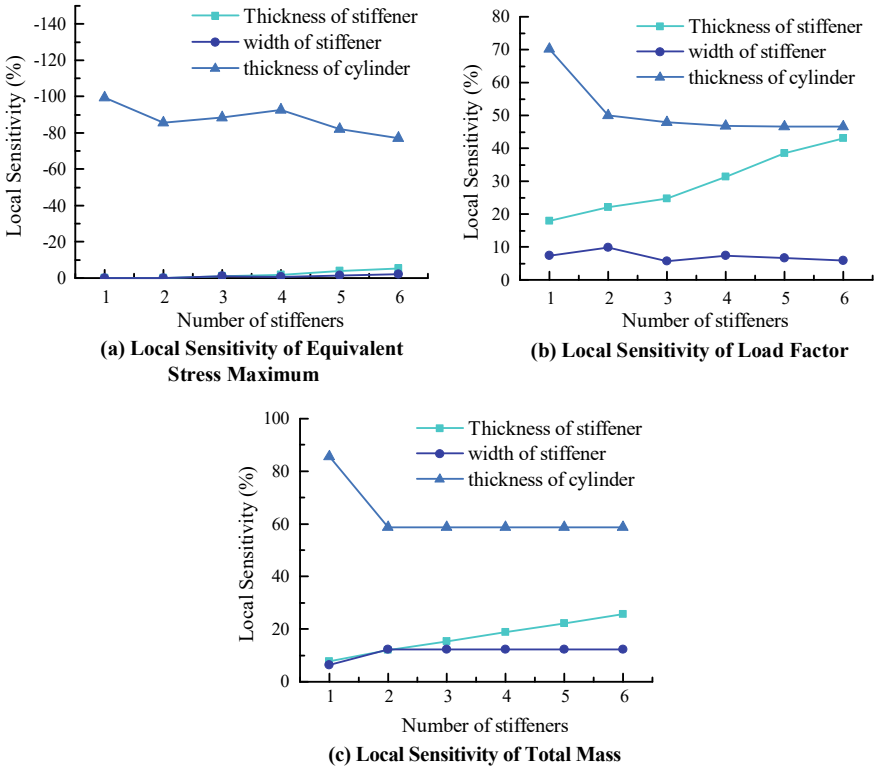


Fig. 4 Local sensitivity curve with the number of stiffeners

on the stiffness of the shell and mass of the shell increases gradually. The width of stiffeners has a small influence on the stiffness. However, the effect of stiffeners on the stiffness of shell is less than that of the wall thickness of shell. The quickest and most convenient way to improve the strength and stiffness of the shell is to increase its wall thickness. The mass of the whole shell will also be greatly increased. The number and thickness of the stiffeners can be increased appropriately in order to improve the stiffness of the body without causing a substantial increase in the mass of the shell.

Then, the parameters of the shell are optimized. The optimization method is MOCA method, and the boundary conditions and optimization objectives are: minimum mass, minimum total shape variable, and load factor greater than 3. The optimization results are: $n = 3$, $H = 9.2975$ mm, $h = 23.134$ mm, $w = 7.218$ mm. Keep one decimal place: $n = 3$, $H = 9.3$ mm, $h = 23.1$ mm, $w = 7.2$ mm. At this time, the maximum stress of the shell is 478.39 Mpa, which is less than the yield stress, and the strength is qualified; Since the calculated critical load is often 3 to 5 times larger than the actual critical load [8], the safety factor should not be less than 3, that is, the load factor ≥ 3 . In the result, the shell load factor is $3.0946 > 3$, and the

stability is qualified; the weight is 10.9 kg, which is 12.5% lighter than before the optimization, achieving the goal of lightening.

The modal analysis of the shell before and after the optimization is carried out. The natural frequency and mode shape results of the first six orders are as shown in Fig. 5 and Table 1. It can be seen from the calculation results that the natural frequency of the optimized shell has been improved, and its dynamic performance has also been optimized.

4 Conclusions

In this paper, the main influence factors on the strength and stiffness of the pressure shell are analyzed by using a finite element analysis method. It seems that the thickness of shell has the greatest influence on the strength, stiffness and total mass of the shell. With the increase of the number of stiffeners, the influence of the thickness of stiffeners on the stiffness of the shell increases greatly. Therefore, in order to lighten the shell, the number and thickness of stiffeners can be increased appropriately. In addition, the existing shell is optimized, and the optimized shell is analyzed too. The results show that the weight of the optimized shell is reduced by 12.5% on the basis of meeting the requirements of strength and stability. Finally, the modal analysis of the optimized shell shows that the natural frequencies of the optimized shell have been improved.

This paper only analyzes the uniformly distributed stiffeners, and obtains better optimization results, which proves that this analysis method is feasible and can be used in practical engineering in the future. In addition, it can also analyze the non-uniform distribution of stiffeners and stiffeners with different cross-section shapes, and find out the best form of stiffeners, and apply it to practical engineering.

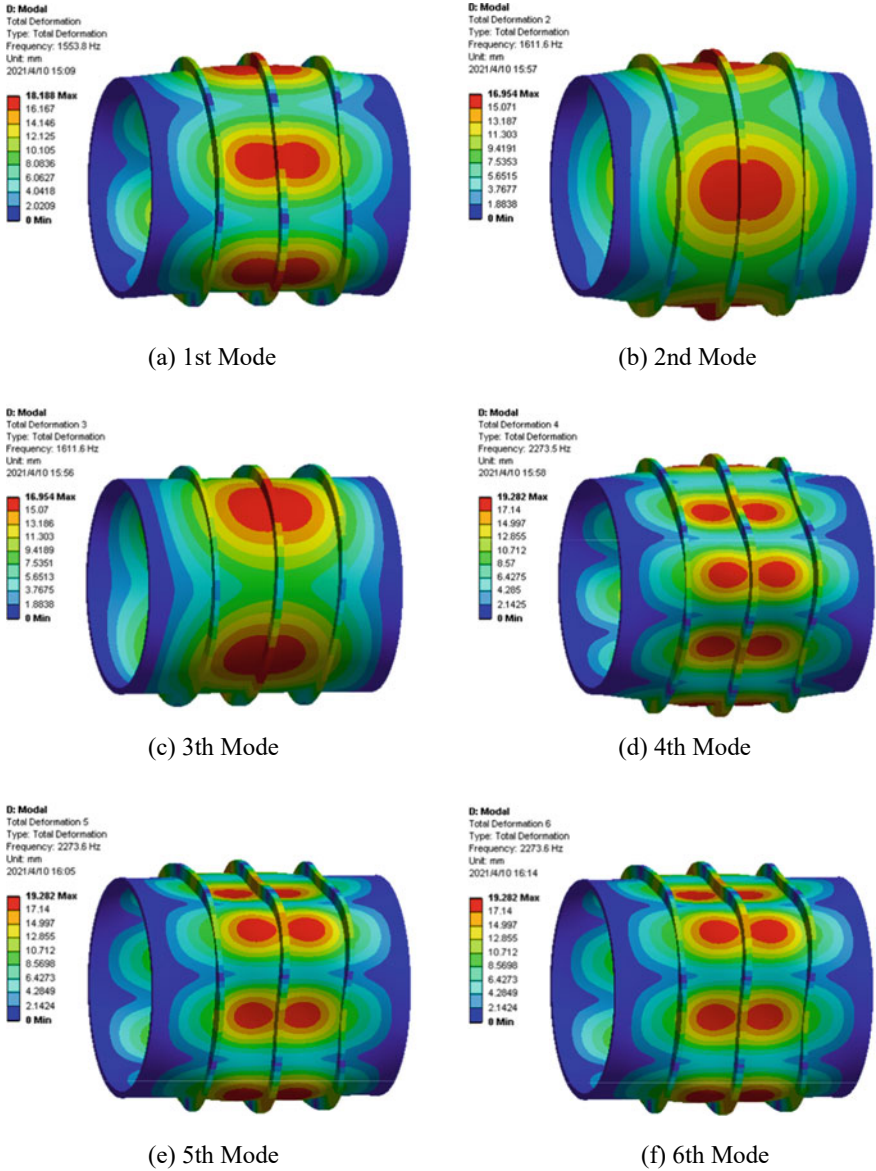


Fig. 5 The front six modes

Table 1 Comparison of the first six natural frequencies before and after optimization

Mode	Frequency after optimization (Hz)	Frequency before optimization (Hz)
1	1553.8	1513.1
2	1553.8	1513.1
3	1611.6	1768.7
4	1611.6	1768.7
5	2273.5	2044.4
6	2273.6	2044.4

Acknowledgements Supported by the Fundamental Research Funds for Central Universities (No. 3102020HHZY030001).

References

1. Bowen, A.D., Yoerger, D.R., Taylor, C., et al.: The Nereus hybrid underwater robotic vehicle. *Underw. Technol.* **28**(3), 79–89 (2009)
2. Vasilikis, D., Karamanos, S.A.: Stability of confined thin-walled steel cylinders under external pressure. *Int. J. Mech. Sci.* **51**(1), 21–32 (2009)
3. Daniel, V., Karamanos Spyros A.: Mechanics of confined thin-walled cylinders subjected to external pressure. *Appl. Mech. Rev.* **66**(1) (2014)
4. Hashemian, R., Mohareb, M.: Finite difference model for the buckling analysis of sandwich pipes under external pressure. *Ocean Eng.* **122**, 172–185 (2016)
5. Razakamiadana, Zidi, M.: Buckling and postbuckling of concentric cylindrical tubes under external pressure. *Mechan. Res. Commun.* **26**(3), 353–362 (1999)
6. Isvandzibaei, M.R., Jamaluddin, H., Hamzah, R.I.R.: Natural frequency characteristics of thin-walled homogeneous and manifold layered cylindrical shells under pressure using energy method. *J. Central South Univer.* **21**(2), 521–532 (2014)
7. Liu, B., et al.: Exact characteristic equations for free vibrations of thin orthotropic circular cylindrical shells. *Compos. Struct.* **94**(2), 484–493 (2011)
8. Von Kármán, T., Tsein, H.-S.: The buckling of thin cylindrical shells under axial compression. *J. Spacecraft Rockets* **40**(6), 898–907 2003

Emulational and Experimental Research on a Sugarcane Field Excitation Device



Hanning Mo, Chen Qiu, Shangping Li, Guiqing He, Bang Zeng, and Daiyun Yang

Abstract The vibration experiment of the sugarcane harvester is of significant value, and it is mainly done in the sugarcane field. This method has low efficiency, poor security and reliability. So, a sugarcane field excitation device is designed in this paper based on the sugarcane field excitation signal already collected. The dynamic characteristics of the sugarcane field excitation device are studied by using theoretical analysis, simulation analysis and experimental research methods. The multi-body dynamic model is studied by using rigid-flexible coupling simulation technology. Based on the simulation results, the sugarcane field excitation device is manufactured. The output frequency of the sugarcane field excitation device is calibrated by the speed calibration method. Finally, based on the experimental optimization results, the function of the sugarcane field excitation device is verified.

Keywords Excitation device · Sugarcane harvester · Dynamic simulation · Speed calibration

1 Foreword

Sugarcanes are one of the main economic crops in the South of China. In China, sugarcanes are mainly planted in hilly areas, but relatively low harvesting mechanization degree is a big problem. It is calculated currently that Guangxi is the leading sugarcane planting region in China, but its mechanized harvesting area is only about 1% in the whole planting area [1]. It is difficult to promote mechanized cutting, which

H. Mo · C. Qiu (✉)

College of Light Industry and Food Engineering, Guangxi University, Nanning 530004, China
e-mail: 303494977@qq.com

School of Mechanical and Material Engineering, Wuzhou University, Wuzhou 543000, China

S. Li

College of Electronic Information, Guangxi University for Nationalities, Nanning 530006, China

G. He · B. Zeng · D. Yang

School of Mechanical Engineering, Guangxi University, Nanning 530004, China

is affected by many factors such as costs, agricultural technologies, planting conditions and so on. Additionally, another important factor cannot always be neglected, that is, at present, mechanized sugarcane cutting may lead to a low ratoon cutting quality. Sugarcane ratoons with a low cutting quality have many shoulders and flaws in their fracture surfaces, which is easy to cause ratoons to break and affects the ratoon budding rate of the next year. It is investigated that nowadays in our country, there are some problems in adopted foreign machines for mechanized sugarcane cutting. Specifically speaking, the sugarcane ratoon breaking rate is generally about 20% to 30%. Sometimes, the rate may be even to 40%. The poor ratoon cutting quality has a severely influence on mass promotion of mechanized sugarcane harvesting [2, 3].

Aiming at what are mentioned above, scholars from home and abroad have done a great amount of basic research. They undertook kinematics analysis on cutters and did experiment research on the effect of cutting forces [4], cutting energies and different cutting edge positions on the cutting quality [5–7]. Sugarcane-pressing rollers, cutters and the interacting model between the cutter installing angle and the space distance among saw teeth of cutting edges were analyzed [8]. Meanwhile, analysis on the effect of penetrating cutting on the cutting quality was also done [9]. In most of researches mentioned above, the cutting quality was studied from the angle of kinematics and cutting parameters [10]. Domestic scholars studied broken forms of sugarcanes suffering from tensions, compressions, bending and torsion loads, discussed broken forms and kinematics of sugarcanes under plain knives [11, 12] and explored the cutting mechanism with combination of characteristics of the sugarcane stem material and mechanical properties of this material [13]. Other scholars turned to the effect mechanism of field and machine structure factors on the ratoon breaking rate [14, 15] and devoted their time to the sugarcane cutting mechanism from the perspective of establishing structural and kinematic parameters of a especially designed emulator of the sugarcane-cutter system [16], which provided a theoretical basis for improving the sugarcane ratoon cutting quality.

Nevertheless, none of these researches were focused on characteristics of sugarcane field roughness and how the sugarcane field excitation can be obtained in labs to avoid the low efficiency, the poor security and the bad reliability during experiments in sugarcane fields. Sugarcane field roughness, cutting forces and the engine cause vibrations of sugarcane harvesters. Among these three factors, sugarcane field roughness plays the most major role in causing vibrations of sugarcane harvesters. It is indicated by experiments and researches underwent by some researchers [17] that vibrations of sugarcane harvesters have bad effects on the sugarcane cutting quality. Therefore, it is necessary to study characteristics of sugarcane field roughness and do simulated sugarcane cutting experiments under the sugarcane field excitation in labs. After all, experiment conditions in labs are relatively much more controllable than those in sugarcane fields which are easily affected by environmental factors such as weathers, temperatures and so on.

The field excitation is transmitted to the frame and cutting system through the tire when the sugarcane harvester is working on the sugarcane field. Due to laboratory conditions, a sugarcane field excitation device is designed to analyze the vibration of the frame caused by field excitation. So, the experiment of the sugarcane harvester

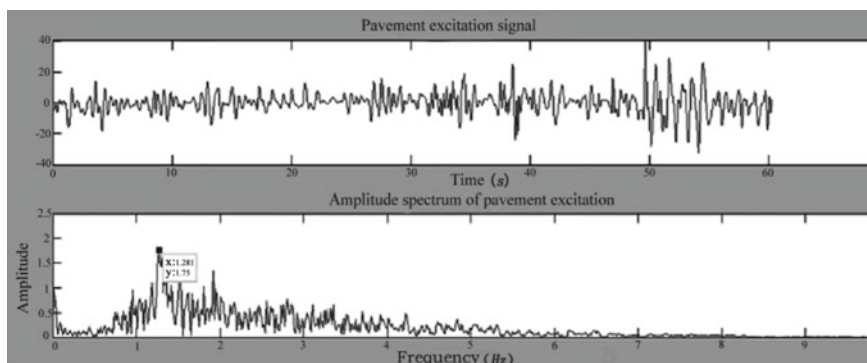


Fig. 1 Field excitation signal with the sampling frequency at 100 Hz

can be carried out in the laboratory instead of sugarcane field. The field excitation signal was collected in the sugarcane field in a suburb of Nanning city [18, 19].

The field excitation signal is low-frequency signal, and the frequency band with the largest contribution is between 0.5 and 1.5 Hz. The energy of the road excitation is most concentrated in the band of 0.5–3.5 Hz where the amplitude of each component is within 1 mm (as shown in Fig. 1).

2 The 3-Dimensional Model of the Sugarcane Field Excitation Device

The field excitation is transmitted to the frame and cutting system through the tire when the sugarcane harvester is working on the sugarcane field. In the laboratory, the unbalanced forces are produced by the eccentric mass blocks that rotate at high speed.

The mass of the sugarcane harvester experiment platform is 1 ton, as showed in Fig. 2. Two simulated sugarcane field excitation devices are placed on the front side of the experiment platform while two shock absorbers are placed on the back side. So, the sugarcane field excitation can be simulated in the laboratory.

The sugarcane field excitation device is a vibration isolation system which consists of multiple parts. The design requirements are as follows: (1) the whole system can work according to the designed motion, and it can ensure that the motion of each part will not interfere with each other; (2) the vibration frequency and amplitude of the sugarcane field excitation device should meet the experiment requirements [20].

According to the analysis results of the field excitation signal, the vibration frequencies of the sugarcane field excitation device are set as 1, 1.5, 2, 2.5 and 3 Hz. The vibration is produced by the rotation of eccentric mass blocks dragged by the motor. The virtual prototype is showed in Fig. 3.



Fig. 2 Sugarcane harvester experiment platform

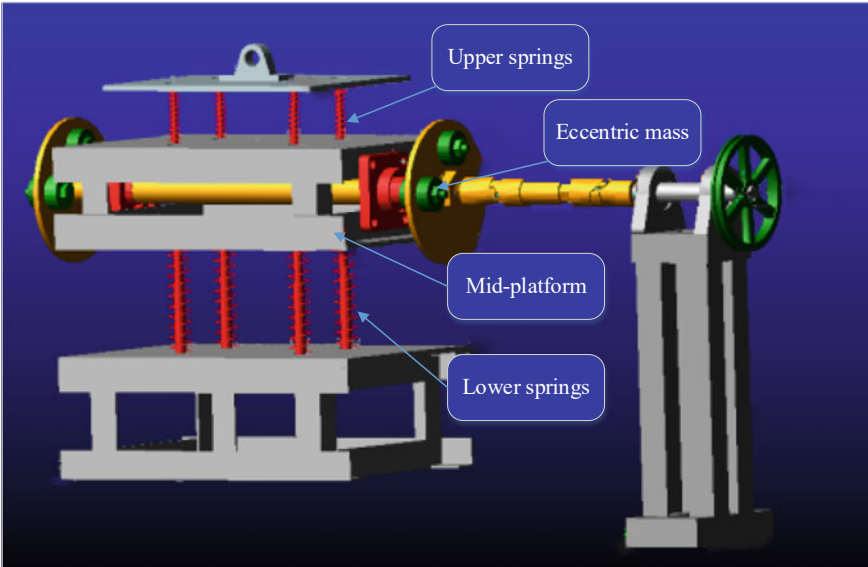


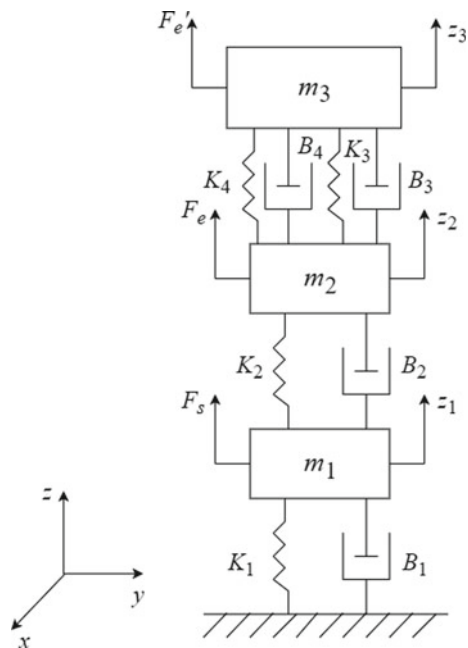
Fig. 3 Virtual prototype of the sugarcane field excitation device

3 The Dynamics Model and the Mathematical Model of the Self-developed Sugarcane Harvester Test Bed with the Sugarcane Field Excitation Device

According to Fig. 2, the mechanics model of the self-developed sugarcane harvester test bed with the simulated sugarcane field exciter can be simplified as a spring-mass system as is shown in Fig. 4. Moreover, in fact, Fig. 4 is equivalent to one fourth of the self-developed sugarcane harvester test bed with the simulated sugarcane field exciter in that a sugarcane harvester is symmetric.

In Fig. 4, the m_1 rectangle is equivalent to one of the four wheels of a sugarcane harvester. The m_2 rectangle is equivalent to one fourth of the body frame of a sugarcane harvester. The m_3 rectangle is equivalent to the engine to drive this sugarcane harvester. F_s is equivalent to the field excitation force suffered by the wheel. F_e is equivalent to the periodical force acting on the body frame of this sugarcane harvester by the engine. F_e' is equivalent to the periodical force acting on the engine by its internal structures. K_1 and B_1 are respectively regarded as the equivalent stiffness coefficient and the equivalent damping coefficient between the wheel, m_1 and the sugarcane field. K_2 and B_2 are respectively regarded as the equivalent stiffness coefficient and the equivalent damping coefficient between the wheel, m_1 and the one fourth of the body frame of this sugarcane harvester, m_2 . K_3 , B_3 , K_4 and B_4 are equivalent to stiffness coefficients and damping coefficients respectively of the rear and front anti-vibration pads of the engine. z_1 , z_2 and z_3 are respectively vertical

Fig. 4 The mechanics model of the self-developed sugarcane harvester experiment platform with the simulated sugarcane field exciter



displacements of m_1 , m_2 and m_3 . According to forces in Fig. 4, there are only linear displacements without any angular ones along the z axis in that all force are parallel to the z axis.

According to the second Newton Law, the mathematical model of Fig. 4 is shown in Eq. (1), in which the three mass blocks were chosen as study objects for force analysis and zero deformation positions of springs were chosen at the equilibrium point of this mechanics model under gravities of mass blocks without any external forces.

$$\begin{cases} m_1 \ddot{z}_1 - B_2(\dot{z}_2 - \dot{z}_1) + B_1 \dot{z}_1 - K_2(z_2 - z_1) + K_1 z_1 = F_s \\ m_2 \ddot{z}_2 + B_2(\dot{z}_2 - \dot{z}_1) - B_3(\dot{z}_3 - \dot{z}_2) - B_4(\dot{z}_3 - \dot{z}_2) + K_2(z_2 - z_1) \\ \quad - K_3(z_3 - z_2) - K_4(z_3 - z_2) = F_e \\ m_3 \ddot{z}_3 + B_3(\dot{z}_3 - \dot{z}_2) + B_4(\dot{z}_3 - \dot{z}_2) + K_3(z_3 - z_2) + K_4(z_3 - z_2) = F'_e \end{cases} \quad (1)$$

Make Eq. (1) into a matrix form as shown in Eq. (2), which is actually a second Newton Law equation.

$$M\ddot{Z} + B\dot{Z} + KZ = F \quad (2)$$

In Eq. (2), M is the mass matrix of the mechanics model, Fig. 4. Z is the displacement matrix. B is the damping coefficient matrix. K is the stiffness coefficient matrix. F is the external force matrix. These five matrixes are shown in Eq. (3) in detail.

$$\begin{cases} M = \begin{bmatrix} m_1 & & \\ & m_2 & \\ & & m_3 \end{bmatrix} \\ Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \\ B = \begin{bmatrix} B_2 + B_1 & -B_2 & 0 \\ -B_2 & B_2 + B_3 + B_4 & -B_3 - B_4 \\ 0 & -B_3 - B_4 & B_3 + B_4 \end{bmatrix} \\ K = \begin{bmatrix} K_2 + K_1 & -K_2 & 0 \\ -K_2 & K_2 + K_3 + K_4 & -K_3 - K_4 \\ 0 & -K_3 - K_4 & K_3 + K_4 \end{bmatrix} \\ F = \begin{bmatrix} F_s \\ F_e \\ F'_e \end{bmatrix} \end{cases} \quad (3)$$

4 Rigid-Flexible Coupling Simulation Based on ADAMS

Import the virtual prototype into the dynamics simulation software MSC.ADAMS, and set the material properties and quality of each component. Add gravity in the Z direction, so that the whole system is in the gravity field [21, 22].

According to the actual working condition of the exciter, the eccentric mass and rotational speed are set as two parameters of the simulation experiment (as listed in Table 1).

During the dynamics simulation, add a Marker point to the ground and the support plate with hinge separately, and the two Marker points are coincident. Then, measure their amplitudes in the Y and Z directions respectively. Set the simulation time to 5 s and the simulation step to 50 steps.

The trend of the above picture is basically the same. The springs sway up and down under the action of alternating force, and the amount of spring compression changes uniformly with time. The specific deformations of the springs are as shown in the above figures.

According to the simulation specific data, the results are shown in Table 2:

It is known from the simulation results:

- (1) Under no-load condition, the exciter can excite the frame at the frequencies of 1, 1.5, 2, 2.5, and 3 Hz, which can meet the influence of field vibration.
- (2) The simulation data indicates that the amplitude of the exciter in the vertical direction is substantially within 2 mm. But in actual work, except for the vibration in the vertical direction, the exciter will vibrate in the horizontal

Table 1 Simulation parameters

Eccentric mass (kg)	Rotational speed (r/min)
0	60
1	90
2	120
3	150
4	180

Table 2 Simulation results

Rotating speed (r/min)	Vibration with 4 kg eccentric mass (mm)	Vibration with 3 kg eccentric mass (mm)	Vibration with 2 kg eccentric mass (mm)	Vibration with 1 kg eccentric mass (mm)
60	2.2198	0.6841	0.5343	0.365
90	1.9361	0.743	0.6008	0.4061
120	1.4261	1.1731	0.9199	0.6553
150	1.1181	1.5592	1.227	0.8497
180	0.8086	1.9286	1.5874	1.0776

direction. The vibration of the experiment platform will be increased. When the simulated road surface vibration experiment is carried out, it can meet the obvious requirements of the vibration of the experiment platform.

5 Experimental Research on the Sugarcane Field Excitation Device

According to the design results, the sugarcane field excitation device manufactured is shown in Fig. 5.

The single-factor experiment is carried out to experiment the no-load vibration performance of the sugarcane field excitation device by changing the eccentric mass and the eccentric shaft speed [23, 24]. The experiment system is shown in Fig. 6.



Fig. 5 The sugarcane field excitation device

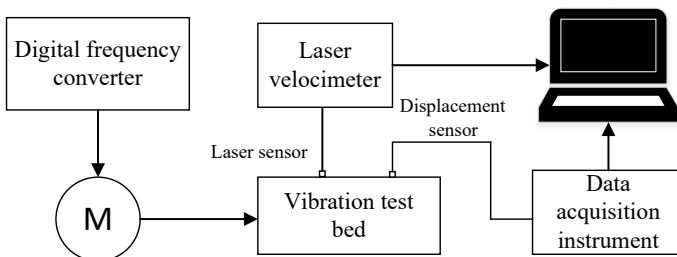


Fig. 6 Experiment system for vibration

Table 3 Experiment arrangement

Sequence	Factor	
	Eccentric mass m (kg)	Eccentric shaft speed v (m/s)
1	0	60
2	1	90
3	2	120
4	3	150
5	4	180

The factors of the vibration experiment are the eccentric mass m and the eccentric shaft speed v . As shown in Table 3.

In order to fully understand the vibration performance of the sugarcane field excitation device, the experiments of the vibration in the vertical direction are arranged.

A laser tachometer is used to measure the eccentric shaft speed in this experiment. By recording the output frequency of the inverter corresponding to the current display value of the tachometer, the frequency value is used as the standard for speed adjustment, which is called speed calibration method. The speed calibration Table 4 is as follows.

The amplitudes with 0 kg eccentric mass, 1 kg eccentric mass, 2 kg eccentric mass, 3 kg eccentric mass, and 4 kg eccentric mass at different speeds are shown in Fig. 7.

The speed calibration table shows that the excitation frequency of the sugarcane field excitation device ranges from 1.3 to 4.8, which satisfies the result of the main energy band of the field excitation of 1–6 Hz.

The experiment results showed that:

Table 4 Speed calibration

Output frequency of the inverter (Hz)	Eccentric shaft speed (r/min)	Excitation frequency (Hz)
5	80	1.3
6	112	1.87
7	134	2.23
8	155	2.58
9	170	2.83
10	190	3.16
11	212	3.53
12	230	3.83
13	251	4.18
14	270	4.5
15	289	4.81

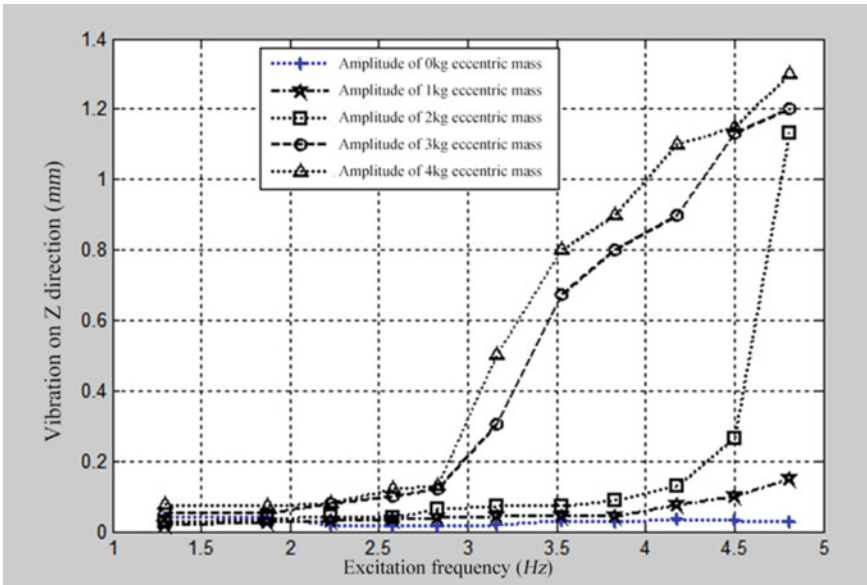


Fig. 7 The vibration in vertical direction

In the vertical vibration, the vibration amplitude ranges from 0.075 to 1.3 mm when the eccentric mass is 4 kg. The vibration amplitude ranges from 0.05 to 1.2 mm when the eccentric mass is 3 kg. The vibration amplitude ranges from 0.025 to 1.13 mm when the eccentric mass is 2 kg. The vibration amplitude ranges from 0.02 to 0.15 mm when the eccentric mass is 1 kg. The no-load experiment data initially shows that low-frequency with high-amplitude and high-frequency with low-amplitude excitation can be achieved by changing the magnitude of the eccentric mass. The excitation performance satisfies the design requirements.

In summary, the sugarcane field excitation device can satisfy the field excitation requirements.

6 Conclusions

1. According to the analysis of field excitation signal, the sugarcane field excitation device was designed.
2. The multi-body dynamic model was studied by using the software MSC.ADAMS. Simulation results show that the vibration exciter can meet the experiment requirements. Based on the simulation results, the sugarcane field excitation device was manufactured.
3. The sugarcane field excitation device was calibrated by speed calibration method, and the vibration of the z phase of the exciter at no load was recorded.

The experiment results show that the sugarcane field excitation device can satisfy the field excitation requirements well.

Acknowledgements Fund projects: 1. Middle-aged and Young Teachers' Basic Scientific Research Ability Promotion Project of Guangxi Universities called Dynamic reverse design research on sugarcane harvesters for hilly areas based on dynamic characteristics of cutters (Project number: 2020KY17008) of China, 2020.

2. Major university-level research project of Wuzhou University called Reverse design method research on sugarcane harvesters for hilly areas based on dynamic characteristics of cutters (Project number: 2020B003) of China, 2020.

3. National Natural Science Foundation Project called Research on cutting system vibration characteristics of sugarcane harvesters under complicated excitations (Grant Number: 51465006) of China, 2014.

References

1. Dongsheng, L., Sijing, T.: The analysis of the accounting and controlling of sugarcane production cost in Guangxi. *Academic Forum* **08**, 28–43 (2013)
2. Lin, H., Lai, R., Zeng, Z.: Study on the technical specification of sugarcane mechanization harvesting. In: *International Academic Conference of Chinese Society for Agricultural Machinery* (2012)
3. Chen, C., He, L., Yan, H.: Research on the several agronomic techniques of sugarcane mechanization harvesting. In: *The Annual Meeting of the Crop Science Society of China* (in Chinese with English abstract) (2014)
4. Oduori, M.F., Sakai, J., Gupta, C.P.: Kinematics of revolving-knife disk-type sugarcane basecutter-1. *AMA* **23**(4), 9–15 (1992)
5. Kroes, S., Harris, H.D.: Variation of cutting energies along a sugarcane internode. In: *Conference on Engineering in Agriculture and Food Processing*. Held University of Queensland, Gatton College, Lawes. *Agricultural-Engineering-Australia*, Vol. 25, 3, 55 (1996)
6. Kroes, S., Harris, H.D.: The specific splitting energy of sugarcane. In: *Proceedings of the 1998 Conference of the Australian Society of Sugar Cane Technologists held at Ballina*. NSW, Australia, pp. 349–356 (1998)
7. da Cunha Mello, R., Harris, H.: Cane damage and mass losses for conventional and serrated hasecutter blades. *Proc. Aust. Soc. Soc. Sugar Cane Technol.* **22**, 84–91 (2000)
8. Kroes, S., Harris, H.: Effects of cane harvester base cutter parameters on the quality of cut. *Proc. Aust. Soc. Sugar Stalk Technol.* 169–179 (1994)
9. Mello, R.D.C., Harris, H.: Performance of basecutter of sugarcane harvester with angled and serrated blades. *Revista Brasileira De Engenharia Agricola E Ambiental* **7**(2), 355–358 (2003)
10. Liu, Q., Ou, Y., Qing, S., et al.: Failure experiments of sugarcane stalks under torsion, compression and tension load. *Trans. Chinese Soc. Agric. Eng.* **22**(6), 201–204 (2006) (in Chinese with English abstract)
11. Liu, Q., Ou, Y., Qing, S., Song, C.: Mechanics analysis on stubble damage of sugarcane stalk during cutting by smooth-edge blade. *Trans. Chinese Soc. Agric. Mach.* **38**(9), 70–73 (2007) (in Chinese with English abstract)
12. Liu, Q., Ou, Y., Qing, S., et al.: Stubble damage of sugarcane stalks in cutting experiment by smooth-edge blade. *Trans. CSAE* **23**(3), 103–107 (2007) (in Chinese with English abstract)
13. Liu, Q., Ou, Y., Qing, S., et al.: Study on the cutting mechanism of sugarcane stem. *J. Agric. Mechanization Res.* (01), 21–24 2007

14. Yang, J., Chen, G.: Experimental study on influencing factors of broken biennial root rate for a single base cutter of sugarcane harvester. *Trans. Chinese Soc. Agric. Mach.* **38**(3), 69–74 (2007) (in Chinese with English abstract)
15. Jian, Y., Zhaoxin, L., Mo, J., et al.: Experimental research on factors affecting the cutting quality of sugarcane cutter. *Trans. Chinese Soc. Agric. Eng.* **21**(5), 60–64 (2005). (in Chinese with English abstract)
16. Wang, Y., Jian, Y., Zenghan, L., et al.: Dynamic simulation experiment on effects of sugarcane cutting beneath surface soil. *Chinese Soc. Agric. Eng.* **27**(08), 150–157 (2011). (in Chinese with English abstract)
17. Zhiqing, C.: *Structure Design and Analysis and Experiment Research of the Carrying Frame in Minitype Sugarcane Harvester*. Guangxi University, Guangxi (2007). (in Chinese with English abstract)
18. Li, J., Zhang, Z., Xu, H., Feng, Z.: Dynamic characteristics of the vibratory roller experiment-bed vibration isolation system: simulation and experiment. *J. Terramechanics Dyn.* **56**, 139–156 (2014)
19. Qiu, C., Li, S., Mo, H., Yang, D., et al.: Collection and analysis of main excitation source signals of sugarcane harvester. In: *Agricultural Mechanization Research* (2018) (in Chinese)
20. Mo, H., Li, S., Zhou, J., Qiu, C., et al.: Axial vibration experiments for cutter frames of sugarcane harvesters with different structures under complicated excitation. *IAEJ* **27**(3), 18 (2018)
21. Liu, F., Wu, J., Gu, F., Ball, A.D.: An introduction of a robust OMA method: CoS-SSI and its performance evaluation through the simulation and a case study. *Shock Vib.* (2019)
22. Soria, L., Peeters, B., Anthonis, J., der Auweraer, H.V.: Operational modal analysis and the performance assessment of vehicle suspension systems. *Shock Vib.* **19**(5), 1099–1113 (2012)
23. Shi, F., Wang, H., et al.: Road spectrum bench experiment research. *J. Vib. Shock* **19** (2012)
24. Geng, Y., et al.: Vehicle road simulation experiment control technology. *Autom. Appl.* **03**, 28–31 (2011)

Experimental Research on Influence Factors of the Sugarcane Ratoon Cutting Quality Under Vibration Conditions



Chen Qiu, Hanning Mo, Shangping Li, Guiqing He, Bang Zeng, and Daiyun Yang

Abstract Aimed at improving the sugarcane ratoon cutting quality of sugarcane harvesters for hilly areas, a sugarcane harvester experiment platform was developed. During sugarcane cutting experiments, flaws may appear in sugarcane ratoons. The flaw number, thickness and length were measured. An experiment index, y was introduced as the comprehensive cutting quality evaluating value of the sugarcane ratoon cutting quality by the improved entropy method with these three parameters. A regressive mathematical model was set up by orthogonal experiments and to study effects of the vibration frequency, the vibration amplitude, the cutter rotating velocity, the sugarcane feeding velocity and the cutter installing angle on the sugarcane ratoon cutting quality. It is shown in experiments that there is a strong linear relationship among y , vibration amplitude and frequency and the amplitude as well as frequency had great effects on y while the moving velocity, the cutter rotating velocity and the cutter installing angle had relatively less significant effects on y . According to the fact whether the effect on y is significant, the significance order is as follows, the vibration amplitude, frequency, the sugarcane feeding velocity, the cutter rotating velocity and the cutter installing angle. Interaction between the vibration amplitude and frequency and that between the amplitude and the cutter rotating velocity also have effects on y . In details, the greater the vibration amplitude and frequency are, the greater y will be, which means the worse the sugarcane ratoon cutting quality will be. The greater the vibration amplitude and the cutter rotating velocity are, the greater y will be. This research was done to study the effect mechanism of the sugarcane ratoon cutting quality and lay the foundation of design and development of sugarcane harvesters with a high sugarcane ratoon cutting quality for hilly areas.

C. Qiu · H. Mo (✉)

College of Light Industry and Food Engineering, Guangxi University, Nanning 530004, China
e-mail: 1433025842@qq.com

School of Mechanical and Material Engineering, Wuzhou University, Wuzhou 543000, China

S. Li

College of Electronic Information, Guangxi University for Nationalities, Nanning 530006, China

G. He · B. Zeng · D. Yang

School of Mechanical Engineering, Guangxi University, Nanning 530004, China

Keywords Sugarcane ratoon cutting quality · Experimental research · Vibration amplitude · Vibration frequency · Linear relationship

1 Introduction

Sugarcanes are one of the main economic crops in the South of China. In China, sugarcanes are mainly planted in hilly areas, but relatively low harvesting mechanization degree is a big problem. It is calculated currently that Guangxi is the leading sugarcane planting region in China, but its mechanized harvesting area is only about 1% in the whole planting area. It is difficult to promote mechanized cutting, which is affected by many factors such as costs, agricultural technologies, planting conditions and so on. Additionally, another important factor cannot always be neglected, that is, at present, mechanized sugarcane cutting may lead to a low ratoon cutting quality. Sugarcane ratoons with a low cutting quality have many shoulders and flaws in their fracture surfaces, which is easy to cause ratoons to break and affects the ratoon budding rate of the next year. It is investigated that nowadays in our country, there are some problems in adopted foreign machines for mechanized sugarcane cutting. Specifically speaking, the sugarcane ratoon breaking rate is generally about 20–30%. Sometimes, the rate may be even to 40%. The poor ratoon cutting quality has a severely influence on mass promotion of mechanized sugarcane harvesting [1–4].

Aiming at what are mentioned above, scholars from home and abroad have done a great amount of basic research. They undertook kinematics analysis on cutters and did experiment research on the effect of cutting forces, cutting energies and different cutting edge positions on the cutting quality. Sugarcane-pressing rollers, cutters and the interacting model between the cutter installing angle and the space distance among saw teeth of cutting edges were analyzed. Meanwhile, analysis on the effect of penetrating cutting on the cutting quality was also done. In most of researches mentioned above, the cutting quality was studied from the angle of kinematics and cutting parameters. Domestic scholars studied broken forms of sugarcanes suffering from tensions, compressions, bending and torsion loads, discussed broken forms and kinematics of sugarcanes under plain knives and explored the cutting mechanism with combination of characteristics of the sugarcane stem material and mechanical properties of this material. Other scholars turned to the effect mechanism of field and machine structure factors on the ratoon breaking rate and devoted their time to the sugarcane cutting mechanism from the perspective of establishing structural and kinematic parameters of a especially designed emulator of the sugarcane-cutter system, which provided a theoretical basis for improving the sugarcane ratoon cutting quality [5, 6]. Nevertheless, none of these researches were focused on the effect of axial vibrations of cutters on the ratoon cutting quality. In sugarcane fields, it is not possible to avoid axial vibrations of cutters because of a great number of vibrations actuating factors. It is indicated by experiments and researches underwent by some researchers that axial vibrations of cutters are closely connected with the ratoon

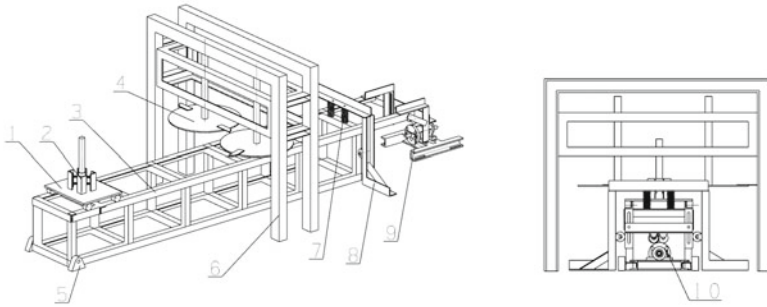
cutting quality. Therefore, it is necessary to do a further research on effects of axial vibrations of cutters and its relative factors on the sugarcane ratoon cutting quality [7, 8].

In this paper, a series of sugarcane cutting experiments were conducted through the experiment platform designed by our research group to study effect principles of axial vibration amplitudes and frequencies of cutters and relative cutting parameters on the ratoon cutting quality. On this basis, effects of following parameters on the sugarcane ratoon cutting quality were further studied, such as the vibration amplitude, frequency, the sugarcane feeding velocity, the cutter installing angle and the cutter rotating velocity etc. In addition, the quantitative relationship expression of the ratoon cutting quality, the amplitude, the frequency, the sugarcane feeding velocity, the cutter installing angle and the cutter rotating velocity was set up, which is an important theoretical basis for the innovative design theory of cutter structures of sugarcane harvesters.

2 The Self-developed Sugarcane Harvester Experiment Platform

In sugarcane fields, there are many factors which may actuate sugarcane harvesters, such as force unbalances of cutters, road roughness, vibrations of rotating components and so on. Axial cutter vibrations may appear, which in fact will lead sugarcanes to the relative displacement effect between a certain frequency and amplitude in the axial direction. In order to study this effect on the sugarcane ratoon cutting quality, an experiment platform was designed to make moving sugarcanes and cutters generate controllable relative displacements with certain frequencies and amplitudes in the axial direction. What's more, this device then made sugarcanes cut down by the cutters, which is analogous to the practical situation in the field. Plus, breakage situations of sugarcane ratoons were observed and relative principles were obtained by the statistical method.

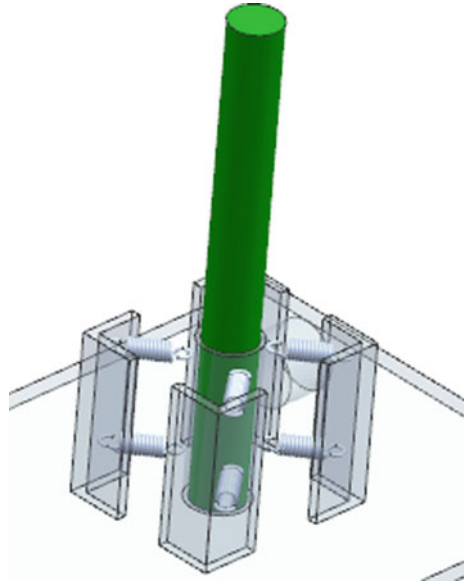
These experiments were conducted through the sugarcane harvester experiment platform designed by our research group. Its structure sketch and the manufactured experiment platform are illustrated in Fig. 1. It is mainly composed of the sugarcane-feeding device, the feeding device, the cutting device and eccentric vibration devices. Sugarcanes were fixed in the sugarcane-feeding device shown as the No. 1 part in Fig. 1 by the sugarcane-clamping device shown as the No. 2 part in Fig. 2 and its stepless velocity controlling and reversing were achieved by a velocity-adjusting motor with a variable frequency controller and a reducer, which can imitate moving velocities of a sugarcane harvester. The structure of the sugarcane-clamping device and the manufactured the sugarcane-clamping device are indicated in Fig. 2. A sugarcane was inserted into a socket which is fixed by 8 springs to imitate the tightening effort of soils. Cutters shown as the No. 4 part in Fig. 1 both consist of 2 disks both with a 450 mm diameter connected with blades all of an 80 mm length and a 5 mm



1. The sugarcane-feeding device 2. The sugarcane-clamping device 3. The feeding platform 4. The cutting device with double blades 5. The fixed hinged support of the platform 6. The cutter holder 7. The hold-down spring 8. The position-limiting device 9. The eccentric vibrating device 10. The eccentric wheels

Fig. 1 The comprehensive structure sketch of the experiment platform

Fig. 2 The sugarcane-clamping device



thickness. Cutter installing angles can be adjusted and the cutters were driven by a stepless velocity-controlling motor. A reverse magnetic value was used to make the cutters begin and stop running as well as positively and negatively rotating. The eccentric vibration device shown as the No. 9 part in Fig. 1 was used to make the feeding platform shown as the No. 3 part in Fig. 1 repeatedly swing around the fixed hinged support shown as the No. 5 part in Fig. 1, which can obtain vibrations in the axial direction just in sugarcane cutting points. Amplitudes and frequencies of this kind of vibrations are dependent on the eccentric distance of eccentric wheels and

Table 1 Calculating expressions of the sugarcane ratoon cutting quality

$x'_{ji} = \frac{x_{ji} - x_{j \min}}{x_{j \max} - x_{j \min}}$	$p_j = \frac{x'_{ji}}{\sum_{i=1}^m x'_{ji}}$	$e_j = -k \sum_{i=1}^m p_{ji} \ln p_{ji}$	$g_j = 1 - e_j$	$w_j = g_j / \sum_{j=1}^n g_j$
-----------------------------------------------------------------	----------------------------------------------	-------------------------------------------	-----------------	--------------------------------

The comprehensive evaluation index $y = \sum_{j=1}^3 \bar{x}_j \cdot w_j$
 x_{ij} is the value of the evaluation index, in which $j = 1, 2, 3$, respectively corresponding to the number of cracks and the crack thickness and length, $i = 1 \dots m$ is the number of experiments

the frequency of the motor, which is used to stimulate axial vibrations generated by cutters and harvesting machines under external excitations in sugarcane fields.

3 Evaluation Indexes

The comprehensive cutting quality evaluating value includes the number of ratoon cracks, crack thicknesses and lengths. These three indexes all have effects on the ratoon cutting quality and they were denoted as x_{ij} during experiment data processes. However, it is difficult to only use one index to evaluate the cutting quality, so it is necessary to use a comprehensive one combining all these three indexes. In this paper, the improved entropy method was used to calculate the comprehensive evaluation index, y of the cutting quality. y is called the comprehensive cutting quality evaluating value. The greater y is, the poorer the sugarcane cutting quality will be. Calculating expressions of y are shown in Table 1.

4 Orthogonal Experiments

In this experiment, No. 20 Guitang sugarcanes just harvested were adopted, which is shown in Fig. 3. Bodies of these sugarcanes were straight and their leaves as well as burrs were wiped off. Their average diameter was 28 ± 3 mm. 1000 mm lengths upward roots were chosen.

During these experiments, a laser displacement-measuring system shown in Fig. 4 was used. It consists of a laser displacement sensor (LK-G150A), a controller (LK-G3001A), a power transformer (MS2-H50), a laptop and related measuring software (LK-H1W). This system was used to measure vibration amplitudes of the feeding device shown as the No. 3 part in Fig. 1 in the axial direction near sugarcane cutting points.

A laser velocity meter (AR926) was used to measure rotation velocities of the cutters shown as the No. 4 part and eccentric wheels shown as the No. 10 part in Fig. 1.

A vernier caliper whose accuracy is 0.02 mm was used to measure crack lengths and thicknesses of incisions of sugarcane ratoons.

Fig. 3 No. 20 Guitang sugarcanes



Fig. 4 Laser displacement measurement system



Sugarcane cutting experiments were made up with orthogonal experiments. The clamping device driven by a motor is fixed on the feeding device. Sugarcanes were inserted in the clamping device. Sugarcane feeding velocities were set up through the velocity-adjusting device with controlling frequencies to imitate the moving velocity of harvesting machines. Meanwhile, the laser displacement-measuring system was used to measure amplitudes of the cutters at sugarcane cutting points. Factors relative with these experiments were amplitudes in the axial direction, vibration frequencies,

Table 2 Factors and levels of the orthogonal experiment

Factors	Levels	
	1	2
Frequency (x_1)/Hz	9	6
Amplitude (x_2)/mm	6.5	3.5
The rolling speed of the cutter (x_3)/rpm	700	600
The moving speed (x_4)/m s ⁻¹	0.6	0.4
The installing angle of the cutter (x_5) ^o	10 ^o	0 ^o

sugarcane feeding velocities which imitate sugarcane harvester moving velocities, cutter installing angles and cutter rotating velocities.

Purposes were to observe effects of vibration amplitudes, frequencies, sugarcane feeding velocities, cutter installing angles and rotation velocities of the cutters on the ratoon cutting quality. Factors and levels of the orthogonal experiment are shown in Table 2 (5 factors and 2 levels). Design on the orthogonal experiment is shown in Table 3. The index of this experiment is the comprehensive evaluation index, y . Results are indicated in Table 4. Variance analysis on results of orthogonal experiments is shown in Table 5.

Table 3 Design on the orthogonal experiment

No	Factors				
	x_1 (Frequency)	x_2 (Amplitude)	x_3 (The rolling speed of the cutter)	x_4 (The moving speed)	x_5 (The cutter angle)
1	1	1	1	1	1
2	1	1	1	2	2
3	1	1	2	1	2
4	1	1	2	2	1
5	1	2	1	1	2
6	1	2	1	2	1
7	1	2	2	1	1
8	1	2	2	2	2
9	2	1	1	1	2
10	2	1	1	2	1
11	2	1	2	1	1
12	2	1	2	2	2
13	2	2	1	1	1
14	2	2	1	2	2
15	2	2	2	1	2
16	2	2	2	2	1

Table 4 Results of orthogonal experiment

No	Factors					
	x_1 (Frequency)	x_2 (Amplitude)	x_3 (The rotation speed of the cutter)	x_4 (The feeding speed)	x_5 (The cutter angle)	y (The comprehensive cutting quality)
1	1	1	1	1	1	1
2	1	1	1	2	2	0.7118
3	1	1	2	1	2	0.6981
4	1	1	2	2	1	0.5191
5	1	2	1	1	2	0.0518
6	1	2	1	2	1	0.25
7	1	2	2	1	1	0.4618
8	1	2	2	2	2	0.1037
9	2	1	1	1	2	0.4672
10	2	1	1	2	1	0.34
11	2	1	2	1	1	0.5382
12	2	1	2	2	2	0.2691
13	2	2	1	1	1	0.1418
14	2	2	1	2	2	0
15	2	2	2	1	2	0.3909
16	2	2	2	2	1	0.2309

Table 5 Variance analysis on results of orthogonal experiments

Source	Sum of squares	Freedom degree	Variance	F	P
Corrected model	0.914	6	0.152	6.353	0.007
Intercept	2.361	1	2.361	98.451	<0.001
Frequency	0.129	1	0.129	5.395	0.045
Amplitude	0.531	1	0.531	22.151	0.001
The rotation speed	0.034	1	0.034	1.414	0.265
The feeding speed	0.114	1	0.114	4.774	0.057
The angle	0.005	1	0.005	0.224	0.647
Error	0.216	9	0.024		
Total	3.491	16			
Corrected total	1.130	15			

It is shown according to Table 5 that under the experiment condition, there are results as follow.

- (1) When the F value of vibration amplitudes was 22.151, $P < 0.01$, there was an extremely significant effect.
- (2) When the F value of vibration frequencies was 5.395, $P < 0.05$, there was a significant effect.
- (3) When the F value of feeding velocities was 4.774, $P < 0.1$, there was a great effect.
- (4) When F values of cutter installing angles and cutter rotating velocities were respectively 0.224 and 1.414, P values were respectively 0.647 and 0.265 which are both greater than 0.1, there was an insignificant effect.
- (5) Influencing degrees of factors of these experiments are as follow, that is, vibration amplitudes > vibration frequencies > feeding velocities > cutter rotating velocities > cutter installing angles.

It turns out according to variance analysis on results of these experiments that when the cutters cut sugarcanes, vibration amplitudes and frequencies will be major influence factors among all factors influencing the cutting quality if large amplitudes and frequencies of the cutters exist.

5 Analysis for Interaction Effects on the Experiment Index

Figure 5 is the curve of effects of interaction between the vibration frequency and amplitude on the comprehensive cutting quality. Other factors should be set to be in the level of 0 when the curve was drawn. According to Fig. 5, when vibration frequencies and amplitudes increase at the same time, their interaction effect will

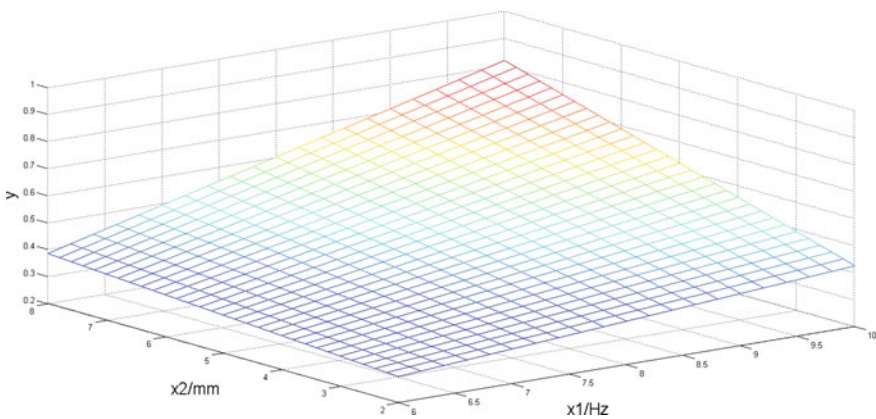


Fig. 5 The simulating calculation results of the interaction effect of the vibration frequency and amplitude

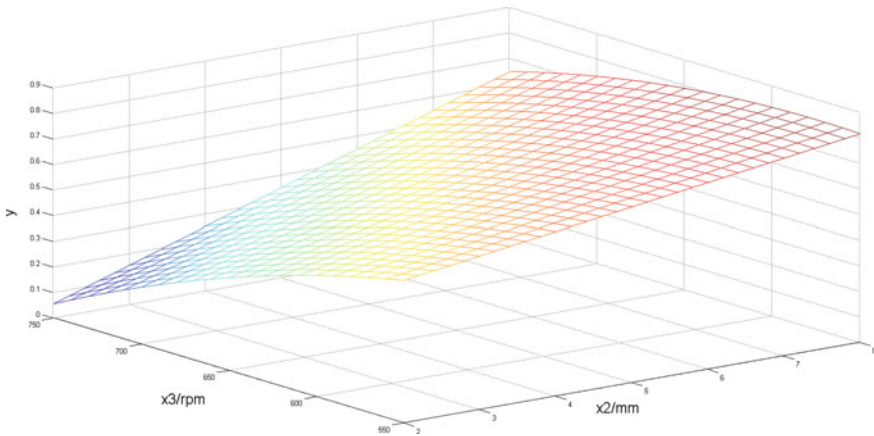


Fig. 6 The simulating calculation results of the interaction effect of the vibration amplitude and cutter rotation speed

make the comprehensive cutting quality evaluating value greater, which means the worse cutting quality. When vibration frequencies and amplitudes decrease at the same time, their interaction effect will make the comprehensive cutting quality value smaller, which means the better cutting quality. With the same vibration frequency, the greater the vibration amplitude is, the greater the comprehensive cutting quality evaluating value will be, that is, the worse the cutting quality will be and vice versa. With the same vibration amplitude, the greater the vibration frequency is, the greater the comprehensive cutting quality evaluating value will be, that is, the worse the cutting quality will be and vice versa.

Figure 6 is the graph of the effect of interaction between vibration amplitudes and cutters rotating velocities by the same handling method. According to Fig. 6, with a certain cutter rotating velocity, it will cause the comprehensive cutting quality evaluating value to be greater when the vibration amplitude increases, that is, the cutting quality will be worse. While with the same vibration amplitude, it will make the comprehensive cutting quality value smaller when the cutter rotating velocity increases, that is, the cutting quality will be better. If the two factors increase at the same time, the comprehensive cutting quality evaluating value will be greater, that is, the cutting quality will be worse.

6 Conclusions

- (1) The improved entropy evaluation method was adopted to take the number of sugarcane cracks, crack thickness and crack length as single comprehensive cutting quality evaluating values of the comprehensive cutting quality. Under experiment conditions, there is a high linear relationship between vibration

amplitudes or frequencies and the sugarcane cutting quality. Linear coefficients between vibration amplitudes or frequencies and the sugarcane cutting quality are respectively 0.997 (in the level of 0) and 0.918 (in the level of 0.05). The greater the vibration amplitude is, the worse the cutting quality will be. The greater the vibration frequency is, the worse the cutting quality will be. In structural design on harvesters, cutter vibrations should be managed to be decreased.

- (2) Effects of vibration amplitudes, vibration frequencies feeding velocities, cutter rotating velocities and cutter installing angles on the comprehensive cutting quality evaluating value were compared by orthogonal experiments. It is shown by analysis results that under experiment conditions, among these 5 effect factors, effects of vibration amplitudes and frequencies are significant while effects of feeding velocities, cutter rotating velocities and cutter installing angles are insignificant. Effect degrees of these 5 factors on the comprehensive cutting quality evaluating value are as follow, i.e., vibration amplitudes > vibration frequencies > feeding velocities > cutter rotating velocities > cutter installing angles.
- (3) Under experiment conditions, there is an interaction effect between vibration amplitudes and frequencies on the comprehensive cutting quality evaluating value. The effect principle is that the greater the vibration amplitude and frequency are, the greater the comprehensive cutting quality evaluating value will be, which means the cutting quality will be worse. The interaction effect between vibration amplitudes and cutter rotating velocities on the comprehensive cutting quality evaluating value is that if the two factors increase at the same time, the comprehensive cutting quality evaluating value will be greater, that is, the cutting quality will be worse.

Acknowledgements Fund projects: 1. Middle-aged and Young Teachers' Basic Scientific Research Ability Promotion Project of Guangxi Universities called Dynamic reverse design research on sugarcane harvesters for hilly areas based on dynamic characteristics of cutters (Project number: 2020KY17008) of China, 2020.

2. Major university-level research project of Wuzhou University called Reverse design method research on sugarcane harvesters for hilly areas based on dynamic characteristics of cutters (Project number: 2020B003) of China, 2020.

3. National Natural Science Foundation Project called Research on cutting system vibration characteristics of sugarcane harvesters under complicated excitations (Grant Number: 51465006) of China, 2014.

References

1. Dongsheng, L., Sijing, T.: The analysis of the accounting and controlling of sugarcane production cost in Guangxi. *Academic Forum* **08**, 28–43 (2013)

2. Lin, H., Lai, R., Zeng, Z.: Study on the technical specification of sugarcane mechanization harvesting. In: International Academic Conference of Chinese Society for Agricultural Machinery (2012)
3. Chen, C., He, L., Yan, h.: Research on the several agronomic techniques of sugarcane mechanization harvesting. In: The Annual Meeting of the Crop Science Society of China (2014) (in Chinese with English abstract)
4. Oduori, M.F., Sakai, J., Gupta, C P.: Kinematics of revolving-knife disk-type sugarcane basecutter-1. *AMA* **23**(4), 9–15 (1992)
5. Kroes, S., Harris, H.D.: Variation of cutting energies along a sugarcane internode. In: Conference on Engineering in Agriculture and Food Processing. Held University of Queensland, Gatton College, Lawes, Agricultural-Engineering-Australia 25, 3, 55 (1996)
6. Kroes, S., Harris, H.D.: The specific splitting energy of sugarcane. In: Proceedings of the 1998 Conference of the Australian Society of Sugar Cane Technologists held at Ballina, NSW, Australia, pp 349–356 (1998)
7. da Cunha Mello, R., Harris, H.: Cane damage and mass losses for conventional and serrated hasecutter blades. *Proc. Aust. Soc. Soc. Sugar Cane Technol.* **22**, 84–91 (2000)
8. Kroes, S., Harris, H.: Effects of cane harvester base cutter parameters on the quality of cut. *Proc. Aust. Soc. Sugar Stalk Technol.* 169–179 (1994)

Research on the Influence of Changes in Particle Sizes on Simulation Results in the Simulation Test Based on the Discrete Element Method



Xuhong Tan, Jingwei Gao, Cheng Hu, Xiaobo Song, Min Zhang, and Peng Zheng

Abstract Taking sands as the research object, the influences of changes in particle sizes on the simulation results were analyzed in the simulation test based on DEM (Discrete Element Method) using tests and comparative simulation methods, so as to study the influences of particle changes on simulation results in the simulation process using DEM. According to the simulation using particles of varied sizes, the angle of repose is changed marginally with minor errors in the test of the angle of repose. Also, particle size has insignificant influence on the simulation results in the test of the angle of repose. That is to say, the simulation test of the angle of repose is less sensitive to the particle size. However, particle size exerts a remarkable influence on the simulation result in the vane shear test. Specifically, when the particle size is changed, then the maximum torque of the vane will be significantly fluctuated, indicating that the vane shear test is highly sensitive to the particle size. When the magnification factor $m \leq 4$, the variation trend of torque will be consistent with reality. In that case, the particle magnification should be less than 4 times.

Keywords Discrete element method · Simulation test · Particle size change

X. Tan · J. Gao (✉) · C. Hu · X. Song · P. Zheng
College of Aerospace Science and Engineering, National University of Defense Technology,
Changsha 410072, China
e-mail: mmjingwei@163.com

X. Tan
e-mail: tanxuhong1991@sina.com

X. Tan
Troop 66400 of the People's Liberation Army, Beijing 100036, China

M. Zhang
Academy of Hi-Tech Research, Hunan Institute of Traffic Engineering, Hengyang 421000, China

1 Introduction

The discrete element method (DEM) is a calculation model proposed for the study of the large-scale gradual movement of the rock system and a numerical calculation method for solving the non-continuum problem first developed by Cundall, P.A and Strack, O.D.L in the 1970s [1–3]. Since then, DEM has been extensively applied in numerical simulations of a variety of fields such as geotechnical engineering, agricultural engineering, industrial engineering, pharmaceutical engineering, and automotive engineering [4, 5].

Sand featuring scatterer, heterogeneity, and natural variability, is a three-phase system consisting of solid particles, water and gas [6]. And the finite element method (FEM) and DEM are primary methods used for analyzing the simulation of the sand model. To be concrete, FEM is to calculate the force in accordance with the continuum-based concept by taking the object as a continuum. Nevertheless, sand is a non-continuum that is comprised of discrete solid particles. In that case, mechanical properties between sand particles cannot be accurately obtained using FEM. In consequence, DEM has been increasingly applied in sand simulation and modeling. YC Zhou, BD Wright, et al. concluded that the rolling friction coefficient between particles might exert an influence on the value of angle of repose based on the analysis of influencing factors of the angle of repose in the numerical simulation of sand accumulation [7]. Liu D. et al. analyzed the influence of varied particle shapes on the simulation results, and concluded that polyhedral particles can better reflect the realistic shape of rock filling than spherical particles, and also realistically simulate the real strength of the rock filling body in the discrete element numerical simulations on the rock filling body using polyhedral particles and spherical particles, respectively [8]. Liu J. et al. simulated the magnified particles in the meso-simulation of discrete elements based on sands of various grain diameters and revealed that the greater the particle size, the greater the sliding friction between particles through the simulation of particle diameter [9]. Dong Y. studied the influence of particle sizes on the simulation results through scaling up in the numerical simulation process of 3D discrete elements [10]. By summing up previous findings, it is found that most studies have improved simulation speed by magnifying particles, whereas the theoretical basis for particle magnification has been rarely proposed, and there is a lack of separate research on the influence of particle amplification itself on the simulation results.

In response to these deficiencies, this paper first measured the physical properties such as apparent density, moisture content, average particle size, and shape of the sand through experiments. Then, the test of angle of repose and the vane shear test were performed. Next, a simulation model conforming to the real test scale was established using the EDEM software. After that, the influence of magnified simulation particles on the simulation results was studied by comparing the test and the simulation result.

The technical roadmap is presented as follows (Fig. 1).

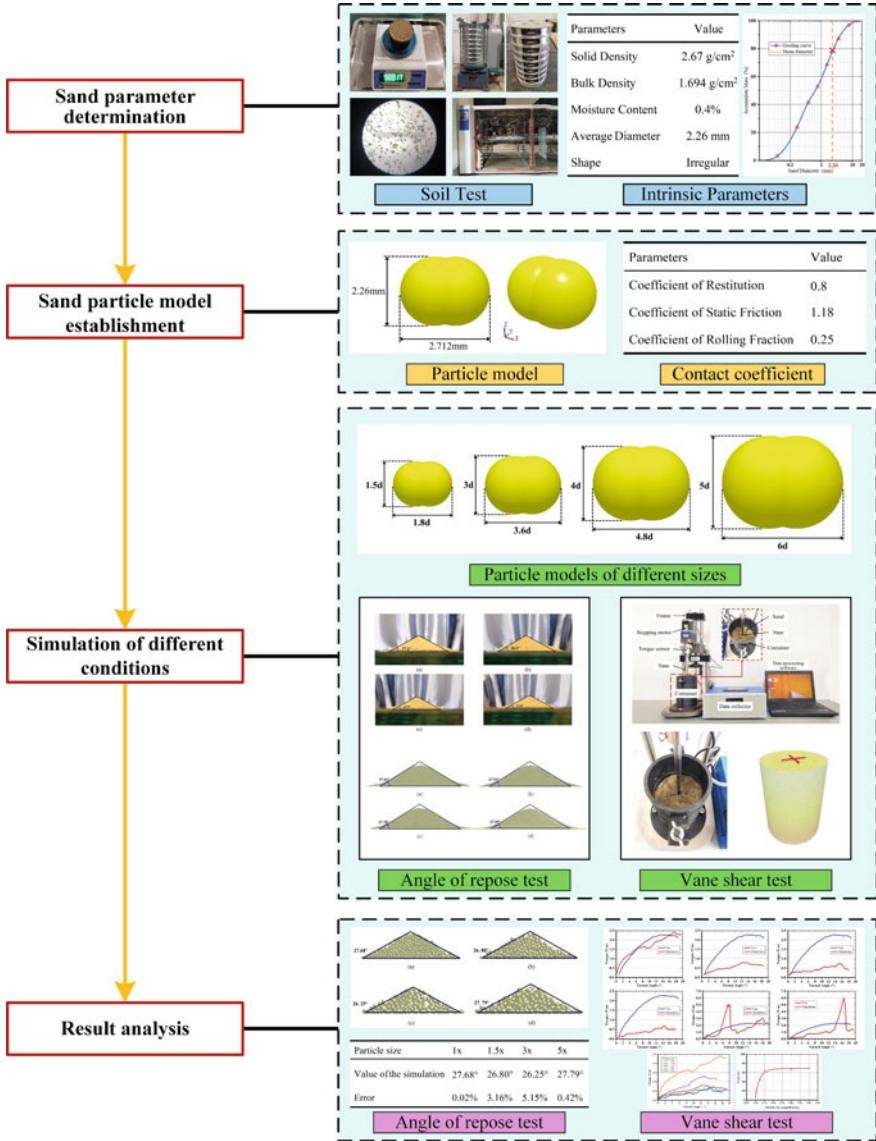


Fig. 1 Technical roadmap

2 Theoretical Analysis

The basic idea behind DEM is to separate the discontinuum into a collection of rigid elements before the iterative calculation of the force and position of the particle based on Newton’s second law equation. And the principle of DEM is the Hertzian-based

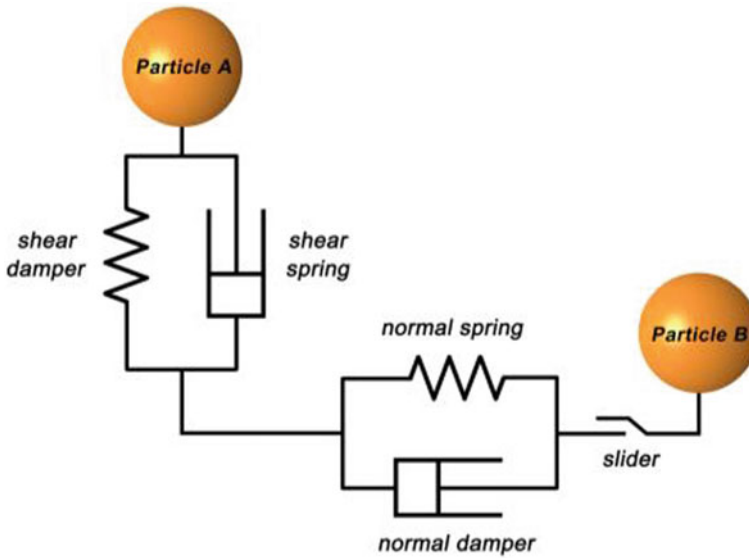


Fig. 2 Vibration model of DEM

normal force contact theory and the tangential force contact theory based on the research result of Mindlin-Deresiewicz. Since damping components can be found in normal force and tangential force, the damping coefficient depend on the coefficient of restitution. Tangential friction conforms to Coulomb's law of friction, and rolling friction is achieved together with the contact-independent directional constant torque model. The normal force and the tangential force between particles can be obtained by calculating the overlap between particles since rigid contact exists between particles [11–14] (Figs. 2 and 3).

3 Test and Model Simulation

Sand, a typical free-flowing material that is suitable for test operations, was taken as the material object of the experimental study. Moreover, sand cannot be simulated in line with its real distribution of particle size in the field of engineering simulation because of its small particle size. In that case, sand was selected as the object of study in this paper.

Fig. 3 Contact model of DEM

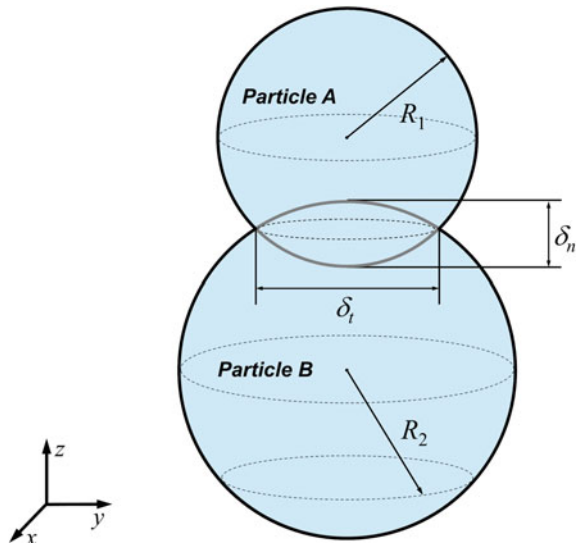


Table 1 Physical parameters of sand

Parameters	Value	Unit
Solid density	2.67	g/cm ³
Bulk density	1.694	g/cm ³
Moisture content	0.4%	–
Average diameter	2.26	mm
Shape	Irregular	–

3.1 Determination of Physical Properties of Sands

Dry sand was sampled under natural conditions for test in order to determine its physical properties parameters and provide a basis for selecting an appropriate simulation contact model. Its physical parameters were obtained through a wide range of soil mechanics tests, including moisture content test, particle analysis test, and accumulation density test (Table 1, Figs. 4, 5 and 6).

3.2 Establishment of the Sand Particle Simulation Model

EDEM software was used as the discrete element simulation software. Simulation parameters for the model should be first defined in the simulation process of sand. By doing so, simulation parameters can be match with the actual parameters. Intrinsic

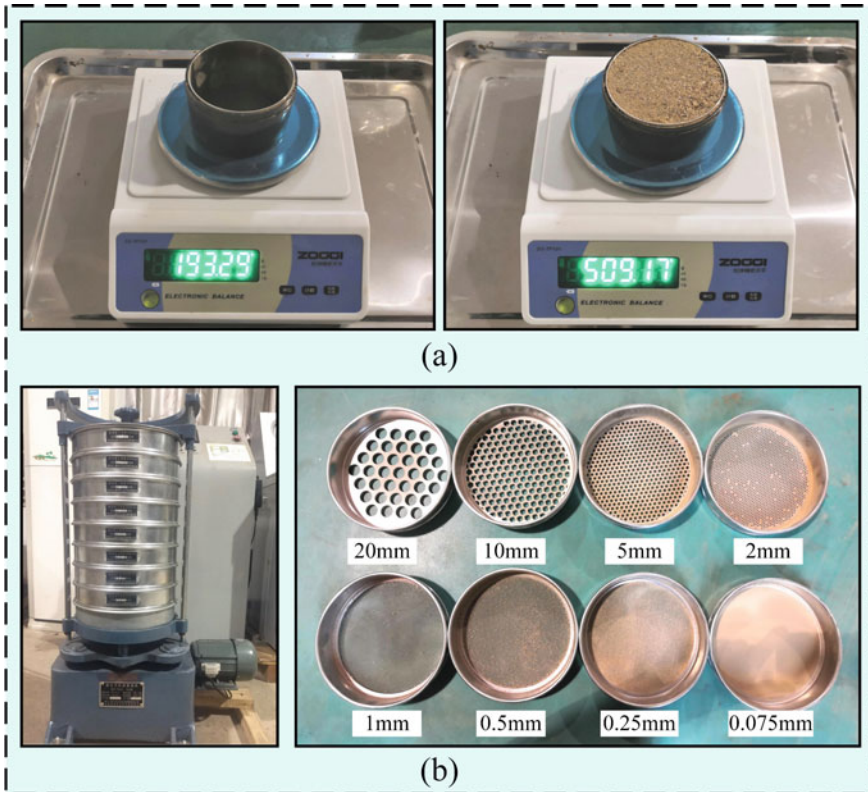


Fig. 4 Soil mechanics test. a Bulk density test, b gradation test

parameters, contact parameters and contact model parameters are major parameters affecting the simulation results.

Among them, intrinsic parameters include particle material density, Poisson's ratio, shear modulus, particle shape, and particle size distribution. The intrinsic parameter of the sand was known based on previous experiments and related literature. Besides, it should be considered that the actual sand shape cannot be characterized in the simulation process due to the characteristics of the discrete element software. Rahul Bharadwaj et al. showed that the use of simplified non-spherical particles in the simulation process has a minor influence on the simulation results for the materials of irregular shapes in related studies [15]. Hence, a double-spherical particle model with the single-sphere diameter of $d = 2.26$ mm and the aspect ratio of 1.2 was constructed in line with the real average particle size measured in the test in the first sand modeling process (Fig. 7).

The coefficient of restitution, static friction coefficient, and rolling friction coefficient are contact parameters between sand particles. According to the DEM model, contact parameters that directly reflect mechanical properties at the macro level exert

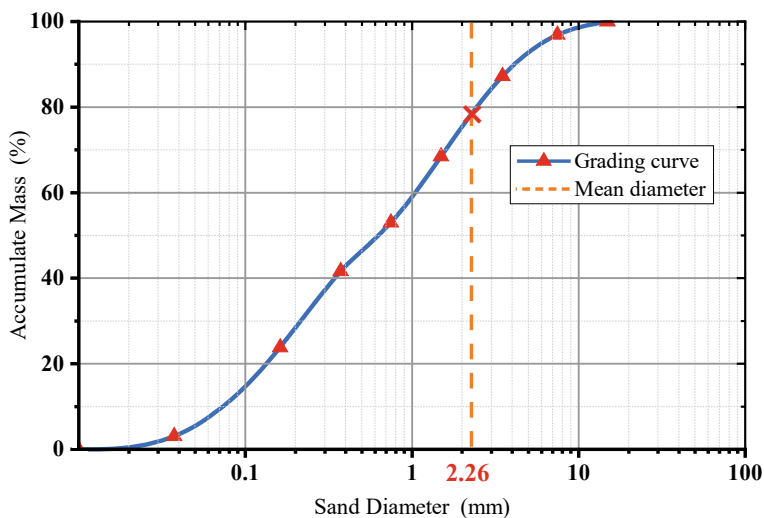


Fig. 5 Sand gradating curve

Fig. 6 Sand grains micrograph

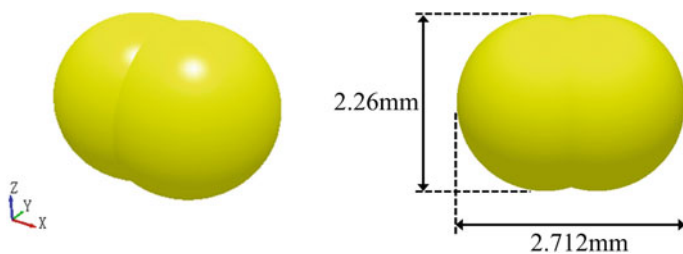
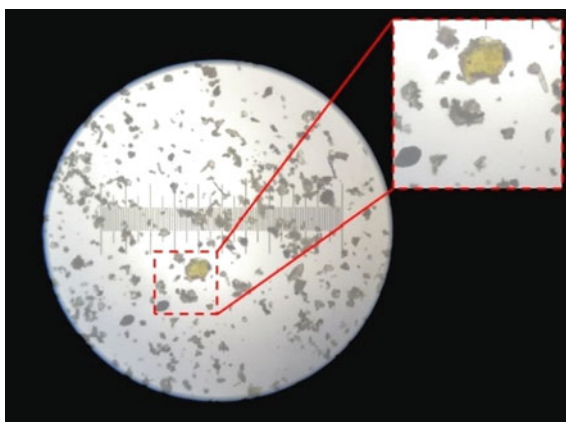


Fig. 7 Simulation particle of Sand grains

Table 2 Contact parameters of the simulation model

Parameters	Value
Coefficient of restitution	0.8
Coefficient of static friction	1.18
Coefficient of rolling fraction	0.25

a direct influence on the simulation results. Contact parameters are determined by referring to related literature [10, 12–18], as shown in Table 2.

The Hertz-Mindlin (no slip) contact model was selected since the micro force such as the liquid bridging force between sand particles can be ignored with the moisture content of sand being 0.4% measured above [19].

3.3 Experiment and Simulation

The test of indoor angle of repose and the vane shear test were conducted for comparison, so as to study the influence of size magnification of simulation particles on the simulation results. The magnification coefficient of the single sphere size of particles in the simulation is set to m (when $m = 1$, no magnification will be performed. It remains the original particle size) as required, and the aspect ratio of the magnification coefficient $1 \leq m \leq 5$ is 1.2 (Fig. 8).

Test and simulation of angle of repose. The test of the indoor angle of repose was performed by lifting a cylindrical material pipeline made of an acrylic pipe with a diameter of 50 mm. At first, 500 g of sand was added to the material pipeline. And then, the material pipeline was lifted at the speed of 1 m/s to make the sand in the pipeline fall freely. After that, the angle of repose was measured when the sand accumulation was stable (Fig. 9). To ensure the accuracy of the test results, contours of the sand accumulation were recorded from two vertical observation directions using cameras in each test. At last, the angle of repose could be determined with the help of the image processing software. After repeating the test for four times (Fig. 10), it can obtain that the average angle of repose of the sand is 27.675° (Table 3).

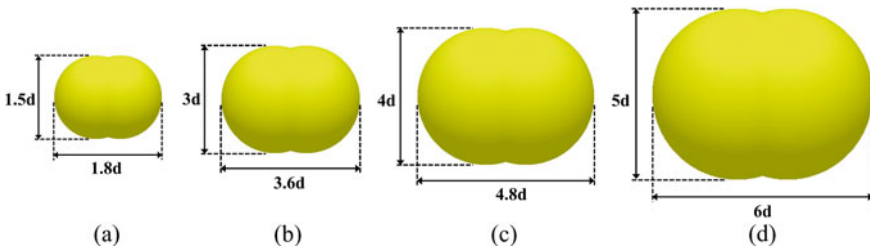


Fig. 8 Different sizes of simulation particles



Fig. 9 Angle of repose test

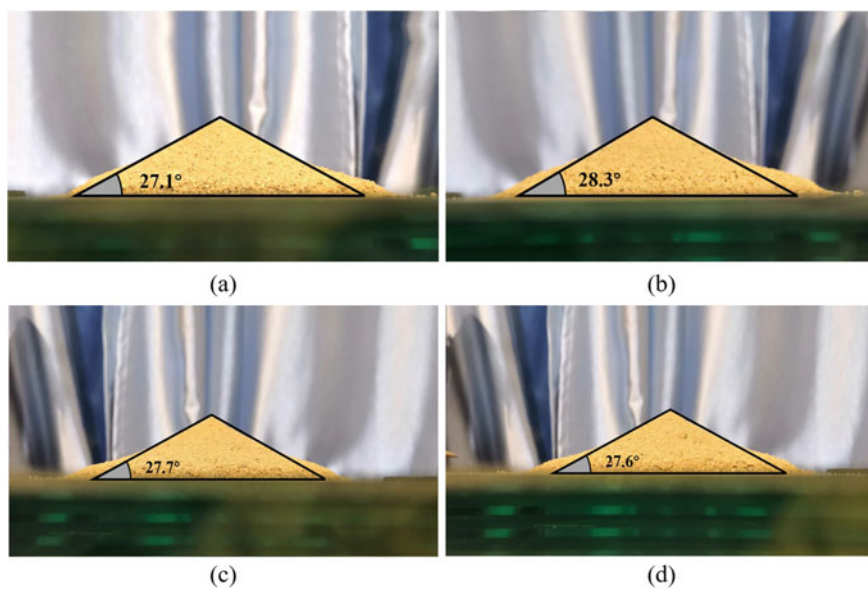


Fig. 10 Angle of repose of sand

Table 3 Result of angle of repose test

Test	(a)	(b)	(c)	(d)
Angle of repose	27.10°	28.30°	27.70°	27.60°
Average	27.675°			

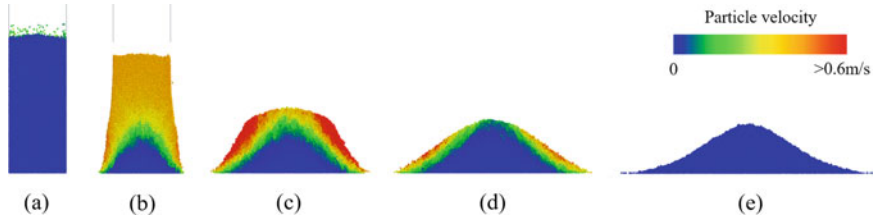


Fig. 11 Scheme of simulation of angle of repose test

In the simulation process of the angle of repose test, particles of specified quality were firstly generated in the material pipeline model. Then the material pipeline model was lifted at a constant speed upon the stable collection of particles. Next, particles flew out of the material pipeline model and accumulated (Fig. 11). The angle of repose was measured after particles accumulation was stable. Intrinsic parameters and contact parameters of simulated particles were set as above. Magnification coefficients of particles were set to $m = 1, 1.5, 3$ and 5 , respectively. At the same time, the lifting speed of the material pipeline and the simulation duration were set to 1 m/s and 2 s , respectively. The angle of repose was measured using the post-processing tool of the software after the simulation is completed, as shown in Fig. 12. Results are shown in Table 4.

In the test of the angle of repose, when the magnification coefficient is within $1 \leq m \leq 5$, a small error will be observed in the simulation result. The maximum value is 5.15% . Apparently, magnifying simulation particles within a certain range has a

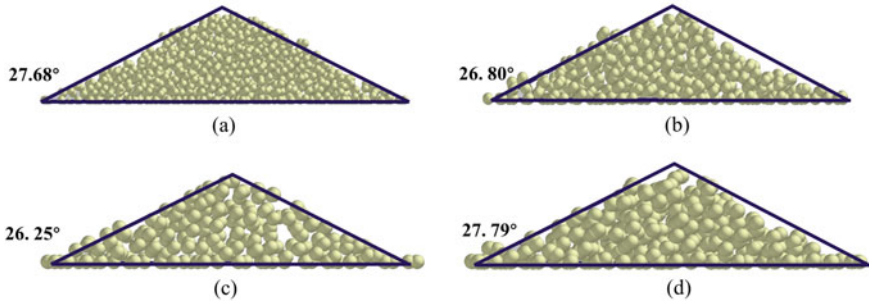


Fig. 12 Simulation results of angles of repose under different particle sizes (magnification). **a** 1 time, **b** 2 times, **c** 3 times, **d** 4 times

Table 4 Simulation result of angles of repose under different particle sizes

Particle size (magnification)	1.0×	1.5×	3.0×	5.0×
Simulation value	27.68°	26.80°	26.25°	27.79°
Error (%)	0.02	3.16	5.15	0.42

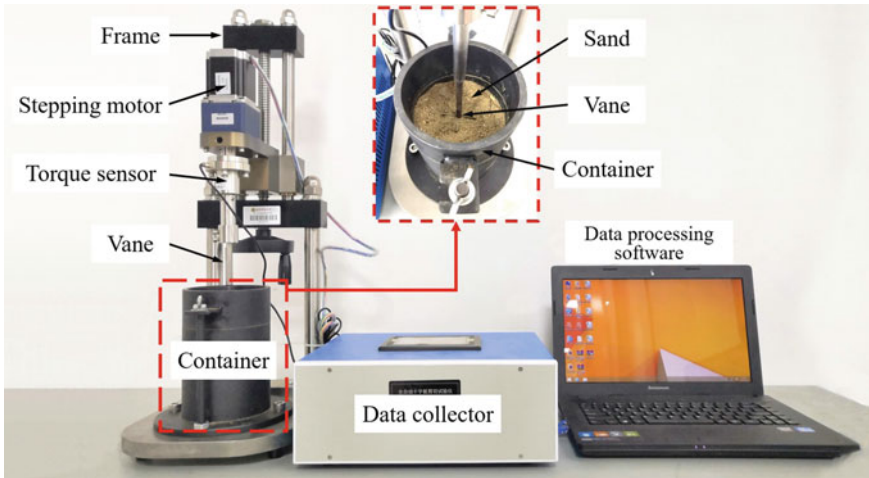


Fig. 13 Vane shear apparatus

small influence on the simulation results. Also, the simulation test of the angle of repose is less sensitive to the diameter of the simulated particle.

Vane shear test. Vane shear test is a test to measure the shear strength of soft soils. The vane shear apparatus used is shown in Fig. 13. Concretely, the stepping motor drives the vane equipped with a torque sensor to rotate at a certain speed. And the torque sensor can measure the real-time torque of the vane and transmits the data to the data collector for recording with the software.

Before the test, 1500 g of sand samples were loaded into a container. At the same time, a vane of the specialization of $\Phi 38.1 \text{ mm} \times h76.2 \text{ mm}$ was selected; the rotational speed was set as $0.1^\circ/\text{s}$; and the test duration was 180 s.

Then, a simulation model was established according to the ratio of 1:1 with the real test. Also, $m = 1, 1.5, 3, 4, 4.5$ and 5 were considered as amplification coefficients of simulated particle sizes, respectively. Relevant parameters are the same as above, and the total mass of particle and motion parameter settings of components are consistent with the real test, as shown in Fig. 14.

Simulation results with different magnification coefficients are compared in Table 5 and Figs. 15 and 16. Evidently, the change in particle size has a significant influence on the simulation results in the simulation of the vane shear test. Besides, the simulation results are highly sensitive to particle size. When the magnification coefficient is $m \leq 4$, the variation trend of the torque curve will be consistent with the actual test curve. When the magnification coefficient is $m > 4$, the variation of the simulation curve will be on the rise as a whole with a sharp increase witnessed at a certain moment. It is greatly different from the variation trend of the real test curve.

When magnification coefficients are $m = 1, 1.5, 3$ and 4, the maximum torques of the vane are 2.41, 0.806, 0.756 and 0.681 Nm. Also, the maximum torque is gradually decreased with the increase of the magnification coefficient. Errors of the simulated



Fig. 14 Vane shear test and simulation model

Table 5 Simulation result of Vane shear test with different particle diameters

Particle size (magnification)	1.0x	1.5x	3.0x	4.0x	4.5x	5.0x
Maximum torque of the text	2.27 N·m					
Value of the simulation (Nm)	2.41	0.806	0.756	0.681	5.98	5.06
Error (%)	6.16	64.5	66.7	70	163.4	122.9

value and the experimental value are 6.16%, 64.5%, 66.7%, and 70% respectively. The variation curve of the error value is presented in Fig. 17. When the magnification coefficient is $1 \leq m \leq 1.5$, the error value will present a fast rate of change. But when $1.5 < m \leq 4$, the rate of changed error value will be slightly on the rise.

4 Conclusions

In this paper, the test of the angle of repose and the vane shear test were performed on the naturally-dried sand to discuss the influences of particle amplification on the simulation results, in response to the deficiencies in particle magnification for the existing discrete element simulation study. Conclusions are made as below:

- (1) In the test of the angle of repose, magnifying simulation particle size within a certain range has a minor influence on the simulation results, and the simulation particle size is less sensitive to the simulation test of the angle of repose.
- (2) In the vane shear test, the change in the simulation particle size has a remarkable influence on the simulation results. Moreover, the change in particle size exerts

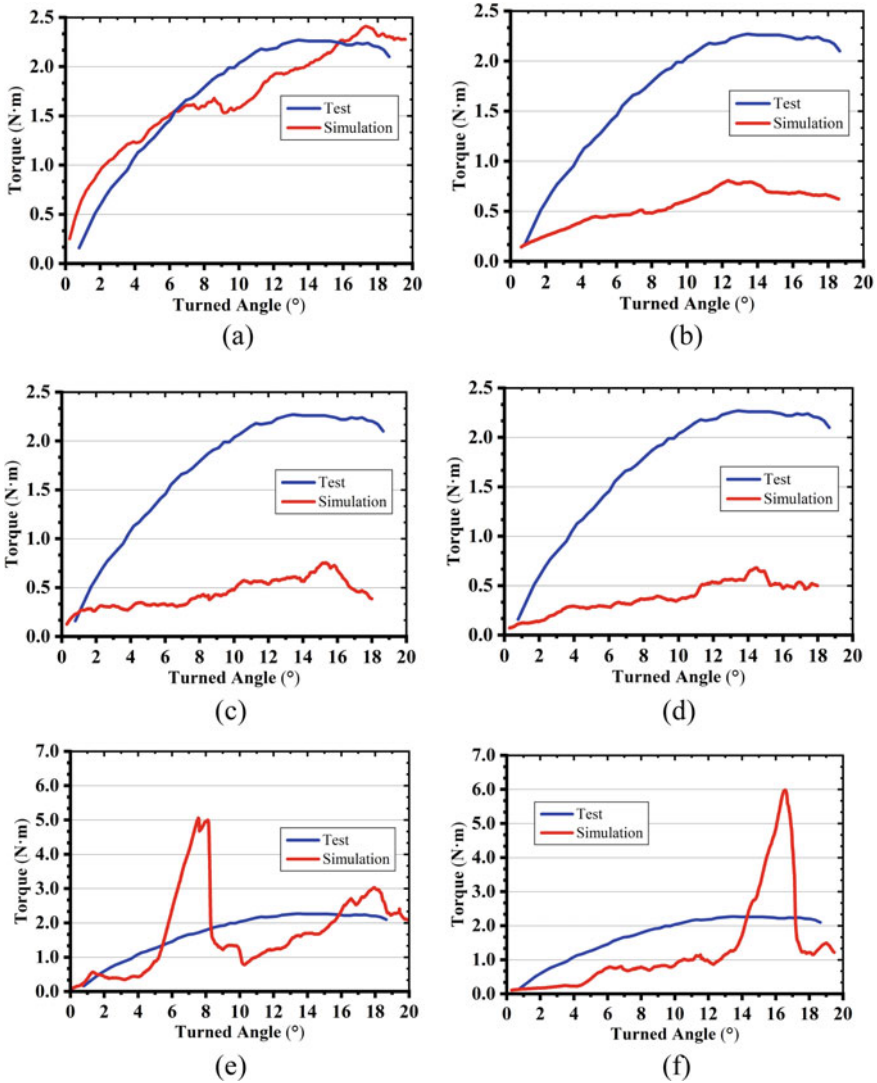


Fig. 15 Results of the vane shear test and simulation at different particle sizes (magnification). **a** 1.0×, **b** 1.5×, **c** 3.0×, **d** 4.0×, **e** 4.5×, **f** 5.0×

a great impact on the friction and the micro force between particles, directly affecting the shear strength of the accumulation of simulation particles.

- (3) Based on the comparison of simulation tests, when particles should be amplified to improve simulation efficiency, the magnification coefficient should be controlled at $m \leq 4$. In the meantime, the simulation parameters should be also re-calibrated using at least two methods. Only by doing so can the simulation

Fig. 16 Comparison of simulation result with different particle sizes

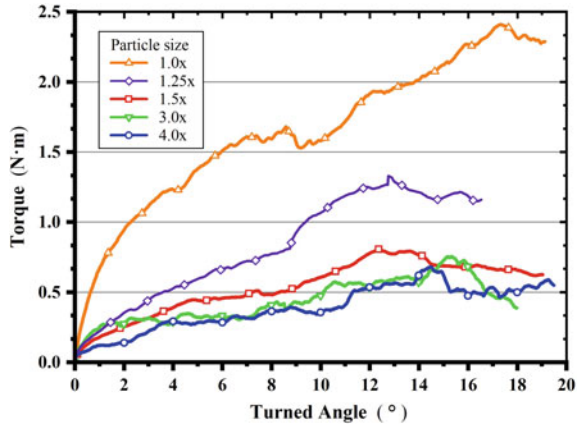
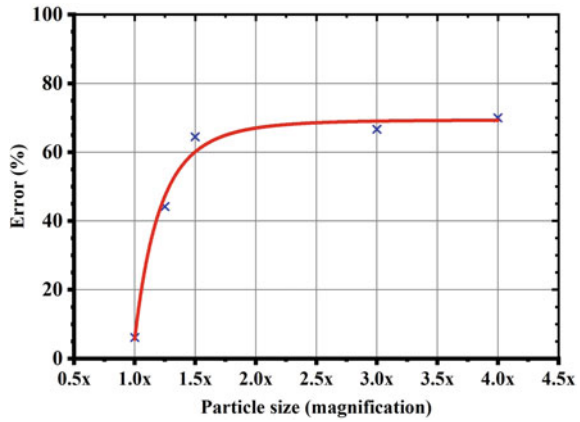


Fig. 17 Error variation curve



be more consistent with the reality, thus raising the accuracy of the simulation results.

References

1. Cundall, P.A., Strack, O.D.L.: A discrete numerical model for granular assemblies. *Géotechnique* **29**(1), 47–65 (1979)
2. Cundall, P.A.: *The Measurement and Analysis of Acceleration in Rock Slopes*. University of Landon (1971)
3. Cundall, P.A.: A computer model for simulation progressive large scale movement in block rock system. *Symposium ISRM* (2), 129–136 (1972)
4. Knuth, M.A., Johnson, J.B., Hopkins, M.A., et al.: Discrete element modeling of a Mars Exploration Rover wheel in granular material. *J. Terramech.* **49**(1), 27–36 (2012)

5. Tsuji, T., Nakagawa, Y., Matsumoto, N., et al.: 3-D DEM simulation of cohesive soil-pushing behavior by bulldozer blade. *J. Terramech.* **49**(1), 37–47 (2012)
6. Zhang, K.G., Liu, S.Y.: *Soil mechanics*, 2nd edn. China Architecture & Building Press, Beijing (2005)
7. Zhou, Y.C., Wright, B.D., Yang, R.Y., et al.: Rolling friction in the dynamic simulation of sandpile formation. *Physica A* **269**(2–4), 536–553 (1999)
8. Liu, D., Xie, T.T., Ma, G., et al.: Numerical simulation of true triaxial test for behavior of rockfill based on grain shape. *Water Resources Power* **29**(9), 68–71 (2011)
9. Liu, J., Sun, T., Zhang, Y.X., et al.: Mesoscopic Simulations by the DEM based on different sand particle diameters. *J. Beijing Univ. Civil Eng. Archit.* **32**(4), 18–22 (2016)
10. Dong, Y.: *Parameter Determination and Numerical Application of Three-dimensional Discrete Element*. Dalian University of Technology (2018)
11. Johnson, K.L., Kendall, K., Roberts, A.D.: Surface energy and the contact of elastic solids. *Proc. R. Soc. Math. Phys. Eng. Sci.* **324**(1558), 301–313 (1971)
12. Zhang, H., Makse, H.: Jamming transition in emulsions and granular materials. *Phys. Rev. E* **72**(1), 011301 (2005)
13. Brilliantov, N., Spahn, F., Hertzsch, J.-M., et al.: Model for collisions in granular gases. *Phys. Rev. E* **53**(5), 5382–5392 (1996)
14. Silbert, L., Ertas, D., Grest, G., et al.: Granular flow down an inclined plane: Bagnold scaling and rheology. *Phys. Rev. E* **64**(5), 51302 (2001)
15. Bharadwaj, R., Ketterhagen, W.R., Hancock, B.C.: Discrete element simulation study of a Freeman powder rheometer. *Chem. Eng. Sci.* **65**(21), 5747–5756 (2010)
16. Roessler, T., Katterfeld, A.: Scaling of the angle of repose test and its influence on the calibration of DEM parameters using upscaled particles. *Powder Technol.* **330**, 58–66 (2018)
17. Hu, C., Gao, J., Diao, J., et al.: Numerical simulation of tire steering on sandy soil based on discrete element method. *AIP Adv.* **11**(1), 015015 (2021)
18. Zeng, H., Xu, W., Zang, M., et al.: Calibration and validation of DEM-FEM model parameters using upscaled particles based on physical experiments and simulations. *Adv. Powder Technol.* **31**(9), 3947–3959 (2020)
19. EDEM: <https://www.edemsimulation.com/resources-learning/ebooks> 2021/3/25

Wind Turbine Condition Monitoring Based on SCADA Data Co-integration Analysis



Chao Zhang, Guanghan Zhao, and Yue Wu

Abstract A wind turbine condition monitoring method based on cointegration analysis is proposed. The co-integration residual obtained by the co-integration process of the SCADA data of the wind turbine is used for monitoring the operation state of the wind turbine. Take the experimental data of a 1.5 MW doubly-fed wind turbine from Jinjie Company in Baotou City, Inner Mongolia, under different environmental and operating conditions, and conduct experiments on the proposed method. The method was tested with known failure cases. The results show that this method can effectively monitor the running status of wind turbines.

Keywords Cointegration analysis · Condition monitoring · Non-stationary · Wind turbine

1 Introduction

Wind turbines are the core equipment for wind power generation. Since wind farms are generally located in areas with complex environments, the generators are greatly affected by the environment during operation. Once they fail, they will cause serious economic losses. Therefore, it is particularly important to carry out the monitoring of the operation status of wind turbines. The existing condition monitoring method is realized by monitoring the mechanical parameters, among which the more widely used parameters are temperature, oil and so on. The SCADA system, that is, the data acquisition and monitoring control system, plays an important role in the field of wind power generation, monitoring parameters including wind speed, rotation speed, temperature, electrical energy, and power. Through the analysis of these data, real-time monitoring and status evaluation of the operating status of the wind turbine can be carried out. However, due to the complex working environment of the wind turbine and the influence of the environment and the operation of the wind turbine itself, the

C. Zhang · G. Zhao (✉) · Y. Wu
College of Mechanical Engineering, Inner Mongolia University of Science and Technology,
Baotou 014010, Inner Mongolia, China
e-mail: 1309527860@qq.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_9

reliability of the SCADA data analysis results of the wind turbine is insufficient, and the fault parameters cannot be accurately found from the data collected by the SCADA system, which will eventually lead to the failure of the wind farm. Effectively grasp the operation level and health status of the wind turbine. Therefore, it is necessary to find a method that can accurately analyze SCADA data to realize the status monitoring of wind turbines [1].

Cointegration theory originated from econometrics. Engle and Granger proposed to analyze the relationship between non-stationary economic variables [2]. This is an effective statistical method to deal with the long-term equilibrium relationship of non-stationary time series. Nowadays, the theory of cointegration is also applied in the field of engineering. The measured signals in the engineering field generally have long-term non-stationary properties, and there may be long-term dynamic equilibrium relationships between these signals, and the co-integration theory can be used to describe the long-term dynamic coordination between non-stationary random processes in the engineering system relationship.

Based on the method research of data trend analysis and process monitoring, this paper proposes a new method of wind turbine condition monitoring based on cointegration analysis [3]. The method of vector regression is used to analyze multiple parameters, and the residuals generated by the nonlinear trend in the data are quickly removed through cointegration calculation. Whether the cointegration residuals obtained by the cointegration analysis of SCADA data is stable or not is used to indicate the operating status of the wind turbine [4]. This method realizes the transition from the analysis of a single process parameter to the automatic interpretation and analysis of a large number of process parameters. Compared with other commonly used data processing techniques such as neural network algorithms, this method has the advantages of simple implementation and small calculation amount [5, 6].

2 Cointegration Theory

2.1 Auto Regression Model

The autoregressive model is the most basic and widely used model in the analysis of random process time series [7]. It describes the linear relationship between the sequence $\{x\}$ at a certain time t and the previous p time sequence, expressed as.

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t \quad (1)$$

Among them, the random sequence $\{\varepsilon_t\}$ is a white noise sequence, and the sequence $\{\varepsilon_t\}$ and the sequence $\{x_k\}(k < t)$ are not correlated. The model (1) is called the p -order autoregressive model, which is denoted as AR.

2.2 Single Whole

In engineering problems, most of the data are non-stationary time series, which often cannot meet the requirements of stability, and it is necessary to transform the non-stationary series into a stationary time series. Single integer definition: If a non-stationary time series $\{x_t\}$ becomes a stationary and reversible stationary time series after d -order difference, and the sequence is still non-stationary after $d - 1$ order difference, then the series is said to have a d -order single Integrity, denoted as $x_t \sim I(d)$.

Therefore, $I(0)$ can represent a stationary sequence, and $I(1)$ can be represented as a first-order single integer. Generally, in engineering problems, the single integer order $d \leq 2$.

2.3 Cointegration

Cointegration describes the long-term equilibrium relationship of the engineering system. It describes the equilibrium relationship of two or more non-stationary time series. Although the mean, variance or covariance of each time series changes with time, the moments of some linear combinations (equilibrium relations) of these series are in some It is immutable at all times. Johansen co-integration test is a multi-variable co-integration test method, which is based on the multi-variable unsteady VAR model [8]. The VAR model is established before the test. Its mathematical expression is:

$$y_t = \beta_1 y_t + \beta_1 y_t + \dots + \beta_1 y_t + c + \varepsilon_t, t = 1, 2, \dots, k \tag{2}$$

Where C -intercept; ε -white noise; β - $n \times n$ parameter matrix; $y_t = (y_1, y_2, \dots, y_n)^T, y_{t-1} = (y_{1-1}, y_{2-1}, \dots, y_{n-1})^T, \dots, y_{t-p} = (y_{1-p}, y_{2-p}, \dots, y_{n-p})^T$, Is an n -dimensional non-stationary I vector, p is the lag term. In addition, the cointegration relationship can be extended to reintegration:

$$\beta^T y_t \begin{Bmatrix} \beta_1^T y_t \\ \beta_2^T y_t \\ \vdots \\ \beta_r^T y_t \end{Bmatrix} + \begin{Bmatrix} c_1 \\ c_2 \\ \vdots \\ c_r \end{Bmatrix} = \begin{Bmatrix} \xi_{1t} \\ \xi_{2t} \\ \vdots \\ \xi_{rt} \end{Bmatrix} \tag{3}$$

The residual ξ is obtained as the condition monitoring model.

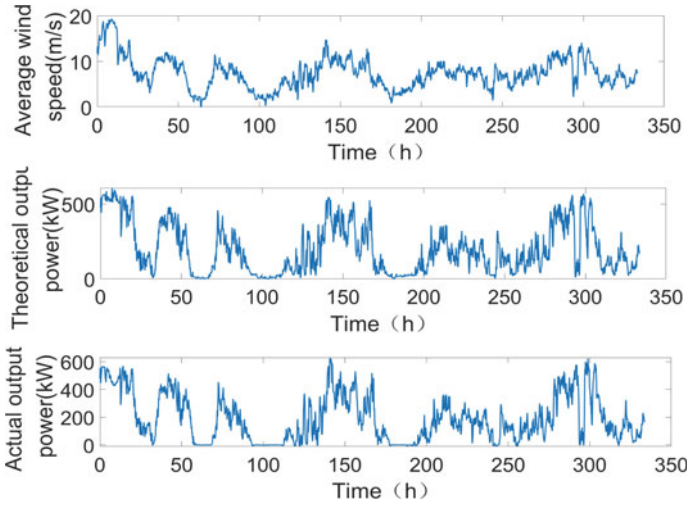


Fig. 1 Wind turbine SCADA data

3 Experiment Analysis

3.1 Experimental Data

The SCADA data of 1.5 MW doubly-fed wind generator from Baotou Jinjie Company is selected as the research object. The unit has a rated wind speed of 11 m/s, a cut-in wind speed of 3.0 m/s, and a cut-out wind speed of 15 m/s. The experiment selects the average wind speed, theoretical output power, and actual output power of the wind turbine in normal operation for co-integration analysis, as shown in Fig. 1.

3.2 Cointegration Model Establishment

Using the method described, a co-integration model is established for the data in Fig. 1, and x_{1t} , x_{2t} , and x_{3t} are the average wind speed, the theoretical output power and the actual output power. Perform a first-order difference on the selected data, and then perform ADF inspection on the original data and the data after the first-order difference. The test results are shown in Table 1.

When the ADF test value in Table 1 is less than the critical value, it means that the time series is a stationary series. When it is greater than the critical value, it means that the time series is non-stationary. It is a stationary variable, which conforms to the definition of single integration, and the cointegration analysis can be continued. Use the above wind turbine data and ordinary least squares regression to estimate Eq. (3) to establish a cointegration model:

Table 1 ADF inspection result

Parameter	ADF inspection	1% threshold	Results evaluation
x_{1t}	-1.07	-2.56	Non-stationary
Δx_{1t}	-37.78	-2.56	Smooth
x_{2t}	-1.53	-2.56	Non-stationary
Δx_{2t}	-37.23	-2.56	Smooth
x_{3t}	-2.14	-2.56	Non-stationary
Δx_{3t}	-24.86	-2.56	Smooth

$$\xi_t = x_{1t} - 0.071285 * x_{2t} + 0.047723 * x_{3t} - 2.608122 \tag{4}$$

Calculate the standard deviation $\sigma = 2.7$ from the above experimental data and the cointegration model Eq. (4). Perform ADF test on the residual. If the test result is a stationary time series, it means that the theoretical useful power and the co-integration model established by the motor power generation have a co-integration relationship. If it is a non-stationary series, it means that the model has failed to establish and there is no between variables. Cointegration. ADF test is performed on the residual ξ_t . Table 2 shows the test results.

The value of residual ξ_t in Table 2 after the ADF test is less than the critical value, indicating that residual ξ_t is a stationary time series and meets the cointegration condition. Since ξ_t conforms to the Gaussian distribution, $\pm 3\sigma$ is selected as the threshold according to the characteristics of high probability events in probability theory. Figure 2 shows the residual sequence of the wind turbine in a healthy state.

As shown in Fig. 2, the residuals in the healthy state are all within the threshold range, and the status of the wind turbine is judged according to whether the residuals are all outside the threshold range.

Table 2 ADF inspection result

Parameter	ADF inspection	1% threshold	Results evaluation
ξ_t	-12.55	-2.57	Smooth

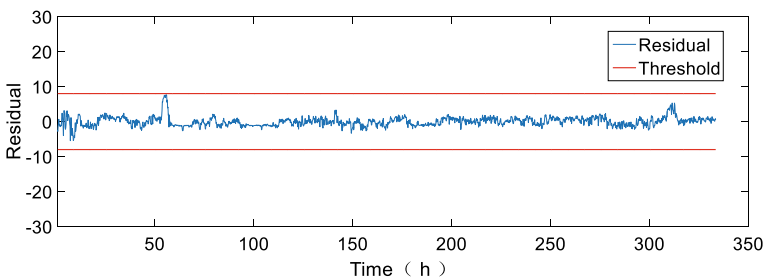


Fig. 2 Cointegration residuals under healthy conditions

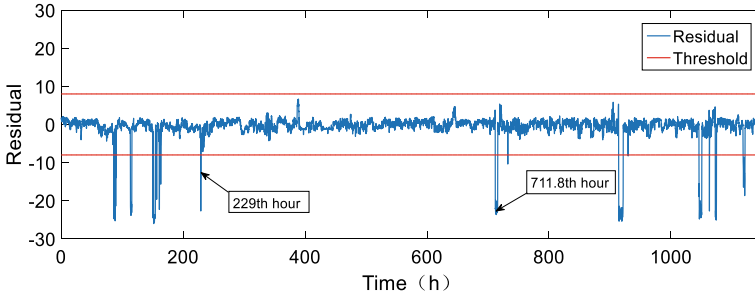


Fig. 3 SCADA data cointegration residual

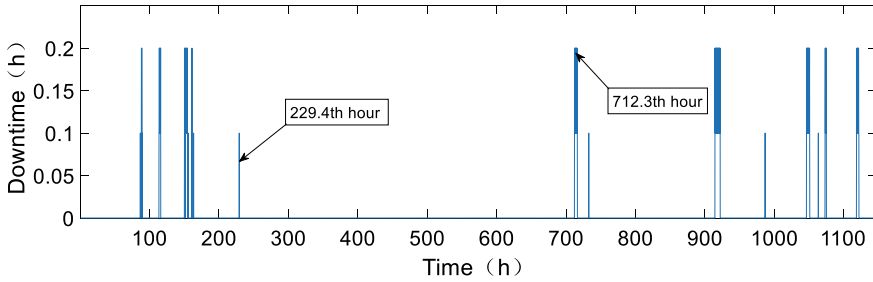


Fig. 4 Wind turbine failure time

3.3 Cointegration Model Verification

Take the SCADA data of the same wind turbine running for 1147 h with faults to verify the model (4), and substitute the data into the cointegration model (4) for verification. The result is shown in Figs. 3, and 4 shows the fault in the wind turbine detection report. Time of occurrence.

Figure 4 shows the fault occurrence time obtained in the wind turbine inspection report. A total of 14 faults have occurred. In Fig. 3, the residual error exceeds the threshold for 14 times, which corresponds to the fault detection report. Among them, the 229th hour and 711.8th hour were earlier than the time of failure in the failure report, and the alarm was about 30 min in advance.

4 Conclusion

A wind turbine condition monitoring method based on wind turbine SCADA data co-integration analysis is proposed. The residual error model is calculated using the wind turbine SCADA data co-integration process, and a known 1.5 MW doubly-fed wind turbine fault is used. The data is used to test the residual model. The results

show that this method can effectively monitor whether the wind turbine operating state is abnormal through the deviation degree of the residual error. Compared with other commonly used data processing techniques and neural network algorithms, this method has the advantages of simple implementation and low cost.

Acknowledgements The authors are grateful to the support from the National Natural Science Foundation of China (No. 51965052) and Inner Mongolia Autonomous Region Science and Technology Plan Project (No. 2018KG007).

References

1. Dao, P.-B., Staszewski W.-J., Tomasz, B., et al.: Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data. *Renew. Energy*, **116**, 107–122
2. Yu, W., Chunhui, Z., Biao, H.: Recursive cointegration analytics for adaptive monitoring of nonstationary industrial processes with both static and dynamic variations. *J. Process Cont.* **92**, 319–332 (2020)
3. Sun, P., Jian, L., Caisheng, W., et al.: A generalized model for wind turbine anomaly identification based on SCADA data. *Appl. Energy*, **168**, 550–567 (2016)
4. Yu, P.Y.: In: *Research on the Application of Co-integration Theory in the Condition Monitoring and Fault Diagnosis of Engineering System*. Nanjing University of Aeronautics and Astronautics (2005)
5. Hameed, Z., Hong, Y.S., Cho, Y.M., et al.: Condition monitoring and fault detection of wind turbines and related algorithms: a review. *Renew. Sustain. Energy Rev.* **13**(1), 1–39 (2009)
6. Márquez, F.P.G., Tobias, A.M., Pérez, J.M.P., et al.: Condition monitoring of wind turbines: techniques and methods. *Renew. Energy* **46**(none), 169–178 (2012)
7. Shi, H., Keith, W., Cross, E.-J.: A cointegration approach for heteroscedastic data based on a time series decomposition: an application to structural health monitoring. *Mech. Syst. Sig. Proc.* **120**, 16–31 (2019)
8. Fan, L.U.: In: *Application Research on Condition Monitoring Method of Complex Dynamic Engineering System Based on Co-integration Theory*. Nanjing University of Aeronautics and Astronautics (2010)

Improvement and Application of YOLOv3 for Smartphone Glass Cover Defect Detection



Yuan Cheng, Jigang Wu, Jun Shaov, and Deqiang Yang

Abstract Smartphone glass covers defects detected by human, which is inefficiency, high costs, low detection accuracy and labour intensive, while the automatic detection methods based on traditional machine vision is poor flexibility, low yield and poor generalisation capability. Therefore, this paper introduces YOLO (You Only Look Once) v3 to smartphone glass cover defects for the first time. The YOLOv3 algorithm was improved for the actual characteristics and specific requirements of defect detection. First of all, the channel attention mechanism SENet (Squeeze and Excitation Networks) was added to the feature extraction network to detect inconspicuous defect features. Moreover, a 104×104 scale detection layer was added to the YOLOv3 detection network to solve the problem of multi-scale defects. Finally, the scaling factor coefficient of the BN (Batch Normalization) layer in the convolutional network is used as the important factor for model pruning to improve the defect detection speed. The improved YOLOv3 algorithm is applied to smartphone glass cover defect detection, and a high accuracy and high detection speed method for smartphone glass cover defects is proposed. 15,914 production site images covering four types of defects, including chipped edges, pits point, soiling and scratches, were obtained from smartphone glass cover manufacturers, 14,321 were annotated as the training set and 1593 were used as the test set to compare and analyse the proposed method and the original YOLOv3 algorithm in this paper. These experiments showed that the mAP (mean average precision) of the detection was 81.0% and the detection speed was 43.1 sheets/s. Compared to the original YOLOv3 algorithm, the mAP of the detection increased by 3% and the detection speed increased by 6.7 frames/s, which meets the need for high precision and efficient detection of defects in the industrial production of smartphone glass covers.

Keywords Mobile phone glass covers · YOLOv3 · Channel attention mechanism · Model pruning

Y. Cheng (✉) · J. Wu · J. Shaov · D. Yang
Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Hunan Province, Xiangtan 411201, China
e-mail: chenghnust@163.com

1 Introduction

Smartphone screen is the core key component of human–computer interaction, generally consisting of three parts: display module, touch module and glass covers, the glass cover is located in the outermost layer of the screen, which is the solid shell and touch medium of the screen. With the rapid development of artificial intelligence and the arrival of the 5G era, smartphones have become a necessary tool [1], and people’s quality requirements for smartphones are getting higher and higher. Smartphone glass covers inevitably produce all kinds of defects during the production process, such as chipped edges, scratches, soiling, pits point and so on. In order to meet the high quality requirements of users for smartphones, manufacturers must carry out 100% quality checks on smartphone glass covers to a high standard. At this stage of mass production, the recognition system depends on a manual inspection with the aid of such tools as bright lights and magnifying glasses. Restricted by human subjective awareness and experience, this inspection method is characterized by inefficiency, high costs, high false detection rates and labour intensive. Therefore, it is important to study the method of detecting defects in mobile phone glass covers to replace manual labour.

Smartphone glass cover defect detection puts forward the following requirements for automatic detection methods: (1) good flexibility of the detection method, which can adapted to various types of defects; (2) strong generalization ability of the detection method, which can adapted to the characteristics of defect features such as inconspicuous and multi-scale; (3) good real-time detection method, which can adapt to the beat of mass production; (4) high detection accuracy, which can completely replace manual detection. Currently, the defect detection of the smartphone cover glass is investigated with differential image method [2, 3], background elimination method [4] and threshold segmentation method [5, 6] in machine vision. The traditional methods generally only detect defects of a certain type or defects with periodic textures, which cannot meet the requirements of flexible inspection. At the same time, these methods are heavily influenced by noise, resulting in poor detection accuracy. In recent years, deep learning-based target detection algorithms [7–10] have made significant improvements in detection accuracy and efficiency relative to traditional methods by building a variety of different network structures, paired with the use of powerful training algorithms to adaptively learn the representation of high-level semantic information in images [11]. In the study of surface defect detection, the YOLOv3 [12] algorithm has shown better detection accuracy and detection speed. Zhang Guangshi et al. [13] used the YOLOv3 algorithm to detect smear marks and missing defects in gears. Weigang [14] and others used the YOLOv3 algorithm to detect defects such as pressed-in iron oxide, patches and cracks on the surface of strip steel. Hongcai et al. [15] used the YOLOv3 algorithm on pharmaceutical glass bottle defect detection and was able to effectively detect defects such as tube end residue, gas lines, bubbles, scratches, stains and stones on glass bottles. Although the YOLO v3 algorithm can provide flexible detection of surface defects, it requires distinctive defect features and a small scale span, and further improvements are needed in

terms of real-time detection. There is no research on mobile phone glass cover defect detection using the YOLOv3 algorithm.

This article introduces YOLOv3 to the defect of the smartphone glass cover for the first time. In view of the actual characteristics and specific requirements of defect detection, the YOLOv3 algorithm is improved. The channel attention mechanism SENet [16] is added to the feature extraction network to solve the problem of unobvious defect features, a 104×104 scale detection layer was added to the YOLOv3 detection network to solve the problem of multi-scale defects, and the scaling factor coefficient of the BN (Batch Normalization) layer of the convolutional network is used as the importance factor for model pruning to improve the defect detection speed. On this basis, the improved YOLO v3 algorithm is applied to the defects of the smart phone glass cover, and a high-precision and high-speed detection method for the defects of the smart phone glass cover is proposed. From the smartphone glass cover manufacturer, 15,914 pictures of the production site covering 4 types of defects such as chipping, pits, dirt and scratches were obtained. 14,321 pictures were marked as training sets and 1593 pictures were used as test sets. The proposed method and the original YOLOv3 algorithm are compared and analyzed.

2 Introduction of YOLOv3

2.1 YOLOv3 Detection Principle

YOLOv3 was proposed by Redmon in 2018, the algorithm works by dividing the image containing the detected target into a $S \times S$ grid, with the width and height of the grid noted as c_x, c_y . When the centre of the target object falls into a grid cell, the coordinates of the relative centroid related to the upper left corner of the grid ($\sigma(t_x), \sigma(t_y)$), as well as the relative width t_w and relative height t_h , would be output by the grid. As a result, the final target prediction frame can be obtained with the actual position, including width and height of the grid [17], which as shown in Fig. 1.

In the figure, the red box is the actual prediction box and the dashed box is the priori bounding box. The center point coordinates (b_x, b_y) , width b_w and height b_h of the prediction box can be obtained by the arithmetic of YOLOv3, where:

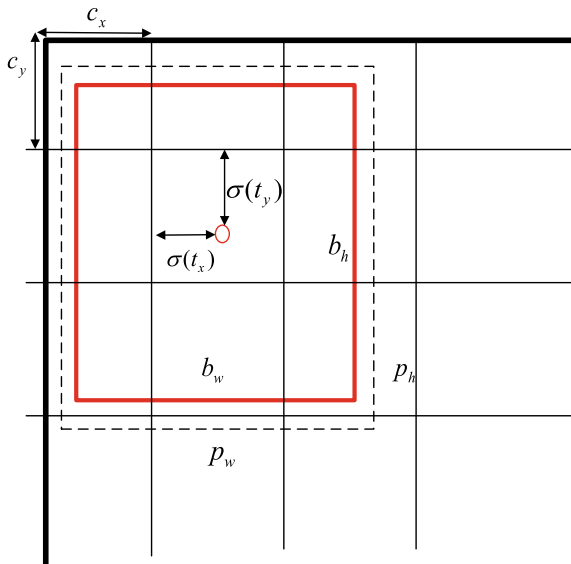
$$b_x = c_x + \sigma(t_x) \quad (1)$$

$$b_y = c_y + \sigma(t_y) \quad (2)$$

$$b_w = p_w e^{t_w} \quad (3)$$

$$b_h = p_h e^{t_h} \quad (4)$$

Fig. 1 Schematic diagram of the YOLOv3 prediction box



2.2 YOLOv3 Network Framework

The darknet-53 network is used to the feature extraction network of YOLOv3, and its network structure is shown in Fig. 2.

During the detection process, three feature maps with different scales would be generated by YOLOv3, whose sizes are 13×13 , 26×26 and 52×52 , with each feature map corresponding to a different field of perception and priori bounding box. The feature map with the size of 13×13 has a larger field of perception and the priori bounding box is relatively large, making it suitable for larger objects. The objects of medium size can be detected by the feature map with the size of 26×26 . The feature map with the size of 52×52 has a smaller field of perception and the corresponding a priori frame is relatively small, which is suitable for the detection of small objects.

3 Improvements in YOLOv3

3.1 Channel Attention Mechanism SENet

The importance level of each channel is not considered in the feature extraction process of YOLOv3, which would induce the poor extraction of useful information. Therefore, the channel attention mechanism SENet is incorporated with a view to improving the feature extraction network of YOLOv3 algorithm.

	Type	Filters	Size	Input	Output
1×	Convolutional	32	3×3/1	416×416×3	416×416×32
	Convolutional	64	3×3/2	416×416×32	208×208×64
	Convolutional	32	1×1/1	208×208×64	208×208×32
	Convolutional	64	3×3/1	208×208×32	208×208×64
	Residual				208×208×64
2×	Convolutional	128	3×3/2	208×208×64	104×104×128
	Convolutional	64	1×1/1	104×104×128	104×104×64
	Convolutional	128	3×3/1	104×104×64	104×104×128
	Residual				104×104×128
	Convolutional	256	3×3/2	104×104×128	52×52×256
8×	Convolutional	128	1×1/1	52×52×256	52×52×128
	Convolutional	256	3×3/1	52×52×128	52×52×256
	Residual				52×52×256
	Convolutional	512	3×3/2	52×52×256	26×26×512
	Convolutional	128	1×1/1	26×26×512	26×26×128
8×	Convolutional	512	3×3/1	26×26×128	26×26×512
	Residual				26×26×512
	Convolutional	1024	3×3/2	26×26×512	13×13×1024
	Convolutional	512	1×1/1	13×13×1024	13×13×512
	Convolutional	1024	3×3/1	13×13×512	13×13×1024
4×	Residual				26×26×1024
	Avgpool		Global		
	Connected		1000		
	Softmax				

Fig. 2 Structure model of the darknet-53 network

The convolutional feature channel interrelationship is adopted by SENet for modelling, and channel responses in specific layers of the convolutional neural network are reassigned to enhance the extraction of useful information. Three main components, namely Squeeze, Excitation and Weight Assignment, are contained in the module, with the basic structure shown in Fig. 3.

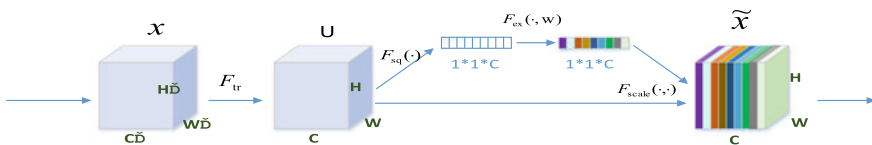


Fig. 3 Basic structure of the SENet

The channel attention mechanism initially performs a conversion operation, which is represented as a convolution operation in practice. In the conversion operation, the input features are x , the output features are U , and the convolution kernel is V .

Squeeze operation encodes the feature map, which is obtained from the conversion operation, compressing the two-dimensional feature map on each channel into a real number with a global perceptual field. The number, which is obtained from the global average pooling formula, represents the original weight of the channel, which can be calculated as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (5)$$

The obtained channel raw weights can be normalized by the excitation operation with a multilayer perceptron containing multiple layers, which is composed of a fully connected layer, ReLU activation function, a fully connected layer and Sigmoid activation function. The final weights of each channel can be expressed as s_c , which can be calculated as follows:

$$s_c = F_{ex}(z, w) = \sigma(g(z, w)) = \sigma(w_2 \delta(w_1 z)) \quad (6)$$

where: δ is the ReLU activation function and σ is the Sigmoid activation function.

The weight assignment operation assigns weights s_c to each channel in the output feature map U , after the conversion operation obtains the final output \tilde{x} , which can be calculated as follows:

$$\tilde{x} = F_{scale}(u_c, s_c) = u_c \otimes s_c \quad (7)$$

\otimes denotes element-by-element multiplication, and SENet enables the assignment of weights to channels in the above manner.

3.2 Improvement of Feature Detection Network

Smaller defects are generated in the manufacture process of the mobile phone glass covers. The smallest perceptual field of YOLOv3 corresponds to the feature map with the size of 52×52 , which is obtained by downsampling the input image with a factor of 8. Therefore, YOLOv3 has a poor performance in detecting targets with pixels within the range of 8×8 .

In this paper, the detection network of the YOLOv3 algorithm has been upgraded, so that its ability to detect small targets can be enhanced. The improved detection layer is based on the original feature maps and continues to perform 4-fold downsample to obtain 104×104 scale of feature maps. Combined with the channel attention module,

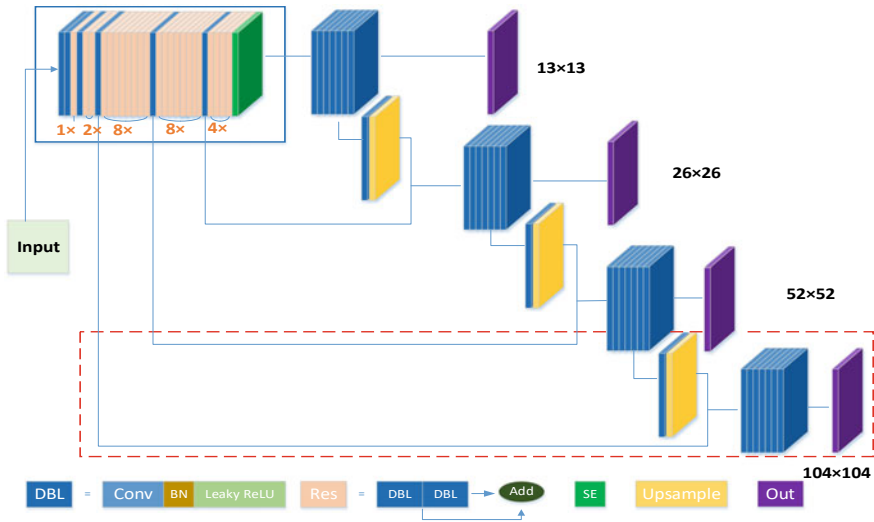


Fig. 4 YOLOv3-improve1 detection network

the improved YOLOv3 network can be obtained and denoted as YOLOv3-improve1, as shown in Fig. 4.

In the figure, Conv as a 5-layers convolutional layer, is composed of 3×3 and 1×1 convolutional with different convolutional kernel sizes. The solid blue boxes refer to the feature extraction network of YOLOv3 and the red dashed boxes refer to the added detection layers. There are 4 dimensional feature maps for the improved YOLOv3, namely 13×13 , 26×26 , 52×52 and 104×104 , each of which is assigned with 12 anchor boxes in descending order. The feature map 104×104 is obtained after the network is improved, which combines the deep information contained in the 109th layer of the network and the shallow information in the 11th layer, thereby providing a further improvement in small target detection.

3.3 Model Pruning

In the industrial manufacture, a certain detection speed shall be achieved during the detection process with the aim of ensuring the balance in the assembly line. The parameters of YOLOv3 are large and computationally intensive, and the computing power has been restricted by the computer terminals on industrial sites. Therefore, the model operations shall be reduced and the detection speed shall be increased with guaranteed detecting accuracy. In this paper, the layer pruning and channel pruning methods proposed by Liu [18] et al. have been adopted.

In network channel pruning, the scaling factor coefficient γ of the BN layer in the convolutional network is regarded as the important factor. When γ is smaller,

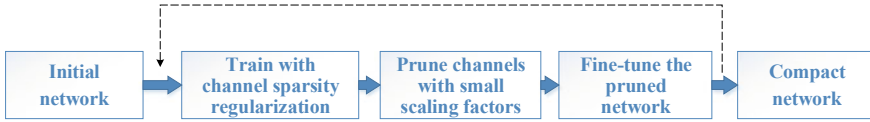


Fig. 5 Diagram of the pruning process of the YOLOv3 model

the corresponding channel is less important and can be pruned. The objective optimisation function of the algorithm as a whole is as follows:

$$L = \sum_{(x,y)} l(f(x, W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \quad (8)$$

where: the first term refers to the model prediction loss and the second term refers to the canonical term about γ . γ is a hyperparameter used for weighing the two terms, generally assigned as $1e-4$ or $1e-5$, $g(*)$ with the expression $g(s) = |s|$, and the L1 paradigm. The overall pruning process is shown in Fig. 5.

4 Experiments and Results Analysis

4.1 Experimental Platform

The experimental platform is Supermicro infreesys server, with operating system: Ubuntu 18.04LTS, CPU: Intel W2123, memory: 32G, graphics card: NVIDIA Geforce RTX2080Ti $\times 2$, and video memory: 16 GB $\times 2$. Deep learning framework: Pytorch.

The dataset used in this experiment was shot at the production site, the types of defects were divided into four categories: chipped, pit point, scratching and soiling, with the specific defect images shown in Fig. 6.

A total of 15,914 images were collected in this dataset, and they were marked with labelling software. The dataset for this study was generated according to the

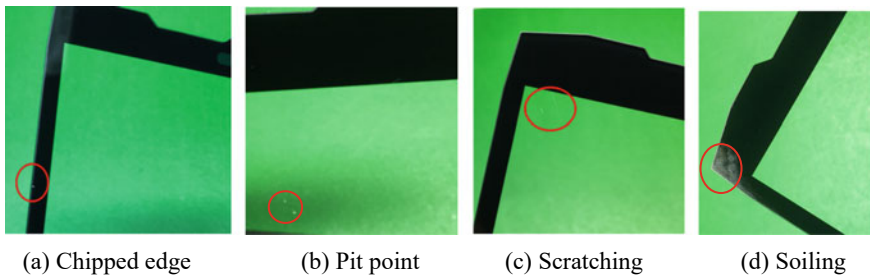


Fig. 6 Defect map of the mobile phone glass covers

Table 1 Priori bounding boxes sizes for each algorithm

Algorithms	The size of priori bounding boxes
YOLOv3	8,11 11,13 9,17 13,12 11,16 15,17 14,21 35,13 39,20
YOLOv3-improve1	8,11 11,13 9,17 13,12 11,16 15,17 14,21 35,13 39,20 16,50 23,39 28,58
YOLOv3-improve2	8,11 11,13 9,17 13,12 11,16 15,17 14,21 35,13 39,20 16,50 23,39 28,58

VOC dataset format required by YOLOv3. The dataset was divided into the training dataset and the test dataset in the ratio of 9:1, with a total of 14,321 photos in the training dataset and 1593 photos in the test dataset.

The algorithm obtained by model pruning based on YOLOv3-improve1 is named YOLOv3-improve2. The dataset was trained and tested with YOLOv3, YOLOv3-improve1 and YOLOv3-improve2 algorithms. The number of samples per batch was set to 16 with subdivision = 8. The input image size was set to $416 \times 416 \times 3$, where 3 was the number of image channels. Besides, the momentum was set to 0.9. The YOLOv3 algorithm contained 3 detection layers, with each layer assigned with 3 priori bounding boxes, and 9 priori bounding boxes were required. In contrast, the YOLOv3-improve1 and YOLOv3-improve2 algorithms contained 4 detection layers, with 3 priori bounding boxes per layer, 12 priori bounding boxes were required, besides, their priori bounding boxes were equal in size. According to the k-means clustering algorithm, the relevant of priori bounding box information is clustered as shown in Table 1.

4.2 Experimental Results

The experiments were conducted on three algorithms with training epochs of 500, and the Adam method was adopted as the parameter optimization method. The bounding box loss values, coordinate loss values, classification loss values, and confidence loss values were included in the trained loss values of the three algorithms, with the total loss value comparison curves of the three algorithms shown in Fig. 7.

In the above figure, the loss value of the YOLOv3-improve2 algorithm is obtained during fine-tuning training, so its initial loss values are relatively low compared to the other two algorithms; while, the trend is similar to the other two algorithms during the training process. During the training process, it can be seen that the YOLOv3-improve1 algorithm and the YOLOv3-improve2 algorithm have approximately similar loss values, and the decline process is gentle, while the YOLOv3 algorithm has an oscillating situation, with large oscillation fluctuations. From the figure, it can be seen that the YOLOv3-improve1 and YOLOv3-improve2 algorithms have overall lower loss values than the YOLOv3 algorithm during the training process. According to the above analysis, it can be concluded that the improved algorithm is more effective than YOLOv3.

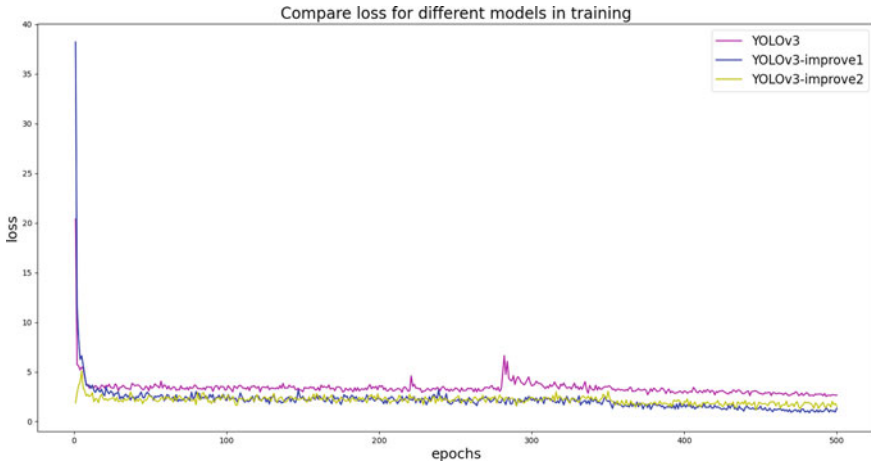


Fig. 7 Comparison of training loss values

Compare the mean average accuracy value mAP of YOLOv3, YOLOv3-improve1, YOLOv3-improve2, and their comparison graphs shown in Fig. 8. Information about the three algorithms, including the trained mAP values and the detection speed (how many sheets per second are detected) is shown in Table 2.

From Fig. 8, it can be seen that the YOLOv3-improve1 and YOLOv3-improve2 algorithms have an overall higher train mAP than the YOLOv3 algorithm. As per contents in Table 2, YOLOv3-improve1 algorithm has 3.3% increase in its mAP value than YOLOv3 algorithm. Due to its deeper network model, its detection speed is 6 sheets/s slower than the original algorithm. The YOLOv3-improve2 algorithm has

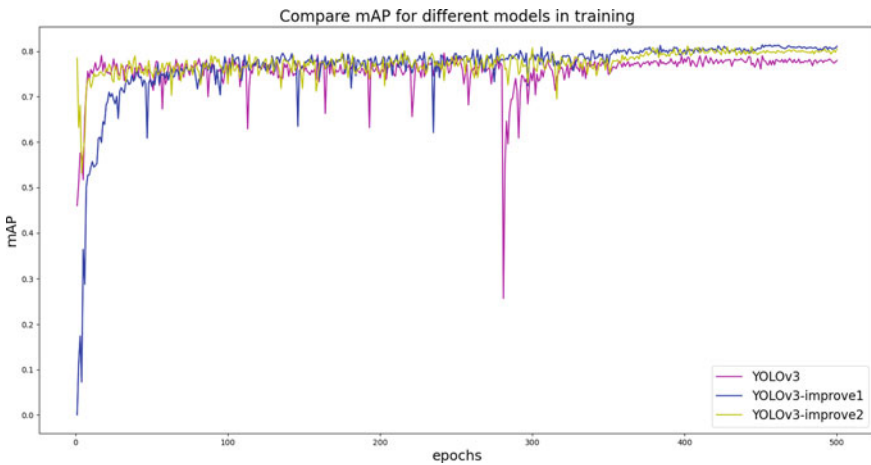
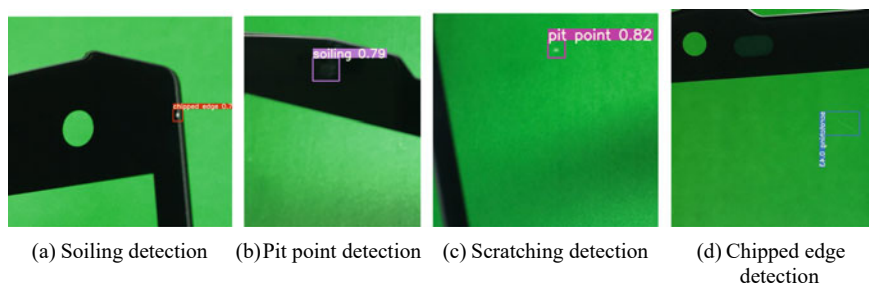


Fig. 8 Comparison of training mAP values

Table 2 Comparison of experimental results

Model	Feature detection layer	Attention mechanism	Network pruning	mAP (%)	Detection speed (sheets/s)
YOLOv3	3 layers	No	No	78.0	36.4
YOLOv3-improve1	4 layers	Yes	No	81.3	30.4
YOLOv3-improve2	4 layers	Yes	Channel crop 60%, layer crop 12 layers	81.0	43.1

**Fig. 9** Mobile phone glass covers defect detection effect

a 0.3% decrease in its mAP value than YOLOv3-improve1, but its detection speed increases by 12.7 sheets/s compared to YOLOv3-improve1, which is a significant improvement with less loss of accuracy. In addition, YOLOv3-improve2 algorithm improves detection accuracy by 3% compared to the YOLOv3 algorithm, while the detection speed also increases by 6.7 sheets/s. This demonstrates that the improved algorithm in this paper has a better performance in detection.

YOLOv3-improve2 training weights were adopted to detect mobile phone glass covers defects, with the results shown in Fig. 9.

According to the above figure, it can be seen that the final YOLOv3-improve2 algorithm provides a more accurate detection method of mobile phone glass covers.

5 Conclusion

The detection of smartphone glass cover defects using manual methods, which is inefficient, costly, low detection accuracy and labour intensive, while the detection methods using traditional machine vision is poor detection flexibility, low yield and poor generalisation capability. Therefore, this paper adopts the YOLOv3 algorithm for defect detection on smartphone glass covers, and improves the YOLOv3 algorithm for the specific requirements and practical characteristics of defect detection.

The channel attention mechanism SENet is added to the Darknet-53 network to solve the problem of inconspicuous defect features, a 104×104 scale detection layer was added to the YOLOv3 detection network to solve the problem of multi-scale defects, and the scaling factor coefficient of the BN layer of the convolutional network is used as the importance factor for model pruning to improve the detection speed. A large number of photographs covering chipped edges, scratches, pits and dirty defects were taken from a smartphone glass cover manufacturing company to make a training dataset and a validation dataset. The proposed method and the original YOLOv3 algorithm are compared and analysed, and the results showed that the algorithm outperforms the original YOLOv3 algorithm in all aspects, not only in terms of real-time performance, but also in terms of detection accuracy, meeting the need for high precision and efficient detection of defects in the industrial production site of smartphone glass covers.

References

1. Chengxuan, W., Qi, L., Ying, T.: Five trends in the development of glass for mobile phones. *Glass* **47**(04), 1–6 (2020)
2. Jian, C., Gao, J., Ao, Y.: Automatic surface defect detection for mobile phone screen glass based on machine vision. *Appl. Soft. Comput.* (2016)
3. Chuanxia, J., Jian, G.: Research on visual inspection method for surface defects of mobile phone glass covers. *Packag. Eng.* **39**(5), 16–21 (2018)
4. Chuanxia, J.: Research on machine vision detection and classification method for mobile phone glass covers surface defects. Guangdong University of Technology (2017)
5. Liang, L.-Q., Li, D., Fu, X., Zhang, W.-J.: Touch screen defect inspection based on sparse representation in low resolution images. *Multimedia Tools Appl.* **75**(5), 2655–2666 (2016)
6. Jianguo, Z., Li, Y., Qi, J.K., Ji, T., Liu, J.: Research on scratch detection on mobile phone screen surface based on machine vision. *Appl. Opt.* **41**, 1–5 (2020)
7. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. *Neural Inf. Proc. Syst.* 91–99
8. He, K., Gkioxari, G., Dollár, P., et al.: Mask R-CNN. In: International Conference on Computer Vision, 2980–2988 (2017)
9. Redmon, J., Divvala, S.K., Girshick, R., et al.: You only look once: unified, real-time object detection. *Comput. Vis. Pattern Recognit.* 779–788 (2016)
10. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot MultiBox detector. *Lect Notes Comput Sci* 21–37 (2016)
11. Li, C., Zhang, X., Huang, Y., Tang, C., Fatikow, S.: A novel algorithm for defect extraction and classification of mobile phone screen based on machine vision. *Comput. Ind. Eng.* 106530
12. Jiang, J., Cao, P., Lu, Z., Lou, W., Yang, Y.: Surface defect detection for mobile phone back glass based on symmetric convolutional neural network deep learning. *Appl. Sci.* **10**, 3621 (2020)
13. Hongjia, S.: Research on Mobile Phone Screen Defect Detection and Classification Algorithm Based on Depth Model. Zhejiang University (2018)
14. Wei, S.: Deep Convolutional Neural Network-Based Defect Detection for Mobile Phone Screens. University of Electronic Science and Technology (2019)
15. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767)
16. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 7132–7141 (2018)

17. Hui, Z., Kunfeng, W., Feiyue, W.: Progress and prospects of deep learning in target vision detection. *Acta Automatica Sinica* **43**(8), 1289–1305 (2017)
18. Liu, Z., Li, J., Shen, Z., et al.: Learning efficient convolutional networks through network slimming. *International Conference on Computer Vision*, 2755–2763

Optimization and Design of Efficiency and Quality of a Company Based on Value Stream Analysis



Guo Jidong, Qiu Zixuan, Huang Zehao, Wu Jiaqi, Zheng Jianxin, Tan Runjia, Lai Lijuan, and Zhou Dawei

Abstract A company mainly produces varistor (e-var). Through observation and analysis, the e-var production line of company a has problems such as low production efficiency, high rate of defective products, more in-process products, nonstandard operation of production line and long waiting time. This paper explores and improves the above problems by the methods of value flow chart analysis, 5W1H method, ECRS principle, standard operation sequence, rapid model change, setting up control experiment and causal chart. The e-var production line is integrated and standard work is formulated. The die changing can be realized quickly in the pressing process, and the quality problems of the products are comprehensively managed. After improvement, company a has improved production efficiency and the rate of defective products decreased significantly, reducing production cost.

Keyword Value flow chart · ECRC principle · Rapid mold change · Defective product rate · Control experiment · Causal char

1 Analysis of Workshop Status of Production Line

In recent years, with the rapid development of global economy, the domestic market demand for varistor (e-var) is increasing year by year, and the competition among enterprises in the industry is becoming more and more fierce. How to improve its competitiveness and become a leading enterprise in the electronic industry is an urgent problem for a company.

The company is a manufacturing enterprise, its main business is the production and sales of e-var products, with an intelligent e-var production line, which adopts pull production mode. There are many problems in the production line, such as long waiting time, low efficiency, nonstandard operation, large WIP inventory and high defective product rate. Nowadays, the market demand for e-var products is increasing.

G. Jidong (✉) · Q. Zixuan · H. Zehao · W. Jiaqi · Z. Jianxin · T. Runjia · L. Lijuan · Z. Dawei
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, Guangdong, China
e-mail: 35189850@qq.com

In order to better meet the market demand, adapt to market changes, achieve high quality and high qualified rate of flexible production, effectively improve the quality of e-var products, improve production efficiency, eliminate waste, and become a leader in the competitive e-var production industry, it is imperative to improve and optimize the company [1]. This paper selects the E-VAR production line as the research object, and uses the innovative method of value flow analysis to make the current value flow diagram of the product, as shown in Fig. 1.

According to the data reflected in the value flow chart, the following problems can be found through careful investigation in the corresponding stations:

1. In the layout of the whole production line, there is no reasonable design and planning. In the production process, it is easy to cause confusion of WIP, unnecessary waste of time and WIP inventory. At the same time, the process distribution is not balanced, many operations do not conform to the principle of action economy in production, the workload increases and the production efficiency is low, so it is necessary to integrate and adjust the process.
2. In the production process of the tablet pressing process, it is observed that the die changing process needs to be carried out manually by the staff, and the whole process from the disassembly of the die to the formal operation is operated by one person, and the staff is responsible for the whole process by one person, and there is no distinction between internal operation and external operation. All the die changing preparations are carried out after the shutdown, which is typical waste in traditional die changing. It is necessary to improve the process of rapid die change [2].
3. The production line lacks unified and standard on-site management. Taking the glass spraying process as an example, due to the great differences in the loading and unloading sequence of different employees, the operators are accustomed to the operation and material placement in the production process. As a result, the total processing time of different batches of parts is inconsistent, and the ideal standard operation sequence is lacking [3].
4. It can be found from the current value stream map that after 8/20 test and 2 ms test, the first pass rate of the product in the film selection post is low, and the rework rate is high, which indicates that there are problems in the previous process. It is found that the top four quality problems that lead to e-var product scrap are crack, deformation, impurity and crack, accounting for 48%, 23%, 12% and 10% respectively.

2 Optimization of Product Efficiency

2.1 Process Improvement

Through the data reflected in the value flow chart and careful investigation in the corresponding work station, it is found that the vehicles of WIP are changed

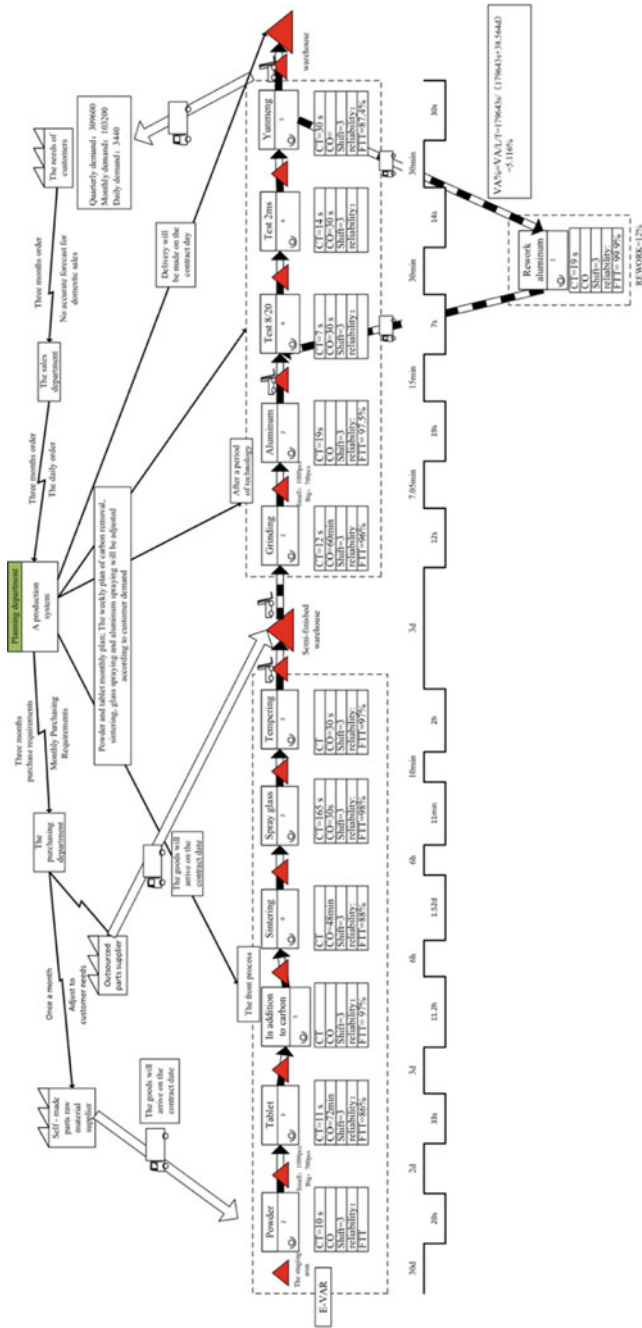


Fig. 1 Current value stream map

frequently from the decarbonization process to the glass spraying process. Among them, the waiting time for glass spraying is as long as 6 h, and the quantity is 17640 Pcs. Due to the limited capacity of WIP in the area to be sprayed, the loading limit is 14112 pcs, so the WIP in the area to be sprayed can only be temporarily stored in the forklift area (Table 1).

In order to explore the problem of WIP accumulation, the following points can be found through 5W1H analysis method:

1. The serious accumulation of work in process (WIP) of glass to be sprayed is due to the long feeding time of employees, the serious waiting waste of glass spraying equipment and WIP, which has great room for improvement.
2. In order to reduce the waste of time, two employees should be responsible for unloading sagger and sweeping powder respectively.
3. The powder sweeping station is mainly responsible for cleaning and packing the WIP. The use of pearl cotton is to prevent the impurities attached to the cleaned WIP. According to the principle of ECRs, the current process is integrated, and the integration content is shown in Table 2.

The comparison of the steps from sintering to glass spraying before and after the improvement is as follows (Figs. 2 and 3).

After the improvement of the process flow, we can see the comparison of the improvement effect as shown in the table below.

According to Fig. 4, the equilibrium rate of the process can be calculated. Before improvement, the balance rate was $(1.03 + 1.79 + 2.45)/(2.45 * 3) = 71.7\%$, after improvement, the balance rate was $(0.94 + 1.28 + 0.9)/(1.28 * 3) = 81.3\%$, and the process balance rate increased by 9.6%. The single cutting time of employees decreased by 46.2%, the single feeding time decreased by 74.6%, the processing time of unit parts decreased by 43.7%, and the process balance rate increased by 9.6%.

Table 1 Vehicle changes in each process

Working procedure	Vehicle changes in WIP	Operator required
Carbon removal	Carbon removal rack	No
A sagger	Carbon removal rack → sagger	Yes
Sinter	Sagger	No
Unloading saggars	Sagger → pearl cotton	Yes
Loading	Pearl cotton → feeding plate	Yes
Spray glass	Loading tray	No
Cutting	Feeding tray → decarbonization rack	Yes

Table 2 ECRS principles process integration table

Before improvement	Project	After improvement
After sweeping powder, employee B put the work in process into pearl cotton and packed it. Then employee C disassembled the pearl cotton and took out the work in process with clean surface to finish feeding	Cancel	In order to prevent dust and other impurities from contacting the work in process, the work in process finished by powder sweeping process is packed and packed with pearl cotton. If the work in process is powder swept before feeding by C staff, the use of pearl cotton can be completely cancelled
After the sagger unloading and powder sweeping and packing are completed in the sintering workshop, the employees transport the products in process to the glass spraying area and wait for glass spraying	Merge	Let the sagger unloading station, powder sweeping station and feeding station be adjacent to each other, and combine the three steps of sagger unloading, powder sweeping and feeding. After employee a completes sagger unloading, employee B sweeps powder immediately, and employee B places the WIP in the transition tray after powder sweeping, which is convenient for employee C to feed. It makes unloading sagger, sweeping powder and feeding a continuous process
Employee C needs to move the WIP to the waiting area and unload the pearl cotton	Rearrangement	Before loading, the loading station needs to transport the WIP to the waiting area and unload the pearl cotton. After the stations were merged, the work of transporting saggars to the waiting area and unloading saggars was handed over to staff a, which improved the operation efficiency of the three stations
When employee C is loading or unloading, only two WIPs can be taken at the same time with both hands until the loading tray is full or empty	Simplify	Using the mobile suction cup instead of manual loading and unloading, employees can operate the mobile suction cup to complete the loading or unloading of 6 WIPs at one time



Fig. 2 Process of WIP from sintering to glass spraying (status quo)



Fig. 3 Process of WIP from sintering to glass spraying (after improvement)

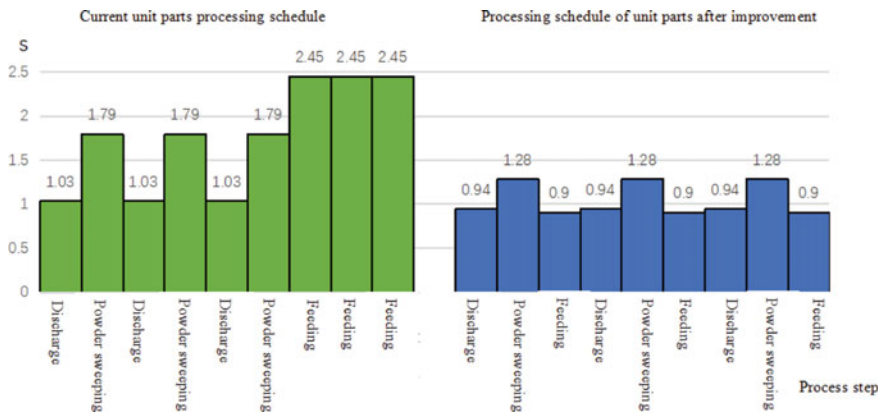


Fig. 4 Comparison of improvement effect of unit parts processing schedule

2.2 Rapid Die Change in Tablet Pressing Process

The tablet pressing machine in the tablet pressing workshop is pneumatic, and each tablet pressing machine is operated by one person. There are more than 100 kinds of pressing molds (upper punching mold, middle punching mold and lower punching mold). Through observation, it is found that the mold changing process needs to be carried out manually by employees, and the whole process from the disassembly of the mold to the formal operation is operated by one person, so there is great room for improvement.

Use video equipment to record the die change of the pressing process, disassemble and record the steps, count the time of the main steps, and draw the statistical Table 3 of the die change step time.

Table 3 Time statistics of die changing steps

Serial number	Step	Time/min	Proportion (%)
1	Remove the mold	20	18.5
2	Clean the mold	5.5	5.1
3	Adjustment	11	10.2
4	Prepare the mold	7	6.5
5	Install the mold	35	32.4
6	Finishing mould	6.5	6
7	Commissioning and adjustment	15	13.9
8	Loading	8	7.4
9	Official operation	–	0
Total	–	108	100



Fig. 5 Ideal die changing procedure

It can be observed that the employees are responsible for the whole process by one person, and there is no difference between internal and external operations. All the preparation for die changing starts after the machine stops, which is typical waste in traditional die changing. Therefore, this paper distinguishes the internal and external preparation time, and designs the ideal die changing process of pressing process. According to the sequence relationship between the pressing steps, the ideal die changing step diagram can be drawn as follows (Fig. 5).

In practice, all external operations, such as taking the mold, adjusting the mold width, sorting materials, etc., can be carried out before the mold change, while cleaning the mold can be carried out after the mold change. There is no conflict between the steps of adjusting the mold and installing the mold. In order to realize the ideal die changing step and shorten the die changing time, the method of parallel operation can be implemented, so that the single person’s pressing process can be changed into two persons parallel operation [4]. The specific plan is to arrange an employee y to prepare for die change and arrange materials when employee x is in the tablet pressing process. When employee x dismantles and installs the mold after completing the pressing process, employee y will recycle the materials, clean the used mold and adjust the machine (Fig. 6).

After the rapid die change, part of the internal operation is converted into external operation. The internal operation time is reduced from 108 to 78 min, and the external operation time is increased to 30 min. finally, the proportion of internal operation time is reduced by 27.8%. According to the ideal die changing process, a single die changing can achieve the goal of reducing the internal operation time of die changing by 30 min. The company’s single die changing machine can change 120 times a month, which can reduce the internal operation time of die changing by 3600 min per month.

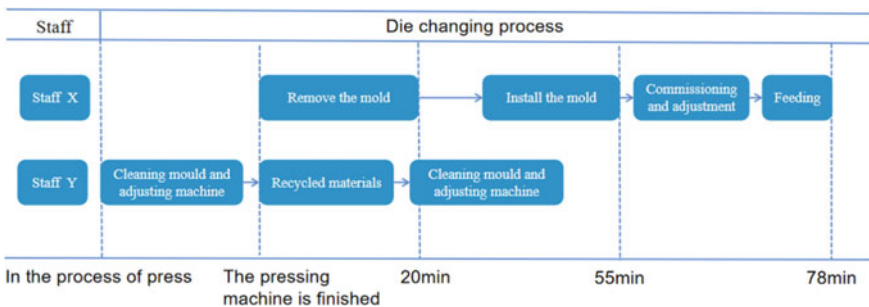


Fig. 6 The process diagram of die changing in the ideal pressing process

2.3 Process Operation Standardization

The glass process is operated by one person. The specific steps are as follows: first, take down the glass sprayed work in process from the charging tray, move it to the decarbonization tray, then put the full amount of decarbonization tray into the decarbonization rack, and then take out the empty decarbonization tray from the decarbonization rack and move it to the position in front of the charging tray before feeding. When feeding, you need to take the WIP from the pearl cotton on the left side, and then take the full amount of the upper material tray to start the machine. It can be observed that employees have a lot of repetitive actions. The proportion of staff loading and waiting for machine conveying tray is the largest in the whole time, which is as high as 47% and 22% respectively, and there is much room for improvement (Table 4).

According to the field observation, it is found that there are two reasons for the long time of glass loading and unloading operation.

1. Employees need to take WIP from the pearl cotton on the left side for feeding, and they can only take 2 WIP at a time until the feeding tray is full. In the process of feeding, there are a large number of employees turning back and forth, squatting and standing up, and each machine loading and unloading position is only operated by one employee, resulting in the action cycle time (CT) in glass spraying process is the longest in all processes.
2. Employees start up the machine after feeding, wait for the machine to finish processing and then send the next tray. Each waiting time is as long as 30 s, there is a serious waste of waiting.

Due to the great difference of loading and unloading sequence of different employees, the total processing time of different batches of parts is inconsistent, and the ideal standard operation sequence is lacking. In order to improve production efficiency and reduce the amount of labor, this paper determines the ideal standard operation sequence.

Through the observation records, it is found that the decarbonization tray full of work in process does not affect the staff's feeding, and there is a lot of waiting waste

Table 4 Comparison of die changing time

Project	Before improvement		After improvement	
	External time	Internal time	External time	Internal time
Time/min	0	108	30	78
Proportion (%)	0	100	27.8	72.2
Total duration/min	108			
Improvement effect	After the improvement, the internal operation time was reduced by 30 min			
Increase proportion	After improvement, the proportion of internal working time decreased by 27.8%			



Fig. 7 Standard operation sequence of glass spraying process

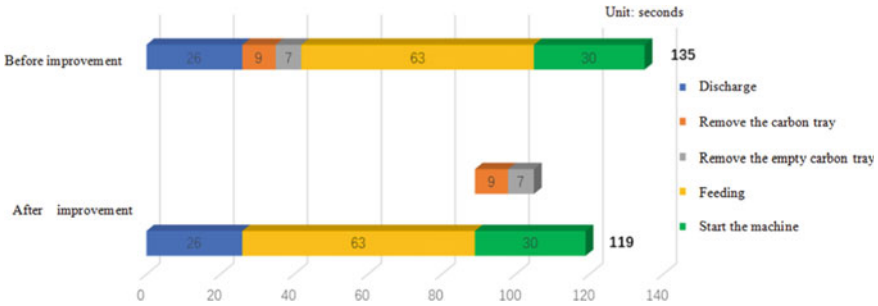


Fig. 8 Time lapse diagram of glass spraying process operation steps

after the staff starts the machine. The staff can complete the feeding immediately after the blanking step, start the machine after the feeding, and use the processing time of the machine to complete the replacement step of the decarbonization tray, which saves a lot of waiting time. The efficiency of glass spraying process is improved. The ideal standard operation sequence is shown in Fig. 7.

According to Fig. 8, after the improvement, the internal time is converted to the external time, and the waiting waste of the equipment is reduced. Finally, the processing time per unit part of glass spraying process is reduced from 135 to 119 s, reducing by 11.9%.

3 Optimization of Product Quality

It can be found from the current value stream map that after 8/20 test and 2 ms test, the first pass rate of the product in the film selection post is low, and the rework rate is high, which indicates that there are problems in the previous process. According to the first pass rate of the value stream diagram, we should pay attention to the pressing and sintering process. The first pass rate of the pressing process is 86%, and the first pass rate of the sintering process is 88%. The proportion of scrapped e-var products due to quality problems is 11%. Statistical analysis of unqualified e-var products shows that the top four quality problems causing scrapped e-var products are crack, deformation, impurity and cracking, accounting for 48%, 23%, 12% and 10% respectively.

Combined with the actual field observation, through the 5W1H analysis method, it is found that the quality problems of e-var products will lead to low product qualification rate, high rework cost and prolonged production cycle [5]. Among

them, if the product is mixed with impurities, it will directly lead to unqualified UV performance; Product deformation will lead to product specification problems; The problems of crack and cracking will lead to the product cannot bear the voltage, so it will be scrapped.

3.1 Process Improvement

(1) Setting up experiments to explore the causes of cracks.

A batch of experimental products were obtained, 10 WIPs were randomly taken out after the pressing process, 10 WIPs were randomly taken out after the decarbonization process, 10 WIPs were randomly taken out after the sintering process, 10 WIPs were randomly taken out after the tempering process, and the rest were put into the semi-finished product warehouse. All the work in process products randomly taken out were grouped according to the process and Hiwave scanning was carried out to obtain the experimental images. Figures 9, 10, 11 and 12 are the images of each experimental group after Hiwave scanning.

Through the experiment, it is found that the crack is a problem existing after pressing, and it can be judged that the crack is a quality problem produced in the pressing process. One third of the in-process products with chip selection and rework have cracks, which leads to the decrease of the first pass rate and rework rate of the chip selection process.

After the investigation of the production equipment, it is found that the oil pump of one tablet press has been replaced. The oil pump of the tablet press is used to observe the pressure, and the amount of oil discharged is used as the display value of the pressure parameters for employee's reference. For the requirements of different batches of products, the displayed parameter values have errors. The reason is that

Fig. 9 Hiwave sound scanning image after pressing

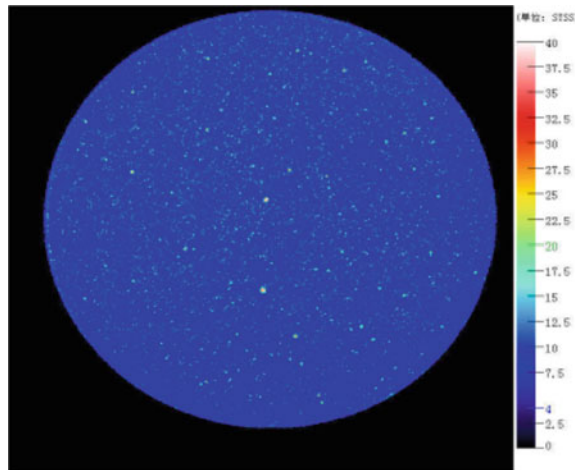


Fig. 10 Hiwave acoustic scanning image after carbon removal

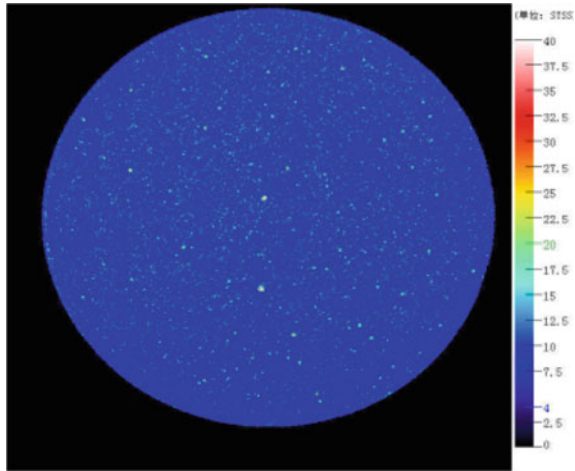
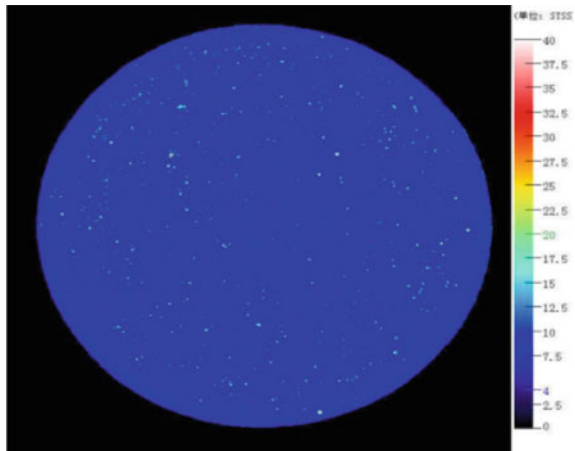


Fig. 11 Hiwave scanning image after sintering

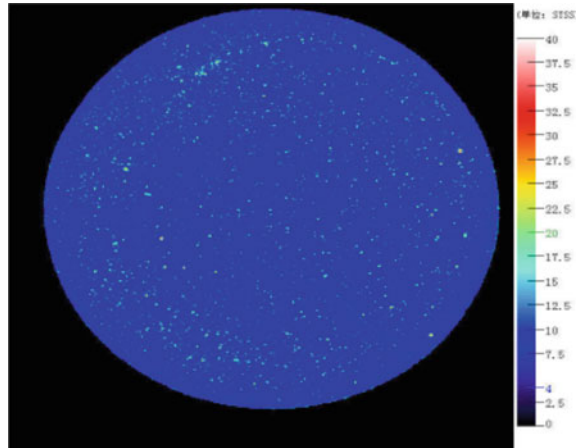


one night the machine broke down. When the maintenance department repaired the machine, it did not communicate with the engineer effectively and directly replaced the matching oil pump. However, the same type of oil pump cannot meet the threshold requirements of the original matching oil pump. For different batches of products, even if the machine parameters are adjusted to the maximum, it still cannot meet the production requirements. The next day, the engineer went to work and did not realize the problem, which led to the product crack quality problem, and entered the next process, the product scrap rate increased.

(2) The control group was set up to explore the causes of impurities and deformation.

A batch of experimental products were obtained and divided into 4 groups on average. A ZnO plate was placed between the product and the sagger in each group,

Fig. 12 Hiwave scanning image after tempering



so that the bottom of the product and the sagger would not contact each other. Put them in the same sintering furnace for sintering process. Table 5 is the control experimental process table to explore the influence of “impurities” and “deformation” on e-var.

Through the control experiment, the experimental data record table can be obtained, as shown in Table 6.

It can be seen from the control experiment that the mean absolute value of the difference between the center point and the plane of the sample in group B is the smallest, that is, the deformation degree of the work in process of the new ZnO pad is the smallest, while that of group A is the larger. It can be judged that using the new ZnO plate during sintering can effectively reduce the influence of the deformation problem on the quality of the work in process. The number of impurities in group C sample is the least after sintering. It can be judged that a layer of ZnO plate on the work in process before sintering can effectively isolate the contact between work in process and impurities, and can effectively reduce the deformation of work in process.

(3) Explore the causes of cracking.

The cause of e-var product cracking was investigated from five aspects of human, machine, material, method and environment by using causality diagram. Figure 13 is a cause-and-effect diagram based on field investigation.

During the 3-day investigation, the team found the following problems:

1. There are inexperienced employees and non-compliance with the provisions of the standard operating manual. According to the standard operation manual, the saggings in process can only be stacked in 5 layers, while some employees stack 6 layers in the process of loading saggings, ignoring the provisions of the standard operation manual.
2. Forklifts or trolleys carrying e-var products vibrate greatly during transportation, which will lead to the displacement of work in process products and make work in process products touch the carbon removal disk or sagger wall, and

Table 5 Comparative experimental process of exploring the influence of impurities and deformation on e-var

Group	Type	Experiment content	Description of control group and experimental group
Group A	Control group	Old ZnO plate is placed between the product and sagger	ZnO board is recycled, too many times of use, ZnO board will appear the phenomenon of bending deformation
Group B	Experience group	A new ZnO plate is placed between the product and the sagger	This group is to explore the influence of ZnO plate deformation on products
Group C	Experience group	The old ZnO plate is placed between the product and sagger, at the same time, a layer of old ZnO board is also padded on the products	One more ZnO plate is used to separate sagger cover from WIP. Because after high-temperature sintering, the powder (impurities) on the sagger cover will adhere to the product, which is difficult to clean through the powder sweeping process
Group D	Experience group	The old ZnO plate is padded between the product and the sagger, and the product is turned over at the same time	After tableting, employees randomly put the work in process on the decarbonization rack. After the work in process enters the decarbonization furnace, the surface becomes smooth, and there will be more powder on the lower surface. If the product is not flipped, the lower surface of the sintering process will adhere to the broken powder. The purpose of product turnover in the experimental group is to explore whether the turnover of the upper and lower surfaces will affect the product quality

Table 6 Relevant experimental data

Project	Group A	Group B	Group C	Group D
Central point	4.657	4.65067	4.65267	4.65033
Plane	4.733	4.646	4.68733	4.735
Absolute value of difference	0.078667	0.022	0.037333	0.084667
Impurity rate (%)	10	6.7	0	6.7

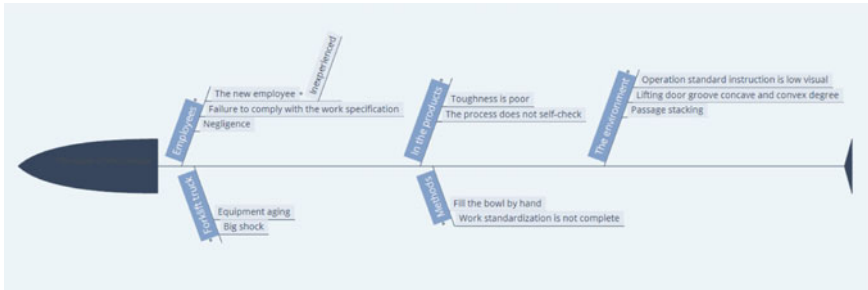


Fig. 13 Cause-and-effect diagram of e-var product quality problem

the collision during transportation will lead to the cracking of work in process products.

3. The toughness of work in process is poor before sintering. If it is impacted by a large force, a gap will be formed.
4. In the process of manual installation of carbon removal rack and sagger, due to negligence, the work in process collides with the plate wall and sagger wall, or the work in process collides with each other, which will lead to the cracking of work in process.
5. The workshop is dusty, the SOP is thick and full of dust, and the visibility is low, which is not conducive for employees to comply with the SOP.

To sum up, “impurity” and “deformation” are produced in sintering process, “crack” is produced in tablet pressing process, “cracking” is produced in the process of loading sagger and removing carbon frame.

3.2 Solutions to Product Quality Problems

3.2.1 Replacing Equipment Parts to Solve Crack Problem

In view of the crack problem, the matching oil pump of the original model should be replaced immediately, and whether the tablet press can meet the production demand should be detected. Set up a maintenance record form. The maintenance department shall make detailed records on the form every time the equipment parts are replaced in the future, and submit feedback to the quality inspection department after maintenance. The quality inspection department should first check the contents of the maintenance record form and confirm the relevant maintenance records when going to work every day. In daily production, the enterprise should check the machinery and equipment every six months to know whether the equipment can meet the actual production demand [6].

3.2.2 Put in Isolation Board to Solve the Problem of Impurity and Deformation

In view of the problem of impurities, a layer of ZnO plate should be covered on the work in process to isolate the work in process and the sagger cover, to prevent the sagger cover from dropping powder and impurities sticking to the surface of the work in process.

In view of the deformation problem, if the employees find that the ZnO plate has bending deformation problem during sagging, they need to replace the new ZnO plate sagging.

Through the contrast experiment, it can be found that a layer of ZnO plate on the work in process can effectively reduce the impurities and deformation of the work in process. Aiming at the problem of using times of ZnO board, this paper obtained a batch of ZnO experimental products, labeled each piece of ZnO, measured the deformation of all ZnO boards after each sintering, took the absolute mean value of the difference between the center point and the plane, and drew the variation diagram of auxiliary production times and deformation of ZnO board, as shown in Fig. 14.

It can be seen from Fig. 14 that the deformation of ZnO plate increases after several times of auxiliary production. After the first 16 times of auxiliary production, the deformation of ZnO did not exceed the deformation index of 0.04. Since the 17th time, the deformation of ZnO board has exceeded 0.04, so it is required to replace the ZnO board after 16 times of use. The cost of ZnO is 1312 yuan/week, and the cost of scrapped WIP due to impurities and deformation is 4527 yuan/week. It can be concluded that the cost can be saved by 3215 yuan per week.

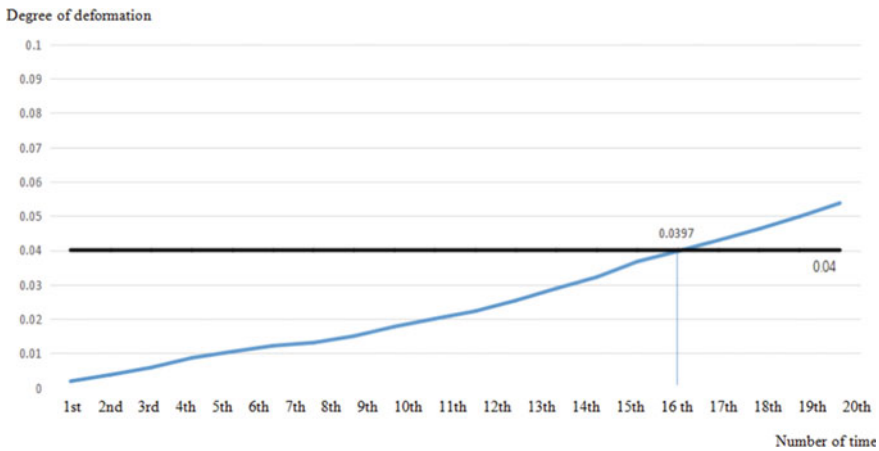


Fig. 14 Variation diagram of times of auxiliary production and deformation degree of ZnO board

3.2.3 Solve the Problem of Cracking for Man Machine Material Method

1. According to Fig. 13, aiming at the problem of inexperienced employees, the original training program is revised.
 - ① Before the new production line is put into operation, staff operation training and assessment shall be carried out, and they are allowed to work after mastering the production process [7].
 - ② The implementation of online teaching, enterprises regularly publish online courses, popularize the latest code of practice. Employees need to study as soon as possible and complete relevant course topics. Online course completion and answer accuracy affect performance to a certain extent [8].
 - ③ The production site will be inspected irregularly by the quality inspection department to check whether the employees consciously comply with the requirements of the standard manual.
2. In order to solve the problem of large vibration of forklifts or trolleys, employees are required to carry in the designated channel and control the handling speed. The sagger layer on the trolley is required to be no more than five layers.
3. Due to the poor toughness of e-var products before sintering, it is easy to produce the cracking due to collision.
 - ① When putting into production, for processes with serious scrap, such as tablet pressing and sintering process, quality inspection procedures for work in process should be added.
 - ② Collision should be avoided as far as possible in the process of transportation. In the process of process handover, the employees of the next process should visually inspect the incoming materials of the work in process of the previous process, and register and eliminate the work in process with the problem of cracking during transportation.
4. In view of the environmental problems, dust removal equipment was added to improve the environment, and employees were arranged to clean the operation site. The standard operation manual should be placed in front of the operator's console to highlight the key contents, facilitate the employees to view the standard operation manual, and remind the employees to consciously abide by the requirements of the standard operation manual [9].

3.2.4 Improvement Effect of Quality Optimization

After improvement, the defective product rate decreased from 11 to 0.2%, a decrease of 10.8%, and the number of defective products per week decreased from 2649 to 49, a decrease of 98.2%. The proportion of crack, deformation, impurities, and cracking was reduced by 98.8%, 97.9%, 97.5% and 97.7% respectively (Fig. 15).

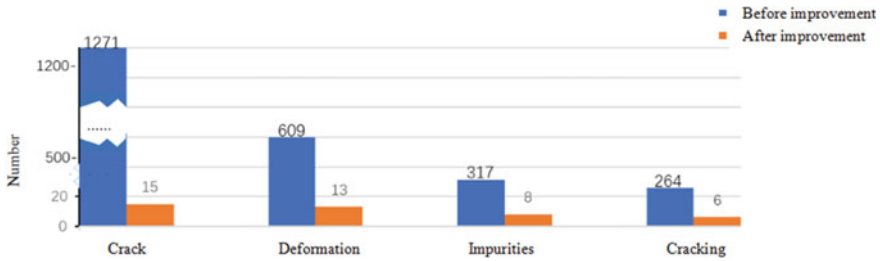


Fig. 15 Variation diagram of times of auxiliary production and deformation degree of ZnO board

4 Conclusion

In the past, the research on the low production efficiency of enterprises is mostly based on a single theory [10]. This study is based on the lean management thinking of industrial engineering and combined with the actual, and uses the idea of expanding from local problems to overall improvement. The improved effect is shown in Table 7 and Fig. 16. Using value flow analysis, experimental demonstration, ECRS principle and other analysis tools, the innovative method of E-VAR production line

Table 7 Comparison of improvement effect

	Project	Before improvement		After improvement		Improvement effect
Product efficiency improvement effect	Total process time	135 s		60 s		Reduce 55.6%
	Waste of time waiting for equipment	105 s		89 s		Reduce 15.2%
	Waste time of WIP waiting	6 h		1 h		Reduce 5 h
	Storage capacity of OEM area	14112 pcs		40320 pcs		Increase 26208 pcs
	Daily cost saving of pearl cotton	491.52 yuan		0 yuan		Reduce 491.52 yuan
	Process balance rate	71.7%		81.3%		Improve 9.6%
	Internal and external classification of rapid die change	External time	Internal time	External time	Internal time	\
Internal and external time of rapid die change	0 min	108 min	30 min	78 min	Internal operation time reduced by 30 min	

(continued)

Table 7 (continued)

	Project		Before improvement		After improvement		Improvement effect
	Internal and external time proportion of quick die change		0	100%	27.8%	72.2%	Internal operation time ratio reduced by 27.8%
Effect of product quality improvement	Defective product rate		11%		0.2%		Reduce 10.8%
	Number of defective products		2649		49		Reduce 98.2%
	Products with crack problems		1271		15		Reduce 98.8%
	Products with deformation problems		609		13		Reduce 97.9%
	Products with impurity problem		317		8		Reduce 97.5%
	Products with cracking problems		264		6		Reduce 97.7%
Comparison of value stream map before and after	Working procedure	Index	\		\		\
	Tableting	FTT	86%		94%		Improve 8%
	Sinter	FTT	88%		96%		Improve 8%
	Spray glass	CT	135 s		60 s		Reduce 55.6%
	Film selection	FTT	87.4%		97.2%		Improve 9.8%
	Rework rate		12.6%		2.8%		Reduce 9.8%
Value added time		0.050991%		0.051256%		Improve 0.5197%	

of A company is studied, to reduce waiting waste, improve production efficiency, standardize operation standards, and improve product quality.

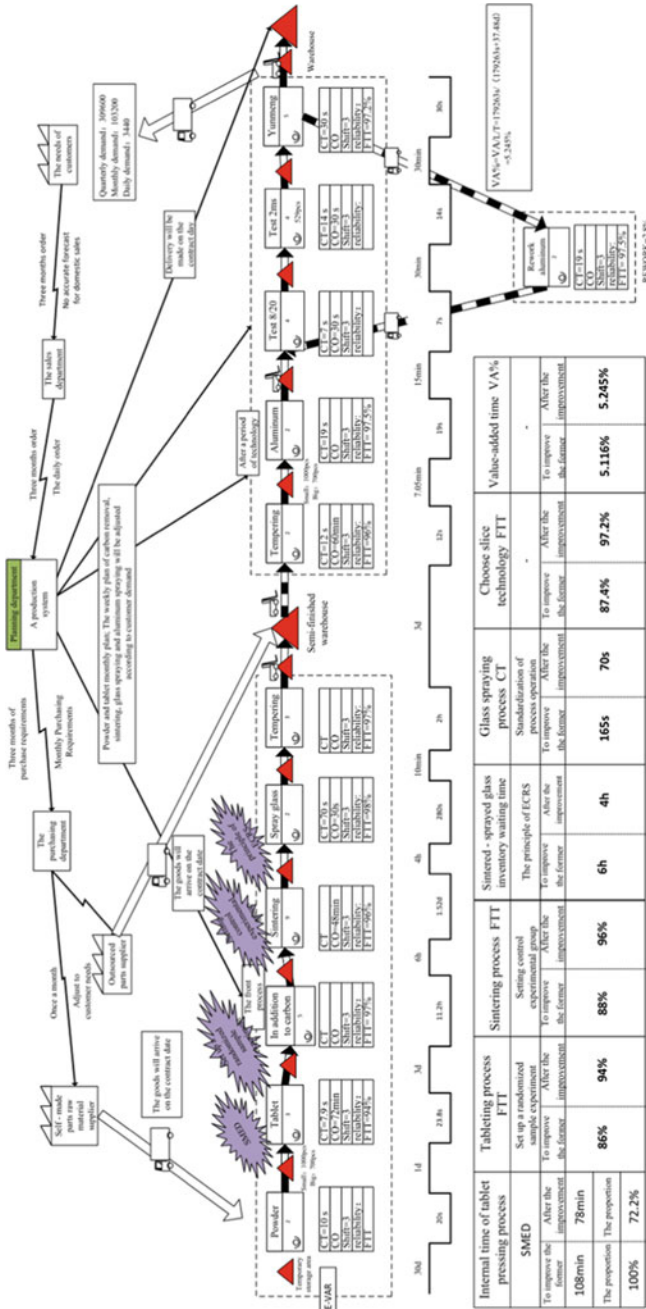


Fig. 16 Improved value stream map

References

1. Ge, L.: Discussion on management innovation mode of manufacturing enterprises based on lean production. *Econom Manage Abstracts* **19**, 62 (2020)
2. Tang, X.: Application of lean production management in enterprise management. *Value Eng* **39**(21), 26–27 (2020)
3. Shaoyun, L.: Modern enterprise standardized operation analysis and its application research. *Enterprise Reform Manage* **18**, 17–18 (2019)
4. Qicheng, T., Binghai, Z.: Improvement of stamping equipment utilization based on rapid die changing technology. *Precis Manuf Automat* **4**, 60–64 (2019)
5. Jing, G.: Quality management status and effective countermeasures of machinery manufacturing enterprises. *Enterprise Reform Manage* **10**, 27–28 (2020)
6. Hongyan, L.: Application of intelligent quality management in machinery manufacturing industry. *Internal Combust Eng Accessor* **15**, 190–191 (2019)
7. Kunbin, Z.: How to improve the effectiveness of quality management system in machinery manufacturing enterprises. *Qual Market* **09**, 31–33 (2020)
8. Yiming, Z., Yuandi, H.: Research and implementation of quality management mode in manufacturing industry. *Internal combustion engine and accessories* **11**, 182–183 (2019)
9. Chao, L: Research on Total Quality Management System Improvement of D Company. Dalian University of Technology (2017)
10. Chuang, G.: Research on Enterprise Environmental Cost Control System Based on PDCA Cycle. Harbin University of Commerce (2017)

Review of Using Operational Modal Analysis for Condition Monitoring



Fulong Liu , Wei Chen, Yutao Men, Xiaotao Zhang, Yuchao Sun, Jun Li, and Guoan Yang

Abstract Modal analysis is critical to better understand structural dynamic vibration characteristics by extracting system's natural frequencies, damping ratios and mode shapes. Modal analysis has been widely used to structure optimization in the design stage, damage detection and structural health monitoring or condition monitoring. According to whether need artificial excitation, the modal analysis techniques can be categorized as experimental modal analysis and operational modal analysis. Conventional experimental modal analysis has to measure the excitation and corresponding response in the meantime, while operational modal analysis measure system's response only during normal operating condition. Therefore, operational modal analysis also called output-only modal analysis methods, which have developed dramatically in recent decades because it is promising as means to achieve structural online monitoring, which is highly desirable for critical mechanical system, important buildings and bridges, etc. This paper made a brief review of the development of popular operational modal analysis techniques and their applications in condition monitoring.

Keywords Operational modal analysis (OMA) · Condition monitoring · Time-domain methods · Frequency-domain methods

F. Liu · W. Chen (✉) · Y. Men · X. Zhang · J. Li
Tianjin Key Laboratory of the Design and Intelligent Control of the Advanced Mechanical System, Tianjin University of Technology, Tianjin 300384, China
e-mail: chenwynn@163.com

National Demonstration Centre for Experimental Mechanical and Electrical Engineering Education, Tianjin University of Technology, Tianjin 300384, China

Y. Sun
Zhonghuan Information College, Tianjin University of Technology, Tianjin 300380, China

G. Yang
Tianjin Langyu Robot Co., Ltd., Tianjin, China

1 Introduction

Operational Modal Analysis (OMA) methods have developed dramatically in recent decades and could be used for monitoring the condition of mechanical or civil structure. In particular, OMA is promising as a means to achieve online monitoring, which is highly desirable for enhancing the reliability, safety and intellectualization of mechanical system and important buildings. The conventional OMA methods are reviewed in this paper.

OMA, also called output-only modal analysis, is used for dynamic system modal parameter identification. In other word, OMA methods only need responses to extract the modal parameters when the system is under ambient excitation. OMA is considered superior to conventional modal analysis techniques: Experimental Modal Analysis (EMA). Compared with EMA, the main advantages of OMA are [1]:

1. OMA is cheaper and more convenient than EMA since there is no need for artificial excitation;
2. OMA can obtain the dynamic characteristics of the whole tested system rather than just a part;
3. OMA can obtain a system's dynamic features under real operational conditions but not experimental conditions;
4. OMA is able to identify close modes;
5. OMA has the capability for online CM.

Because of these advantages, numerous OMA methods have been developed, and these are usually divided into frequency-domain and time-domain methods. The time-domain techniques can be further classified as two-stage and one-stage. The framework and processes of the most popular OMA methods are presented in Fig. 1. The relevant studies and applications are reviewed below.

2 OMA Methods

2.1 Frequency-Domain Methods

For the frequency-domain methods, Peak-Picking (PP) is the simplest technique to estimate the modal parameters. PP is named after the key step of the method: picking the peaks from a spectrum plot as the identified eigenfrequencies [2]. The major issue with the PP method is the accuracy of the identified results, especially for a system with close modes.

Frequency Domain Decomposition (FDD) has been proposed based on the PP method [3]. The FDD technique overcomes the disadvantages of PP by decomposing the spectral density matrix into a set of Single-Degree of Freedom (SDOF) systems via Singular Value Decomposition (SVD). The FDD was widely used, and many

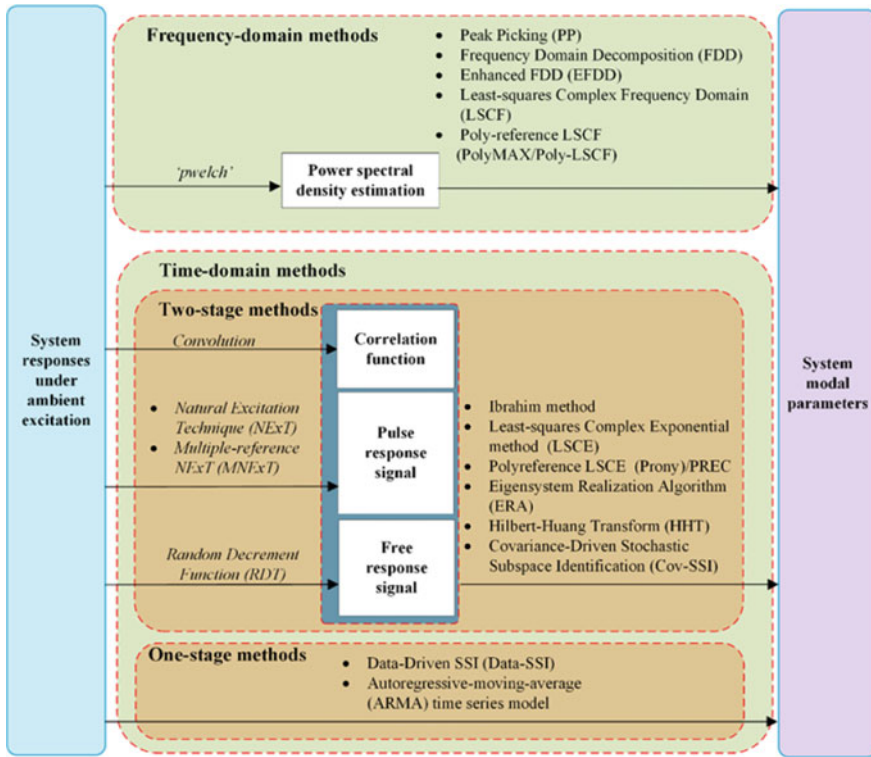


Fig. 1 Framework and processes of popular OMA methods

enhanced methods were proposed based on FDD, such as Enhanced FDD (EFDD) [4], and Frequency-spatial Domain Decomposition (FSDD) [5].

In 2001, the Least Squares Complex Exponential (LSCE) method, a frequency domain method, was presented to cope with high system orders and high modal overlap, and to make the selection of the model order and relevant physical system poles easier [6]. Moreover, a Poly-reference LSCE method was proposed later to provide considerably improved pole stabilization [7]. It is worth to note that Poly-reference LSCE is widely used in commercial modal test systems: LMS Test.Lab [7].

In 2004, an updated Poly-reference LSFCF, called PolyMAX is employed by LMS system because this method has the capability to identify the closely space modes and high damping modes compared with other conventional approaches [7]. Such capability is on the basis of the clear Stabilization Diagram (SD) which contains frequency, damping and participation information. The SD identified by PolyMAX is clearer than the ones identified by other approaches. In order to increase the noise suppression ability, a PolyMAX Plus method was proposed by the same research team in 2012 [8]. The main development of PolyMAX Plus method is adding some maximum likelihood estimation features to proper handle the effects of uncertainty

and estimate the confidence bounce, which can improve the identification results in case of very noisy data [8].

2.2 *Time-Domain Methods*

As shown in Fig. 1, time-domain methods can be divided into two groups: two-stage and one-stage methods. The first stage for the two-stage methods is usually to obtain the correlation, pulse response or free-response functions, and then the obtained functions are used to extract modal information in the second stage.

A good example of the two-stage approach is the Ibrahim Time Domain (ITD) method. The ITD method was first proposed in 1977 based on the Random Decrement Technique (RDT) for modal identification of structures [9], where RDT was employed in the first stage to obtain a free-response signal for the system under random excitation.

In 1985, Juang et al. proposed the Eigensystem Realization Algorithm (ERA) [10] which also uses the free-response signal as obtained in the first stage. In 1995, the Natural Excitation Technique (NExT) was presented as a means of modal testing the permitted structures to be investigated in their actual environments. NExT, as ERA, also uses pulse response function obtained in the first stage [11]. The Covariance-Driven SSI (Cov-SSI) was proposed in 1999 [12]. This is a robust OMA method employing the correlation function of system response as input.

For the one-stage methods, Data-Driven SSI (DD-SSI) [12] and Autoregressive Moving Average (ARMA) methods are the two most popular approaches, using the collected raw data for modal identification.

2.3 *Other Methods*

Besides the classical time-domain and frequency-domain OMA methods reviewed, there have been many related techniques developed in recent years. For instance, a Bayesian approach was proposed, and appears to be becoming popular, for OMA [13, 14]. The uncertainty of system identification is addressed by the Bayesian approach, as the Bayesian approach take modal identification as an inference problem where probability is used as a measure for the relative plausibility of outcomes given both a model of the system and measured data [13, 15].

In addition, transmissibility measurements for OMA are drawing increasing attention because of its ability to successfully extract modal parameters in the presence of harmonics [16–20]. While the poles of transmissibility measurements did not match with the poles of the measuring system, it was shown that modal parameters could be extracted by measuring the system's transmissibility when the system was under subject to different excitations [16].

Recently, an output-only damage identification method was presented in [21, 22], which was also developed on the fundamentals of OMA theory. This method combined with PCA theory can be a baseline-free approach. Its performance has been proved by numerical and experimental studies to detect and locate a damage site.

Even though OMA has proved to be effective and robust, we should note that OMA methods still facing some challenges which have to be solved. The first one is most of OMA techniques are developed under the assumption of identifying a linear and time-invariant system. It is apparent that this do not comply with the real engineering situation. Many researchers have made efforts to cope with this challenge. For instance, a linear approximation of nonlinear and nonstationary systems was proposed in [23] for correlation-driven OMA in terms of estimated natural frequencies, damping ratios and mode shapes. Similarly, the correlation signal was grouped and then averaged in subset in [24, 25] for the same purpose of reducing system's nonlinear and nonstationary effects. The second challenge is the assumption of white noise excitation using OMA. Although many approaches have been investigated to remove the harmonics effects in OMA [26, 27], it is still a hot topic since this problem is not solved completely.

3 Applications of OMA for CM

As mentioned earlier, OMA has been widely used for CM in different areas and has been used to estimate the health status of historic buildings [28, 29]. OMA is popular in the field of civil engineering for monitoring the health of structures, because it is hard to artificially excite buildings, whereas state-of-the-art OMA can accurately extract modal parameters. For example, SSI was employed in [12] for SHM of a steel mast excited by wind and a bridge subject to normal traffic usage. The condition of an in-service three-span highway bridge was successfully assessed by SSI combined with a wireless sensor networks [30]. Also the dynamic behaviour of the Tamar Suspension Bridge have been investigated by SSI which took into account operational and environmental influences [31]. Generally, a real-time monitoring system is required for the monitoring of important buildings or bridges, with many of the monitoring systems based on OMA [32].

Secondly, OMA is widely used with wind turbines as a powerful means of CM [26, 33–36]. In fact, the NExT was initially developed to identify the modal parameters of a parked wind turbine [11]. Moreover, there are many improved OMA algorithms to make them suitable for wind turbines operating under real conditions, these include the automated OMA methods presented in [35].

Last but not least, OMA is a powerful technique for the CM of mechanical systems, and numerous methods have been investigated [10, 24, 25, 27, 37–40]. For example, the Poly-reference LSCE and SSI were employed to identify the modal characterisation of the rear suspension of a family car during road tests as early as 1999 [37]. Recently, an improved SSI method was successfully applied to identify the dynamic

characteristics of a car and a railway vehicle suspension system with the aim of using it for CM [24, 25].

Besides, some researches compared the performance between different OMA approaches by implementing to identify the same structure using the same response. In [41], SSI, PolyMAX, ERA and FDD were employed to monitor a confederation bridge using the same datasets. The identification results from these four methods were compared with the modes calculated by the finite element model of the bridge. The identified results show a good correlation with the modal properties calculated by finite element model. Among the four methods, SSI is more consistent in the estimation of frequency, damping and mode shape. Furthermore, it is worth to highlight that SSI outperforms PolyMAX method in mode shape estimation. However, it is noticeable that all four methods exhibit higher variance in damping estimation especially for ERA and FDD. Another comparative study can be found in which presents similar conclusions.

4 Summary

It can be seen from this short review, OMA is a powerful and robust approach for validating and updating finite element models, SHM or CM, load estimation of a system under ambient excitation or operating scenario. The popular OMA techniques mainly include FDD, SSI, PolyMAX, Bayesian and transmissibility measurements. Each method has its advantages and disadvantages, so we have to select the proper approaches according to application scenario. Besides, it is still an open research area to counter the assumptions of white noise excitation and linear system when employing OMA.

Acknowledgements This research was supported by the Tianjin Natural Science Foundation of China (Grant No.: 18JCYBJC95200), Science & Technology Pillar Program of Tianjin (Grant No.:19YFZCSF01150) and Innovation cultivation Foundation 19-163-12-ZT-006-007-06.

References

1. Bin Zahid, F., Ong, Z.C., Khoom S.Y.: A review of operational modal analysis techniques for in-service modal identification. *J. Braz. Soc. Mech. Sci. Eng.* **42**(8), Jul. 2020, Art. no. 8 (2020). <https://doi.org/10.1007/s40430-020-02470-8>
2. Peeters, B., De Roeck, G.: Stochastic system identification for operational modal analysis: a review. *J. Dyn. Syst. Meas. Control-Trans. Asme* **123**(4), 659–667 (2001). <https://doi.org/10.1115/1.1410370>
3. Brincker, R., Zhang, L.M., Andersen, P.: Modal identification of output-only systems using frequency domain decomposition. *Smart Mater. Struct.* **10**(3), 441–445 (2001). <https://doi.org/10.1088/0964-1726/10/3/303>
4. Jacobsen, N.-J., Andersen, P., Brincker, R.: Using EFDD as a robust technique to deterministic excitation in operational modal analysis. In: 2nd International Operational Modal Analysis

- Conference, IOMAC 2007, April 30, 2007 - May 2, 2007, Copenhagen, Denmark, 2007, p. Aarhus School of Engineering; Bruel and Kjaer
5. Zhang, L., Wang, T., Tamura, Y.: A frequency-spatial domain decomposition (FSDD) method for operational modal analysis. *Mech. Syst. Signal Process.* **24**(5), 1227–1239 (2010). <https://doi.org/10.1016/j.ymsp.2009.10.024>
 6. Van der Auweraer, H., Guillaume, P., Verboven, P., Vanlanduit, S.: Application of a fast-stabilizing frequency domain parameter estimation method. *J. Dyn. Syst. Meas. Control-Trans. Asme* **123**(4), 651–658 (2001). <https://doi.org/10.1115/1.1410369>
 7. Peeters, B., Van der Auweraer, H., Guillaume, P., Leuridan, J.: The PolyMAX frequency-domain method: a new standard for modal parameter estimation? *Shock Vib.* **11**(3–4), 395–409 (2004). <https://doi.org/10.1155/2004/523692>
 8. Peeters, B., El-kafafy, M., Guillaume, P.: The new PolyMAX Plus method: confident modal parameter estimation even in very noisy cases, pp. 17–19. Belgium, Leuven (2012)
 9. Ibrahim, S.: Random decrement technique for modal identification of structures. *J. Spacecr. Rockets* **14**, 696 (1977). <https://doi.org/10.2514/3.57251>
 10. Juang, J.-N., Pappa, R.S.: An eigensystem realization algorithm for modal parameter identification and model reduction. *J. Guid. Control Dyn.* **8**(5), 620–627 (1985). <https://doi.org/10.2514/3.20031>
 11. James, G., Carne, T., Laufer, J.: The natural excitation technique (NExT) for modal parameter extraction from operating structures. *J. Anal. Exp. Modal Anal.* **10** (1995)
 12. Peeters, B., de Roeck, G.: Reference-based stochastic subspace identification for output-only modal analysis. *Mech. Syst. Signal Process.* **13**(6), 855–878 (1999). <https://doi.org/10.1006/mssp.1999.1249>
 13. Au, S.-K., Zhang, F.-L., Ni, Y.-C.: Bayesian operational modal analysis: theory, computation, practice. *Comput. Struct.* **126**, 3–14 (2013). <https://doi.org/10.1016/j.compstruc.2012.12.015>
 14. Zhang, F.-L., Au, S.-K., Lam, H.-F.: Assessing uncertainty in operational modal analysis incorporating multiple setups using a Bayesian approach. *Struct. Control Health Monit.* **22**(3), 395–416 (2015). <https://doi.org/10.1002/stc.1679>
 15. Zhu, Z., Au, S.-K., Li, B., Xie, Y.-L.: Bayesian operational modal analysis with multiple setups and multiple (possibly close) modes. *Mech. Syst. Signal Process.* **150**, 107261 (2021). <https://doi.org/10.1016/j.ymsp.2020.107261>
 16. Devriendt, C., Guillaume, P.: The use of transmissibility measurements in output-only modal analysis. *Mech. Syst. Signal Process.* **21**(7), 2689–2696 (2007). <https://doi.org/10.1016/j.ymsp.2007.02.008>
 17. Devriendt, C., De Sitter, G., Guillaume, P.: An operational modal analysis approach based on parametrically identified multivariable transmissibilities. *Mech. Syst. Signal Process.* **24**(5), 1250–1259 (2010). <https://doi.org/10.1016/j.ymsp.2009.02.015>
 18. Yan, W.-J., Ren, W.-X.: Operational modal parameter identification from power spectrum density transmissibility. *Comput.-Aided Civ. Infrastruct. Eng.* **27**(3), 202–217 (2012). <https://doi.org/10.1111/j.1467-8667.2011.00735.x>
 19. Araujo, I.G., Laier, J.E.: Operational modal analysis using SVD of power spectral density transmissibility matrices. *Mech. Syst. Signal Process.* **46**(1), 129–145 (2014). <https://doi.org/10.1016/j.ymsp.2014.01.001>
 20. Yan, W.-J., Ren, W.-X.: An enhanced power spectral density transmissibility (EPSDT) approach for operational modal analysis: theoretical and experimental investigation. *Eng. Struct.* **102**, 108–119 (2015). <https://doi.org/10.1016/j.engstruct.2015.08.009>
 21. Cao, S., Ouyang, H., Cheng, L.: Baseline-free adaptive damage localization of plate-type structures by using robust PCA and Gaussian smoothing. *Mech. Syst. Signal Process.* **122**, 232–246 (2019). <https://doi.org/10.1016/j.ymsp.2018.12.017>
 22. Cao, S., Ouyang, H., Cheng, L.: Baseline-free multidamage identification in plate-like structures by using multiscale approach and low-rank modelling. *Struct. Control Health Monit.* **26**(2), e2293 (2019). <https://doi.org/10.1002/stc.2293>
 23. Friis, T., Tarpø, M., Katsanos, E.I., Brincker, R.: Best linear approximation of nonlinear and nonstationary systems using operational modal analysis. *Mech. Syst. Signal Process.* **152**, 107395 (2021). <https://doi.org/10.1016/j.ymsp.2020.107395>

24. Liu, F., Wu, J., Gu, F., Ball, A.D.: An introduction of a robust OMA method: CoS-SSI and its performance evaluation through the simulation and a case study. *Shock Vib.* **2019**, 6581516 (2019). <https://doi.org/10.1155/2019/6581516>
25. Liu, F., Zhang, H., He, X., Zhao, Y., Gu, F., Ball, A.D.: Correlation signal subset-based stochastic subspace identification for an online identification of railway vehicle suspension systems. *Veh. Syst. Dyn.* **58**(4), 569–589 (2020). <https://doi.org/10.1080/00423114.2019.1589534>
26. Agneni, A., Coppotelli, G., Grappasonni, C.: A method for the harmonic removal in operational modal analysis of rotating blades. *Mech. Syst. Signal Process.* **27**, 604–618 (2012). <https://doi.org/10.1016/j.ymssp.2011.09.013>
27. Storti, G., Machado, T.: The use of operational modal analysis in the process of modal parameters identification in a rotating machine supported by roller bearings. *J. Mech. Sci. Technol.* **35**(2), Art. no. 2 (2021). <https://doi.org/10.1007/s12206-021-0105-3>
28. Ramos, L.F., Marques, L., Lourenco, P.B., De Roeck, G., Campos-Costa, A., Roque, J.: Monitoring historical masonry structures with operational modal analysis: two case studies. *Mech. Syst. Signal Process.* **24**(5), 1291–1305 (2010). <https://doi.org/10.1016/j.ymssp.2010.01.011>
29. Gentile, C., Saisi, A., Cabboi, A.: Structural identification of a masonry tower based on operational modal analysis. *Int. J. Archit. Herit.* **9**(2), Art. no. 2 (2015). <https://doi.org/10.1080/15583058.2014.951792>
30. Whelan, M.J., Gangone, M.V., Janoyan, K.D., Jha, R.: Operational modal analysis of a multi-span skew bridge using real-time wireless sensor networks. *J. Vib. Control* **17**(13), Art. no. 13 (2011). <https://doi.org/10.1177/1077546310373058>.
31. Cross, E.J., Koo, K.Y., Brownjohn, J.M.W., Worden, K.: Long-term monitoring and data analysis of the tamar bridge. *Mech. Syst. Signal Process.* **35**(1–2), Art. no. 1–2 (2013). <https://doi.org/10.1016/j.ymssp.2012.08.026>
32. Reynders, E., Maes, K., Lombaert, G., De Roeck, G.: Uncertainty quantification in operational modal analysis with stochastic subspace identification: validation and applications. *Mech. Syst. Signal Process.* **66–67**, 13–30 (2016). <https://doi.org/10.1016/j.ymssp.2015.04.018>
33. Chauhan, S., Hansen, M., Tcherniak, D.: Application of operational modal analysis and blind source separation/independent component analysis techniques to wind turbines. *Conf. Proc. Soc. Exp. Mech. Ser.* (2009)
34. Chauhan, S., et al.: Operational modal analysis of operating wind turbines: application to measured data. In: *Conference proceedings of the society for experimental mechanics series*, vol. 5, pp. 65–81 (2011)
35. Devriendt, C., Magalhaes, F., Weijtjens, W., De Sitter, G., Cunha, A., Guillaume, P.: Structural health monitoring of offshore wind turbines using automated operational modal analysis. *Struct. Health Monit.- Int. J.* **13**(6), Art. no. 6 (2014). <https://doi.org/10.1177/1475921714556568>
36. Hansen, M.H., Thomsen, K., Fuglsang, P., Knudsen, T.: Two methods for estimating aeroelastic damping of operational wind turbine modes from experiments. *Wind Energy* **9**(1–2), Art. no. 1–2 (2006). <https://doi.org/10.1002/we.187>
37. Hermans, L., Van der Auweraer, H.: Modal testing and analysis of structures under operational conditions: industrial applications. *Mech. Syst. Signal Process.* **13**(2), Art. no. 2 (1999). <https://doi.org/10.1006/mssp.1998.1211>
38. Qi, K., He, Z., Li, Z., Zi, Y., Chen, X.: Vibration based operational modal analysis of rotor systems. *Measurement* **41**(7), Art. no. 7 (2008). <https://doi.org/10.1016/j.measurement.2007.11.009>
39. Pierro, E., Mucchi, E., Soria, L., Vecchio, A.: On the vibro-acoustical operational modal analysis of a helicopter cabin. *Mech. Syst. Signal Process.* **23**(4), Art. no. 4 (2009). <https://doi.org/10.1016/j.ymssp.2008.10.009>
40. Dollon, Q., Antoni, J., Tahan, A., Gagnon, M., Monette, C.: Operational modal analysis of hydroelectric turbines using an order based likelihood approach. *Renew. Energy* **165**, 799–811 (2021). <https://doi.org/10.1016/j.renene.2020.11.086>
41. Rahman, M.S., Lau, D.T.: A comparative study of system identification techniques under ambient vibration. *Top. Dyn. Civ. Struct.* **4**, 145–154 (2013). https://doi.org/10.1007/978-1-4614-6555-3_17

Vibration Analysis of the Rudder Drive System of an Underwater Glider



Liming Guo, Jing Liu, Guang Pan, Baowei Song, Yonghui Cao, Yong Cao, Yujun Liu, and Hengtai Ni

Abstract Underwater gliders are mainly used to monitor the marine environment. In order to reduce the interference of self-vibrations, the rudder drive system of gliders is analyzed. Firstly, the meshing frequencies of the gears are calculated. Then, the vibrations of the actuator of rudder drive system are tested; and a finite element model of the rudder drive system of underwater glider is developed. Finally, the gear meshing frequencies and the vibration frequencies of the actuator are compared with the results from the finite element model analysis. The results show that the gear meshing frequencies and the vibration peak frequencies of the actuator are different from the natural frequencies of the rudder drive system, the system has no resonance and the structure design is reasonable one.

Keywords Vibration analysis · Vibration test · Underwater glider · Rudder drive system

1 Introduction

Underwater gliders are relatively new types of underwater vehicle. It relies on the change of its own buoyancy to achieve the floating and diving. It can glide forward through the horizontal component of the lift underwater [1]. Figure 1 shows the trajectory of the underwater glider.

The research of this paper is based on the flying wing underwater glider platform. Compared with the cylindrical underwater glider, it has a larger glider wing and higher lift drag ratio. There is a rotatable rudder blade on the wing, as shown in Fig. 2. The roll attitude of the glider can be changed by changing the angle of the rudder blade.

L. Guo · J. Liu (✉) · G. Pan · B. Song · Y. Cao · Y. Cao · Y. Liu · H. Ni
School of Marine Science and Technology, Northwestern Polytechnical University, Xian 710072, China
e-mail: jliu0922@nwpu.edu.cn

L. Guo
e-mail: guoliming@mail.nwpu.edu.cn

Fig. 1 Glide trajectory of underwater glider

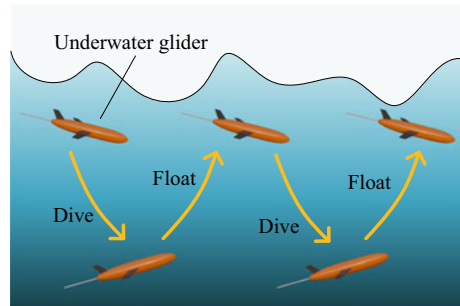
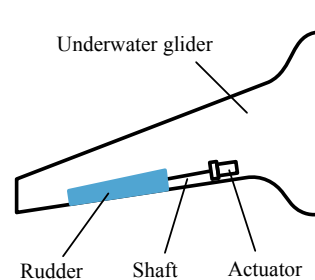


Fig. 2 Rudder drive system of underwater glider



Suppressing methods of the harmful vibrations in different mechanical systems is an important research field. Ishida and Inoue [2] investigated the internal resonance phenomena of an asymmetrical shaft theoretically and experimentally in the vicinities of the major critical speed, and twice and three times the major critical speed. Łuczak [3] studied mechanical resonance in direct drive servo-mechanical systems by using frequency analysis method. Xiang et al. [4] carried out a comparative study of the rudder transmission vibration characteristics of electric drive and traditional hydraulic drive. The results indicate that the vibration intensity has a positive correlation with the rudder blade loading force and angular velocity at the zero balance position. The electric drive can more effectively reduce the overall vibration of the rudder transmission. Mansoor and Al-shammari [5] studied the vibration of gas turbine rotor with the existence of cracks and without them. Huang et al. [6] established a simplified lumped-mass model to investigate coupled torsional-longitudinal vibrations of a ship's propeller shaft. It is found that the natural frequencies are unaffected while the maximum acceleration are increased with the rotational speed as well as the loading.

Because of the long span and thin thickness of the rudder blade, if resonance occurs during operation, it's quite easy to cause structural damage and interference noise. Therefore, this paper analyzes the mode of rudder drive system. The vibration frequencies of the actuator system are tested, which is used to verify the rationality of structural design.

Table 1 Parameters of gearbox

Components	Part	Gear module	Teeth number of gear	Pressure angle
One-stage gear drive	Motor gear	0.4	18	20°
	One-stage large gear	0.4	60	20°
Two-stage gear drive	Two-stage small gear	0.4	16	20°
	Two-stage large gear	0.4	60	20°
Worm drive	Worm	0.8	1	20°
	Worm wheel	0.8	40	20°

Table 2 Meshing frequencies of gears

Components	Meshing frequencies (Hz)
One-stage gear drive	3.87
Two-stage gear drive	232
Worm drive	870

2 Calculation of Gear Meshing Frequency of Actuator

The meshing frequencies of the gear of the actuator are analyzed, and the parameters of the gears are shown in Table 1.

The rotation speed of rudder plate is 5.8 r/min. The meshing frequencies of gearbox can be calculated as shown in Table 2.

3 Vibration Test and Analysis of Actuator System

In order to study the influences of the vibrations of the actuator system on the rudder drive system, the vibration frequencies of the actuator system under the working conditions are tested and analyzed. As shown in Fig. 3, two acceleration sensors are used to test the two positions of the side and top face of the actuator system. The test signal is transmitted to the computer through the acquisition card. Then, the time domain signal is transformed into the frequency domain signal by using the fast Fourier Transform method. Table 3 shows the peak vibration frequencies of the actuator system.

In order to avoid the accidental error misleading the experimental results and analysis in the process of experiment, eight groups of experiments are carried out in vibration test, one of the experimental data are shown in Fig. 4.

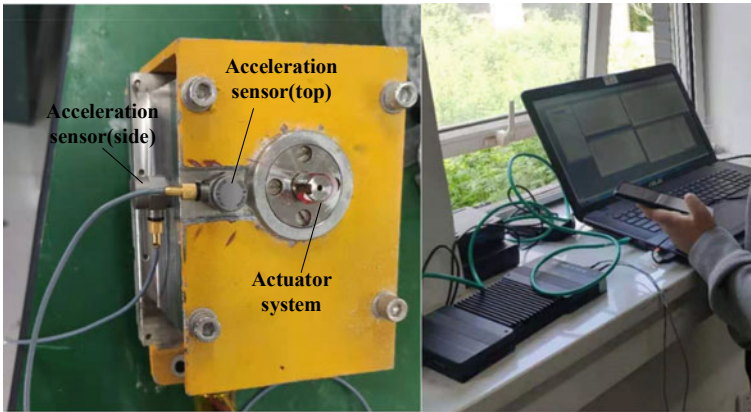


Fig. 3 Vibration test experiment of the actuator system

Table 3 Vibration frequency of the actuator system

Measuring position	Peak vibration frequency of actuator/Hz				
Top face	137.5	800	1600	2125	4327.5
Side face	50	800	1600	2125	4327.5

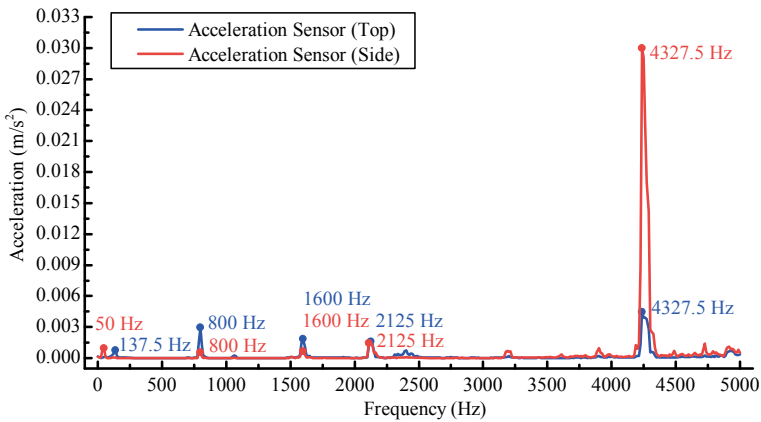
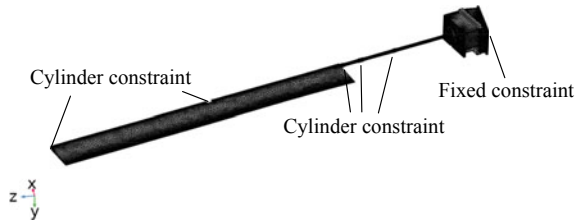


Fig. 4 Vibration spectrum of the actuator system

As shown in Fig. 4, the peak frequencies of the vibration signals of the top face of the drive system are larger than those of the side. The main peak frequencies include 50, 137.5, 800, 1600, 2125, and 4327.5 Hz.

Fig. 5 Vibration model of the rudder drive system



4 Simulation Results

In order to study the resonance phenomenon of the rudder drive system of underwater glider, the mode analysis of the rudder drive system is carried out to explore its natural frequencies.

4.1 A Finite Element Model of Rudder Drive System

The rudder drive system is mainly composed of actuator system, drive shaft and rudder. The rudder and drive shaft are connected by a coupling. The drive shaft and rudder are relatively fixed. During operation, the rotating shaft drives the rudder to rotate through the drive shaft. The finite element method is used for mode analysis of the rudder drive system. The boundary constraints of the whole system are shown in Fig. 5.

4.2 Modal Analysis Results of Rudder Drive System

According to the frequency analysis of actuator system in Sect. 3, the frequencies of large amplitude include 50, 137.5, 800, 1600, 2125, 4327.5 Hz. Therefore, the mode frequencies of rudder drive system in the range of 0–5000 Hz are analyzed. The mode frequencies are shown in Table 4.

As shown in Table 4, in the range of 0–200 Hz, mode frequencies of rudder drive system include 80.604, 186.52, 197.17 Hz; in the range of 700–900 Hz, mode frequencies of rudder drive system include 727.3, 762.19, 831.11 Hz; in the range of 1500–1700 Hz, the mode frequencies of rudder drive system include 1647.3 Hz; in the range of 2000–2200 Hz, the mode frequencies of rudder drive system include 2050.1 Hz, 2056.2 Hz; in the range of 4300–4400 Hz, the mode frequencies of rudder drive system include 4360.1, 4378.4, 4380.3 Hz; The mode frequencies of rudder drive system are different from the peak frequencies of the actuator system, which indicates that there is no resonance in the whole structure of the rudder drive system,

Table 4 Mode frequencies of the rudder drive system

Frequency range/Hz	Mode	Frequency/Hz	Actuator vibration frequency/Hz
0–200	1	80.604	137.5
	2	186.52	
	3	197.17	
700–900	11	727.3	800
	12	762.19	
	13	831.11	
1500–1700	23	1647.3	1600
2000–2200	27	2050.1	2125
	28	2056.2	
4300–4400	55	4360.1	4327.5
	56	4378.4	
	57	4380.3	

The structure design of rudder drive system is reasonable. The mode shape of each mode of the rudder drive system is shown in Fig. 6.

5 Conclusions

In this paper, the vibrations of the rudder drive system in an underwater glider are analyzed. The meshing frequencies of the gears in gearbox of the actuator system are calculated. The main vibration frequencies test of the actuator system is carried out. Compare the mode frequencies of the rudder drive system with the peak vibration frequencies of the actuator system. The result showed that the vibration peak value of the actuator system does not match the mode frequencies of the rudder drive system. Therefore, no resonance occurred on the whole structure, the structure is reasonable designed.

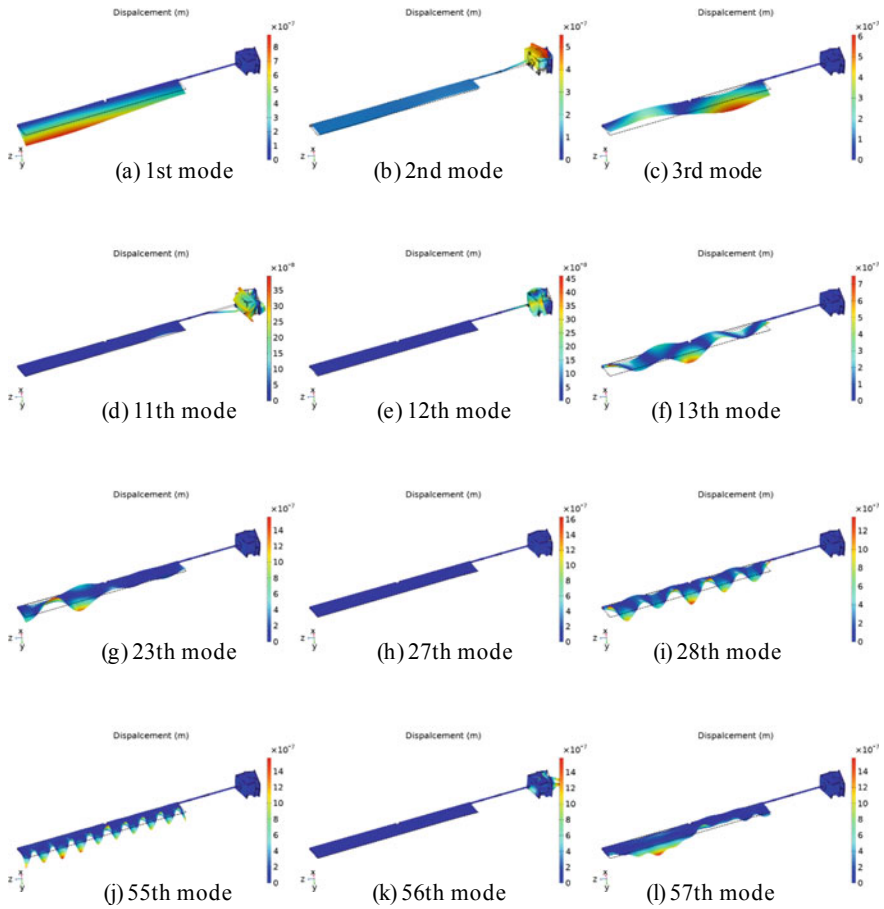


Fig. 6 Mode shape of each mode of rudder drive system

Acknowledgements Supported by the Fundamental Research Funds for Central Universities (No. 3102020HHZY030001).

References

1. Zhang, Y., Zhang, Z., Quan, Z., et al.: Hydrodynamic performance and calculation of lift–drag ratio on underwater glider. *J. Mar. Sci. Technol.* **26**(1), 16–23 (2021)
2. Ishida, Y., Inoue, T.: Internal resonance phenomena of an asymmetrical rotating shaft. *J. Vib. Control* **11**(9), 1173–1193 (2005)
3. Łuczak, D.: Frequency analysis of mechanical resonance in direct drive. In: 2012 12th IEEE International Workshop on Advanced Motion Control (AMC). IEEE, 2012, pp. 1–5

4. Xiang, K., Li, W., Lan, Q.: Comparative study of rudder transmission vibration characteristics of electric and hydraulic drive. *Chinese J. Ship Res.* **12**(6)
5. Mansoor, H.I., Al-shammari, M.A., Al-Hamood, A.: Experimental analysis of cracked turbine rotor shaft using vibration measurements. *J. Mech. Eng. Res. Dev.* **43**(2), 294–304 (2020)
6. Huang, Q., Yan, X., Wang, Y., et al.: Numerical modeling and experimental analysis on coupled torsional-longitudinal vibrations of a ship's propeller shaft. *Ocean Eng.* **136**, 272–282 (2017)

Online Method for Assessment and Tracking of Wear in Kaplan Turbine Runner Blades Operating Mechanism



Oscar García Peyrano, Daniel Vaccaro, Rodrigo Mayer,
and Matías Marticorena

Abstract Kaplan turbines rely on an operating mechanism inside the runner hub to control blade angles. Contact surfaces of the moving parts on these mechanisms are constantly subjected to frictional and contact forces, inflicting wear which can lead to malfunctioning and performance reduction. In this paper, a novel method for wear assessment in individual blade joints of the runner operating mechanism is presented. The technique consists in monitoring blade angles separately during turbine operation through inductive proximity probes mounted on the discharge ring. These angles are contrasted with the operating mechanism positioning data at several instants and the performance of each joint is evaluated. This technique has been implemented on a 92 MW Kaplan turbine. In October 2018, excessive clearances in three blade joints were detected and an inspection was recommended during the next programmed maintenance. The runner hub was later disassembled and all joints inspected, which confirmed those joints had been worn down and were replaced. This result shows that the proposed method can effectively assess clearances on blade joints during operation, providing an early detection method to anticipate mechanism malfunction and incorporate in the Condition-Based Maintenance plan for production optimization.

O. García Peyrano · R. Mayer · M. Marticorena (✉)
Vibration Division, Bariloche Atomic Center, National Atomic Energy Commission, San Carlos de Bariloche, Argentina
e-mail: matiasmarticorena@cab.cnea.gov.ar; matias.marticorena@cab.cnea.gov.ar;
matias.marticorena@ib.edu.ar

O. García Peyrano
e-mail: garciapeyrano@yahoo.com.ar

O. García Peyrano
GPSignal, San Carlos de Bariloche, Argentina

O. García Peyrano · R. Mayer · M. Marticorena
Balseiro Institute, Cuyo National University, San Carlos de Bariloche, Argentina

D. Vaccaro
Pichi Picún Leufú Hydro Power Plant, Pampa Energía SA, Piedra del Águila, Argentina

Keywords Condition-based maintenance · Kaplan turbine · Operating mechanism · Machine health monitoring

1 Introduction

Kaplan hydro turbines are widely utilized throughout the world, as they can achieve efficient performance in high flow, low head applications thanks to their combination of adjustable runner blades with adjustable wicket gates.

This work focuses on the blade operating mechanism in charge of automatically adjusting runner blade angles with respect to the direction of water flow so as to achieve maximum efficiency at a given set of conditions. As any collection of moving parts in a machine, the individual components of this mechanism are prone to wear [1–3]. Moreover, runner blades are normally subjected to intense fluctuating hydrodynamic forces, which are transmitted to these components and therefore increase mechanical solicitations [4, 5]. These factors have been shown to limit the lifetime of operating mechanism parts [6–8] and even causing severe damage to the machine, leading to substantial losses due to unexpected shutdowns, prolonged downtime and asset replacement [9, 10].

Thus, great effort has been put into failure analysis and lifetime estimation of runner blades and operating mechanism [6–11]. In contrast, condition monitoring of these components has received comparatively less attention comparatively. While various types of instrumentation and diagnosis techniques are commonly implemented on hydraulic turbines, the focus is mainly put into monitoring bearing condition and behavior [12], structural integrity and on detecting the presence of rubs and cavitation [13]. The runner blade operating mechanism, on the other hand, poses a further challenge for online monitoring due to being located inside the runner hub. Assessment of its condition, therefore, often falls to scheduled maintenance at the risk of fault development going undetected.

This work presents a novel method for identifying increased clearances in blade operating mechanism components using previously installed instrumentation on a Kaplan turbine. In addition to enabling online evaluation of said components and anticipating failure, this approach provides useful feedback for lifetime estimations and condition-based maintenance.

2 Methods and Measuring Equipment

The proposed method was implemented on a 92 MW vertical Kaplan turbine with five runner blades and a runner diameter of 7800 mm. Figure 1 shows a 3D model of the turbine blade operating mechanism and its components.

The mechanism consists of a hydraulic piston that moves a cylinder body attached to a crosshead along the axial direction. The crosshead is connected to five fork

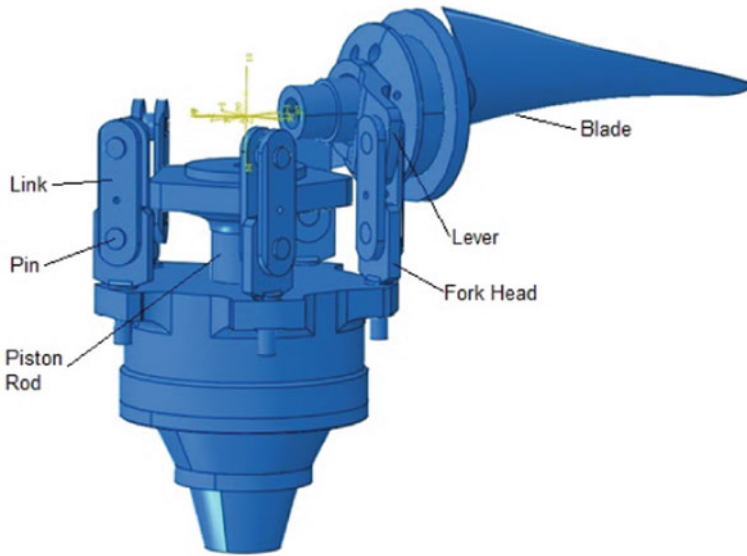


Fig. 1. 3D model of runner blade operating mechanism

heads, each of which is linked to a lever that can pivot around a radial direction. As the crosshead moves upwards or downwards, each lever will rotate a certain amount alongside with its corresponding blade. Lever angles β can vary in this way from -16° (closed runner position) to 14° (opened runner position), as shown in Fig. 2a and 2b, respectively.

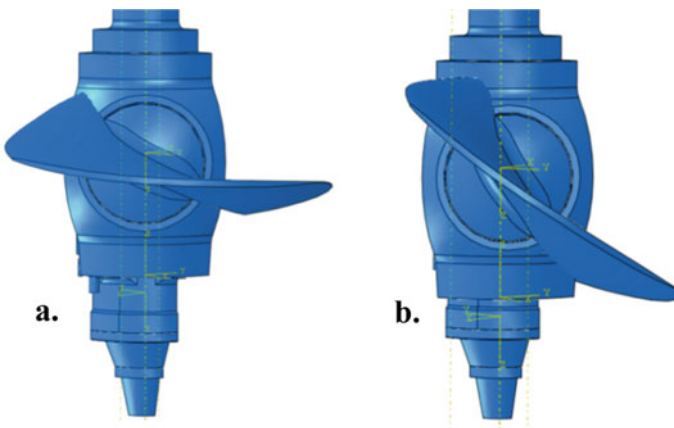


Fig. 2 Blade angle maximum and minimum positions. **a** Closed runner ($\beta_c = -16^\circ$). **b** Opened runner ($\beta_o = 14^\circ$)

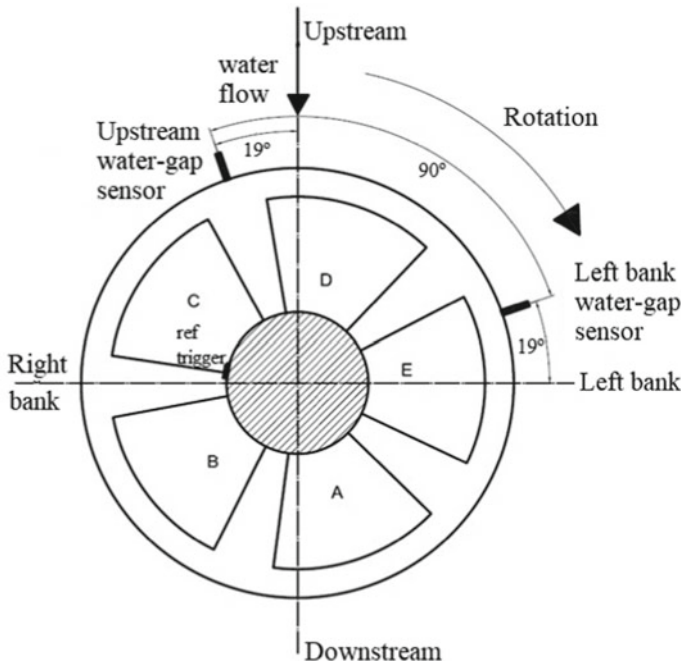


Fig. 3 Schematic showing blade nomenclature and water-gap sensors positions

2.1 Instrumentation

The turbine is equipped with magnetic pickup sensor serving as a 1X reference signal.

Figure 3 shows an upper view schematic of the runner. Between the blade lips and the discharge ring there is a water gap which must be monitored to ensure it remains within the range of 2 and 8 mm. Two proximity probes have been installed to that end and integrated to the machine's protection system. One of them is located on the upstream side of the discharge ring and the other, 90° apart from it, on the left bank side. Runner blades have been labeled from A to E.

Runner opening is inferred from the axial position of the cylinder body and expressed as percentage of maximum opening at lever angle $\beta_o = 14^\circ$ (Fig. 2b). Data were acquired at 4096 samples/s.

2.2 Blade Angle Estimation

When all mechanism components are in optimal conditions, a smooth transition between two blade angles should take place, following cylinder body movement. If a significant amount of clearance exists on one or several mechanism links, an

uneven transition is expected. This can be caused either by a time delay between cylinder and lever movement or by hydrodynamic forces acting upon the blade (whose movement would be relatively unrestrained due to looseness on the linkage). Therefore, the condition of each linkage could theoretically be assessed by comparing cylinder position data to blade angle data during mechanism actuation.

In order to estimate individual blade angles, data from the upstream water-gap sensor was employed. This method takes advantage of the pulse-like time signal obtained from the proximity probes located on the turbine’s discharge ring every time one of the blades passes in front of them, as Fig. 4 shows.

Each individual pulse will have an associated pulse width, measured in seconds, which will depend on blade lip width e , blade angle α and runner angular speed ω , as illustrated on Fig. 5.

Let D be the runner radius (measured from the center of the shaft to the blade lip), ω the angular speed and Δt the measured pulse width. The length a that the blade lip will present to the proximity probe, corresponding to the pulse width, can be expressed as

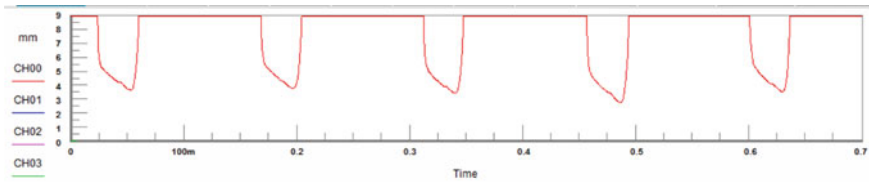
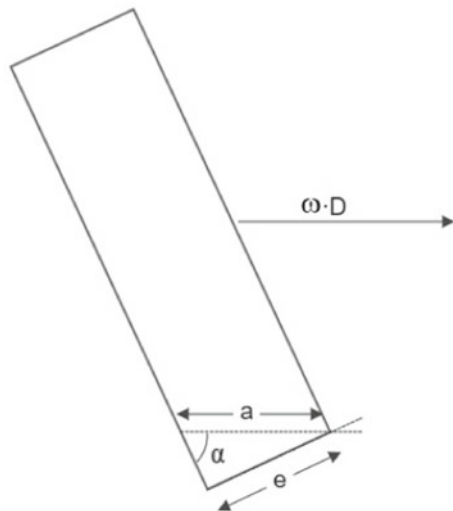


Fig. 4 Water-gap sensor time signal

Fig. 5 Blade angle estimation from pulse width measurement



$$a = \omega \cdot D \cdot \Delta t = \frac{2\pi}{T} \cdot D \cdot \Delta t \quad (1)$$

where T is the runner's rotation period. Given that

$$\sin\alpha = \frac{e}{a} \quad (2)$$

Therefore α can be expressed in terms of the measured pulse width as

$$\alpha = \sin^{-1} \frac{e}{\omega \cdot D \cdot \Delta t} \quad (3)$$

Δt was measured from the instant the signal from the proximity probe decreases from its base value of 9 mm (sensor maximum range) until the instant it returns to this value.

The lever angle of each blade is related to α by the expression:

$$\beta_n = \alpha_n + k \quad (4)$$

where n identifies each blade and k is a known calibration constant. The value of e is obtained from Eq. (3) by calculating α with Eq. (4) at the appropriate runner opening percentage.

2.3 Lever Angle Uncertainty

As stated above, data acquisition was performed at 4096 samples/s, which means there is a 0.2 ms time lapse between measurements. Therefore, pulse width uncertainty can be calculated as

$$\sigma_{\Delta t} = \sqrt{2}(0.2 \text{ ms}) = 0.28 \text{ ms} \quad (5)$$

Let σ_e be the uncertainty of variable e and σ_ω that of ω . At a frequency of $f = 1.38 \text{ Hz}$ the rotation period $T = 0.725 \text{ s}$. As with Δt , the uncertainty of T is given by the data acquisition sample rate, thus $\sigma_T = \sigma_{\Delta t}$. Error propagation of Ec. (1) yields:

$$\sigma_\omega = \sqrt{\left(\frac{2\pi\sigma_{\Delta t}}{T^2}\right)^2} = 0.003 \text{ s}^{-1} \quad (6)$$

and

$$\sigma_a = \sqrt{(D\Delta t\sigma_\omega)^2 + (\omega D\sigma_{\Delta t})^2} = 9.6 \text{ mm} \quad (7)$$

Table 1 Blade lip widths calculated from pulse widths Δt , measured at 22.1% runner opening

Blade	Pulse width (Δt) (ms)	Blade width (e) (mm)
A	17.1	65.78
B	16.9	64.71
C	17.0	71.09
D	17.0	68.96
E	17.3	66.73

Table 1 lists the values of e calculated from Eq. (3) using values of Δt measured at a runner opening of 22.1% and blade angles $\alpha = 6.57^\circ$ for all blades, according to Eq. (4).

These values can be used with Eq. (2) to obtain blade width uncertainty σ_e and blade angle uncertainty σ_α :

$$\sigma_e = \sqrt{(\sin\alpha \cdot \sigma_a)^2} = 1.095 \text{ mm} \tag{8}$$

$$\sigma_\alpha = \sqrt{\left(\frac{\sigma_e}{a\sqrt{1 - (\frac{e}{a})^2}}\right)^2 + \left(\frac{-e\sigma_a}{a^2\sqrt{1 - (\frac{e}{a})^2}}\right)^2} = 0.2^\circ \tag{9}$$

Finally, lever angle uncertainty can be expressed from Eq. (4) as:

$$\sigma_\beta = \sigma_\alpha = 0.2^\circ \tag{10}$$

3 Operating Mechanism Test

A test was performed on October 2018 with the purpose of evaluating the state of blade operating mechanism joints. It consisted in measuring runner opening percentage (computed from cylinder position) alongside with blade angles estimated from pulse width measurements as described above. After reaching steady state operation at a rotation frequency of $f = 1.38Hz$, runner opening was slowly decreased from 22 to 16% by means of the operating mechanism, along a 70 s time period. Blade angle data were then computed and analyzed looking for signs of mechanism looseness.

Figures 6, 7, 8 and 9 show percentage of runner opening and estimated lever angle during the test. The first two minutes were used to reach steady state operation and do not appear in the graphs.

Blade A shows significant data dispersion on estimated angle throughout the test. Maximum variation between two adjacent points is 0.2° .

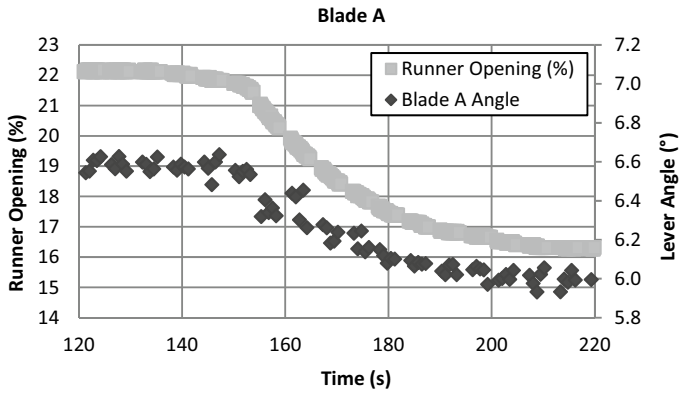


Fig. 6 Percentage of runner opening and estimated lever angle versus time on blade A

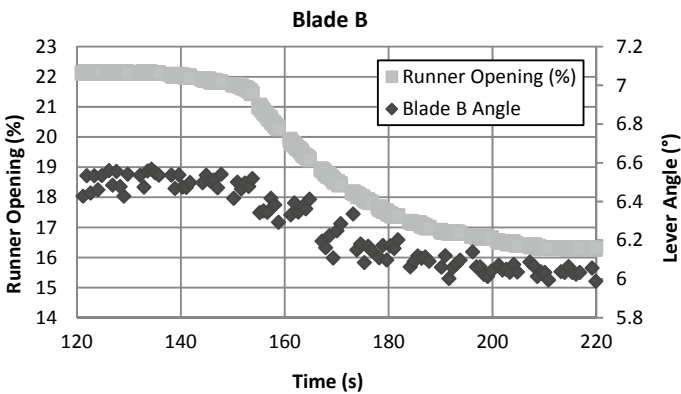


Fig. 7 Percentage of runner opening and estimated lever angle versus time on blade B

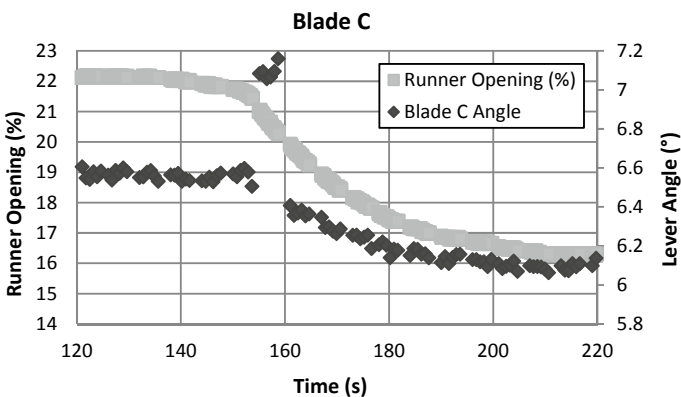


Fig. 8 Percentage of runner opening and estimated lever angle versus time on blade C

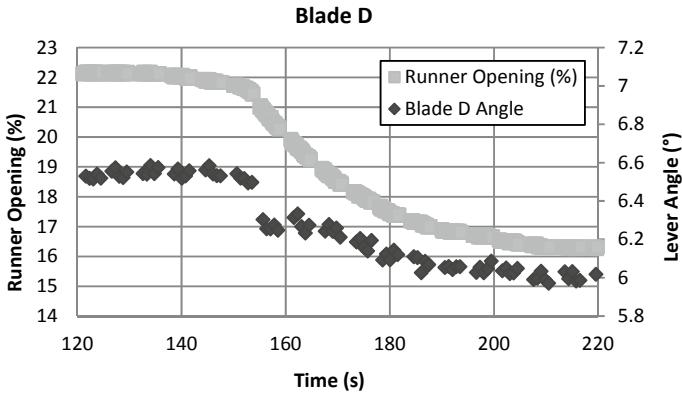


Fig. 9 Percentage of runner opening and estimated lever angle versus time on blade D

Blade B evidences higher angular data dispersion than A, with variations between adjacent points up to 0.4° .

Blade C angular data is less dispersed throughout the test, but shows a pronounced discontinuity between 155 and 159 s, of up to 0.8° .

Blade D also has lower data dispersion than A and B, and a visible discontinuity at 156 s of 0.2° (Fig. 10).

Finally, blade E shows lower data dispersion previous to runner opening decreasing and higher dispersion afterwards. A discontinuity in blade angle of 0.3° can be seen at 155 s, shortly after mechanism actuation begins.

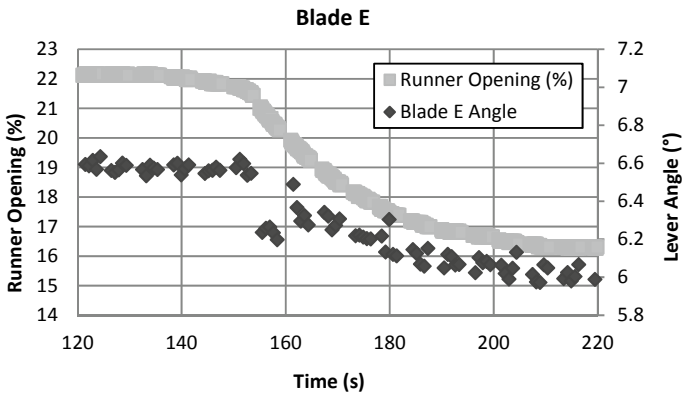


Fig. 10 Percentage of runner opening and estimated lever angle versus time on blade E

4 Joint Condition Assessment

After the above method had been implemented and data from the operating mechanism test was analyzed, the authors recommended the inspection of all blade joints. This was done during maintenance tasks performed on March 2020, where several joints were found to have sustained significant wear.

Fork head and lever joints are coated with 0.5 mm self-lubricated Karon V liners. In order to estimate joint clearances, diameters were measured on every link joint and its respective pin corresponding to each blade, then liner thickness was assessed. Clearances were therefore calculated based on the amount of Karon V material that had been worn out. If no liner material remained on the joint surface, a clearance of 1 mm was assumed for that component. Total clearances result from adding up all estimated clearances on a given linkage (fork head joint orifice and pin plus lever joint orifice and pin).

4.1 Wear Measurement

Results from maintenance performed on the operating mechanism are listed on Table 2. Blade A showed little signs of wear (0.04 mm total calculated clearance). Blades B, D and E had sustained greater damage, with at least one of their joint components completely stripped of liner material (total clearances between 1 mm and 1.5 mm). Blade C was in the worst condition, with no liner material left on two of its joint components (2.05 total clearance).

Table 2 Linkage joints analysis performed on March 2020

Blade	Joint	Fork head	Lever	Total calculated clearance (mm)
		Calculated clearance (mm)	Calculated clearance (mm)	
A	Link	0.02	0.00	0.04
	Pin	0.02	0.00	
B	Link	0.05	0.00	1.01
	Pin	0.06	0.90	
C	Link	1.05	1.00	2.05
	Pin	0.00	0.00	
D	Link	1.05	0.40	1.49
	Pin	0.04	0.00	
E	Link	0.03	0.00	1.05
	Pin	1.02	0.00	

5 Discussion

Discontinuities in blade angle transitions found in blades C, D and E result from an abrupt change in measured pulse width while the operating mechanism is being actuated, which is in turn considered as indicative of looseness. It is worth noting that, while C shows the highest angle difference between adjacent data, it also displays the least dispersion, which implies that both these parameters are influenced by individual blade behavior and properties (such as blade lip regularity and mechanism looseness) rather than by factors common to all of them, such as rotor vibration.

Angular data dispersion on all blades is consistent with the angular uncertainty calculated on Eq. (10). Blade A, displayed little to no signs of mechanism wear but higher dispersion, which suggests that variations of blade lip geometry could constitute a source of error in blade angle estimation. The fact that this dispersion affects some blades but not others, regardless of the amount of clearance found in their mechanism components, indicates again that other possible sources of error, such as torsional, lateral or axial vibrations, are not the cause. This also shows that data dispersion is not a reliable predictor of mechanism wear, as it is lower in blades C and D, which were in worse condition. Instead, it should be attributed to random measuring uncertainty.

Maximum angular variation between adjacent data (i.e. discontinuities in blade angle transitions) proved to be a consistent sign of mechanism looseness, as shown with blades C, D and E. However, data dispersion in the vicinity has to be taken into account in order to avoid misdiagnosing joint wear. In case of blade B a certain amount of mechanism looseness was detected during inspection, but wasn't evident in blade angle performance. It should be taken into account that there was a 18 months interval between performance test and mechanism inspection, during which blade B linkage elements could have deteriorated. It is therefore advisable to inspect all blade joints if looseness was detected in at least one of them, such that maintenance was required.

6 Conclusions

A method for online monitoring and evaluation of the runner blade operating mechanism of Kaplan turbines has been developed and tested on a 92 MW machine.

Results show that discontinuities in blade angle transitions constitute a reliable indicator of excessive clearances in mechanism joints, product of wear. Hence, continuous monitoring of this type provides a valuable resource for machine health assessment, early failure detection and condition-based maintenance. Moreover, this technique utilizes proximity probes installed on the turbine's discharge ring, thus avoiding the need for intrusive interventions inside the runner hub.

Acknowledgements Special thanks are due to Pampa Energía SA and personnel of Pichi Picún Leufu Hydropower Station for their contribution to the experimental activities of this work.

References

1. Liu, X., Luo, Y., Wang, Z.: A review on fatigue damage mechanism in hydro turbines. *Renew. Sustain. Energy Rev.* **54**(1), 1–14 (2016)
2. Luo, Y., Wang, Z., Zeng, J., Lin, J.: Fatigue of piston rod caused by unsteady, unbalanced, unsynchronized blade torques in a Kaplan turbine. *Eng. Fail. Anal.* **17**(1), 192–199 (2010)
3. Liu, X., Luo, Y., Wang, Z.: Fatigue analysis of the piston rod in a kaplan turbine based on crack propagation under unsteady hydraulic loads. *IOP Conf. Ser.: Earth Environ. Sci.* **22**(1) (2014)
4. Wang, Z.W., Luo, Y.Y., Zhou, L.J., Xiao, R.F., Peng, G.J.: Computation of dynamic stresses in piston rods caused by unsteady hydraulic loads. *Eng. Fail. Anal.* **15**, 28–37 (2008)
5. Soltani Dehkharghani, A., Engström, F., Aidanpää, J., Cervantes, M.: An indirect measurement methodology to identify load fluctuations on axial turbine runner blades. *Sensors* **20**(24), 7220 (2020)
6. Budai, A.M., Campian, V.C., Frunzaverde, D., Miclosina, C., Pepa, D.: Lifetime Estimations of the operating mechanism of kaplan turbine runner blades/Pročjene trajanja radnog mehanizma lopatica rotora Kaplan turbine. *Tehnicki Vjesnik - Technical Gazette* **24**(2), 271+ (2017)
7. Larin, O., Trubayev, O., Vodka, O.: The fatigue life-time propagation of the connection elements of long-term operated hydro turbines considering material degradation. *PNRPU Mech. Bulletin*, 1, 167–193 (2014)
8. Budai, A.M., Campian, V.C., Cojocar, A., Korca, Z., Dumbrava, C.: Estimation of Lifetime Duration for a lever pin of runner blade operating mechanism using a graphic—analytic method. In: *Ann “Eftimie Murgu” University from Resita XXII*(2), 67–76 (2015)
9. Zhang, M., Valentín, D., Valero, C., Egusquiza, M., Egusquiza, E.: Failure investigation of a Kaplan turbine blade. *Eng. Fail. Anal.* **97**(1), 690–700 (2019)
10. Urquiza, G., García, J.C., González, J.G., Castro, L., Rodríguez, J.A., Basurto-Pensado, M.A., Mendoza, O.F.: Failure analysis of a hydraulic Kaplan turbine shaft. *Eng. Fail. Anal.* **41**(1), 108–117 (2014)
11. García Peyrano, O., Marticorena, M., Japaz, N., Giraudo, P., Sanchez Sarmiento, G.: Virtual machine diagnosis: dynamic behavior monitoring of a 92 MW Kaplan turbine rotor mechanism during machine production. *HydroVision Brazil*, Rio de Janeiro, Brazil (2012)
12. Monkova, K., Monka, P., Hric, S., Kozak, D., Katinić, M., Pavlenko, I., Liaposhchenko, O.: Condition monitoring of kaplan turbine bearings using Vibro-diagnostics. *Int. J. Mech. Eng. Robot. Res.* **9**(8), 1182–1188 (2020)
13. Lahdelma, S., Juuso, E.K.: Vibration analysis of cavitation in Kaplan water turbines. In: *IFAC Proceedings* **41**(2), 13420–13425, Seoul, Korea (2008)

Oscar García Peyrano, PhD is Associate Professor at Balseiro Institute, Cuyo National University of Argentina. He is also the director of GPSignal condition monitoring company and consultant of the Vibration Laboratory of the National Atomic Energy Commission of Argentina (CNEA).

Daniel Vaccaro is the manager of Pichi Picún Leufú Hydro Power Plant, Pampa Energía SA.

Rodrigo Mayer is a member of the Vibration Division of CNEA and postgraduate student at Balseiro Institute.

Matías Marticorena is the Head of the Vibration Division of CNEA and Instructor at Balseiro Institute, Cuyo National University of Argentina.

Bearing Fault Diagnosis Based on Improved Residual Network



Haofei Du, Chao Zhang, and Jianjun Li

Abstract In the wind power generation system, the bearing plays a very important role. Whether it can run stably directly determines the quality of the electricity produced and has a great influence on the efficiency of power generation. Due to the harsh working environment, the bearing has become one of the most vulnerable components in the entire wind turbine system. Therefore, bearings of wind turbines need to be maintained regularly. However, it needs to be shut down every time for maintenance, which will incur high maintenance cost. So, the fault diagnosis of the bearing is particularly important. A fault diagnosis method is proposed based on deep learning in this paper. This method is based on the residual module to construct a new ResNet model and embeds the attention mechanism in it to select information that is more critical to the current task goal from a lot of information. In addition, a long short-term memory is added to the network to extract the long-term dependence of the vibration signal and ensure that the information on the time series will not be lost as the training progresses. The experimental results show that the method proposed in this paper is very effective for the fault classification of fan bearings.

Keywords Residual network · Attention mechanism · Long short-term memory network · Fault diagnosis

1 Introduction

In recent years, with the national attention to environmental protection, new energy power generation has been greatly supported by the state. Compared with traditional power generation, new energy power generation is more green and environmentally

H. Du · J. Li (✉)

School of Information Engineering, Inner Mongolia University of Science and Technology,
Baotou 014010, China
e-mail: xidianjj@163.com

C. Zhang

School of Mechanical Engineering, Inner Mongolia University of Science and Technology,
Baotou 014010, China

friendly. Wind energy is a clean and renewable energy, so wind power is promoted and constructed in many countries. As a major carbon emitter, China must vigorously develop new energy power generation in order to achieve the goal of carbon neutrality in the future. Therefore, in recent years, more and more wind farms have been built across the country. With the widespread use of wind turbines, related problems have also followed. Since wind turbines are mostly installed on the tops of mountains, deserts or islands, their working environment is relatively harsh. The internal parts of the turbines may be damaged due to the harsh environment, which affects the efficiency of power generation. Among wind turbine components, the most easily damaged are the gearbox. This is because wind turbines had to operate day and night in strong wind. Data show that more than 70% of gearbox failure are bearing failure. However, due to the high position of the bearing, this brings a lot of troubles for the maintenance. Every maintenance needs to be shut down, which increases the maintenance costs. Therefore, in order to save costs and improve the operating efficiency of wind turbines, the fault diagnosis of the bearings is particularly important.

Since the bearing is very important to the stable operation of the entire wind turbine system, more and more people are devoted to the research of bearing fault diagnosis and have achieved many results. In the past, fault diagnosis mainly used traditional feature engineering methods. Traditional feature engineering methods mainly use signal processing methods to extract features from input signals, and then use a classifier to classify the extracted features. Signal processing methods mainly include wavelet transform [1], frequency spectrum analysis [2] and so on. Classification models mainly include support vector machine [3], hidden Markov model [4] et al. Although traditional methods have achieved good classification results, it relies too much on experience in the process of feature extraction and has certain limitations. In recent years, with the development of deep learning, more and more fields have begun to use deep learning to solve various problems. Deep learning has been widely used in computer vision [5], document classification [6], and other fields. Because deep learning has powerful feature extraction and classification capabilities, bearing fault features can be extracted without relying on human experience in the process of fault diagnosis. So many researchers began to use deep learning to solve fault diagnosis problems. Zhang et al. [7] proposed a method combines an one-dimensional convolutional neural network (1DCNN), support vector machine (SVM) classifier. The extracted features are input into the SVM classifier, and particle swarm optimization (PSO) is used to optimize the SVM classifier. Chen et al. [8] proposed a method that Transferable Convolutional Neural Network (TCNN). Firstly, construction and pre-train an one-dimensional convolutional neural network based on a large source task datasets. Then a transfer learning strategy is adopted to train a deep model on target tasks by reusing the pre-trained network. This method not only utilizes the learning power of deep network but also leverages the prior knowledge from the source task. Yuan et al. [9] proposed an intelligent industrial process monitoring and fault diagnosis method based on the discrete wavelet transform and deep learning. First, this method uses the discrete wavelet transform to present the multiscale representation of the raw data. Second, using multiple convolution neural network extract the features at each scale, and then the extracted multiple features are fused by the

long short-term memory network to further reduce useless information and retain useful information. Yang et al. [10] proposed an efficient fault diagnosis method without additional denoising procedures. First, an one-dimensional deep residual shrinkage network, which directly takes the raw vibration signals contaminated by noise as input, is developed to realize end-to-end fault diagnosis. Then, to further enhance the noise immunity of the diagnosis model, the first layer of the model is set to a wide convolution layer to extract short time features. Besides, an adaptive batch normalization algorithm is introduced into the diagnosis model to enhance the adaptability to noise. Li et al. [11] proposed a novel wavelet-driven deep neural network termed as WaveletKernelNet (WKN), where a continuous wavelet convolutional (CWConv) layer is designed to replace the first convolutional layer of the standard CNN. This enables the first CWConv layer to discover more meaningful kernels. Lei et al. [12] proposed a novel fault diagnosis framework based on an end-to-end Long Short-term Memory (LSTM) model, to learn features directly from multivariate time-series data and capture long-term dependencies through recurrent behavior and gates mechanism of LSTM. Experiments verify that the proposed model has good robustness. Zhang et al. [13] proposed a method of converting raw signals into two-dimensional images. This method can extract the features of the converted two-dimensional images and eliminate the impact of expert’s experience on the feature extraction process. It follows by proposing an intelligent diagnosis algorithm based on Convolution Neural Network (CNN), which can automatically accomplish the process of feature extraction and fault diagnosis. Qian et al. [14] proposed a new deep transfer learning network based on convolutional auto-encoder (CAE-DTLN) to implement the mechanical fault diagnosis in target domain without labeled data. In this method, CAE is used as the feature extractor as it has the ability of noise removal. Both CORrelation Alignment (CORAL) loss and domain classification loss are integrated to enhance the effect of domain confusion. Although the use of deep learning in the fault diagnosis of wind turbine bearings has achieved high accuracy, there is still room for improvement.

A method of fault diagnosis is proposed in this article. Our main contributions in this research are as following:

1. A new ResNet model is constructed based on the residual module. This model is used to extract the characteristic of rolling bearing vibration signal and can prevent the gradient from disappearing or exploding during the training process.
2. The attention model is added to the ResNet model to increase the weight of important feature information and reduce the interference of unimportant features to improve the operating efficiency of the model.
3. LSTM is added to the network to extract the long-term dependence and ensure that the characteristic information on the time series will not be lost.

The rest of this article consists of the following. Section 2 introduces the fault diagnosis method based on ResNet, attention mechanism and LSTM. Section 3 introduces related experimental data and evaluation standards. Section 4 introduces the experimental results and analysis. In Sect. 5 summary and feature prospects are introduced.

2 Methodology

The structure of the model proposed in this paper is shown in Fig. 1. The input of the model is the vibration signal of the wind turbine bearing, and the output is the classification result. The model is mainly composed of three parts: the residual network, attention mechanism and LSTM. The residual network contains convolutional layer, pooling layer, etc., which are mainly used for feature extraction of input data. The attention mechanism is used to analyze the extracted features, increasing the weight of salient features and suppressing insignificant features. Through this method, the operating efficiency and classification accuracy of the model can be improved. LSTM is used to extract long-term dependent features on the time series and ensure that these features will not be lost as the training progresses. The three parts are described in detail below.

2.1 2DCNN

Convolutional neural network is a feedforward deep neural network that contains convolution operations. Because of its characterization learning ability, it can classify the input information according to the hierarchical structure, so it is widely used in the fields of classification and recognition. Convolutional neural networks generally include three layers: convolutional layer, activation layer and pooling layer. In recent years, due to the rapid development of image recognition, more and more researchers have begun to apply convolutional neural networks to this field, and have proposed many classic models based on convolutional neural networks. The classic neural network models proposed in recent years mainly include: GoogleNet [15], LeNet [16], VGG [17], ResNet [18], etc. Because most of these classic networks are proposed in processing pictures, basically two-dimensional convolutional neural

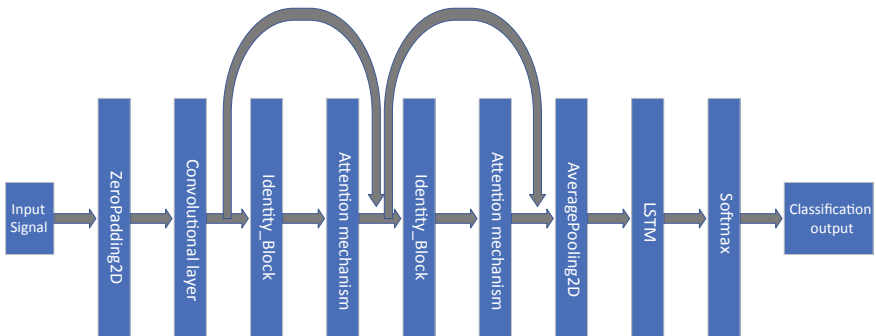


Fig. 1 Architecture of the proposed model

networks are used. However, the vibration signal of the wind turbine bearing is one-dimensional. To apply it to a two-dimensional convolutional neural network. Each sample needs to be deformed. The basic architecture of 2DCNN will be introduced in this section.

2.1.1 Convolutional Layer

The convolutional layer is the core of the neural network. Each convolutional layer is composed of several convolutional units. The parameters of each convolutional layer are optimized through the back propagation algorithm. In addition, another advantage of convolutional neural networks is that they can share parameter. It means that in the process of feature extraction, a model uses the same parameter among multiple parameters. Doing so can reduce the parameters in the convolutional neural network and improve the efficiency of model calculations. The purpose of the convolution operation is to extract different features of the input. The first layer of convolution mainly extracts some rough features. As the depth of the network increases, more detailed features can be extracted. In this article, the original wind turbine vibration signal is sampled to make a data set. The one-dimensional vector of each sample in the data set is converted into a two-dimensional matrix and input into the convolutional neural network to perform the convolution operation. The convolution process is shown as the Eq. (1).

$$s(t) = (x * w)(t) \quad (1)$$

In the convolution formula, the parameter x is the input data, the parameter w is the kernel function and $s(t)$ is the output.

2.1.2 Active Layer

The function of the activation layer is to map features to high-dimensional nonlinear intervals for interpretation and solve problems that cannot be solved by linear models. The commonly used functions mainly include: Sigmoid, Relu, TanH, Softmax. Because the Relu function converges fast and can prevent the gradient from disappearing, the Relu activation function is selected in this article. The formula is shown as Eq. (2).

$$f(x) = \max(0, x) \quad (2)$$

In this formula, $f(x)$ is the activation value.

In addition, the model uses the Softmax function as the activation function during classification. The Softmax function is also called the normalized exponential function. It can compress a k -dimensional vector z containing any real number into another k -dimensional vector $\sigma(z)$. After compression, the value of each element

is between zero and one and the sum of all elements is 1. The formula is shown as Eq. (3).

$$P(z_i) = \frac{e^{z_i}}{\sum_{c=1}^c e^{z_c}} \quad (3)$$

In this formula, z_i is the output value of the i -th node, and c is the number of output nodes which is the number of classification categories.

The Sigmoid function is used in the attention model. This function is used for the output of hidden layer neurons and can map real numbers to the (0, 1) interval. In the attention model, the Sigmoid function is placed after the fully connected layer to calculate the output, so that the output value obtains a normalized weight between 0 and 1. The formula is shown as Eq. (4).

$$s(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

where x is the output value of the fully connected layer, and $s(x)$ is the calculated weight.

2.1.3 Pooling Layer

The pooling layer is usually placed after the convolutional layer. The role of the pooling layer is to reduce the size of the model, compress feature dimensions and prevent overfitting. After the wind turbine bearing vibration signal is input to the convolutional layer, the characteristic signal will be extracted. Then the characteristic signal will be sent to the pooling layer to reduce redundant features. Pooling generally has two types: average pooling and maximum pooling. In this study, average pooling and maximum pooling are used in different locations on the network. The formula for average pooling and maximum pooling is shown as Eqs. (5) and (6).

$$z_j = \frac{\sum_{\substack{(j-1)r+1 \leq l \leq jr \\ (j-1)r+1 \leq h \leq jr}} y_{lh}}{r * r} \quad (5)$$

$$z_j = \max_{\substack{(j-1)r+1 \leq l \leq jr \\ (j-1)r+1 \leq h \leq jr}} \{y_{lh}\} \quad (6)$$

In this formula, z_j represents the output value after pooling operation; y_{lh} represents the feature value of a point on the extracted feature map; r represents the width of the pooling area.

2.2 Residual Network

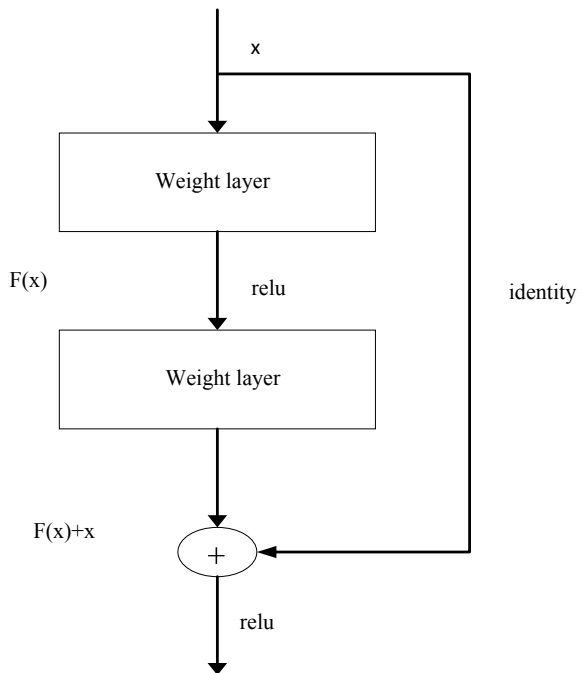
The essence of convolutional neural network is to fit a function that meets the corresponding target. Therefore, in order to achieve the goal, the researchers began to deepen the network architecture to infinitely fit the objective function, but the study found that simply stacking the network would not improve the network performance too much. This is because as the network deepens, the problem of gradient disappearance begins to occur, so the deep network is difficult to train. In order to solve the problem of the disappearance of the gradient of the deep network, the residual network was proposed. The framework of the residual module is shown in Fig. 2.

As can be seen from Fig. 2, compared with the traditional neural network architecture, the residual network has added an identity mapping layer. Doing so can prevent the gradient decay or disappear as the network depth increases. In the figure, $F(x)$ represents the residual, and $F(x) + x$ is the final mapping output, the final output of the network can be obtained as $H(x) = F(x) + x$. Therefore, the calculation process of the residual frame can be written as following three formulas.

$$H_1(x) = \text{relu}(w_1 * x) \tag{7}$$

$$H_2(x) = \text{relu}(w_2 * H_1(x)) \tag{8}$$

Fig. 2 Residual framework module



$$H(x) = H_2(x) + x \tag{9}$$

In the formula, x is the input data, w_1 and w_2 are the corresponding convolution operation, H_1 is the output of the first layer, H_2 is the output of the second layer.

A fault diagnosis model is built based on the residual network module in this study. Because the wind turbine bearing vibration signal data set is small, this network only uses two-layer residual modules. The purpose of using fewer network layers is to avoid redundant features extracted from too many network layers and interfere with the classification performance of the model. The experimental results also prove that too many layers will not improve the classification accuracy of small data set samples too much and even reduce its classification accuracy. Figure 1 show the model’s framework. In this model, the first layer of the convolution kernel with a size of 14×14 . Using a large convolution kernel can obtain a large feature receptive field and filter out some noise interference.

2.3 Attention Mechanism

Attention mechanism was originally used in the field of machine translation. The essence of the attention mechanism is similar to the human visual selective attention mechanism. The core goal is to select the most important information for the task goal from numerous information. There are two forms of realization of the attention mechanism, namely spatial attention and channel attention. In this research, the channel attention model is embedded in the residual module of the residual network to apply attention weight to different channels, thereby improving the classification ability of the model. The channel attention model is shown in Fig. 3.

It can be seen from the figure that the features extracted by convolution will be sent to the channel attention model. There are two fully connected layers of the model. In order to prevent the model from overfitting, dropout is added to the model. The reduce_mean layer in the figure is used to compress the input data. Shape $[-1]$ is the number of channels of the input feature. Reduction_ratio is a dimensional change parameter. Reduction_ratio is a dimensional change parameter, which can reduce the input data to a specified dimension. Then return

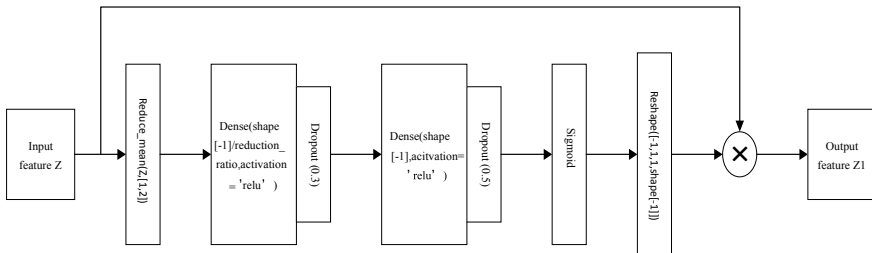


Fig. 3 Channel attention mechanism model

to the original dimension through a fully connected layer to obtain more nonlinear changes during transformation and reduce the amount of calculation. After two fully connected layers, the result is input into the sigmoid function to calculate the weight. Then multiply the weight and the input features to output the result. From the above figure, the calculation formula can be obtained as showing below.

$$\alpha = \delta(\sigma(\sigma(Z))) = [\alpha_1, \alpha_2, \dots, \alpha_L] \tag{10}$$

$$O = Z\alpha \tag{11}$$

In the formula, σ represents the Relu function; δ represents the Sigmoid; L represents the number of channels, α represents the weight; O is the result of multiplying the input and the weight.

2.4 LSTM

Simple RNN cannot remember information seen before many time steps and cannot learn long-term dependent information. Therefore, when extracting features of long-term sequence information, information may be lost, which will affect the accuracy of classification. In order to solve this problem, LSTM was proposed. In this study, because the fault signal of the bearing has a certain dependence on time, LSTM is added to the proposed model to better preserve the information on the time series. Figure 4 shows the LSTM structure.

As can be seen from Fig. 4, LSTM has three gates, namely the forget gate, the input gate and the output gate. $c_{(t)}$ and $h_{(t)}$ are the long-term state and the short-term state respectively. The forget gate determines how much information from the previous moment needs to be retained in the current state. The calculation process is

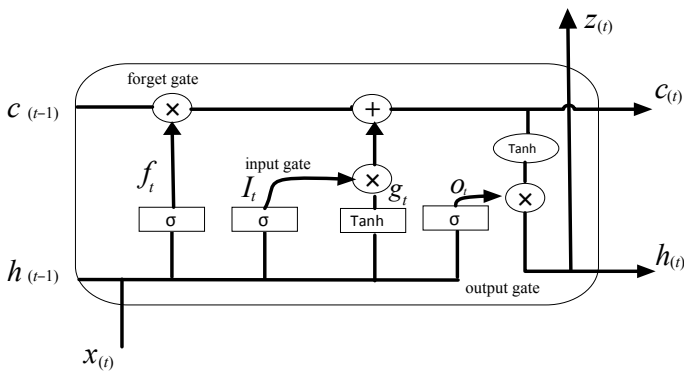


Fig. 4 The structure of LSTM

shown as Eq. (12).

$$f_t = \sigma(W_f[h_{(t-1)}, x_{(t)}] + b_f) \quad (12)$$

The input gate determines how much input the current moment needs to be saved in the unit. The calculation process is shown in Eqs. (13) and (14).

$$I_t = \sigma(W_I[h_{(t-1)}, x_{(t)}] + b_I) \quad (13)$$

$$g_t = \tanh(W_g[h_{(t-1)}, x_t] + b_g) \quad (14)$$

The output gate is used to determine how much of the current unit status is to be output to the output value. The calculation process is shown in Eqs. (15) and (16).

$$o_t = \sigma(W_O[h_{(t-1)}, x_t] + b_o) \quad (15)$$

$$h_t = o_t * \tanh(C_t) \quad (16)$$

where σ is the Sigmoid function; W is the weight; b is the offset value.

After the convolutional layer, the extracted features are sent to the LSTM network to further extract the dependency on the time series of the feature information. Finally, the extracted features are sent to the classifier for classification, which makes the wind turbine bearing fault diagnosis more accurate.

3 Experiment

The simulation platform of this experiment is configured as Intel core i5-1135g7 @2.40 Hz, 16 GB running memory, and the graphics card is NVIDIA mx450. The simulation model is implemented using the Python deep learning framework TensorFlow. The maximum number of iterations of the experiment is 30, and each batch of training samples is 128.

3.1 The Dataset

Because it is difficult to collect data on bearing faults of wind turbine gearboxes, this study uses the data set of the bearing data center of Case Western Reserve University (CWRU). CWRU's data acquisition platform is shown in the Fig. 5.

As shown in Fig. 5, the data acquisition platform consists of a 2 hp motor (left), a torque encoder (center), and a dynamometer (right). The test bearing supports the motor shaft. In this study, the data of the bearing drive end was selected as the

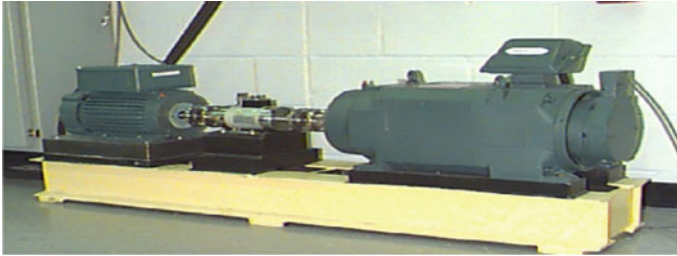


Fig. 5 Data acquisition platform

Table 1 Bearing type description

Fault labels	Fault location	Fault diameters (mil)
0	Inner race	7
1	Ball	7
2	Outer race	7
3	Inner race	14
4	Ball	14
5	Outer race	14
6	Inner race	21
7	Ball	21
8	Outer race	21
9	Normal	Normal

experimental dataset. The bearing data set mainly has fault data at three locations, namely outer ring fault, inner ring fault and rolling element fault. The fault diameters are 7 mils, 14 mils and 21 mils (1 mil = 0.001 inches) respectively. In this study, each fault type at the same speed is divided into a data set. In addition to a set of normal operating conditions, there are a total of ten types of bearing conditions. CWRU uses acceleration sensors to collect data. The acceleration sensor is placed on the drive end to collect vibration signals. The sampling frequency is 12 kHz and 48 kHz respectively (0, 1, 2, 3, 4, 5, 6, and 7 hp). In this study, bearing vibration signals at two sampling frequencies were selected for experimentation. Ten bearing types are shown in Table 1.

3.2 Experiment Description

3.2.1 Data Preprocessing

The bearing vibration signal collected by the sensor needs to be pre-processed to meet the requirements of model training and testing. In this study, 70% of the bearing vibration signals were divided into the training set; 20% were divided into the verification set and 10% of the data set were divided into the test set. Since there are ten bearing types, one thousand samples of each type are generated during the preprocessing. Each sample is generated by a sliding window with a step size of 28 and a sampling point of 1024. In addition, the model proposed in this article is a two-dimensional convolutional neural network, so in order to meet the input format, each sample needs to be changed from one-dimensional to two-dimensional.

3.2.2 Parameters of the Model

The fault diagnosis structure of the wind turbine bearing proposed in this paper is shown as Fig. 1. The parameter settings of each module in the model are shown in Table 2.

The input layer of the network is the pre-processed wind turbine bearing vibration signal and the output layer is the result classified by the Softmax classifier. There are two layers in Table 2 as `se_moudel` and this module is the attention module. Since the cross-entropy loss function is simple to derive, the loss is only related to the probability of the correct category and the convergence speed is faster. Therefore, the cross-entropy loss function is often used in multi-classification problems and the formula is shown in Eq. (17).

$$L = - \sum_{i=0}^9 y_i \log(p_i) \quad (17)$$

In the formula, y_i is the true label of the wind turbine bearing; i is the number of bearing types; p_i is the network output.

In convolutional neural networks, the choice of optimizer has significant impact on network performance. Because the Adam optimizer has the advantages of simple implementation, efficient calculation, and parameter updates are not affected by gradient scaling transformation. This research chooses Adam optimizer.

4 Result and Analysis

In neural networks, the settings of the learning rate have a significant influence on the performance of the model. Whether the learning rate is appropriate determines

Table 2 The layers of the network

Layer	Output shape	Parameters
Zero_padding2d	(None, 38, 38, 1)	0
Conv2d	(None, 25, 25, 64)	12,608
Batch_normalization	(None, 25, 25, 64)	256
Max_pooling2d	(None, 13, 13, 64)	0
Conv2d_1	(None, 13, 13, 64)	36,928
Batch_normalization_1	(None, 13, 13, 64)	256
Conv2d_2	(None, 13, 13, 64)	36,928
Batch_normalization_2	(None, 13, 13, 64)	256
SE_moudel	(None, 13, 13, 64)	0
Add	(None, 13, 13, 64)	0
Activation	(None, 13, 13, 64)	0
Conv2d_3	(None, 13, 13, 64)	36,928
Batch_normalization_3	(None, 13, 13, 64)	256
Conv2d_4	(None, 13, 13, 64)	36,928
Batch_normalization_4	(None, 13, 13, 64)	256
SE_moudel_1	(None, 13, 13, 64)	0
Add_1	(None, 13, 13, 64)	0
Activation_1	(None, 13, 13, 64)	0
Average_pooling2d	(None, 6, 6, 64)	0
Flatten	(None, 2304)	0
Tf_op_layer_Expandims	(None, 1, 2304)	0
Lstm	(None, 1, 128)	1,245,696
Flatten_1	(None, 128)	0
Dense	(None, 10)	1290

whether the neural network can converge to the global minimum. Therefore, in order to select the appropriate learning rate, this paper takes the learning rate at 0.01, 0.001, 0.0001 and 0.00001 and uses the same data set to do four sets of experiments. The training set and validation set loss function values are displayed in the figure below.

The network is trained for a total of 30 rounds. In Fig. 6, the value of the verification set loss function has been oscillating when the learning rate is 0.01. When the learning rate is 0.001, although the loss function value of the verification set has a tendency to converge. There are still several oscillations during the convergence process. When the learning rate is 0.0001, it can be seen that the value of the verification set loss function quickly converges to the minimum without large fluctuations. Although the loss function of the verification set is converging, the convergence speed is too slow, and it takes a long time to train to reach the optimal when the learning rate is 0.00001. Therefore, setting the learning rate of the wind turbine bearing fault diagnosis model to 0.0001 can make the bearing fault diagnosis more accurate and efficient.

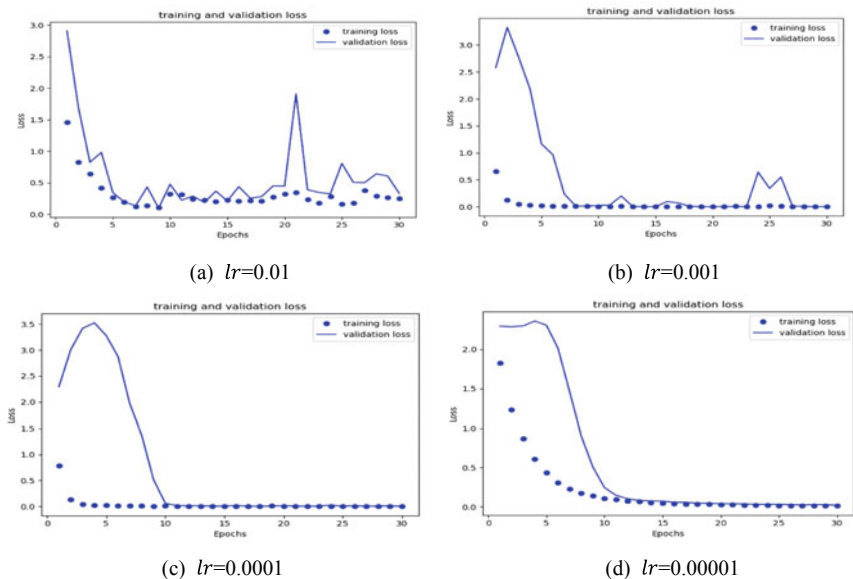


Fig. 6 Graph of loss function for different learning rates

The evaluation indicators are used in this experiment mainly include accuracy, precision, recall and f1 score. Accuracy is a common model evaluation index. Generally, the higher the accuracy, the better the performance of the model. Precision represents the proportion of samples classified as positive cases that are actually positive cases. The recall is used to measure how many of all positive cases are predicted as positive cases. The F1 score is used to measure the overall performance of the model. The calculation formulas of the four evaluation criteria are as following:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

$$Precision = \frac{TP}{TP + FP} \quad (19)$$

$$Recall = \frac{TP}{TP + FN} \quad (20)$$

$$F1score = 2 * \frac{P * R}{P + R} \quad (21)$$

In the formula, TP is true positive, which means that a sample is a positive class and is predicted to be a positive class. FP is false positive, which means that a sample is negative but is predicted to be positive. FN is the false negative, which means that

a sample is positive but is predicted to be negative. TN is true negative, which means that a sample is negative and is predicted to be negative.

At 12 kHz sampling frequency bearing vibration signal, the experiment compared the Se-Resnext [19] model, CNN-ISVM [20] model, Teod-Mcsdae model [21], and proposed model. The data set is 2 hp and the result of the four models is shown in Table 3.

It can be seen from Table 3 that the classification accuracy of the method proposed in this paper is better than Se-Resnext and CNN-ISVM on the 2 hp data set. Both the Teod-Mcsdae and the model proposed in this paper achieve 100% classification accuracy, but the Teod-Mcsdae model is complicated in calculation and puts forward higher requirements on the hardware configuration.

For bearing vibration signals at sampling frequency of 48 kHz, the experiment compared the 1DCNN-PSO-SVM model [5], CNN-LSTM model [22], and proposed model. The data set is 7 hp and the result of the three models is shown in Table 4.

In Table 4, three other evaluation criteria for model performance have been added. In terms of accuracy, the method proposed in this paper reaches 99.2%, which is higher than 98.2% of 1DCNN-PSO-SVM and 98.7% of CNN-LSTM. Therefore, from the perspective of accuracy evaluation indicators, the model proposed in this article is better than the other two models. Precision represents the probability that all samples that are predicted to be positive are actually positive samples. In the Table, the precision of the model proposed in this paper is 1.2% higher than that of 1DCNN-PSO-SVM, reaching 99.2%. Recall value represents the probability of being predicted as a positive sample in a sample that is actually positive. In this study, the proposed method is higher than other methods in recall value. For F1 score, it can measure the performance of the entire model. From the table, the F1 score of the method proposed in this article is higher than that of other methods. Among the four evaluation indicators, the performance of the model proposed in this paper is better than other models, indicating that the proposed model has strong fault diagnosis capabilities.

In order to prove the good classification effect of the proposed model, this paper compared the LSTM model, Resnet18 model, Attention Mechanism-ResNet18-LSTM model, and the proposed model. A vibration signal with a sampling frequency

Table 3 Comparison of different models (12 kHz)

Model	Se-Resnext	CNN-ISVM	Teod-Mcsdae	Proposed
Acc (%)	99.83	99.83	100.0	100.0

Table 4 Comparison of different models (48 kHz)

Model	Acc (%)	Precision (%)	Recall (%)	F1 score (%)
1DCNN-PSO-SVM	98.2	98.0	98.0	98.0
CNN-LSTM	98.7	–	–	–
The proposed method	99.2	99.2	99.2	99.2

Table 5 Comparison of different models (6 hp)

Model	Acc (%)	Precision (%)	Recall (%)	F1 score (%)
LSTM	78.9	79.4	78.9	78.7
ResNet18	94.1	94.3	94.1	94.0
Attention Mechanism-ResNet18-LSTM	95.2	95.7	95.2	95.1
The proposed method	99.2	99.2	99.2	99.2

of 48 kHz is used as the dataset. The number of neurons in LSTM is set to 128. The number of training samples in each batch of the experimental model is 128. The experimental results are shown in Table 5.

Table 5 shows that the model proposed in this paper is better than the other three models in fault classification performance. The four evaluation indicators of the LSTM model are all less than 80%, which are much lower than the other three models. The ResNet18 model is superior to LSTM in fault classification performance, and its four evaluation indicators Acc, Precision, Recall and F1 score all reach 94%. The Attention Mechanism-ResNet18-LSTM model adds an attention module and a long and short-term memory network on the basis of ResNet18, so that the model can extract more detailed and key features. Therefore, four evaluation indicators of this model are all higher than the ResNet18 model by more than 1%. The four indicators of the proposed model reached 99.2%, which is higher than LSTM and ResNet18. The indicators are also three to four percentage points higher than ResNet18 with attention module and LSTM. The reason is that the bearing data set used in the experiment is small, and the number of layers in the ResNet18 network is large. Too many layers will extract redundant features and affect model performance.

The confusion matrix of the proposed model is shown in Fig. 7. In Fig. 7, the horizontal axis represents the predicted label, and the vertical axis represents the true label. The numbers on the coordinate axis represent ten different types of bearings. Each type has 100 test samples, a total of 1000 test samples. The values on the main diagonal of the matrix represent the number of correctly classified samples of each type. The larger the value, the better the classification effect of the model.

5 Conclusion

Energy shortage and environmental protection have contributed to the development of new energy. In the field of new energy power generation, wind power occupies a very important position. As more and more wind farms are built and put into use, their maintenance becomes extremely important.

This research proposes a bearing fault diagnosis method based on ResNet-Attention Mechanism-LSTM. The one-dimensional bearing signal is pre-processed into a two-dimensional sample and input the ResNet. The attention model is embedded in the ResNet to increase the weight of important feature channels and

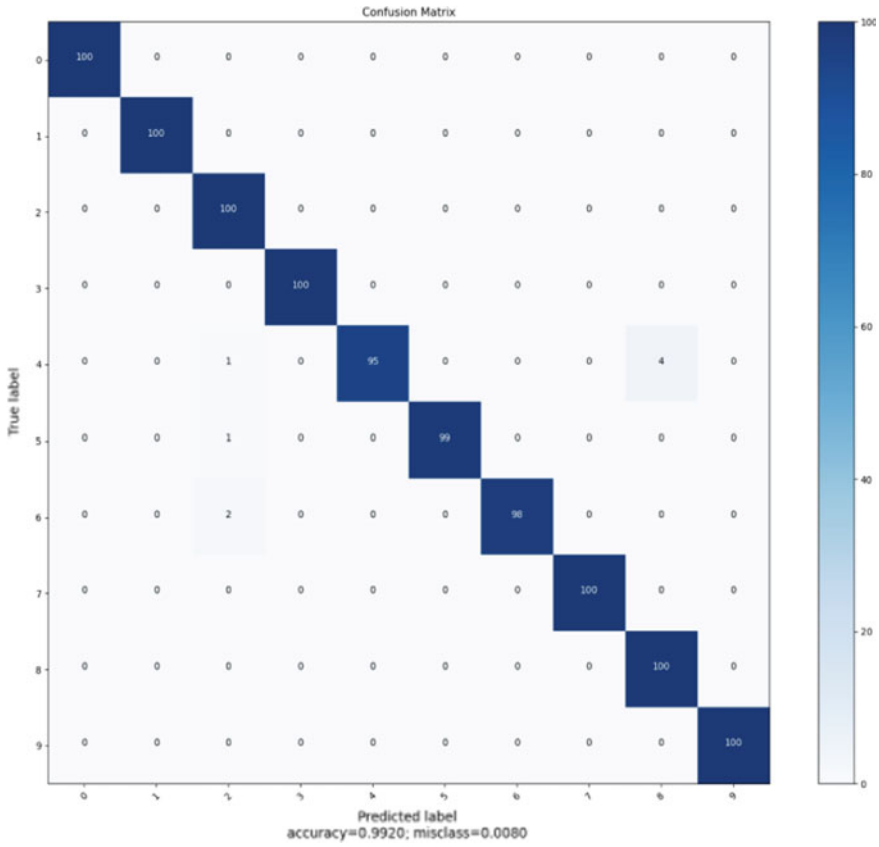


Fig. 7 Confusion matrix of the proposed method

reduce the interference of unimportant features. LSTM is added to the network to extract the long-term dependence on the signal, making the classification feature richer.

This method is used for bearing fault diagnosis and compared with other methods. The results show that the method proposed in this paper outperforms other methods in bearing fault diagnosis and avoid the problem of relying too much on manual experience in traditional methods. Since the wind turbine vibration signal will inevitably be mixed with noise during the acquisition process, the next step will be focused on the removal of noise in the signal.

Acknowledgements This work was supported by the National Nature Science Foundation of China (NO:51965052).

References

1. Deng, F.: A fault diagnosis method of rolling element bearings with adaptive frequency window empirical wavelet transform. *J. Xi' An Jiao Tong Univ* **52**(8), 22–29
2. Tao, L.: Rolling element bearing fault diagnosis based on frequency band entropy. *J. Vibrat. Shock* **33**(1), 77–80 (2014)
3. Demidova, L.A.: Two-stage hybrid data classifiers based on SVM and KNN algorithms. *Symmetry* **13**(4), 615 (2021)
4. Hao, F.: Fault diagnosis of rolling bearing based on continuous hidden markov model. *Chinese J. Construct. Mach.* **17**(2), 184–188 (2019)
5. Liu, Z.: Automatic detection of transformer components in inspection images based on improved faster R-CNN. *Energies* **11**(12), 3496 (2018)
6. Huang, Y.: Hierarchical multi-attention networks for document classification. *Int. J. Mach. Learn Cyber*, 1–9 (2021). <https://doi.org/10.1007/s13042-020-01260-x>
7. Zhang, X.: Research on bearing fault diagnosis of wind turbine gearbox based on 1DCNN-PSO-SVM. *IEEE Access* **8**, 192248–192258 (2020)
8. Chen, Z.: Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network. *IEEE* **16**(1), 339–349 (2019)
9. Yuan, J.: A multiscale feature learning scheme based on deep learning for industrial process monitoring and fault diagnosis. *IEEE Access* **7**, 151189–151202 (2019)
10. Yang, J.: Fault diagnosis of rotating machinery based on one-dimensional deep residual shrinkage network with a wide convolution layer. *Hindawi* **2020**, 1–12 (2020)
11. Li, T.: WaveletKernelNet: an interpretable deep neural network for industrial intelligent diagnosis. *IEEE Trans.* 1–11 (2021). <https://doi.org/10.1109/TSMC.2020.3048950>
12. Lei, J.: Fault diagnosis of wind turbine based on long short-term memory networks. *Renew. Energy* **133**(2019), 422–432 (2018)
13. Zhang, J.: A new bearing fault diagnosis method based on modified convolutional neural networks. *Chinese J. Aeronaut.* **33**(2), 439–447 (2019)
14. Qian, Q.: A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis. *Measurement* **178**(2021), 1–9 (2021)
15. Szegedy, C., Liu, W.: Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9 (2015). <https://doi.org/10.1109/CVPR.2015.7298594>
16. LeCun, Y.: Gradient-based learning applied to document recognition. *IEEE* **86**(11), 2278–2324 (1998)
17. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference of Learning Representation (ICLR), 1–14 (2015). [arXiv: 1409.1556](https://arxiv.org/abs/1409.1556)
18. He, K., Zhang, X.: Deep residual learning for image recognition. In: 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
19. Hu, X.: Fault diagnosis method of rolling bearing based on SE-ResNeXt. *Comput. Measurement Control*, 1–8 (2021)
20. Kang, S.: Fault diagnosis method of rolling bearing under varying loads based on deep online transfer. *Cont. Dec.* 1–10 (2021)
21. Pei, X.: Intelligent bearing fault diagnosis based on teager energy operator demodulation and multiscale compressed sensing deep autoencoder. *Measurement* **179**(2021), 1–15 (2021)
22. Shen, T.: CNN-LSTM method with batch normalization for rolling bearing fault diagnosis. *Comput. Integrated Manuf. Syst.* 1–16 (2021)

Study on Optimization and Improvement of Production Line of H Product



Guo Jidong, Liang Yuyan, Ma Zenan, Qiu Zijian, Mo Yuwei, Li Limin, and Zhou Dawei

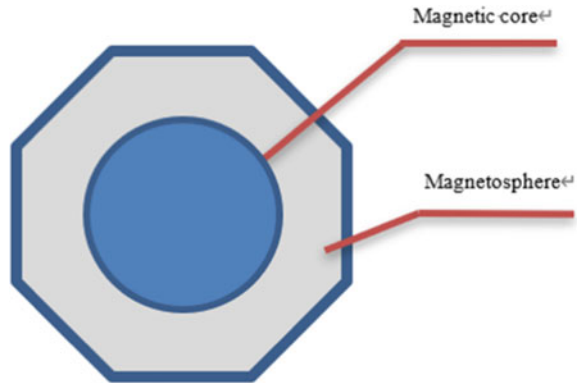
Abstract This case makes full use of the knowledge of IE in various aspects, such as work research, ergonomics, and other methods to put forward four different improvement proposals, including two work table design, one fixture design and one operator man–machine operation design, solving the problem of one company H product has a large inventory of WIP and so on. After the proposal was put forward, the Flexsim simulation technology and other methods were used to evaluate the scheme which is proposed in this case and verify the feasibility of the scheme. The final improvement proposal resulted in a total reduction about 2709 s in appearance inspection engineering cycles, at the same time the new improved workstation made it easier for employees to work, to reduce physical injury to employees due to long working hours; Four employees were reduced in the electrode magnetic coil project, and the annual wage expenditure was saved by about 198,144 yuan.

Keywords Flexsim simulation technology · Work research · Ergonomics · Fixture design

1 Introduction

With the further development of social production, the demand for balance of production line in modern processing and manufacturing industry is gradually improved. This paper focuses on the inventory backlog of WIP caused by poor management and poor spatial layout in the production process of H product of the company, which will lead to the decline of productivity of the enterprise. Aimed at the problems of the company, respectively, suit the remedy to the case: firstly, work measurement and motion analysis was carried out on the appearance inspection engineering, and industrial engineering and other theory methods for process optimization; secondly,

G. Jidong (✉) · L. Yuyan · M. Zenan · Q. Zijian · M. Yuwei · L. Limin · Z. Dawei
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, Guangdong, China
e-mail: 35189850@qq.com

Fig. 1 H product diagram

design fixture, optimize and improve the bottleneck process by using image analysis method, improve productivity of this process and reduce WIP inventory.

2 Production Status and Problem Analysis

2.1 The Introduction of H Product

A company mainly produces electronic products. H product is one of the electronic products of the company, made of magnetic core and magnetic coil, and its volume is small, as shown in Fig. 1. There are 4 seed products under H product, and the process flow of the sub products is the same. Therefore, the optimization and improvement of 1 seed product in this case is applicable to other sub products.

The general process flow chart of H product is shown in Fig. 2. Plastic product box is used for transportation and appearance inspection engineering.

2.2 Production Status of H Product

Through field observation, there is a large amount of WIP inventory between the product changing plate and the appearance inspection engineering. Table 1 shows the ratio of personnel per unit production line and WIP inventory at a certain time.

To sum up, the appearance inspection process is the bottleneck process of H product line.

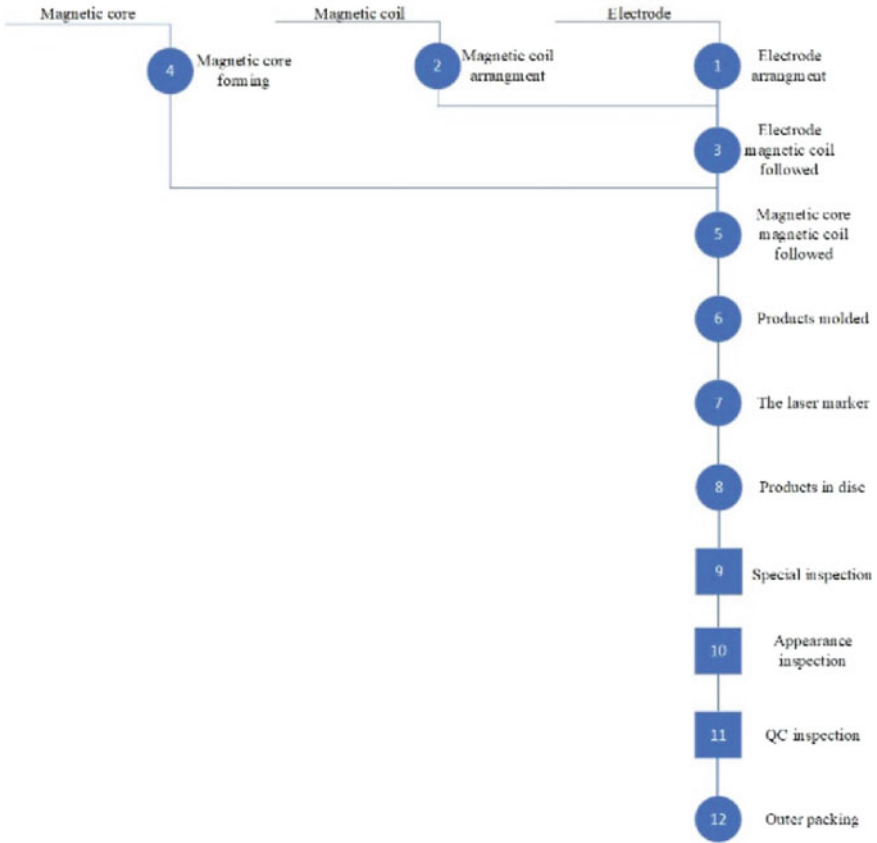


Fig. 2 General process flow chart

2.3 The Problem Analysis

Based on the bottleneck process and the problem of long idle time of electrode and magnetic coil employees, this case analyzes the balance ratio and operation ratio.

2.3.1 The Analysis of Capacity Balance Rate of Each Process in Production Line

The stopwatch analysis method was used to conduct the operation measurement, calculating the standard time, and summarizing the product process diagram of H, as shown in Table 2.

The formula for calculating the balance rate of the production line [1] is as follows:

Table 1 Production status record

No.	Engineering	Personnel	Work in process	Headcount	Note
1	Electrode arrangement	1	4 disc-720 pcs	8	Statistical time: 10/6 15:30
2	Magnetic coil arrangement			8	
3	Electrode magnetic coil followed			8	
4	Magnetic core forming	1	4 disc-720 pcs	8	
5	Magnetic core magnetic coil followed				
6	Products molded	1	0	8	
7	The laser marker		0		
8	Products in disc	1	0	8	
9	Special inspection	1	0	8	
10	Appearance inspection	6	25 group-5000 pcs	48	

$$\eta = \text{Total time of each process} / (\text{CT} * \text{Number of operations} * 100\% \quad (1)$$

CT is the standard time of bottleneck process.

According to the formula (1), the balance rate of H product production line is 34%, which is in line with the reality.

2.3.2 Products in Disc Process—Status Analysis

Through interviews and video analysis, it was found that the operators’ movements such as straightening their arms and detouring path did not conform to the action economy principle [2].

2.3.3 Analysis of the Present Situation of Appearance Inspection Working Table

Through interview and field investigation, we found that the following problems existed in the appearance inspection process.

Table 2 H product flow chart

Job name:			Statistics					
Handler:			Item			Number	Times/s	
Investigator:			Operation	○	8	2401.5		
Start:			Inspection	□	2	1330.2		
End:			Transport	⇒	0			
			Temporary storage or delay	D	0			
			Storage	▽	0			
NO.	Job description	time/s	Process series					Unit
			Operation	Inspection	Transport	Temporary storage or delay	Storage	
1	Electrode arrangement	360	●	□	⇒	D	▽	192pcs
2	Magnetic coil arrangement	300	●	□	⇒	D	▽	180pcs
3	Electrode magnetic coil followed	240	●	□	⇒	D	▽	180pcs
4	Magnetic core forming	240	●	□	⇒	D	▽	180pcs
5	Magnetic core magnetic coil followed	300	●	□	⇒	D	▽	180pcs
6	Products molded	225	●	□	⇒	D	▽	180pcs
7	The laser marker	270	●	□	⇒	D	▽	180pcs
8	Products in disc	466.5	●	□	⇒	D	▽	180pcs
9	Special inspection	227.7	○	■	⇒	D	▽	180pcs
10	Appearance inspection	1102.5	○	■	⇒	D	▽	180pcs

Problem 1: there are some movements in the work, such as feet can't touch the ground and pedal high to pick up pieces, etc.

Problem 2: in the appearance inspection process, four people share a work table, and there is no gap or boundary between the work stations, so there exists a risk of goods mixing.

Problem 3: when employees share a set of tools, there exist problems that have a long distance to get the tools and wait in line.

2.3.4 The Analysis of the Current Situation of Fixture

The length, width and height of long suction bars are 195 mm, 36 mm and 9 mm respectively, while the length, width and height of short suction bars are 85 mm, 30 mm and 6 mm respectively. Through the operation observation, it is found that taking and putting down the suction bars for many times during the appearance inspection process, resulting in repeated actions and increasing the workload of employees. Therefore, it is necessary to improve the current situation of the suction bars.

Using the mod timing method [3] to record the inspection time of 60 H products, it was found that in the current method, the mod value of long suction bar was 1436, and that of short suction bar was 1322. There are many simple repetitive actions in the process of operation, and there is room for improvement.

2.3.5 Man-machine Operation Status Analysis

Field observation found that there was a long spare time for the staff of the Electrode magnetic coil followed project, so the project was continuously observed and analyzed.

According to the observation data, the operation analysis of the Electrode magnetic coil followed engineering operator is carried out:

- ① The waiting time of employees in the process of operation accounted for 57.9% of the total time.
- ② Excluding the time of non-cycle work content, the work rate of employees is 24.9%.

In order to eliminate the situation of other people's assistance in operation, the operation data of other people in this station is recorded at the same time, and it is concluded that the number of other people's assistance in operation is 0, that is, there is no machine interference.

To sum up, the waiting time is too long and the operation rate is low, so the design of human-machine operation can be optimized.

2.4 Production Line Analysis Based on Flexsim

Combined with the production status of H product, summarize the data and conduct Flexsim status analysis and simulation [4]. In order to facilitate the analysis, the Flexsim model is set up with multiple simulation premises, and the current production line model is established (Fig. 3).

In the simulation results, the idle time column of special inspection temporary storage area is 0%, the empty is 1.02%, the input of special inspection temporary

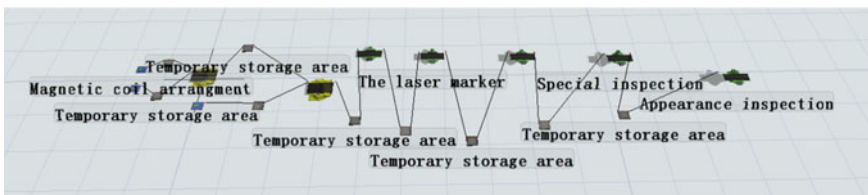


Fig. 3 Production line status model

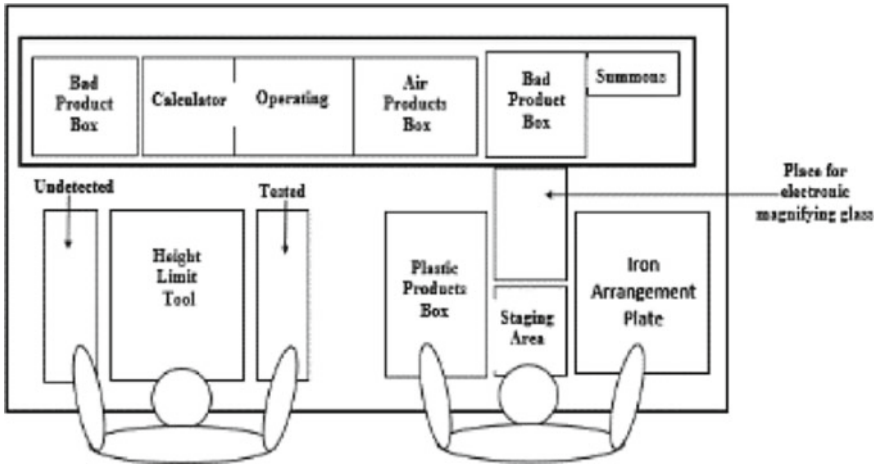


Fig. 4 Improve the front work table

storage area is 460, and the output is 195, so the WIP inventory between special inspection and appearance inspection process is serious, which is consistent with the actual situation.

3 H Product Optimization and Improvement Plan

3.1 Improvement Scheme of the Appearance Inspection Engineering

3.1.1 Products in Disc-Work Desktop Improvement

According to the principle of action economy, the layout of the worktable was improved, before and after improvement are shown in Figs. 4 and 5.

3.1.2 Design Scheme of Appearance Inspection Table

In this case, according to the working space design principles proposed in Ergonomics Principles of Working System Design [5] and the human body size provided in Chinese Adult Body Size [6], the appearance inspection table was redesigned.

Design the height of the table top

The design of table top height needs to design the height of the seat first, and then determine the height of the table top.

- ① The design of the seat height

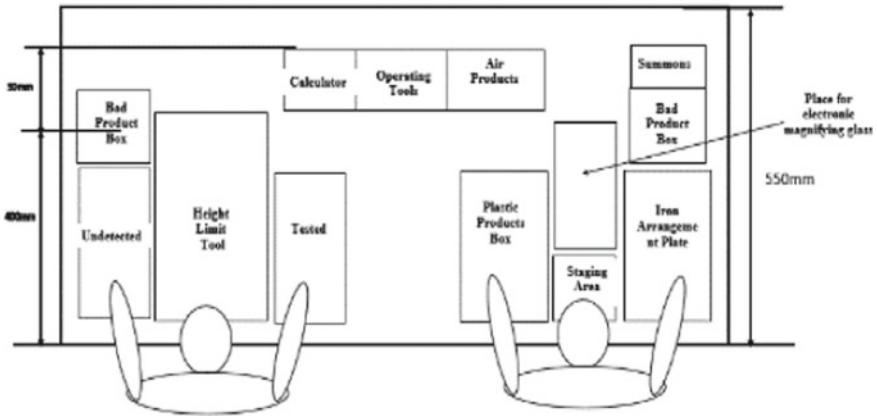


Fig. 5 Improved worktable

Since there is no qualitative regulation on gender and age for the job requirement of this process, the design size of the seat height adopts the data of medium male and female figure. $\text{Seat height} = \text{“calf plus foot height”} + \text{correction number of shoes worn} + \text{correction amount of pants worn—thickness of hip}$ [7].

Calculated the comprehensive seat height = $(422 \text{ mm} + 386 \text{ mm}) / 2 = 404 \text{ mm}$.

② The design of the desktop height

$\text{Desktop height} = \text{comprehensive seat height} + \text{table and chair height difference} = \text{comprehensive seat height} + \text{sitting height} / 3$.

Calculated table height = $697.8 \text{ mm} \approx 700 \text{ mm}$, as shown in Fig. 6(unit mm).

Fig. 6 Schematic diagram of table height size



Design the layout of the working table.

Based on the fact that the platform is so high that the staff need to press the desk with their arms to access the work, this case designs the height of the platform which used to place semi-finished products.

In order to make it easier for more people to work, the height of the platform will be the same as the height of the employees' shoulders when they are sitting, and will be designed in the 5th percentile of the adult female body size.

According to the relevant formula, the relative height of the platform was calculated to be 220 mm, as shown in Fig. 6 (unit mm).

The layout of the table surface design

Optimization of production site management is the premise and basis for enterprise management to achieve overall optimization [8]. Therefore, this case will carry out plane layout design for the worktable according to 5S principle and the action economy principle, as shown in Fig. 7. A 10 cm high partition is added to eliminate the risk of mixed goods. Its layout is shown in Fig. 8. Defect product boxes are designed to make it more convenient for employees to take out H products, shown in Fig. 9.

General design drawing of appearance inspection work table

The three-dimensional diagram of the current working table of appearance inspection process is shown in Fig. 10, and the redesigned working table of appearance inspection working process is shown in Fig. 11.

Fig. 7 Tool area division diagram

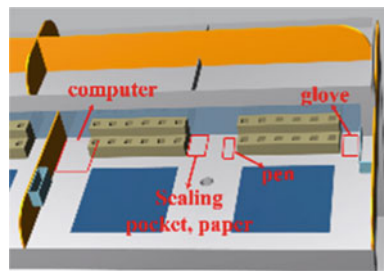


Fig. 8 Baffle layout

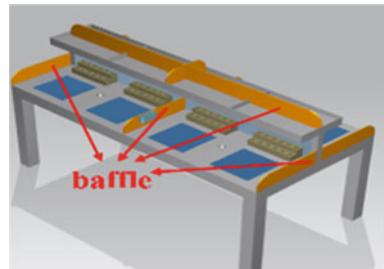


Fig. 9 Rejects box

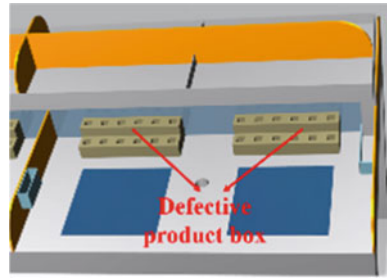


Fig. 10 The current work table

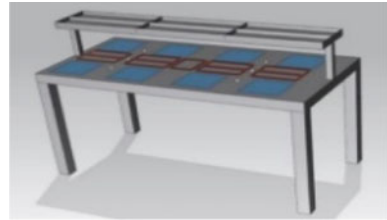
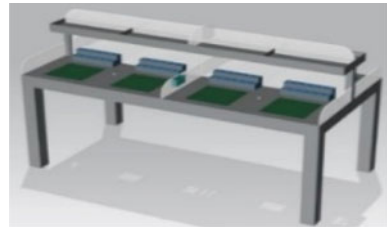


Fig. 11 The redesigned work table



3.1.3 Design Fixture for Appearance Inspection Process

Design concept

According to Sect. 2.3, the current suction bar operation does not conform to the principle of action economy, so the ECRS principle is used to redesign the long suction bar and short suction bar, as shown in Figs. 12 and 13.

Fig. 12 Long suction bar

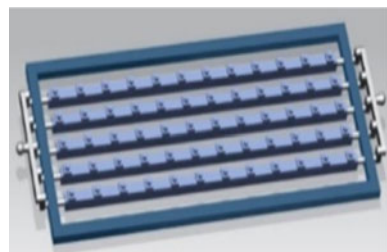
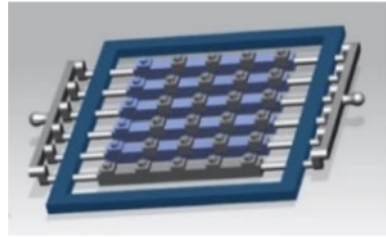


Fig. 13 Short suction bar



The newly designed long suction strip tray and short suction strip tray can respectively suck up one and half of the H product at a time for bottom inspection, reducing the action of multiple fetching and releasing.

3.2 Man–Machine Operation Improvement

According to the man–machine operation analysis in Sect. 2.3.5, this case analyzed and measured the leisure capacity of employees in electrode magnetic coil engineering, designed a man–machine operation table with one person holding two machines, and proved its feasibility at the same time.

3.2.1 Analysis of Leisure Sapacity

According to equipment operation observation table and staff operation observation table in Sect. 2.3, leisure capacity analysis is carried out:

$$N = \frac{M + t}{t} \tag{2}$$

According to Eq. (2), the leisure capacity of the operator can be obtained as follows:

$$N = \frac{(M + t)}{t} = \frac{373 + 204}{204} = 2.828 > 2$$

Therefore, the electrode magnetic coil followed engineering staff can be a person with two assignments.

Table 3 Man–machine operation schedule

The work object	The job content	Time/s	Beat/min
Operators	Exchange the electrode plate	8	2
	Exchange magnetic coil plate	8	4
	Semi-finished plate to hardening machine	10	4
	Take empty semi-finished products to the next machine	6	4
	The semi-finished tray trolley is moved to the cooler	28	24
Electrode magnetic ring to the machine	The electrode plate operate	2 min	
	Magnetic coil plate operate	4 min	
	Intermediate plate output	4 min	

3.2.2 Man–Machine Operation Measurement

The electrode magnetic coil followed engineering was followed by the stopwatch method for engineering equipment, to obtain the time of each work content, as shown in Table 3.

3.2.3 Man–Machine Operation Cycle Design

According to the operator's operation and equipment time parameter table, combined with the man–machine analysis diagram, the operation improvement of one person holding two machines can be realized. This method is also applicable to other production lines of H product.

4 Conclusion

Through the improvement of the layout of the work table, the image analysis method is used to record the standard working time of the staff before and after the improvement of the work table, which changes from 31.1 s to 20.6 s, and the work efficiency is increased by 33.8%. The additional improvement of the partition on the work table improves the working efficiency of the staff, reduces the risk of product mixing, and provides more working areas for the staff.

In the appearance inspection process of the bottleneck process, the fixture was improved. The MOD value of the process was reduced from 1436 to 1314, and the time to complete the process was shortened by 15.738 s after the improvement.

Through man–machine operation analysis, the change from one person holding one device to two, the work efficiency is increased by 100%, the personnel ratio of

Table 4 Overall operation cycle design

<i>Cycle operation</i>							
Operation content	Frequency				Unit time/s		Total time/s
	One cycle	5 cycles	Remain 90s	Total cycle			
Spare electrodes, magnetic coils and semi-finished plates	2	0	0	2	11	5.5	22
Exchange the electrode plate	34	170	0	170	8	4	1360
Exchange magnetic ring plate	16	80	0	80	8	4	640
Semi-finished disc moved to hardening machine	16	80	0	80	10	5	800
Take empty semi-finished products to the next Machine	16	80	0	80	6	3	480
Ove the semi-finished plate to the cooler	2	10	0	10	28	14	280
Total	86	420	0	422	71	35.5	3822
<i>Non-cyclic operation</i>							
Operation content	Ratio (%)			Processing time/s			
Contact fault corresponding	4.5			503.55			
Harden fault corresponding	0.2			22.38			
Write summon	3.0			335.7			
Collect defective products	1.0			111.9			
Cooler	2.1			234.99			
Total	10.8			2417.04			
<i>Data aggregation</i>							
Total working time of operator/s				5999.04			
Total design cycle time/s				11190			
Ratio of operator operating 2 equipments/s				53.61%			

the corresponding process can be reduced by 50%, and the annual wage expenditure is reduced by 198,144 yuan.

The balance rate of the production line was increased from 34 to 38%, an increase of 4%, and the work efficiency was improved by 33.8%. The time consumed by the bottleneck process was reduced, the ratio of personnel was reduced, and the labor cost was reduced. The improvement effect was obvious.

References

1. Shuping, Y., Fu, G.: *Fundamental Industrial Engineering*, 2nd edn, China Machine Press, Beijing (2018)
2. Wei, S.: Research on the application of industrial engineering in the optimization of production line. *Tech. Innov.* **9**, 144–145 (2020)
3. Fu, G.: Application of the MOD method in production line capability balancing. *J. Northeastern Univ.* **5**(1), 31–33 (2003)
4. Jiangbo, W., Kun, Y., Jieming, R.: Research on balance improvement of turbine bearing production line based on IE and Flexsim. *Mech. Eng. Autom.* **220**(03), 38–40 (2020)
5. GB/T16251–2008/ISO:6385:2004. *Ergonomic Principles of Work System Design*. Standards Press of China, Beijing (2008)
6. GB/T10000–1988. *Chinese Adult Body Size*. Standards Press of China, Beijing (1989)
7. Ruan baoxiang: *Industrial Design Ergonomics*. 3rd edn. China Machine Press, Beijing, (2016)
8. Xiulan, R.: On optimization strategy of enterprise production site management. *J. Baotou Vocat. Tech. College* **21**(3), 48–50 (2020)

Towards Data Driven Dynamical System Discovery for Condition Monitoring a Reciprocating Compressor Example



Ann Smith and W. T. Lee

Abstract A viable data driven approach for determining dynamical systems describing engineering processes would be a valuable tool in condition monitoring. The application of the SINDy algorithm for dynamical system discovery is investigated in the context of a reciprocating compressor. A feasibility study was carried out in which an attempt was made to recover a model of the compressor from synthetic data obtained from that model. A simplified model of the compressor with two degrees of freedom was developed from an existing model. Following the SINDy approach a parsimonious model was constructed from a large library of functions using sparse regression. This model has the same structure as and similar coefficients to the original model thus demonstrating the potential of this approach.

Keywords SINDy · Dynamical systems · Condition monitoring · Reciprocating compressors · Digital twin

1 Introduction

Condition Monitoring (CM) is concerned with preventing, or at the very least predicting, impending component failure. Quality management through continuous monitoring of process outputs aims to detect and identify deviations from normal operation at onset thus ensuring optimal performance, safety, and productivity [1]. A key problem in condition monitoring is having a quantitative understanding of the system being monitored. One solution is found in data reduction methods such

A. Smith (✉)

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK
e-mail: a.smith@hud.ac.uk

A. Smith · W. T. Lee

Centre for Mathematics and Data Science, University of Huddersfield, Huddersfield HD1 3DH, UK

W. T. Lee

MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

199

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_17

as principal component analysis (PCA). However, the gold standard for this would be a predictive model of the system implemented as a dynamical system. However, for many condition monitoring problems deriving such a model from fundamental electrical or mechanical principles is difficult or impossible.

As machine complexity continues to increase with respect to both individual components and complete systems, so does the intricacy and cost of predictive maintenance programmes. Sensors too have become smaller and more affordable, thus the ability to gather, process and communicate information is growing. Large-scale industrial processes are monitored by a network of 1000's of sensors each capturing data reflecting the process condition. Whilst fault signature analysis of sensor fault relationships can assist in identifying optimal sensor placement for focus, detailed holistic analysis of the process utilising complex prognostic models is often unfeasible.

A recent advance in this area is the development of an alternative data driven route to the determination of dynamical systems [2]. Preliminary results towards the application of these principles to engineering plant are reported here. An instrumented reciprocating compressor for which both experimental data and a dynamical system model are available [3] is taken as an example system. As an initial check of the feasibility of this approach recovering a simplified model of the compressor from synthetic experimental data is investigated.

This paper presents original research in the field of dynamical system discovery. Initial investigations are presented which demonstrate an implementation of the Sparse Identification of Nonlinear Dynamics (SINDy) methodology [2, 4]. The SINDy algorithm is utilised to determine the underlying dynamical system of a component or process. A subset of the dynamical system representing a reciprocating compressor model was solved using a Runge-Kutta method with adaptive time stepping (MATLAB's ode45). The resulting solution was incorporated into the SINDy algorithm and the original system recovered. A simplified system of differential equations representing the cylinder action of a reciprocating compressor formed the base model under investigation. Subsequently, noise terms were incrementally added to mimic a typical search for an unknown dynamical system.

2 Theoretical Background

The key challenge in discovering dynamical systems from experimental data is the large space of possible models that must be systematically investigated. To overcome this problem Kutz et al. developed the Sparse identification of nonlinear dynamics (SINDy) methodology [2]. This relies on the use of a library of test functions allowing linear regression to be used despite the nonlinear nature of the equations—essentially all nonlinearity is contained in the test functions. As standard linear regression will generate nonzero coefficients in front of every test functions due to experimental error a sparse regression procedure must be used to generate a parsimonious model avoiding the overfitting problem.

A very encouraging application of the SINDy algorithm in our context is an application to vortex shedding by a cylinder at moderate Reynolds numbers [4]. In this feasibility study a single (simulated) experimental measurement was used to determine a human interpretable dynamical system describing the dynamics of vortex shedding. Furthermore, with the introduction of further data the full fluid dynamical state of the system can be recovered. The long-term aim of this research is to see if a similar process is feasible in engineering plant such as the reciprocating compressor.

3 Methodology

3.1 *Mathematical Models*

Due to their prevalence and importance in industrial processes there is naturally much interest in the detection and diagnosis of Reciprocating Compressor (RC) faults. RCs are critical components in many high-pressure processes and their failure can have a serious impact on both the process operation and its equipment.

A highly simplified model based on a reciprocating compressor focussing on the mechanics was developed based on the model developed by Elhaj et al. [3]. In addition to the assumptions introduced by the original modelling process this model makes the following assumptions.

- Constant drive torque
- Only a single cylinder is considered.
- The pressure in the cylinder is taken as constant.

The result is a two-dimensional set of ordinary differential equations with four free parameters. Random values of these parameters of order unity were chosen.

Initial conditions were taken to be $\theta = 0$, $\omega = 0$, i.e. simulating a cold start. Note that simulating the steady state operation of the device might have led to some terms not appearing in the model determined by fitting as these may only be important in the dynamics as the device approaches steady state.

3.2 *Synthetic Data Set*

To generate a synthetic dataset of measurements of theta and omega the differential equations were solved numerically and then noise was added. A Runge-Kutta numerical integrator was used to solve these equations. To turn the results of the simulation into synthetic data artificial noise was added in the form of uniformly distributed pseudorandom values with zero mean on the interval $[-0.5, 0.5]$.

The SINDy algorithm also requires $\frac{d\omega}{dt}$ data. Following Champion et al. [2] this was obtained by substituting the synthetic data for ω and θ into the dynamical system model.

3.3 RC Models

A reciprocating compressor was chosen as the focus from this study as in addition to its importance in industrial settings a mathematical model of a Broom Wade TS9 RC rig has been developed from first principles [3]. The TS9 is a two-stage RC with compression cylinders arranged in V-shape formation. The rig has a maximum working pressure of 1.379 MPa (13.8 bar) and a crank speed of 440 rpm. A simplified version of this model was developed for this preliminary investigation.

The equation of motion of a crankshaft in a simple model of a reciprocating compressor with a single cylinder is given by

$$J \frac{d\omega}{dt} = \frac{P_c r^2 s_c}{L} \cos \theta \sin \theta + P_c r s_c \sin \theta - \left(m_p r^2 + \frac{m_{cr} r^2}{2} \right) \omega^2 \cos \theta \sin \theta + T_m \quad (1)$$

$$\frac{d\theta}{dt} = \omega \quad (2)$$

where J is the moment of inertia of the power unit, ω the angular velocity of the crankshaft, P_c the cylinder pressure, r the radius of the crank, s_c the cross-sectional area of the cylinder, L is the length of the connecting rod, m_p is the piston mass, m_{cr} mass of the connecting rod and T_m is the driving torque from the electric motor respectively.

The SINDy algorithm promotes sparsity by eliminating terms with small coefficients. In order for the size of coefficients to be meaningful the equations of motion must be written in dimensionless form. In dimensionless form the equations of motion of the crankshaft is given by

$$\frac{d\omega}{dt} = -D \cos \theta \sin \theta \omega^2 + C \cos \theta \sin \theta + B \sin \theta + A \quad (3)$$

$$\frac{d\theta}{dt} = \omega \quad (4)$$

with coefficients arbitrarily chosen as $A = 0.5$, $B = 0.75$, $C = 0.75$, $D = 0.25$. For simplicity the same symbols are used for dimensional and dimensionless variables.

3.4 SINDy

In order to recover the dynamical system from the synthetic data the SINDy algorithm was used [2]. From $\theta(t)$ and $\omega(t)$ data a library of functions of the form $\cos(m\theta)\omega^n$ and $\sin(m\theta)\omega^n$ was generated. The $d\omega/dt$ data was regressed against all members of this library of functions. To create a parsimonious model, sparsity parameter λ was introduced and all terms from the function library with a coefficient less than λ were removed. In this case λ was chosen by hand as the aim of the study was to determine whether it was possible in principle to recover the original model by this process. (More generally λ would be determined by a formal sparse regression technique.) The regression was then carried out against the reduced function library.

4 Results

The synthetic dataset is shown in Figs. 1 and 2. By visual inspection of the results of the regression it was determined that values of λ in the range $0.0046 < \lambda < 0.2451$ would eliminate terms that were not in the original model. By repeating the regression with these terms eliminated a model of the same form as the original with

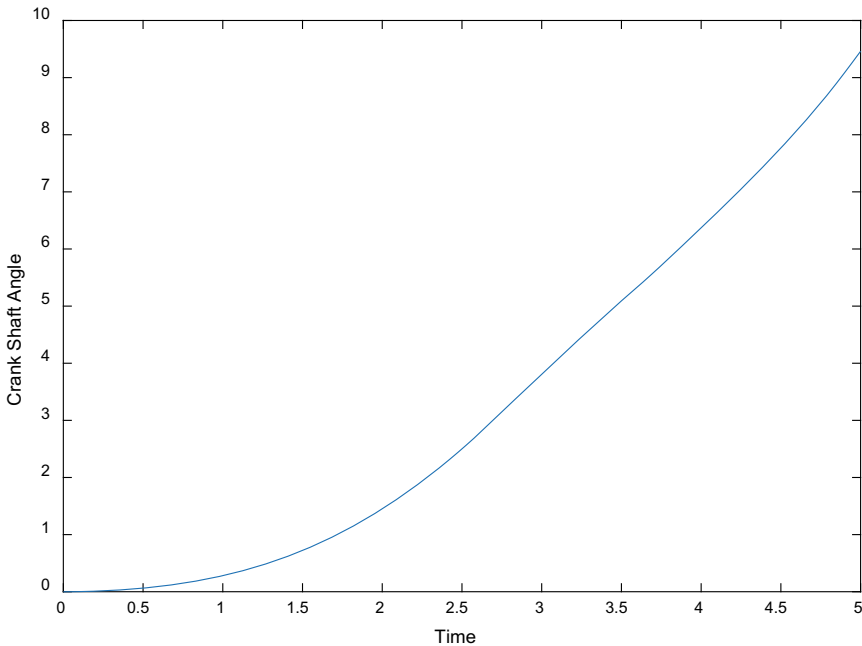


Fig. 1 Synthetic data: crankshaft angle θ in radians given as a function of dimensionless time. Time is given in dimensionless units

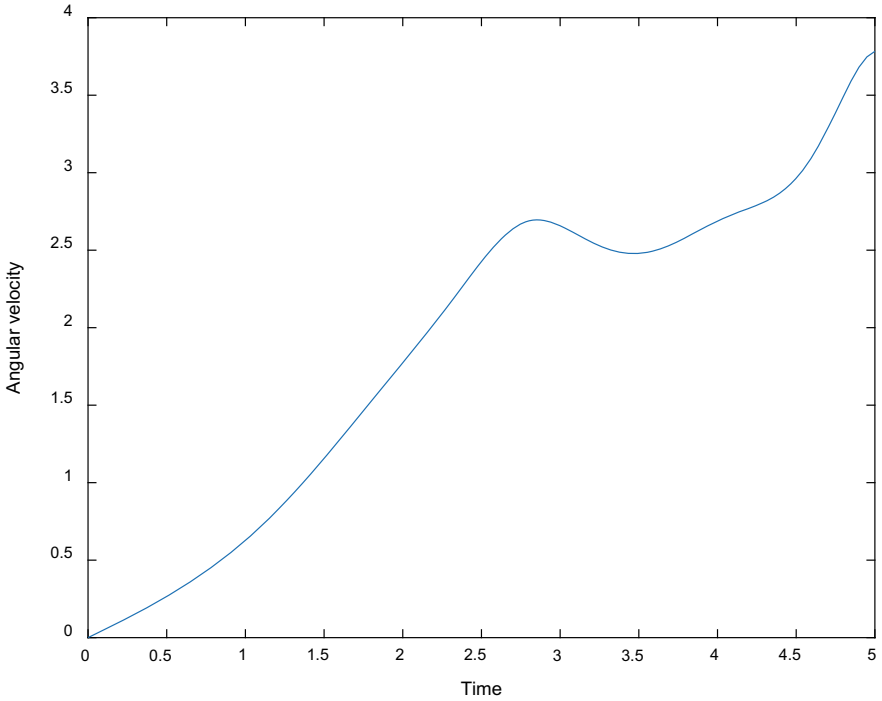


Fig. 2 Synthetic data dimensionless angular velocity ω of crankshaft as a function of dimensionless time

coefficients close to the original values was obtained as shown in Table 1. Good agreement is obtained both qualitatively in which terms should be included in the model and quantitatively: the values of the coefficients recovered agree closely with the original coefficients.

Table 1 Comparison of model coefficients with those recovered via the SINDy algorithm

Original model	SINDy model
0.5	0.5006
0.75	0.7533
0.75	0.7306
0.25	0.2489

Good agreement is obtained between the sets of coefficients

5 Conclusions and Further Directions

The ability to use data to determine dynamical systems describing engineering plant would be an important step forward for condition monitoring. The first step towards demonstrating the feasibility of this concept is whether existing models of such systems can be recovered from simulated data generated by the models themselves. A demonstration that this is possible in the context of a simplified model of a reciprocating compressor is reported here. Future research will extend this first to simulated data from a complete model of the device and then move to investigating the results of using this process with experimental data.

References

1. Wang, J., Wang, Z., Ma, X., Smith, A., Gu, F., Zhang, C., Ball, A.: Locating sensors in large-scale engineering systems for fault isolation based on fault feature reduction. *J. Franklin Inst* **357**(12), 8181–8202 (2020)
2. Champion, K., Lusch, B., Kutz, J.N., Brunton, S.L.: Data-driven discovery of coordinates and governing equations. *Proc. Natl. Acad. Sci.* **116**(45), 22445–22451 (2019)
3. Elhaj, M., Gu, F., Ball, A.D., Albarbar, A., Al-Qattan, M., Naid, A.: Numerical simulation and experimental study of a two-stage reciprocating compressor for condition monitoring. *Mech. Syst. Sig Process* **22**(2), 374–389 (2008)
4. Loiseau, J.C., Noack, B.R., Brunton, S.L.: Sparse reduced-order modelling: sensor-based dynamics to full-state estimation. *J. Fluid Mech.* **844**, 459–490 (2018)

Rolling Bearing Remaining Useful Life Prediction Based on LSTM-Transformer Algorithm



Xinglu Tang, Hui Xi, Qianqian Chen, and Tian Ran Lin

Abstract Bearings are the most critical components in modern industrial rotating machinery. If a bearing is damaged, it can lead to serious consequences such as an interruption to a production line and financial losses. It is important to monitor the bearing operation condition and to predict the remaining useful life (RUL) of bearings so that a scheduled maintenance can be planned ahead. In order to improve the accuracy of a bearing RUL prediction, a new data-driven RUL prediction technique based on Long Short-Term Memory (LSTM) network and Transformer network is proposed. Firstly, a total of 8 degradation characteristics in both time and frequency domains are extracted from the bearing data to be used as the input features. After the data preprocessing steps such as normalization and sliding window interception, the degradation characteristic dataset is obtained. Then, the proposed LSTM-Transformer technique is applied to the characteristic dataset for training and prediction. The prediction result shows that the proposed technique can effectively overcome the information loss of LSTM network caused by the increase distance between the input and output sequences to produce a more accurate RUL prediction. The RUL prediction obtained using the proposed technique is compared with those using existing techniques such as GRU, LSTM and CNN networks for an evaluation of the effectiveness and efficiency of the proposed technique. It is confirmed that the proposed technique can yield a more accurate bearing RUL prediction than the existing techniques.

Keywords Rolling bearing · Data driven · Remaining useful life · LSTM · Transformer

X. Tang · H. Xi · Q. Chen · T. R. Lin (✉)
Centre for Structural Acoustics and Machine Fault Diagnosis, Qingdao University of Technology,
Qingdao 266520, China
e-mail: trlin@qut.edu.cn

X. Tang
e-mail: tangxinglu_qut@163.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_18

207

1 Introduction

Rolling element bearings are one of the most critical components in modern industrial rotating machinery [1]. Due to the harsh operating conditions in practical applications, rolling element bearings are prone to failure which can cause substantial economic losses and even human casualty [2]. Therefore, it is necessary to carry out effective preventive maintenance of bearings.

RUL is a key index to predict the potential machine failure in a production line [3]. RUL prediction techniques can be generally grouped into two categories: physical model-based techniques and data-driven techniques [4]. Physical model-based techniques usually require rich professional knowledge and experience to establish a mathematical model to describe the physical relationship between subsystems in the analysis of the system degradation process [5]. For example, Liao [6] utilized an enhanced Paris-Erdogan (PE) model with a state-space model for the RUL prediction of an equipment. El-Tawil and Jaoude [7] developed an analytic prognostic methodology based on nonlinear damage laws to determine the RUL of a system. The physical model based techniques can achieve an accurate RUL prediction when the interrelationship between the system components and the working characteristics of the equipment are well defined. However, due to the increasing complexity of modern mechanical systems, it becomes harder and harder to establish an accurate physical model.

On the contrast, data-driven RUL prediction techniques use the performance degradation data of a system instead of analyzing the degradation process of the system [8]. With the fast development and expansion of artificial intelligent techniques, several machine learning and deep learning techniques have been successfully employed for the RUL prediction of rotating machinery. For example, Jaseena and Koor [9] proposed a decomposition-based hybrid deep Bidirectional Long Short Term Memory models with skip connections for the wind speed prediction, and shows that the proposed method can produce a better prediction accuracy than other models. Vaswani et al. [10] proposed the Transformer technique, based on an attention mechanism and dispensing with recurrence and convolutions concurrently. The application of the proposed technique on two machine translation tasks shows that the model is more parallelizable, requires substantially less time to train.

Aiming to overcome the drawback of LSTM network such as the information loss problem and parallel computing problem, this paper combines the LSTM network and Transformer technique for a more accurate bearing RUL prediction. The effectiveness of the technique is verified using a set of published bearing degradation data.

2 An Introduction of the Related Techniques

2.1 LSTM Neural Network

LSTM neural network is a special type of recurrent neural network (RNN), which alleviates the gradient vanishing and gradient explosion problems of traditional RNN model to a certain extent. Figure 1 shows the basic unit structure of a LSTM.

The basic update formula of LSTM is as follows:

$$\mathbf{F}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \tag{1}$$

$$\mathbf{N}_t = \sigma(\mathbf{W}_n \mathbf{x}_t + \mathbf{U}_n \mathbf{h}_{t-1} + \mathbf{b}_n) \tag{2}$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) \tag{3}$$

$$\mathbf{C}_t = \mathbf{F}_t \cdot \mathbf{C}_{t-1} + \mathbf{N}_t \cdot \tilde{\mathbf{C}}_t \tag{4}$$

$$\mathbf{O}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \tag{5}$$

$$\mathbf{h}_t = \mathbf{O}_t \cdot \tanh(\mathbf{C}_t) \tag{6}$$

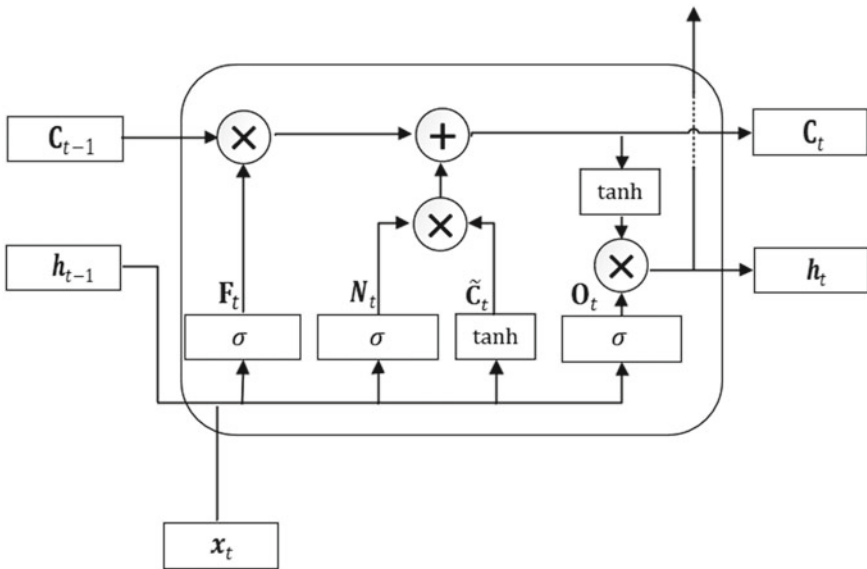


Fig. 1 An illustration of the LSTM unit structure

where F_t , N_t , O_t denote forgetting gate, input gate and output gate respectively, h_{t-1} and h_t are the output of the previous time and the current time respectively, x_t is the input of the current time, W and U are the weight matrix, b is the deviation vector, $\sigma()$ is the sigmoid activation function, \tilde{C}_t denotes the candidate value vector, $\tanh()$ is the tanh activation function, “ \cdot ” represents the product of the corresponding elements of two vectors participating in the operation.

2.2 Transformer Model

A transformer model is a network architecture based on an attention mechanism [10]. The transformer model adopts an encoder-decoder architecture. Figure 2 shows the

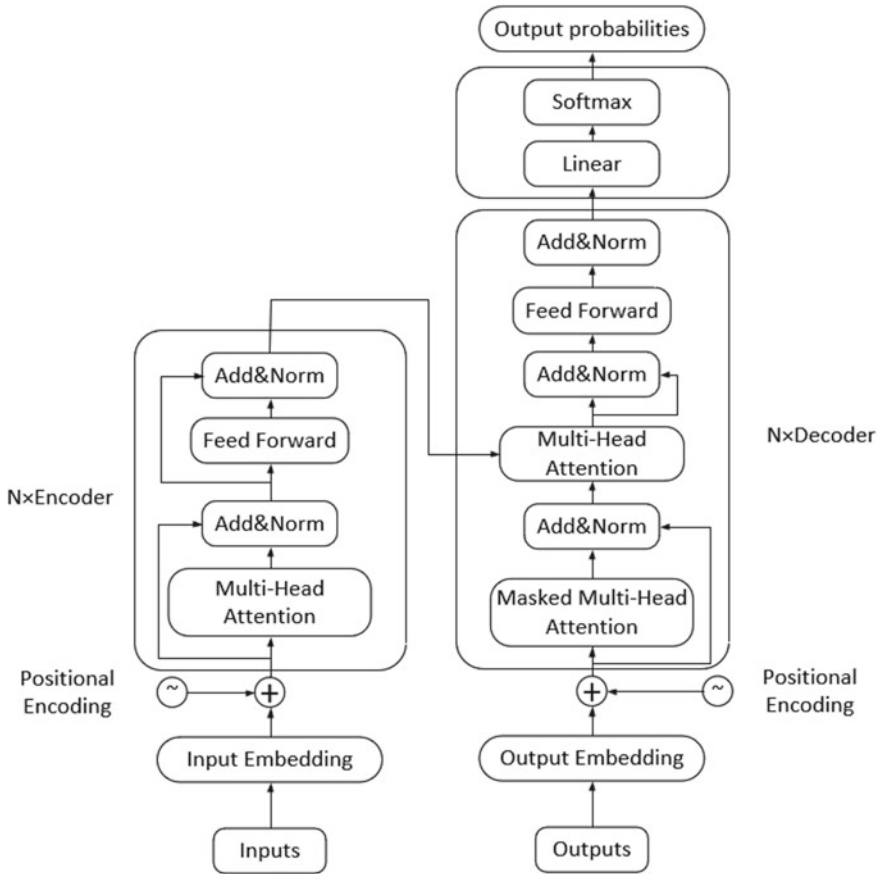


Fig. 2 An illustration of a transformer unit structure [11]

internal structure for a pair of encoder and decoder. A description of the algorithm of the transformer is given below.

A transformer model does not have the capacity to describe the sequence of input sequences where all inputs are passed into the model at once. In order to deal with this problem, an additional vector positioning encoding is added to the input of the encoder layer and the decoder layer to determine the position of the current data.

An encoder consists of two layers, a self-attention layer and a feed-forward neural network layer. A residual connection is used to connect the two sublayers. In the self-attention layer, a multi-head attention mechanism is added. Because the Attention has different distributions in different subspaces, the multi-head attention actually seeks for the correlation between sequences from different angles and combines the correlation captured in different subspaces.

The masked multi-head attention layer at the decoder ensures the consistency between the training stage and the reasoning stage. The rest of the decoder layer is basically the same as the encoder unit.

Finally, a simple fully connected neural network with softmax layer is used for the output of the result.

3 The Proposed LSTM-Transformer Technique

In this paper, LSTM network and Transformer network are combined together in the RUL prediction of roller bearings. A LSTM network is an effective approach in the modeling of time series, though the hidden layers of the network cannot be calculated parallelly at the same time leading to low computational efficiency. On the contrast, a Transformer network does not have such problem since the model input can be fed into the network at once. In addition, Transformer focuses on the relevance of input data rather than the distance, which can also alleviate the information loss problem typically found in a LSTM network. The flow structure of the proposed technique is shown in Fig. 3.

In this approach, a two-layer LSTM is used to capture the temporal information of the input of a time series where a Relu activation function is used in each network to increase the operation speed of LSTM networks, as well as to solve the problem of gradient disappearance up to a certain extent. The sequential characteristics of the input time series can be effectively extracted by using the two-layer LSTM network. The sequential characteristics are then fed into the Transformer network comprising 6 pair-layers of encoders and decoders. Through the self-attention mechanism in the encoders, the degradation trend information of sequence characteristics can be captured. Whilst the decoders can link the degradation information with the degradation stage. Finally, the output matrix having the bearing degradation stage information from the transformer network is used as the input in a dense layer (i.e., a full connection layer) which converts the output of the Transformer into a one-dimensional RUL prediction result.

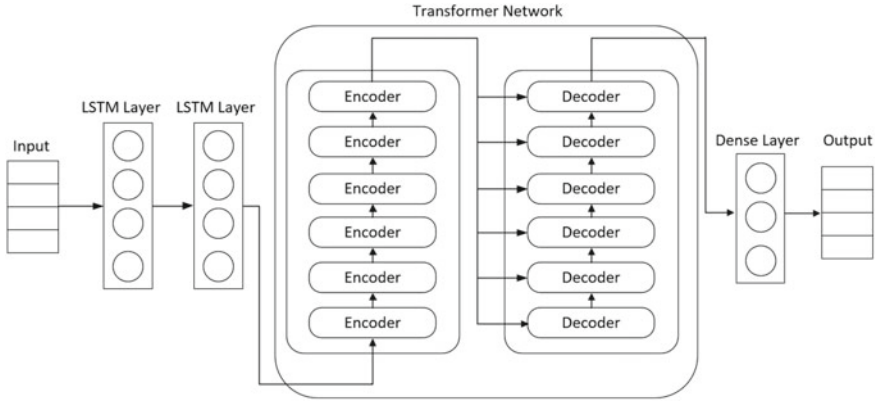


Fig. 3 A flow chart of the proposed LSTM-transformer technique

4 The Parametric Setting and Result Analysis

4.1 Dataset

The bearing degradation dataset from the published PHM2012 Prognosis Challenge bearing data is used in the numerical evaluation of the proposed RUL prediction technique [12]. The Bearing 1-1 and Bearing 1-3 data from the published dataset are used as the training and the test dataset in this study. The operation conditions (i.e., the shaft speed and the external loading) of the bearing degradation test for this two selected bearing datasets are 1800 RPM and 4000 N.

4.2 The Evaluation Metrics

For sake of convenience in the evaluation and quantification process, the root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) are used as the evaluation indices. The formula used in calculating these indices are as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \tag{7}$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \tag{8}$$

$$R^2 = 1 - \left(\frac{\sum_{i=1}^m (y_i - \bar{y}_i)^2}{\sum_{i=1}^m (y_i - \hat{y}_i)^2} \right) \quad (9)$$

where m is the number of data points, y_i is the forecast data, \hat{y}_i is the original data, \bar{y}_i and is the mean value of the forecast data.

RMSE and MAE are two of the most common measures of continuous variables. A smaller value of these indices indicates a better fitting effect. Whilst the value of R^2 reflects the effectiveness of a regression model, a higher value is expected.

4.3 The Network Parametric Setting

Eight typical characteristics in both time and frequency domain such as peak value, variance, root mean square, peak-to-peak value, waveform factor, pulse factor, mean frequency and frequency standard deviation are extracted from the bearing data. In order to improve the convergence speed and the prediction accuracy of the model, the prediction output data are normalized between 0 and 1 where 1 represents that the bearing is in the new condition, and 0 represents that it has reached the end of life.

During the network training, the maximum number of iterations is set at 200, and the batch size is set at 32, An Adam optimizer is used in the process where the mean squared error is used as the loss function. After every 50 iterations, the learning rate is reduced to 1/5 of the original rate so that a large learning rate can be maintained at the initial stage of training and the convergence can be accelerated after the initial stage. When the optimization is approaching the optimal value, a small learning rate is used for the training to avoid possible crossing of the optimal value.

4.4 Results and Discussion

After the optimization and training of the model, a RUL prediction of the test bearing dataset (Bearing 1-3) is obtained which is shown in Fig. 4. The straight orange line in the figure represents the actual degradation life of the bearing, while the blue curvy line represents the RUL prediction of the bearing using the proposed technique. It is shown that the RUL prediction using the current technique is in good agreement with the actual bearing degradation life.

In order to further illustrate the advantage of the RUL prediction technique proposed in this paper, a comparison study is undertaken where the RUL prediction using the proposed technique is compared to those using LSTM, GRU, CNN networks. Table 1 lists the three evaluation indices from the RUL prediction results using these networks.

Fig. 4 LSTM-transformer predicting results

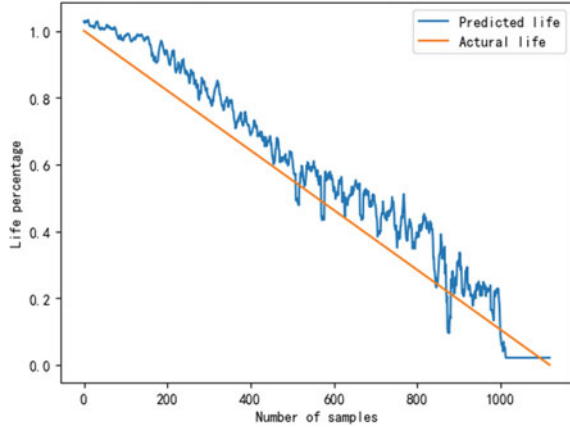


Table 1 Comparison of model effects

	RMSE	MAE	R ²
GRU	0.093	0.078	0.910
LSTM	0.101	0.084	0.903
CNN	0.104	0.091	0.890
LSTM-transformer	0.088	0.076	0.921

It is shown that the prediction result using the proposed technique has the smallest RMSE and MAE values within the four network models implying the proposed technique can produce more accurate RUL prediction than the other network models. The proposed technique also yields the highest value of R² indicating that the proposed is a superior regression model which can effectively capture the degradation trend of the bearing.

5 Conclusion

An effective bearing RUL prediction technique was presented in this paper by combining two LSTM networks with a Transformer network. The proposed technique can effectively overcome the information loss of LSTM networks to produce a more accurate bearing RUL prediction. The comparison study presented in the paper showed that the proposed technique performs better than three other commonly used techniques in the bearing RUL prediction.

References

1. Yu, K., Lin, T.R., Tan, J.W.: A bearing fault diagnosis technique based on singular values of EEMD spatial condition matrix and Gath-Geva clustering. *Appl. Acoust.* **121**, 33–45 (2017)
2. Yu, K., Lin, T.R., Tan, J.: A bearing fault and severity diagnostic technique using adaptive deep belief networks and Dempster-Shafer theory. *Struct. Health Monit.* **19**(1), 240–261 (2020)
3. Kang, Z., Catal, C., Tekinerdogan, B.: Remaining useful life (RUL) prediction of equipment in production lines using artificial neural networks. *Sensors* **21**(3), 932 (2021)
4. Wang, Q., Xu, K., Kong, X., Huai, T.: A linear mapping method for predicting accurately the RUL of rolling bearing. *Measurement* **176**, 109127 (2021)
5. Wang, Z.: Using GMM-HMM model and parallel computing for health estimation and prognosis of turbofan engines. In: 2018 International Conference on Computer Modeling, Simulation and Algorithm (CMSA 2018). Atlantis Press (2018)
6. Liao, L.: Discovering prognostic features using genetic programming in remaining useful life prediction. *IEEE Trans. Ind. Electron.* **61**(5), 2464–2472 (2013)
7. El-Tawil, K., Jaoude, A.A.: Stochastic and nonlinear-based prognostic model. *Syst. Sci. Control Eng. Open Access J.* **1**(1), 66–81 (2013)
8. Ji, Y., Chen, Z., Shen, Y., Yang, K., Wang, Y., Cui, J.: An RUL prediction approach for lithium-ion battery based on SADE-MESN. *Appl. Soft Comput.* **104**, 107195 (2021)
9. Jaseena, K.U., Koor, B.C.: Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. *Energy Convers. Manag.* **234**, 113944 (2021)
10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, pp. 5998–6008 (2017)
11. Source code interpretation of transformer. <https://zhuanlan.zhihu.com/p/110800071>. Last accessed 14 May 2021
12. Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., Varnier, C.: PRONOSTIA: an experimental platform for bearings accelerated degradation tests. In: *IEEE International Conference on Prognostics and Health Management, PHM'12*, pp. 1–8. IEEE Catalog Number: CPF12PHM-CDR (2012)

Research and Application of Order Analysis Technology Without Tachometer Under Variable Speed Condition



Ruibao Yang and Jianguo Wang

Abstract Under the condition of variable speed, the traditional signal processing method cannot accurately determine the fault location of motor bearing. This paper proposes and studies the order analysis method without tachometer. Through the short-time Fourier transform and speed tracking of the collected fault signal, the speed signal of the motor fault bearing is obtained indirectly. The extraction process of the speed signal is completed and the order analysis is carried out to obtain the bearing fault diagnosis results. The results show that this method can effectively judge the fault location of motor bearing.

Keywords Data processing · Nonstationary signal · Order ratio analysis · Resampling · Speed tracking

1 Introduction

With the increasing innovation of industrial technology, (the development of machinery in the process of modernization of society is also more and more important, and gradually to the direction of precision development). Bearings and gears are the most vulnerable components in the operation of the machine. Therefore, it is necessary to detect the status of these devices, predict fault information in advance, and deal with it in time to the maximum extent, which is the original intention of troubleshooting. Order ratio tracking is the key to order ratio analysis, its development step has been steadily moving forward, it can be divided into the following processes: the traditional order ratio analysis technology is hardware scale analysis [1]; The order ratio analysis technology of the tachometer is available [2, 3]; Order ratio analysis technology without tachometer [4, 5].

Zhile and Bo [6] proposed to window the signal first to solve the vibration signal under the non-stable operating conditions, and then used the spectral cliff to obtain

R. Yang (✉) · J. Wang
College of Mechanical Engineering, Inner Mongolia University of Science and Technology,
Baotou 014010, Inner Mongolia, China
e-mail: yrb121311@163.com

the windowing signal of the complex envelope, and then used the order gravity sampling to obtain the order ratio spectrum information. Guanqi et al. [7] combines time–frequency extrusion and order-ratio analysis to diagnose the occurrence of bearing failure, and verifies it through simulation experiments and example analysis, and in the absence of a tachometer, the results of this method are also very ideal. Jianxin et al. [8] proposed that the overall average experience model and order ratio analysis method should be combined to analyze the non-stable signal generated under the condition that the speed is variable and not constant.

2 Order Analysis

In engineering practice, the speed of equipment is generally an unshophonic signal, in which case the order ratio analysis plays its advantages. The key to order ratio analysis is to transform an unsmoothed time domain signal into a medium-smooth angle signal by resampling at an equal angle, which is the essence of the order ratio analysis. For example, in the course of motor operation, the signal extracted at a constant speed does not reflect the state of the entire working process, on the contrary, it is a process of increasing speed, reducing speed, not a state of constant speed. Order-to-scale analysis does have advantages over other traditional spectrum and time–frequency analysis methods in the analysis of non-smooth signals because its results are not affected by changes in speed, whereas other analytical methods are not.

2.1 Definition of Order Analysis

The rotation of mechanical equipment cannot avoid the change of speed naturally, and the order ratio analysis because of its own resampling process and dilute the impact of speed, can show the operation state of mechanical equipment, with this characteristic, in some large-scale equipment operation status monitoring and fault characteristics analysis has a more objective understanding. The number of points sampled in each lap is the same when the order ratio is analyzed, and the position remains the same, and a data is collected whenever a certain angle is turned, which is the most essential part of the order ratio analysis. The horizontal coordinate of the final result of the order ratio analysis is the scale ratio, and the ordinate is the magnitude intensity. The order ratio is defined as the number of fluctuations per revolution of the reference axis: order = the number of fluctuations/per revolution, the relationship between the order ratio and frequency f is:

$$o = \frac{60f}{n} \quad (1)$$

where n -the speed of the reference shaft (r/min); f -represents frequency (Hz). Set the swing frequency $f_n = n/60$, in turn/second (r/s), then:

$$f = o * f_n \tag{2}$$

As can be seen from the definition of the above order ratio analysis, the change of speed has no effect on the results of the order ratio analysis, so it can analyze the operating state of the equipment well, and is suitable for analyzing the fault characteristics affected by the change of speed. In general, in the spectrum analysis, the signal will be divided into equal time segments, that is, every once in a while to collect a data, so that the final sampling results will change with the original signal change trend, the original smooth signal result is a smooth signal, the original signal is variable, non-smooth signal so the final results can be imagined. The difference between order ratio analysis and its difference is that no matter what the original signal is, the sample it ends up taking is a smooth angle signal, which is the biggest difference, so the order ratio analysis is more used to process the complex signal changeable.

2.2 Second Order Specific Gravity Sampling

Order-specific sampling is the essence of order ratio analysis, a process in which the vibration signal at the equivalent interval of the synchronously collected speed pulse signal is resampled at equal angle intervals. Here the speed pulse signal has two functions, the first is by the speed pulse signal to obtain the speed information of the device, determine the moment of resampling, the second function is to determine the frequency of filtering, and then according to the determined point in time to collect data, and then interpolate, and finally get the result of sampling. Finally, it is transformed quickly, and the order ratio spectrum is analyzed. Determining the moment of resampling is a particularly important step in resampling, so its accuracy is the key to smoothing the signal.

Assuming that the hinge is constant at an angle acceleration motion for a short period of time, the quadrangle equation of the cumulative angular $\theta(t)$ is expressed as:

$$\theta(t) = b_0 + b_1t + b_2t^2 \tag{3}$$

Among (b_0, b_1, b_2) is unknown and needs to be solved. Suppose three consecutive pulse points are reached (t_1, t_2, t_3) , The respective corners are $(0, \Delta\theta, 2\Delta\theta)$

$$\theta(t_1) = 0 \quad \theta(t_2) = \Delta\theta \quad \theta(t_3) = 2\Delta\theta \tag{4}$$

In order to make the corner increment $\Delta\theta = 2\Delta\pi/z$, z is the number of pulses per revolution, the upper incoming type available:

$$\begin{bmatrix} 0 \\ \Delta\theta \\ 2\Delta\theta \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \quad (5)$$

Solve the equation to find the equation coefficient (b_0, b_1, b_2):

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \Delta\theta \\ 2\Delta\theta \end{bmatrix} \quad (6)$$

you get to turn any corner θ_i ($\theta \in [0, 2\Delta\theta]$) the corresponding point in time is t_i :

$$t_i = \frac{1}{2b_2} \left[\sqrt{b_1^2 + 4b_2(\theta_i - b_0)} - b_1 \right] \quad (7)$$

When a pulse arrives at the actual calculation, the calculation is recalculated to reduce the cumulative error of calculating the resampling time. The pulse moment at this time is the most t_3 , the first two pulse moments are the most t_1, t_2 , so that there will be repeated calculation points. To avoid duplication, only part of the data point is calculated at each calculation, usually taking the middle part, that is: $\pi \leq \theta \leq 3\pi$. Discrete the crucible and replace it with $k\Delta\theta$. The publicity becomes:

$$t_i = \frac{1}{2b_2} \left[\sqrt{b_1^2 + 4b_2(k\Delta\theta - b_0)} - b_1 \right] \quad (8)$$

K is the interpolation coefficient, and the value range is:

$$\frac{\pi}{\Delta\theta} \leq k \leq \frac{3\pi}{\Delta\theta} \quad (9)$$

By using the point in time obtained by the formula to interpolate the transmission vibration signal, we can get the map with the angle as the horizontal coordinate and the amplitude as the ordinate coordinate, and then complete the resampling process of the signal, and then use FFT to obtain the order ratio spectrum.

3 Short Time Fourier Transform

The difference between the short-term Fourier transformation and the Fourier transformation is that the short-term Fourier transformation is a segmented Fourier transformation, not an overall one, and its key lies in the segmentation. First of all, to carry out a short-term Fourier transformation must first select a suitable window function, the original signal for a series of segments, and then for each small segment of the

signal obtained Fourier transformation, the results of each segment of the transformation in the original order stitched together, this is the complete short-term Fourier transformation process. The formula is:

$$x(t, \omega) = \int_{-\infty}^{\infty} x(\tau)h^*(\tau - t)e^{-j\omega\tau} d\tau \tag{10}$$

In the form, $h^*(\tau - t)$ is the co-choke of window function $A(r - r)$, and window function $h(\tau - t)$ is used to segment intercept the signal $h(\tau - t)$. Intercepts the signal of the t -moment, and after the four-moment transformation, the transient spectrum reflects the frequency component of the r -moment signal. Therefore, all these collections of instantaneous spectrums reflect the global time frequency distribution of the signal. Short-term Fourier transformation in the processing of non-smooth signals also has certain advantages, because its theoretical process is relatively simple, easy to understand, has appeared in various fields of research.

4 Research on Variable Speed Fault Diagnosis Method Without Speed Signal

The data used in this section is a set of data for the motor at the speed reduction is a vibration signal measured in the gearbox, a set of data is a vibration signal measured in the motor, because the speed of the motor in the process is not constant, so the collected data is an impulsive signal, so the use of order ratio analysis of the signal analysis, sampling rate of $25.6k * 2.56$, the fault bearing type is 6215, the parameters are as follows in Table 1.

4.1 Research Objects

The illustration shows the train wheel gearbox, with an external motor with a faulty bearing (6215) connected to the motor and a pinion in the gearbox (Figs. 1 and 2).

Table 1 Fault characteristic frequency

Mode	Cage	Running pulley	Outer ring	Inner ring
6215	0.145	2.852	4.564	6.436

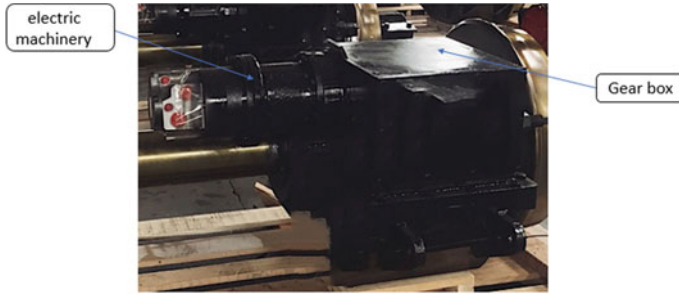
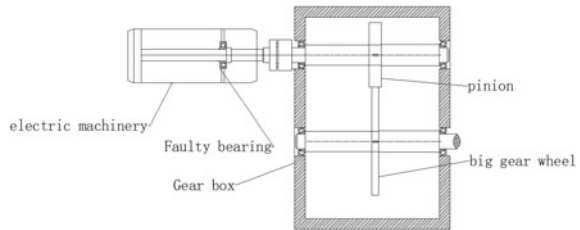


Fig. 1 Picture of motor fault bearing

Fig. 2 Structure diagram



4.2 Realization Principle of Order Analysis Without Tachometer

First of all, the structural diagram analysis: the fault bearing is the bearing of the motor, it is connected with the gearbox coaxial, that is, the speed of the two (frequency conversion) is the same, the speed signal of the pinion also got the speed signal of the motor fault bearing, so the following to get the fault bearing speed signal as the main line analysis experiment, the specific operation process and theoretical part is as follows (Fig. 3):

Step 1: The use of short-term Fourier transformation process vibration signal, first of all, the original signal full transformation. Then the vibration signal is intercepted by window function, and the result of short-term Fourier transformation is obtained by constantly moving the window function. (As shown in Fig. 4, the relationship between time and amplitude, Fig. 5 is a three-dimensional diagram after the short-term Fourier transformation.)

Step 2: Make a short-term Fourier transformation top view (Fig. 6), find out the meshing frequency similar to the plot line in the figure, and speed matching tracking, the tracking steps are as follows:

1. Delineate an area (Fig. 7), the area is the extent of the track.
2. Fixed time axis, in the same case to find the maximum magnitude point A_{max} .
3. Calculate the frequency f_m corresponding to the A_{max} .

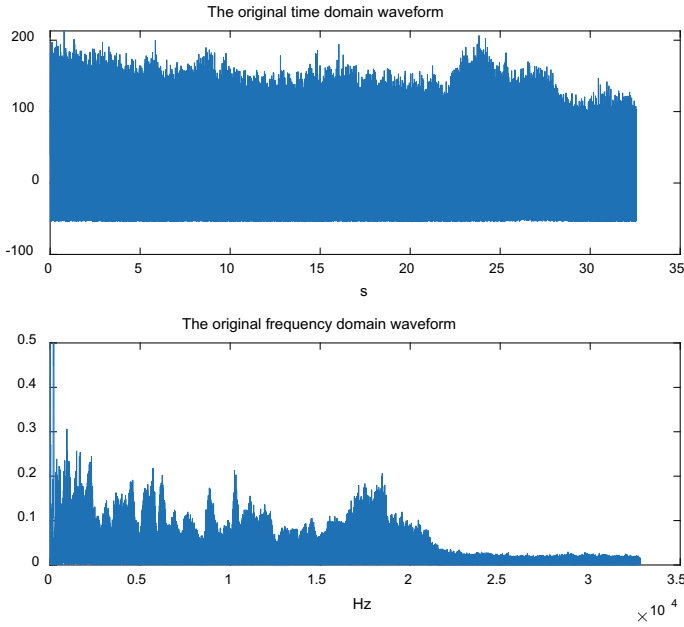
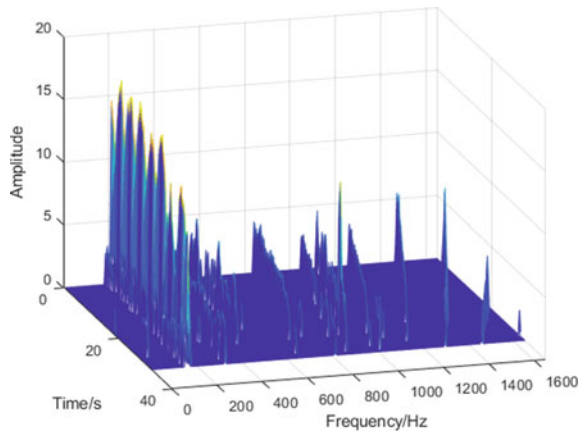


Fig. 3 The original waveform

Fig. 4 A three-dimensional graph of the time frequency after the transformation of the short-term Fourier



4. Slide time to find the frequency f_{mi} for each time period t_i , n as the number of window sliding times.
5. The conversion of all short-term vibration data into the entire time period vibration data is recorded as f .

Step 3: After obtaining the change curve of the speed (Fig. 8), it is necessary to

Fig. 5 Time frequency map after short Fourier transformation (top view)

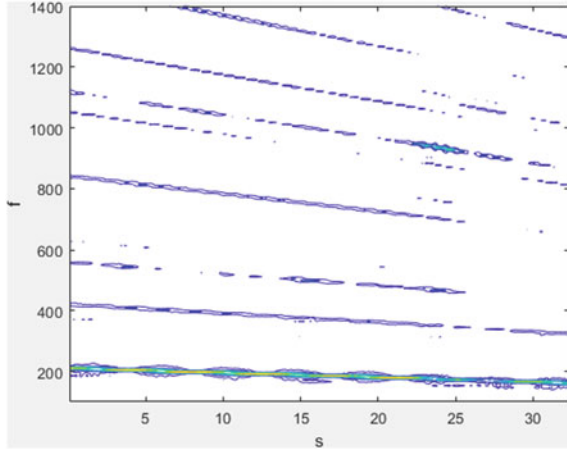
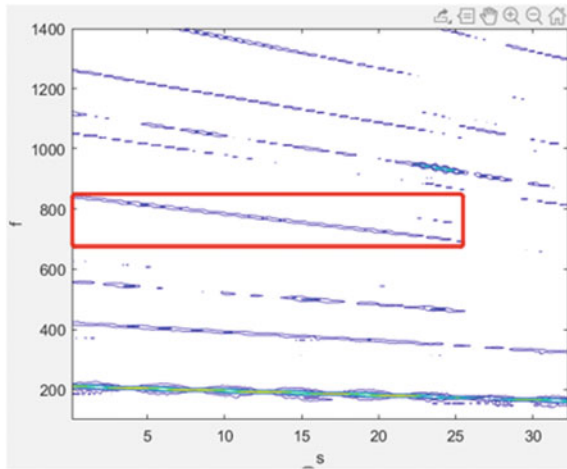


Fig. 6 Speed tracking demarcated areas



subduct the relationship between time and angle, and then to sample the same angle, and finally to analyze the data after the peer angle sampling.

4.3 An Example of Bearing Fault Diagnosis

As shown in the figure, the original signal time domain and frequency domain waveform diagram of the gearbox vibration signal are transformed to obtain its three-dimensional time frequency map for a short period of time, and a top view is made to find its meshing frequency and speed tracking in the figure.

Fig. 7 Speed tracking curve

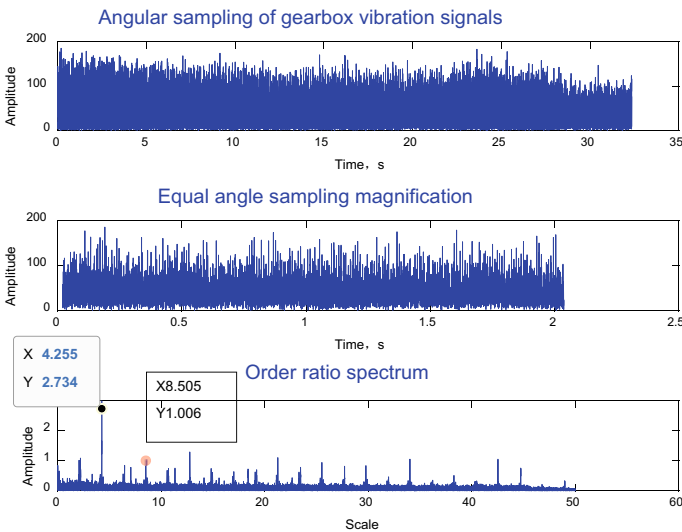
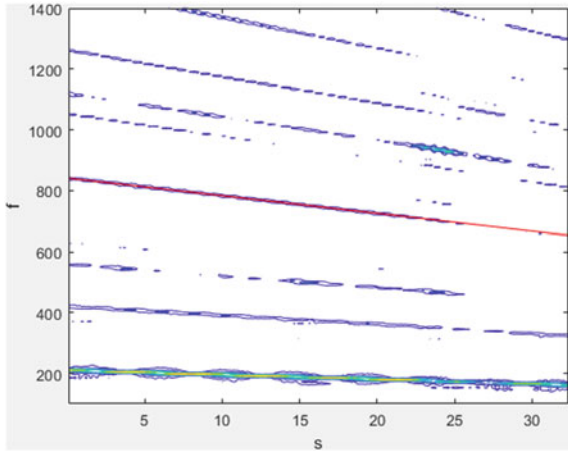


Fig. 8 Order ratio

Find out the demarcated area of the line similar to the meshing frequency and carry out the speed matching tracking results as shown in Fig. 6, the speed tracking completed the meshing frequency of the line, that is, the red line shown in Fig. 7, using the relationship between the engaged frequency and the frequency to obtain the frequency, that is, the speed signal of the motor, and then can carry out order ratio analysis, as shown in Fig. 8.

As can be seen from the order spectrum (Fig. 8), theoretically at 4.564, 9.128, the order ratio will show obvious performance characteristics and peaks of its octave, while the actual peak will appear at 4.25, 8.501, the order ratio, due to the human

error when collecting signals and the accuracy error of the installation of experimental devices and other factors exist, resulting in the theoretical value and the real value of the error, but still can be judged from this phenomenon in the figure at this time the rolling bearings have an outer ring fault.

5 Conclusion

In this paper, taking the fault bearing in the motor as the research object, combining the advantages of short-time Fourier transform, speed matching tracking and order analysis, a tachometer free order analysis method suitable for variable speed condition is proposed. The experimental results show that the method can effectively extract the speed signal of the fault bearing, and the fault location of the bearing can be detected by reasonable analysis.

References

1. Ning, C.: Study on the application of step ratio analysis in the diagnosis of rotating mechanical faults. Tianjin University of Technology (2018)
2. Lingli, C., Xin, W., Huaqing, W., et al.: Bearing fault feature extraction method based on improved switch Kalman filtering. *J. Mech. Eng.* **7**, 44–51 (2019)
3. Jiang, W., Hui, Z., Wanjun, T., et al.: Calculating order ratio tracking technology and its application in vibration monitoring. *J. Anhui Inst. Electr. Eng.* **3**, 98–103 (2015)
4. Liang, T.: Summary of Gabor's expanded and transformed research. *J. Anhui Univ. Nat. Sci.* **2**, 2–13 (2017)
5. Yuebo, J., Yu, G.: The development and application of the step ratio analysis technology and prospect of the modern manufacturing engineering **11**, 123–126 (2007)
6. Zhile, W., Bo, W.: Solar wheel fault feature extraction based on envelope ratio analysis. *Solar J.* **9**, 318–323
7. Guanqi, G., Weiguo, H., Ning, L., et al.: The fault detection method of variable speed bearing based on time-frequency extrusion and order ratio analysis. *Vibr. Shock* **3**, 205–210
8. Jianxin, L., et al.: Bearing fault diagnosis based on step ratio analysis and EEMD. *Mach. Tool Hydraul.* **9**, 170–172 (2014)

Simulation Analysis of Tooth Surface Wear Considering Axis Parallelism Error



Ruiliang Zhang and Yandong Shi

Abstract Axis parallelism error cause edge contact and stress concentration, which lead to uneven load distribution and uneven wear of tooth surface, and seriously reduce service life. In this paper, a tooth contact analysis (TCA) model and a gear wear analysis model considering the axis parallelism error were established based on the basic equation of contact problem and Archard's wear equation, and the change of contact characteristics and wear depth were analyzed. The results show that when the axis parallelism error exists, the tooth load is uneven, and the load distribution becomes more uneven with the increase of the error, and the uneven contact is improved with the increase of load. The wear depth decreases in the tooth width direction, and the wear depth decreases firstly and then increases along the line of action, and the wear depth of the pitch point is zero. The wear depths reach their maximum value at the root of the pinion where the stress concentration occurs, and the maximum wear depth increases non-linearly.

Keywords Axis parallelism error · Gear · Contact analysis · Tooth surface wear

1 Introduction

Gear transmission is one of the transmission forms that play a huge role in modern mechanical equipment [1], however, the gear system failure caused by gear failure also seriously affects the service life of mechanical equipment. The assembly error is inevitable in the assembly process of gear mechanism, especially the axis parallelism error in the vertical plane has a greater influence on the contact characteristics in the tooth width direction of the gear [2]. Axis parallelism error will cause contact line deviation and uneven contact in the tooth width direction, and uneven load distribution caused by uneven contact will also affect the wear distribution of tooth surface, all of which have a great impact on the service life of gear. Therefore, it

R. Zhang (✉) · Y. Shi
College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan
030024, China
e-mail: rl_zhang@163.com

is of great significance to study the gear contact characteristics considering the axis parallelism error and its influence on the wear distribution of the tooth surface.

Yuesheng et al. [3], Xue et al. [4], Shao et al. [5] built the tooth surface equation considering the assembly error to analyse the influence of the assembly error on the gear contact path and contact stress. Liu et al. [6], Dalei et al. [7] and Shukun et al. [8] analyzed the influence of assembly error on tooth surface contact characteristics based on finite element theory and finite element software. But none of the above studies considered the influence of assembly error on gear wear. Zhang and Liu [9] analyzed the gear wear distribution under different meshing deviations, but the calculation of contact stress was based on finite element software, whose calculation accuracy and time were greatly related to the number and quality of mesh. Kumar et al. [10] and Chen and Ji [11] studied the influence of radial clearance and corner contact on tooth surface wear distribution based on the established gear contact model, but related studies show that the axis parallelism error in the vertical plane has a greater influence on tooth contact characteristics [2]. Therefore, it is necessary to study and analyze the wear distribution of the tooth surface considering the axis parallelism error in the vertical plane.

In this paper, the axis parallelism error is firstly transformed into the contact clearance along the line of action, and then the tooth contact analysis (TCA) model considering the axis parallelism error is established. Finally, the tooth contact analysis model was combined with Archard's wear equation to establish a tooth wear analysis model considering the axis parallelism error, and the change of tooth contact characteristics and wear depth were analyzed in the presence of axis parallelism error.

2 Tooth Contact Analysis Model Considering Axis Parallelism Error

2.1 Contact Clearance

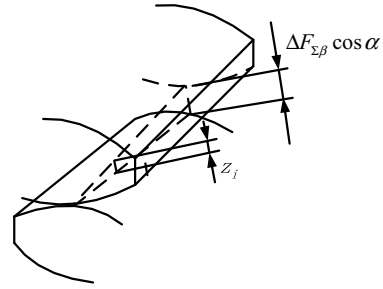
When the axis parallelism error exists, the gears are engaged with each other on one transverse plane, but separated from each other on the other transverse plane. The normal distance of the two meshing gears along the line of action on the side where the separation occurs is defined as the contact line deviation. The axis parallelism error $\Delta F_{\Sigma\beta}$ will generate contact line deviation Δn_{β} in the meshing plane and Δc_{β} in the plane perpendicular to the meshing plane.

Literature [2] found that the contact line deviation Δc_{β} can be ignored, and then the contact line deviation is:

$$\Delta l_{\beta} = \Delta n_{\beta} = \Delta F_{\Sigma\beta} \cos \alpha \quad (1)$$

where α is the pressure angle at the contact point. This contact line deviation is the initial clearance of the gear pair, as shown in Fig. 1.

Fig. 1 Schematic diagram of contact clearance



2.2 Numerical Calculation Model

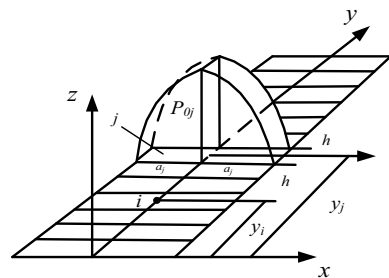
In the process of gear meshing, the effective contact width is far less than the curvature radius of the meshing point, so the meshing area can be treated as an elastic half-space. Because the gear contact is often simplified to finite line contact, one-dimensional simplification should also be made of gear contact: the gear is sliced in the tooth width direction and divided into n strip elements on average. In strip elements j , the stress is assumed to be uniformly distributed in the tooth width direction, and the stress is assumed to conform to the Hertz stress distribution in the involute direction, as shown in Fig. 2. In the figure, Y direction is the tooth width direction, and X direction is the involute direction.

In addition, the contact stress of strip elements j at any location in the involute direction is [12]:

$$p_j = p_{0j} \sqrt{1 - \left(\frac{x'}{a_j}\right)^2} \tag{2}$$

where p_{0j} is the maximum contact stress at the center of the element j , a_j is the semi contact width. According to the Boussinesq solution of the elasticity theory [13], the displacement generated by the stress on element j at the center of element i is:

Fig. 2 Assumption of stress distribution in tooth width direction



$$\omega_{ij} = \frac{1}{\pi E'} D_{ij} p_{0j} \quad (3)$$

where D_{ij} is the flexibility coefficient, and E' is the equivalent elastic modulus. Then the basic equation of the contact problem is satisfied in the tooth width direction [14]:

$$\pi \sum a_j h_j p_{0j} = Q \quad (4)$$

$$\frac{1}{\pi E'} \sum D_{ij} p_{0j} = \delta - z_i(y_i), \quad i = 1, 2, \dots, n \quad (5)$$

where Q is the total load, δ is the total deformation, and z_i is the initial contact clearance of each element. Equations (4) and (5) form a $n + 1$ order equation set with total of $n + 1$ unknowns of p_{0j} and δ . The deformation δ and stress p_{0j} can be obtained by solve the above equations with numerical solution.

3 Wear Modelling

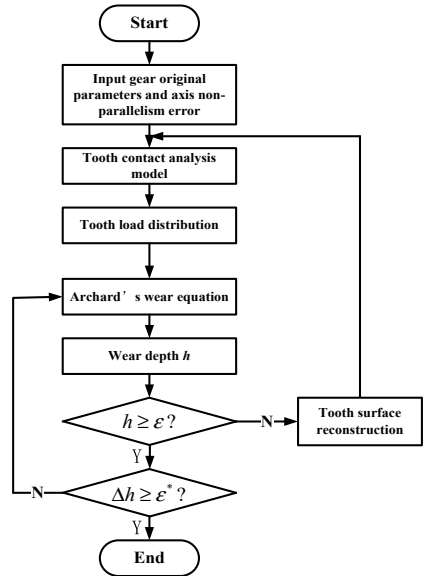
Archard's wear equation is widely used in the numerical simulation analysis of tooth surface wear [15]. Because gear transmission is a dynamic process, and every contact point's load and sliding distance is changing, so the gear wear model is simplified into a discrete quasi-static model. Then, the discrete Archard's wear equation is as follows:

$$h = \int_0^s k p ds \quad (6)$$

where k is the wear coefficient, p is the contact stress, and s is the relative sliding distance. Among them, the contact stress p is calculated by the contact analysis model of the upper section, but the average contact stress of the meshing point is taken when calculating the wear depth.

Because the tooth surface wear is a dynamic process, the tooth surface needs to be reconstructed and updates tooth surface geometry when the wear depth of a meshing point reaches wear threshold ε . When the wear depth of a meshing point reaches the maximum allowable wear ε^* , the simulation ends. The wear calculation process is shown in Fig. 3.

Fig. 3 Flow chart tooth surface wear calculation



4 Results and Discussion

According to the above contents, the numerical simulation was carried out to analyze the tooth surface contact and wear condition in the case of axis parallelism error. The gear parameters used in the analysis were shown in Table 1.

4.1 Tooth Contact Analysis Considering Axis Parallelism Error

The contact stress of the gear surface under different axis parallelism error and the load is 1000 N m as shown in Fig. 4. It can be seen from the figure, the contact stress of tooth surface is unevenly distributed in the tooth width direction. This is because

Table 1 Parameters of gear

Parameter	Value	Parameter	Value
Module/mm	5	Tip clearance coefficient	0.25
Tooth number	30/40	Pressure angle/°	20
Tooth width/mm	30	Young's modulus/GPa	206
Addendum coefficient	1	Poisson's ratio	0.3

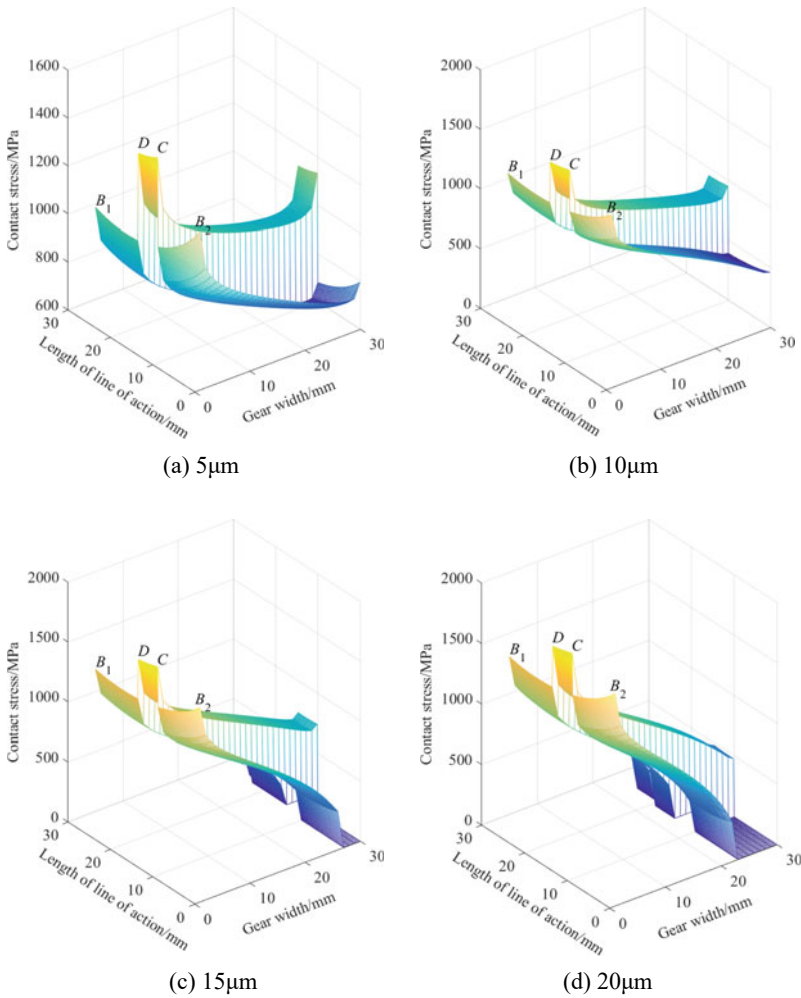


Fig. 4 Contact stress under different axis parallelism error

errors cause stress concentration and uneven load distribution on one transverse plane, but lead to the initial contact clearance on other transverse plane, which causes stress reduction. As the axis parallelism error increases, the contact stress is zero in Fig. 4c and d, which indicates that the gear contact separation occurs here. The uneven distribution of load is aggravated with the increase of axis parallelism error.

The contact stress of double-tooth meshing zone B_2C and DB_1 is obviously smaller than single-tooth meshing zone CD along the line of action, and the contact stress curve of CD is more flat than that of B_2C and DB_1 , which show that the uneven load distribution is improved. This is because CD zone is a single-tooth meshing zone and

the load is larger than the double-tooth meshing zone B_2C and DB_1 , so the increase of load can improve the uneven load distribution.

4.2 Wear Analysis Considering Axis Parallelism Error

When the load is 1000 N m and parallelism error is 20 μm , wear depth distribution of pinion as shown in Fig. 5. It can be seen from the figure that the overall wear depth increases with the increase of wear cycles. The wear depth decreases firstly and then increases along the line of action. Because the sliding distance at the pitch

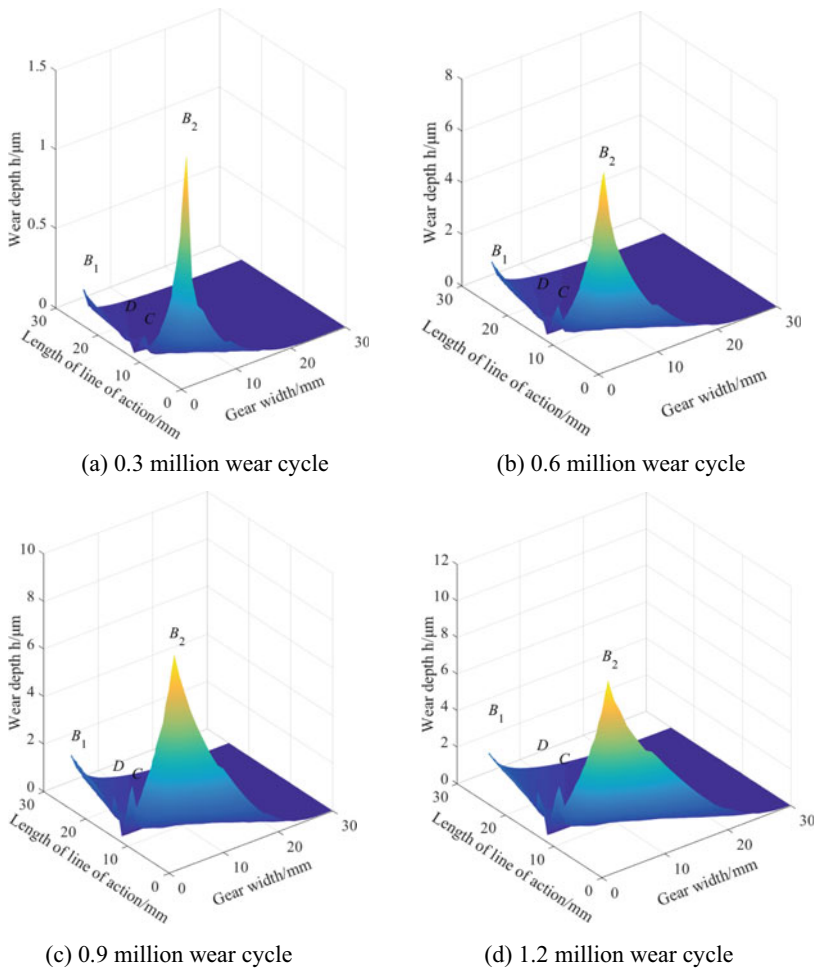


Fig. 5 Tooth surface wear depth distribution of driving wheel

point is zero due to the pure rolling, the wear depth at the pitch point is zero. The wear depth at the dedendum of the pinion is greater than the addendum due to larger sliding distance and contact stress near the dedendum. At the same time, the wear depth has a sudden change in the alternating position of the single teeth-meshing and the double teeth-meshing, which is caused by the sudden change of the contact stress here. When the wear cycle is 0.3 million, 0.6 million, 0.9 million and 1.2 million, the maximum wear depth is 1.46 μm , 7.74 μm , 9.12 μm and 10.68 μm , respectively. It shows that the maximum wear depth increases gradually, but the amplitude of increase decreases gradually, so the increase is nonlinear.

The wear depth of the pinion decreases gradually in the tooth width direction. This is because when the axis parallelism error exists, the contact stress gradually decreases in the tooth width direction, leading to the decrease of the wear depth calculated accordingly. Some parts have zero wear depth due to the gear contact separation occurs here. It can also be seen from Fig. 5 that, due to the influence of parallelism error, the wear amount of the gear on the side where the stress concentration occurs is far greater than the wear amount at the center of the gear. Therefore, the influence of parallelism error on the gear should be considered in the wear prediction of the gear.

5 Conclusions

Based the established tooth contact analysis model and gear wear model considering the axis parallelism error, this paper carries out numerical simulation analysis on the tooth contact and gear wear, and obtains the following main conclusions:

1. When the axis parallelism error exists, the load distribution in the tooth width direction is uneven. The increase of the error will aggravate the uneven load distribution and the increase of the load will improve the uneven contact.
2. When the axis parallelism error exists, the wear depth decreases gradually in the tooth width direction. Along the line of action, the wear depth decreases firstly and then increases, and the wear depth of the pitch point is zero. The wear depths reach their maximum value at the root of the pinion where the stress concentration occurs, and the maximum wear depth increases non-linearly. The wear depth on the side where the stress concentration occurs at the root of the driving gear increases obviously, so it is necessary to consider the influence of the parallelism error in the wear prediction.

References

1. Xiaxia, H., Bingbing, S., Xiaoxia, D., et al.: Research of gear contact based on Hertz contact theory. *J. Zhejiang Univ. Technol.* **44**(1), 19–22 (2016)

2. Jing, X., Yuehai, S.: Longitudinal contact analysis of gear pair with alignment error. *Mach. Design Res.* **31**(6), 44–47 (2015)
3. Yuesheng, Y., Jiong, C., Zhenyu, Z., et al.: Analysis of influence of axis non-parallelism on tooth surface meshing performance of herringbone gear. *J. Mech. Transm.* **45**(01), 129–132 (2021)
4. Xue, C., Han, D., Wen, S.: Adaptive data-driven collaborative optimization of both geometric and loaded contact mechanical performances of nonorthogonal duplex helical face-milling spiral bevel and hypoid gears. *Mech. Mach. Theory*, **154** (2020)
5. Shao, N., Ding, X., Liu, J.: Tolerance analysis of spur gears based on skin model shapes and a boundary element method. *Mech. Mach. Theory*, **144**, 103658 (2020)
6. Liu, S., Song, C., Zhu, C., et al.: Concave and convex modifications analysis for skewed beveloid gears considering misalignments. *Mech. Mach. Theory*, **133** (2018)
7. Dalei, L., Anmin, L., Erliang, Z.: A general calculation method for time-varying meshing stiffness of gear. *J. Chongqing Univ. Technol. (Nat. Sci.)*, **33**(10), 61–66 (2019)
8. Shukun, L., Lei, H., Xiang, C., et al.: Design and experimental research of heavy duty gear modification based on ROMAX. *Coal Mine Mach.* **42**(01), 8–10 (2021)
9. Zhang, J., Liu, X.: Effects of misalignment on surface wear of spur gears. *ARCHIVE Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* 1994–1996 (vols. 208–210) **229**(9), 1350650115574732 (2015)
10. Kumar, P., Hirani, H., Agrawal, A.K.: Modeling and simulation of mild wear of spur gear considering radial misalignment. *Iran. J. Sci. Technol. Trans. Mech. Eng.* **43**(July suppl 1), S107–S116 (2019)
11. Chen, Z., Ji, P.: Study on wear in spur gears based on an improved load distribution model considering the effects of corner contact. *Eng. Fail. Anal.* **115**, 104605 (2020)
12. Hertz, H.: On the contact of elastic solids and on hardness. *Miscellaneous Papers*. MacMillan, London (1896)
13. Jiwei, L., Tianyu, L.: *Analysis Calculation and Application of Rolling Bearing*, pp. 1–2. Machinery Industry Press, Beijing (2009)
14. Yanwei, W., Jiwei, L., Liping, C.: Analysis of the contacts of taper roller bearings. *J. Huazhong Univ. Sci. Technol. (Nat. Sci. Ed.)* **35**(9), 19–21 (2007)
15. Archard, J.F.: Contact and rubbing of flat surfaces. *J. Appl. Phys.* **24**, 981–988 (1953)

A TFG-CNN Fault Diagnosis Method for Rolling Bearing



Hui Zhang, Shuying Li, and Yunpeng Cao

Abstract It is difficult to obtain enough data to train a robust diagnosis model for different rolling bearing faults, and the existing intelligent bearing fault diagnosis algorithms have insufficient generalization ability. Therefore, a rolling bearing fault detector based on the time–frequency graph and convolution neural network (TFG-CNN) is introduced to improve the generalization performance of the fault diagnosis algorithm as much as possible under the condition of considering the diagnosis accuracy and sample size. The specific implementation method is to use Fast Fourier transform (FFT) to transform the vibration data of rolling bearing into a two-dimensional network graph, and then use CNN to classify them. Finally, the performance of the proposed method is analyzed by using the rolling bearing fault datasets of Case Western Reserve University, and analysis results show that the proposed method can simultaneously diagnose the fault location and severity of rolling bearing, and has good cross-domain diagnosis ability and anti-noise performance.

Keywords Rolling bearing · Graph · Convolution neural network · Cross-domain diagnosis

1 Introduction

Time–frequency analysis is an effective method to extract features of rolling bearing vibration signals. Such methods include short-time Fourier transform [1], Wigner Ville Distribution [2], wavelet transforms [3], empirical mode decomposition [4], local singular value decomposition [5], etc. Then, some intelligent classifier algorithms, such as support vector machine (SVM), Bayesian classification, and random

H. Zhang · S. Li · Y. Cao (✉)

College of Power and Energy Engineering, Harbin Engineering University, Harbin 150001, China
e-mail: caoyunpeng@hrbeu.edu.cn

H. Zhang

e-mail: Zhanghui_heu@163.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

237

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_21

forest method, are used to identify the fault type and severity [6–8]. However, the accuracy of those methods is not sufficient and the generalization ability is poor.

In recent years, deep learning models such as deep belief network [9], deep stack self-manipulator [10], convolution neural network [11], cyclic neural network [12], long-term memory neural network [13] are applied to fault diagnosis. While those presented algorithms rely on a large data set to improve the diagnosis rate, and the limited samples of the actual rotating equipment are usually ignored. Using standard data set, When the size of the training set is reduced to 20% of the original size of the training set, the accuracy of the traditional CNN model is 72.4% [14], the Adaptive CNN network combined with an adaptive algorithm is 87.9% [15], and the SN-SSGAN network [16] uses SVM as a classifier is 97.65%. Unfortunately, when the length of the training set is less than 1% of the data set, the accuracy of the network plummets. As one of the most successful intelligent algorithms for rolling bearings, the recognition rate is 82% when the sample size suddenly drops to 10% [17]. At the same time, the existing neural network models have little research on the classification performance of data sets with various external factors of Background noise and change of working conditions. Although there are some breakthroughs in the research on the adaptability of the model under load changes. However, the introduction of such methods is bound to require extensive data for model correction, which will directly affect the stability of the model in a limited sample.

In this paper, a bearing fault diagnosis algorithm based on TFG-CNN is established. The purpose is to improve the generalization performance of the algorithm as much as possible under the condition of both diagnosis accuracy and sample size.

The specific work and contributions of this paper are as follows:

1. The TFG is used to fully utilize the bearing spectrum characteristics and pays more attention to the relationship between different sequences.
2. CNN is used to classify the characteristic diagram of rolling bearing, and the highest accuracy is 100%.
3. TFG-CNN has excellent performance under the minimum sample, thus solving the problem of insufficient data in actual operation.
4. Aiming at the problem of insufficient adaptability of the model in variable load operation of equipment, a specific solution is proposed.

2 Method

2.1 Time–Frequency Graph

It is widely known that FFT can extract the time–frequency information of rolling bearing quickly and effectively. Consequently, the paper adopts FFT to analyze the signal of rolling bearing to further discover the difference between different fault states.

$$x(k) = \sum_{n=0}^{N-1} x(n) \left(\cos 2\pi k \frac{n}{N} - j \sin 2\pi k \frac{n}{N} \right) (k = 0, 1, 2 \dots N - 1) \quad (1)$$

where $x(n)$ is the sampling signal and N is the signal length.

Based on the experience of rolling bearing fault diagnosis and considering the fluctuation caused by roller sliding, the method of setting bandwidth is used to generate spectrum sequence. In addition, according to the fault characteristic frequency and rotation frequency information, the characteristic information of rolling bearing can be recognized effectively by choosing the appropriate bandwidth.

Therefore, this paper proposes to generate m spectrum sequences with N Hz bandwidth to identify the bearing fault state. The formula is as follows:

$$E_h = \left(\sum_{k=1}^n A_k^2 \right) / N \quad (2)$$

where $k = 1, 2, \dots, n$, n represents the number of frequencies contained in each frequency band, $h = 1, 2, \dots, m$, m represents the number of frequency bands, A_k is the amplitude of the k th frequency in the frequency band, E_h is the average energy in this frequency band. And N is the bandwidth. The Fourier change diagram and spectrum sequence diagram are shown in Fig. 1.

This paper proposes a TFG model as shown in Fig. 2 which ignores the overall relationship and turns to consider the relative relationship between different sequences. Specifically, a one-dimensional array is used to store the data of all vertices in the graph, on the other hand, a two-dimensional array is used to store the data of the relationship between vertices. These two aspects of information constitute an adjacency matrix. In this way, the graph model is established with each frequency segment as

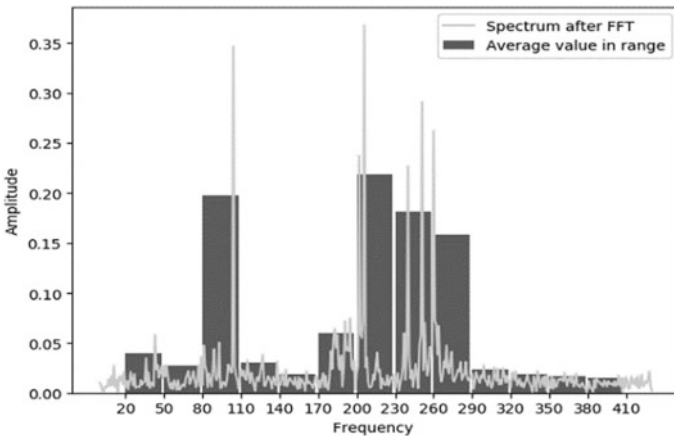
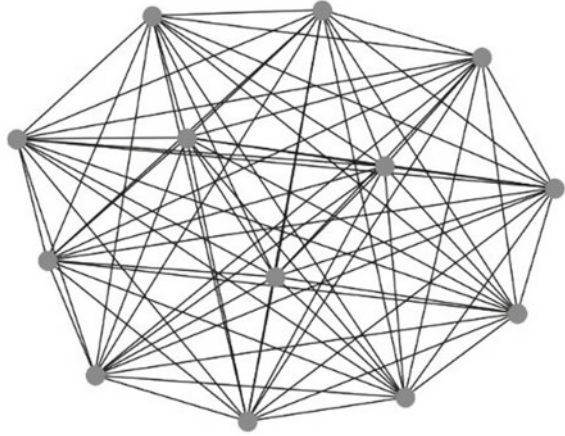


Fig. 1 Fourier transform diagram and spectrum sequence diagram

Fig. 2 TFG model

the vertex and the average amplitude difference of each frequency segment as the weight. Ultimately, the spectrum is converted to a fully connected network graph.

2.2 Convolution Neural Network

CNN is a kind of deep feedforward neural network with the characteristics of local connection and weight sharing. Compared with a feedforward neural network, CNN has fewer parameters. For this paper, two-dimensional convolution is used to transform the fully connected network graph into a higher level and more abstract representation. To enhance the expression ability of the network, dropout is used in each layer. After convolution, BN is normalized in batches. The basic CNN components in this paper are as follows.

Convolution layer

The function of the convolution layer is to extract the feature of a local region. Feature mapping is the feature extracted by convolution of bearing vibration sequence, and each feature mapping can be regarded as a class of extracted image features. The output feature map is calculated as follows:

$$Z^P = W^P * X + b^P \quad (3)$$

$$Y^P = f(Z^P) \quad (4)$$

where $*$ is the convolution operator, $f(\cdot)$ is nonlinear activation function, W^P is convolution kernel, b^P is scalar offset, and Y_p is the feature map and Z^P is the net input of the convolution layer.

Pooling layer

The pooling layer can effectively reduce the size of the model, improve the calculation speed, and improve the robustness of the extracted features. In this paper, maximum pooling is used to reduce the convolution sequence, and the maximum value is used to replace the value of a region of the output feature map. The calculation formula is as follows:

$$y_{m,n}^d = \max(x_i) \tag{5}$$

where $i \in R_{m,n}^d$, x_i is the activity of each neuron in the region R_K^d .

Fully connected layer

The connection layer plays the role of classifier in the whole convolutional neural network. The full connection layer maps the learned distributed feature representation to the sample tag space. When the fully connected layer is used as the output layer, the activation function is Softmax:

$$y = softmax(W^T x) = \exp(W^T x) / (1_C^T \exp(W^T x)) \tag{6}$$

where W is a matrix composed of weight vectors of C classes, 1_C^T is the all-1 vector of C dimension, and Y is the vector composed of prediction conditional probabilities of all categories.

2.3 The Diagnostic Flowchart of TFG-CNN

The flow chart of the TFG-CNN proposed in this paper is shown in Fig. 3, which mainly includes the following two parts: the production of neural network data set and the structure of convolution neural network.

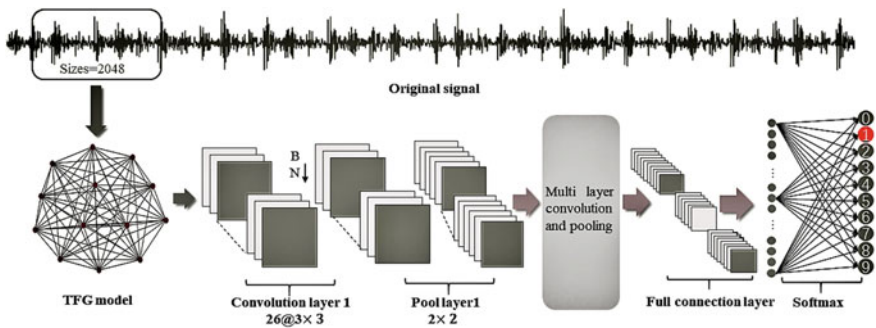


Fig. 3 The diagnosis flow chart of the model proposed in this paper

Step1: data processing

To train CNN neural network, the noise data is normalized. The normalization formula is as follows:

$$x = (x - x_{min}) / (x_{max} - x_{min}) \quad (7)$$

The random sampling method is used to sample the bearing fault sequence. Compared with overlapping sampling, random sampling reduces the number of training samples, but it will greatly improve the randomness of the data, and the generated model can be better generalized to unknown data.

3 Results and Discussion

This paper uses Case Western Reserve University (CWRU) rolling bearing data to show the performance of the proposed method by contrast with current algorithms. The experimental platform is shown in Fig. 4.

The drive end bearing model is SKF6205, and the bearing is a single point damaged by EDM. In addition, an acceleration sensor is respectively placed above the bearing pedestal at the fan end and the drive end of the motor to collect the vibration acceleration signal of the faulty bearing. In this paper, the vibration signal with the sampling frequency of 12k Hz is selected. The geometric dimensions of the drive end bearing and the failure frequency of each component are listed in Table 1, where the rotation frequency is f .

To find the difference between different fault states better. Three groups were randomly selected as the experimental group from the signals. At the same time, the samples representing different fault types and different fault degrees were selected as the control group. For a better visual experience, the adjacency matrix is transformed into a gray image. The TFG-model of vibration signal of rolling bearing is shown in Fig. 5. It can be getting some meaningful results as followed:

Fig. 4 Experimental setup of Western Reserve University

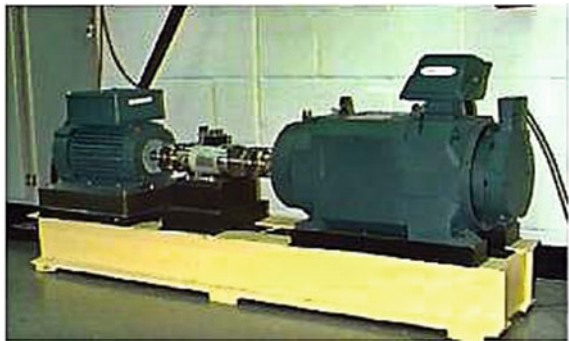


Table 1 Information of rolling bearing data set

Fault types	Data sets	Load	Fault frequencies	Fault diameters	Classification
Normal	A, B, C, D	0hp, 1hp, 2hp, 3hp	–	0	0
Inter-race	A, B, C, D	0hp, 1hp, 2hp, 3hp	5.4152f	0.007	1
	A, B, C, D	0hp, 1hp, 2hp, 3hp		0.014	2
	A, B, C, D	0hp, 1hp, 2hp, 3hp		0.021	3
Outer-race	A, B, C, D	0hp, 1hp, 2hp, 3hp	3.5848f	0.007	4
	A, B, C, D	0hp, 1hp, 2hp, 3hp		0.014	5
	A, B, C, D	0hp, 1hp, 2hp, 3hp		0.021	6
Ball	A, B, C, D	0hp, 1hp, 2hp, 3hp	4.7135f	0.007	7
	A, B, C, D	0hp, 1hp, 2hp, 3hp		0.014	8
	A, B, C, D	0hp, 1hp, 2hp, 3hp		0.021	9

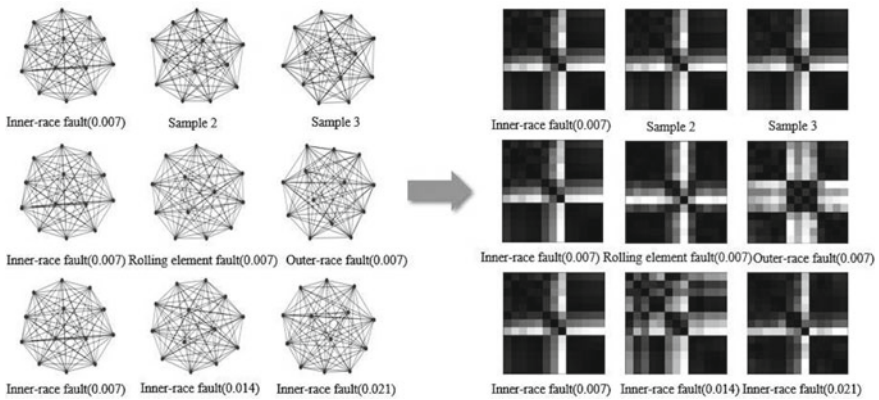


Fig. 5 The TFG model (first from left) of rolling bearing and display of gray image

1. There is a slight difference between TFGs under the same state.
2. There are great differences between TFGs with different fault locations.
3. There are great differences between TFGs with different fault degrees.

Table 2 Structural parameters of CNN

Layer name	Config.			Padding?
Conv1	Conv.	26@3 × 3	Stride = 1	Y
	Pooling	26@2 × 2	Stride = 2	N
Conv2	Conv.	52@3 × 3	Stride = 1	Y
	Pooling	52@2 × 2	Stride = 2	N
Conv3	Conv.	104@3 × 3	Stride = 1	Y
	Pooling	104@2 × 1	Stride = 2	N
Connected layer		100		
Softmax		10		

Therefore, it is feasible to diagnose the fault state of the rolling bearing by the TFG model in theory. Python is used to build a convolution neural network in the Tensor-Flow environment. The parameters of the neural network are shown in the figure, which includes three convolution layers, three maximum pooling layers, and one full connection layer. Simultaneously, the network is optimized by Adam optimizer and dropout is used to prevent over fitting (Table 2).

3.1 Accuracy of the TFG-CNN

The sample size of the training set is 4000, and that of the test set is 1500. Each data set is trained 20 times to ensure the reliability of the results. The recognition rate of the convolution neural network is shown in Fig. 6.

The results show that the convolutional neural network can achieve more than 99% accuracy in each data set. In particular, the accuracy is the most stable on the C data set, and the recognition rate is not less than 99.75%. Therefore, the model proposed in this paper has high accuracy.

3.2 Effect of Sample Size on TFG-CNN

In the actual operation of rotating machinery, most of the time will be in a healthy status. This has contributed to the fault data that can be collected is limited. Consequently, fault diagnosis in the case of unbalanced data is a problem to be faced at present.

To observe the performance of the model under the limit sample, experiments with different sample sizes were carried out. In the experiments, training samples with the sizes of 20, 40, 70, 140, 210, 350, 700, 1400 were used to train the model, and the verification set with the sample size of 700 was used to verify the model. The

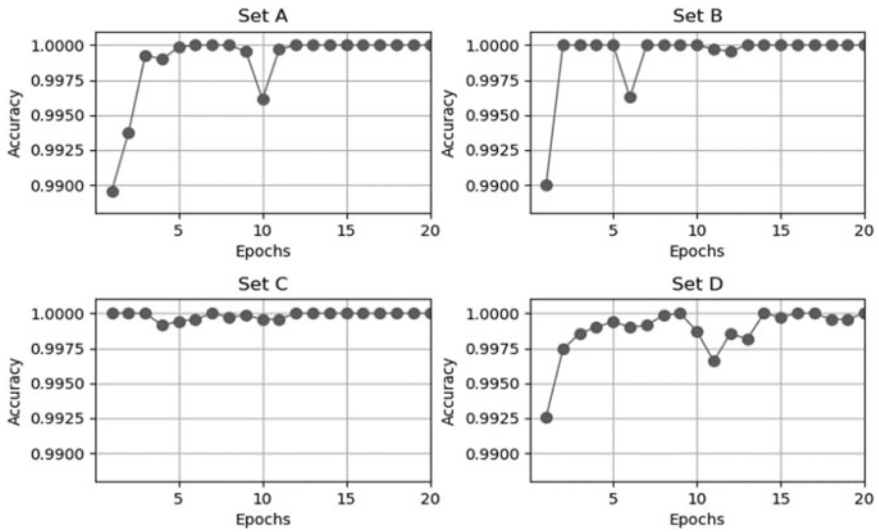


Fig. 6 The accuracy of the model proposed in this paper

test was repeated 20 times. The experimental results are compared with (WDCNN), the best algorithm at present. The results are shown in Fig. 7. The results show that the proposed model not only performs well in the case of a small sample size but also has an excellent performance in the case of limited samples. This provides a good reference for solving the problem of rolling bearing data imbalance.

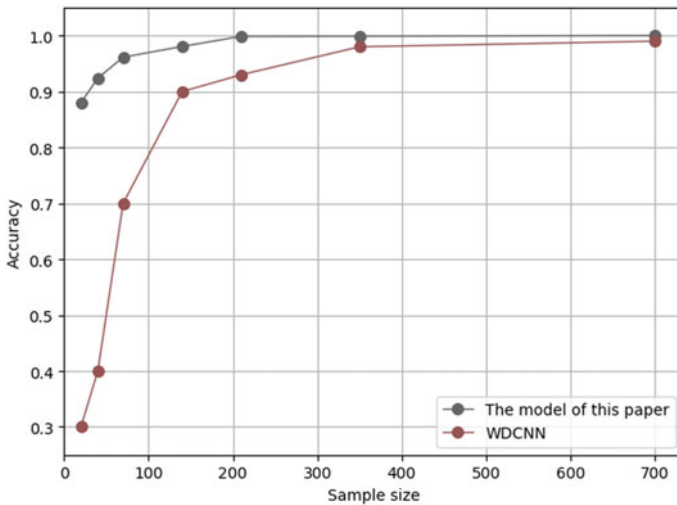


Fig. 7 Influence of training set size on accuracy of different models

3.3 Anti-noise Ability of the TFG-CNN

Because of the complex working environment, huge dynamic load, and continuous high-speed operation state of rotating machinery, the vibration signals of axle box bearings are often submerged in strong noise, which makes it difficult to extract the impact components containing fault information. Therefore, reducing the influence of environment and system noise, extracting the relevant components from the background noise for the efficient diagnosis, has become an important step in rolling bearing fault diagnosis.

To test the anti-noise performance of the model, the signal of the test set is coupled with different degrees of white noise to simulate the noise pollution of the industrial environment. In this experiment, the sample sizes of the training set are 7000, 4000, and 700 respectively, and the size of the verification set is 1500. The test results are shown in Table 3.

Compare this model with the DNN model, MLP model, and WDCNN model in the same way. The comparison results are shown in Fig. 8. It can be found that the accuracy of TFG-CNN is higher than 80% under different degrees of noise background, and the anti-noise performance of this model is better than that of the DNN network and MLP network. In addition, in the weak noise environment, the diagnosis rate is similar to that of the WDCNN model, but in the strong noise environment, the diagnosis rate is slightly weaker than that of the WDCNN model.

Table 3 The diagnostic rate of this model under different levels of noise and data volume

Size of the training set	Load	SNR (dB)					
		0 (%)	2 (%)	4 (%)	6 (%)	8 (%)	10 (%)
7000	0	82.4	86.6	91.7	93.4	95.1	97.0
	1	79.8	88.2	94.7	97.2	98.3	98.6
	2	83.2	90.1	95.6	98.4	99.8	99.9
	3	83.1	89.7	95.3	99.7	99.9	99.9
4000	0	81.8	86.1	90.6	92.5	94.2	96.6
	1	78.5	87.1	93.2	95.5	97.6	98.1
	2	81.1	89.5	95.1	97.6	99.5	99.7
	3	82.8	89.4	94.5	99.1	99.8	99.9
700	0	73.2	81.4	88.6	92.3	93.4	95.2
	1	74.4	79.7	92.3	94.1	97.7	98.5
	2	76.2	82.4	90.2	96.5	97.6	99.3
	3	78.1	81.6	92.4	97.5	98.3	99.5
Maximum		83.2	90.1	95.6	99.1	99.9	99.9

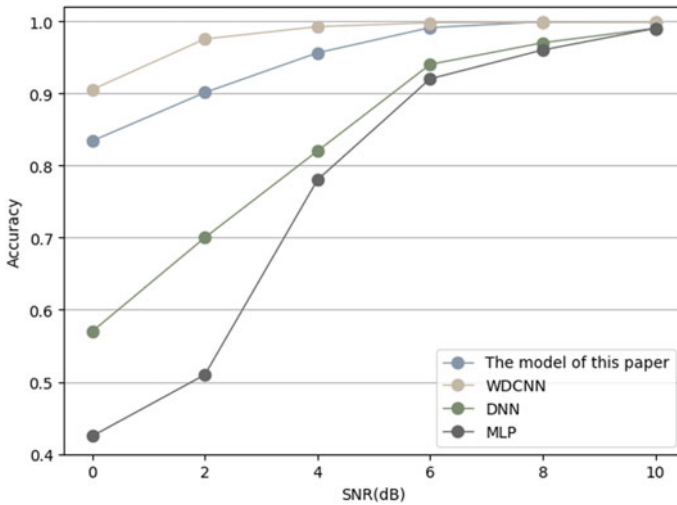


Fig. 8 Accuracy comparison of each model in different noise environments

3.4 Variable Load Adaptability of the TFG-CNN

For rotating machinery, load change is very common. When the load changes, the amplitude, phase, and period information of the vibration signal will change, so it is very important to verify the cross-domain diagnosis ability of the model under the condition of variable load.

To verify the cross-domain diagnostic ability of the model. In this paper, all the data sets following the different proportions of a new data set, and expect to rely on neural network powerful feature extraction ability to learn common features. There is, in addition, A, B, C, and D in the table below represent the data of rolling bearing under different loads. For example, ABC indicates that the training set of the model is the combination of vibration data of rolling bearing under 0hp, 1HP, and 2HP. The experimental results are shown in Table 4.

It can be seen that different sample sizes have little effect on accuracy. When using 0hp and 3HP as the training set, the model has strong adaptability to the test set of other loads, which also shows that the model can extract the feature information of rolling bearing well. However, the model trained by low load vibration signal has poor adaptability to the high load test signal, and the model trained by high load vibration signal has strong adaptability to the low load test signal. In addition, the phase and amplitude of the 3HP load signal are quite different from other loads. This is because some features of high load vibration signal are more dominant to the model, and the model gives these features higher weight in the process of learning, but for low load vibration signal, the information contained in these features will be weakened. Therefore, it is more comprehensive to describe the vibration characteristics of rolling bearings by using the high load vibration signal training model. In the actual operating

Table 4 Accuracy of combination of different data sets

Set	Sizes	A (%)	B (%)	C (%)	D (%)	Set	A (%)	B (%)	C (%)	D (%)
AB	14,000	98.3	99.5	98.8	45.2	AC	98.6	98.7	99.8	39.2
	5600	98.1	98.1	97.3	41.5		96.9	97.5	98.3	38.1
	1400	97.4	96.5	95.1	37.8		95.7	95.3	96.4	37.5
AD	14,000	99.7	98.2	96.7	99.5	BC	88.7	99.1	99.6	29.3
	5600	98.9	96.1	94.7	99.1		86.1	98.3	98.5	27.8
	1400	97.8	93.4	92.1	98.6		83.2	93.4	96.8	25.4
BD	14,000	80.6	99.3	77.1	99.7	CD	78.8	87.2	98.2	99.1
	5600	79.2	97.8	76.4	98.1		75.4	93.5	96.4	98.2
	1400	76.1	95.4	73.5	96.5		73.1	80.3	93.8	96.2
ABC	14,000	98.3	99.4	99.8	31.2	ABD	97.2	99.1	96.3	99.2
	5600	97.2	98.1	97.9	30.5		96.6	97.7	93.6	98.4
	1400	94.9	94.7	95.4	28.4		93.9	93.4	91.7	95.8
BCD	14,000	88.1	99.6	99.9	99.1	ABCD	98.2	99.4	99.9	99.6
	5600	86.5	98.2	98.5	98.1		97.4	98.2	98.6	98.3
	1400	84.7	93.7	96.6	96.7		94.1	92.8	96.7	96.4

conditions of rotating machinery, combined with high and low load data as the training set, higher accuracy can be obtained.

4 Conclusion

In this work, this paper proposes a rolling bearing fault diagnosis method based on TFG-CNN and then uses the data of Case Western Reserve University to analyze its accuracy, sample size adaptability, and cross-domain diagnosis performance to verify the reliability of the model. The conclusions are as follows:

- The model in this paper is used for bearing fault diagnosis with small parameters, strong expression ability, and is easy to design. The model has a high recognition rate in CWRU bearing database.
- The model proposed in this paper has a low requirement for the size of the training set and can have an accuracy rate of 87% under the limit capacity, which provides a reference for other engineering applications.
- The TFG-CNN has good anti-noise performance, and the accuracy of the model can reach more than 80% in the environment of $SNR = 0$. Given the variable load situation, this paper also gives the corresponding solutions.

In future work, there are two prospects:

- Using GNN to diagnose and analyze the network composed of adjacency matrix directly.
- Transfer learning is introduced to further improve the cross-domain diagnosis ability of the model.

References

1. Wang, Y., Deng, S.E., Wang, H.D.: Fault diagnosis method for rolling bearings based on SG filter and STFT with combining window functions. *Bearing*
2. Shi, L., Zhang, Y., Mi, W.: Application of Wigner-Ville-distribution-based spectral kurtosis algorithm to fault diagnosis of rolling bearing. *J. Vibr. Measur. Diagn.* **31**(1), 27–31 (2011)
3. Kankar, P.K., Sharma, S.C., Harsha, S.P.: Fault diagnosis of rolling element bearing using cyclic autocorrelation and wavelet transform. *Neurocomputing* **110**(8), 9–17 (2013)
4. Cai, Y.P., Ai-Hua, L.I., Shi, L.S., et al.: Roller bearing fault detection using improved envelope spectrum analysis based on EMD and spectrum kurtosis. *J. Vibr. Shock* **30**(2), 168–172 (2011)
5. Gao, Y., VILLECCO, F., Li, M., et al.: Multi-scale permutation entropy based on improved LMD and HMM for rolling bearing diagnosis. *Entropy* **19**(4), 176 (2017)
6. Sun, W., Zhao, R., Yan, R., Shao, S., Chen, X.: Convolutional discriminative feature learning for induction motor fault diagnosis. *IEEE Trans. Ind. Inf.* **13**(3), 1350–1359 (2017)
7. Wang, D., Miao, Q.: Some improvements on a general particle filter based Bayesian approach for extracting bearing fault features. *J. Vibr.* **137**(4) (2015)
8. Yang, B.-S., Di, X., Han, T.: Random forests classifier for machine fault diagnosis. *J. Mech. Sci. Technol.* **22**(9), 1716–1725 (2008)
9. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. *Neural Comput.* **18**(7), 1527–1554 (2006)
10. Cabrera, D., et al.: Automatic feature extraction of time-series applied to fault severity assessment of helical gearbox in stationary and non-stationary speed operation. *Appl. Soft Comput.* **58**, 53–64 (2017)
11. Zhang, W., Li, C., Peng, G., Chen, Y., Zhang, Z.: A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.*
12. Mikolov, T., Kombrink, S., Burget, L., Khudanpur, S.: Extensions of recurrent neural network language model. In: *Proceedings of the IEEE International Conference on Speech Signal Processing (ICASSP)*, pp. 5528–5531, May 2011
13. Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y.: Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. C Emerg. Technol.* **54**, 187–197 (2015)
14. Lu, C., Wang, Z., Zhou, B.: Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. In: *Advanced Engineering Informatics*, vol. 32, pp. 139–151 (2017). (View)
15. Guo, X., Chen, L., Shen, C.: Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* **93**, 490–502 (2016)
16. Zhao, D., Liu, F., He, M.: bearing fault diagnosis based on the switchable normalization SSGAN with 1-D representation of vibration signals as input. *Sensors* **19**(9), 2000 (2019)
17. Zhang, W., Peng, G., Li, C., et al.: A new deep learning model for fault diagnosis with anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **17**(2), 425 (2017)

A Gas Turbine Gas Path Digital Twin Modeling Method



Junqi Luan, Yun Peng Cao, Shuying Li, and Ran Ao

Abstract Degradation of gas path performance is the focus of gas turbine condition monitoring. In order to realize gas turbine maintenance cycle performance monitoring, diagnosis and prediction, combining mechanism knowledge and data information, a gas turbine gas path performance digital twin modeling method is proposed, and a health monitoring framework of the gas turbine gas path performance digital twin is constructed. Taking the split-shaft gas turbine as the research object, the gas turbine performance calculation model is established without relying on component characteristic information, the parameter matching method based on differential evolution is studied, and the gas turbine performance digital twin model is developed. The simulation test was carried out the results show that the gas turbine gas path digital twin model realizes the quantitative characterization of the performance degradation of gas path components during the maintenance period, and provides a basis for the gas turbine gas path performance fault diagnosis.

Keywords Gas turbine · Digital twin · Gas path diagnosis · Performance simulation · Differential evolution

1 Introduction

Gas turbines have the advantages of fast start-up, stable operation, high thermal efficiency, etc., and have become the core power equipment in aviation, shipbuilding, and electric power industries. However, due to the increase in the use time and the harsh operating environment, performance degradation will inevitably occur, and in severe cases, the safety margin of the gas turbine may even be reduced [1]. Therefore,

J. Luan · Y. P. Cao (✉) · S. Li · R. Ao
College of Power and Energy Engineering, Harbin Engineering University, Harbin 150001, China
e-mail: caoyunpeng@hrbeu.edu.cn

J. Luan
e-mail: luanjunqi@hrbeu.edu.cn

it is necessary to carry out research on the direction of performance degradation of gas turbines to ensure its safety and reliable operation.

Establishing an appropriate model is the basis for model-based fault diagnosis of gas turbines. Thirunavukarasu et al. [2] proposed a gas turbine modeling and simulation method under different conditions in a virtual test-bed environment. Li et al. [3] proposed a gas turbine dynamic model based on a causal sequence diagram and simulated a 30 kW gas turbine. Hou et al. proposed an improved fuzzy modeling method to obtain higher tracking performance and interference suppression capabilities. Hosseinalipour et al. [4] proposed a static and linear dynamics model to simulate the performance of a micro gas turbine, reducing the amount of calculation and the complexity of the equation. Xu et al. [5] proposed a new improved hybrid modeling method based on machine learning, which can effectively improve modeling accuracy. Liu and Karimi [6] developed a modeling method based on machine learning to predict the performance of compressors and turbines, and achieved good results. In the study of gas turbine performance degradation, Chaobing et al. [7] proposed a gas turbine performance degradation prediction method based on multiple non-linear regression models, which quantitatively diagnosed gas turbine performance degradation. Tülin and Bülent [8] used an artificial neural network (ANN) method to identify exhaust temperature and developed a neural network with the smallest prediction error. In order to achieve a condition-based maintenance strategy and a more reliable assessment of institutional integrity, the US Air Force Research Laboratory has launched the Airframe Digital Twin (ADT) program. It is designed to help reduce the number and duration of aircraft certification tests, reduce repair and maintenance costs, and shorten product development cycles [9–11]. The purpose of the concept is to pass through the virtual entities and subsystems of physical equipment, and this connection is not one-way and static, but is connected throughout the life cycle of the product. Aiming at the traditional modeling methods that rely excessively on component characteristics and lack the ability of adaption and parameter matching, we first establishes a steady-state model for the gas path performance calculation of a gas turbine, and then studies the method of parameter matching based on the differential evolution. In the Labview environment, a digital twin model for gas turbine performance is developed. By inputting different parameters, it can simulate different operating conditions and status of the gas turbine in different scenarios online, which provides a basis for gas turbine performance fault diagnosis. Then the failure factor was implanted on the basis of the gas path performance model to establish the failure model of the gas turbine. After obtaining the operating data of the failure, the diagnosis model of the gas turbine was established to diagnose the performance degradation of the gas turbine.

2 Method

This paper establishes a gas turbine digital twin model, as shown in Fig. 1, to simulate the state of the engine under different operating conditions and environments,

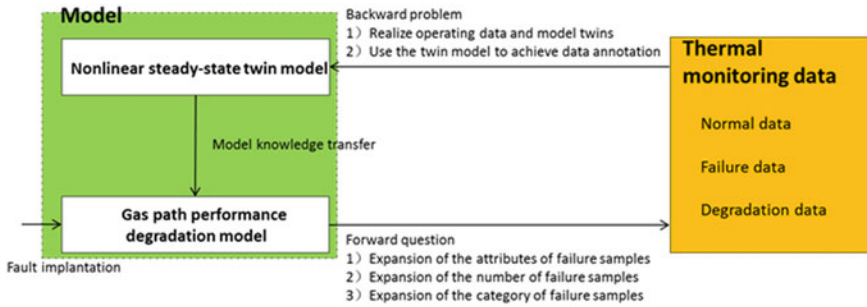


Fig. 1 Digital twin model of gas turbine

and then extract performance degradation characteristics to achieve the degradation simulation of the full life cycle. The steady-state twin model is used for knowledge transfer to establish a gas path performance degradation model. After the fault is implanted, the degradation data of the gas turbine model can be obtained directly, it is used to reproduce the latent law in the process of gas path performance degradation, and to expand the attributes and quantity of the fault data, which can obtain data that is difficult to obtain during the experiment. In the inverse problem, various types of data are twinned with the model, and the model is identified to obtain the factors that change in the process of gas path performance degradation.

Take the split-shaft gas turbine as an example. Its main components include compressor, combustion chamber, high-pressure turbine and power turbine. Its structure is shown in Fig. 2.

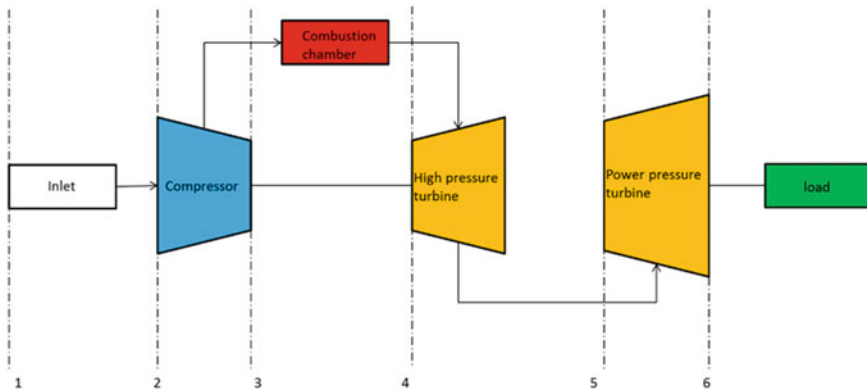


Fig. 2 Split-shaft gas turbine structure

2.1 Nonlinear Steady-State Twin Model of Gas Turbine

The steady-state simulation is to find the steady-state common operating point, so it is necessary to establish the common operating equation, that is, the balance equation of the gas turbine. When the balance condition is met, the gas turbine can be considered to be in a steady state. At this time, the various parameters of the gas turbine can be considered as the state parameters under the operating conditions. The steady-state twinning model described in this paper includes four equilibrium equations:

1. Balance equation between compressor outlet flow, fuel flow and high-pressure turbine inlet flow G_{tin}

$$G_{tin} - G_{cout} - G_f = 0 \quad (1)$$

where G_{tin} is the high-pressure turbine outlet flow, G_f is the fuel flow.

2. Balance equation between the high-pressure turbine outlet flow rate and the power turbine inlet flow rate

$$G_{ptin} - G_{tout} = 0 \quad (2)$$

where G_{ptin} is the inlet flow of the power turbine, and G_{tout} is the outlet flow of the high-pressure turbine.

3. Power balance equation of compressor and high-pressure turbine

$$\eta_{mt} W_t - W_c = 0 \quad (3)$$

where η_{mt} is the mechanical efficiency of the gas generator shaft, W_c is the power consumption of the compressor, W_t is the work done by the high-pressure turbine.

4. Power balance equation of power turbine load shaft

$$\eta_{mpt} W_{pt} - W_e = 0 \quad (4)$$

where η_{mpt} is the mechanical efficiency of the power turbine shaft, W_e is the load power, and W_{pt} is the work done by the power turbine.

error is used to describe the deviation of each balance equation:

$$\begin{cases} error_1 = (G_{cout} + G_f)/G_{tin} - 1 \\ error_2 = G_{ptin}/G_{tout} - 1 \\ error_3 = \eta_{mt} W_t/W_c - 1 \\ error_4 = \eta_{mpt} W_{pt}/W_e - 1 \end{cases} \quad (5)$$

$$error = error_1^2 + error_2^2 + error_3^2 + error_4^2 \quad (6)$$

When the sum of squares of each deviation reaches the minimum, it can be considered that the iterative calculation is over at this time, and the twin model has reached the equilibrium state. The specific algorithm will be given in the Sect. 2.3.

2.2 Component Characteristics Correction Method

The component characteristics in the thermodynamic model have an important influence on the accuracy of the model. However, the current mechanism modeling of gas turbines faces the difficulty of obtaining component characteristic lines, especially the compressor characteristic lines. Therefore, it is necessary to find a way to accurately establish a gas turbine component model even when there is only a general characteristic line. Literature [12] gives a characteristic curve correction method based on particle swarm algorithm. After the compressor characteristic diagram is organized into a general relative reduced form, its flow and efficiency characteristic diagrams are graphs showing the distribution of different relative reduced speed lines. The correction coefficient of the compressor characteristic line is defined as follows:

$$\begin{aligned} S_g &= G_{c,cor}^* / G_{c,cor} \\ S_\eta &= \eta_c^* / \eta_c \end{aligned} \quad (7)$$

where * represents the modified characteristic parameters.

The commonly used expression of correction coefficient is in the form of quadratic function, which is defined in this paper as the form of quadratic function with rotation speed as the independent variable. As follows:

$$S = a + b(1 - n_{cor,rel}) + c(1 - n_{cor,rel})^2 \quad (8)$$

where $n_{cor,rel}$ is the relative corrected speed, and a, b, c is the parameter to be identified.

The corrected characteristic parameters of gas turbine components are

$$\begin{aligned} G_{c,cor}^* &= S_g G_{c,cor} = S_g f_1(n_{cor}/n_{cor,0}, \pi_c/\pi_0) \\ \eta_c^* &= S_\eta \eta_c = S_\eta f_2(n_{cor}/n_{cor,0}, \pi_c/\pi_0) \end{aligned} \quad (9)$$

Therefore, the digital twin modelling becomes a parameter identification problem, which mainly affects the overall deviation of the characteristic line, and mainly affects the shape correction of the characteristic line. In the paper, the parameters that need to be identified and the parameters that need to be solved in the gas turbine model are regarded as variables of the optimization function.

Obtain component characteristic data from a gas turbine characteristic diagram, select 9 speed lines with equivalent relative speeds of 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1 respectively, and select a number of relative pressure ratio points for each

line. After sorting, it is used as the original interpolation table, and then the correction coefficient is introduced, and the final result is used as the final characteristic parameter of the compressor.

2.3 Differential Evolution Optimization Algorithm

Differential Evolution (DE) was proposed by Storn and Price in 1995. It was originally used to solve the Chebyshev polynomial problem. The principle of DE is very similar to that of a genetic algorithm, but it uses real number coding, and the evolution process is the same as the genetic algorithm, including mutation, crossover and selection. The selection strategy in DE algorithm is usually tournament selection, and the crossover operation method is basically the same as the genetic algorithm, but the difference strategy is adopted in the mutation operation, that is, the difference vector between individuals in the population is used to perturb individuals to achieve individual mutation. The mutation method of DE effectively utilizes the distribution characteristics of the population, improves the search ability of the algorithm, and avoids the shortcomings of the mutation method in the genetic algorithm [13]. Compared with other evolutionary algorithms, the DE algorithm has the following advantages: (1) It exhibits strong robustness in non-convex, multimodal, nonlinear, and continuous non-differentiable function optimization problems. (2) The convergence speed is faster. (3) Good at solving multi-variable function optimization problems. (4) Simple operation and easy implementation. The specific process is shown in Fig. 3. The specific optimization process is as follows:

1. Initialization

Enter the corresponding parameters, population number M , crossover factor $CR \in [0, 1]$, mutation factor $F \in [0, 2]$, DE algorithm generates M real-valued parameter vector with dimension D in each generation as the population, and the individuals in the population are expressed as x_i^t ($i = 1, 2, \dots, M$, t representing the number of evolution), usually with uniform probability. The random function of the distribution generates an initial population randomly from a given boundary.

2. Variation

The vector and of two different individuals $x_{r_2}^t$ and $x_{r_3}^t$ in the population are subtracted, and the resulting difference vector is weighted and added to another randomly selected individual vector $x_{r_1}^t$ to generate a mutation vector. The mutation vector generated for each target vector is

$$v_i^{t+1} = x_{r_1}^t + F(x_{r_2}^t - x_{r_3}^t) \quad (10)$$

where r_1, r_2, r_3 represents three different individuals in the population. The randomly selected individuals r_1, r_2, r_3 and the current target vector individuals i are different from each other, so the population size $M \geq 4$. F is called the

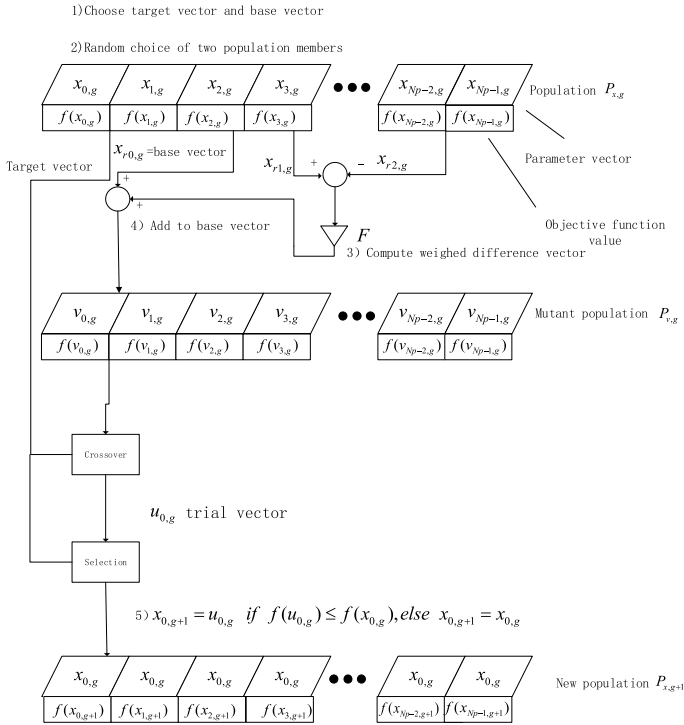


Fig. 3 DE algorithm flow chart

variation factor, and its value range is generally [0, 2], and its function is to control the amplification ratio of the deviation variable in the variation species.

3. Crossover

The variation vector v_i^{t+1} and the target vector individual x_i^t are subjected to a parameter mixing operation, that is, mutation, to generate a test vector u_i^{t+1} . The crossover operation can increase the diversity of the vector. Random selection is used to ensure that at least one parameter u_i^{t+1} is obtained from v_i^{t+1} , so as to ensure that the target individual x_i^t is evolved. The test vector is

$$u_i^{t+1} = (u_{i1}^{t+1}, u_{i2}^{t+1}, \dots, u_{iD}^{t+1}) \quad (11)$$

$$u_{ij}^{t+1} = \begin{cases} v_{ij}^{t+1}, & rand(j) \leq CR \text{ or } j = randn(i) \\ x_{ij}^{t+1}, & rand(j) > CR \text{ and } j \neq randn(i) \end{cases} \quad (12)$$

where $rand(j) \in [0, 1]$ represents the random number that generates the j -th variable, CR is the cross factor, the value range is [0, 1], $randn(i) \in [1, 2, \dots, D]$ is the dimension index obtained by random selection, and the function is to ensure that the test variable u_i^{t+1} obtains at least one parameter

from the variation vector v_i^{t+1} . The larger CR is, the more v_i^{t+1} contributes to u_i^{t+1} , which promotes the diversity of the population and facilitates the global search. In actual use, the value is determined according to multiple experiments.

4. Selection

The DE algorithm compares u_i^{t+1} with x_i^t according to a greedy search strategy. If u_i^{t+1} has better fitness than x_i^t , it will enter the next generation. This operation is called selection. The new population generated by this random disturbance can ensure that the objective function is obtained. Good convergence. Otherwise, x_i^t will enter the next generation. If the objective function is to be minimized, the selection process is

$$x_i^{t+1} = \begin{cases} u_i^{t+1}, & f(u_i^{t+1}) < f(x_i^t) \\ x_i^t, & f(u_i^{t+1}) \geq f(x_i^t) \end{cases} \quad (13)$$

The population number M is selected as 50, the crossover factor CR is 0.4, the variation factor F is set to 0.5, and the objective function is selected as the sum of the squares of each deviation, namely *error*, for optimization calculation.

3 Forward Problem Study: Gas Turbine Performance Simulation

The simulation model in this paper is established under the NI/LabVIEW environment.

3.1 Construction of the Steady-State Twin Model

The steady-state twin model is divided into two parts, the first part is the steady-state mechanism model of the split-shaft gas turbine, and the second part is the optimization model based on the DE optimization algorithm.

We introduce the steady-state mechanism model of the gas turbine and establish component-level models of the compressor, combustion chamber, high-pressure turbine, power turbine, and balance equation, and the input and output of each module in turn connected.

The model needs to call the VI of the mechanism model of the gas turbine, and set the initial parameters, boundary values, population size, scaling factor (variation factor), crossover probability (crossover factor), and the maximum number of cycles and other parameters.

3.2 Gas Turbine Performance Simulation Process

First, select the variables that need to be solved for the steady-state model: compressor pressure ratio π_c , high-pressure turbine expansion ratio π_t , and power turbine expansion ratio π_{pt} . Since this paper selects 1.0 operating condition as an example, the correction coefficients are zero for the first and second terms, therefore, only the constant term of the correction coefficient is retained, so the compressor and turbine characteristic line correction coefficient $a_1, a_2, a_3, a_4, a_5, a_6$ is selected and used as an array for iterative calculation. These variables need to be given the guess value and upper and lower boundaries based on experience. The output of the model is the objective function *error*. Boundary conditions and speed and other parameters can also be input into the model as variables, so as to develop a digital twin software for gas path performance simulation of split-shaft gas turbines to meet the simulation process under different working conditions and different environmental conditions. Provide a basis for subsequent quantification of performance degradation characteristics and fault diagnosis.

3.3 Simulation Results

The simulation results obtained according to the above method process are shown in Table 1.

It can be seen from the above table that the error between the simulation value and the design value is within 1.39%, and the model has a certain degree of credibility.

At this time, the optimized variables are shown in Table 2.

Table 1 Steady-state simulation results under 1.0 working condition

	Design value <i>T/K</i>	Simulation value <i>T/K</i>	Error/%	Design value <i>P/Pa</i>	Simulation value <i>P/Pa</i>	Error/%
3	702.77	703.17	0.057	1,743,170	1,734,408	0.503
4	1399	1404	0.358	1,668,213	1,659,825	0.503
5	1049.4	1041.0	0.800	396,175	401,679	1.390
6	777.2	779.9	0.347	101,255	101,255	0

3 represents the compressor exit cross-section parameter, 4 represents the combustor exit cross-sectional parameter, 5 represents the high-pressure turbine exit cross-sectional parameter, and 6 represents the power turbine exit cross-sectional parameter

Table 2 Variable results after optimization

π_c	π_t	π_{pt}	a_1	a_2	a_3	a_4	a_5	a_6
17.33	4.11	3.69	1.05	0.98	1.00	0.98	1.00	1.05

After the model is established, in addition to simulating the state of the gas turbine under different operating conditions and environments, it can also simulate the performance degradation of the gas turbine by implanting fault factors.

3.4 Gas Turbine Performance Degradation Model

On the basis of the gas turbine steady-state model established above, model knowledge transfer is carried out, and fault factors are implanted to simulate the state of the gas turbine after performance degradation. The fault factors are shown in Table 3.

The parameters and correction coefficients of the steady-state model have been solved in the previous article, and the fault model of the gas turbine is established based on this. The failure factor K is implanted in the performance parameters of the gas turbine, for example, K_1 represents a 5% decrease in the compressor flow rate. The measurement parameters and performance parameters of each section obtained in this way are the parameters in the fault state. Take the compressor flow rate drop by 5% as an example. Since the flow rate drop has a low impact on the pressure, only the outlet temperature of the section is given, as shown in Table 4.

It can be seen that when the compressor flow rate decreases, the outlet temperature of each component changes. Therefore, the exhaust temperature can be used as a criterion for performance degradation, but a reasonable characterization range needs to be set for each exhaust temperature, which is more complicated. Here, the failure factors introduced in the previous article are selected as quantitative features for diagnosis. Therefore, it is necessary to establish a diagnosis model on the basis of the mechanism model and the failure model.

Table 3 Gas turbine failure factor implantation

	G_C	G_T	G_{PT}	η_C	η_T	η_{PT}
K_1	-5%	0	0	0	0	0
K_2	0	-5%	0	0	0	0
K_3	0	0	-5%	0	0	0
K_4	0	0	0	-5%	0	0
K_5	0	0	0	0	-5%	0
K_6	0	0	0	0	0	-5%

Table 4 Outlet temperature of each section under fault condition

T'_3/K	T'_4/K	T'_5/K	T'_6/K
697.79	1448.00	1074.22	797.86

4 Inverse Problem Study: Gas Path Diagnosis

4.1 Gas Turbine Diagnostic Model

The difference between the diagnosis model and the mechanism model is that the parameters that need to be optimized in the mechanism model are all given in the diagnosis model. Because the system is in a balanced state under the given optimized parameters, the objective function needs to be reset. The sum of the squares of the temperature difference between the outlets of each component in the healthy state and the fault state is selected as the objective function, which is defined as

$$E = (T_3 - T'_3)^2 + (T_4 - T'_4)^2 + (T_5 - T'_5)^2 + (T_6 - T'_6)^2 \quad (14)$$

where E represents the objective function, and the superscript $'$ represents the parameters in the fault state.

The objective function is optimized with the fault factor as the optimization variable. When the objective function is the smallest, it indicates that the normal state and the fault state have been matched. At this time, observe the change of the fault factor K . Since the optimized result has slight fluctuations, it is performed five times in total.

4.2 Gas Turbine Diagnostic Results

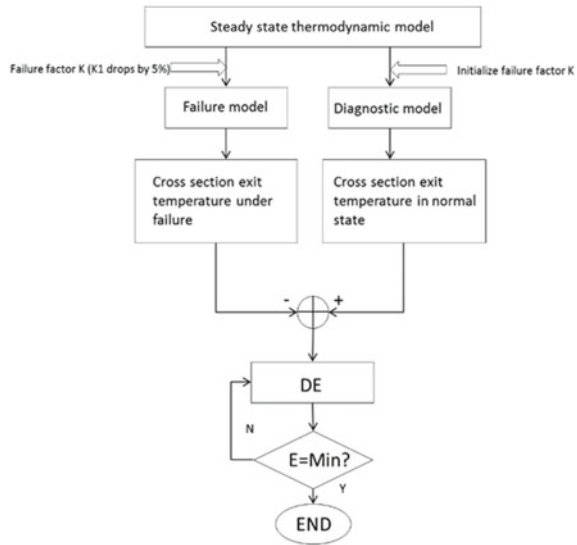
The diagnosis process is shown in Fig. 4. Simulation experiment to observe the results, as shown in Fig. 5.

It can be seen from the above figure that the failure factor is all 1 under normal conditions. When the compressor flow rate drops, the failure factor K_1 is reduced by 5% in the five experiments, while the rest of the failure factors fluctuate within an acceptable range, which is in line with expectations. Match. Therefore, it can be considered that it is reasonable to quantify the performance degradation characteristics by the failure factor.

5 Conclusion

1. This article first established a steady-state model of the gas path performance of a split-shaft gas turbine in the Labview environment, introduced correction coefficients to correct and fit the general characteristic line, and solved the unknown variables using the differential evolution algorithm to establish a gas path for the gas turbine. Compared with the experimental value of the performance digital

Fig. 4 Twin model reverse diagnosis process



twin model, the simulation result has a maximum error of 1.39%, which verifies the accuracy of the model.

2. Implant fault factors, establish the fault model of the split-shaft gas turbine, and obtain the parameters of each section in the fault state, so as to establish the diagnosis model of the gas turbine. Taking the compressor flow rate drop by 5% as an example, five simulation experiments have been carried out. The results show that under this failure, only K_1 dropped by 5%, and the rest of the failure factors are floating within an acceptable range. Therefore, the change in the failure factor is used to quantify the gas turbine the performance degradation characteristics are reasonable.

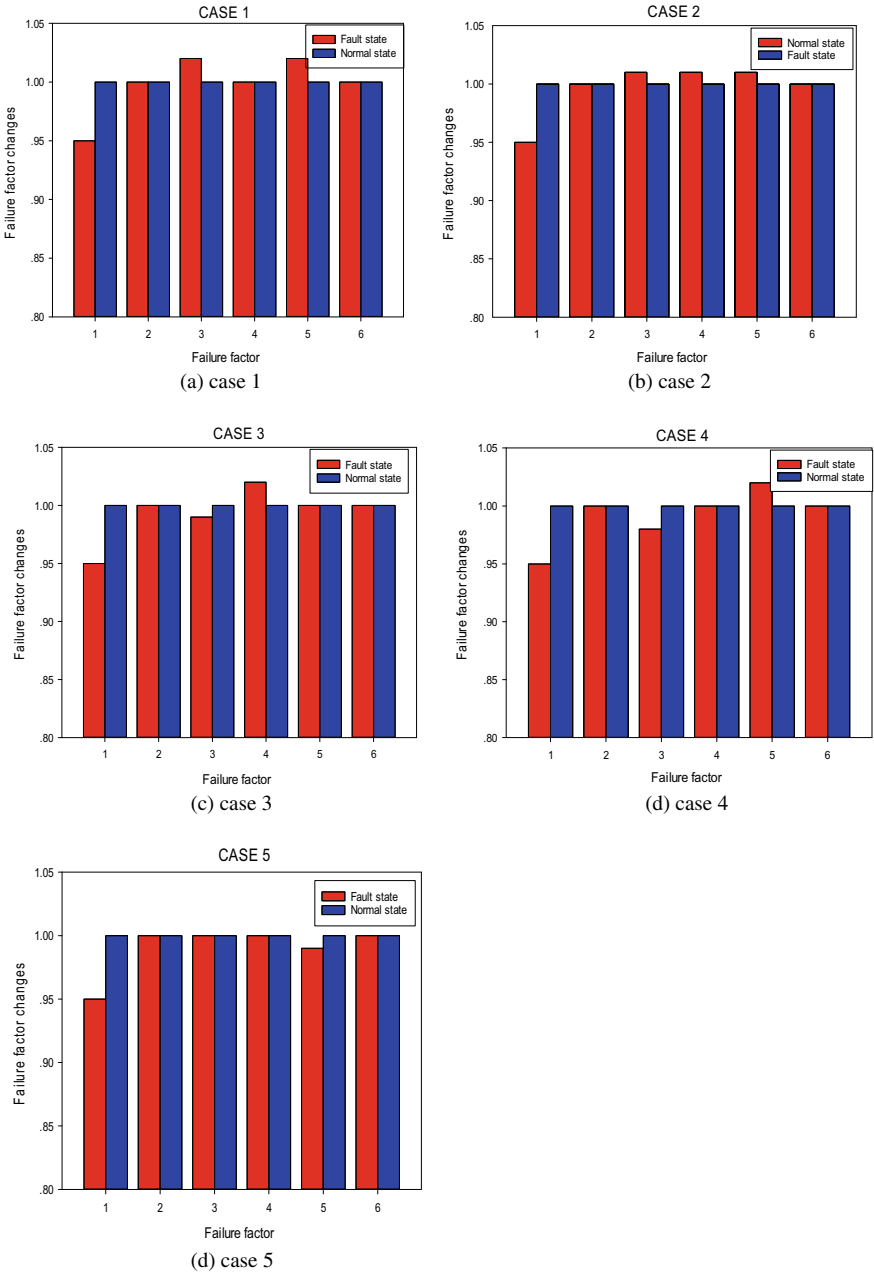


Fig. 5 Gas path fault diagnosis results

References

1. Gas turbine plant thermal performance degradation assessment. ASME Turbo Expo.: ASME, pp. 43–51 (2008)
2. Thirunavukarasu, E., Fang, R., Khan, J.A., et al.: Modeling and simulation of gas turbine system on a virtual test bed (VTB). In: International Mechanical Engineering Congress and Exposition: ASME, pp. 337–346 (2012)
3. Li, P., Degobert, B., Francois, B., et al.: Modeling and control of a gas micro turbine generator by using a causal ordering graph. In: 2006 The Proceedings of the Multiconference on Computational Engineering in Systems Applications, Beijing, China, pp. 271–277 (2006)
4. Hosseinalipour, S.M., Abdolahi, E., Razaghi, M.: Static and dynamic mathematical modeling of a micro gas turbine. *J. Mech.* **29**(2), 327–335 (2013)
5. Xu, M., et al.: An improved hybrid modeling method based on extreme learning machine for gas turbine engine. *Aerosp. Sci. Technol.* **107**, 106333 (2020)
6. Liu, Z., Karimi, I.A.: Gas turbine performance prediction via machine learning. *Energy* **192**(Feb 1), 116627.1–116627.10 (2020)
7. Chaobing, H., Hongluo, Z., Weidong, H., et al.: Performance degradation analysis of a gas turbine after offline water washing of the compressor. *J. Chin. Soc. Power Eng.* (2019)
8. Tülin, Y.M., Bülent, K.: Aircraft gas turbine engine health monitoring system by real flight data. *Int. J. Aerosp. Eng.* **2018**, 1–12 (2018)
9. Shafto, M., Conroy, M., Doyle, R., et al.: Modeling, simulation, information technology and processing roadmap. NASA (2010)
10. Glaessgen, E.H., Stargel, D.S.: The digital twin paradigm for future NASA and U.S. air force vehicles: AIAA 20121-1818. AIAA (2012)
11. Grieves, M., Vickers, J.: Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen, F.J., Flumerfelt, S., Alves, A. (eds.) *Transdisciplinary Perspectives on Complex Systems*, pp. 85–113 (2017)
12. Changduk, K., Jayoung, K.: Components map generation of gas turbine engine using genetic algorithms and engine performance deck data. *J. Eng. Gas Turbines Power* **129**(2), 312–317 (2007)
13. Storn, R., Price, K.: Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces. *J. Global Optim.* **23**(1) (1995)

Small Sample MKFCNN-LSTM Transfer Learning Fault Diagnosis Method



Yonglun Guo, Guoxin Wu, and Xiuli Liu

Abstract Aiming at the problem that there are all kinds of noise interference in the planetary gearbox of wind turbine in the general experimental scene, the vibration data obtained is less and the fault characteristics are not obvious. A MKFCNN-LSTM migration learning algorithm based on multi-kernel fusion convolution neural network (MKFCNN) and long and short time memory neural network (LSTM) is proposed to realize the fault diagnosis of wind turbine planetary gearbox. Firstly, the MKFCNN is constructed to extract the multi-scale spatial features of the sample signal, and then it is connected in series with LSTM to extract the corresponding time information of the sample signal. In view of the associated fault feature information between the rolling bearing data set and the planetary gearbox data set, the rolling bearing vibration signal of the Western Reserve University is input into the MKFCNN-LSTM as the source domain sample data, and iterative training is used to update the network weight and offset value. The pre-trained MKFCNN-LSTM is obtained, and then fine-tuned by inputting the vibration data of the planetary gearbox with small samples in the target domain, the weight and offset values are transferred from the source domain to the target domain, and finally the accuracy of fault recognition based on the number of small samples in the target domain is improved. The experimental results show that the proposed method can apply the original fault diagnosis knowledge to the vibration data set of the planetary gearbox in the laboratory. Compared with stack autoencoders (SAE), support vector machine (SVM) algorithm, the accuracy of fault identification and classification is improved to a certain extent.

Keywords MKFCNN · LSTM · Transfer learning · Rolling bearing · Planetary gearbox · Fault diagnosis

Y. Guo · G. Wu (✉) · X. Liu

The Ministry of Education Key Laboratory of Modern Measurement and Control Technology, Beijing Information Science & Technology University, Beijing 100192, China
e-mail: wgx1977@bistu.edu.cn

1 Preface

Planetary gearbox is widely used in wind turbines, aviation equipment and other large and complex mechanical and electrical equipment. It works in a bad environment and is prone to all kinds of failures. Under the application of various load forces, the key components are prone to problems [1]. Therefore, fault monitoring and diagnosis can effectively find and solve the problems in the process of operation.

With the rapid development of science and technology, deep learning with its complex nonlinear network structure and deep feature extraction ability, more and more experts and scholars apply it to the field of mechanical and electrical equipment fault diagnosis. Reference [2] combines unsupervised pre-trained deep wavelet convolutional self-encoder with supervised trained LSTM network to identify various fault types and fault degrees of bearings. In reference [3], the fault diagnosis of gear transmission chain is realized by using deep belief network (DBN). The reference [4, 5] converts the one-dimensional vibration signal of the planetary gearbox into a two-dimensional characteristic graph and inputs it into the convolution neural network, which achieves a better correct rate of fault diagnosis.

Transfer learning refers to applying the features or knowledge learned from one source domain to other target domains and solving problems in this field, in which the features that have been learned by the algorithm are called source domains, and the areas that need to be solved are called target domains. Transfer learning can associate the source domain sample features with the target domain sample features, learn cross-domain knowledge with the help of existing knowledge and experience, and further improve the generalization ability of the model. The transfer learning method has been put into practice in the field of fault diagnosis in the literature [6, 7], and a good diagnosis result has been obtained.

The above methods have achieved good results. But in the actual scene, the problems are as follows: (1) Some need to extract fault features manually, and then transfer learning for fault diagnosis, and the credibility of the results is affected. (2) The deep network structure is relatively simple, in the scene of noise interference and insufficient data, it is difficult to extract comprehensive and effective essential features, and the final recognition results are also affected.

In view of the above problems, when the vibration data of planetary gearbox in the target domain contains the influence of noise, the feature is not obvious, and the amount of data is insufficient, by combining the advantages of multi-scale feature recognition extraction of MKFCNN network, LSTM's sensitivity to time series and the ability of transfer learning method to learn cross-domain knowledge, a planetary gearbox fault diagnosis method based on MKFCNN-LSTM transfer learning is proposed.

2 The Spatial Multi-scale Feature Learning of MKFCNN

MKFCNN network, inspired by Google Inception V1 network [8], is a multi-scale feature learning framework based on one-dimensional convolution neural network (1DCNN). It consists of convolution layer, pooling layer, and basic MKFCNN module. The basic MKFCNN module is shown in Fig. 1 and consists of four branches: Branch 1 is a convolution layer with a convolution core of 1×1 . Branches 2 and 3 are two series convolution layers. Branch 4 is a series of maximum pool layer and convolution layer with convolution core size of 1×1 . The MKFCNN network structure is shown in Fig. 2.

The detailed configuration information of the basic MKFCNN network module is shown in Table 1. The relevant parameters of the MKFCNN network are shown in Table 2.

The realization of spatial multi-scale feature extraction mainly depends on the basic convolution-pooling operation. Given the original vibration data X , and Eq. (1) shows:

$$X = \{x_1^0, x_2^1, \dots, x_N^t\} \tag{1}$$

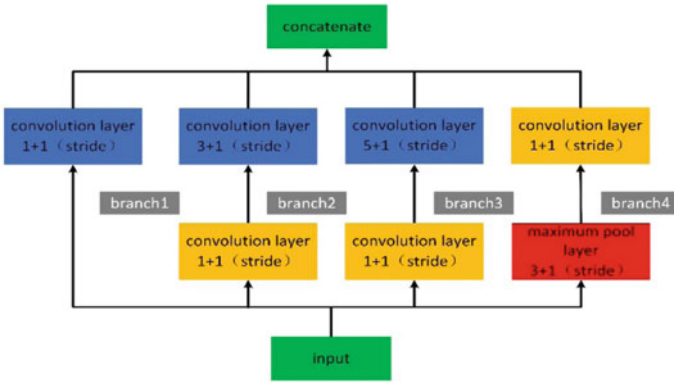


Fig. 1 Basic MKFCNN module structure

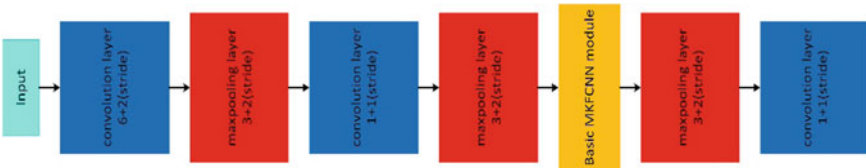


Fig. 2 MKFCNN network structure

Table 1 Basic MKFCNN network module parameters

Branches	Number of layers	Convolution kernel numbers	Nuclear size	Strides
Branch 1	1	32	1	1
Branch 2	1	16	1	1
	2	32	3	1
Branch 3	1	32	1	1
	2	64	5	1
Branch 4	1	0	3	1
	2	32	1	1

Table 2 MKFCNN network configuration parameters

Number of layers	Network layer type	Convolution kernel numbers	Nuclear size	Stride
1	Convolution layer	16	6	2
2	Maximum pool layer	0	3	2
3	Convolution layer	32	1	1
4	Maximum pool layer	0	3	2
5	Basic MKFCNN network module			
6	Maximum pool layer	0	3	2
7	Convolution layer	32	1	1

x_i^t represents the position of the collected data under time t , and a total of N data points have been collected. Data X is selected as the input of the first convolution layer, in which the convolution kernel with window length m is used for local feature extraction, and the convolution operation is completed by moving the convolution kernel in the whole input area, thus the corresponding feature graph is obtained. For the calculation of the z_i output of i nodes in the feature graph, and Eq. (2) shows:

$$z_i = \sigma(\mathbf{w}^T y_{i:i+m-1} + \mathbf{b}) \tag{2}$$

\mathbf{w} represents the convolution kernel vector, \mathbf{b} represents the offset value, $y_{i:i+m-1}$ represents the intercepted data segment with length m from the i point of the input data, and $\sigma(\cdot)$ is the nonlinear activation function. Here the ReLU activation function is selected to prevent the gradient from disappearing and accelerate the convergence process of the model.

The output scalar z_i in Eq. (2) can be regarded as the activation output of the convolution check signal. By constantly moving the convolution kernel on the input data X , and the characteristic graph output of the j convolution kernel is shown in Eq. (3):

$$\mathbf{Z}_j = [z_1, z_2, \dots, z_{L-m+1}] \tag{3}$$

After that, the pooling layer further processes the feature map of the convolution layer output in order to extract the important and location-invariant features. The local maximum of the input feature map is calculated by using the maximum pooling operation with a pooling length of p , and the k -th pooling feature is shown in Eqs. (4), (5):

$$h_k = [h_1, h_2, \dots, h_{\frac{L-m}{p}+1}] \tag{4}$$

$$h_j = \max_{(j-1)p+1 \leq i \leq jp} \{z_i\} \tag{5}$$

3 Time Feature Learning of LSTM

The fault of planetary gearbox of wind turbine is often caused by some long-term slow and small changes. Therefore, LSTM is used to further mine the time relevance of vibration data and enrich fault features.

The LSTM network includes a memory unit c_t , and a hidden unit h_t , as shown in Fig. 3.

Among them, c_t is mainly controlled by three gates, including input gate i_t , output gate o_t , and amnesia gate f_t . Iterative update equations are shown in Eqs. (6), (7), (8), (9), (10) and (11).

$$i_t = \text{sigmoid}(U_i h_{t-1} + W_i x_t + b_i) \tag{6}$$

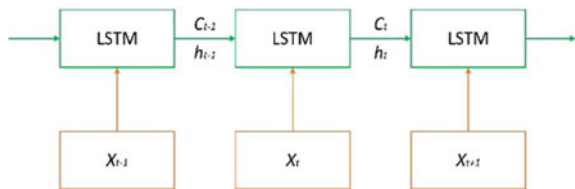
$$f_t = \text{sigmoid}(U_f h_{t-1} + W_f x_t + b_f) \tag{7}$$

$$o_t = \text{sigmoid}(U_o h_{t-1} + W_o x_t + b_o) \tag{8}$$

$$\tilde{c}_t = \tanh(U_c h_{t-1} + W_c x_t + b_c) \tag{9}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{10}$$

Fig. 3 LSTM framework



$$h_t = o_t \odot \tanh(c_t) \tag{11}$$

In the equations: U, W, b are the parameters of network iterative update learning, and the operator \odot represents the product operation between variables.

The structure principle of the LSTM unit is shown in Fig. 4. First, the forgetting gate f_t obtains the new input x_t and the previous hidden state h_{t-1} . When the value of the forgetting gate is close to 1, the information of the last memory unit c_{t-1} will be retained. Secondly, the new input information and the previous hidden state information are merged into the input gate, and the input gate updates the information and generates a new memory unit c_t . Finally, Output determines that part of the information in the memory unit forms a new hidden state h_t , and part of the information is output as a model.

Combining the MKFCNN network with the LSTM network, the combined MKFCNN-LSTM network is shown in Fig. 5.

Fig. 4 The structure principle of LSTM unit

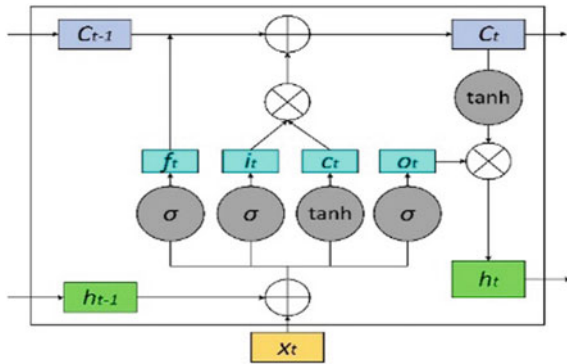
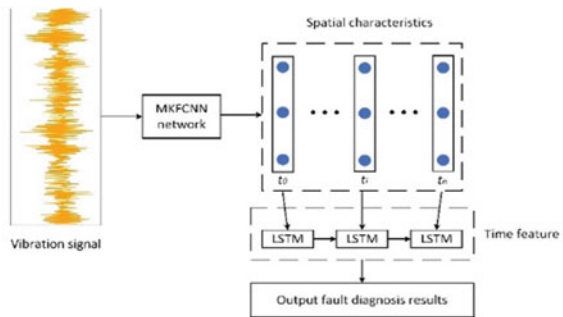


Fig. 5 Network structure of MKFCNN-LSTM



4 Fault Diagnosis Process of Planetary Gearbox Based on MKFCNN-LSTM Transfer Learning

The fault diagnosis flow chart of MKFCNN-LSTM transfer learning is shown in Fig. 6. The specific process is as follows:

Step 1 Get the sample and preprocess it. The vibration data sets of five fault types of rolling bearings in Western Reserve University are collected as the source domain sample data set, and the self-collected planetary gearbox vibration data set as the target domain sample data set. The above two data sets are divided into training set and test set respectively. Add noise with a signal-to-noise ratio of 10 to the source domain data.

Step 2 The MKFCNN-LSTM network is pre-trained under the source domain sample data. Initialize the number of layers, the number of neurons, weights and other parameters of the network, input the source domain sample data for iterative training, and finally get the MKFCNN-LSTM pre-training network.

Step 3 Fine-tuning MKFCNN-LSTM Network under small sample data in Target Domain. Freeze part of the neural network layer of the network, input small sample data to fine-tune the weights and bias values of the rest of the network layer, update the

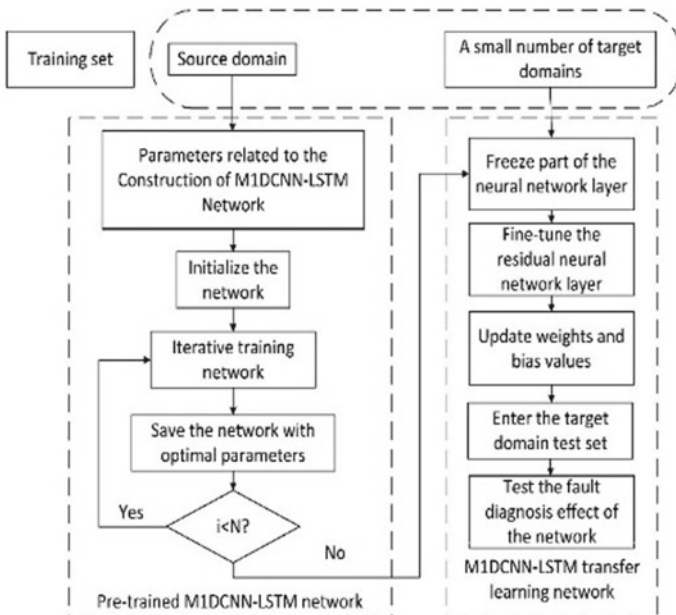
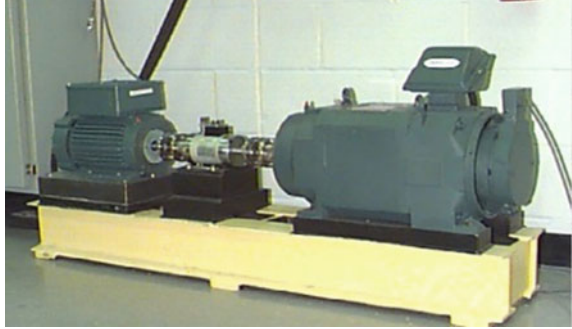


Fig. 6 Fault diagnosis process of MKFCNN-LSTM transfer learning

Fig. 7 Rolling bearing test-bed of Western Reserve University



network through the back propagation algorithm, and finally obtain the MKFCNN-LSTM migration learning model suitable for small sample data in the diagnostic target domain.

5 Experimental Verification Analysis

5.1 Obtain Rolling Bearing Data of Source Western Reserve University

The rolling bearing vibration data of the Western Reserve University are obtained as the source domain sample data, and the experimental platform for collecting the experimental data is shown in Fig. 7. The sampling frequency is 1797 rpm, the motor load is 0 HP, and the defect size is 0.007 in., 0.014 in., 0.021 in. and 0.028 in. respectively. Each defect degree corresponds to a fault type, which is inner ring damage, rolling body damage, outer ring 6-point direction damage, outer ring 3-point direction damage, and normal vibration data. The time domain diagram of vibration signal is shown in Fig. 8.

There are 1420 samples for each type of fault, and each sample contains 512 data points. The vibration time domain diagram is shown in Fig. 8, and the basic data is shown in Table 3.

5.2 Obtaining Vibration Data of Planetary Gearbox in Target Area

The feasibility and effectiveness of the proposed method are verified by collecting the vibration data of the planetary gearbox. The experimental equipment for fault diagnosis research is shown in Fig. 9, which is composed of torque meter, motor,

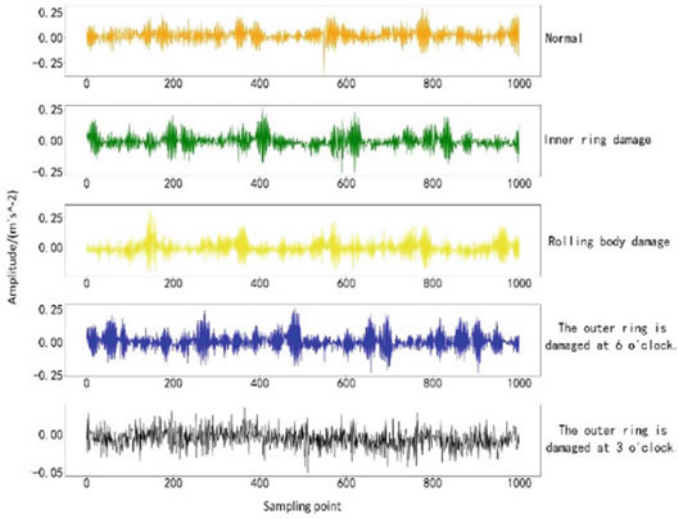


Fig. 8 Time domain map of vibration data of Western Reserve University

Table 3 Fault types of rolling bearings

Rolling bearing state	Damage degree/in.	Number of samples	Labels
Normal	0	1420	1
Inner ring damage	0.028	1420	2
Rolling body damage	0.021	1420	3
The outer ring is damaged at 6 o'clock	0.014	1420	4
The outer ring is damaged at 3 o'clock	0.007	1420	5

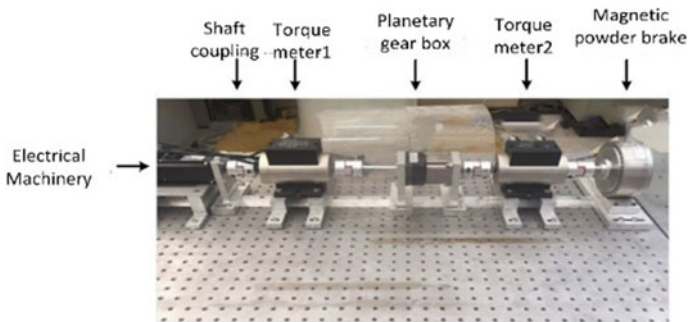


Fig. 9 Planetary gearbox test bench

Table 4 Basic parameters of planetary gearbox

Parameter name	Numerical value
Number of solar gear teeth	24
Number of planetary gear teeth	24
Number of inner ring teeth	72
Module of planetary gear train/mm	0.6
Pressure angle/(°)	20
Number of bearing rollers	8

magnetic powder brake, planetary gear box, etc., in which the axial load is output by magnetic powder brake, and the radial load on the input shaft and output shaft of planetary gearbox is output by torque meter. The basic parameters of the planetary gearbox are shown in Table 4.

The artificial machining damage of planetary gear and rolling bearing in planetary gearbox includes: tooth surface wear, tooth root crack, planetary gear tooth fracture and tooth root crack, planetary gear tooth fracture and rolling body missing. The sampling frequency is 20.48 kHz, the motor speed is 1980 rpm, and 327,675 vibration data points are collected in each group. Figure 10 shows the time domain diagram of the vibration signal. The size of the sliding partition window is 512 data points, the moving step is 256 data points, the number of samples for each type of fault is 1278, and the data length of each sample is 512. The basic data is shown in Table 5.

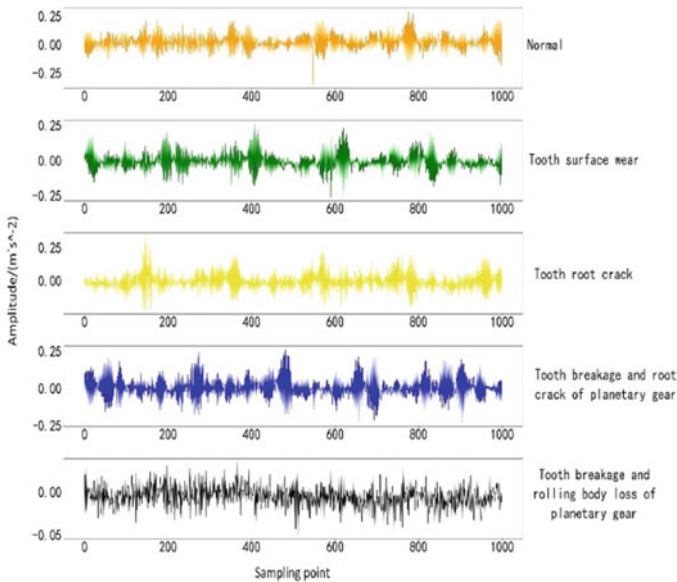
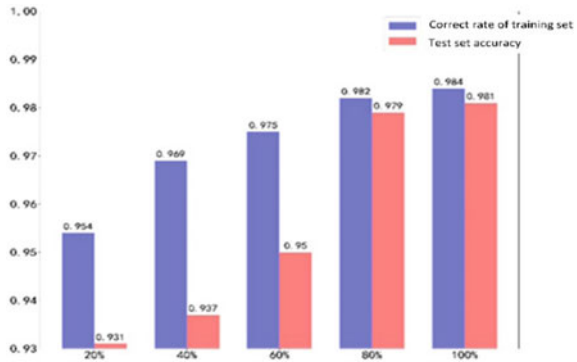


Fig. 10 Time domain diagram of vibration signal

Table 5 Fault type of planetary gearbox

Planetary gearbox status	Number of samples	Labels
Normal	1278	1
Tooth surface wear	1278	2
Tooth root crack	1278	3
Tooth breakage and root crack of planetary gear	1278	4
Tooth breakage and rolling body loss of planetary gear	1278	5

Fig. 11 The relationship between the number of small samples and the performance of model



5.3 Fault Diagnosis Analysis of Planetary Gearbox in Small Sample Scene

By inputting 20%, 40%, 60%, 80% and all small samples of vibration data into MKFCNN-LSTM transfer learning algorithm, the proportion of training set and test set is 4–5 and 1–5, respectively, to verify the influence of different small sample data sets on the fault diagnosis accuracy of the model. The classification accuracy of the training set and the test set is the average of the last five times in each experiment, and the bar chart is shown in Fig. 11.

As can be seen from Fig. 11 with the increase of the amount of small sample data, the average fault identification accuracy of M1DCNN-LSTM model training set and test set increases gradually, until it reaches 80% of the small sample fault data, and the classification results tend to be stable, eliminating the problem of over-fitting.

5.4 Comparative Analysis of Diagnostic Methods

This method is compared with SAE network and SVM method in order to verify the effectiveness and superiority of this method. The method configuration in SAE network reference [9] shows that the number of neurons in the two hidden layers is

Table 6 Comparative analysis of the effect of multi-method fault diagnosis

Arithmetic	Correct rate of training/%	Verification accuracy/%
SAE	99.79	92.94
SVM	95.05	94.13
MKFCNN-LSTM transfer learning	98.43	98.25

256 and 128 respectively, and the number of iterations is 30. The SVM method is set up by reference [10]. The error penalty factor C is 0.4 and the kernel function radius γ is 0.2.

The basic experimental results are shown in Table 6, in which the correct rates of training set and test set of each method are the average of the last five training and testing rates of their respective methods, respectively.

From the data in the table, it can be concluded that the classification accuracy of the training set of the SAE network is higher than that of the test set, which leads to the problem of over-fitting. MKFCNN-LSTM transfer learning exceeds 97% in both training set diagnosis accuracy and test set diagnosis accuracy, which verifies that the model has no over-fitting or underfitting problems, and has a high fault diagnosis accuracy.

5.5 Visual Effect of All Kinds of Fault Diagnosis by Confusion Matrix

The confusion matrix can show the proportion of various fault recall rates in the planetary gearbox data set. The above different methods are applied to the vibration data test set of the planetary gearbox in the target domain, and the display is based on the obfuscation matrix. The horizontal axis represents the forecast category, the vertical axis represents the actual category, the diagonal position represents the recall rate, and the rest is the proportion of prediction errors, and the sum of each row is 1, as shown in Fig. 12.

It can be seen from the figures that the first two comparison algorithms do not classify the second, third and fourth types of faults clearly, and about 0.01–0.21% of them are mistakenly classified into other fault types and normal types, and the normal types are misjudged as fault types. The recall rate of all kinds of fault diagnosis of the proposed method is 95% or more, and the second, third and fourth types of faults will still be misjudged, but the proportion remains in the range of 0.01–0.05. Generally speaking, the effect of fault classification is good.

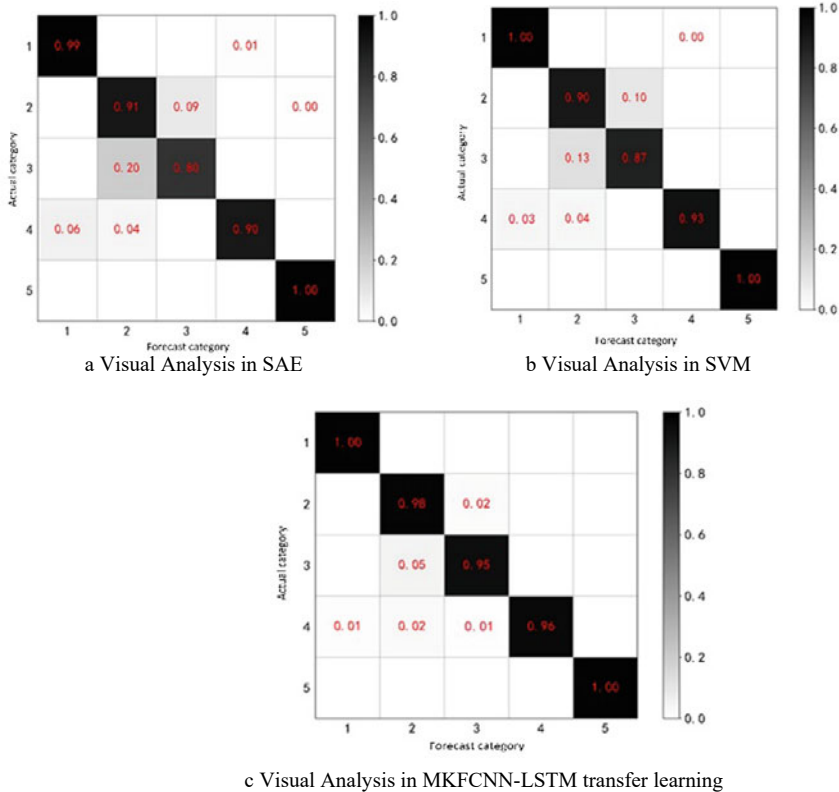


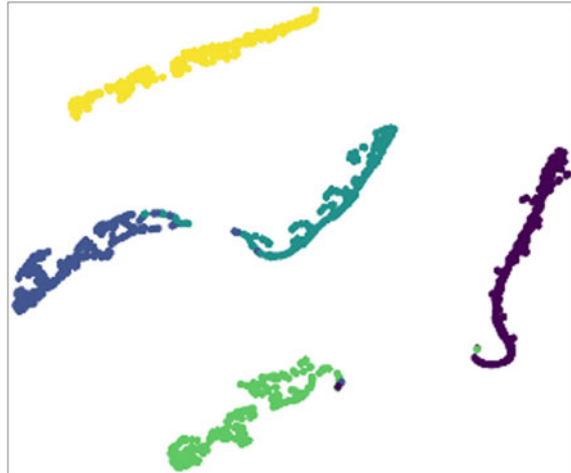
Fig. 12 Visual fault classification effect of confusion matrix

5.6 Visual Learning Results of Fault Features Based on T-SNE

T-SNE algorithm is a nonlinear dimensionality reduction algorithm, which mainly reduces the dimensionality of high-dimensional data to 2-D or 3-D, which is convenient for naked eye observation. The final result of fault identification and classification of M1DCNN-LSTM migration model is visualized as shown in Fig. 13. Each color in the diagram represents a class of fault categories. It can be clearly seen from the diagram that all kinds of faults gather together respectively, with a clear boundary and a small amount of overlap and mixin.

The above experimental results show that M1DCNN-LSTM transfer learning combines the multi-scale spatial domain feature extraction ability of M1DCNN and the strong time domain perception ability of LSTM, and carries out fault diagnosis and recognition based on the pre-training model, which can effectively extract the deep nonlinear and comprehensive features of fault signals.

Fig. 13 Visual learning results of fault features based on T-SNE



6 Conclusion

In this study, aiming at the diagnosis of slight fault, single fault and compound fault of planetary gearbox, a MKFCNN-LSTM migration learning algorithm is proposed. Through practical engineering practice, the superiority and effectiveness of the method are verified. The experimental results show that:

1. The proposed method combines the advantages of extracting fault signal features in spatial domain with the ability to obtain signal time domain information in time domain, and further improves the performance of mining comprehensive and deep fault features.
2. In the proposed method, the data distribution is used as a link to transfer different electromechanical equipment for fault feature learning, and the original fault diagnosis knowledge is applied to the fault identification of planetary gearbox by extracting different damage degrees and different types of fault features of rolling bearings of Western Reserve University in the source region.
3. Compared with SAE network, SVM method, the fault diagnosis accuracy of the proposed method is improved to a certain extent. How to use the proposed method to diagnose three or more compound faults of planetary gearbox is the direction of future research.

Acknowledgements This research is supported by Qin Xin Talents Cultivation Program, Beijing Information Science & Technology University.

References

1. Liu, R., Yang, B., Zio, E., Chen, X.: Artificial intelligence for fault diagnosis of rotating machinery: a review. *Mech. Syst. Signal Process.* **108** (2018)
2. Du, X., Chen, Z., Xu, X., et al.: Fault diagnosis of bearing based on wavelet convolutional auto-encoder and LSTM network. *J. Mech. Electr. Eng.* **36**(07), 663–668 (2019)
3. He, J., Yang, S., Gan, C.: Unsupervised fault diagnosis of a gear transmission chain using a deep belief network. *Multi. Digit. Publ. Inst.* **17**(7) (2017)
4. Jing, L., Zhao, M., Li, P., Xu, X.: A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* **111** (2017)
5. Chen, Z.Q., Li, C., Sanchez, R.V., Wang, D.: Gearbox fault identification and classification with convolutional neural networks. *Shock Vibr.* **2015** (2015)
6. Duan, L., Xie, J., Wang, K., et al.: Gearbox diagnosis based on auxiliary monitoring datasets of different working conditions. *J. Vibr. Shock* **36**(10), 104–108+116 (2017)
7. Chen, C., Shen, F., Yan, R.: Enhanced least squares support vector machine-based transfer learning strategy for bearing fault diagnosis. *Chin. J. Sci. Instrum.* **38**(01), 33–40 (2017)
8. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9 (2015)
9. Xie, Y., Wang, F., Fu, Z.: Mechanical fault diagnosis of power transformer based on stacked auto encoder. *High Voltage Appar.* **56**(09), 46–53+59 (2020)
10. Xue, Z., Zheng, X., Qiu, X., et al.: The online diagnosis method of marine induction motor bearing fault based on SVM. *Ship Ocean Eng.* **49**(05), 1–5+9 (2020)

Prediction of Sensor Values in Paper Pulp Industry Using Neural Networks



João Antunes Rodrigues, José Torres Farinha, António Marques Cardoso, Mateus Mendes, and Ricardo Mateus

Abstract The economic sustainability of any industry is directly linked to the management and efficiency of its physical assets. The maintenance of these assets is one of the key elements for the success of a company since it represents a relevant part of its Capital and Operational Expenses (CAPEX and OPEX). Due to the importance of maintenance, a lot of research has been done to improve the methodologies aiming to maximize physical assets' availability at the most rational costs. The introduction of Artificial Intelligence in the world of maintenance increased the quality of prediction on equipment failures, namely when associated to continuous equipment monitoring. This paper presents a case study where a neural network is proposed to predict the future values of various sensors installed on a paper pulp press. Data from the following variables is processed: electric current; pressure; temperature; torque; and speed.

J. A. Rodrigues (✉) · R. Mateus

Research Centre in Industrial Engineering, Management and Sustainability, Universidade Lusófona, Campo Grande 376, 1749-024 Lisboa, Portugal

e-mail: p5942@ulusofona.pt

R. Mateus

e-mail: p5768@ulusofona.pt

J. A. Rodrigues · A. M. Cardoso

CISE, University Beira Interior, 6201-001 Covilhã, Portugal

e-mail: ajmcardoso@ieee.org

J. T. Farinha · M. Mendes

Polytechnic of Coimbra—ISEC, Quinta da Nora, 3030-199 Coimbra, Portugal

e-mail: torres.farinha@dem.uc.pt

M. Mendes

e-mail: mmendes@isec.pt

J. T. Farinha

CEMMPRE, Coimbra University, DEM, Polo 2, 3030-290 Coimbra, Portugal

M. Mendes

University of Coimbra—ISR, Coimbra, Portugal

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_24

Keywords Predictive maintenance · Condition monitoring · Neural networks · Forecasting · Artificial intelligence

1 Introduction

Cost reductions along with regulations and concerns for safety and environmental impacts play a major role in the success of industries [1].

Industries have been looking for and investigating new techniques and equipment management tools that provide a competitive advantage in the quality or cost of their products, processes or services. Industrial equipment maintenance is therefore a key issue [1].

Kumar and the British Standards Institute (2015) define maintenance as the combination of all technical and administrative activities necessary to keep equipment, facilities and other physical assets in the desired operational condition or to restore them to comply with its function with quality [2].

The four main objectives of maintenance are: safety, quality, cost and availability.

There are many different approaches to maintenance. The predictive approach is one of the most important and effective, aiming to maximize the equipment's availability at a minimum cost.

This type of maintenance mainly involves predicting system failures, the main task of which is to detect the first signs of failure. Taking these signs into account, predictive maintenance aims to warn when faults are likely to occur and thus to suggest making scheduled stops on the asset. The advancement of technology and new techniques in the field of predictive maintenance have been making predictive maintenance more efficient, applicable and accessible to industries.

Predictive maintenance aims to predict the occurrence of failures before they happen, using data from consistent and constant monitoring of the conditions and operation of the target equipment.

Unwanted conditions, such as wear and tear of equipment components, are observed and/or predicted using forecasting algorithms to optimize when preventive interventions in the asset should be scheduled, thus avoiding breakdowns and reducing repair costs and production losses, and consequently increasing the asset availability.

Predictive maintenance has been adopted by various sectors in the manufacturing and service industries, in order to improve reliability, safety, availability, quality, as well as promoting environmental sustainability, since predictive maintenance reduces production surpluses as well as non-compliant products [3].

Quality measures how well an asset performs its function properly, while reliability measures how that asset maintains its original level of quality steady over time, under the various operating conditions to which it is exposed.

Predictive maintenance techniques are increasingly associated with sensor technologies because to make good predictive maintenance decisions it is necessary to have good quality information regarding the past and current operations of the asset.

Therefore, it is paramount to calibrate sensors properly and to process the resulting data in a reliable way [4].

The present paper focus on a predictive maintenance approach aiming to identify the current equipment state and predict its future operating conditions through collected past data processed using Artificial Intelligence (AI) algorithms.

Prediction is made for the next 90 days, a timeframe which allows the industry to adequately prepare and schedule maintenance interventions, thereby avoiding loss of production and optimizing stopping time. The company's competitiveness advantage is achieved by reducing maintenance downtime and increasing production time.

A multi-layer perceptron Artificial Neural Networks (ANN) model is presented for predicting future data from various variables. The algorithm presented in this paper is implemented in Python, using the supervised learning model MLP Regressor from Scikit-learn (Sklearn).

Sklearn is an open-source machine learning library for the Python programming language. It offers numerous functions for processing data, performing operations such as sorting, regression and grouping algorithms, including Support Vector Machines (SVM), Random Forest, gradient augmentation and k-means. It is designed to work with the numerical and scientific Python libraries NumPy and SciPy. Sklearn is written in Python and uses Numpy extensively for high-performance linear algebra and array operations.

2 Literature Review

Rodrigues et al. [5] use Feed Forward Neural Networks to classify the level of degradation of lubricants of Diesel engines. The results show that neural network models can classify oil conditions, achieving more than 90% precision compared to the performance of human experts, and thus allowing the process to be automated in the future [5].

It is important to have continuous and efficient maintenance in order to keep assets as available as possible and with no accidents. Allah Bukhsh et al. [6] developed predictive models that used existing data from a railway and produced interpretable results. Such predictive models were supported by classification techniques based on Decision Trees, Random Forest and Fault Trees with gradient increase; these tools allow to predict the need for maintenance on the railway [6].

Hongxiang et al. develop an algorithm using Artificial Neural Networks to analyze spectroscopy data from an oil. Results show that the mining of oil spectroscopy data by ANN methods can be used to classify types of lubricant and distinguish routine conditions of a Diesel engine from operating conditions [7].

Prediction in maintenance area support making better decisions. Okoh et al. [8] present an approach to determine when a system needs to undergo maintenance, repair, and overhaul, before a failure occurs. The novelty in this study is the development of the through-life performance approach [8].

One of the main maintenance challenges is to increase equipment availability. Makridis et al. [9] describe that predictive maintenance extends vessel lifetimes in the maritime sector, while reducing overall maintenance costs as well. The authors present a machine learning approach for detecting anomalies in the data collected through sensors installed on the vessels, hence predicting the condition of specific parts of the vessel's main engine [9].

Upgrading equipment and making it smarter is a common goal for many managers around the world. However, leveraging Artificial Intelligence methods from older equipment is sometimes extremely difficult, as these assets must be equipped with technical diagnostic tools and sensors for data collection. Vlasov et al. [10] discuss methods for maintaining industrial equipment, with a focus on predictive maintenance and the principles for building wireless sensor networks and data transmission protocols to collect information. The purpose of this study is to demonstrate the feasibility and reliability of using wireless sensors as technical diagnostic tools and as decision support tools for prediction. Main advantages include cost reductions and real-time information and analysis of equipment's state [10].

The use of Internet of Things (IoT) technologies to allow the exchange of information among sensors, machines, servers, and processing units is currently revolutionizing the industrial world, also providing innovation to the maintenance sector. Fernandes et al. [11] describe a related system for the prediction of failures in a metallurgical industry. There was no history of failures, so, learning takes place in an unsupervised way. Failures are predicted through moving average models, integrated in autoregressive models, using data from the sensors installed in the equipment, thus allowing the monitoring of different machine components and parameters [11].

3 Methods

3.1 Dataset and Data Preprocessing

The present case study focuses on the prediction of target variables on an industrial paper pulp press. The company in question provided a three-year dataset containing the history of six variables: Electric Current Intensity (Sensor 1); Pressure (Sensor 2); Rotation Speed (Sensor 3); Temperature (Sensor 4); Torque (Sensor 5); and Velocity (Sensor 6). All sensors collected data with a sample frequency of one minute.

The dataset contains several repeated values as well as discrepant samples which may be due to reading errors: the upper outliers may be the result of errors in the sensors reading; the lower outliers are a consequence of the same causes previously mentioned, as well as possibly downtimes programmed or not programmed.

Repeated values were removed using a Python algorithm developed by Mateus et al. [12]. The lower and upper outliers were also removed and subsequently replaced by the average of each variable in question.

Data was then transformed into histogram bins and predefined statistics computed from sliding windows. These statistics are the input of the ANN model Network. It should be noted that all data were normalized using the StandardScaler library from Sklearn before feeding into the ANN model.

To evaluate the performance of the Neural Network prediction algorithm, two different evaluation metrics were used: Mean Absolute Percentage Error (MAPE); and Mean Squared Error (MSE).

3.2 ANN Architecture

For predicting future values, the chosen architecture model was the Multi-Layer Perceptron (MLP), which is one of the most popular feed forward architectures, using Sklearn's MLPRegressor.

MLPRegressor uses various hyper parameters to optimize the generalization capacity of the network model for prediction. Various combinations of hyper-parameters were tested during training to find the best possible network configuration, so that the best forecast could be achieved.

To avoid overfitting, MLPRegressor might include a regularization term added to the loss function to reduce the number of model parameters.

The algorithm chosen for weight optimization was Adam solver. The Adam solver is an optimizer algorithm based on a stochastic gradient proposed by King-ma, Diederik and Jimmy Ba. This solver is recommended for large data sets [13].

In order to choose the ideal number of hidden layers, tests were carried out with one, two, and three hidden layers.

The initial approach for the input vector of the ANN considered only data from histogram bins of a sliding window, one for each sensor, along with their average values. However, the prediction results were very irregular. Therefore, it was decided to introduce additional metrics regarding the variance and median in the input vector, which made the prediction results more stable.

Results from tests carried out with two and three hidden layers were very similar and better than with one layer only. Hence, a network with two hidden layers (150, 75) was chosen. Figure 1 depicts the chosen architecture.

4 Experiments and Results

This section describes the experiments carried out during the development of this research study along with some of the best results obtained through several tests.

Tests were carried out with various window sizes applying. Histograms computed from those sliding windows, along with the respective mean, median, and variance, compose the input vector ($S_{11}, S_{12}, \dots, S_{1w}, \dots, S_{nw}, AVG_1, \dots, AVG_n, M_1, \dots, M_n, V_1, \dots, V_n$) of the ANN model. S_{ij} represents the value of sensor i at time j of the

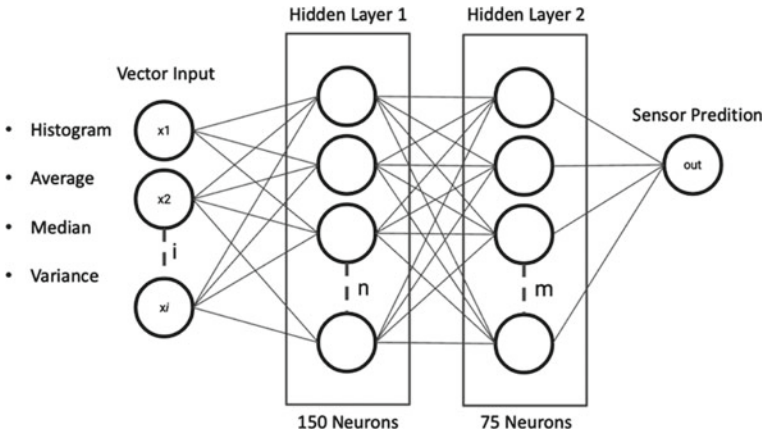


Fig. 1 Network architecture. The network receives histograms, average, median and variance from each sensor for a given sliding window, and outputs predictions for that sensor

window of size w . The window started with the first w samples of the time series and slid to the end of the series, in steps of 1, for an overlapping window, or steps of w samples for a non-overlapping window. AVG_k is the mean value of sensor k in the window. M_k is the median value of sensor k for the window and V_k is its variance.

Data were resampled to speed up processing. Results shown in Tables 1 and 2 were obtained with a period of 10 samples. The Neural Network performed up to 200 learning epochs in all tests. Table 1 shows the results of the tests developed using the overlapping window.

It can be seen from Table 1 that the window size of 2440 samples is the one that obtains the worst loss and MSE results for all variables. In general, the window can

Table 1 Results of the prediction tests with overlapping sliding window

Window size (samples)	Oil		Current		Temperature		Torque		Pressure		Velocity	
	MSE	Loss	MSE	Loss	MSE	Loss	MSE	Loss	MSE	Loss	MSE	Loss
2440	10.51	0.70	1.73	0.13	3.26	0.49	0.62	0.06	9.53	1.69	2.14	0.25
1440	7.23	0.48	1.67	0.11	3.67	0.29	0.80	0.06	7.57	1.18	2.84	0.19
720	5.94	0.19	1.56	0.09	3.55	0.19	0.54	0.06	4.74	0.78	1.84	0.14
360	5.22	0.07	1.45	0.07	3.28	0.13	0.56	0.05	5.39	0.51	1.68	0.04
180	5.44	0.13	1.15	0.06	2.59	0.10	0.55	0.04	4.94	0.36	1.43	0.08
144	5.69	0.12	1.15	0.06	2.46	0.09	0.54	0.04	4.62	0.34	1.51	0.08
90	5.02	0.11	1.08	0.05	2.46	0.09	0.56	0.04	4.53	0.31	1.43	0.07
45	5.24	0.20	1.08	0.06	2.23	0.11	0.56	0.03	4.97	0.37	1.46	0.07
24	5.60	0.41	1.14	0.07	2.29	0.19	0.52	0.03	4.96	0.69	1.49	0.08
12	5.76	0.86	1.10	0.11	2.16	0.34	0.53	0.03	4.70	1.25	1.61	0.13

Table 2 Results of the prediction tests using non-overlapping sliding windows

Window size (samples)	Oil		Current		Temperature		Torque		Pressure		Velocity	
	MSE	Loss	MSE	Loss	MSE	Loss	MSE	Loss	MSE	Loss	MSE	Loss
2440	66.93	2016.18	20.61	156.63	26.90	348.65	6.52	9.13	8.12	17.95	1.00	0.19
1440	66.19	1985.44	19.06	153.46	25.40	332.34	5.34	9.03	8.31	18.00	0.65	0.36
720	65.94	1945.29	18.79	153.35	25.32	329.65	4.99	9.08	8.21	20.09	1.19	0.59
360	54.69	1327.35	10.13	44.62	15.58	135.72	1.03	0.34	4.50	10.41	1.13	0.75
180	43.93	836.25	4.64	7.71	8.25	42.00	0.53	0.14	4.25	8.76	1.07	0.73
144	43.95	837.15	4.69	7.81	8.28	41.90	0.52	0.14	4.16	8.04	1.14	0.83
90	25.99	267.54	1.30	0.65	1.61	3.39	0.46	0.12	4.04	7.98	1.18	0.76
45	8.60	26.05	1.02	0.53	1.59	1.92	0.47	0.13	3.80	7.21	1.15	0.57
24	4.64	11.07	0.74	0.44	1.58	1.68	0.49	0.11	4.15	3.58	1.41	0.17
12	5.45	3.61	0.78	0.26	1.77	0.78	0.48	0.10	4.72	1.13	1.56	0.09

be reduced to six hours (360 samples) or in some cases up to 12 min (12 samples), since they all present good prediction results, which indicate that these will be the best window sizes to use for 10 samples resample rate in this data set.

Figure 2 shows that the learning of the Neural Network is quite fast: in just three epochs the network has already learned up to a small error.

Fast learning and the low data loss make it possible to achieve a good prediction fit as shown in Fig. 3. This prediction presents a Mean Squared Error of 1.10 (see Table

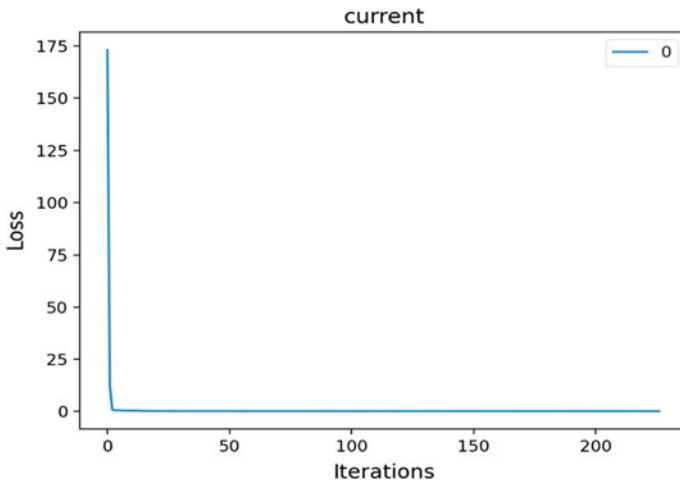


Fig. 2 Learning history of current’s intensity, using an overlap window size of 12 samples. The curve shows that the model learns in the first epochs

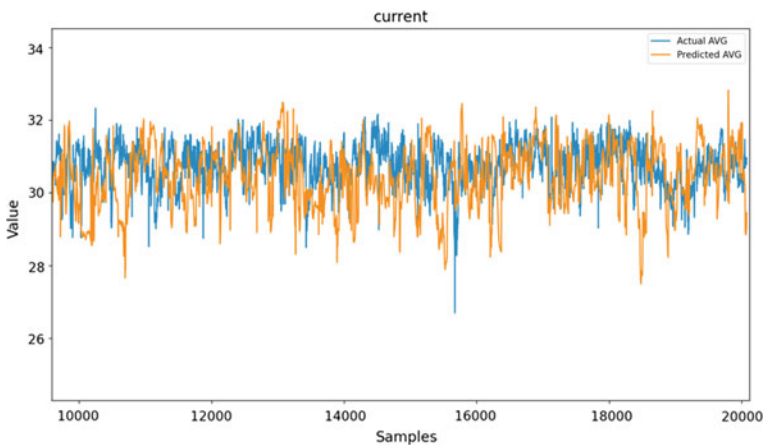


Fig. 3 Current’s intensity prediction using an overlapping window size of 12 samples

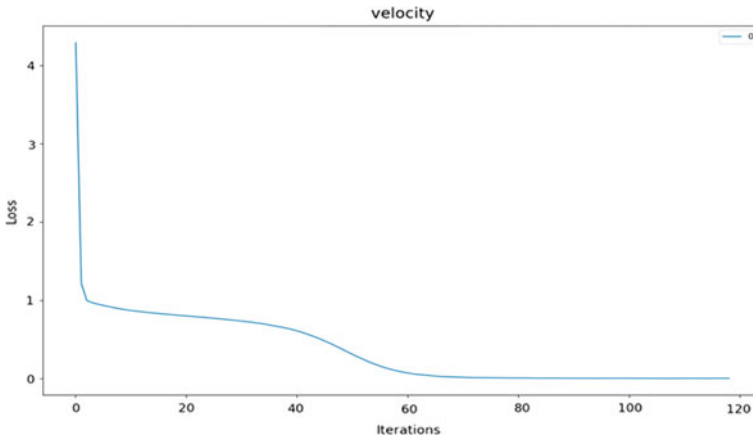


Fig. 4 Learning curve of the velocity using a non-overlapping window size of 12 samples

1). Table 2 shows corresponding results for the tests developed with non-overlapping sliding windows.

Table 2 shows the learning problems resulting from Neural Network when using large size windows without overlapping. However, window sizes with 24 and 12 samples generate results that are already satisfactory.

In the absence of the overlapping technique, the Neural Network receives much less input data per epoch, which delays the learning process. This problem is in part overcome by reducing the window size, for the Network receives more input data samples per epoch. However, when using smaller windows, it can be more difficult to catch larger patterns.

Figure 4 illustrates the network learning history for predicting velocity using a 12-sample overlapping window size. A slow slope is evident, demonstrating a slower learning rate when compared to other variables.

Note though that albeit learning is slow, the Network can learn, presenting a loss value of 0.09 in the window of 12 samples for the variable in question.

Figure 5 presents the corresponding results of the velocity prediction using a window size of 12 samples and 200 learning epochs. It should be noted that for the test shown in this figure the outliers from a programmed stop were not removed.

5 Conclusion

Prediction is very important for better decisions in maintenance and other areas. Data resampling can make the prediction process much faster since it reduces the dataset considerably.

The use of sliding windows over time series is necessary for training. Overlapping windows offer learning in less epochs. Larger windows make it easier to catch

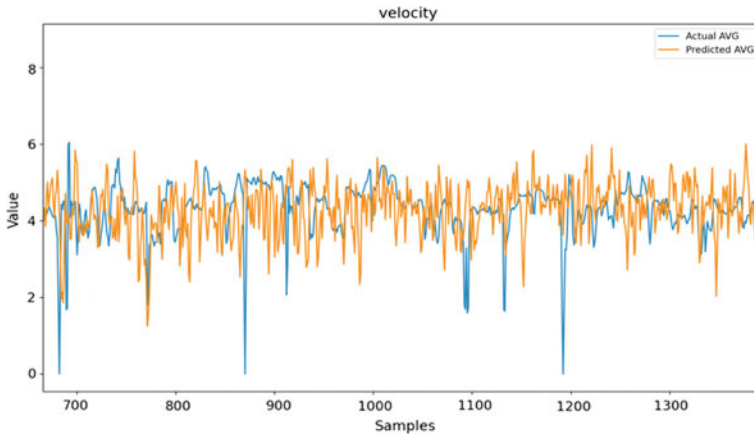


Fig. 5 Velocity prediction using a non-overlapping window size 12 samples

long trends, but the optimal window size needs to be determined experimentally. Overlapping windows offer more input data to the Neural Network in each epoch and thus generating faster learning rates and better prediction results.

Another drawback of non-overlapping windows is the limitation imposed by small datasets. Another major disadvantage of this type of sliding windows is the slower learning rate, especially for small window sizes. Its great advantage though is the speed of processing when the input vector is created.

Future work includes additional experiments to improve the input vectors and optimize other neural network hyperparameters.

Acknowledgements This research is sponsored by FEDER funds through the program COMPETE—Programa Operacional Factores de Competitividade—and by national funds through FCT—Fundação para a Ciência e a Tecnologia—under the project UIDB/00285/2020. This work was produced with the support of INCD funded by FCT and FEDER under the project 01/SAICT/2016 no 022153.

Conflicts of Interest The authors declare no conflict of interest.

Funding The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 871284 project SSHARE and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-029494, and by National Funds through the FCT—Portuguese Foundation for Science and Technology, under Projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

References

1. Cachada, A., et al.: Maintenance 4.0: intelligent and predictive maintenance system architecture. In: 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Sept 2018, vol. 1, pp. 139–146 (2018). <https://doi.org/10.1109/ETFA.2018.8502489>
2. Kumar, U., Galar, D., Parida, A., Stenström, C., Berges, L.: Maintenance performance metrics: a state-of-the-art review. *J. Qual. Maintenance Eng.* **19**(3), 233–277 (2013). <https://doi.org/10.1108/JQME-05-2013-0029>
3. Selcuk, S.: Predictive maintenance, its implementation and latest trends. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **231**(9), 1670–1679 (2017). <https://doi.org/10.1177/0954405415601640>
4. Martins, A.B., Torres Farinha, J., Marques Cardoso, A.: Calibration and certification of industrial sensors—a global review. *WSEAS Trans. Syst. Control* **15**, 394–416 (2020). <https://doi.org/10.37394/23203.2020.15.41>
5. Rodrigues, J., Costa, I., Farinha, J.T., Mendes, M., Margalho, L.: Predicting motor oil condition using artificial neural networks and principal component analysis. *Sci. Technol.* **22**(3), 9 (2020)<http://ein.org.pl/sites/default/files/2020-03-06.pdf>
6. Allah Bukhsh, Z., Saeed, A., Stipanovic, I., Doree, A.G.: Predictive maintenance using tree-based classification techniques: a case of railway switches. *Transp. Res. Part C Emerg. Technol.* **101**, 35–54 (2019). <https://doi.org/10.1016/j.trc.2019.02.001>
7. Hongxiang, T., Yuntao, L., Xiangjun, W.: Application of neural network to diesel engine SOA. In: 2011 Third international conference on measuring technology and mechatronics automation, Jan 2011, vol. 1, pp. 555–558. <https://doi.org/10.1109/ICMTMA.2011.141>
8. Okoh, C., Roy, R., Mehnen, J.: Predictive maintenance modelling for through-life engineering services. *Procedia CIRP* **59**, 196–201 (2017). <https://doi.org/10.1016/j.procir.2016.09.033>
9. Makridis, G., Kyriazis, D., Plitsos, S.: Predictive maintenance leveraging machine learning for time-series forecasting in the maritime industry. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Sept 2020, pp. 1–8. <https://doi.org/10.1109/ITSC45102.2020.9294450>
10. Vlasov, A.I., Grigoriev, P.V., Krivoshein, A.I., Shakhnov, V.A., Filin, S.S., Migalin, V.S.: Smart management of technologies: predictive maintenance of industrial equipment using wireless sensor networks. *Entrepreneurship Sustain. Issues* **6**(2), 489–502 (2018). [https://doi.org/10.9770/jesi.2018.6.2\(2\)](https://doi.org/10.9770/jesi.2018.6.2(2))
11. Fernandes, M., Canito, A., Corchado, J.M., Marreiros, G.: Fault detection mechanism of a predictive maintenance system based on autoregressive integrated moving average models. In: Distributed Computing and Artificial Intelligence, 16th International Conference, Cham, pp. 171–180 (2020). https://doi.org/10.1007/978-3-030-23887-2_20
12. Mateus, B.C., Mendes, M., Farinha, J.T., Cardoso, A.M.: Anticipating future behavior of an industrial press using LSTM networks. *Appl. Sci.* **11**(13) (2021). <https://doi.org/10.3390/app11136101>
13. sklearn.neural_network.MLPRegressor—scikit-learn 0.24.2 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html. Accessed 16 May 2021

YOLOV4-Based Wind Turbine Blade Crack Defect Detection



Xin Yan, Guoxin Wu, and Yunbo Zuo

Abstract Wind turbine blade is an important component of wind turbine. Wind turbine blade crack damage will cause hidden danger to the operation of wind turbine. The current wind turbine blade defect detection mainly relies on manual inspection, and the image detection technology can improve the inspection efficiency and reduce the unit maintenance cost. In view of the existing wind turbine blade crack defect detection algorithm with low recognition rate and low accuracy, a YOLOv4-based wind turbine blade crack detection method is proposed. First establish the wind turbine blade crack image dataset, then the anchor box parameters in YOLOV4 are optimized by K-means++ algorithm to make the anchor box parameters match the crack defect size; BiFPN is used instead of PANet to achieve better feature fusion, and finally the Focal Loss function is introduced to balance the number of small size defect samples in the data. The comparison tests show that the AP of the improved YOLOv4 algorithm reaches 93.49, which is better than the original YOLOv4 and the other three comparison algorithms, and has better efficiency and practicability.

Keywords Deep learning · Object detection · Blade fault · YOLOv4 · BiFPN · Focal loss · K-means++

1 Preface

With the global economic growth, the energy crisis and the environmental degradation caused by fossil energy are becoming more and more serious, and the active research and development of new energy sources have become the focus of attention of countries around the world. Wind energy has many advantages such as huge reserves, wide distribution, no pollution, and renewable, wind power has gradually become one of the most mature, most developed and promising power generation methods in the field of new energy [1]. According to the paper “The Future of

X. Yan · G. Wu (✉) · Y. Zuo

The Ministry of Education Key Laboratory of Modern Measurement and Control Technology, Beijing Information Science & Technology University, Beijing 100192, China
e-mail: [wxg1977@bistu.edu.cn](mailto:wgx1977@bistu.edu.cn)

Wind Energy: Deployment, Investment, Technology, Grid Connection and Socio-Economic Impact” published by International Renewable Energy Agency (IRENA), the global installed wind power capacity has reached 564 GW, and will reach 5044 GW by 2050. But due to the blade itself material factors, blade processing and installation process damage and blade work in the harsh environment and other reasons. Crack defects often appear on the blade, serious crack defects may lead to wind turbine out of service. Crack defects reduce the wind energy conversion rate and lifetime of wind turbine blades as well as wind turbines [2], so the detection of wind turbine blade defects is necessary.

At present, wind turbine blade defect detection means mainly focus on vibration detection technology, acoustic emission technology, infrared thermal imaging technology and machine vision detection technology. Wu Jianzhong used the short-time Fourier transform to analyze the blade in the health state and different crack damage state vibration signal change law, which provides a certain theoretical basis and method for the large wind turbine blade damage diagnosis [3]; however, because the defect detection based on vibration signal requires additional pickup and other sensor installation, which increases the complexity of the wiring and electrical system, Tang et al. used acoustic Tang used acoustic emission monitoring technology to achieve online monitoring of the structural health of wind turbine blades [4], the drawback of this method is that acoustic emission is very sensitive to environmental factors, and noise is difficult to exclude, so it is difficult to quantitatively analyze the defects; because when the wind turbine blade is subjected to dynamic mechanical load will generate stress and cause the temperature of the special area to rise, so that the image obtained by infrared thermal imaging technology in this area has different from the normal The problem of this method is that the working environment of wind turbine blades in service can interfere with the infrared thermographic signal [5].

The current wind farm for wind turbine blade defect detection mainly relies on telescope observation and rope dropping for repair, this manual interpretation of whether there is a defect method has a greater subjectivity, the disadvantage is time-consuming, large detection workload, low detection efficiency, subjectivity, is not conducive to the objective detection and assessment of blade health status.

During the past few years, with the development of deep learning theory and the improvement of arithmetic power of hardware devices, many representative object detection algorithms have been proposed and validated on public datasets [6]. The current mainstream object detection algorithms can be divided into two categories according to their design principles: two-stage object detection algorithms and one-stage object detection algorithms. The former first generates a series of bounding boxes for the detection points in the feature map before classifying the samples by convolutional neural networks, and common two-stage object detection algorithms include such as R-CNN [7], Fast R-CNN [8], and Faster-RCNN [9]. One-stage object detection algorithms do not generate bounding boxes and directly extract features to predict object types by convolutional neural networks, and common one-stage object detection algorithms include YOLO [10], YOLO v2 [11], YOLO v3, YOLOv4, SSD. The above algorithms are proposed and applied to provide a new idea for the defect detection of wind turbine blades.

In this paper, we establish a wind turbine blade crack image dataset and expand it by offline enhancement, and then improve the YOLOv4 object detection algorithm to achieve efficient and fast detection of wind turbine blade crack defects. For small-size crack defects in wind turbine blades, we first improve and optimize the anchor box size in YOLOv4 by K-means++ clustering algorithm, use BiFPN network to enhance the fusion capability of the algorithm for different feature layers to enhance the detection capability of the model for small-size crack defect targets, and use Focal Loss function to control the ratio of positive and negative samples. The number of small size defects is enriched by using Mosaic data enhancement to finally achieve excellent detection capability for small size crack defects of wind turbine blades. Compared with previous visual inspection methods for wind turbine blades, this method achieves higher accuracy and recall rate with greatly reduced number of model parameters and training time.

2 Dataset

At present, there is no open source data set for crack defect images of wind turbine blades. The main reasons are as follows: firstly, the number of cracked wind turbine blades is relatively small among in-service wind turbine blades; secondly, the current wind turbine blade cracks still largely rely on inspection workers to pass the naked eye. It is difficult to detect the small cracks by the naked eye before the wind turbine fails. At the same time, due to the harsh working environment of the wind turbine, the on-site collection of the wind turbine blade image data set requires high collection equipment and environment. Because the number of images of the crack defects of the fan blades is very scarce and cannot meet the training of the object inspection model, it is necessary to simulate the defects of the fan blades in a laboratory environment and obtain a data set. This paper constructs the BladeFault-VOC-V1 image data set of fan blade crack defects in VOC format for target detection, and trains and tests the target detection algorithm based on this data set. The BladeFault-VOC-V1 data set has a total of 320 images, collected and expanded in a well-built laboratory environment. This data set contains two types of data, the positive sample of the defective fan blade image and the negative sample of the normal fan blade, in the form of the defect. It is a crack defect.

2.1 Wind Turbine Blade Defect Marking System

This system mainly consists of wind turbine and defective blade, industrial camera, image acquisition card, PC and other parts. The industrial camera is selected as MV-E7000M industrial camera, which has the advantages of high definition, high precision and good color reproduction and is widely used in the field of artificial intelligence. The workflow of the labeling system is as follows: firstly, the simulated

cracked fan blade is photographed under sunlight conditions by the industrial camera, and the image acquisition system with the camera as the core component acquires continuous image information on the blade part to be detected; then the A/D signal is converted by the image acquisition card, and then the converted signal is converted into digital image information recognized by the computer, and the image information is transmitted back to the computer, through Labellmg labeling tool to blade image of the crack for labeling.

2.2 Build Dataset

The established BladeFault-VOC-V1 data set was collected from a key experiment of the Ministry of Education of Beijing Information Science and Technology University of modern measurement and control technology. The image resolution is $3000 * 4000$, the shooting angle is 90° , 60° , 45° , and the distance from the blade is 0.8 and 0.5 m. A total of 32 images of leaves with cracks were taken. There are more than 3 crack defects in each sample image. Considering that the actual inspection process of wind turbine blades is mainly carried out on sunny days with good light, this article also carried out under sunshine conditions when the simulated data set was collected. The image was annotated with the PASCAL VOC format data set using Labellmg annotation tool. Each image was accurately marked with the location of the crack, including 116 crack defects in total.

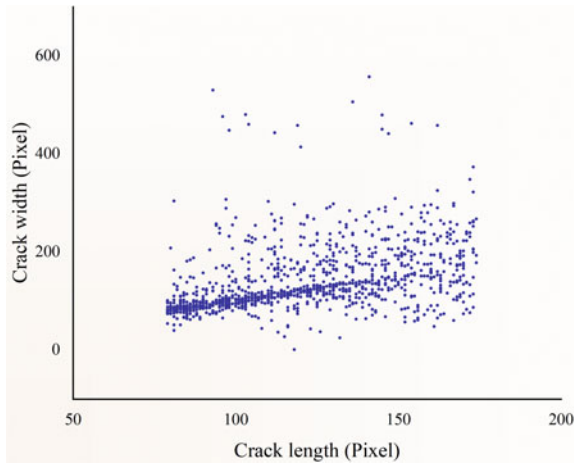
To address the overfitting problem that may be caused by the small sample size of the dataset, we augment the dataset with data before training including: flipping, scaling, random cropping, and various filtering operations to expand the dataset by 10 times. The limited data is used to produce the equivalent value of more data, thus improving the training effect. The final BladeFault-VOC-V1 dataset was obtained after the expansion. The dataset contains a total of 1160 crack defects with an average width and height of 120 Pixel and 139 Pixel, and an average area of 17,528, accounting for 0.146% of the original map area. The specific defect size distribution diagram is shown in Fig. 1.

3 Algorithm

As a classical one-stage object detection algorithm, the YOLO (You Only Look Once) family of object detection algorithms has received a lot of attention from academia and industry. Although YOLOv4 does not make revolutionary changes, it adds many practical techniques to the YOLOv3 algorithm. Compared with YOLOv3, the AP and FPS of YOLOv4 are improved by 10% and 12% respectively.

The YOLOv4 object detection algorithm can be divided into three parts: the backbone feature extraction network CSPDarkNet53, the enhanced feature extraction

Fig. 1 Defect size scatter diagram



network SPP+PANet, and YOLO head. Compared with the YOLOv3 model, its main improvement ideas are as follows:

1. Input: some improvement operations were done in the training phase of the model, mainly including Mosaic data augmentation, Cross mini-Batch Normalization, Self-Adversarial Training.
2. Backbone: Using CSPDarkNet53, Mish activation function and dropblock.
3. Neck: Adding SPP module and FPN+PAN structure.
4. Head: The main improvement is the loss function GIoU loss during training, and the NMS filtered by the predicts bounding boxes becomes DIOU NMS.

In order to improve the detection capability of YOLOv4 for small size crack defects, this paper improves the YOLOv4 algorithm to achieve effective detection of crack defects in wind turbine blades. The general flow of the improved algorithm is as follows.

Firstly, K-Means++ [12] clustering algorithm is used to analyze the crack defect size and get the parameters of suitable bounding boxes, CSPDarkNet53 is selected as backbone feature extraction network for initial feature extraction, then BiFPN is used to replace PANet for feature fusion and use Focal Loss to balance the contribution of positive and negative samples to the algorithm, YoloHead is finally used to predict and get the detection result. The flow chart of the improved YOLOV4 algorithm is shown in Fig. 2.

3.1 Clustering Analysis

The YOLOv4 model generates a series of Anchor Boxes for object detection, and then adjusts the parameters of the Anchor Boxes to generate prediction frames, so

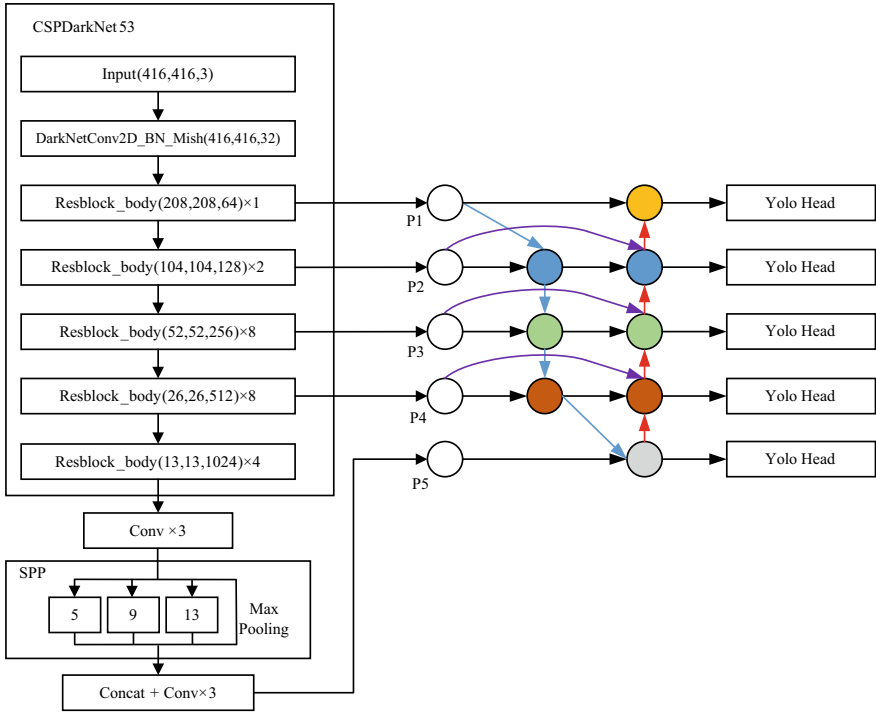


Fig. 2 Improved YOLOv4 algorithm model

the choice of Anchor Boxes has a great influence on the model prediction box. The Anchor Box is a more suitable Anchor based on the COCO dataset, but it is not suitable for the small-sized wind turbine blade crack defects in our BladeFault-VOC-V1 dataset. Therefore, in this paper, we use K-means++ method to cluster the small size wind turbine blade crack defect size in BladeFault-VOC-V1 dataset, get the suitable Anchor size and adjust it to get the suitable Anchor box size for our dataset. The original K-means clustering method requires artificially setting K points as preset clustering centers, and different clustering centers may lead to not completely different clustering results.

The K-means++ clustering algorithm we choose is superior to the K-means algorithm, it does not need to set the initial clustering center artificially, its implementation process is: first randomly select a point as the first initial class cluster centroid, Then select the point furthest from the centroid of the first cluster as the center point of the second cluster, and then select the point furthest from the center points of the first two clusters among the remaining points as the third cluster centroid. In this way, a total of K initial cluster centroid is selected. After the centroid of the initial cluster is selected, clustering is performed according to the K-means algorithm to complete the implementation of the K-means++ clustering algorithm. The Anchor Box sizes

obtained after K-means clustering algorithm analysis are: (61, 65) (68, 73) (79, 81) (93, 235) (108, 431) (110, 244) (139, 101) (142, 439) (168, 187).

3.2 BiFPN

One difficulty in the current object detection field is how to represent and process multi-scale features. In order to solve this problem, FPN [13], NAS-FPN and PANet have appeared in recent years. FPN proposed a top-down fusion of multi-scale features. According to this idea, PANet adds a bottom-up path aggregation network based on FPN. The recent NAS-FPN uses neural structure search to automatically design Feature network topology, although its performance is excellent, the exploration process often requires thousands of GPUs.

BiFPN is an enhanced version of FPN. Compared with ordinary FPN, the construction of BiFPN is more complicated. First, it removes the nodes that have only one input edge that contributes little to the fusion of different features, and then adds an additional edge to the original input and output in the same layer, and adds an additional edge to the original input and output in the upper and lower layers. Connected separately, and finally repeat each top-down and bottom-up bidirectional path to achieve more advanced feature fusion. At the same time, a channel attention mechanism is set up to balance the weights of different feature layers. The previous feature fusion methods treat all input features equally, but different input features have different contributions to output features at different resolutions. For this reason, BiFPN adds a weight to each input, allowing the network to judge the importance of different inputs. Finally, BiFPN realizes the feature fusion of different scales and balances the feature information of different scales through two-way cross-scale connection and fast normalization fusion. As a specific example, Eqs. (1) and (2) are the fourth-level feature fusion formulas:

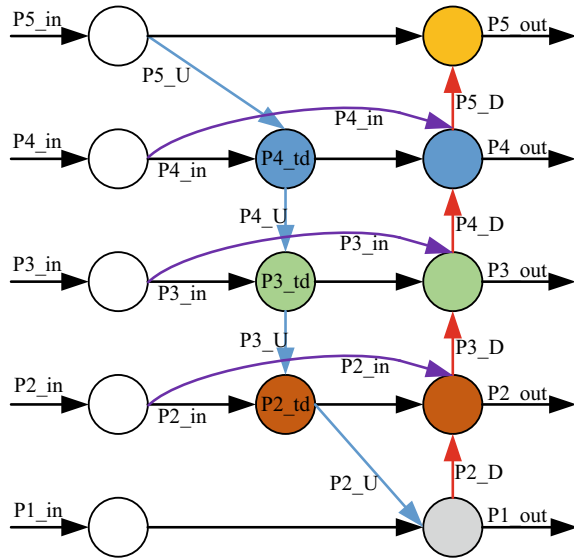
$$P_4^{td} = Conv \left(\frac{\omega_1 \cdot P_4^{in} + \omega_2 \cdot Resize(P_5^{in})}{\omega_1 + \omega_2 + \varepsilon} \right) \quad (1)$$

$$P_4^{Out} = Conv \left(\frac{\omega_1' \cdot P_4^{in} + \omega_2' \cdot P_4^{td} + \omega_3' \cdot Resize(P_3^{out})}{\omega_1' + \omega_2' + \omega_3' + \varepsilon} \right) \quad (2)$$

Among them, P_4^{td} is the fourth-level intermediate feature on the top-down path, and P_4^{Out} is the fourth-level output feature, ω is the weight of each node.

The working principle of BiFPN is shown in Fig. 3.

Fig. 3 BiFPN working principle diagram



3.3 Focal Loss

Focal Loss is a new Loss calculation scheme proposed by He Kaiming. After an image is input to the network, a large number of candidate frames will be generated, but only a few of them contain objects. Therefore, the number of positive samples with objects is often much smaller than the number of negative samples without objects. At the same time, for the object detector in this article, large-size cracks are simple samples that are easy to be classified, while small-size cracks are difficult samples that are difficult to be classified. Their optimization capabilities for the model are different. Focal Loss reduces the weight of a large number of simple negative samples in training and solves the problem of the imbalance of the positive and negative sample ratio in the one-stage object detector. It has two main features: one is to control the weight of positive and negative samples, and the other is to control the weight of samples that are easy to classify and difficult to classify. The mathematical definition of Focal loss is expressed in Eqs. (3), (4).

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \tag{3}$$

$$p_t = \begin{cases} p & x = 1 \\ 1 - p & otherwise \end{cases} \tag{4}$$

4 Experiments and Data Analysis

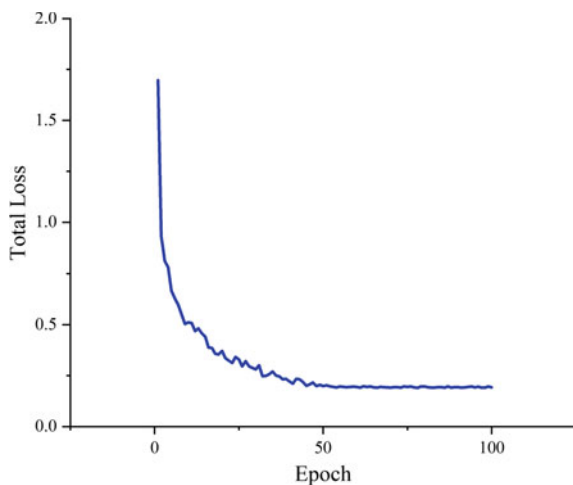
In order to evaluate the performance of the algorithm in this paper, we design a comparative test to verify it. The test data set uses the self-built blade fault image data set BladeFault-VOC-V1. The software and hardware information used by the training platform is shown in Table 1.

In terms of training parameters, the Batch size is set to 4, and the initial learning rate is set to 0.001. When the number of iterations reaches 2000 and 5000, the learning rate is reduced to 0.0001 and 0.00001. The convergence curve of the loss value during the algorithm training process is shown in Fig. 4.

Table 1 Software and hardware information

	Environment	Description
Hardware environment	Windows10	Operating system
	Inter(R) i7-10870H CPU @2.20 GHz	CPU
	RTX 3070 8G	GPU
	DDR4 16G 3200 MHz	Memory
Software environment	Python3.8.8	Python version
	Keras2.4.3	Python library version
	Tensorflow-gpu 2.4.1	Python library version
	Opencv-python 3.4.10	Python library version
	CUDA 11.1	Python library version

Fig. 4 Training loss graph



4.1 Algorithm Evaluation Indicators

The evaluation indicators commonly used in object detection algorithm include: Precision, Recall, Average Precision (AP). Among them, AP is considered to be an important index for the evaluation of algorithms in the target detection field, so this paper chooses AP as the evaluation index.

The recall rate represents the proportion of all defective samples that the classifier considers to be defective samples that are indeed defective samples. The calculation formula is as Eq. (5):

$$R = \frac{X_{TP}}{X_{TP} + X_{FN}} \quad (5)$$

Accuracy represents the proportion of samples that the classifier considers to be defective and that are indeed defective samples to account for the proportion of samples that the classifier considers to be defective. The calculation formula is as Eq. (6):

$$P = \frac{X_{TP}}{X_{TP} + X_{FP}} \quad (6)$$

Among them: X_{TP} represents the number of defective samples that have been correctly classified; X_{FN} represents the number of defective samples that have been misclassified; X_{FP} represents the number of non-defective samples that have been misclassified.

AP refers to the area under the P-R curve of the defect target, which is used to measure the average classification accuracy of defects. The calculation formula is as Eq. (7):

$$AP = \int_0^1 P(R)dt \quad (7)$$

4.2 Detection Result of Crack Defects

Select two fan blades cracks defect image, using the original and improved algorithms YOLOv4 detection algorithm, the detection results are shown in Fig. 5.

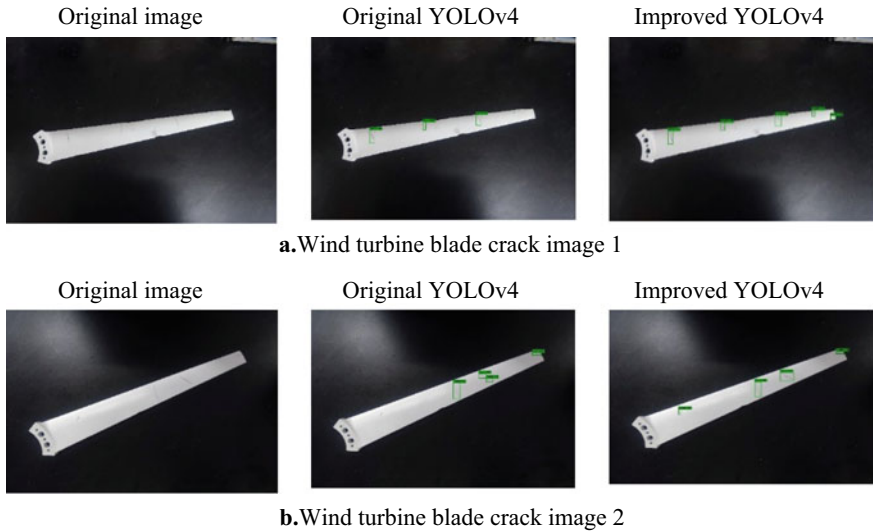


Fig. 5 Test result comparison

4.3 Comparative Experiment

In order to verify that the algorithm in this paper can effectively detect the small-size crack defects of the blade, we set up a set of comparative tests. Train and test the algorithm in this paper with Faster R-CNN, SSD and the original YOLOv4 object detection algorithm. The data set used is the BladeFault-VOC-V1 data set. Table 2 shows the comparative experimental results of the above detection algorithms.

It can be seen from the table that the AP of Faster R-CNN is only 11.25, which is not suitable for detecting crack defects. At the same time, due to the characteristics of the two-stage object detection algorithm, its FPS is only 12, while the one-stage object detection The FPS of the algorithm is generally higher than that of the two-stage object detection algorithm. In the one-stage object detection algorithm, the main difference between SSD-500 and SSD-300 is the resolution of the input image. The AP of the SSD-500 inputting $500 * 500 * 3$ image is compared with the inputting $300 * 300 * 3$ SSD. The AP of -300 is 17.59 higher, but its AP value is only 33.88,

Table 2 Experimental results

Algorithm	AP/%	FPS/(f s ⁻¹)
Modified-YOLOv4	93.49	51
YOLOv4	90.26	58
SSD-300	16.29	45
SSD-500	33.88	37
Faster R-CNN	11.25	12

which cannot be applied to actual detection. Compared with the original YOLOv4, the detection speed of the algorithm in this paper is reduced, but the AP value for crack defects has increased by 3.23, which is far more than other object detection algorithms.

5 Conclusion

5.1 A Subsection Sample

Aiming at the problem that the small-size crack defects of fan blades are difficult to be detected by conventional object detection algorithms, this paper improves the YOLOv4 object detection algorithm. First, use the K-means++ clustering algorithm to select a more suitable Anchor Box, then use BiFPN to improve the feature fusion ability of the algorithm, and finally use the Focal Loss function to balance the positive and negative samples in the data. Experimental verification shows that the improved YOLOv4 object detection algorithm can effectively detect small-size crack defects in fan blades. The AP value reaches 93.49 and the detection speed reaches 45 f s^{-1} .

Although the YOLOv4 object detection algorithm has been improved in this paper and good experimental results have been obtained, the defects of fan blades are of various forms, and many defects often appear at the same time in practical applications. Therefore, in the future, more different types of fan blade defects will be added and the algorithm will be further improved to enhance the practicality of the algorithm.

References

1. Bo, L., Zhijia, H., Hao, J.: Current status and development trend of wind power generation. *J. Northeast Electr. Power Univ.* **36**(02), 7–13 (2016)
2. Gang, L., Hongli, H., Yani, L.: Recent developments in wind turbine blade condition monitoring and fault diagnosis technology. *Ind. Instrum. Autom.* **05**, 16–20+55 (2017)
3. Jianzhong, W., Yi, T.: Crack damage detection of fan blades based on short-time Fourier transform. *Chin. J. Eng. Mach.* **12**(02), 180–183 (2014)
4. Tang, J., Soua, S., Mares, C., Gan, T.-H.: An experimental study of acoustic emission methodology for in service condition monitoring of wind turbine blades. *Renew. Energy*, **99** (2016)
5. Infra-red thermography for condition monitoring of composite wind turbine blades: feasibility studies using cyclic loading tests. *NDT & E Int.* **29**(6), 395–395 (1996)
6. Shun, Z., Yihong, G., Jinjun, W.: The development of deep convolutional neural networks and their applications in the field of computer vision. *Chin. J. Comput.* **42**(03), 453–482 (2019)
7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014). <https://doi.org/10.1109/CVPR.2014.81>
8. Girshick, R.: Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448 (2015). <https://doi.org/10.1109/ICCV.2015.169>

9. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017). <https://doi.org/10.1109/TPAMI.2016.2577031>
10. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016). <https://doi.org/10.1109/CVPR.2016.91>
11. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517–6525 (2017). <https://doi.org/10.1109/CVPR.2017.690>
12. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 881–892 (2002). <https://doi.org/10.1109/TPAMI.2002.1017616>
13. Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944 (2017). <https://doi.org/10.1109/CVPR.2017.106>

State-of-Art of Metal Debris Detection in Online Oil Monitoring



Dingxin Yang and Xiaorong Liu

Abstract Metal debris detection technology in online oil monitoring has drawn significant industrial attention recently since it can provide information about wear, lubrication and friction conditions of friction pairs. In the paper, the principles of several types of metal debris sensors based on photoelectric or imaging, X-ray luminescence spectrum, ultrasonic detection and electric impedance measurement are reviewed. Especially, the inductive debris sensors which gain advantages of simple structure, complete flow measuring and the ability to distinguish ferromagnetic and non-ferromagnetic metal particles, have received extensive attention. The developing progress, detection principle, typical sensor structure, and industrial applications are presented in detail. We also provided the prototype of debris sensor we developed and the performance and application in bearing health status evaluation and remain life prediction. Finally, the main problems confronted and the development trend for debris sensor using in industrial applications are also proposed.

Keywords Online oil monitoring · Debris detection · Inductive debris sensor

1 Introduction

Online metal debris detection in lubricating oil has received more and more focus because it is of great significance to get the real-time wear information from metal debris in lubricating oil. This information can be used in early prediction of catastrophic failure for large and complex equipment such as a ship power plant and an aeroengine.

Support bearing is an important part of aero-engine. It usually works under high load, high speed and high temperature. According to the statistics of engine bearing faults, the excessive rolling and erosion faults caused by metal wear particles account for 59% of the total faults [1, 2]. Therefore, on-line detection of wear metal particles

D. Yang (✉) · X. Liu

Laboratory of Science and Technology on Integrated Logistics Support, National University of Defense Technology, Changsha 410073, China
e-mail: yangdingxincn@163.com

in the oil of aero-engine bearing is of great significance to ensure the normal operation of the equipment and avoid irreparable damage to the aero-engine.

According to the wear law of mechanical friction pairs, in the normal wear stage, the size of metal particles is generally less than 20 μm . When the wear process develops from normal wear stage to sharp wear stage, the particle size increases to the range of 50–200 μm . According to experimental results, most of the normal wear metal particle size of F119 engine bearing is less than 100 μm . The size of abnormal wear metal particles is in the range of 100–150 μm , or even large than 200 μm . Moreover, relevant experiments show that the distribution trend of wear particle size has no correlation with the size of bearing itself [3]. So wear metal debris monitoring, especially particles with the size from 50 to 200 μm is of the key to condition monitoring of engine machines, and friction system. The wear information can be used for early wear state assessment and fault prediction of key engine components.

2 Online Metal Debris Detection Technology

2.1 *The Main Principles for Online Metal Debris Detection Technology*

The metal wear particles in lubricating oil will change many physical or chemical properties of the oil, such as the dielectric constant, permeability, conductivity, alkalinity acidity and optical transmission characteristics etc. The online metal debris sensor can detect the changes of these parameters in the oil to determine whether there are wear particles in the oil, and to obtain further information such as size, shape, ferromagnetic properties and even elementary composition of wear particles. According to the different detection principles, the metal wear particle sensors can be divided into optical sensors, electromagnetic sensors, acoustic sensors and other types of sensors. The principle of these sensors is briefly introduced below.

Photoelectric or imaging based Technology. The light transmission characteristic of lubricating oil is affected by presence of oil debris. So the optical detection device can be used to measure the optical transmission to detect the particles in the oil. Light scattering and micro imaging are two types of optical principles. Wear debris in oil can be quantitatively and qualitatively detected by measuring the attenuated light density after light has transmitted through an oil sample [2].

Light scattering system usually consists in a laser LED as light source and a receiving photo diode as light detector. The incident light intensity can be absorbed and scattered by the wear debris suspending in lubrication oil, so the light intensity changes can represent the debris concentration. The attenuated light density is detected by the receiving diode, which can be used to calculate the debris size and concentration in the lubrication oil. The system developed by Wuhan University of

Technology uses the light scattering characteristics of abrasive particles in oil to monitor the size of solid particles in oil [4].

In addition to the optical detection sensor based on the principle of light transmission, the optical imaging system can also be used to obtain the size and shape information of debris directly. Researchers from Xi'an Jiaotong University focused on the on-line ferrography analyzer. They designed an imaging wear particle detection sensor system, which consists of a light source, a fluid channel and a CMOS image sensor [5, 6]. The light source is used to illuminate the fluid channel, and the CMOS image sensor can take high-speed photos of wear particles passing through the light source. These images are transmitted to PC and processed by software to obtain the size and shape information of particles. The debris detection results are relatively direct and reliable. The system also needs high-performance hardware and efficient image processing algorithm to accurately distinguish the wear particles.

X-ray Based Photoelectric or Imaging Based Technology. X-ray fluorescence spectrometry (XRF) analysis is one of the most commonly used oil analysis methods in the laboratory. Based on the principle that wear particles are excited by X-ray energy to produce dispersive fluorescence spectrum. The content of metal elements in wear particles can be detected. This method can provide very reliable and accurate information about the oil sample.

However, due to the complex structure and operation process, the traditional spectral analysis is only used for off-line analysis of lubricating oil. Later, Spectro Analytical Instruments Inc. [7] proposed an XRF system that can monitor wear particles online. The X-ray emission source emits X-ray to a fluid chamber. When a metal wear particle passes through the fluid chamber, the X-ray will be reflected by the particle and detected by the X-ray detector. This technology allows the user to measure concentrations of 12 elements and up to six different elements at a time. The advantages of XRF technology are that it is nondestructive, noninvasive and it typically needs minimal services. But the relatively expensive cost and complex processing process make it difficult for this technique to be widely used. Another in-line XRF system has also been developed at the Pacific Northwest National Laboratory (PNNL) [8].

Ultrasonic Wave Based Technology. Based on the principle that wear particles in oil can reflect the incident ultrasonic wave, the acoustic detection technology of wear particles has been developed. The acoustic sensor uses a focused pulse echo ultrasonic transducer to act as an ultrasonic transmitter and receiver at the same time.

When the incident sound wave meets the wear particles, it will produce the acoustic pulse reflection, and the receiver receives the echo generated by the particles. The amplitude of the ultrasonic echo signal is closely related to the size, shape and material of the wear particles. Then, after a period of time, we will receive the echo from the back wall. There is always a fixed time interval between the incident pulse and the reflected pulse, and the reflected signal of particles can be accurately extracted in this time interval. In addition, because the acoustic reflection coefficient of bubbles in oil is negative and that of solid metal particles is positive, when the incident pulse encounters bubbles, the waveform phase will be reversed. Therefore, bubbles and solid particles can be distinguished according to the pulse phase of reflection [9].

However, this method can not distinguish the ferromagnetic or non-ferromagnetic of solid metal particles because the acoustic reflection coefficients of both types of solid wear particles are very close. Moreover, the selection of ultrasonic frequency and vibration noise will also affect the detection performance of the sensor.

Electrical Impedance Based Detection Technology. The principle of electric impedance measurement is widely used in online wear debris detection. The metal debris in the lubrication oil will lead to the change of oil electric impedance. The electric parameters of oil include conductivity, dielectric constant and permeability, which can correspond to the impedance parameters of resistance, capacitance and inductance respectively. The change of these parameters has a certain relationship with the size and concentration of particles in the oil. Therefore, if the designed sensor can measure the change of the corresponding impedance, the information of particles in the oil will be obtained. The types of sensors include resistance sensor [10], capacitance sensor [11], electrostatic sensor [12] and inductive sensor.

Among these sensors, the inductive debris sensor can be directly connected to the oil pipe, and perform full- flow measurement of the oil pipe without bypass. It can count the metal wear debris in the oil, make a statistics of the debris size range, and distinguish ferromagnetic and non ferromagnetic metal particles. Therefore, it is one of the most practical metal debris detection technologies, and has been used widely in industrial applications. The following focuses on the inductive online metal debris detection technology.

2.2 State-of-Art of Inductive Metal Debris Sensor

According to the literature, K. W. Chambers designed an inductive wear debris monitoring sensor in 1988, and studied the influence of induced current, oil temperature, viscosity, flow rate and particle size on the output signal [13]. MetalSCAN debris sensor developed by Gastops Ltd. of Canada is the most widely used online oil debris sensor [8]. The sensor is directly installed in the oil pipe and connected with the control unit through the connecting cable. There are three inner coils surrounding the inside bore of the sensor. Two outer coils are reversely wound and driven by AC power supply. The magnetic fields generated by them are in opposite directions. The midpoint between the two coils cancels each other. The sensor coil in the middle of the core creates magnetic disturbance due to the passing of metal debris. The amplitude of magnetic field change is converted into voltage value to determine the particle size. The phase change is used to determine the properties of the particles (ferromagnetic or non ferromagnetic).

Depending on the type and magnitude of the magnetic disturbance, the control unit determines the type of particle and the particle size. MetalSCAN has been claimed to be the first full flow wear debris sensor. MetalSCAN sensor has been successfully used in the condition monitoring of F22, Apache helicopter, and other power systems. The latest Metalscan4110 can detect ferromagnetic particles and non ferromagnetic particles with sizes of more than 100 μm and 305 μm , respectively [14]. The inner

diameter of the pipe used is 7.6 mm. At the same time, it also provides sensors with different pipe diameters.

Similar to MetalSCAN metal debris monitoring sensor, there are the TechAler™ 10 wear debris sensor developed by MACOM technologies of the United States and the FG online wear particle sensor developed by Kittiwake of the United Kingdom. TechAler™ 10 can provide wear particle size distribution and image information at different failure stages of the machine, and has a patented screening technology to eliminate false alarms caused by blisters and bubbles. The minimum ferromagnetic particle size it can detect is 50, 150 μm for non-ferromagnetic particles. However, the sensor needs to be connected from the lubrication system by-pass when it is installed [15].

Akron university has designed a high throughput inductive pulse sensor based on inductive Coulter counting principle for detecting metallic wear debris in lubrication oil [16]. The sensor is composed of two layers of coils. The oil passes through the middle of the coil. It is proved that the sensor can monitor and distinguish ferromagnetic particles and non ferromagnetic particles with high sensitivity. In addition, an inductive microfluidic monitoring sensor is also designed, and its detection range is 50–125 μm [17]. However, this kind of sensor can only adapt to the situation of low oil flow rate.

In practical application, the GE90 high bypass ratio turbofan engine used by Boeing 777 is the first commercial aeroengine equipped with debris monitoring system. The system can collect and calculate large size ferromagnetic particles in oil. The practical application shows that under the maximum engine speed, the efficiency of the debris monitoring system to capture the bearing rotational fatigue debris is as high as 90%. While the capture rate of the traditional magnetic plug is less than 30%.

Since the 1990s, several universities and research institutions in China have been carrying out the research work of oil debris monitoring sensor. Central South University has studied the internal magnetic field of the electromagnetic abrasive sensor, established the mathematical model of the internal magnetic field, verified the correctness of the theoretical model of the output characteristics of the sensor through the experimental data [18]. Beijing Hangfeng Ltd. has developed an online metal particles sensor for lubricating oil monitoring. The technical manual shows that the minimum detectable ferromagnetic particle is 125 μm , non-ferromagnetic particle size is 450 μm with the pipe diameter of 8 mm [19]. Ren Yijun from University of Science and Technology of China have studied the optimization method of the structural parameters of the sensor and developed an ultrasensitive inductive metal debris sensor. The experiment results show the sensor can successfully detect 134 μm ferromagnetic particles and 230 μm non-ferromagnetic particles in oil tube of 43 mm diameter [20]. There are also many other similar studies, which will not be introduced one by one. We will present our own research work on debris sensor in the following.

3 The Debris Detection Sensor We Developed

According to the experimental statistics, the on-line monitoring of metal wear particles of size of about $100\ \mu\text{m}$ can provide effective data support for the early fault prediction of complex mechanical systems. In order to effectively detect the ferromagnetic metal debris with size equal to or less than $100\ \mu\text{m}$, our research team has carried out on-line monitoring technology research on inductive micro metal wear particle sensor. Firstly, the magnetic field distribution and output characteristics of the sensor are analyzed theoretically and simulated numerically. On this basis, the structural parameters of the sensor are optimized based on the heritage algorithm.

When the metal particles equal to or less than $100\ \mu\text{m}$ pass through the sensor, the output characteristic signal is very weak and easy to be disturbed by noise. Aiming at this problem, the weak characteristic signal enhancement detection methods are studied. A weak aperiodic signal enhancement transmission algorithm based on stochastic resonance principle is studied and verified by numerical simulation. The wavelet method and stochastic resonance method are used to enhance the signal transmission of small wear particle sensor. The results show that both methods can enhance the characteristics of aperiodic impact signal of small metal wear particle when it passes the sensor. The method of stochastic resonance has a better effect to detect the weak output characteristic signal.

Based on the results of sensor optimization design and the method of weak characteristic signal extraction, two kinds of metal wear particle detection sensors with 8 mm diameter and 15 mm diameter were developed. The developed sensors prototypes are shown in Fig. 1.

By making different size of metal particles samples, the sensitivity detection experiment of metal particles was carried out. The experimental results show that the metal debris sensor with flow hole diameter of 8 mm can effectively detect $70\ \mu\text{m}$ ferromagnetic metal particles. While the sensor with 15 mm diameter can detect $150\ \mu\text{m}$



Fig. 1 The inductive metal debris sensor prototypes we developed (Left: for 8 mm diameter pipe, Right: for 15 mm diameter pipe)

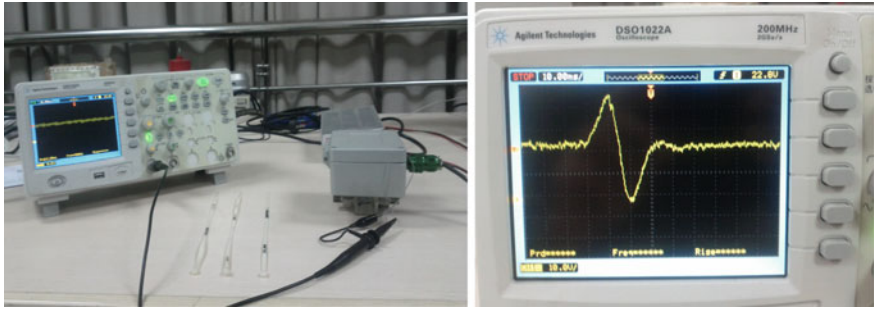


Fig. 2 The captured output waveform of the sensor when 100 μm ferromagnetic metal particle is passing through the sensor

ferromagnetic ones. The output waveform of sensor prototype when the ferromagnetic particle passing through the sensor is shown in Fig. 2. The experiment results have proven that the effective detection of 100 μm ferromagnetic metal particles.

Based on the theoretical analysis and numerical simulation of the magnetic field distribution and output characteristics of the sensor, the structural parameters of the sensor are optimized, and the principle prototype of the multi-layer spiral tube double excitation electromagnetic on-line wear particle monitoring sensor is developed. The adaptive filtering algorithm is designed to filter the output signal, which improves the sensitivity to small particles. The 10 mm diameter sensor can effectively detect the diameter of 120 mm μM. Sensor principle prototype and maximum diameter 120 μm. The output signal waveform of the sensor when the ferromagnetic metal particles passing through is shown in the figure below, which realizes the effective detection of 100 μm diameter ferromagnetic metal particles.

4 Conclusion

The on-line monitoring technology of oil wear particles is one of the effective methods of mechanical condition monitoring. Based on the different working principles, this paper introduces a variety of on-line oil debris monitoring technologies and their development status, and analyzes their advantages and disadvantages. At present, the technology of on-line monitoring system of oil wear particles based on inductive sensor is mature and has high reliability.

The current development trend of wear particle on-line detection technology can be summarized as two aspects: First, the new sensing principle and structure design are adopted to continuously improve the detection ability of metal wear particles in the size range of 50–100 μm. The second is to improve the ability to distinguish and identify the types of wear particles by the fusion of different sensing technologies. For example, inductive wear particle sensing technology can distinguish ferromagnetic and non ferromagnetic metal particles, but it can not effectively detect non-metallic

particles. Ultrasonic abrasive testing technology can detect non-metallic particles and bubbles, but can not distinguish metal, non-metallic metal particles. Through the fusion or integration of the above two sensing technologies, the recognition ability of different types of wear particles can be improved.

References

1. Zhan, H.Q., Song, Y., Zhao, H., Gu, J., Yang, H.Y., Li, S.: Study of the sensor for on-line lubricating oil debris monitoring. *Sens. Trans.* **175**(7), 214–219 (2014)
2. Wu, T.H., Wu, H., Du, Y., Peng, Z.: Progress and trend of sensor technology for on-line oil monitoring. *Sci. China Technol. Sci.* **56**(12), 2914–2926 (2013)
3. Miller J.L., Kitaljevich, D.: In-line oil debris monitor for aircraft engine condition assessment. In: 2000 IEEE Aerospace Conference, pp. 49–56. IEEE, New York (2000)
4. Yonghui, Y., Xinpin, Y., Hanlian, X.: Development on optic-fiber transducer in oil contamination monitoring. *Inst. Tech. Sens.* **11**, 3–4 (2006)
5. Wu, T.H., Mao, J.H., Wang, J.T.: A new on-line visual ferrograph. *Tribol. T.* **52**(5), 623–631 (2009)
6. Wu, T.H., Wang, J.Q., Peng, Y.P.: Description of wear debris from on-line Ferrograph image by their statistical color. *Tribol. Lubr. Technol.* **68**(11), 69–77 (2012)
7. Toms, A.M., Cassidy, K.: Filter debris analysis for aircraft engine and gearbox health management. *J. Fail. Anal. Prev.* **8**(2), 183–187 (2008)
8. Edmonds J., Resner M.S., Shkarlet K.: Detection of precursor wear debris in lubrication systems. In: 2000 IEEE Aerospace Conference Proceedings, pp. 73–77. IEEE, New York (2000)
9. Xu, C., Zhang, P.L., Wang, H.G., Li, Y.L., Lv, C.: Ultrasonic echo waveshape features extraction based on QPSO-matching pursuit for online wear debris discrimination. *Mech. Syst. Sig. Proc.* **60–61**, 301–315 (2015)
10. Kiziroglou, M. E., Wright, S.W., Yeatman, E.M.: Coil and core design for inductive energy receivers. *Sens. Actuators A: Phys.* **313**, 112206 (2020)
11. Itomi, S.: Oil condition sensor. U.S. Patent, 7151383 (2006)
12. Harvey, T.J., Morris, S., Wang, L., Wood, J.K.: Real-time monitoring of wear debris using electrostatic sensing techniques. *Proc. Inst. Mech. Eng. Part J: J. Eng. Tribol.* **221**(1), 27–40 (2007)
13. Chambers, K.W., Arneson, M.C., Waggoner, C.A.: An on-line ferromagnetic wear debris sensor for machinery condition monitoring and failure detection. *Wear* **128**(3), 325–337 (1988)
14. Gastops Products page, <https://30nbnb10r4n0173ti73ov30l-wpengine.netdna-ssl.com/wp-content/uploads/2020/05/c010024-ms4110-brochure.pdf>. Last accessed 10 May 2021
15. Wei, F., Minjie, C., Shizhong, H.: Oil on-line monitoring sensor technology. *Lubr. Eng.* **37**(1), 99–104 (2012)
16. Du, L., Zhe, J.: A high throughput inductive pulse sensor for online oil debris monitoring. *Tribol. Int.* **44**(2), 175–179 (2011)
17. Du, L., Zhu, X., Han, Y., Zhao, L., Zhe, J.: Improving sensitivity of an inductive pulse sensor for detection of metallic wear debris in lubricants using parallel LC resonance method. *Meas. Sci. Technol.* **24**(7), 075106 (2013)
18. Yan, H.Z., Zhang, Y.J.: The design of an on-line monitoring sensor of wear metal particles and the analysis of its characteristic. *Chinese J. Sens. Actuators* **4**, 333–338 (2002)
19. BeiJing Hangfeng Webpage, http://www.hangf.com/goods_show.aspx?id=113. Last accessed 10 May 2021
20. Ren, Y.J., Li, W., Zhao, G.F.: Inductive debris sensor using one energizing coil with multiple sensing coils for sensitivity improvement and high throughput. *Tribol. Int.* **128**(7), 96–103 (2018)

Regression Prediction of Performance Parameters in Ship Propulsion Equipment Simulation Model Based on One-Dimensional Convolutional Neural Network



Liangyuan Huang and Guoji Shen

Abstract Deep learning methods such as the one using Convolutional Neural Network (CNN) have made remarkable achievements in computer vision and natural language processing. Compared with the conventional neural network structures, CNN features low complexity, fewer parameters, and higher degree of nonlinearity. As the sizes of sensor signal input are often different from those of image input, using CNN to monitor the equipment status is a new issue compared with image recognition. To examine the impacts of various one-dimensional CNN structures on the regression of performance parameters, this paper conducts a preliminary study on the application of CNN in equipment status recognition, and utilizes published simulation datasets of ship propulsion equipment to train and test one-dimensional CNN models with different structures. The results show that the size of convolution kernels hinges on the attributes of input features when one-dimensional CNN is used for data regression prediction. In the case of independent and direct feature input, the training effect can be effectively improved by using 1×1 convolution kernels and the Network In Network (NIN) structure.

Keywords One-dimensional convolutional neural network · Ship propulsion equipment simulation model · Performance parameters · Regression prediction

1 Introduction

Based on the measurable information obtained on equipment, building corresponding physical or data model to achieve the quantitative real-time monitoring of the equipment running state and identify potential anomalies and faults and implement corresponding operation, can reduce the performance degradation and avoid dangerous

L. Huang (✉) · G. Shen
College of Intelligence Science and Technology, National University of Defense Technology,
Changsha 410073, China
e-mail: huangliangyuan@nudt.edu.cn

G. Shen
e-mail: shenguoji@nudt.edu.cn

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_27

315

situations [1], which is of great significance to ensure the reliability and safety of mechanical equipment system, especially large mechanical equipment. Due to the complex structure, strong coupling and limited measuring points, compared with the simple mechanical structure, large mechanical equipment puts forward higher requirements on the selection of model and measuring points, the feasibility of implementation scheme and the reduction of maintenance cost. Considering that it is difficult to construct accurate physical models for large mechanical equipment, most current researches adopt data-driven modeling methods, i.e., various machine learning models. ZhiQiang Chen et al. [2] used vibration time-domain signal indicators such as standard deviation and skewness and the RMS of each frequency band as feature input vectors, and constructed convolutional neural networks (CNN) to verify the fault diagnosis experiment of gearbox. Chen Lu et al. [3] used sparsity representation and data destruction in the iterative learning of the stacked denoising autoencoder method to obtain higher-order features with better robustness. Yaguo Lei et al. [4] used an unsupervised two-layer neural network as a sparse filter to directly learn features from mechanical vibration signals and then perform SoftMax regression. Its effectiveness was verified by the motor bearing data set and the locomotive bearing data set. Turker Ince et al. [5] used the motor current signal as the direct input of a one-dimensional convolutional neural network, which verified the effectiveness of the method for real-time monitoring of the motor state. Wei You et al. [6] proposed a hybrid model using convolutional neural networks and support vector machines for feature extraction and multi-classification tasks. Miao He et al. [7] built a large memory storage retrieval neural network and input the signal spectrum matrix obtained by short-time Fourier transform for training verification. Compared with other bearing fault diagnosis methods, it has better diagnostic performance at lower speeds. Wei Zhang et al. [8] proposed a new fault diagnosis model TICNN, which directly processes the original vibration signal without denoising preprocessing, and can still achieve high diagnostic accuracy when workload changing. Haidong Shao et al. [9] used a variety of autoencoders with different characteristics to build an integrated deep autoencoder for unsupervised feature learning of vibration signals. Ruonan Liu et al. [10] built a dislocated time series convolutional neural network structure by adding a dislocation layer, which can extract relevant features between signals at different intervals in periodic mechanical signals. Haidong Shao et al. [11] adopted compressed sensing to reduce the amount of data, and proposed an improved convolutional deep belief network structure based on compressed sensing for rolling bearing fault feature learning and fault diagnosis. Feng Jia et al. [12] proposed a deep normalized convolutional neural network for imbalanced fault classification of mechanical equipment. Jiedi Sun et al. [13] proposed to use nonlinear projection to achieve compressed acquisition and build a stacked sparse autoencoder network. Long Wen et al. [14] developed a signal-image conversion method to convert signals into two-dimensional images, and established a Lenet-5 based convolutional neural network for fault diagnosis, which has high diagnostic accuracy on motor bearing dataset, self-priming centrifugal pump dataset and axial piston hydraulic pump dataset. Wentao Mao et al. [15] proposed a deep output kernel learning method for collaborative diagnosis of multiple bearing fault types, which

improved the accuracy of bearing fault diagnosis and training efficiency. Liang Guo et al. [16] proposed a deep convolution transfer learning network, which realized the transfer learning of fault diagnosis through two modules of state recognition and domain adaptation. Zhuyun Chen et al. [17] proposed a fault diagnosis method combining convolutional neural network and extreme learning machine. However, at present, the existing research results mostly focus on the extraction and diagnosis of fault features of basic mechanical structures such as gears and bearings. Research on the selection of machine learning indicators and model construction for large-scale complex equipment is still to be developed.

As a continuation of machine learning, since Geoffrey E. Hinton [18] proposed the use of deep auto-encoding networks to initialize weights and unveiled the prelude of deep learning research, deep learning theory represented by Convolutional Neural Networks (CNN) has achieved considerable development and extensive, which most of the research on deep learning methods focuses on machine vision, natural language processing and image processing, such as the Alex-Net model proposed by Alex Krizhevsky et al. [19], the ResNet model proposed by Kaiming He et al. [20], and the cyclic continuous translation model combining CNN and RNN proposed by Nal Kalchbrenner et al. [21] Moreover, some researches on applying machine learning models to condition monitoring and fault diagnosis of mechanical equipment have also been launched [22]. Considering the strong adaptability and nonlinearity, low complexity, and good generalization ability of deep convolutional neural networks, convolutional neural networks are applied to the research of data-driven state monitoring and prediction of mechanical equipment. It is of great significance to simplify the network model structure and improve the calculation efficiency of the model.

The machine learning method provides a relatively simple tool for real-time condition monitoring and performance degradation prediction of mechanical equipment [23], without the need to build a physical simulation model of complex equipment, while the method of equipment maintenance has gradually changed from regular maintenance to preventive maintenance in the past few decades. Traditional machine learning methods such as backpropagation neural networks and support vector machines have achieved satisfactory results in previous studies. Despite the fact that deep learning method represented by Convolutional Neural Network (CNN) originated from image and language processing, it still be of significance to study its application in the field of equipment condition monitoring and prediction. This paper uses the ship propulsion simulation model data set disclosed in the UCI database to construct a one-dimensional CNN with different structures. By comparing the training effects of different structural models, the experience of applying one-dimensional CNN to equipment state prediction is analyzed.

1.1 Data Set

The data set is generated by the Simulink simulation model of the diesel electric and natural gas hybrid propulsion device of the frigate built by Andrea Coraddu et al.

Table 1 Main simulation output

Variable	Symbol	Units
Lever position	l_p	(\cdot)
Ship speed	v	Knot
Gas turbine shaft torque	GT_T	kN m
Gas turbine rate of revolutions	GT_n	r/min
Gas generator rate of revolutions	GG_n	r/min
Starboard propeller torque	T_s	kN
Port propeller torque	T_p	kN
HP turbine exit temperature	T_{48}	$^{\circ}\text{C}$
GT compressor inlet air temperature	T_1	$^{\circ}\text{C}$
GT compressor outlet air temperature	T_2	$^{\circ}\text{C}$
HP turbine exit pressure	P_{48}	bar
GT compressor inlet air pressure	P_1	bar
GT compressor outlet air pressure	P_2	bar
GT exhaust gas pressure	P_{exh}	bar
Turbine injection control	TIC	%
Fuel flow	m_f	kg/s

[24]. Three parameters of ship speed, GT Compressor decay state coefficient and GT Turbine decay state coefficient are used to describe the status of the propulsion device model.

Ship speed v : This parameter is controlled via the control lever and has a linear relationship with the position of the control lever.

GT Compressor decay state coefficient kM_c : This parameter describes the reduced value of airflow flow M_c and isentropic efficiency η_c over service hours.

GT Turbine decay state coefficient kM_t : This parameter describes the gas flow rate reduction factor over service hours.

Table 1 lists the main simulation outputs of the model, and the parameters of these outputs can be obtained through the ship's automation system in practical applications. HP and GT are the high pressure and the gas turbine.

Andrea Coraddu et al. describe the state space of the propulsion device through the form of a triple grid of ship speed and two attenuation coefficients, that is, 16 simulation outputs are obtained by inputting 3 different input parameters:

$$lp_i = i, i \in \{1, \dots, 9\} \quad (1)$$

$$kM_c^i = 1 - i \cdot 0.001, i \in \{0, 1, \dots, 50\} \quad (2)$$

$$kM_t^i = 1 - i \cdot 0.001, i \in \{0, 1, \dots, 50\} \quad (3)$$

The state space contains $9 \cdot 51 \cdot 26 = 11934$ conditions. A data set of 11,934 samples can be obtained after running the simulation model on the Simulink for each triple $[lp, kM_c, kM_t]_i$, which each sample contains the 16 characteristic values in Table 1.

Since the ship speed can be directly obtained in the actual situation, the parameters that need to be regression fitting are the GT compressor decay state coefficient kM_c and the GT turbine decay state coefficient kM_t . Same as the original author of the data set, we use 16 parameters obtained through simulation as input features and 2 attenuations as model output for machine learning model regression, while we use CNN network model for training.

1.2 One-Dimensional Convolutional Neural Network

The sensor signals used for asset status monitoring are usually one-dimensional data. That is to say, when CNN is applied to one-dimensional data, it uses one-dimensional convolution kernel, usually called one-dimensional CNN. The overall structure of a one-dimensional CNN usually includes input layer, convolutional layer, pooling layer, fully connected layer, as well as output layer, which is similar to a two-dimensional neural network. One dimension here refers to the data whose input dimension is $N \times 1$ or $1 \times N$, and its network structure is shown in Fig. 1.

In order to extract local features of an image, a CNN used for image recognition generally requires a certain size of convolution kernel. However, there is a special setting of a convolution kernel, that is, a convolution kernel with a size of 1×1 ,

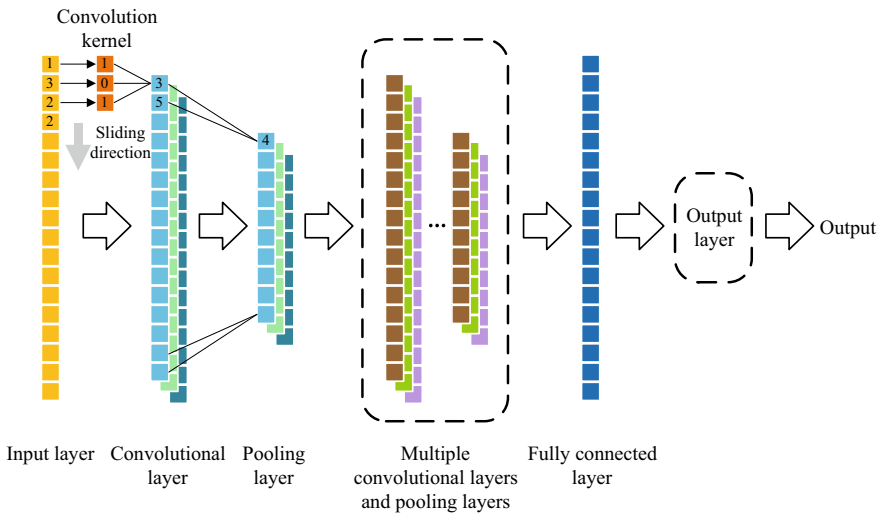


Fig. 1 Basic structure of one-dimensional CNN

which was first seen in the Network In Network (NIN) network structure proposed by Min Lin et al. [25] Aiming at the problem that the convolution operation of the convolutional layer is a linear operation and cannot extract nonlinear features, the NIN structure proposes to add a small neural network after each convolutional layer to improve the nonlinear feature extraction ability of the CNN. It is indicated that the combination of two sets of convolutional layers with a certain number of 1×1 convolution kernels have the characteristics of a small fully connected neural network, furthermore, it can achieve a deeper network structure with a small amount of calculation.

2 One-Dimensional Convolutional Neural Network Model Construction

This paper uses the Deep Learning Toolbox provided by MATLAB to build a one-dimensional convolutional neural network model.

2.1 Build a Variety of CNN Model Structures

With reference to the basic structure of the CNN, we built a CNN network model with a $[c_{1 \times 3}^{10} - p_{1 \times 2} - c_{1 \times 3}^{10} - p_{1 \times 2} - f_{10} - f_1]$ structure where $c_{1 \times 3}^{10}$ is convolutional layer with a convolution kernel of 10; $p_{1 \times 2}$ is pooling layer with a pooling area of 1×2 ; f_{10} and f_1 denote fully connected layers with 10 and 1 neurons, respectively. On the basis of the originally constructed convolutional neural network structure, the network structure is modified according to the regression influence of each network on the compressor decay prediction, which means the CNN model structure is gradually changed in the research process. The specific results and analysis will be reported in next chapter. All the constructed CNN structures are shown in Table 2.

Table 2 All the constructed CNN structures

Number	Structure
1	$[c_{1 \times 3}^{10} - p_{1 \times 2} - c_{1 \times 3}^{10} - p_{1 \times 2} - f_{10} - f_1]$
2	$[c_{1 \times 3}^{10} - \alpha_{ReLU} - c_{1 \times 3}^{10} - f_{10} - f_1]$
3	$[c_{1 \times 3}^{10} - f_{10} - \alpha_{ReLU} - c_{1 \times 3}^{10} - f_{10} - f_1]$
4	$[c_{1 \times 1}^{10} - f_{10} - \alpha_{ReLU} - c_{1 \times 1}^{10} - f_{10} - f_1]$
5	$[c_{1 \times 1}^{10} - f_{10} - \alpha_{ReLU} - c_{1 \times 1}^{10} - c_{1 \times 1}^{10} - f_{10} - f_1]$

2.2 Data Preprocessing

Before inputting the 16 eigenvalues into the network, firstly, the maximum and minimum normalization of the data was carried out and then the constant features were eliminated, which avoids the influence of the feature value range and constant features on the regression performance of the model. Correspondingly, the target values are also normalized to the maximum and minimum.

2.3 Model Training and Verification Parameter Settings

The prediction task is actually a regression task. This article adopts unified training parameter settings, and the training parameters are shown in Table 3. The validation set used in the training process is the training set itself.

In the research process of this paper, the CNN structure listed in Table 2 is used to train the sample set with only the change of GT compressor decay state coefficient. In this sample set, the GT turbine decay state coefficients are all 1, and the number of samples is 459. Then the sample set is still used to input the trained CNN model to test the regression prediction effect of the model.

For the CNN with the best regression prediction effect, the sample set was uniformly and randomly divided into training set and test set according to the ratio of 1/2, 1/3, 1/4, 1/5 and 1/8, so as to observe the regression effect when the training set accounted for different proportions. Likewise, the CNN structure with the best regression effect was used to train and test the sample set with only changes in the decay state coefficient of GT turbine.

Eventually the CNN structure is used to train all available samples and simultaneously predict the two parameters of GT compressor decay state coefficient and GT turbine decay state coefficient, according to different training set proportions.

Table 3 Unified training parameter settings

Training parameter	Value
Batch size	32
Initial learning rate	0.01
Verification frequency	50
Validation tolerance	20
The maximum number of iterations	1000

3 Results and Discussion

3.1 Regression of GT Compressor Decay State Coefficient

The visualization results of regression prediction on the sample set of the compressor attenuation single parameter change using the CNN structure No. 1 in Table 2 are shown in Fig. 2.

The performance of regression of attenuation coefficient of GT compressor using No. 1 CNN structure is not ideal, and the verified RMSE reached by the final training is 0.1445. The regression results are distinguished according to ship speed, as shown in Fig. 3, obviously abnormal output exists when ship speed is 3 and 6. In practice, the network model can be limited to a specific range to reduce the prediction output error. With this in mind, we excluded the samples with ship speed of 3 and 6, and the sample number of the sample set was 357.

After improving the CNN model structure several times according to the effect of each regression, the regression effect has been significantly improved. The results of the full sample training of the GT compressor attenuation decay state regression task with different CNN structures are shown in Fig. 4.

Significantly, there are two structures that have significantly improved the regression effect. Firstly, the pooling layer was cancelled and a nonlinear activation layer was added in the process of improving structure No. 1 to structure No. 2, reducing the verification RMSE from 0.1306 to 0.0526. Secondly, the size of the convolution kernel was changed from 1×3 to 1×1 in the process of improving structure No. 3 to structure No. 4, verifying that the RMSE was reduced from 0.0297 to 0.0140.

For the two obvious effect improvements, we infer that the reason may be the contradiction between the nature of the input features and the structure of the CNN model. The function of the $p_{1 \times 2}$ pooling layer is to summarize the features in multiple

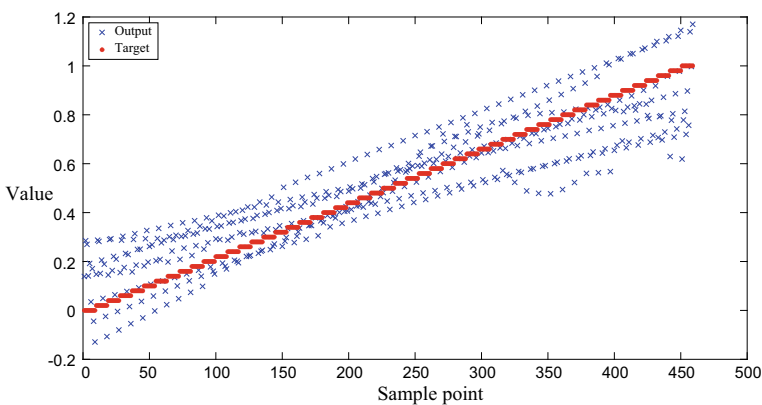


Fig. 2 Performance of No. 1 CNN structure in GT compressor attenuation coefficient regression task

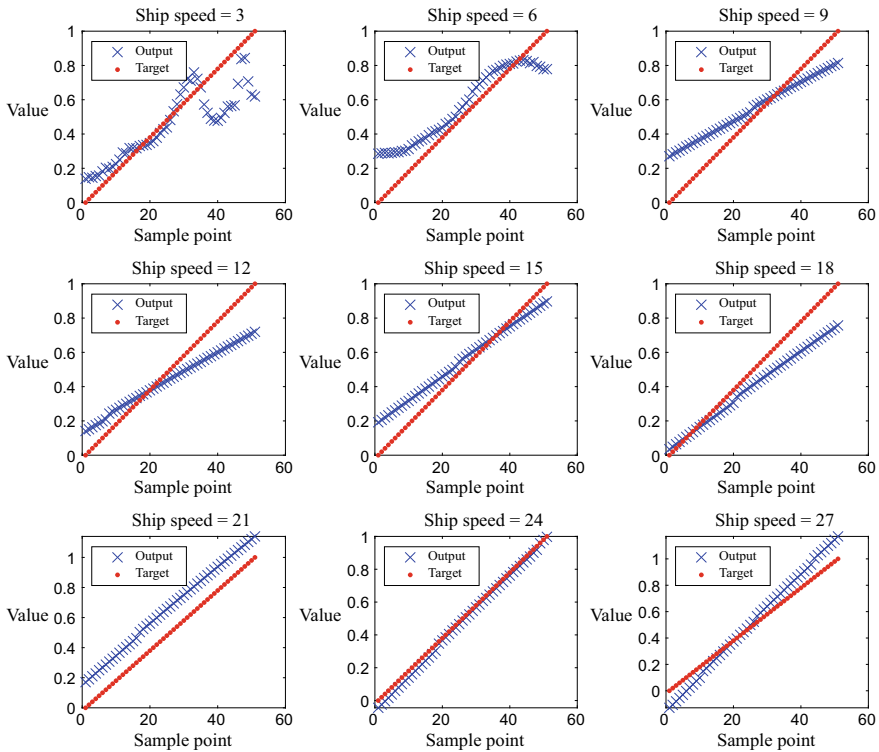


Fig. 3 Separate prediction results based on ship speed

1×2 regions, which is effective for eliminating noise information in the data. For the original vibration signal, the characteristics of the signal do not exist in a single sampling point, but are scattered and hidden in a set of vibration data. Such a situation of direct correlation between data points can be considered that the distribution of their characteristics in the original signal is loose, so it is necessary to extract the features contained in the signal data. However, in this paper, the data set is the characteristic data of each measuring point generated by the simulation model which differ from the original vibration signal data. It is the same one-dimensional data input but the data points are not directly related so can be regarded as a high-level summary of the characteristics represented. The use of the pooling layer may obscure adjacent features, resulting in the loss of necessary information after passing through the pooling layer. Similarly, the use of a 1×3 convolution kernel may introduce the influence of adjacent features during the convolution operation, and relatively independent functional inputs will be subject to further interference. The 1×1 convolution kernel avoids the influence of adjacent parameters in the convolution operation, and increases the nonlinearity of the structure to adapt to the nonlinear characteristics in the input data simultaneously.

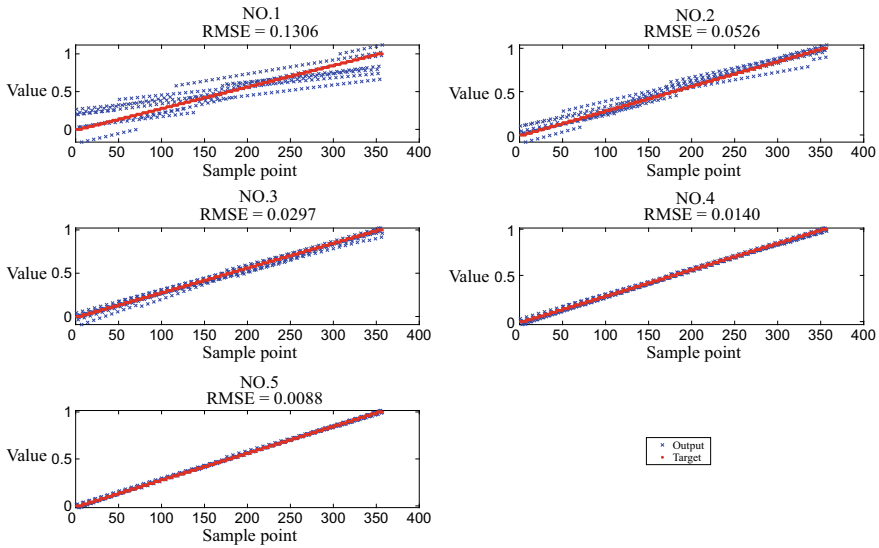


Fig. 4 The performance of different CNN structures in the regression task of GT compressor attenuation coefficient

Table 4 RMSE under different training set ratios in GT compressor decay state coefficient regression task (average of 5 times)

Proportion of training set	1/2	1/3	1/4	1/5	1/8
Verify RMSE	0.0199	0.0218	0.0255	0.0228	0.0384
Test RMSE	0.0214	0.0263	0.0284	0.0306	0.0549

The sample set is divided into training set and test set randomly according to different proportions, and multiple regression predictions are performed and the average results are shown in Table 4.

3.2 Regression of GT Turbine Decay State Coefficient

The number of samples used for the regression task of GT turbine decay state coefficient is 182 after removing the samples with ship speeds of 3 and 6. Table 5 shows the results of regression prediction of GT turbine decay state coefficient using the No. 5 structure CNN structure according to different training set ratios.

Due to the small number of training samples, the effect of GT decay regression prediction is more likely to be adversely affected by the proportion of smaller training sets.

Table 5 RMSE under different training set ratios in GT turbine decay state coefficient regression task (average of 5 times)

Proportion of training set	1/2	1/3	1/4	1/5	1/8
Verify RMSE	0.0223	0.0573	0.0381	0.0902	0.0459
Test RMSE	0.0255	0.0661	0.0600	0.1558	0.2350

Table 6 RMSE under different training set ratios in GT compressor decay state coefficient and GT turbine decay state coefficient regression task (average of 5 times)

Proportion of training set	1/2	1/3	1/4	1/5	1/8
Verify RMSE	0.0520	0.0453	0.0502	0.0452	0.0682
Test RMSE	0.0368	0.0320	0.0360	0.0323	0.0488

3.3 Double Decay State Coefficient Regression

In order to meet the needs of the task of double-decay regression prediction using the No. 5 structure, the final fully connected layer needs to be adjusted, specifically, the number of neurons should be modified to 2. The revised structure is $[c_{1 \times 1}^{10} - f_{10} - \alpha_{ReLU} - c_{1 \times 1}^{10} - c_{1 \times 1}^{10} - f_{10} - f_2]$. After excluding samples with ship speeds of 3 and 6, the number of samples is 9282. The regression prediction results are shown in Table 6.

It can be found that when the sample size is much larger than the single attenuation coefficient regression, there is no significant difference in the effect of the double attenuation coefficient regression task under the five training set ratios, which corresponds to the practical experience and characteristics of the CNN model that depends on the number of training sets.

3.4 Discussion

In this work, by constructing and improving the CNN model structure, the influence of different network layer characteristics on the regression task of ship propulsion model simulation data is studied. The simulation data set of the ship propulsion equipment used has relatively independent input characteristics.

The results show that when one-dimensional CNN is used for data regression prediction tasks, whether to use the pooling layer and the size of the convolution kernel used should depend on the nature of the input features. In contrast to this paper, when the original vibration signal is directly used as the input, the signal's fault characteristics are hidden in the signal data sequence, and the sampling points are directly related in the time domain, which is an indirect feature input. Regional feature extraction can be performed through a certain size of convolution kernel to help identify fault features.

As to the regression task studied in this paper, the input is standardized parameters. In practical applications, the value of the parameter is the value measured by each sensor. Therefore, the input is a direct feature which means there is no direct correlation between the parameters. Using convolutional layers and pooling layers for CNN models of this type of input will have a negative effect. The introduction of the 1×1 convolution kernel not only avoid the influence of adjacent parameters in the convolution operation, but also increase the nonlinearity of the one-dimensional CNN structure to adapt to the nonlinear characteristics of the input data.

References

1. Gao, Z., Cecati, C., Ding, S.X.: A survey of fault diagnosis and fault-tolerant techniques—Part I: fault diagnosis with model-based and signal-based approaches. *IEEE Trans. Industr. Electron.* **62**(6), 3757–3767 (2015)
2. Chen, Z., Li, C., Sanchez, R.V.: Gearbox fault identification and classification with convolutional neural networks. *Shock. Vib.* **2015**, 1–10 (2015)
3. Lu, C., Wang, Z.Y., Qin, W.L., et al.: Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Sig. Process.* **130**, 377–388 (2017)
4. Lei, Y., Jia, F., Lin, J., et al.: An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Trans. Industr. Electron.* **63**(5), 3137–3147 (2016)
5. Ince, T., Kiranyaz, S., Eren, L., et al.: Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Industr. Electron.* **63**(11), 7067–7075 (2016)
6. You, W., Shen, C., Guo, X., et al.: A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery. *Adv. Mech. Eng.* **9**(6), 1687814017704146 (2017)
7. He, M., He, D.: Deep learning based approach for bearing fault diagnosis. *IEEE Trans. Ind. Appl.* **53**(3), 3057–3065 (2017)
8. Zhang, W., Li, C., Peng, G., et al.: A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Sig. Process.* **100**, 439–453 (2018)
9. Shao, H., Jiang, H., Lin, Y., et al.: A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mech. Syst. Sig. Process.* **102**, 278–297 (2018)
10. Liu, R., Meng, G., Yang, B., et al.: Dislocated time series convolutional neural architecture: an intelligent fault diagnosis approach for electric machine. *IEEE Trans. Industr. Inf.* **13**(3), 1310–1320 (2017)
11. Shao, H., Jiang, H., Zhang, H., et al.: Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. *Mech. Syst. Sig. Process.* **100**, 743–765 (2018)
12. Jia, F., Lei, Y., Lu, N., et al.: Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mech. Syst. Sig. Process.* **110**, 349–367 (2018)
13. Sun, J., Yan, C., Wen, J.: Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Trans. Instrum. Meas.* **67**(1), 185–195 (2018)
14. Wen, L., Li, X., Gao, L., et al.: A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Industr. Electron.* **65**(7), 5990–5998 (2018)
15. Mao, W., Feng, W., Liang, X.: A novel deep output kernel learning method for bearing fault structural diagnosis. *Mech. Syst. Sig. Process.* **117**, 293–318 (2019)

16. Guo, L., Lei, Y., Xing, S., et al.: Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Trans. Industr. Electron.* **66**(9), 7316–7325 (2019)
17. Chen, Z., Gryllias, K., Li, W.: Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. *Mech. Syst. Sig. Process.* **133**, 106272 (2019)
18. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017)
20. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
21. Kalchbrenner, N., Blunsom, P.: Recurrent continuous translation models. In: 2013 Conference on Empirical Methods in Natural Language Processing (2013)
22. Nangolo, F.N., Kimera, D.: Maintenance practices and parameters for marine mechanical systems: a review. *J. Qual. Maint. Eng.* **26**(3), 459–488 (2019)
23. Coraddu, A., Oneto, L., Ghio, A., et al.: Machine learning approaches for improving condition-based maintenance of naval propulsion plants. *Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ.* (2014)
24. Carvalho, T.P., et al.: A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* **137**(106024) (2019)
25. Lin, M., Chen, Q., Yan, S.: Network In Network. *Comput. Sci.* (2013)

Research on Numerical Simulation Study of Solidification Heat Transfer Coefficient of Extra-Wide Slab



Xinxia Qi, Qi Jia, and Wenxiong Wu

Abstract Because the extra-wide slab continuous caster has larger sectional width, heat emission condition is different between the length and width. Therefore, it is inevitable that the water distribution is regional separation along the casting slab width and length. It brings a lot of difficulties to determine process parameters and control continuous casting process, and causes quality defect. Based on 3250 mm × 150 mm extra-wide slab casting and production process, the paper established the finite element model, calculates the temperature distribution and shell thickness of the slab with different casting speed and superheat, to study influence of process control parameters of wide slab quality.

Keywords Extra-wide slab · Solidification and heat transfer · Numerical simulation

1 Technical Characteristics and Related Parameters of the Casting Machine

The extra-wide slab continuous caster has a straight arc with a radius of 6.67 m. It is multi-point bending and multi-point straightening. The secondary cooling zone of it includes foot roller, 9 secondary cooling sections, water distribution in vertical and horizontal direction, fansement 0–8. Plate mould size: 3450 mm × 950 mm. The whole process is non-oxidized pouring.

X. Qi

School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, Guangdong, China

e-mail: 1410134965@qq.com

Q. Jia

Beijing Infrastructure Investment Co., LTD, Beijing 100101, China

W. Wu (✉)

School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China

e-mail: 895386202@qq.com

2 Establishment of Solidification Heat Transfer Model

2.1 Fundamental Assumption

The following assumptions about the model were made [1–3]:

(1) After the casting speed is stable, the heat transfer condition won't change with time. (2) The heat transfer conditions of inner arc and outer arc of casting billet are similar. (3) The heat transfer caused by the temperature gradient along the casting billet direction is very small, so it can be ignored. (4) Thermal physical properties of steel are only related to temperature. (5) The radiation heat transfer of the surface of the second cold casting billet, the contact heat transfer with the support roller and the cooling heat transfer of the second cold water are considered by the comprehensive heat transfer coefficient. (6) The effect of vibration on solidification heat transfer is ignored. According to the assumed conditions, the solidification heat transfer equation is simplified into a two-dimensional unsteady heat conduction Eq. (1):

$$\rho c_p \frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left(k \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) + q_v \quad (1)$$

in the Eq. (1) ρ : density of liquid steel/kg m⁻³; c_p : heat capacity/J kg K⁻¹; k : thermal conductivity/W m⁻¹ K⁻¹; q_v : internal heat source/J m⁻³ K⁻¹; T : temperature/°C; t : time/s.

2.2 Selection of Physical Parameters and Boundary Conditions

- (1) Physical parameters. This paper uses the simulated steel grade Q460 and calculate liquidus temperature and solid phase temperature according to the empirical formula [4]. The formula $\lambda_0 = 69.78 - 10.12 C - 16.75 Mn - 33.73 Si$ is used to calculate the thermal conductivity of Q460, and the thermal conductivity of Q460 is $\lambda_0 = 31.13 \text{ W m}^{-1} \text{ }^\circ\text{C}^{-1}$; Comprehensive thermal conductivity λ_{eff} is used in the liquid phase zone from 400 to 1200 °C. When $T \leq T_S$, $\lambda_{eff} = \lambda_0$; When $T \geq T_L$, $\lambda_{eff} = 7 \lambda_0$; When $T_S < T < T_L$ †, $\lambda_{eff} = 5 \lambda_0$.

The equivalent specific heat was used to replace the specific heat of steel in the two-phase region. The specific heat and latent heat of solidification of Q460 which is below 1300 °C are similar to values of 1.5%Mn steel. By extrapolating or interpolating the obtained data, this paper obtains the data at the liquidus temperature and solid phase temperature.

$$C_L = 837 \text{ J kg}^{-1} \text{ }^\circ\text{C}^{-1}, C_S = 715.83 \text{ J kg}^{-1} \text{ }^\circ\text{C}^{-1}.$$

Table 1 Mold cooling parameters

	Right	left	Inner arc	Outer arc
q_m	$640-55.1 \sqrt{t}$	$640-53.7 \sqrt{t}$	$640-54.8 \sqrt{t}$	$640-54.8 \sqrt{t}$

(2) Boundary condition: value of the average heat flux density of the mold at the production site is put into the Eq. (2):

$$q_m = (640 - \beta\sqrt{t}) \times 4.187 \times 10^3 \tag{2}$$

t: The corresponding time of the casting billet at a certain position in the mold/s. The heat flux of this simulation is shown in Table 1. The instantaneous heat flux of the mold is loaded into the model as the boundary condition.

According to the onsite process situation, this paper calculates the slab area, water flow rate, water flow density and wide surface heat transfer coefficient of foot roll area and 0–8 section. Radiative heat transfer in the air-cooled zone is treated as the third type of boundary condition, and this paper calculates the comprehensive heat transfer coefficient $h = 0.31/\text{kW m}^{-2} \text{ k}^{-1}$.

3 The Results and Analysis of Simulation

Based on the assumption of heat transfer symmetry between inner and outer arcs, a finite element model was established by taking 1/4 cross section of the casting billet to simulate the temperature field of the solidification process of the wide slab, and the model was verified by actual measurement of the surface temperature of the casting billet.

3.1 Research on Temperature Distribution of Upper Surface Center of Casting Billet

When the casting speed is 1.10, 1.20, 1.25 and 1.30 m min^{-1} , and the casting temperature is 1530 °C, the temperature change curve of the center of the upper surface of the simulated extra-wide slab is shown in Fig. 1. The simulated values and the actual measured values of the infrared instrument are shown in Table 2.

It can be seen from Table 2 that the error range between the simulated temperature and the measured temperature is less than 2.20%, indicating that the simulation results are valid.

According to Fig. 1 and Table 2, it can be seen that: (1) When the casting speed increases from 1.10 to 1.30 m min^{-1} , the straightening temperature of the upper surface center will increase from 957 to 1030 °C. And the temperature at the upper

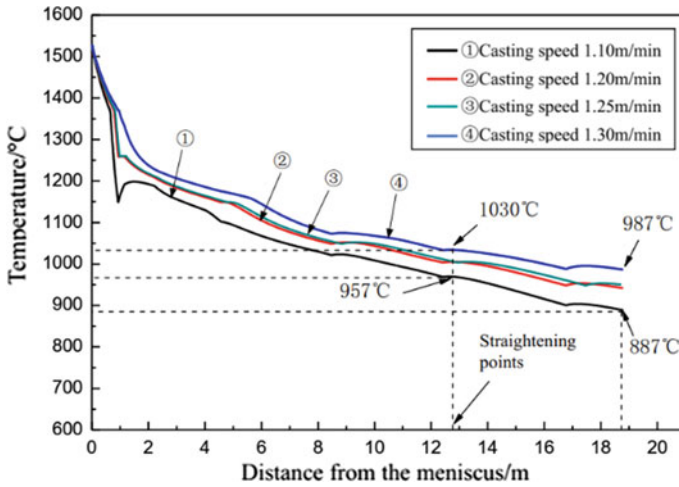


Fig. 1 Temperature distribution at the center of upper surface of the slab at different casting speeds

Table 2 Temperature distribution at the upper surface of the casting billet at different casting speeds

	Casting speed/m min ⁻¹	1.10	1.20	1.25	1.30
Straightening temperature/°C	Simulation	957	983.4	1005	1030
	Actual measurement	942	965	985	1011
Temperature of the billet/°C	Simulation	887	931	953	987
	Actual measurement	/	922	932	/

surface center of the casting billet will increase from 887 to 987 °C, when the billet is ejected. In other words, for each 0.10 m min⁻¹ increase in the casting speed, the surface temperature of the casting billet in the straightening section will increase by about 36.5 °C and the straightening temperature of the upper surface center will increase by about 50 °C. That means the casting speed has a great effect on the surface temperature of the billet. (2) Only when the casting speed is greater than 1.20 m min⁻¹, can the straightening temperature be greater than 960 °C and the occurrence probability of transverse cracks on the surface of the casting billet be reduced.

It can be concluded that the temperature of casting billet when it was ejected can be increased obviously by increasing the casting speed. Appropriate increase of the drawing speed is beneficial to increase the surface temperature of the extra-wide slab when it be straightened.

3.2 Simulation of the Temperature Field Distribution

(1) Temperature field distribution at the outlet of mold

The temperature distribution at the outlet of the wide slab mold at different casting speeds is shown in Fig. 2, and the temperature of molten steel is simulated at 1530 °C. When the casting speed is 1.10, 1.20 and 1.30 m min⁻¹, the temperature distribution corresponds to (a), (b) and (c) in Fig. 2, and the shell thicknesses corresponds to 16.4 mm, 14.5 mm and 11.6 mm, respectively.

According to Fig. 2, it can be seen that: (1) The shell thickness is reduced, with the increase of casting speed [5]. (2) The thickness of the billet shell will be reduced by 2.4 mm on average when the casting speed is increased by 0.1 m min⁻¹.

(2) Temperature field distribution at the straightening section

In the straightening section of the slab at different casting speeds, the temperature distribution is shown in Fig. 3, and the molten steel temperature is simulated at 1530 °C. When the casting speed is 1.10, 1.25, and 1.30 m min⁻¹, the temperature distribution corresponds to (a), (b) and (c) in Fig. 3, respectively.

According to Fig. 3, the solidification of the casting billet at the straightening point is: (a) When the casting speed is 1.10 m min⁻¹, the casting billet is completely solidified; (b) When the casting speed is 1.25 m min⁻¹, the casting billet is basically solidified, and there is a small amount of paste area; (c) When the casting speed is 1.30 m min⁻¹, the solidification is not complete, and there is a small liquid core.

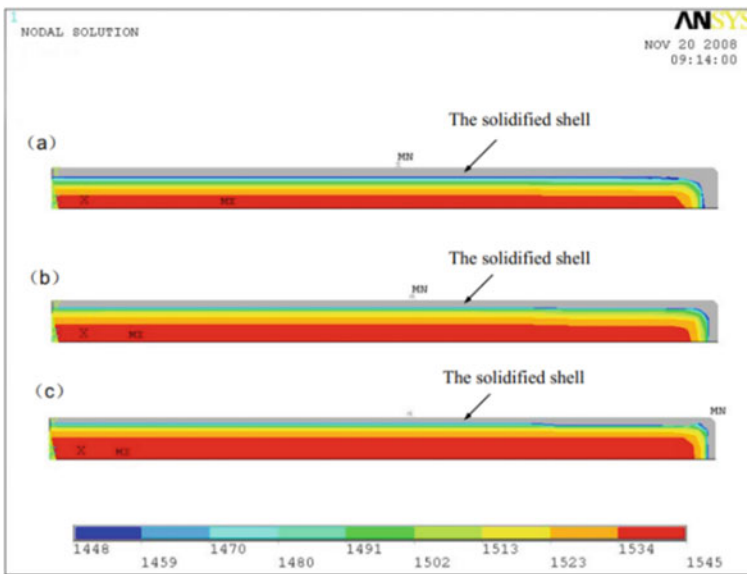


Fig. 2 Temperature distribution diagram of mold outlet

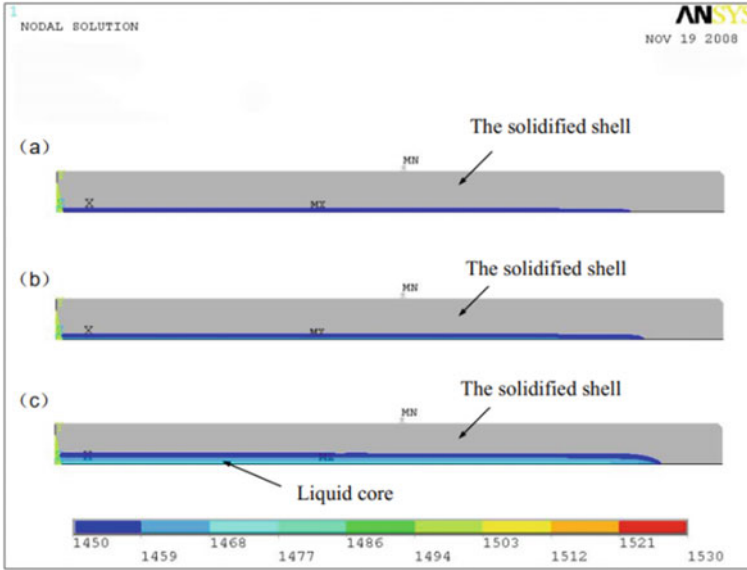


Fig. 3 Temperature distribution of straightening section

3.3 Simulation of Temperature Distribution When the Billet Speed is 1.20 m min^{-1}

When the casting speed is 1.20 m min^{-1} and the casting temperature is $1530 \text{ }^\circ\text{C}$, the temperature distribution of the upper surface center and core of the wide slab is shown in Fig. 4.

According to Fig. 4, when the casting speed is 1.20 m min^{-1} , at the straightening point, the center temperature of the upper surface of the wide slab is $983.40 \text{ }^\circ\text{C}$, the center temperature is $1446 \text{ }^\circ\text{C}$ and the casting billet temperature is $931 \text{ }^\circ\text{C}$.

3.4 Simulation of Temperature Distribution in the Center of Casting Billet at Different Casting Speeds

When the casting speed is $1.10, 1.20, 1.25$ and 1.30 m min^{-1} , and the casting temperature is $1530 \text{ }^\circ\text{C}$, the temperature distribution curve of the central part of the wide slab is shown in Fig. 5.

According to Fig. 5, it can be seen that: (1) when the casting speed is 1.10 and 1.20 m min^{-1} , the solidification end point is less than 12.8 m , the straightening is under full solidified situation, which can avoid cracks inside the casting billet. When the casting speed is faster than 1.25 m min^{-1} , the solidification end point is greater than 12.8 m , and the liquid core appears when straightened. (2) With the increase

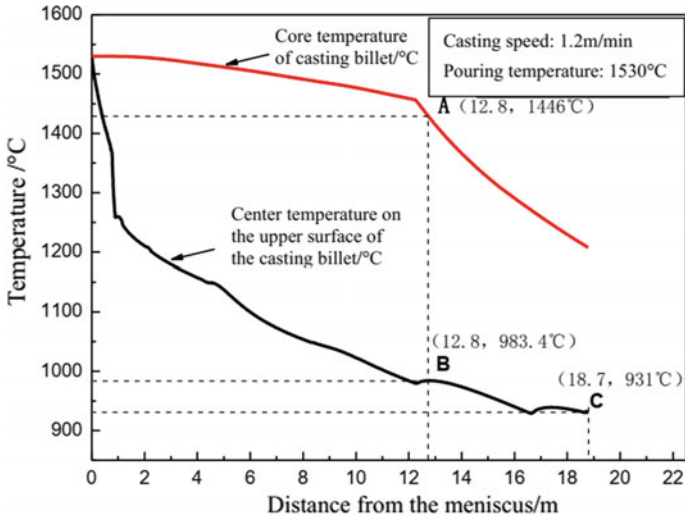


Fig. 4 Temperature distribution when the casting speed is 1.20 m min^{-1}

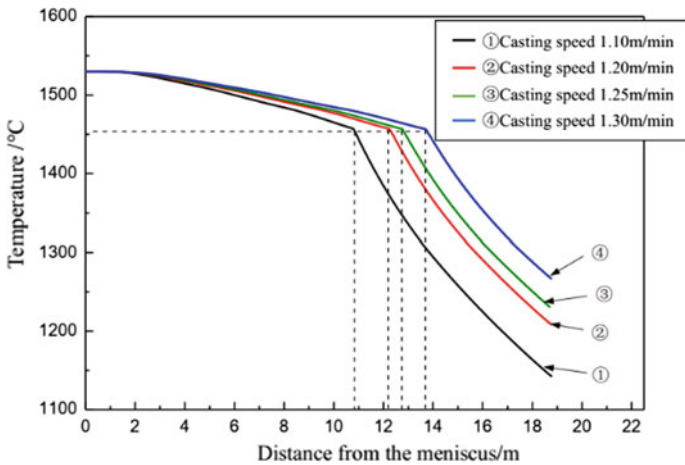


Fig. 5 Center temperature distribution at different casting speeds

of the casting speed, the solidification end of the wide slab moves backward, and the wide slab is changed from full solidification to straightening with liquid core, which increases the probability of inside cracks and increases the degree of central segregation [6, 7].

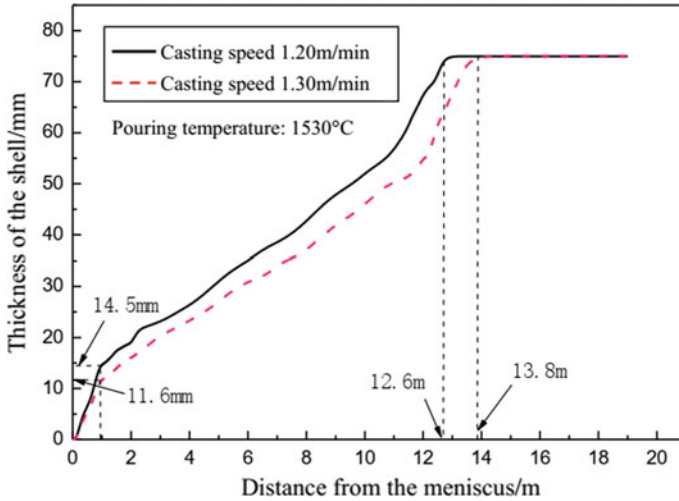


Fig. 6 Shell thickness distribution at different casting speeds

3.5 Simulation of Thickness Distribution of Billet Shell

When the casting speed is 1.20 and 1.30 m min^{-1} , and the casting temperature is 1530 $^{\circ}\text{C}$, the thickness change curve of the billet shell is shown in Fig. 6.

According to Fig. 6, when the casting speed increases from 1.22 to 1.30 m min^{-1} , the length of liquid core will increase from 12.6 to 13.8 m and the thickness of the billet shell when it's gets out of the mould will decrease from 14.5 to 11.6 mm. That means when the casting speed increases by 0.1 m min^{-1} , the length of the liquid core will increase by 1.2 m and the thickness of the billet shell out of the mould will decrease by 2.9 mm.

3.6 Simulation Results of Different Superheat Degree

When the casting speed is 1.20 m min^{-1} and the superheat of the molten steel is 10, 15, 20 and 25 $^{\circ}\text{C}$, the temperature change curve of the upper surface center of the extra-wide slab is shown in Fig. 7. Change curve of center temperature of casting billet is shown in Fig. 8.

According to Fig. 7, when the casting speed remains unchanged and the superheat increases from 10 to 25 $^{\circ}\text{C}$, the billet temperature of the upper surface center at the straightening point will increase from 996 to 1022 $^{\circ}\text{C}$. That means the billet temperature of the upper surface center at the straightening point will increase by 1.73 $^{\circ}\text{C}$ for each 1 $^{\circ}\text{C}$ superheat increase. It can be seen that the influence of superheat

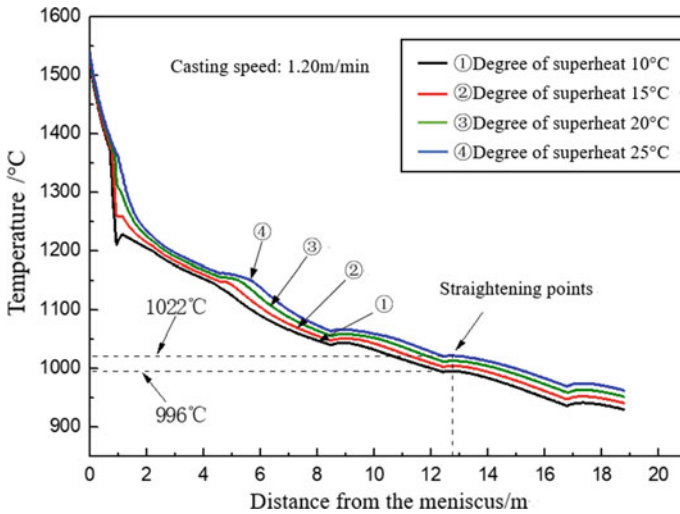


Fig. 7 Distribution of upper surface center of the wide slab

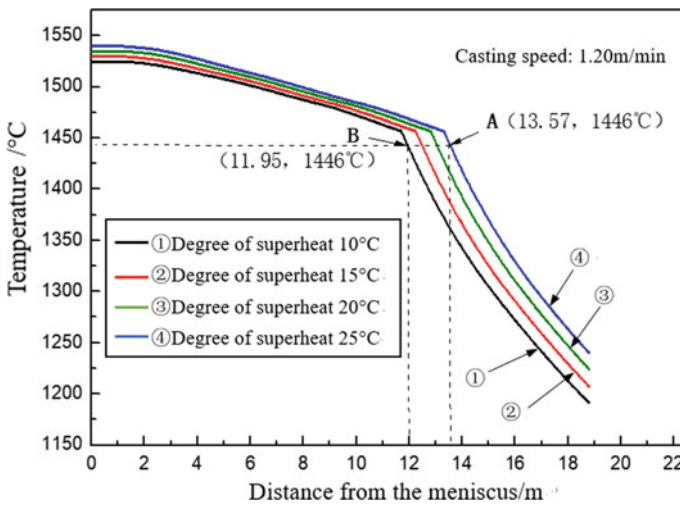


Fig. 8 Distribution of central temperature of the wide slab

change on the surface temperature of the casting billet is much smaller than that of the casting speed.

According to Fig. 8, it can be seen that: (1) When the casting speed is 1.20 m min^{-1} and the superheat of molten steel increases from 10 to 25 °C, the length of the liquid core will increase from 11.95 to 13.57 m, and the way to straighten the casting billet will change from full solidification to straightening with the liquid core. Every 1 °C increase in the superheat is equivalent to extending the length of the liquid core by

0.11 m. (2) When the casting speed is 1.20 m min^{-1} and the superheat is $20 \text{ }^\circ\text{C}$, the length of the liquid core will be greater than 12.8 m and the casting billet will be straightened with the liquid core; When the superheat is $15 \text{ }^\circ\text{C}$, the casting billet will be straightened with the full solidification. (3) With the increase of superheat, the solidification end will move backward, and the center of the casting billet will be loose and the center segregation will be aggravated.

4 Conclusion

The simulation results show that:

- (1) The process control range of the extra-wide slab continuous caster is very narrow. When the casting speed ranges from 1.20 to 1.25 m min^{-1} and the superheat exceeds $20 \text{ }^\circ\text{C}$, the casting billet will be straightened with liquid core. The internal center segregation and center porosity are serious.
- (2) With the increase of the casting speed, the shell of the billet when it's gets out of the mould will become thinner, and the thickness of the shell decreases by 2.4 mm for each 0.1 m min^{-1} increase of the casting speed.
- (3) When the drawing speed is 1.10 m min^{-1} and the superheat is $15 \text{ }^\circ\text{C}$, the straightening temperature of the upper surface of the wide slab will be less than $950 \text{ }^\circ\text{C}$, and the transverse crack will be easy to occur on the surface of the wide slab.
- (4) Considering the surface quality and internal quality of the casting billet, we suggested that the casting speed of the wide slab continuous caster for producing microalloy Q460 steel should be $1.20\text{--}1.25 \text{ m min}^{-1}$, and the superheat should be less than $20 \text{ }^\circ\text{C}$. At this time, the casting billet may still be straightened with liquid core, and the center segregation and central porosity may still be serious.
- (5) Through the simulation study on the solidification heat transfer of the wide slab, this paper has found the variation rules of temperature, shell thickness and core length of wide slab under different conditions of continuous casting process, and the process measures to improve the quality of wide slab, which provided a theoretical basis for the determination of the process parameters of continuous casting.

References

1. Liu, K., Sun, Q., Wang, L., Zhang, J.: Optimum total soft reduction amount for continuous casting slab with different thickness. *J. Iron Steel Res.* **29**(02), 111–116 (2017)
2. Deng, J., Wang, R.: Mechanical properties of casting slab of high strength cold rolled steel 1500MS for auto at high temperature. *Spec. Steel* **39**(01), 54–56 (2018)

3. Zhong, G., Wu, C., Ni, J., Du, W., Sun, J., Zhai, Q.: Thermal simulation of solidification process of Cr-saved ferritic stainless steel in continuous casting process. *J. Iron Steel Res.* **27**(05), 30–34 (2015)
4. Wang, L., Sun, Y., Niu, A., Li, Y., Chen, C.: Numerical simulation of heat transfer and solidification in X80 slab. *Iron Steel Vanadium Titanium* **39**(06), 143–149 (2018)
5. Shen, Y., Bai, M., Gao, P., Chen, J.: Numerical simulation of heat-transfer and strain during solidification in ultra-thick rectangular slab with different shape of corner. *Iron Steel* **51**(04), 38–43 (2016)
6. Qian, H., Du, C., Hu, P., Zhang, J., Liao, H.: Free thermal shrinkage behavior during the solidification process in slab casters. *J. Iron Steel Res.* **28**(08), 27–32 (2016)
7. Xie, J., Luo, G., Liu, L., Wang, C.: Numerical simulation calculation of Solidification and heat transfer behavior of steel Q235B slab in continuous casting. *Spec. Steel* **41**(02), 10–14 (2020)

Numerical Study of Motor Electrical Signature for Condition Monitoring of Gear Tooth Breakage in a Motor-Gear System



Funso Otuyemi, Xiuquan Sun, Fengshou Gu, and Andrew D. Ball

Abstract Gearboxes are the most significant components of an electromechanical system. They are often exposed to various abnormal working condition which causes damage to the gear. These damages can be of any form such as tooth breakage, tooth wear. Furthermore, this breakage will cause changes in gear tooth mesh stiffness, thereby reducing the gear dynamics. Recently, several attempts have been performed for the detection of localized gear tooth faults using electric signatures from induction motor with promising results. However, the interaction between the motor-gear dynamics to detect and diagnose this fault has not been fully investigated. Therefore, this study proposed a 6° of freedom (DOF) electrical motor model integrated with an 18° of freedom (DOF) gear dynamic model to fulfil the detection of gear faults using motor current signature. In this model, a comparison of electric torque and constant torque as the input for the gear model has been investigated to study the influence of electric motor torque on gear dynamic. Furthermore, this study considered different severities of tooth breakage to demonstrate the performance of motor current in gear tooth breakage detection and its location. The numerical results show an increase in amplitudes at the frequency of f_s , f_{r1} and f_s , f_{r2} as the severity of the fault increases at different stage of the gearbox which can reflect the presence of tooth breakage. This proposed numerical analysis does have a good agreement with the experimental validation.

F. Otuyemi · X. Sun · F. Gu (✉) · A. D. Ball
Centre for Efficiency and Performance Engineering (CEPE), University of Huddersfield,
Huddersfield HD1 3DH, UK
e-mail: f.gu@hud.ac.uk

F. Otuyemi
e-mail: funso.otuyemi@hud.ac.uk

X. Sun
e-mail: xiuquan.sun@hud.ac.uk

A. D. Ball
e-mail: a.d.ball@hud.ac.uk

Keywords Condition monitoring · Dynamic behaviour · Induction motor · Electric torque · Motor current signal analysis · Tooth breakage

1 Introduction

Gearboxes play a critical role in electromechanical systems because they are often used in the transmission of power and changing the speed of motion. In most electromechanical systems, gearbox form a part of the mechanical load housing that is coupled to an electrical drive which usually is an induction motor. However, they are susceptible to faults thereby causing discontinuity in production schedules in industries resulting to lower productivity or unscheduled downtime. Failures in gearboxes can be tooth-related such as pitting, spalling, crack or broken tooth, which can take place frequently and lead to a complete failure of the gearbox. One of the most common causes of gear failure is gear tooth fatigue due to rotational or cyclic loading. Furthermore, resulting in progressive damage to gear teeth and ultimately leads to the complete failure of the gear. Hence, the ability for early detection and diagnosis of gear tooth fatigue crack has always been one of the major technical challenges. When such faults occur, additional mechanical impacts are generated at the rotational frequency in the vibration signal. The shape of mechanical impacts is related to the mechanical structure resonance excited by the tooth localized fault when the damaged tooth is engaged. This is often caused by inadequate lubrication and extreme working conditions. However, to improve the current techniques of gearbox vibration monitoring and diagnosis, many researchers have worked on the gear dynamic modelling to ascertain the effect of tooth breakage on gears [1].

Bartelmus [2] did a mathematical modelling and computer simulation to gearbox dynamic examinations. Shao et al. [3] established three-dimensional finite element models of gears with tooth breakage and investigated the effects of breakage position and length on the dynamic features. Bai [4] proposed a method to detect and locate faults in a gear tooth by putting into account the whole tooth width or the whole tooth thickness which causes a reduction in gear tooth stiffness resulting in a more significant dynamic impact. Chen and Shao [5] proposed a calculation method of the mesh stiffness for a gear with tooth breakage, the effects of the tooth breakage on the gear dynamic response are assessed using two statistical indicator such as RMS and kurtosis.

Few research has focused on the dynamic mechanism of gear system with failures using a time varying motor electric torque as an input for the gear dynamics for detection and diagnosis. However, the research on gear tooth breakage and dynamic models of a gear system in this paper combined with experimental gearbox can supply a theoretical basis to fault diagnosis of gearbox.

This study proposed 6-DOF induction motor model integrated with 18-DOF gear dynamic model to investigate the behaviours of the gear system with tooth defects. Hence, sidebands relating to fault signatures are used in detection. Also, you will

know which stage of the gearbox there is fault, thereby making detection and diagnosis much easier to detect. The robustness and effectiveness of the proposed technique to detect gearbox faults under different operating conditions are demonstrated through both model simulations and experiments.

2 Dynamic Model of a Motor-Gear System

In this section, the dynamic characteristics of the electromechanical coupling system are attained through simulation by using a combined motor and gear model. In this motor-gear system, the driving torque is applied on the gear system and the power is transmitted from the motor to the gear system through the shaft as shown in Fig. 2. Therefore, the rotational and translational displacements of the motor and each of the gear are chosen as the generalized coordinates in the dynamic model of the motor-gear system.

2.1 Induction Motor Dynamic Model

In order to model the dynamic of an induction motor, a simple 6-DOF was put forward. Deriving this model equations can be generated from the $dq0$ equivalent of the induction motor. Therefore, all the voltages, currents and flux linkages between the stationary stator as well as the moving rotor are modelled. According to Aktaibi and Ratnani [7], the $dq0$ flux linkage and current equations were implemented and can be demonstrated below as:

$$\begin{cases} \frac{d\psi_{qs}}{dt} = \omega_b \left[v_{qs} + \frac{\omega_e}{\omega_b} \psi_{ds} + \frac{R_s}{X_{ls}} (\psi_{md} - \psi_{ds}) \right] \\ \frac{d\psi_{ds}}{dt} = \omega_b \left[v_{ds} + \frac{\omega_e}{\omega_b} \psi_{qs} + \frac{R_s}{X_{ls}} (\psi_{md} - \psi_{ds}) \right] \\ \frac{d\psi_{0s}}{dt} = \omega_b \left[v_{\theta s} - \frac{R_s}{X_{ls}} \psi_{\theta s} \right] \\ \frac{d\psi_{qr}}{dt} = \omega_b \left[v_{qr} - \frac{(\omega_e - \omega_r)}{\omega_b} \psi_{dr} + \frac{R_r}{X_{lr}} (\psi_{mq} - \psi_{qr}) \right] \\ \frac{d\psi_{dr}}{dt} = \omega_b \left[v_{dr} - \frac{(\omega_e - \omega_r)}{\omega_b} \psi_{qr} + \frac{R_r}{X_{lr}} (\psi_{mq} - \psi_{dr}) \right] \\ \frac{d\psi_{0r}}{dt} = \omega_b \left[v_{\theta r} - \frac{R_r}{X_{lr}} \psi_{\theta r} \right] \end{cases} \tag{1}$$

where subscript $d, q, 0$ denotes the axis of the resolving magnetic motive force of the three-phase; subscript s and r represent both stator and rotor reference frame; $\psi_{qs} \psi_{ds} \psi_{0s}$ is the magnetic flux linkage of q -axis, d -axis and 0 -axis components of the stator; $\psi_{qr} \psi_{dr} \psi_{0r}$ is the magnetic flux linkage of q -axis, d -axis and 0 -axis components of the rotor; R_s is the stator resistance; R_r is the rotor resistance; ω represent the angular speed of the motor; $v_{qs} v_{ds}$ denotes the q -axis and d -axis components

of the stator voltage vector v_s ; v_{qr} v_{dr} denotes the q -axis and d -axis components of the rotor voltage vector v_r ; i_{qs} i_{ds} is the q -axis and d -axis components of the stator current vectors i_s ; i_{qr} i_{dr} is the q -axis and d -axis components of the rotor current vectors i_r .

$$\psi_{mq} = X_{ml} \left[\frac{\psi_{qs}}{X_{ls}} + \frac{\psi_{qr}}{X_{lr}} \right] \quad (2)$$

$$\psi_{md} = X_{ml} \left[\frac{\psi_{ds}}{X_{ls}} + \frac{\psi_{dr}}{X_{lr}} \right] \quad (3)$$

$$X_{ml} = 1 / \left(\frac{1}{X_m} + \frac{1}{X_{ls}} + \frac{1}{X_{lr}} \right) \quad (4)$$

In order to find the current, substitute the values of the flux linkages:

$$i_{qs} = \frac{1}{X_{ls}} (\psi_{qs} - \psi_{mq}) \quad (5)$$

$$i_{ds} = \frac{1}{X_{ls}} (\psi_{ds} - \psi_{md}) \quad (6)$$

$$i_{qr} = \frac{1}{X_{lr}} (\psi_{qr} - \psi_{mq}) \quad (7)$$

$$i_{dr} = \frac{1}{X_{lr}} (\psi_{dr} - \psi_{md}) \quad (8)$$

Based on the above equation, the electromagnetic torque equation is developed from the electric power supplied to the rotor in the $dq0$ axis. Hence, the torque and rotor speed can be determined as follows:

$$T_e = \frac{3}{2} \left(\frac{P}{2} \right) \frac{1}{\omega_b} (\psi_{ds} i_{qs} - \psi_{qs} i_{ds}) \quad (9)$$

$$\frac{d\omega_r}{dt} = \frac{P}{2J} (T_e - T_L) \quad (10)$$

$$\omega_r = \int \frac{P}{2J} (T_e - T_L) \quad (11)$$

where T_e is electric torque, P is the number of pole pairs of motor.

After deriving the electric torque and rotor speed equations, the $dq0$ axis transformation should now be applied to the induction motor stator current [7]. However, this current and torque equations that describe the dynamic behaviour of an induction

motor are time varying which can be successfully as an input for the gear dynamic model [6].

The instantaneous values of the stator and rotor currents in three-phase system are ultimately calculated using the following transformation as:

$$\begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} i_d \\ i_q \end{bmatrix} \tag{12}$$

Figure 1 illustrates the transformation motor current signal in *abc* reference frame to *dq0* frame.

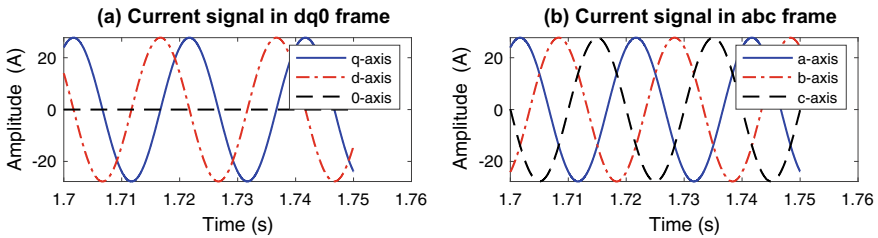


Fig. 1 Transformation of *abc* axis to *dq0* axis

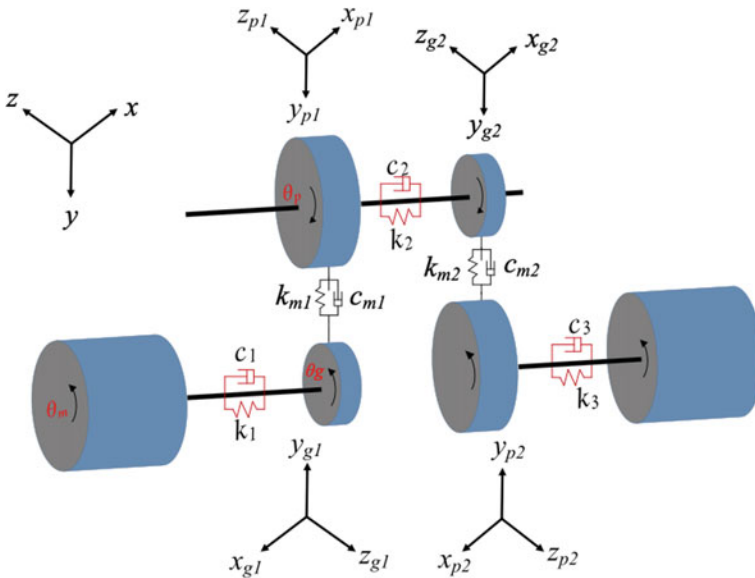


Fig. 2 Dynamic model of motor-gear transmission system

2.2 Gearbox Dynamic Model

The model developed in this study is based on a two-stage reduction gearbox. There is a total of 18-DOF in the model and it includes of six inertias, namely induction motor, load generator, driving and driven gears. The translational motion X, Y, Z of gears are considered into this model. To study the dynamic responses of the gear caused by tooth breakage, an integrated motor-gear dynamic model was put forward as demonstrated in Fig. 2. Furthermore, the damping and stiffness are also taken into consideration.

The governing equation of the 18-DOF gear dynamic model dynamic can be expressed as:

$$\left\{ \begin{array}{l}
 I_m \ddot{\theta}_m + (\theta_m - \theta_{g1})k_1 + (\dot{\theta}_m - \dot{\theta}_{g1})c_1 = T_e \\
 I_{g1} \ddot{\theta}_{g1} + (\theta_{g1} - \theta_m)k_1 + (\dot{\theta}_{g1} - \dot{\theta}_m)c_1 - F_{m1}r_{bg1} = 0 \\
 m_{g1} \ddot{x}_{g1} + c_{bx} \dot{x}_{g1} + k_{bx} x_{g1} = F_{x1} \\
 m_{g1} \ddot{y}_{g1} + c_{by} \dot{y}_{g1} + k_{by} y_{g1} \\
 m_{g1} \ddot{z}_{g1} + c_{bz} \dot{z}_{g1} + k_{bz} z_{g1} = F_{z1} \\
 I_{p1} \ddot{\theta}_{p1} + F_{m1}r_{bg1} + k_2(\theta_{p1} - \theta_{g2}) + c_2(\dot{\theta}_{p1} - \dot{\theta}_{g2}) = 0 \\
 m_{p1} \ddot{x}_{p1} + c_{bx} \dot{x}_{p1} + k_{bx} x_{p1} = -F_{x1} \\
 m_{p1} \ddot{y}_{p1} + c_{by} \dot{y}_{p1} + k_{by} y_{p1} = -F_{y1} \\
 m_{p1} \ddot{z}_{p1} + c_{bz} \dot{z}_{p1} + k_{bz} z_{p1} = -F_{z1} \\
 I_{g2} \ddot{\theta}_{g2} - k_2(\theta_{p1} - \theta_{g2}) - c_2(\dot{\theta}_{p1} - \dot{\theta}_{g2}) - F_{m2}r_{bg2} = 0 \\
 m_{g2} \ddot{x}_{g2} + c_{bx} \dot{x}_{g2} + k_{bx} x_{g2} = F_{x2} \\
 m_{g2} \ddot{y}_{g2} + c_{by} \dot{y}_{g2} + k_{by} y_{g2} = F_{y2} \\
 m_{g2} \ddot{z}_{g2} + c_{bz} \dot{z}_{g2} + k_{bz} z_{g2} = F_{z2}n \\
 I_{p2} \ddot{\theta}_{p2} + F_{m2}r_{bg2} + k_3(\theta_{p2} - \theta_L) + c_3(\dot{\theta}_{p2} - \dot{\theta}_L) = 0 \\
 m_{p2} \ddot{x}_{p2} + c_{bx} \dot{x}_{p2} + k_{bx} x_{p2} = -F_{x2} \\
 m_{p2} \ddot{y}_{p2} + c_{by} \dot{y}_{p2} + k_{by} y_{p2} = -F_{y2} \\
 m_{p2} \ddot{z}_{p2} + c_{bz} \dot{z}_{p2} + k_{bz} z_{p2} = -F_{z2} \\
 I_L \ddot{\theta}_L - k_3(\theta_{p2} - \theta_L) - c_3(\dot{\theta}_{p2} - \dot{\theta}_L) = -T_L
 \end{array} \right. \tag{13}$$

where I_i is the moment of inertia, subscript $i = m, p, g, L$ denotes motor, pinion, gear, load generator; m is the mass of the gear; subscript $m = p, g$; $\theta_i/\dot{\theta}_i/\ddot{\theta}_i$ represents the angle/velocity/acceleration of rotation; b are for first coupling, second coupling and bearing; x_i, y_i and z_i denotes the translational displacement; c_j and k_j respectively denotes the damping constant and stiffness constant, the subscript $j = 1, 2, 3$ denotes the damping and stiffness coefficient; F_m is the meshing force; T_e, T_L denotes the torque from the driving motor and the load generator; r_{bg}, r_{bp} are the radius of the base circle of gear and pinion.

Gear Mesh Stiffness

Mesh stiffness is one of the core vibrations sources of the gear system when it comes to tooth fault. Therefore, the accurate modelling of the mesh stiffness is important for the simulation of other important aspects of the gear teeth in order to investigate its dynamic behaviour. The gear mesh stiffness model described in this study was based on the work by Sánchez [13] and can be calculated by using an approximated normalized pair stiffness as:

$$k_m(\xi) \cong k \cdot \cos(b_0(\xi - \xi_m)) \tag{14}$$

Note, the value of ξ_m cannot change because function $k_m(\xi)$ keeps the symmetry respect to the midpoint of the interval of contact.

Where k is the averaged gear mesh stiffness. However, Eq. (14) remains valid for describing the tooth-pair stiffness, ε_α represent constant ratio, but parameters b_0 and ξ_m should be computed considering the values of the contact ratio and the inner point of contact parameter corresponding to a standard gear with the same geometry given as:

$$b_0 = \left[\frac{1}{2} \left(1.11 + \frac{\varepsilon_\alpha}{2} \right)^2 - 1.17 \right] \tag{15}$$

$$\xi_m = \xi_{inn} + \frac{\varepsilon_\alpha}{2} \tag{16}$$

Also, Sun et al. [12] stated that the average tooth stiffness k depends on the contact point. However, in order to explain the contact point, the involute profile parameter ξ_{inn} is used and can defined as:

$$\xi_{inn} = \frac{z}{2\pi} \sqrt{\frac{r_c^2}{r_b^2} - 1} \tag{17}$$

where z is the number of teeth, r_b and r_c denotes the base radius and contact radius respectively.

Gear Meshing Force

As stated in Sanchez et al. [13] the dynamic gear meshing force can be represented as:

$$F_m = k_m \delta + c_m \dot{\delta} \tag{18}$$

where, $c_m = 2\eta\sqrt{k_m m_e}$ and c_m is the gear mesh damping coefficient, m_e is the equivalent mass of gear and pinion; δ is the gear relative displacement which can be illustrated as:

$$\delta = (y_g + r_{bg}\theta_g) - (y_p + r_{bp}\theta_p) \quad (19)$$

where r_{bg} , r_{bp} are the radius of the base circle of gear and pinion.

2.3 Coupling Between Motor Model and Gearbox Model

This proposed 6° of freedom (DOF) electrical motor model integrated with an 18° of freedom (DOF) gear dynamic model with electric torque from the induction motor, coupled together via simple torsional stiffness k_m and damping coefficient c_m . As a result, the equation of motion for the motor-gear system becomes:

$$\omega_r = \dot{\theta}_m \quad (20)$$

$$T_{in} = T_e \quad (21)$$

where ω_r is the rotor speed of the motor model which is equal to $\dot{\theta}_m$ used as an input in the gear model. T_e denotes the electric torque of the motor model which is equal to T_{in} used an input to the gear model.

2.4 Numerical Solutions of the Model

The numerical solution for the proposed 6-DOF of the induction motor model integrated with an 18-DOF of the gear dynamic model has been illustrated in the flow chart diagram as shown in Fig. 3. Firstly, $t = 0$ represents the starting time, and the initial condition of the combined model was set as: $\psi[1 : 6] = 0$, $y[1 : 18] = 0$. The motor rotor speed ω_r equals the input speed of the gear model $y_i[1]$. Afterwards, the flux linkage was obtained by solving the different equations of the motor model. Then the motor current in $dq0$ frame can be calculated using this flux linkage. Secondly, in the gear model, the relative displacement, the gear mesh stiffness and the mesh force were calculated according to the real-time displacement $y_i[1 : 18]$. These calculated parameters were used to solve the 18-DOF of the gear dynamic model using ODE solver in Matlab. Besides, the electric torque derived from motor model was the input for the gear dynamic model to achieve the coupling effect.

After solving this 18-DOF gear dynamic model, the output y_i was obtained. When $t \leq T$, the output will be used as an input to repeatedly solve the 6-variable parameter for the flux linkage of the induction motor model and 18-variable parameter of the dynamic of the gear model. But when $t > T$, the output will be derived as the final results, and this end the calculation (Fig. 3).

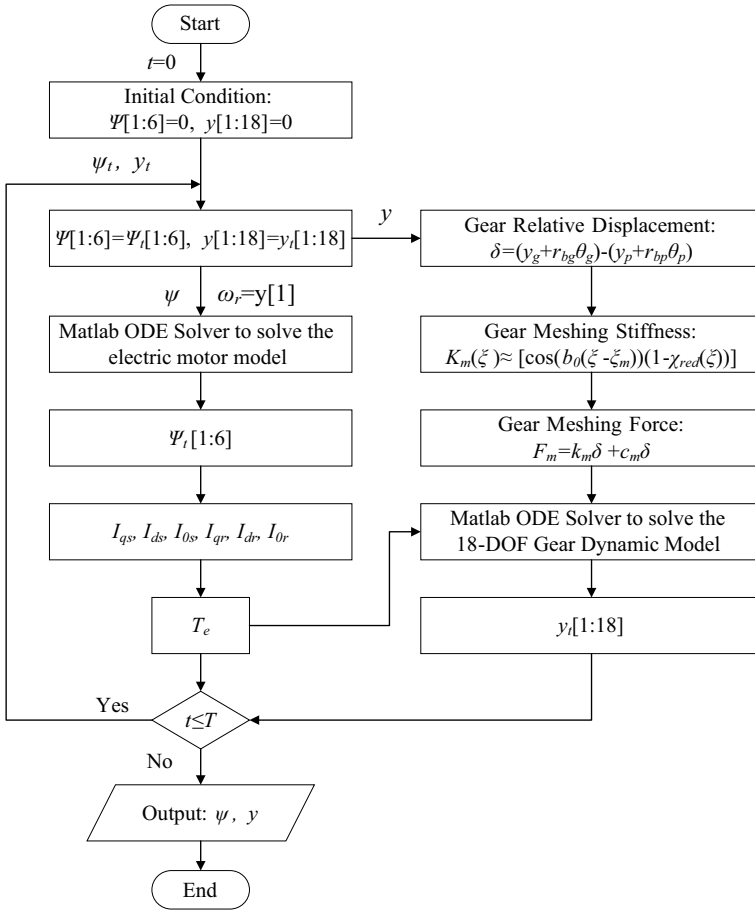


Fig. 3 The flowchart of the main program of numerical motor-gear model

3 Effect of Electric Dynamics on System Responses

To study the effect of electric dynamics on the motor-gear dynamic, the following comparison between constant torque and electric torque in Fig. 4 are both extracted for better diagnosis. The reason for this comparison is to show the behavioural effect of this torques on the system and their response.

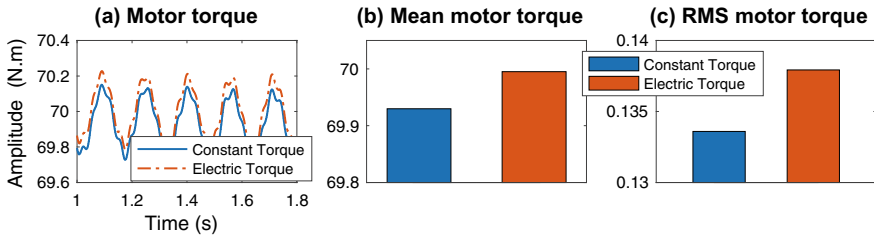


Fig. 4 Motor torque characteristics difference

3.1 Torque Differences

In order to have a more distinctive indication about how the electric dynamics influence the gear dynamics, the amplitude of the constant torque and electric torque were extracted from the spectra shown in Fig. 4a. The variations of the amplitude were shown in Fig. 4b, c. The amplitude of the mean motor torque under electric torque is higher than the constant torque thereby causing a fluctuation in the electric torque. However, for better and more distinctive indication, the RMS motor torque was extracted, and it reflects the deviation of the signal under electric torque. High RMS represents more modification effect between the shaft frequency and supply frequency, which can benefit the condition monitoring of faults occurred in the gear-motor system.

Figure 5 shows the difference in motor torque. However, amplitude of their sidebands was extracted. It can be seen that the amplitude of the motor torque under electric torque is slightly higher than the constant torque. This indicates that electric torque can reflect the fluctuation of gear or motor rotation motion more clearly which

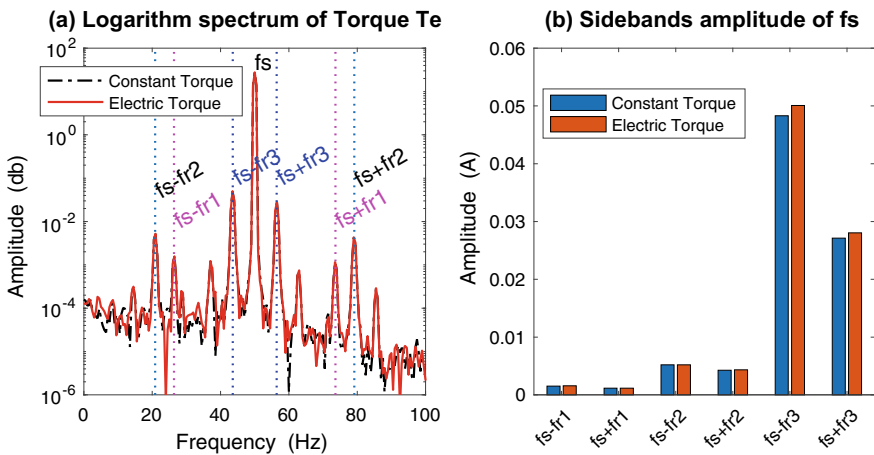


Fig. 5 Motor torque difference with their sideband

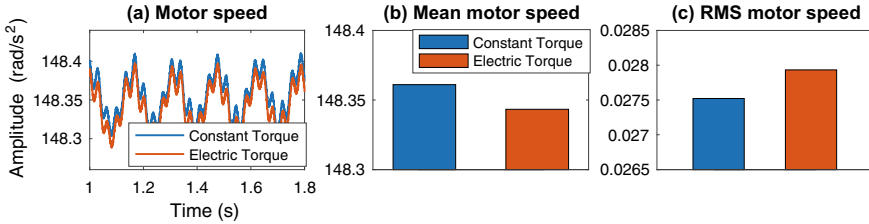


Fig. 6 Motor speed characteristics difference

enhances the modification effect of the sidebands around the supply frequency for easy detection.

3.2 Speed Differences

For more distinctive indication on how the electric dynamics influence the gear dynamics, Fig. 6 illustrates the time variation change in the motor speed between constant torque and electric torque with their respective amplitudes. It can be seen that the amplitude of motor speed under electric torque is slightly less than that under the constant torque. This is as a result of the fluctuation in the shaft which causes the fluctuation in motor speed. However, in Fig. 6c the amplitude of the RMS motor speed under electric torque is slightly higher than constant torque which indicates that electric torque is more sensitive to the fluctuation of motor speed and the gear rotation that is why we can get more fault frequency features.

3.3 Gear Torsional Responses

Wang et al. [14] studied the gear torsional stiffness response which depends on the torsional stiffness of the shaft and is also one of the main parameters for a nonlinear dynamic of a gear system. Thus, the study of the influence torsional stiffness has on the dynamic characteristics of a gear is necessary.

As can be seen in Fig. 7b, c, the mean and RMS value of middle shaft speed were extracted, and the amplitude of the RMS value of the shaft speed under electric torque is slightly higher than the constant torque. This indicates that electric torque is more sensitive to the fluctuation of the motor speed which makes it easier for more fault frequency features to be detected.

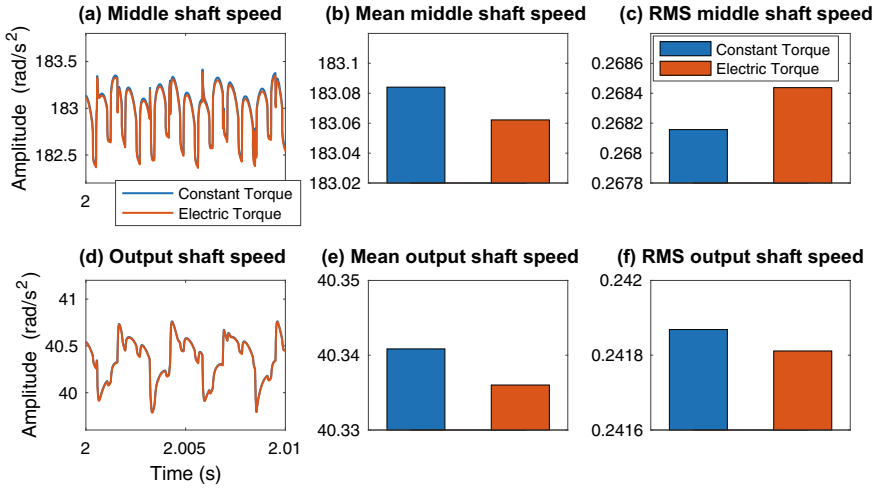


Fig. 7 Middle and outer shaft characteristics difference

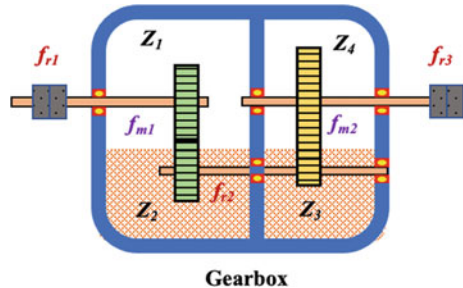
4 Simulation of Tooth Breakage

In order to verify the performance of the electric current signature for condition monitoring for detection, gear tooth breakage was chosen as the simulated fault for this study. This is because gear tooth breakage is one of the most common faults often found in gears. Furthermore, to simulate this fault, the schematic diagram of a gearbox has to be studied and this can be shown in Fig. 8.

In order to simulate this tooth breakage for investigation, a piecewise function of index was put forward to describe the reduction in gear mesh stiffness when the broken tooth come into mesh, which can be illustrated as:

$$\chi_{red}(\xi) = \begin{cases} 0, & (0 \leq \text{mod}(\xi, 2\pi) < z_i\alpha_p) \\ p_{per}, & (z_i\alpha_p \leq \text{mod}(\xi, 2\pi) \leq z_i\alpha_p + \varepsilon\alpha_p) \\ 0, & (z_i\alpha_p + \varepsilon\alpha_p < \text{mod}(\xi, 2\pi) \leq 2\pi) \end{cases} \quad (22)$$

Fig. 8 Schematic diagram of a gearbox



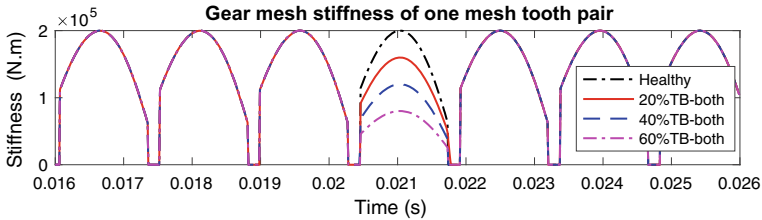


Fig. 9 Time varying gear mesh stiffness excitation

where, p_{per} is the percentage of tooth breakage of one tooth, which can equal the value located in the range of [0 1], z_i denotes the number of broken teeth, α_p represents the pitch angle, ε represents the contact ratio. The gear mesh stiffness considering tooth breakage can be described as:

$$k_m(\xi) \cong k \cdot [\cos(b_0(\xi - \xi_m))(1 - \chi_{red}(\xi))] \tag{23}$$

In this study, p_{per} was selected as 0.2, 0.4, 0.6 which represents the percentage of tooth breakage as 20%, 40%, 60% respectively.

Figure 9 depicts the time-varying mesh stiffness which varies periodically due to the continuous engagement of the teeth pair when it comes into contact with each other. It can also be noticed that the mesh stiffness displays a decreasing trend with the growth of the tooth breakage, which mainly occurs due to the reduction in gear tooth contact [12].

4.1 Tooth Breakage in the First Stage of Gearbox

In order to have a more distinctive indication on how gear tooth breakage influence the motor current, the amplitude sidebands surrounding the fundamental frequency were extracted from the spectra as shown in Fig. 10b–d respectively. The amplitude sideband $f_s f_{r1}$ shows a significant gradual increase with the fault severity. This is due to the fault in the first stage of the gearbox which is on the first shaft thereby giving a significant change in the motor current. Furthermore, the amplitude of sideband $f_s f_{r2}$ shows a slight increase because the second shaft is a bit far away from the motor compared with the first shaft, and the influence induced by the second shaft will cause more attenuation. However, the amplitude sideband $f_s f_{r3}$ remains constant throughout because the third shaft was not affected by the tooth breakage in the first stage.

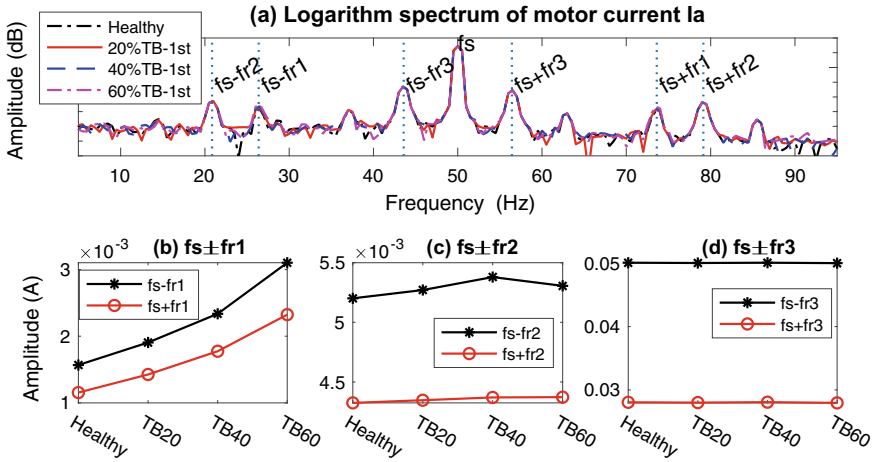


Fig. 10 Change in amplitude of the motor current under different severity of gear tooth defects in the gearbox first stage

4.2 Tooth Breakage in the Second Stage of Gearbox

To get a more distinguishable indication on how gear tooth breakage impacts the motor current, the amplitude sidebands around the supply frequency were extracted from the spectra as shown in Fig. 11b–d respectively. The amplitude sideband $f_s \pm fr_1$ around the supply frequency doesn't give a significant indication. This is because the fault is in the second stage of the gearbox. However, $f_s \pm fr_2$ shows a gradual increase

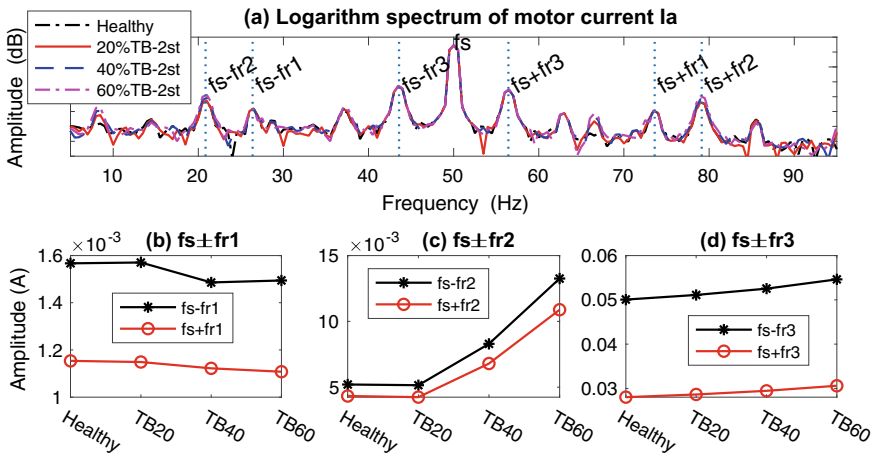


Fig. 11 Change in amplitude of the motor current under different severity of gear tooth defects in the gearbox second stage

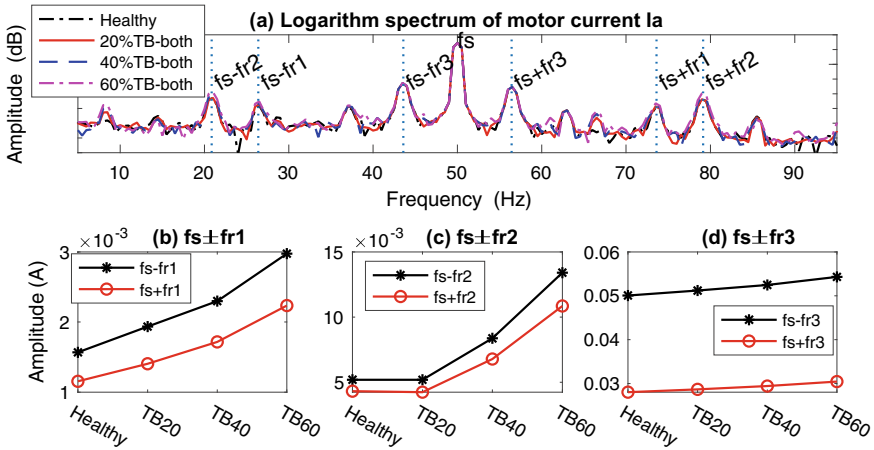


Fig. 12 Change in amplitude of the motor current under different severity of gear tooth defects in both stage of the gearbox

with the fault severity. This is as a result of the fault simulation in the second stage which is related to the second shaft, thereby giving a significant change in the motor current. Furthermore, the amplitude sideband $f_s f_{r3}$ shows a slight increase because of the influence induced by the third shaft will be attenuated more.

4.3 Tooth Breakage in Both Stage of Gearbox

In order to have a more robust indication about how gear tooth breakage affects the motor current signature, the amplitude sidebands around the supply frequency were extracted from the spectra as shown in Fig. 12b–d correspondingly. The amplitude sidebands $f_s f_{r1}$ and $f_s f_{r2}$ increases gradually with an increase in the fault severity. This is due to the fault in the first and second stage of the gearbox which is on the first and second shaft thereby giving a significant change in the motor current. However, the amplitudes sideband $f_s f_{r3}$ does not give that much of a significant increase because of the influence induced by the third shaft therefore resulting into more attenuation.

5 Experimental Validation

In order to verify and validate the simulated results obtained from the numerical analysis, an experimental test was carried out to investigate the electric dynamic responses with the simulation of gear tooth breakage as shown in Fig. 13. Figure 14a



Fig. 13 Gearbox test rig

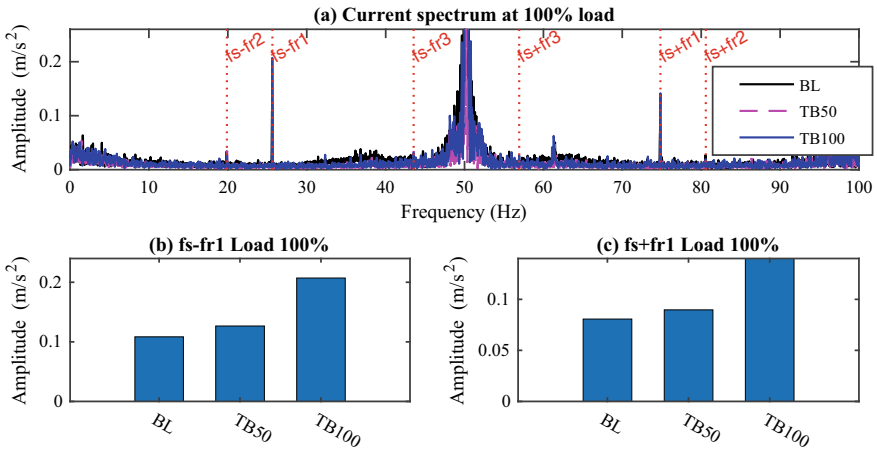


Fig. 14 Change in amplitude of the motor current under different severity of gear tooth defects

illustrates the motor current spectrum with the amplitude sidebands relating to the shaft frequencies in this study. Hence, the amplitudes of the sidebands surrounding the fundamental frequency were extracted from the spectra as shown in Fig. 13b, c. The amplitude sideband $f_s - f_{r1}$ relating to the first shaft with the gear tooth breakage increases with fault severity when the gear comes into contact. With this amplitude increase, there is a good consistency with the numerical result.

6 Conclusion

This study investigates the performance of electric motor current in detecting gear tooth breakage. In order to achieve this goal, a 6 DOF motor model combined with 18 DOF gear model was put forward. Electric torque is preferred in the model because it can enhance the modification effect. Furthermore, a simulation work and experimental work with tooth breakage occurred in different stages in gearbox had been carried out, several conclusions can be drawn as:

- (1) Through the comparison of the electric torque and constant torque, it can be found that the electric torque can effectively reflect the modification effect induced by the fluctuation of the shaft, which can be beneficial to the condition monitoring of faults occurred in the motor-gear system.
- (2) Through the study of the tooth breakage, it can be seen that the tooth breakage causes a continual increase in the amplitude of the corresponding sidebands around the supply frequency with the increase in fault severity, which can reflect the occurrence of tooth breakage. Besides, the change in the amplitude of the sidebands can be used to localise or identify the position of the broken tooth in the gearbox. For example, clear increase in f_s f_{r1} and slight increase in f_s f_{r2} reflect the tooth breakage in the first stage, and clear increase in f_s f_{r2} and slight increase in f_s f_{r3} reflect the tooth breakage in the second stage.
- (3) The experimental results do have a good agreement with the numerical results indicating that the numerical model can give a reliable and robust predication of motor-gear dynamics. Also, it provides a reliable theoretical basis for condition monitoring and accurate fault diagnosis of gear tooth breakage using motor current signature.

References

1. Jaspreet, Z.J., Dhupia, S., Gajanayake, C.J.: Stator current analysis from electrical machines using resonance residual technique to detect faults in planetary gearboxes. *IEEE Trans. Ind. Electron.* **62**(9), 5709–5721 (2015)
2. Bartelmus, W.: Mathematical modelling and computer simulations as an aid to gearbox diagnostics. *Mech. Syst. Sig. Process.* **15**(5), 855–871 (2001)
3. Shao, R., Dong, F., Wang, W., Jia, P.: Influence of cracks on dynamic characteristics and stress intensity factor of gears. *Eng. Fail. Anal.* **32**, 63–80 (2013)
4. Wenyu, B., Qin, D., Wang, Y., Lim, T.C.: Dynamic characteristics of motor-gear system under load saltations and voltage transients. *Mech. Syst. Sig. Process.* **100**, 1–16 (2018)
5. Zaigang, C., Shao, Y.: Dynamic simulation of spur gear with tooth root crack propagating along tooth width and crack depth. *Eng. Failure Anal.* **18**(8), 2149–2164 (2011)
6. Anand, B., Aspalli, M.S.: Dynamic dq model of induction motor using simulink. *Int. J. Eng. Trends Technol. (IJETT)* **2s4**(5), 252–257 (2015)
7. Adel, A., Rahman, M.A.: Dynamic simulation of a three-phase induction motor using matlab simulink. *Currents* **8**, 9 (2011)
8. Jidong, Z., Dhupia, J.S., Gajanayake, C.J.: Stator current analysis from electrical machines using resonance residual technique to detect faults in planetary gearboxes. *IEEE Trans. Ind. Electron.* **62**(9), 5709–5721 (2015)
9. Murat, I., Kahraman, A.: A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors. *J. Sound Vib.* **329**(4), 371–393 (2010)
10. Far, Z.D., Henao, H., Capolino, G.A.: Planetary gearbox effects on induction machine in wind turbine: modeling and analysis. In: 2012 XXth International Conference on Electrical Machines, pp. 1790–1796. IEEE (2012)
11. Punit, L.R., Thosar, A.G.: Mathematical modelling of a 3-phase induction motor using MATLAB/simulink. *Int. J. Modern Eng. Res. (IJMER)* **4**(6), 62–67 (2014)
12. Xiuquan, S., Wang, T., Zhang, R., Gu, F., Ball, A.D.: Numerical modelling of vibration responses of helical gears under progressive tooth wear for condition monitoring. *Mathematics* **9**(3), 213 (2021)

13. Miryam, B.S., Pleguezuelos, M., Pedrero, J.I.: Approximate equations for the meshing stiffness and the load sharing ratio of spur gears including hertzian effects. *Mech. Mach. Theor.* **109**, 231–249 (2017)
14. Wang, J., Zhang, J., Yao, Z., Yang, X., Sun, R., Zhao, Y.: Nonlinear characteristics of a multi-degree-of-freedom spur gear system with bending-torsional coupling vibration. *Mech. Syst. Sig. Process.* **121**, 810–827 (2019)

Analysing the Fault Behavior of a Complex Mechanical System for Diagnosis: A Bond Graph-Based Approach



Jinxin Wang, Shenglei Zhao, Xiuzhen Ma, and Fengshou Gu

Abstract Multiple faults diagnosis is a critical problem in mechanical fault diagnosis. Fault behavior analysis aims to find out the response characteristics of operating parameters under different faults and is the most primitive task for multiple faults diagnosis. This paper proposes a bond graph-based approach to analyse the mechanical fault behavior for diagnosis. The analytical model of an engineering system is firstly established via bond graph. The temporal causal graph is derived from the bond graph model to depict the analytic relationships system variables. The operating parameters response characteristics under different faults are then derived by representing the faults with abnormal change of system variables. The approach is illustrated via an engine lubrication system. The presented approach avoids time-consuming formula transformation which is necessary in mathematical model-based approach, and therefore provides an efficient for fault behavior analysis.

Keywords Multiple faults diagnosis · Fault behavior analysis · Bond graph · Temporal causal graph · Engine lubrication system

J. Wang (✉)

School of Safety Engineering, China University of Mining and Technology, Xuzhou 221116, China

e-mail: wangjinxin@cumt.edu.cn

J. Wang · X. Ma

College of Power and Energy Engineering, Harbin Engineering University, Harbin 150001, China

S. Zhao

College of Electrical and Automation Engineering, Shandong University of Science and Technology, Qingdao 266590, China

F. Gu

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

359

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_30

1 Introduction

A complex mechanical system will inevitably suffer from various failures during its service cycle. Condition monitoring and fault diagnosis (CMFD) technique aims to pinpoint the origin of an abnormality timely if it presents, which provides an effective means to ensure the safe and high-efficient operation of a complex mechanical system.

The diagnosis of multiple faults diagnosis is a crucial task in CMFD [1–4]. It can be divided into 4 different steps: fault behavior analysis, fault description and feature selection, fault detection, and fault isolation. Fault behavior analysis aims to find out the fault propagation path and the response characteristics of operating parameters under different faults, which is the most primitive task among the 4 tasks.

Correlation analysis and fault tree are two primary approaches for fault behavior analysis. Correlation analysis takes advantage of mathematical statistics theory to deal with the sample data of two state variables, and investigates the correlation of variables according to the data distribution regularities. This approach has been successfully used to study the causal relationships of mechanical faults and symptoms for some equipment [5, 6]. However, a large number of typical fault sample data is the prerequisite for this method. In practice, high-quality data is not always available for different machines, which limits this method to being widely used in real-world.

Fault tree is another method for fault behavior analysis [7–9]. Different from correlation analysis, this method takes the most undesired fault as Top Event, and then describes the propagation path of the target fault via an inverted tree structure. The causal relationships of faults and symptoms are therefore derived by exploring this tree. Fault tree gets rid of the dependency on sample data, however, the fault propagation path are actually modeled according to experts' subjective experience. Thus, different experts may give different diagnostic results. Besides, a fault tree for multiple fault diagnosis often requires a very complete expert knowledge base, which is usually difficult to obtain in practice.

Some researchers also use the commercial software to investigate the mechanical fault behavior [10], e.g. AMESim for hydraulic valve leakage, GT-Power for engine misfire. Although this approach provides an efficient way for fault behavior analysis, the fault type that commercial software can deal with limited by the component libraries the software can provide. Thus, commercial software-based approach is found limited in some cases.

Mathematical model-based approach is an alternative way to pursue the causal relationships of faults and symptoms for diagnosis. This approach firstly describes the working process of a mechanical system via a series of analytical redundancy relations (ARRs), and then introduces the interested fault by changing the state variable, the operating parameters response characteristics under different faults can be got accordingly. Despite this approach gets rid of the dependency on sample data, and avoids the influence of experts' subjective experience at the same time, mathematical model-based approach suffers severely from the enormous cost of modeling and calculation. Take a diesel engine as an example, up to 1640 ARRs can

be generated from the 39 differential equations and algebraic equations (required to model the dynamic behaviors of the engine) to depict the dependency among the variables [11–13].

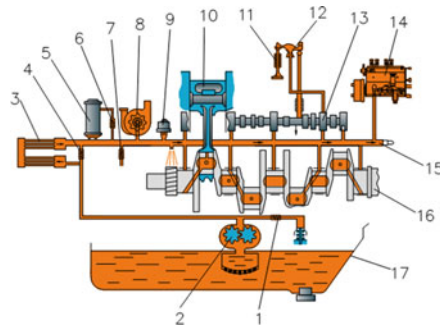
Bond graph is a graphical modeling methodology. It models an engineering system by considering the energy flows between the components. Different bond elements are used to describe the storage, release, transfer and conversion of the energy. A complete model of an engineering system is constructed via combining the various bond elements according to the connection of components. Bond graph is a domain-independent modeling method (it can describe the system from different domains, e.g. electrical, mechanical, hydraulic, in the same way), therefore it is a powerful tool for modeling engineering systems when different physical domains are involved. Besides, not constructing the ARRs directly, bond graph firstly takes use of graphical elements to represent the interaction of system components and then derives the mathematical analytical equation from the graphical model. Compare with the conventional mathematical model-based approach, bond graph simplifies the construction of ARRs and therefore provides an effective way for fault behavior analysis.

This paper presents a bond graph-based approach to analyse the mechanical fault behavior for diagnosis. The analytical model of an engineering system is firstly established via bond graph. The temporal causal graph is derived from the bond graph model to depict the analytic relationships system variables. The operating parameters response characteristics under different faults are then derived by representing the faults with abnormal change of system variables. The approach is illustrated via an engine lubrication system. The rest of this paper is organized as follows. Section 2 introduces the structure and working principle of engine lubrication system. Section 3 describes the presented approach and illustrated using an engine lubrication system. The model and the presented approach are verified in Sect. 4. Finally, Sect. 11 summaries the paper.

2 Example: Engine Lubrication System

Lubrication system is one of the most crucial systems in diesel engine, which supplies enough oil to the engine friction pair to guarantee the safe and reliable operation of a diesel engine. Fault diagnosis is a vital task to guarantee the safe operation of engine lubrication system.

A schematic diagram of a typical marine diesel engine lubrication system is shown as Fig. 1. The gear pump is driven by crankshaft through a transmission gear or a belt. A relief valve is usually installed after the oil pump to prevent the damage caused by excessive oil pressure. The lubricating oil then enters the oil cooler to keep a certain temperature. The oil cooler is usually connected in parallel with a thermostatic valve to control the oil temperature. The cooled lubricating oil enters the filter. Each filter is connected in parallel with a bypass valve to prevent the engine lubrication failure caused by filter clogging. The main oil gallery is usually installed with a pressure relief valve to adjust the pressure of the lubricating oil. Besides lubricating the friction



1-relief valve of gear pump, 2-gear pump, 3-oil cooler, 4- thermostatic valve, 5-oil filter, 6-bypass valve, 7-pressure regulating valve, 8-supercharger, 9-oil pressure gauge, 10-piston, 11-air valve, 12-rocker arm, 13-camshaft, 14-fuel injection pump, 15-main oil gallery, 16-crankshaft, 17- oil sump

Fig. 1 A typical marine diesel engine lubrication system

pair, the lubricating oil can also take heat away from the engine body to make the pistons, crankshafts and other components working at a certain temperature.

3 Fault Behavior Analysis Using Bond Graph-Based Approach

3.1 Modeling the System Dynamic Behavior Based on Bond Graph

3.1.1 Bond Graph Theory

In bond graph, four generalized variables: effort $e(t)$, flow $f(t)$, momentum $p(t)$ and displacement $q(t)$ are defined to represent the energy flow in any systems. The effort $e(t)$, flow $f(t)$ are two basic variables, where the power of the instantaneous energy flow is product of these two variables, see Eq. (1).

$$Power(t) = e(t) \cdot f(t) \quad (1)$$

Momentum $p(t)$ and displacement $q(t)$ are defined as

$$\begin{cases} \frac{dp(t)}{dt} = e(t) \\ \frac{dq(t)}{dt} = f(t) \end{cases} \tag{2}$$

Six basic elements are defined to model the components in real systems, see Fig. 2. Figure 2a shows the effort source S_e and flow source S_f , which often used to describe the energy source with a specific effort/flow level. Figure 2b shows the three passive elements: R-element, C-element and I-element. R-element is used to describe the components which dissipate energy, e.g. dampers, frictions and electric resistors in real-life systems. C-element and I-element are used to represent the components which store all kinds of energy. Specifically, in C-element, the energy is stored by accumulating the net flow, and in I-element, the energy is stored by accumulating the net effort.

Figure 2c, d show the two-port elements: transformer and gyrator respectively, which are usually used to represent the energy conversion both in same energy form and different energy form. Variables m and r are called transformer module and gyrator ratio respectively.

Figure 2e, f are two junctions, which are used to represent the conservation of energy. The 0-junction is also named common effort junction, which is used to connect the elements with equal efforts. The 1-junction is also named common flow junction, which is used to connect the elements with equal flows.

Although bond graph had been successfully applied in electrical, mechanical, hydraulic domains, it was found limited in thermodynamics domain. To solve this

Fig. 2 Basic elements of bond graph



(a) Sources (b) Passive elements

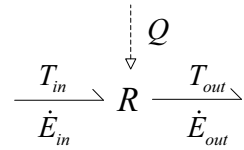


(c) Transformer (d) Gyrator



(e) 0-Junction (f) 1-Junction

Fig. 3 The 2-port R-element representing convection in pseudo bond graphs



problem, Karnopp proposed pseudo bond graphs [14]. Different from classical bond graph (or power bond graph), temperature T and heat flow rate \dot{E} are chosen as effort source S_e and flow source S_f in pseudo bond graphs (\dot{E} is already a power term). The convection of heat is therefore represented via a two-port R-element, see Fig. 3. The variables T_{in} and \dot{E}_{in} represent the inlet temperature and heat flow rate, while T_{out} and \dot{E}_{out} represent the outlet temperature and heat flow rate. The differential equation can be written as Eq. (3). The thermodynamic equations derived by this way match the Euler method very well, which makes pseudo bond graphs being widely used in thermodynamics domain.

$$\begin{aligned} \text{if } Q > 0, \quad \dot{E}_{in} &= \dot{E}_{out} = \rho c Q T_{in} \\ \text{if } Q < 0, \quad \dot{E}_{in} &= \dot{E}_{out} = \rho c Q T_{out} \end{aligned} \tag{3}$$

3.1.2 Mechanical Domain

We model the dynamic behavior of engine lubrication system based on bond graph theory. The working of engine lubrication system involves three domains: mechanical, hydraulic and thermodynamic. We model the mechanical domains first. The crankshaft is modeled by a flow source S_f . The gear pump is connected with crankshaft via a set of transmission gear or a belt. The transmission is modeled via a transformer TE where the transformer module i represents the transmission ratio. Parameter R_{effic} describes the energy loss while transmission. Since the pump has a certain mass, it will consume some energy while working. We describe this process via I_{pump} . The bond graph model of mechanical domains can be seen as Fig. 4.

3.1.3 Hydraulic Domain

The pump converts the mechanical energy into hydraulic energy, which can be modeled via a transformer TE. R_{leak} represents the leakage of oil pump. The oil flow through the cooler is governed by the parameters R_{cooler} and C_{cooler} , which represent the fluidic resistor and hydraulic volume respectively. The thermostatic valve presents a narrow opening and is modeled via a resistance R_{thermo} , which is also connect in parallel with cooler. The filter and bypass valve has similar physical effect with cooler and thermostatic valve, so that they have Similar model structure.

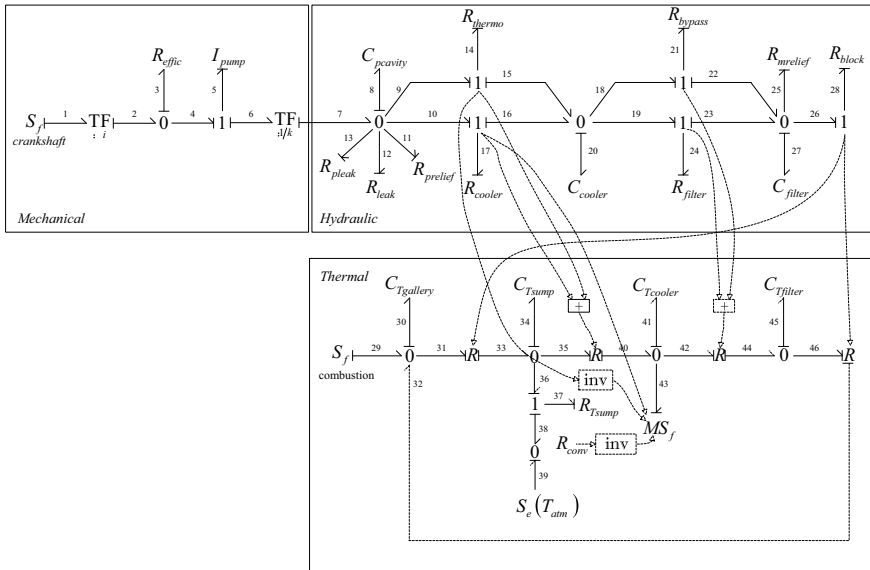


Fig. 4 The bond graph model of engine lubrication system

We regard the engine body as a resistor for oil flow, therefore it is represented via R_{block} . $R_{mrelief}$ describes the fluidic resistor of pressure regulating valve.

3.1.4 Thermal Domain

The working engine can be seen as a heat source which is therefore represented as S_f . The components of engine is modeled as $C_{Tgallery}$. C_{Tsump} is the heat of oil sump. The oil in sump will dissipation to air, we model the heat dissipation via R_{Tsump} . The temperature of environment is thought as constant, so it is modeled as S_e . The oil flows into cooler under the working of pump. We model the heat capacity of cooler as $C_{Tcooler}$. The dissipated heat is represented by MS_f . The model of oil filter is similar with cooler. The transmitted heat is not only related to temperature, but also to mass flow. So the thermodynamic model is connected with hydraulic part via two-port R-element.

The bond graph model of engine lubrication system is shown as Fig. 4. R_{leak} represents the leak oil through oil pipe, and it is quantize by Q_{leak} .

3.2 Analysing the Parameters Response Characteristics Using Temporal Causal Graph

Although we can derive the analytical redundancy relations and analyse the fault behavior through the ARRs, this work is thought to be time-consuming and error-prone. Temporal Causal Graph (TCG) is a signal flow diagram, in which the variables are represented as vertices and the relations between the variables as directed edges, and the relations type are described via a set of labels $L = \{1, -1, =, \lambda, 1/\lambda, \lambda dt, \lambda/\lambda dt\}$. Specifically, $1, \lambda$ and $1/\lambda$ describe the positive correlation, in which the coefficients of association are $1, \lambda$ and $1/\lambda$ respectively; -1 represents the negative correlation; $=$ means two variables are equationual in number; λdt and $1/\lambda dt$ represent the integral relationships in which the parent nodes are the time derivatives of child nodes. The TCG can be derived directly from the bond graph model, see [15].

The TCG model of engine lubrication system derived from the bond graph is shown as Fig. 5. Since the TCG describes the functional relationships of variables intuitively, it can conveniently be used to analyse the parameters response characteristics under different faults. Taking filter blocking as an example, the fault behavior can be shown as Fig. 6. The loop in TCG represents the negative feedback, e.g. the filter blocking will decrease the Q_{27} , as for loop $Q_{27} \rightarrow p_{27} \rightarrow p_{25} \rightarrow Q_{25}$, we have $Q_{27} \downarrow \rightarrow p'_{27} \downarrow \rightarrow p'_{25} \downarrow \rightarrow Q'_{25} \downarrow \rightarrow Q'_{27} \uparrow$. The first order derivative of Q_{27} increases, therefore Q_{27} will finally reach a constant value. We investigate 11 faults of engine lubrication system and the response of 8 working parameters, see Tables 1 and 2. The fault behavior is shown as Table 3, in which $+1$ represents the occurrence of fault would increase the parameter (also marked with red), and -1 is present otherwise (also marked with blue).

4 Experimental Validation

In this section, we introduce some faults (in total 7 faults) shown in Table 1 and the working parameters in Table 2 are detected accordingly. The engine is an in-line 2 cylinders water-cooled marine engine, see Fig. 7. The engine is connected with a reduction gear box. The propeller is installed in a pipe loop which is filled with water. A flow adjusting valve is set at the middle of the pipe loop to give an adjustable load.

The experiments are carried out at 2000 r/min and 25%, 50%, 75% load respectively. The parameters at 75% load are presented as Table 4.

It can be seen from Table 4 that the working parameters present abnormal changes under different faults, and the abnormal changes got from experiment data meet well with Table 3. Take filter blocking as an example, Table 4 shows that lubrication oil pressure after pump and the pressure before filter increase 0.1 bar compared with the one at engine normal condition ($+1$ in Table 3). On the contrary, the pressure after filter and pressure before engine decrease 1.5 bar (-1 in Table 3). As for the

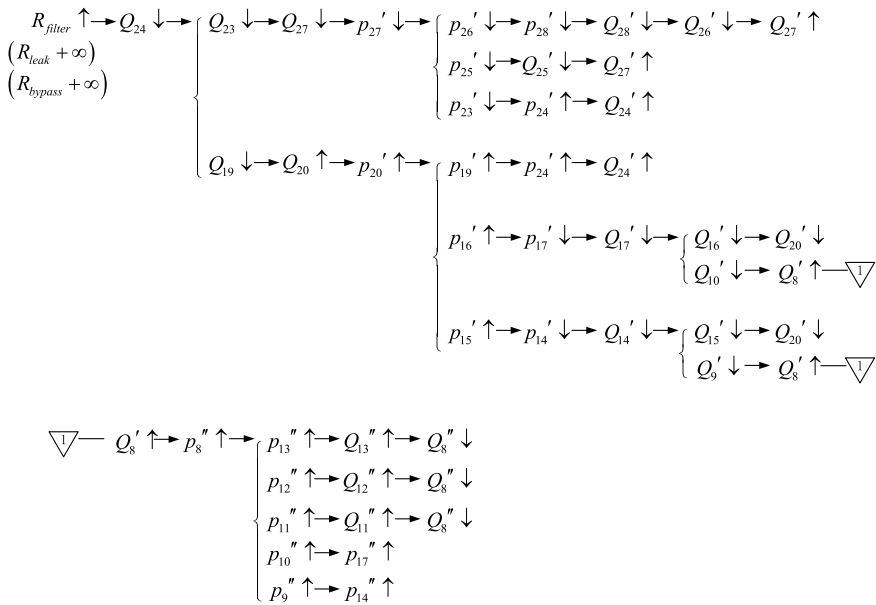


Fig. 6 Failure behavior analysis of filter blocking

Table 1 Common faults of a diesel engine lubrication system

Symbols	Faults	Symbols	Faults
f_1	filter blocking	f_7	Cooler fouling
f_2	Pipe leakage	f_8	Relief valve of pump leakage
f_3	Thermostatic valve stuck	f_9	Relief valve of pump stuck
f_4	Lubrication oil shortage	f_{10}	Main oil gallery blockage
f_5	Bypass valve leakage	f_{11}	Main oil gallery leakage
f_6	Thermostatic valve leakage		

temperature, the temperature at different position all increase compared with normal values (+1 in Table 3). The results show that the fault behavior can be got using bond graph combining with TCG.

Table 2 Observable parameters of an engine lubrication system

Symbols	Working parameters	Variables in TCG
s_1	Lubrication oil pressure after pump	p_8
s_2	Lubrication oil pressure before filter	p_{20}
s_3	Lubrication oil pressure after filter	p_{27}
s_4	Lubrication oil pressure before engine	p_{28}
s_5	Lubrication oil temperature after pump	T_{34}
s_6	Lubrication oil temperature after cooler	T_{41}
s_7	Lubrication oil temperature after filter	T_{45}
s_8	Lubrication oil temperature before engine	T_{30}

Table 3 Fault-symptom causal relationship of engine lubrication system

Num	Faults	Simulation	Working parameters							
			s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
1	f_1	$R_{filter} \uparrow$	+1	+1	-1	-1	+1	+1	+1	+1
2	f_2	$R_{leak} : +\infty \rightarrow R$	-1	-1	-1	-1	+1	+1	+1	+1
3	f_3	$R_{thermo} \uparrow$	+1	-1	-1	-1	-1	-1	-1	-1
4	f_4	$Q_7 \downarrow Q_{T_{summ}} \downarrow$	-1	-1	-1	-1	+1	+1	+1	+1
5	f_5	$R_{by\ pass} : +\infty \rightarrow R$ $R_{filter} \rightarrow +\infty$	-1	-1	+1	+1	-1	-1	-1	-1
6	f_6	$R_{thermo} \downarrow$	-1	+1	+1	+1	+1	+1	+1	+1
7	f_7	$R_{cooler} \uparrow$	+1	-1	-1	-1	+1	+1	+1	+1
8	f_8	$R_{prelief} \downarrow$	-1	-1	-1	-1	+1	+1	+1	+1
9	f_9	$R_{prelief} \uparrow$	+1	+1	+1	+1	-1	-1	-1	-1
10	f_{10}	$R_{block} \uparrow$	+1	+1	+1	+1	+1	+1	+1	+1
11	f_{11}	$R_{block} \downarrow$	-1	-1	-1	-1	-1	-1	-1	-1

5 Conclusions

In this paper, we present a bond graph-based approach to analyse the fault behavior for diagnosis. The approach is demonstrated via an engine lubrication system. Results show that the presented approach can exactly derive the response characteristics of operating parameters under different faults. This approach avoids time-consuming

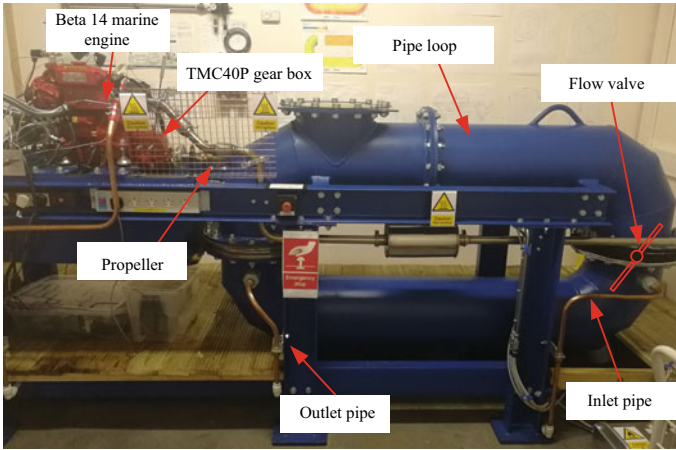


Fig. 7 Test cell of a Beta 14 marine engine

Table 4 Working parameters of a Beta 14 diesel engine lubrication system

Num	Condition	s_1/bar	s_2/bar	s_3/bar	s_4/bar	$s_5/^\circ\text{C}$	$s_6/^\circ\text{C}$	$s_7/^\circ\text{C}$	$s_8/^\circ\text{C}$
1	Normal	4.03	3.86	3.68	3.53	87.40	79.58	81.45	80.26
2	f_1	4.15	4.03	3.17	3.04	90.16	82.59	84.58	83.62
3	f_2	3.62	3.48	3.32	3.18	91.85	83.98	85.98	85.88
4	f_3	4.08	3.80	3.61	3.48	81.91	76.79	78.51	74.88
5	f_4	3.78	3.64	3.47	3.35	94.19	86.63	88.40	86.50
6	f_5	3.88	3.78	3.76	3.61	81.44	73.37	75.76	74.17
7	f_6	3.99	3.93	3.74	3.60	93.06	86.97	88.33	85.89
8	f_7	4.12	3.78	3.62	3.48	89.58	83.19	84.03	82.70

formula transformation which is necessary in mathematical model-based approach, and therefore provides an efficient for fault behavior analysis.

Acknowledgements This work was supported by the Fundamental Research Funds for the Central Universities (Grant No: 2021QN1089).

References

1. Li, Z., Jiang, Y., Hu, C., Peng, Z.: Recent progress on decoupling diagnosis of hybrid failures in gear transmission systems using vibration sensor signal: a review. *Measurement* **90**, 4–19 (2016)
2. Zhang, S., He, Q., Ouyang, K., Xiong, W.: Multi-bearing weak defect detection for wayside acoustic diagnosis based on a time-varying spatial filtering rearrangement. *Mech. Syst. Sig. Process* **100**, 224–241 (2018)
3. Shao, H., Jiang, H., Lin, Y., Li, X.: A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mech. Syst. Sig. Process* **102**, 278–297 (2018)
4. Diehl, E.J.: Predictive modeling of a two-stage gearbox towards fault detection. *Shock Vib.* **2016**(17), 1–13 (2016)
5. Chen, P., Toyota, T., He, Z.J.: Automated function generation of symptom parameters and application to fault diagnosis of machinery under variable operating conditions. *IEEE Trans. Syst. Man Cybern. Part A Syst. Humans* **31**(6), 775–781 (2001)
6. Sekhar, A.S.: Multiple cracks effects and identification. *Mech. Syst. Sig. Process.* **22**(4), 845–878 (2008)
7. Bobbio, A., Portinale, L., Minichino, M., Ciancamerla, E.: Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliab. Eng. Syst. Saf.* **71**(3), 249–260 (2001)
8. Khakzad, N., Khan, F., Amyotte, P.: Safety analysis in process facilities: comparison of fault tree and Bayesian network approaches. *Reliab. Eng. Syst. Saf.* **96**, 925–932 (2011)
9. Bartlett, L.M., Hurdle, E.E., Kelly, E.M.: Integrated system fault diagnostics utilizing diagraph and FT-based approach. *J. Reliab. Eng. Syst. Saf.* **94**, 1107–1115 (2009)
10. Guowei, S.: Study on the Method of Diesel Engine Fault Diagnosis Based on Ant Clustering Algorithm. Harbin Engineering University, Harbin (2014)
11. Wahlström, J., Eriksson, L.: Modelling diesel engines with a variable-geometry turbocharger and exhaust gas recirculation by optimization of model parameters for capturing non-linear system dynamics. *PI Mech. Eng.* **225**(7), 960–986 (2011)
12. Svärd, C., Nyberg, M., Frisk, E., Krysander, M.: Automotive engine FDI by application of an automated model-based and data-driven design methodology. *Control Eng. Pract.* **21**(4), 455–472 (2013)
13. Chi, G., Wang, D., Le, T., Yu, M., Luo, M.: Sensor placement for fault isolability using low complexity dynamic programming. *IEEE Trans. Autom. Sci. Eng.* **12**(3), 1080–1091 (2015)
14. Karnopp, D.: State variables and pseudo bond graphs for compressible thermofluid systems. *J. Dyn. Syst. Meas. Contr.* **101**, 201–204 (1979)
15. Mosterman, P.J., Biswas, G.: Diagnosis of continuous valued systems in transient operating regions. *IEEE Trans. Syst. Man Cybern. Part A Syst. Humans* **29**(6), 554–565 (1999)

Dynamic Responses of Clearance Induced Impacts in Big End Bearing Condition Monitoring of Diesel Engines



Solomon Okhionkпамwonyi, Fengshou Gu, and Andrew D. Ball

Abstract Enlarged clearance in connecting rod big end bearing is known to be a typical fault in internal Combustion (IC) engines such as diesel engines. Enlarged clearance causes inefficient operation that leads to reduced performance of the engine. In this work, the influence of enlarged clearances in big end bearing on the vibration of the system is studied. The kinematics and dynamics of the big end bearing were explored using numerical simulation to investigate the effect of the clearance on the associated response of acceleration signals obtained by multiply the simulated bearing forces with frequency response function (FRF). Different clearance and engine speed conditions were simulated for analysis of various degree of bearing clearance fault. Results obtained for these cases are evaluated for vibration analysis which shows that simulation model developed is able to simulate the vibration signals issuing from the dynamics of engine system with various clearance phenomenon.

Keywords Diesel engine · Vibration analysis · Big-end bearing · Enlarged bearing clearance · Journal bearings simulation

1 Introduction

In internal combustion engines, the mechanism which consist of the piston, connecting rod and crankshaft is employed to convert the reciprocating motion of piston to angular rotation of the crankshaft though the transmission of combustion force. The mechanism which is mostly referred to as slider-crank mechanism, requires the clearances at the links between the components to ensure movement

S. Okhionkпамwonyi · F. Gu (✉) · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, UK
e-mail: f.gu@hud.ac.uk

S. Okhionkпамwonyi
e-mail: Solomon.okhionkпамwonyi@hud.ac.uk

A. D. Ball
e-mail: a.d.ball@hud.ac.uk

and functionality [1]. In the slider-crank mechanism, these links are formed between the piston and connecting rod, connecting rod and crankshaft with adequate clearance. Journal-bearing component pair is a typical example of a revolute joint with clearance that allows relative motion between the coupled parts. In order to ensure a functional and durable journal-bearing coupling, relative motion must be permitted to avoid contact between surfaces that could lead to excessive vibration, friction and wear and ultimately bearing damage [2–4].

Due to the rotating and reciprocating dynamic loads imposed on journal bearings used in internal combustion engines, these engine bearings are more complex. Particularly, the connecting rod big end bearing which is more exposed in comparison to other engine bearings such as the crankshaft main bearings. The variation in speed and load is consequential in big end bearings behaviour as it leads to changes to the lubrication regime from boundary lubrication experienced at engine start-up and shutdown, to oil film separation whilst in operation [3]. In the advent of impulsive loading, the oil film become thinner with less capability to separate both surfaces. The resulting effect induces the enlargement of the clearance between the journal and bearings shells leaving the components vulnerable to wear as the impact between them increases and continuously advances the severity of wear damage to possibly failure. Vibration analysis in the way of measured signals ensuing from engine body can be implemented for the corresponding bearing fault diagnosis.

Several research have been aimed at study of the dynamic behaviour of mechanical systems with revolute joint and clearance [2, 4–12]. Majority of these investigations are exclusive to analysis of the dynamic response of the slider crank mechanism using the reaction forces in the clearance joint, velocities, and accelerations of the slider under various models in simulation of clearance revolute joints. Earlier studies in this field have studied the impact forces in clearance using dry contact models that excludes the effect of friction and lubrication [5, 8]. Bai and Zhao, adopted the slider crank mechanism to study the effect of clearance on dynamic response of the system using an improved hybrid contact force model and friction force model. Using improved nonlinear stiffness and damping coefficients unrestricted by coefficient of restitution and size of clearance, they compared experimental and simulation results to analyse the systems dynamic behaviour with different clearance [12]. Subsequent works in this area took into consideration inclusion of lubrication in dynamic analysis of the performance of revolute joint with clearance [1, 9, 11]. Schwab, Meijaard and Meijers presented a comparative study of three clearance models including a contact force model and hydrodynamic lubrication model to investigate the effect of joint clearance in dynamic response of the slider crank mechanism [8]. Similarly, Flores et al. also presented a simulation methodology using dry contact with and without friction and a lubricated model to study the influence of the clearances in the responses from these models through analysis of reaction forces and moments [1]. Daniel and Cavalca included the hydrodynamic lubrication in dynamic analysis of the slider-crank mechanism using two combined model; equation of motion and Lagrange method respectively for practical results of the system interaction with the bearing lubrication [11]. Majority of the above research have exclusively focused on slider-connecting rod-crank mechanism and have provided an in-depth study into the

dynamic behaviour of the system but only a limited number have considered vibration analysis for condition monitoring of bearing clearance faults particularly in IC engine connecting rod big end bearing. Fault diagnosis and detection in IC engines are quite challenging because of the complex operation of the engine systems and their associated transient and non-stationary signals [13]. Vibrations ensuing from engine operation are also weak and nonstationary owing to the variable load and sliding speed during operation. However, vibration signals measured from the engine are second order cyclostationary signals which contain information associated with various engine impacts such as those due to clearance in bearings, piston slap and engine valves opening and closures etc. Therefore, analysis of method for condition monitoring and diagnosis of bearing clearance to implement an effective monitoring strategy and obtain bearing operation information from vibration signals, an in depth understanding of the dynamic behaviour is necessary. In this work, the basis of diagnostics is the inclusion of vibration analysis in clarifying the influences of clearance in big end bearing on vibration responses using amplitudes of impacts and corresponding angular position. The impact forces of different enlarged clearance were simulated and after which the accelerations were obtained from the measured frequency transfer functions (FRFs) from the big end journal and the engine body.

2 Dynamics and Kinematics of Piston-Connecting Rod—Crank Assembly

IC engine connecting rod modelled as a slider-crank mechanism transform the piston translation motion to crank rotation motion. Due to higher operating speeds of IC engines, the dynamic effects play a key part in the kinematic and dynamic characteristics of the system mechanism. When considering the simulation of wear induced enlarged clearances in big end bearing, the system can be modelled as two separate subsystems involving the piston and connecting as one part and the other being the crank. The kinematic analysis of the piston connecting rod system requires calculation of the position of mass centre of the joint motion and subsequently their associated velocities and accelerations. These can be derived from the schematic representation of the engine kinematic mechanism shown in Fig. 1.

Where r is the crank radius, l is connecting rod length and φ is the angular position of connecting rod. The connecting rod angular speed and angular acceleration can be determined by speed of crank journal. The increase of φ is negative relative to the X–Y coordinated system attached to engine body, it can be expressed as;

$$\sin(\varphi) = l + r - l \cos(\varphi) - r \cos(\theta) \quad (1)$$

$$\sin(\varphi) = -r/l \sin(\theta) \quad (2)$$

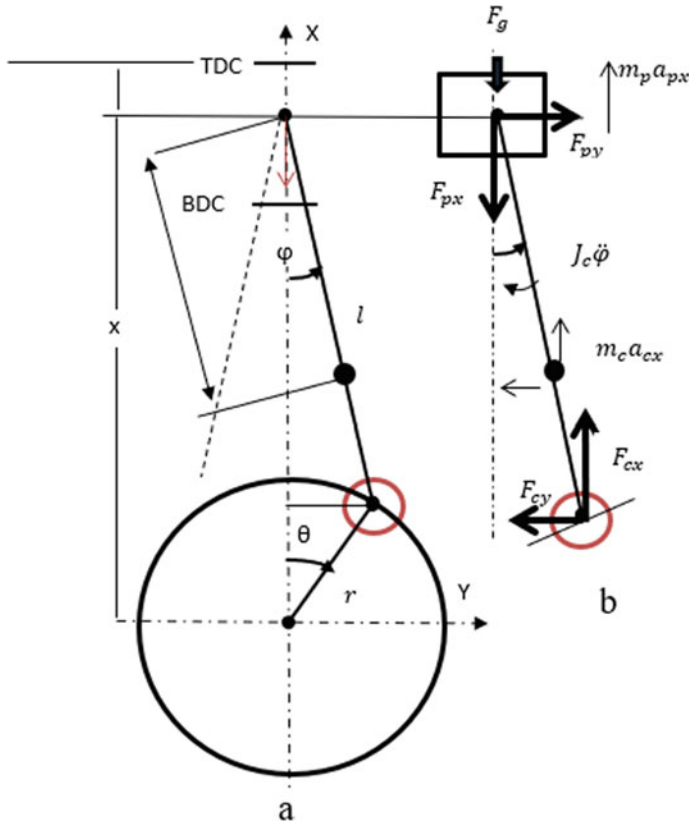


Fig. 1 Schematic representation of engine kinematics **a**; expanded view showing forces of big end bearing **b**

The angular velocity and acceleration of connecting rod can be derived from the displacement of the connecting rod in y and x directions, respectively. The displacement is given as

$$r \sin(\theta) = -l \sin(\phi) \quad \text{and} \quad \cos(\phi) = \frac{1}{l} \sqrt{l^2 - (r \sin(\theta))^2} \tag{3}$$

$$r \cos(\theta) \dot{\theta} = l \cos(\phi) \dot{\phi} \tag{4}$$

The angular velocity and acceleration of connecting rod denoted by $\dot{\phi}$ and $\ddot{\phi}$ are given as

$$\dot{\phi} = -\frac{r \dot{\theta} \cos(\theta)}{l \cos(\phi)} = \frac{-r \dot{\theta} \cos(\theta)}{\sqrt{l^2 - (r \sin(\theta))^2}} \tag{5}$$

$$\begin{aligned}
\ddot{\varphi} &= -\frac{r \cos(\varphi)[\cos(\theta)\ddot{\theta} - \sin(\theta)\dot{\theta}^2] + \dot{\theta}\dot{\varphi} \cos(\theta) \sin(\varphi)}{l \cos^2(\varphi)} \\
&= -\frac{r}{l} \left[\frac{\cos^2(\varphi)[\cos(\theta)\ddot{\theta} - \sin(\theta)\dot{\theta}^2]}{\cos^3(\varphi)} + \frac{r\dot{\theta}^2 \cos^2(\theta) \sin(\varphi)}{\cos^3(\varphi)} \right] \\
&= \frac{r}{l \cos^3(\varphi)} \left\{ \cos^2(\varphi) \sin(\theta)\dot{\theta}^2 - \cos^2(\varphi) \cos(\theta)\ddot{\theta} + \frac{r}{l} \dot{\theta}^2 \cos^2(\theta) \sin(\varphi) \right\} \quad (6)
\end{aligned}$$

In the kinematics of the journal, the motion can be described by the angular motion of the crankshaft. The moment balance equation on the journal can be given as;

$$J_c \ddot{\theta} = r F_{cx} \sin(\theta) + r F_{cy} \cos(\theta) - T_l \quad (7)$$

where J_c is the moment of inertia of the crank, F_{cx} and F_{cy} are the resultant forces in the vertical (x) and horizontal (y) directions, T_l is the driving torque and the friction torque is neglected.

The connecting rod motion in vertical x and horizontal y direction is given as;

$$J_c \ddot{\varphi} - F_{cx} l \sin(\varphi) + F_{cy} l \cos(\varphi) = 0 \quad (8)$$

$$m_c a_{cx} + F_{cx} - F_{px} = 0 \quad \text{and} \quad m_c a_{cy} + F_{cy} - F_{py} = 0 \quad (9)$$

As piston moves downward in X direction.

$$m_p a_{px} - F_g + F_{px} = 0 \quad (10)$$

From the above equations, m_c and m_p denotes the mass of connecting rod and piston, respectively. F_g is the combustion force.

$$F_{px} = F_g - m_p a_{px} \quad (11)$$

The resultant force at the crank journal and bearing in vertical direction is expressed as:

$$F_{cx} = F_g - m_p a_{px} - m_c a_{cx} \quad (12)$$

The bearing force pair in Y direction can be obtained from Eq. 7 as;

$$F_{cy} = \frac{F_{cx} l \sin(\varphi) - J_c \ddot{\varphi}}{l \cos(\varphi)} \quad (13)$$

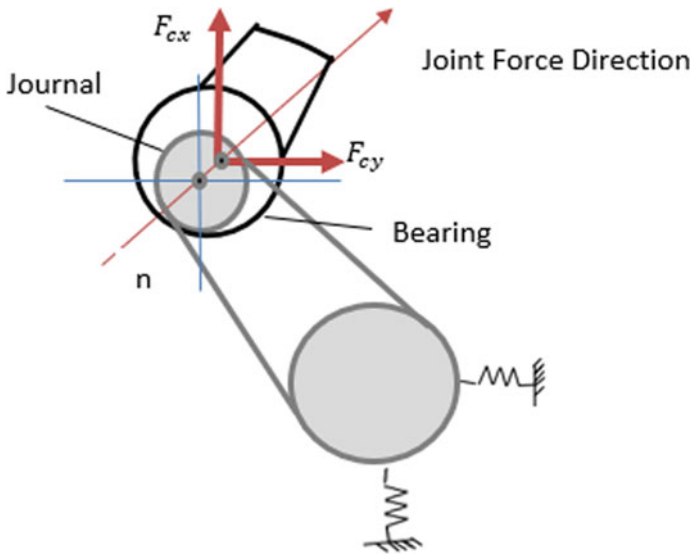


Fig. 2 Illustration of joint force and bearing clearance model

2.1 *Vibro-Impact Analysis of the Big End Bearing Clearance Model*

To observe the impact behaviour between the crank journal and big end bearing, a vibro-impact bearing model is used, in the presence of clearance between the journal and the bearing as shown in Fig. 2. The joint forces ensuing from the inertial effects of the joints can be obtained by dynamic force analysis. In the event of contact between the journal and bearing, creates an impact force and the direction of the force coincides with the normal contact force denoted in Fig. 2 as n . The model assumes a continuous contact between the bearing and journal. The resultant forces F_{cx} and F_{cy} at the crank-rod joint both make up the overlying effects of the applied torque and the associated inertial.

2.2 *Methodology of Numerical Simulation*

The bearing impacts was simulated numerically by solving the kinematic equations from the dynamic model using MATLAB ode45 solver. The differential equations were solved by integration using initial conditions in fixed time step mode. The entire output simulation was sampled at 13 kHz. The simulation was run for different clearance of 0.1, 0.25 and 0.5 mm to simulate enlarged bearing clearance. Three different engine speeds of 1000, 1400 and 1800 rpm were selected typical to medium speed heavy duty engine. The selected output loads of the engine are 40, 80 and

Table 1 Parameter for dynamic simulation of bearing impacts

Parameters	Value	Units
Stroke	0.132	m
Bore	0.103	m
Crank radius (r)	0.066	m
Connecting rod length (l)	0.205	m
Mass of connecting (m_c)	2.5	kg
Moment of inertial of connecting rod (J_c)	0.041	Kg m
Mass of piston + piston pin (m_p)	2.3	kg

160 Nm. The measured cylinder pressure in the simulation was created using a hybrid pressure model. A MATLAB polynomial curve fitting for averaged cylinder pressure as a function of crank angle was used on cylinder pressure measured from experiment. The obtained cylinder pressure is subsequently applied in computation of system indicated torque and total torque resulting from combustion/gas pressure at different speed and load conditions. The equations of motions are solved numerically with the indicated torque as input to the system. The simulation continued with the calculation of the displacement of the bearing and journal using system parameters show in Table 1. The main output of the simulation is the resultant bearing forces and the simulation was run for all three operating parameter of varying speed, load and clearance to study their effect on the resulting bearing force.

3 Results and Discussion

3.1 Effect of Torque on Bearing Performance

The influence of the external torque applied to the system is studied, which is referenced to the cylinder pressure or combustion forces that drives the piston and hence the connecting rod in the slider motion. The external load linearly affects the cylinder pressure during combustion and consequently all related forces. The effect of the torque on the bearing forces was considered in bearing performance analysis via selection of varying torques of 40, 80 and 160 Nm under constant speed and radial clearance of 1400 rpm and 0.1 mm, respectively. Figure 3 shows the bearing forces in both vertical and horizontal directions and the effect of cylinder pressure with variable load increment. As a result of the increasing cylinder pressure, the associated combustion peaks at 360° is seen to increase with higher load. The variation in bearing forces due to the load increase appears only around the combustion event and outside this, the result indicates that the external torques have non-existent or very subtle effect on the bearing forces. The main reason being that the bearing force is predominantly associated with the inertia forces of the slider-crank system components i.e., piston and conrod rather than the combustion forces. This isolated

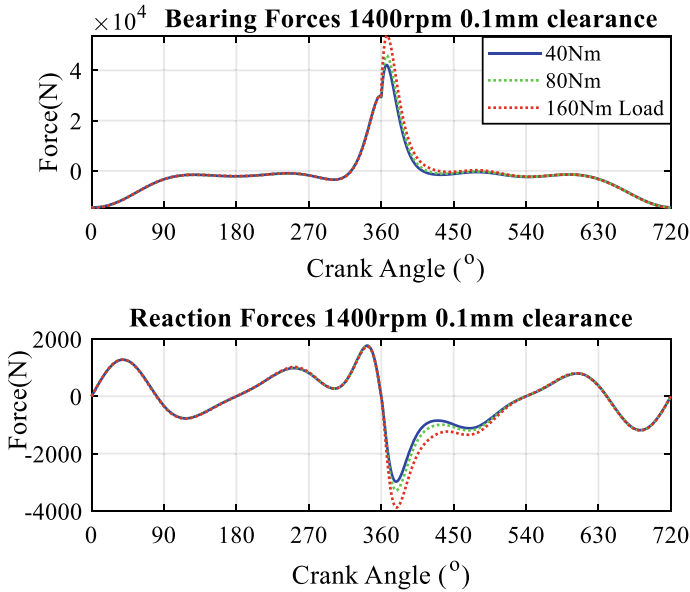


Fig. 3 Connecting rod bearing force in vertical (x) and Horizontal (y) directions at different loads

variation of bearing forces observed in the combustion region might be because of the effect of generated combustion forces on the journal motion which may abruptly shift via the clearance.

3.2 Effect of Speed on the Bearing Performance

Contrarily to the subtle effect of load variation on the bearing forces, the variation of engine speed have more noticeable influence on the bearing force. Figure 4a, b shows the effect of different speeds 1000, 1400 and 1800 rpm on the bearing performance under constant load of 80 Nm, and radial clearance of 0.1 mm. Under the effect of speed, the amount of occurrence where the forces change with direction is largely influenced by the varying engine speed. However, this speed variation is not so obvious in the combustion region. This is as expected, as the main force in the other three strokes, intake, compression, and exhaust are attributed to the inertial force which is solely dependent on the speed of engine. The reaction force (Fig. 4b) is the bearing force in the horizontal direction and is also regarded as the side-thrust force of the connecting rod.

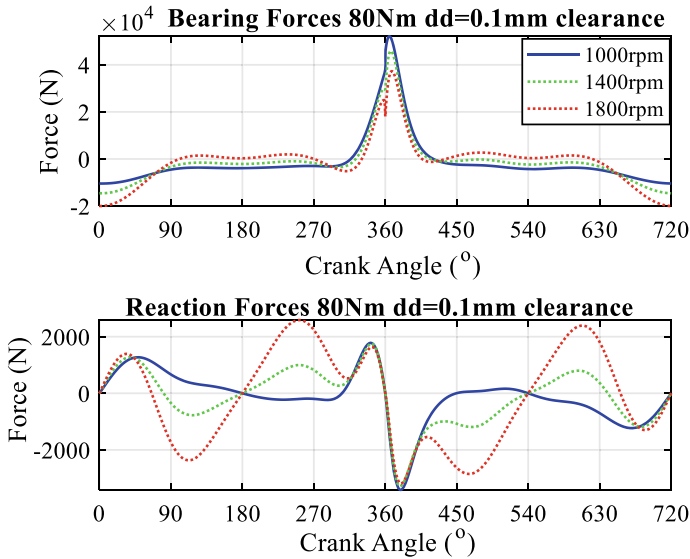


Fig. 4 Connecting rod bearing force vertical (x) and Horizontal (y) directions at different engine speeds

3.3 Effect of Radial Clearance on Bearing Performance

The impact forces in the vertical X and Horizontal Y directions (Fig. 5) were simulated to study their behaviour under various oversized clearance conditions, 0.1, 0.25 and 0.5 mm. These clearances apply to the connecting rod big end bearing and crank journal. An obvious increase in impact frequency is noticed when the clearance is increase to 10 times the normal or minimum clearance. The increase in impacts with increasing clearance appears more under observation of the horizontal impacts which clearly is related to the inertia forces that are associated with bearing impacts. In Fig. 5a the impact between 180 and 270 is hardly noticed but becomes apparent in the horizontal impacts around this same crank position. This very subtle impact noticed in the vertical impact is due to the vertical combustion forces acting in opposition to the inertial forces.

3.4 Simulated Vibration Signal Analysis Using FRF

As the scope of this work is concerned with condition monitoring of engine bearing operation, vibration analysis was included to study bearing impacts using simulated bearing forces. The Frequency response function (FRF) tool was employed to aid the vibration analysis. This was used in obtaining the natural frequencies of the structural test engine similar to the simulated one. Following simulation of the bearing forces

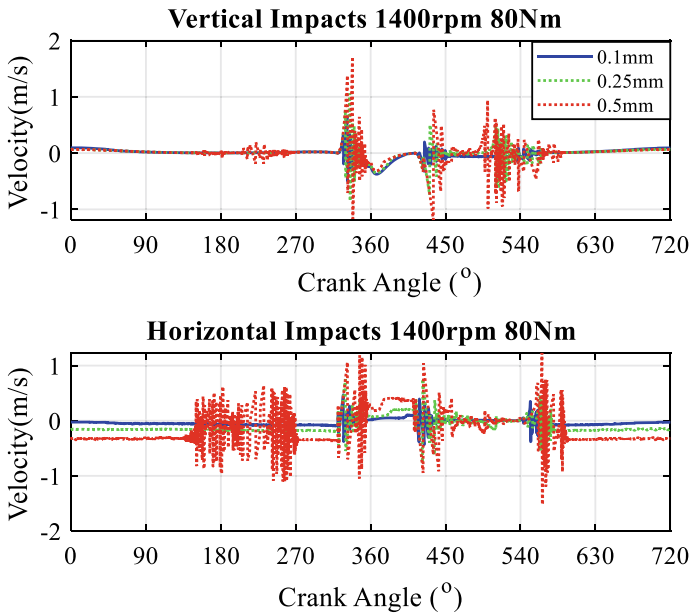


Fig. 5 Velocity of vertical and horizontal bearing impacts at various bearing clearance

in the vertical (x) direction and horizontal (y) directions, their accelerations were calculated by multiplying the bearing forces and measured FRF in the frequency domain. An inverse Fourier transform was subsequently used to produce the accelerations in time domain. Three magnetic pick-up accelerometers were mounted on the crankshaft end, the surface of big end bearing and the engine body, respectively. Excitation force was applied on the big end bearing and journal using hammer hitting test method and the corresponding acceleration was measured from these three points. The vibrations for the bearing forces of increasing clearance are given in Figs. 6 and 7 these are shown with corresponding responses measured from the three accelerometers pick up point (Crankshaft, Big end bearing and Engine body). Slightly higher vibration amplitudes are obtained from accelerometer mounted on the engine bed compared to other measurement points. When the clearance increases from 0.1 to 0.5 mm for simulation case of 1800 rpm engine speed at 80 Nm, the vibration, and amplitudes for the case with higher clearance is shown to slightly increase. The vibration characteristic signals appears to be periodic which may give an indication of reoccurring contact between the journal and bearing with increased clearance. In addition, under the case of increased clearance, the noise level associated with vibration is higher compared to the normal clearance condition.

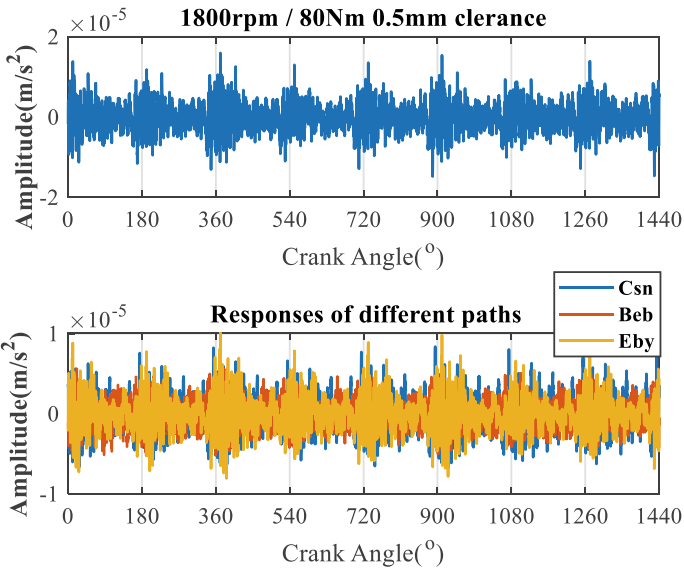


Fig. 6 Bearing impact vibrations for 0.5 mm clearance at 1800 rpm engine speed/80 Nm load. **(Plot a)** Summed responses from measurements points. **(Plot b)** Responses from different points, Crankshaft (csn), Big end bearing (Beb) and Engine body (Eby)

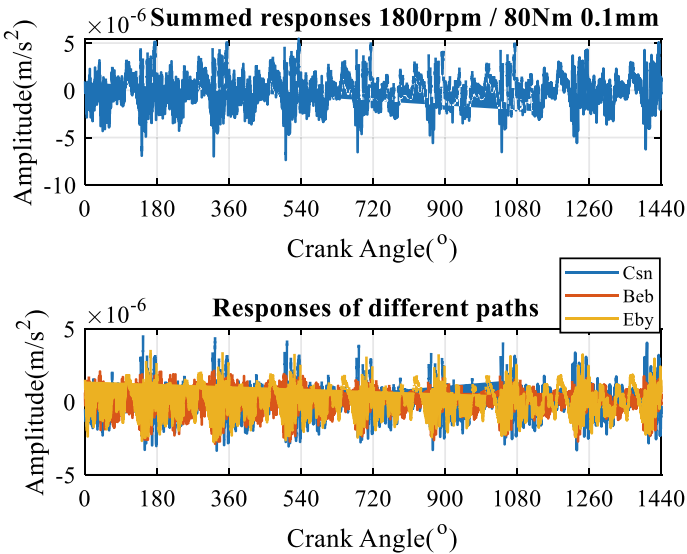


Fig. 7 Bearing impact vibrations for 0.1 mm clearance at 1800 rpm engine speed/80 Nm load. **(Plot a)** Summed responses from measurements points. **(Plot b)** Responses from different points, Crankshaft (csn), Big end bearing (Beb) and Engine body (Eby)

4 Conclusion

The primary focus of this work is to study the influence of enlarged clearances in the dynamic response of IC engine bearing operation for condition monitoring of bearing performance and associated faults in heavy duty diesel engines using dynamic simulation of the bearing impact forces. The model adopted is applied to connecting rod big end bearing with clearance between the bearing and the crank journal. The results presented shows that there are higher effect in the reaction force or forces in Y directions of the bearing force pair which consequently leads to vibrations of the system. The increase of clearance between the bearing and journal can influence the dynamic response of the engine system and vibration analysis to some extent can capture the various effects observed in the dynamics of the system.

References

1. Flores, P., et al.: A study on dynamics of mechanical systems including joints with clearance and lubrication. *Mech. Mach. Theor.* **41**(3), 247–261 (2006)
2. Erkaya, S., Uzmay, İ: Experimental investigation of joint clearance effects on the dynamics of a slider-crank mechanism. *Multibody Sys. Dyn.* **24**(1), 81–102 (2010)
3. Chen, J., Randall, R.B.: Intelligent diagnosis of bearing knock faults in internal combustion engines using vibration simulation. *Mech. Mach. Theor.* **104**, 161–176 (2016)
4. Flores, P., et al.: Numerical and experimental investigation on multibody systems with revolute clearance joints. *Nonlinear Dyn.* **65**(4), 383–398 (2011)
5. Soong, K., Thompson, B.S.: Theoretical and experimental investigation of the dynamic response of a slider-crank mechanism with radial clearance in the gudgeon-pin joint. *J. Mech. Transm. Autom. Des.* **112**(2), 183–189 (1990)
6. Lankarani, H.M., Nikravesh, P.E.: Continuous contact force models for impact analysis in multibody systems. *Nonlinear Dyn.* **5**(2), 193–207 (1994)
7. Bauchau, O.A., Rodriguez, J.: Modeling of joints with clearance in flexible multibody systems. *Int. J. Solids Struct.* **39**(1), 41–63 (2001)
8. Schwab, A.L., Meijaard, J.P., Meijers, P.: A comparison of revolute joint clearance models in the dynamic analysis of rigid and elastic mechanical systems. *Mech. Mach. Theor.* **37**(9), 895–913 (2002)
9. Flores, P., Ambrósio, J., Claro, J.P.: Dynamic analysis for planar multibody mechanical systems with lubricated joints. *Multibody Sys. Dyn.* **12**(1), 47–74 (2004)
10. Khemili, I., Romdhane, L.: Dynamic analysis of a flexible slider-crank mechanism with clearance. *Eur. J. Mech. A/Solids* **27**(5), 882–898 (2008)
11. Daniel, G.B., Cavalca, K.L.: Analysis of the dynamics of a slider-crank mechanism with hydrodynamic lubrication in the connecting rod-slider joint clearance. *Mech. Mach. Theor.* **46**(10), 1434–1452 (2011)
12. Bai, Z.F., Zhao, Y.: Dynamic behaviour analysis of planar mechanical systems with clearance in revolute joints using a new hybrid contact force model. *Int. J. Mech. Sci.* **54**(1), 190–205 (2012)
13. Antoni, J., Daniere, J., Guillet, F.: Effective vibration analysis of IC engines using cyclostationarity. Part I—a methodology for condition monitoring. *J. Sound Vibr.* **257**(5), 815–837 (2002)

Impeller Wear Diagnosis in Centrifugal Pumps Under Different Flow Rate Based on Acoustic Signal Analysis



Alsadak Daraz, Fengshou Gu, and Andrew D. Ball

Abstract Centrifugal pumps are commonly used in pipelines with moderate head and discharge requirements for hydraulic transportation of liquids and solids over long distances. The performance characteristics of the pump and erosion wear are the most significant design and selection parameters. Mechanical wear can cause considerable damage to the impeller in the pump, reducing the pump's lifetime and efficiency. This paper investigates the impeller wear of a centrifugal pump using an acoustic mechanism at different flow rates. As the peripheral velocity or circling radius at the impeller surface increases, a small amount of wear in the impeller inlet region rapidly develops. The uniform corrosive wear area is defined as the portion of the impeller where the impact velocity is less than the critical value, and it expands as the impeller velocity increases. In fact, wear mechanisms differ from one region to the next; when the tangential portion of the impact velocity is high, the impeller wears out faster. Because of the high capability of noise reduction when modulating the signals, modulated signal bispectrum (MSB) analysis is used for extracting the incipient fault signature. The experimental results show that the diagnostic features produced by modulated signal bispectrum allow for early detection of the initial impeller wear fault. Furthermore, the MSB study of acoustic signals demonstrates the efficacy of airborne sound-based monitoring. It offers a strong attestation of complete distinction between safe and defective conditions at various flow rates.

Keywords Impeller wear inlet vane · MSB analysis · Condition monitoring

A. Daraz · F. Gu (✉) · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK
e-mail: f.gu@hud.ac.uk

A. Daraz
e-mail: Alsadak.Daraz@hud.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_32

385

1 Introduction

In order to follow the progress of technology, the centrifugal has become a significant fluid machine in modern industries such as petroleum, chemicals, metallurgy, and aerospace [1, 2]. In the meantime, centrifugal compressors are becoming increasingly large-scale, high-speed, and remotely controlled [3]. Centrifugal pumps condition monitoring (CM) is critical for avoiding unanticipated failure, lowering disruption expense and improving the availability of the machinery [4].

Numerous academics have devised various methods for detecting centrifugal pumps faults [5, 6]. The vibration signature analysis methodology for detecting and diagnosing defects in machine components is the most common and commonly utilised methodology for CM [5, 7]. The existing monitoring methods such as simple vibration analysis are insufficient for identifying potential pump failures and prevent periodic breakdowns and service interruptions that cause large pumps to shut down. Machine vibrations because of the flow excitation and structure-borne sound in roller bearings are used to track the conditions of the parts such as impellers, bearings, shafts of the pump, and drive pieces. Process variables such as water flow, pressure (suction, discharge), velocity, motor speed, temperatures, bearings, and leakage measurement are also examined. The machine receives condition details automatically [8].

The blade loss, on the other hand, is a common occurrence. The impeller is the most critical component since it converts kinetic energy into pressure energy. However, the impeller is subjected to centrifugal forces, friction forces, and gas pressure, both of which can cause cracks. As stated by statistics, the blades are responsible for 65% of centrifugal compressor failures. Furthermore, 40% of blade fatigue failures are still unknown [9]. Blade fatigue failures are also caused by fluid-induced vibration, which is evident in acoustic resonance, unsteady flow, revolving stalls, and flutter [10, 11]. High local pressures are produced when cavitation bubbles implode on the impeller or other pump components, which can exceed the material's fatigue strength, yield point, or ultimate strength. If this happens, the material will pit, a process known as cavitation erosion. With more corrosive media, material degradation accelerates [12]. Bross and Addie [13] constructed a computational model to investigate impeller suction sealing wear behaviour as a function of impeller design factors. They found that the impeller suction sealing area experiences the most wear. When the leakage flow rate increases, the intensity of local wear for the nose sealing gap grows, reducing the pump's life and efficiency. Walker and Bodkin [14] found that using the weight loss approach to quantify the wear in the field did not accurately anticipate the localised wear pattern. They measured the maximum depth of material loss to assess side liner erosion wear. They also evaluated the performance of side liner impellers with three different designs, finding that the wear rate of a high-efficiency impeller is proportional to the square of the particle size. The influence of particle size on standard and decreased eye diameter impellers is negligible.

Furthermore, some wear occurs at the place where the impeller and casing tolerances almost come into touch regularly. The erosion induced by liquid leaking through this little clearance causes this form of wear. Furthermore, a tiny particle

that is not apparent to the naked eye might induce impeller wear and erosion. Again, bigger particles such as sand, rust, and boiler scale and dissolved minerals in hard water can all contribute to the formation of these types of faults. The two issues will necessitate a substantial quantity of real-time data analysis and a sophisticated network of acquisition and interpretation software. In contrast, the first three problems are primarily reliant on hardware quality and performance. Kim et al. [15, 16] demonstrate how a single microphone combined with a low-profile monaural construction can be used to locate the sound source.

The measurement of vibration, acoustic, temperature, speed, wear, pressure, and torque characteristics has improved fault identification and diagnosis methodologies for CM. Vibration and acoustic techniques are intrinsic to the measuring operation, and they provide multiple benefits in terms of non-intrusive signal measurement, rich information from many sources of signals, and high jumps in computational power in the arena of advanced signal approaches [17]. The most challenging aspect of the vibration procedure is trying to locate and implementing the accelerometer in the correct location for CM. Airborne sound, on the other hand, has the attractive advantage of being a non-contact approach that does not require direct access to the equipment. Furthermore, in comparison to accelerometers, the microphone, which can cover a large variety of comprehensive information on mechanical parts and requires a small number of sensor channels [18, 19].

This study focuses on the study of using acoustic signals of diagnostic impeller faults in centrifugal pumps. In particular, the MSB analysis will be the main signal process tools for suppressing random contents of acoustic signals due to turbulent flows and background noise and thereby enhancing the deterministic modulations due to periodic interaction between impeller and low stream. The acoustic sources and its signal behaviours are overviewed analytically under common impeller fault conditions. The monitoring performances is then evaluated based a dataset from a laboratory test system.

2 Sources of Hydraulic and Mechanical Noise

Mechanical Noise sources are including pump surfaces or vibrating parts generated by pressure changes in fluid or air. The impeller, seal rubs, volute casing, faulty rotators, unbalanced rotors or stators and vibrating pipe, or bearings walls are samples of mechanical noise causes [20, 21]. The inappropriate assembly of coupling often causes mechanical noise when the speed of the pumping is doubled. The high vibration resulting friction from the movement of the impeller, bearing or seals is caused when the pumping speed is almost close to the high speed. The high-pitched noise is a signal of recognition if frictional action occurs [21].

Fluid noise sources are the pressure vibrations caused by the fluid's unstable flow. The impellers rotating in a guided volute case contains centrifuge pumps. The fluid movement within a pump Like turbulent as well as three-dimensional [20, 22]. Due to this turbulent fluid behaviour, pressure fluctuations are generated. Resulting, the

flow-induced noise that is passed through the pipes will therefore be produced. The pulses might produce wideband frequency or discrete frequency. The broadband energy is mostly driven by a high flow velocity and the pump impeller is the reason for discrete energy. Due to turbulence, pulsation of pressure, a water hammer, and cavitation, as well as the interaction between impellers and cut wear, flux-induced noise can emerge [20, 22].

A hydraulic pumping system generates audible noise due to a combination of hydrodynamic and mechanical forces. The two types of background noise are transported to the surrounding air as airborne noise through the system through its liquid and structure. Because the pump is the essential participant, it is critical to recognise the pump's contribution to the noise produced [23, 24]. Pump noise is determined by the type of pump, its geometry (structure and size), and operating conditions (load and speed). The interactions between the pump's impeller edge and diffuser vanes are the main sources of noise in a centrifugal pump. Blade passage frequency (BPF), pressure variations due to turbulence, flow friction, flow separation, and vortices in the axial and radial clearances, notably between open or semi-open rotors and the fixed component of the casing, are all key sources of noise. Depending on the pump head being produced and the distance the site of operation is from the BEP, the magnitude and frequency of background noise varies from one pump to another, even across two seemingly identical pumps [23, 24].

The pressure pulsation produced in an operating pump cavity is strongly related to its vibration as well as noise levels due to a present trend in rising rotating speed and power; its ability to be separated into hydrodynamic phenomena in the pump's operating cavities have been categorized into [25, 26].

2.1 Hydrodynamic Relationships Between Impeller Flow and Volute

Because the fluid interacts with the impeller's vanes as well as stationary volute chambers, the fluid's pressure pulsates predominantly with the impeller's structure. As a result, in particular, the vane number will be linked to the pulsation rate [27].

At this point, the flow velocity is equal to the vane angle, resulting in less flow contact with the impeller. Any BEP activities can cause a misalignment of the liquid velocity and the impeller's vane angle. As a result, additional friction and vibrations will occur [26].

To help impeller defect identification, appropriate features can be designed based on these properties. The frequency with which the blade passes frequency (BPF) characteristic is presented in the Equation:

$$BPF = kzfr \tag{1}$$

where k denotes the number of harmonics and z is the blade number, and f_r is the speed of pump rotor. While the centrifugal pump is operating at its maximum capacity, the amplitude of pressure pulsation is also kept to a minimum (BEP) [28].

2.2 *The Circulation of a Vortex*

Due to the high flow speed, the fluid flow becomes unstable and generates a local vortex when the flow cross-section mutation occurs. While turbulent flow and vortex are present, vibration is caused by water particle action, mutual effect, engaging with the pump, valve body wall, pipeline, or other interactions [25, 28]. Such vortex can also produce vortex shedding and result in an oscillating flow at the vortex shedding frequency calculated by

$$VSF = S_n V / D$$

where D denotes dimension of obstruction, V is flow velocity and S_n is Strouhal number with a value range between 0.213 and 0.533. As the flow field in impeller passages varies greatly, the VSF oscillations can be in a wide frequency range and often exhibit unsteady behaviours.

2.3 *Cavitation*

Whenever a centrifugal vibration is clear over background noise, it is often observed that cavitation is a problem [29]. Cavitation occurs when the pressure is less than the local vapour pressure in a location in a liquid. It is regarded as a natural event when local pressure lowers due to external influences, which lead to fluid vaporisation and then a sudden collapse of the bubbles. The quick collapse and uneven disintegration of these bubbles leads to surface damage because of the high concentration of the generated energy [30]. In the design of equipment such as pumps, hydrofoils and turbines [31] this phenomenon has to be guarded against. The implementation and productivity of the pump would be adversely affected if cavitation appears. The presence of cavitation specifically affects pump productivity and can result in erotic, vibrative, hydraulic and pumping performance impairment [32, 33].

2.4 *Wear by Fluid Velocity*

Because of internal fluctuations, turbulence and wall friction, the fluid velocity within the volute is not constant. These are the prime causes why fluid noise is produced

in centrifugal pumps. Hence the fluid-induced noise and the fluid's unstable flux conduct in a centrifuge pump [20]. Small impellers with high motor speeds may create the necessary pump pressure. As a result, large fluid velocities are produced, which raises the wear on pump parts beyond what is ideal. Furthermore, the impeller is subjected to rapid wear as a result of the high tip fluid velocity. Moreover, high-velocity fluid is responsible for approximately 95% of the time for abrasive wear in pump parts [34].

2.5 Pressure Pulsations

Pressure pulsations are produced by ordinary pumping systems and are caused mainly by unsteady flow and fluid turbulence which caused by rapid rotor movement within the pump [22]. The magnitude of these pulses depends on the distance between rotor blade tips the volute lips [20]. Moreover, a large number of parameters allowing for a pressure pulse to be generated. Even when the pump working at highest efficiency point (BEP), and its functioning nominally (NPSH etc.), turbulence at the impeller tips and high flow velocity still allows to produce pulsations. When the operating condition varies, however, due to different sources, such as cavitation, secondary flow (recirculation) and unsteady flow etc., pulsations may emerge in the signal [20, 22].

3 Fault Simulation

The main causes of impeller loss are solid–fluid interactions, which result in wear and corrosion on the material's surface. Two defects were simulated in this study: impeller inlet vane wear with two severity levels: low and high in a closed type impeller of a centrifugal pump. By eliminating a few of the materials in the impeller's side, an imbalance was introduced into the exterior diffuser vanes, as seen in Fig. 1.

The approach for fault identification based on airborne sound signal analysis is explained in this study. The structure shown in Fig. 2 has three stages: the first shows the simulate impeller defect, which is inlet vane wear; the second displays the way of analysis, which is dependent on popular spectrum analysis and MSB analysis; and the third step compares these approaches. The final stage of the proposed method is fault detection.

4 Test Procedure

The test rig was designed to assess the monitoring of a pump system's output using various measures such as vibration and acoustic signals. The goal was to see how impeller wear influenced the centrifugal pump's efficiency. The experimental data

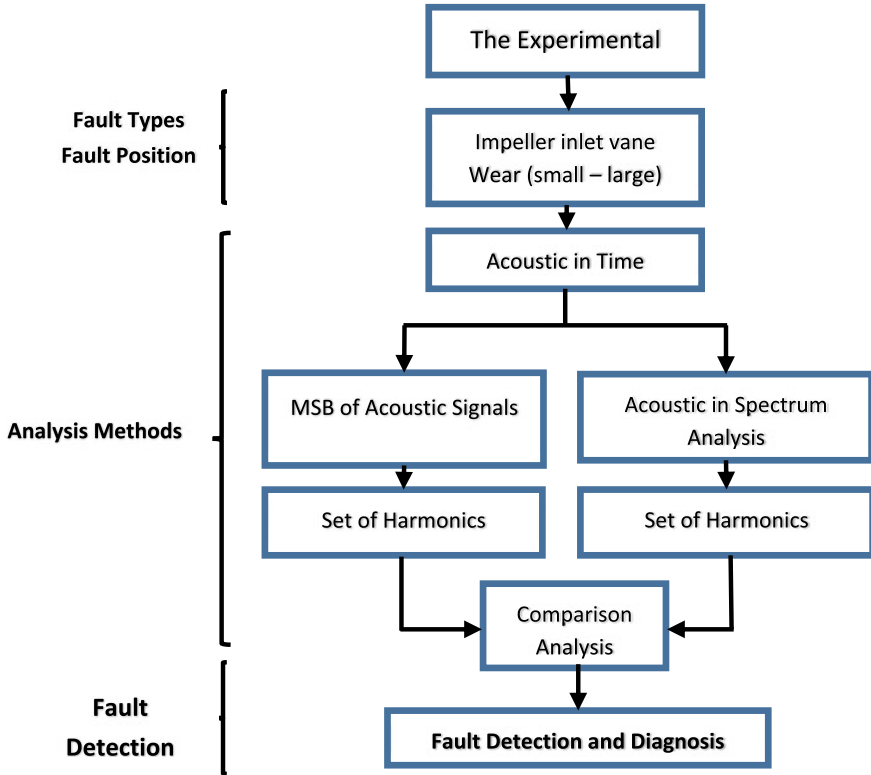


Fig. 1 Framework of the procedure for impeller fault detection

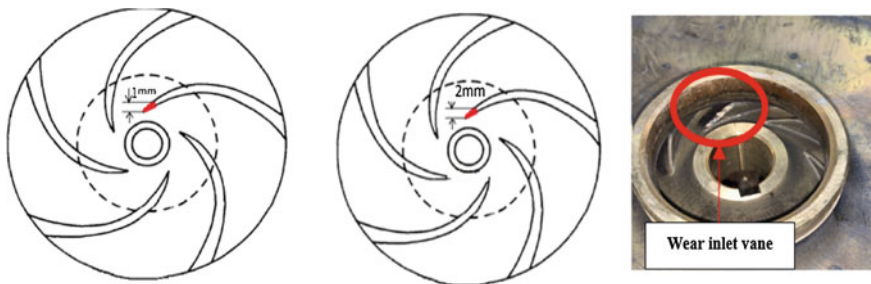


Fig. 2 Wear initial vane fault simulation of impeller

was obtained from impeller wear in the form of airborne sound. The test-rig shown in Fig. 3 was used in this experiment.

As shown in Fig. 3, a test rig for acoustic monitoring consists of a centrifugal pump, pips, water tank, suction valve, two pressure transducers, hydrophone, and flow metre. The experimental datasets were gathered from two test cases, both of

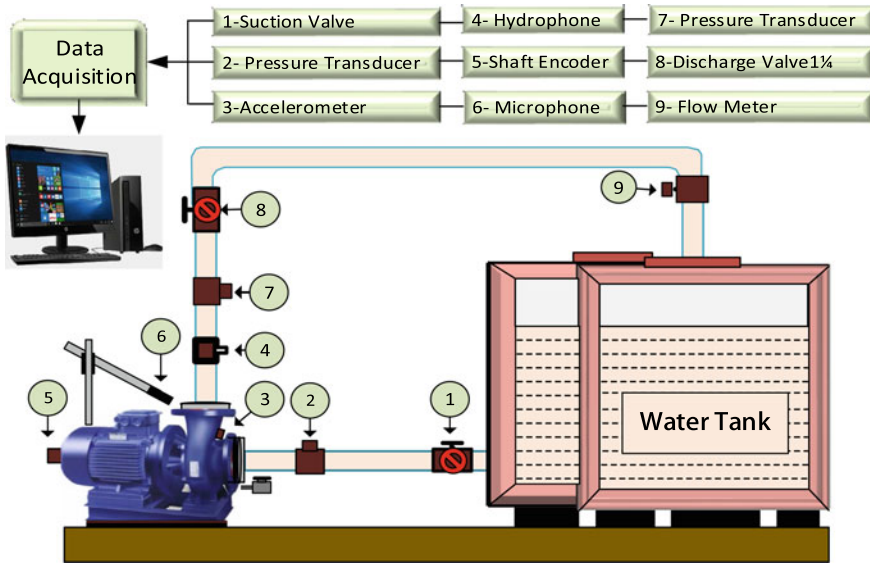


Fig. 3 Schematic diagram of the test-rig

which were free of impeller wear. The calculated data was collected using a data acquisition method (DAS). DAS was linked to seven networks. With a sampling rate of 96 k and a fixed motor speed of 2900 rpm, datasets of 40 s were collected. Under 10 bar of pressure, the pump will produce 250 L/min of water flow. This study looked at the efficiency of a pump at nine different flow rates (0, 50, 100, 150, 200, and up to 450) l/min. In both the suction and discharge lines, two pressure sensors were mounted. An accelerometer was also mounted vertically on the pump casing to measure vibration. On the discharge line, the hydrophone was mounted. The vibration and sound of pump components was measured using a microphone placed 50 mm away from the pump body. Furthermore, a flow sensor was mounted on the discharge line to calculate the volume of liquid flow rate. The rotational speed of the centrifugal pump was measured using an encoder mounted on the motor shaft.

5 Results and Discussion

5.1 Impact Impeller Wear on Performance Curve

The efficiency of a centrifugal pump against a head as a function of flow rate in both healthy and faulty cases. Figure 4 depicts the impeller wear of a centrifugal pump at various flow speeds and under various conditions. Since the pump head and flow rate are directly proportional to the pump's efficiency, the flow rate curve of impeller

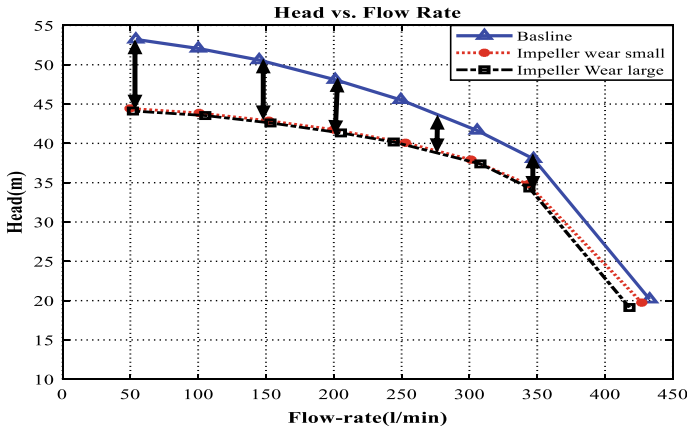


Fig. 4 Pump impeller wear on performance curve

wear fault is steadily declining at low flow rates. However, the pump curve is clarified by the degradation in pump output due to impeller wear defect.

5.2 Time Domain of Acoustic Signals

Figure 5 depicts the airborne sound signals of impeller wear in the time domain in both the safe and defective cases at various flow speeds. The disparity between the amplitude of healthy and defective cases can be observed, with the amplitude of the faulty case being higher than baseline at various flow rates, which may indicate a system failure. In addition, the acoustic signal generated by the sliding surface’s deformation. Because of the waveform’s high randomness, it’s difficult to pinpoint the source of the abnormality.

5.3 Spectrum of Acoustic Signals

Figure 6 depicts the frequency domain of the airborne signal in both healthy and defective impeller wear fault cases at various flow rates. The typical fault frequencies the shaft rotatong frequency f_0 (48.3 Hz) and blade or vane passing frequency νpf (338.3 Hz), as well as their associated harmonics, are clearly visible and substantially higher than the baseline levels. As a result, the acoustic signal’s spectrum reveals the existence of an impeller wear fault. These frequency peaks, like the vibration signal, may be polluted by noise produced by fluid flow or other device components. However, background noise from fluid and other pump components can contaminate and influence the fault frequencies.

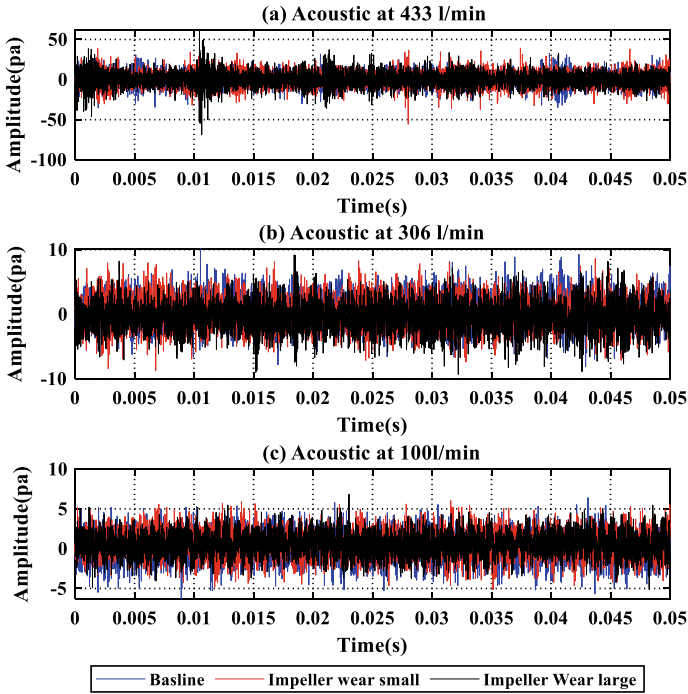


Fig. 5 Acoustic signal in the time domain

5.4 Impeller Inlet Vane Defect Detection Based MSB Analysis

This part intends to examine the detection of impeller wear under different flow rates based on MSB.

(1) Impeller Wear Fault Detection and Diagnosis by MSB Analysis

MSB analysis applied for impeller failure under various conditions to improve fault diagnosis and identification. Figure 7 indicates the magnitude and coherence of the MSB of the acoustic signals at a flow rate of 450 l/m. The MSB’s merits use its superiority to eliminate random wideband noise, so it is able to boost the discrete components as corresponding MSB coherences, the prominent peaks matching to the fundamental frequency of 48.3 Hz and its harmonics, as well as the vane passing frequency of 338.3 Hz, are particularly noteworthy. Where the baseline magnitude is less noise than the defective one suggesting the existence of impeller fault. In addition, as in Fig. 8, the higher pulses are seen due to mechanical and hydraulic effect. Therefore, the consequence of MSB coherence obtained from the average of the two components was achieved. The findings show that the impeller wear fault tends to be simple peaks relative to the frequency of the fault, while in the case of

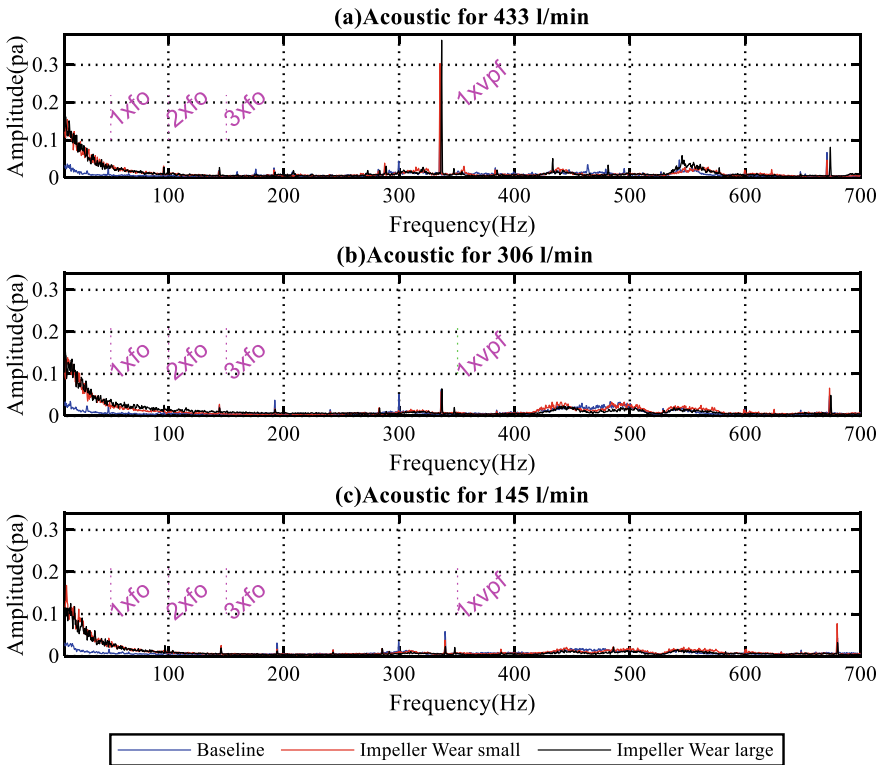


Fig. 6 Acoustic signal in the spectrum domain

baseline, no peaks exist. Furthermore, the outcome of the magnitude and coherence of MSB indicates the distinctive difference between good and deficient cases, thus lower peaks of coherence of MSB. In addition, the higher peaks and pulses triggered by high hydraulic asymmetry and mechanical pulses suggest that the impeller wear fault is present.

(2) **Comparison of Fault Detection of Impeller Wear Based on Spectrum and MSB Analysis**

Figure 9 plots the magnitudes of the first five harmonics of the drive shaft’s rotational frequency of 48.3 Hz as a function of flow rate for the baseline and two seeded faults. The first harmonic separates the baseline and the defective conditions in a clear and meaningful way, but it does not effectively distinguish the two fault conditions. The second harmonics display a consistent distinction between baseline and defective conditions as well, but the difference is much smaller than the first harmonic. The second harmonic can’t tell the difference between the two fault conditions either. However, the baseline signal produces a higher amplitude peak than any fault condition for the 3rd, 4th, and 5th harmonics.

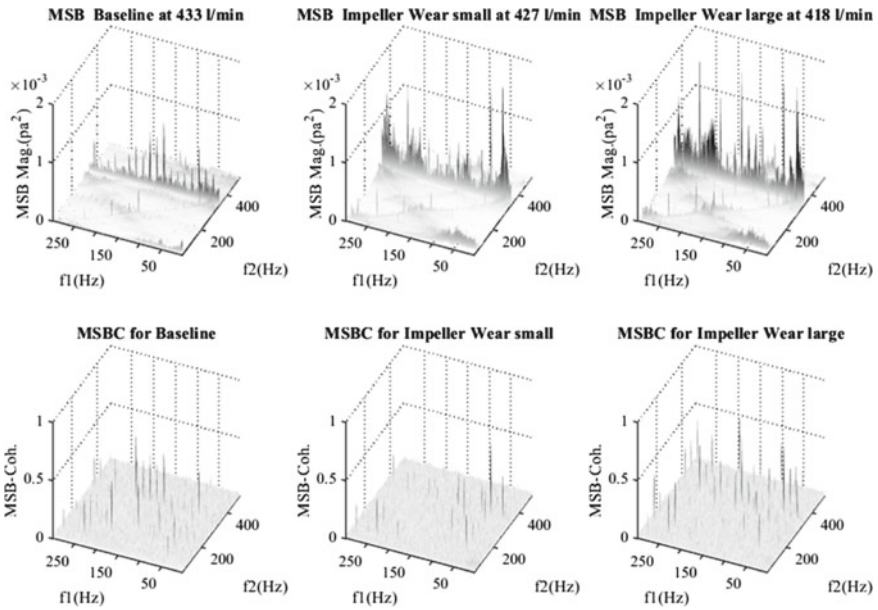


Fig. 7 MSB magnitude and coherence of acoustic signals for at flow rate 433 l/min

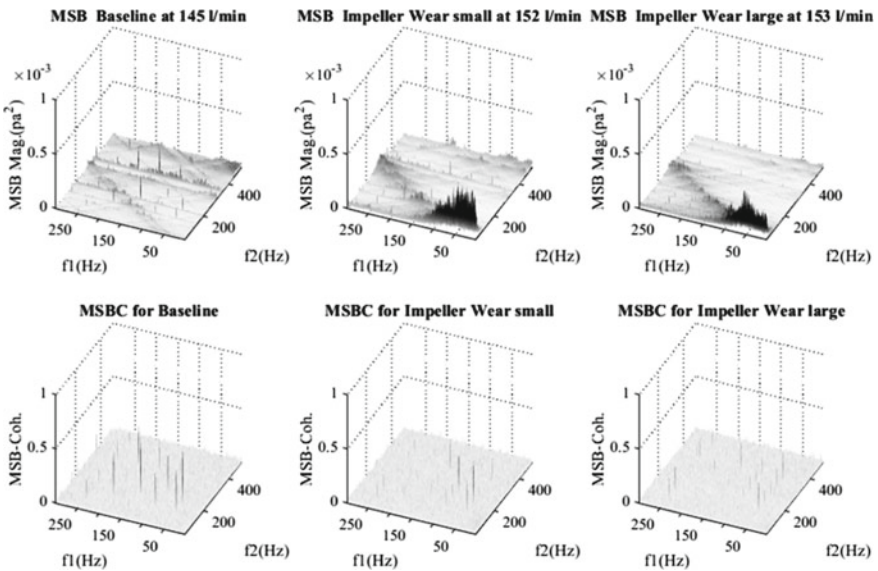


Fig. 8 MSB magnitude and coherence of acoustic signals for at flow rate 150 l/min

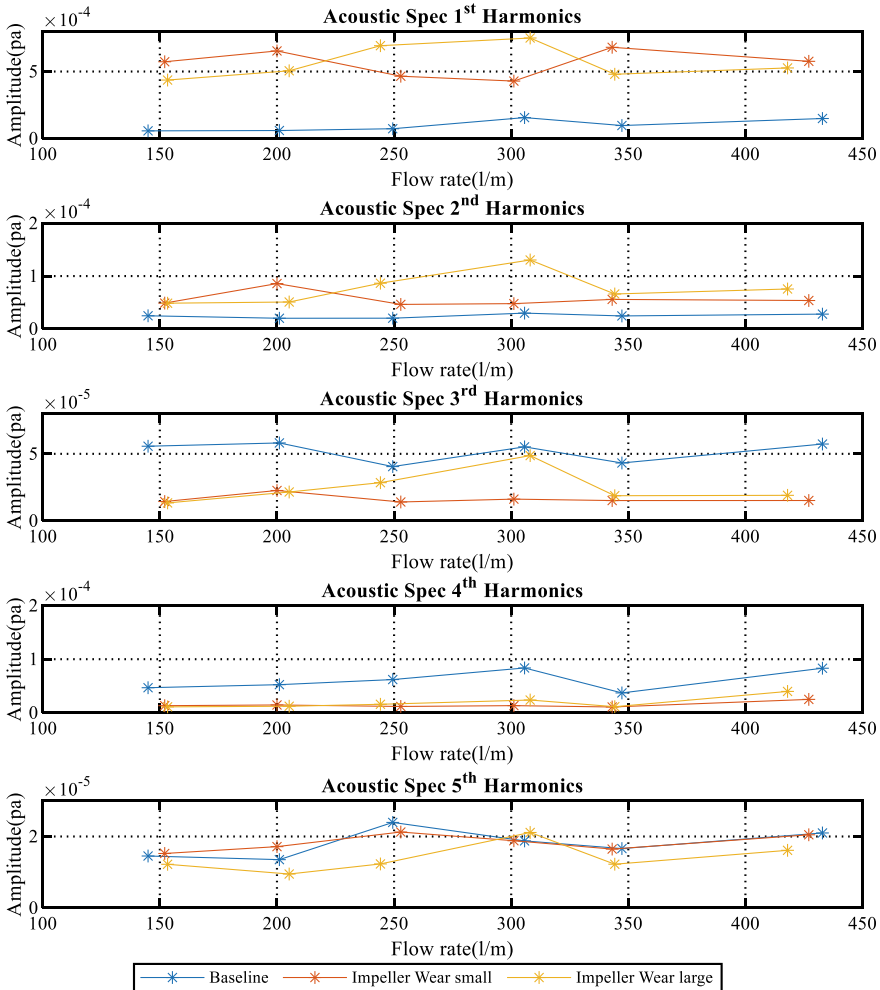


Fig. 9 Acoustic power spectrum plots for the first five harmonics of 48.3 Hz with flow rate

The magnitudes of the MSB peaks for the first five harmonics of the shaft drive frequency (48.3 Hz) derived from the baseline and two seeded faults are shown in Fig. 10. Furthermore, compared to a spectrum study under various flow speeds, the MSB result of impeller wear provides useful data and a good curve pattern for separating all harmonics in both safe and defective circumstances. There is a strong separation between baseline and fault conditions for the first, particularly the fifth harmonics (except at a flow rate of 350 l/min for the first harmonic). However, there is no consistent separation of the two faults. At a flow rate of 350 l/min, the 2nd, 3rd, and 4th harmonics display no noticeable separation of the minor fault and baseline and only distinguish the baseline and large fault.

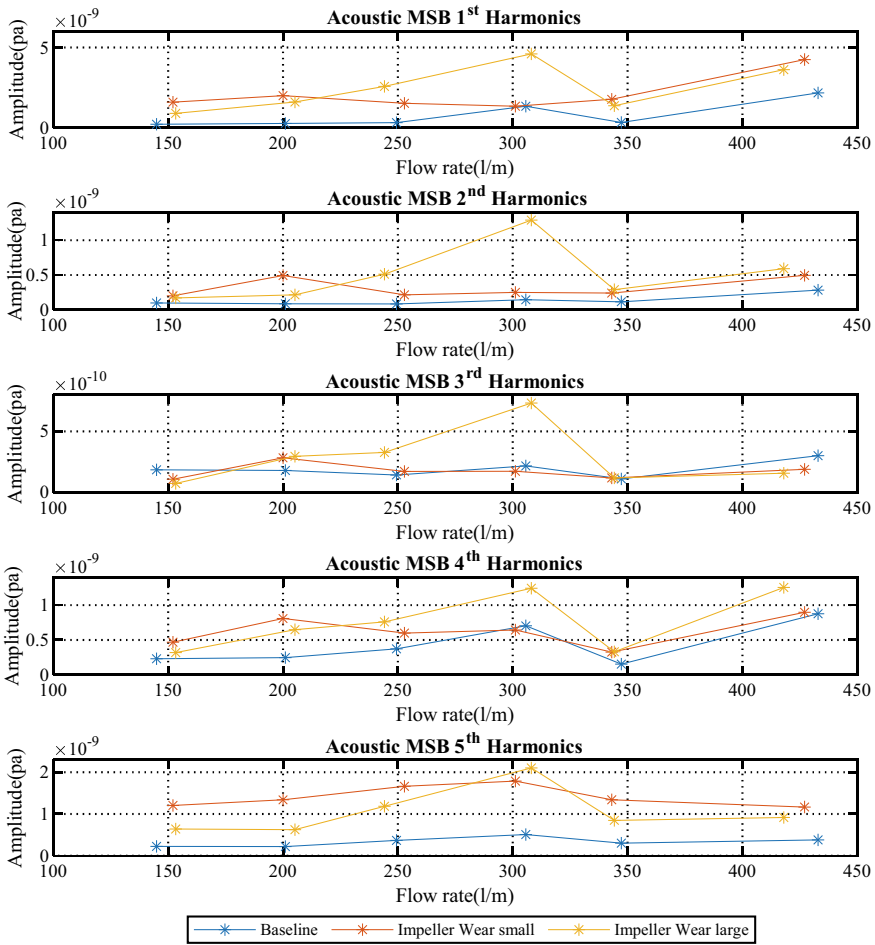
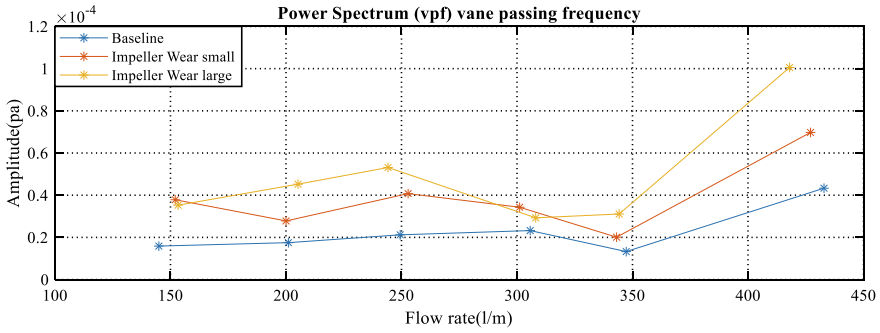


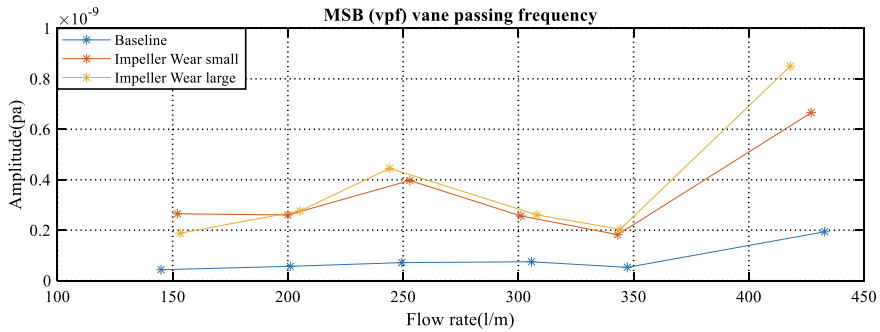
Fig. 10 The MSB of the first five harmonics of 48.3 Hz in the acoustic signals with flow rate

The MSB performs better at low frequencies for all flow rates due to its ability to eliminate wideband, low-frequency noise. As the flow pressure drops, the vibration amplitude for the large impulses of the impeller wear fault of the pump increases. As a result, it demonstrates that MSB analysis can extract characteristic modulations indicative of impeller wear by largely removing noise effects on signals.

Figure 11 compares the power spectrum and MSB examination of the acoustic signals for the baseline and two seeded wear faults at the vane passing frequency (338.3 Hz). There is a clear distinction between baseline and impeller wear defects for both the power spectrum and MSB plots. The findings in Fig. 11b show a more thorough separation and a good distinct distinction between the health and two plots of impeller faults. The MSB results indicate a strong separation between the baseline and faults for all flow speeds, but not between the two-seeded defects. Despite the



(a)



(b)

Fig. 11 a Power spectrum and b MSB for the vane passing frequency (338.3 Hz) of the acoustic signal

fact that the sound signals were distorted by background noise, it demonstrated the ability to extract modulation characteristics of impeller wear faults. The results of the experiment show that the acoustic signal outperforms vibration analysis, whether analysed using MSB analysis or the power spectrum.

6 Conclusion

The experimental analysis of airborne sound signals linked to impeller small and large inlet vane wear was discovered in this study. Furthermore, fault identification and diagnosis approaches were carried out separately using common spectrum analysis and advanced MSB analysis. At the fundamental shaft frequency and their harmonics as well as the blade passing frequency. The BMS produces better results than the spectrum. It also provides a good separation between all harmonics at different flow rates, which is useful for online monitoring. As a result, the proposed approach will successfully identify and diagnose the impeller wear fault, which is triggered

artificially by the test pump. Furthermore, MSB's unique ability to reduce noise and demodulate nonlinearity allows for this high diagnostic performance quality. As a result, a good distinction of impeller wear defects can be achieved by averaging MSB peaks in the low-frequency range. Furthermore, the impeller defect caused by fluid flow interaction through the inlet vane suggests more definite changes due to the wear fault, which causes a significant decrease in the flow rate curve.

References

1. Huang, Z., Wang, X.: Turbine Compressor. Chemical Industry Press, Beijing, China (2004)
2. Daraz, A., et al.: Detection and diagnosis of centrifugal pump bearing faults based on the envelope analysis of airborne sound signals. In: 2018 24th International Conference on Automation and Computing (ICAC). IEEE (2018)
3. Kammerer, A.: Experimental Research into Resonant Vibration of Centrifugal Compressor Blades. ETH Zurich (2009)
4. Hartigan, K.: Inside monitoring to extend pump life. *J. World Prehistory* **2008**(502), 28–32 (2008)
5. Sakthivel, N., Sugumaran, V., Babudevasenapati, S.: Vibration based fault diagnosis of monoblock centrifugal pump using decision tree. *Exp. Syst. Appl.* **37**(6), 4040–4049 (2010)
6. Kamiel, B., et al.: Impeller fault detection for a centrifugal pump using principal component analysis of time domain vibration features (2005)
7. Peck, J.P., Burrows, J.: On-line condition monitoring of rotating equipment using neural networks. *ISA Trans.* **33**(2), 159–164 (1994)
8. Albraik, A., et al.: Diagnosis of centrifugal pump faults using vibration methods. *J. Phys. Conf. Ser.* (2012). IOP Publishing
9. Wen, B., et al.: Theory and experiment of nonlinear dynamics for rotating machinery with faults (2004)
10. Baumgartner, M., Kameier, F., Hourmouziadis, J.: Non-engine order blade vibration in a high pressure compressor. In: Twelfth International Symposium on Airbreathing Engines (1995)
11. Lei, Y., et al.: A method based on multi-sensor data fusion for fault detection of planetary gearboxes. **12**(2), 2005–2017 (2012)
12. Pumps, S.: Sulzer Centrifugal Pump Handbook. Elsevier (1998)
13. Bross, S., Addie, G.: Prediction of impeller nose wear behaviour in centrifugal slurry pumps. *Exp. Therm. Fluid Sci.* **26**(6–7), 841–849 (2002)
14. Walker, C.I., Bodkin, G.C.: Empirical wear relationships for centrifugal slurry pumps: Part 1: side-liners. *Wear* **242**(1–2), 140–146 (2000)
15. Stan, M., et al.: Centrifugal pump monitoring and determination of pump characteristic curves using experimental and analytical solutions. **6**(2), 18 (2018)
16. Kim, Y., Kim, K.: Near-field sound localization based on the small profile monaural structure. *Sensors* **15**(11), 28742–28763 (2015)
17. Gu, F., et al.: The condition monitoring of diesel engines using acoustic measurements Part 1: acoustic characteristics of the engine and representation of the acoustic signals. SAE Technical Paper (2000)
18. Ahmida, A., et al.: Gear wear process monitoring using acoustic signals (2014)
19. Benesty, J., Chen, J., Huang, Y.: *Microphone Array Signal Processing*, vol. 1. Springer Science & Business Media (2008)
20. Si, Q., et al.: Flow-induced noises in a centrifugal pump: a review. **11**(7), 909–924 (2019)
21. Birajdar, R., Patil, R., Khanzode, K.: Vibration and noise in centrifugal pumps—sources and diagnosis methods. In: 3rd International Conference on Integrity, Reliability and Failure (2009)

22. Wachel, J. Szenasi, F.J.P.H.: *Vibration and Noise in Pumps*, 1st ed. McGraw-Hill, pp. 9–87 (1976)
23. Čudina, M.: Pumps and pumping system noise and vibration prediction and control. In: *Handbook of Noise and Vibration Control*, pp. 897–909 (2007)
24. Gülich, J.F.: *Centrifugal Pumps*, vol. 2. Springer (2008)
25. Chen-xiao, N., Xu-she, Z.J.C.H., *PNEUMATICS: Study on vibration and noise for the hydraulic system of hydraulic hoist*. 2013(2), 64
26. Timouchev, S., Tourret, J.: Numerical simulation of BPF pressure pulsation field in centrifugal pumps. In: *Proceedings of the 19th International Pump Users Symposium*. Texas A&M University. Turbomachinery Laboratories (2002)
27. Li, H., et al.: Weak defect identification for centrifugal compressor blade crack based on pressure sensors and genetic algorithm. **18**(4), 1264 (2018)
28. Li, H., et al.: Pressure pulsation signal analysis for centrifugal compressor blade crack determination. **2014** (2014)
29. Fraser, W.H.: Centrifugal pump hydraulic performance and diagnostics. In: *Pump Handbook*. McGraw-Hill, New York, New York. Negative Flows,” Eleventh International Conference of the British Pump Manufacturers’ Association New Challenges-Where Next (1985)
30. Evans, J.: *Cavitation a Largely Misunderstood Phenomenon*, August 2005
31. Schiavello, B., Visser, F.C.: Pump cavitation: various NPSHR criteria, NPSHA margins, impeller life expectancy. In: *Proceedings of the 25th International Pump Users Symposium*. Texas A&M University. Turbomachinery Laboratories (2009)
32. Grist, E., *Cavitation and the Centrifugal Pump: A Guide for Pump Users*. CRC Press (1998)
33. Thai, Q., Lee, C.: The cavitation behavior with short length blades in centrifugal pump. *J. Mech. Sci. Technol.* **24**(10), 2007–2016 (2010)
34. Bachus, L., Custodio, B.: *Know and Understand Centrifugal Pumps*. Elsevier Inc (2003)

Fault Diagnosis for Gas Turbine Rotor Using MOMEDA-VNCMD



Yingjie Cui, Hongjun Wang, and Xinghe Wang

Abstract It is important rotating machinery for gas turbines in aviation, ship-building, and other industries. Given the high failure rate of the gas turbine rotor system, fault diagnosis of the rotor system is completely vital. Aiming at the fault diagnosis of the gas turbine rotor, we adopt a method based on Multipoint Optimal Minimum Entropy Deconvolution Adjusted (MOMEDA)—Variational Nonlinear Chirp Mode Decomposition (VNCMD) in this paper. For the gas turbine rotor test rig data, the original data is first analyzed for effective value, the fault signal is extracted, the fault signal is filtered by MOMEDA, the processed filtered signal is subjected to VNCMD decomposition, and the signal is reconstructed according to the magnitude of spectral kurtosis, and passed Envelope analysis to extract fault characteristics. This paper analyzes the data of the gas turbine rotor test bench, and the results show that the proposed method has achieved excellent results in the fault diagnosis of the gas turbine rotor.

Keywords Gas turbine rotor · Variational nonlinear chirp mode decomposition · Multipoint optimal minimum entropy deconvolution adjusted · Fault diagnosis

Y. Cui · H. Wang (✉) · X. Wang
School of Mechanical and Electrical Engineering, Beijing Information Science and Technology University, Beijing 100192, China
e-mail: wanghongjun@bistu.edu.cn

Y. Cui
e-mail: Cyingjie123@163.com

Beijing International Science Cooperation Base of High-end Equipment Intelligent Perception and Control, Beijing 100192, China

H. Wang
Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, Beijing 100192, China

1 Introduction

Gas turbine is very important in industrial production. It has a complicated structure and a special working environment [1], and it runs continuously for a long time. During the operation of gas turbines, rapid start and stop and rapid change of working conditions often occur. When the working conditions change rapidly, that is, the rotor speed changes rapidly, the torque that the rotor bears will increase greatly. Under the impact of alternating stress for a long time, Gas turbine rotors are prone to failure. Therefore, it is very important to analyze rotor vibration signals and perform state monitoring and fault diagnosis on the rotor system.

With the development of signal processing technology, people have higher and higher requirements for fault diagnosis technology. In recent years, a variety of fault diagnosis methods have emerged at home and abroad, including time series analysis, empirical decomposition, wavelet analysis, neural network, etc. For non-stationary and noisy signals, one of the more commonly used methods is the Empirical Mode Decomposition (EMD) [2]. Empirical mode decomposition is an adaptive algorithm, suitable for analyzing non-stationary and nonlinear phenomena because it can adaptively decompose the signal into intrinsic mode functions and then use Hilbert transform to select appropriate internal mode functions to construct the Envelope spectrum. However, empirical mode decomposition has the problem of modal aliasing. In order to overcome this problem, Ensemble Empirical Mode Decomposition (EEMD) [3] is proposed, and it is widely used in fault diagnosis. Although EMD has been improved in many versions, there are always modal aliasing and end effects. Recently, Peng Zhike's team proposed a Variational Nonlinear Chirp Mode Decomposition (VNCMD) [4] method, which is based on Variational Modal Decomposition (VMD) [5], which can characterize the signal as multiple FM and AM component phases. Overlay, use bandwidth variational constraints to simultaneously filter out target components, thereby avoiding backward errors caused by traditional recursive filtering, and is widely used in fault diagnosis of rotating machinery. Guo et al. [6] used bearing fault diagnosis under variable speed conditions to accurately locate the type of bearing defect; Niu et al. [7] used it in the lathe cutting process to effectively identify the existence and degree of cutting chatter; Chen et al. [8] Used in early fault detection and multi-feature extraction for rotor rubbing.

In the acquisition of the gas turbine rotor signal, since the sensor is installed outside the casing, the measured signal is the signal attenuated by the system, and it is difficult to extract the feature. Sun et al. [9] used Minimum Entropy Deconvolution (MED) and envelope cepstrum methods to extract early fault feature frequencies. Recently, Mc Donald et al. [10] proposed a new MED algorithm- Multipoint Optimal Minimum Entropy Deconvolution Adjusted (MOMEDA) in Optimized Minimum Entropy Deconvolution (OMED). This method takes the maximum D norm as the goal, and constructs the target vector according to the pre-estimated fault frequency, and does not use the iterative method to find the coefficients of the optimal filter for understanding the convolution. Zhang et al. [11] used the combination of the

EMD method and the MOMEDA method to detect parallel shaft gearbox failure, and verified that this method is superior to the traditional method.

This article combines the MOMEDA method with the VNCMD method. According to the gas turbine rotor vibration signal, the fault signal is processed by MOMEDA filtering, and then VNCMD is used to decompose and reconstruct the fault signal, and the fault signal is analyzed through the envelope spectrum.

2 Methods

2.1 A Brief Introduction of MOMEDA

Multipoint Optimal Minimum Entropy Deconvolution Adjusted (MOMEDA) is a deconvolution algorithm proposed by Geoff L to extract periodic pulses. It mainly solves the optimal filter through a non-iterative method to extract continuous periodic pulses. The main process of the algorithm is:

Vibration signal $y(x)$ collected by the sensor:

$$y(s) = h(s)x(s) + n(s) \tag{1}$$

In Eq. (1), $n(s)$ is noise, $x(s)$ is the impact sequence, $h(s)$ is the transfer function, and the purpose of MOMEDA is to find an FIR filter to restore the output result $y(s)$ as much as possible. Impact signal.

In order to obtain continuous shock pulses, MOMEDA is a deconvolution algorithm for multi-pulse target recognition at a certain position. The concept of MDN is introduced as Eq. (2), and the maximum value is MOMEDA, as Eq. (3):

$$MDN(\vec{y}, \vec{t}) = \frac{1}{\|\vec{t}\|} \cdot \frac{\vec{t}^T \vec{y}}{\|\vec{y}\|} \tag{2}$$

$$MOMEDA : \max MDN(\vec{y}, \vec{t}) = \max_{\vec{f}} \frac{\vec{t}^T \vec{y}}{\|\vec{y}\|} \tag{3}$$

Among them, \vec{t} is to determine the target vector of the shock pulse position d , and normalize the optimal target obtained by the solution. The target solution will change as the sampling frequency changes, and pulses of different periods can be extracted at the same sampling frequency. At the same time, the target vector \vec{t} can distinguish the impact component and noise component of the original signal.

The extreme value of Eq. (3) can be obtained by taking the derivative of the filter coefficients ($\vec{f} = f_1, f_2, \dots, f_L$) and simplifying it:

$$t_1 \vec{M}_1 + t_2 \vec{M}_2 + \dots + t_{N-L} \vec{M}_{N-L} = X_0 \vec{t} \tag{4}$$

Set Eq. (4) equal to zero, it will become:

$$\|\vec{y}\|^{-1} X_0 \vec{t} - \|\vec{y}\|^{-3} \vec{t}^T \vec{y} X_0 \vec{y} = \vec{0}, \text{ that is : } \frac{\vec{t}^T \vec{y}}{\|\vec{y}\|^2} X_0 \vec{y} = X_0 \vec{t} \tag{5}$$

Since $\vec{y} = X_0 \vec{f}$ and assuming that $(X_0 X_0^T)^{-1}$ exists:

$$\frac{\vec{t}^T \vec{y}}{\|\vec{y}\|^2} \vec{f} = (X_0 X_0^T)^{-1} X_0 \vec{t} \tag{6}$$

The MOMEDA filter and output solution can be simply summarized as Eqs. (7), (8) and (9):

$$\vec{f} = (X_0 X_0^T)^{-1} X_0 \vec{t} \tag{7}$$

$$X_0 = \begin{bmatrix} X_L & X_{L+1} & X_{L+2} & \cdots & X_N \\ X_{L-1} & X_L & X_{L+1} & \cdots & X_{N-1} \\ X_{L-2} & X_{L-1} & X_L & \cdots & X_{N-2} \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ X_1 & X_2 & X_3 & \cdots & X_{N-L+1} \end{bmatrix}_{L \text{ by } N-L+1} \tag{8}$$

$$\vec{y} = X_0^T \vec{f} \tag{9}$$

Therefore, the multiple of Eq. (7) is the solution of the MOMEDA filter coefficients, and then the convolution is performed according to the obtained coefficients, which avoids the influence of iteration on the solution of its coefficients. Since the filter can effectively extract periodic connected pulses, it avoids the influence of not being an integer when selecting periodic parameters.

2.2 A Brief Introduction of VNCMD

The VNCMD method characterizes the signal as the superposition of multiple frequency and amplitude modulation components. For a multi-component broadband modulation signal, the original broadband modulation signal is converted into a narrowband signal through demodulation technology, which can be characterized as:

$$x(t) = \sum_{i=1}^L \{a_i(t) \cos\left(2\pi \int_0^t f_i(s) ds\right)\}$$

$$+ b_i(t) \sin\left(2\pi \int_0^t \tilde{f}_i(s) ds\right) \} + n(t) \quad (10)$$

It is composed of L modes, and $n(t)$ is the background noise component. Where $\tilde{f}_i(t)$ and $f_i(t)$ represent the estimated instantaneous frequency and actual instantaneous frequency of the i -th mode, respectively, $a_i(t)$ and $b_i(t)$ are the two demodulation components:

$$a_i(t) = A_i(t) \cos\left(2\pi \int_0^t (f_i(s) - \tilde{f}_i(s)) ds + \phi_i\right) \quad (11)$$

$$b_i(t) = -A_i(t) \cos\left(2\pi \int_0^t (f_i(s) - \tilde{f}_i(s)) ds + \phi_i\right) \quad (12)$$

According to Eqs. (11) and (12), if the estimated instantaneous frequency is equal to the actual instantaneous frequency, $\tilde{f}_i(t) = f_i(t)$, the bandwidth of the demodulated component is the smallest, so a constraint model is constructed to evaluate the bandwidth of the demodulated signal:

$$\begin{aligned} & \min_{\{a_i\}\{b_i\}\{\tilde{f}_i\}} \left\{ \sum_{i=1}^L \|a_i''\|_2^2 + \|b_i''\|_2^2 \right\} \\ \text{s.t.} \quad & \left\| x(t) - \sum_{i=1}^L (a_i(t) \cos\left(2\pi \int_0^t \tilde{f}_i(s) ds\right) + b_i(t) \sin\left(2\pi \int_0^t \tilde{f}_i(s) ds\right)) \right\|_2 \leq \varepsilon \quad (13) \end{aligned}$$

In order to solve the constrained optimization problem, Lagrangian multipliers and quadratic penalty terms are introduced to optimize the reconstruction problem. The discrete form is:

$$\begin{aligned} L(\{a_i\}\{b_i\}\{f_i\}, \omega, \lambda) = & \sum_{i=1}^L (\|\Omega a_i\|_2^2 + \|\Omega b_i\|_2^2) \\ & + \lambda^T \left(\omega + \sum_{i=1}^L (C_i a_i + D_i b_i) - x \right) \\ & + \frac{\alpha}{2} \left\| \omega + \sum_{i=1}^L (C_i a_i + D_i b_i) - x \right\|_2^2 \quad (14) \end{aligned}$$

Among them, $\varphi_i(t) = 2\pi \int_0^t \tilde{f}_i(s) ds$, Ω is the second-order difference operator, λ and α are the Lagrange multiplier operator and the second-order penalty term parameters, respectively, ω is the error term.

Use the alternating direction multiplier optimization algorithm to optimize the instantaneous frequency and solve the modal components, and update the

demodulation components:

$$a_i^{L+1} = \left(\frac{2}{\alpha} \Omega^T \Omega + C_i^T C_i \right)^{-1} C_i^T \left(x - \sum_{m \neq i} C_m a_m - \sum_m D_m b_m - \omega - \frac{1}{\alpha} \lambda \right) \quad (15)$$

$$b_i^{L+1} = \left(\frac{2}{\alpha} \Omega^T \Omega + D_i^T D_i \right)^{-1} D_i^T \left(x - \sum_{m \neq i} C_m a_m - \sum_{m \neq i} D_m b_m - \omega - \frac{1}{\alpha} \lambda \right) \quad (16)$$

Firstly, the $a_i(t)$ and $b_i(t)$ can be calculated by the two demodulation components of Eqs. (15) and (16) to calculate the instantaneous frequency increment of the modal, and update the instantaneous frequency $\tilde{f}_i(t)$:

$$\begin{aligned} \Delta \tilde{f}_i^{L+1}(t) &= -\frac{1}{2\pi} \frac{d}{dt} \left(\arctan \left(\frac{b_i^{L+1}(t)}{a_i^{L+1}(t)} \right) \right) \\ &= \frac{b_i^{L+1}(t) \cdot (a_i^{L+1}(t))' - a_i^{L+1}(t) \cdot (b_i^{L+1}(t))'}{2\pi \left((a_i^{L+1}(t))^2 + (b_i^{L+1}(t))^2 \right)} \end{aligned} \quad (17)$$

$$f_i^{L+1} = f_i^L + \gamma \cdot \Delta f_i^{L+1} \quad (18)$$

Among them, γ is the penalty parameter, and $0 < \gamma < 1$ is the proportional parameter used to stabilize the algorithm.

Finally update the Lagrange multiplier:

$$\lambda^{L+1} = \lambda^L + \alpha \left(\omega^{L+1} + \sum_{i=1}^L x_i^{L+1} - x \right) \quad (19)$$

The above iterative decomposition process is repeated until the iterative stop condition is satisfied, and L broadband eigenmode components are obtained.

2.3 Fault Diagnosis Based on MOMEDA-VNCMD Method

In this paper, a certain type of gas turbine rotor system is studied, and the fault diagnosis of the gas turbine rotor system is realized by the method of combining VNCMD and MOMEDA. First obtain the experimental data through the test bench, perform effective value analysis on the obtained experimental data, extract the fault time period through the effective value analysis, and extract the fault data; first perform the MOMEDA filter processing on the extracted fault data; The signal undergoes short-time Fourier transform, extracts the high-energy frequencies inside,

and performs VNCMD decomposition according to these high-energy frequencies to obtain multiple modal components; then calculates the spectral kurtosis value of each modal component, and extracts the largest three. The modal component corresponding to the spectral kurtosis is reconstructed, the reconstructed signal is enveloped, and the fault feature is extracted from the envelope spectrum for fault diagnosis. The flowchart of this method is shown in Fig. 1.

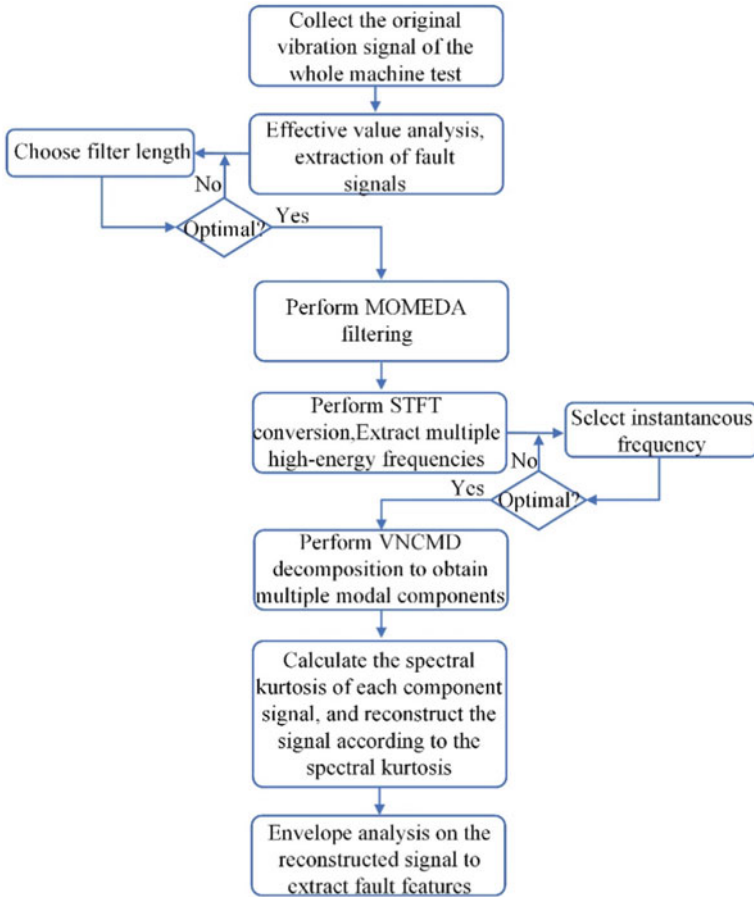


Fig. 1 MOMEDA-VNCMD flow chart

3 Data Analysis and Verification

3.1 Data Acquisition of Gas Turbine Rotor Failure

In order to verify the effectiveness and practicability of the method proposed in this article, data from a gas turbine rotor test bench of a certain company is used for analysis. Before leaving the factory, the gas turbine will be tested on the test bench to test the vibration characteristics of the gas turbine under various working conditions to detect whether there are manufacturing problems and whether the performance quality meets the requirements. Due to the installation of sensors at different positions and different angles, the measured values vary greatly. After years of experimentation, the company has determined the installation position of the sensor. As shown in Fig. 2, a speed sensor is installed at the radial position of the front end of the low-pressure compressor casing. As the front measuring point, a speed sensor is installed at the radial position of the casing between the high-pressure compressor and the combustion chamber as a rear measuring point, and a sensor for measuring the speed of the high-pressure rotor is installed inside the gas engine.

For a certain type of dual-rotor gas turbine of the company, the vibration data during the test on the test bench was collected. The sampling frequency is 6000 Hz, and the experiment time is 2–3 h. When the gas turbine is tested on the test bench, due to the relatively long experiment time and the large amount of data, the vibration limit set by the company is an effective value of 8 mm/s. Based on an analysis of experimental data, the high-pressure rotor speed changes continuously under multiple working conditions, the sampling frequency is 6000 Hz, and the experimental time lasts 8000 s. The effective value analysis of the obtained data is shown in Fig. 3. The dotted line is the change of the high-voltage rotor frequency, the dashed line is the

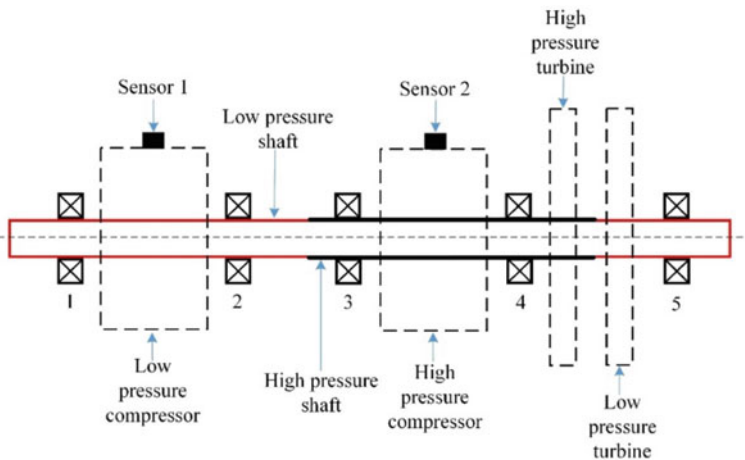


Fig. 2 Schematic diagram of gas turbine vibration test

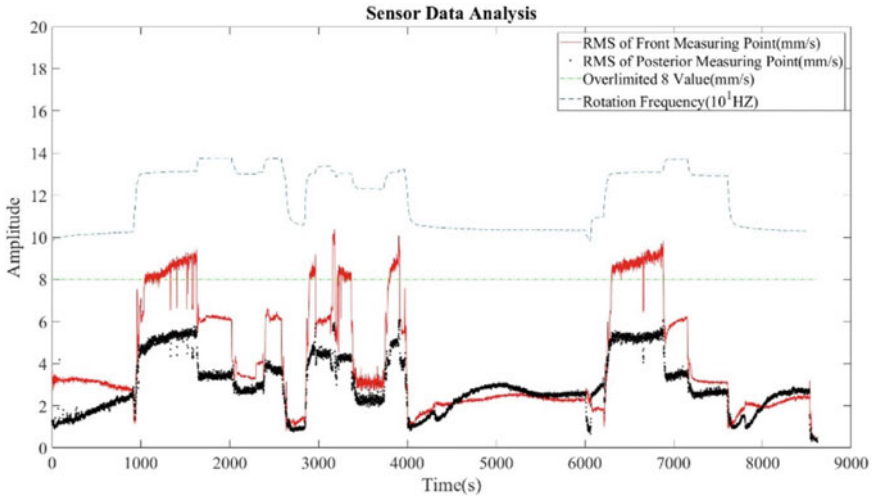


Fig. 3 Gas turbine sensor effective value diagram

vibration limit value, the solid line is the effective value of the vibration data at the previous measurement point, and the dotted line is the post measurement point. The effective value of the vibration data. It can be found from the gas turbine sensor data effective value analysis graph that the rotational speed changes in various working conditions during the experiment. The effective value of the vibration of the front measuring point and the effective value of the vibration of the rear measuring point are changing with the working condition, and the effective value of the vibration of the front measuring point is Vibration exceeding the limit occurred in several working conditions, and there was no excessive vibration at the rear point.

3.2 Analysis of VNCMD-MOMEDA Method

Since the vibration exceeds the limit at the front point, we use the filtered data of the front point for the next analysis. First, we extract the data that exceeds the limit in the first segment of the front-side point filtering, and we resample the original fault data. Figure 4 is the time domain signal diagram after we extract the resampled fault data. The accuracy of VNCMD mainly depends on the choice of the initial instantaneous frequency caused by noise and other aspects. Therefore, we first use the MOMEDA filtering method to preprocess the original signal to reduce the impact of noise. The most important thing in the use of the MOMEDA algorithm is the selection of the filter length. Here we have obtained through experiments the filter length selection 710 is a better value. Figure 5 is a time-domain diagram after we MOMEDA the fault signal.

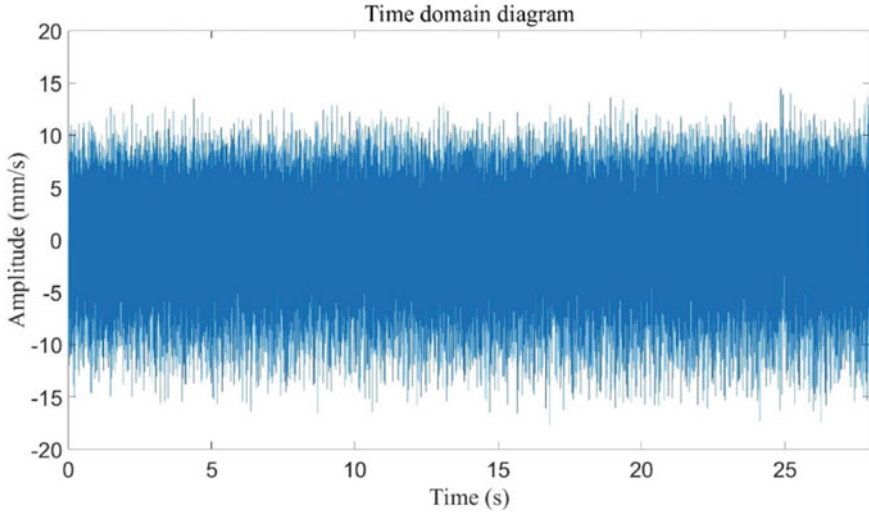


Fig. 4 Time domain diagram of fault signal

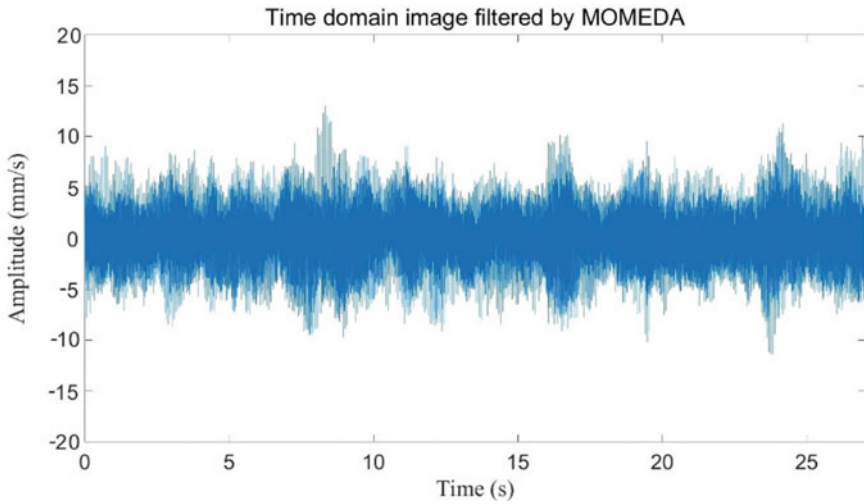


Fig. 5 Time domain image after MOMEDA filtering

We perform time–frequency analysis on the filtered data and do a short-time Fourier transform to obtain a time–frequency graph. The transformed time–frequency graph is shown in Fig. 6. The energy of the signal increases from low frequency to high frequency first and then attenuates gradually. Several mid-frequency high-energy curves can be found in the time–frequency image. The intermediate frequency fluctuates with the speed. The higher the frequency, the greater the fluctuation and

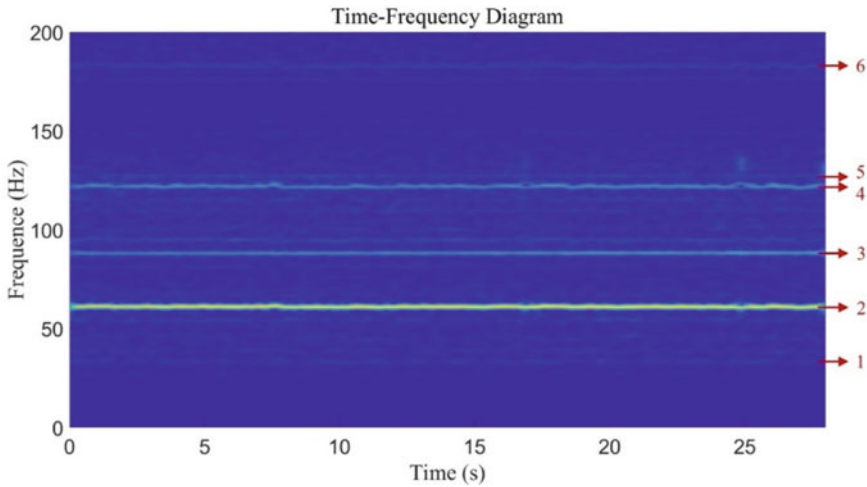


Fig. 6 Time–frequency diagram of fault signal

the smaller the energy. From the time–frequency diagram, we can see that the energy of the instantaneous frequency is concentrated in 50–150 Hz (here, 50 and 150 are just a fuzzy concept). The next step is VNCMD decomposition. VNCMD decomposition needs to set the initial instantaneous frequency. The setting of the initial instantaneous frequency will affect our subsequent analysis, so we need to extract multiple high-energy fault frequencies. According to the actual situation and through many experiments and analyses, the effect of extracting 6 fault frequencies is better than other effects, so here we extract 6 fault frequencies. The choice of these 6 fault frequencies is mainly selected from the energy concentrated in the instantaneous frequency. We choose four instantaneous frequencies from 50 to 150 Hz, one for each below 50 Hz and one above 150 Hz. Here, the estimated instantaneous frequency of VNCMD is set to these 6 high-energy frequencies, and then the signal is subjected to variational nonlinear modal decomposition to obtain 6 components, and the time domain and frequency domain diagrams of the 6 components are drawn separately. The drawing result is shown in Fig. 7.

Next, we calculate the kurtosis value of each signal decomposed, and the kurtosis value is shown in Table 1. According to the value of kurtosis, through multiple experiments, we select the first, third, and fifth components for reconstruction.

The time–frequency diagram of the reconstructed signal is shown in Fig. 8, and the envelope spectrum is shown in Fig. 9. From Fig. 9, we can see the frequency multiplier component of the signal. We combine the effective value map of the original data and do the same analysis on the following fault signals. The abnormal vibration is mainly caused by the low-pressure rotor. This gas turbine has a rotor. Temporary bending, which is the same as the predicted result. Since the experimental data is the data on the test bench, it is less affected by the noise effect. If it is in practical application, the effect on noise reduction will be better.

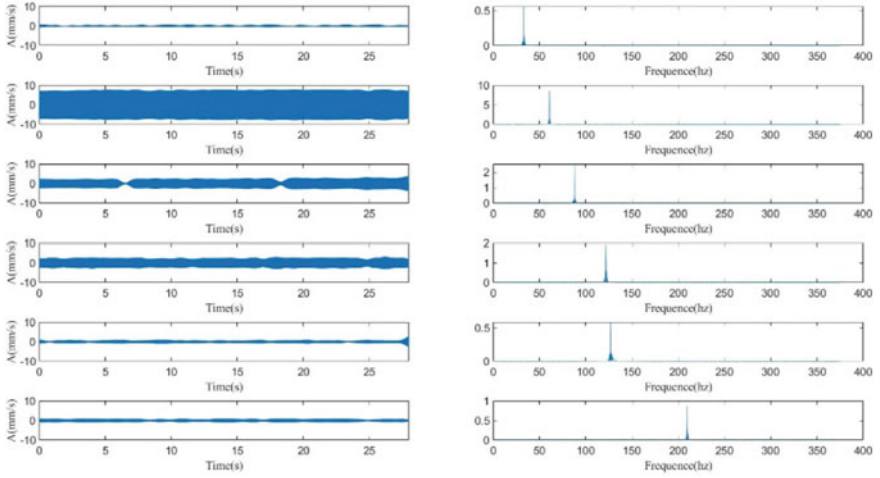


Fig. 7 VNCMD component decomposition

Table 1 The kurtosis value of each component

Portion	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
Kurtosis	1.9525	1.5034	1.6398	1.5409	2.7322	1.6046

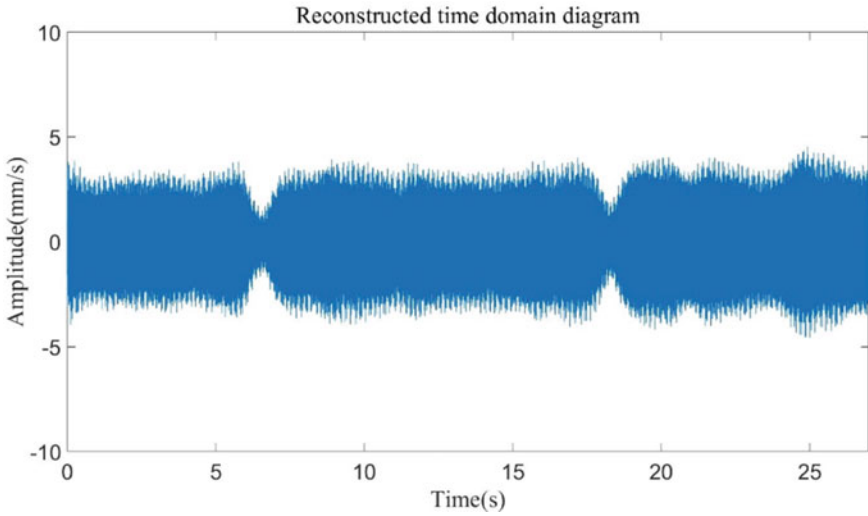


Fig. 8 Reconstruct the time domain diagram

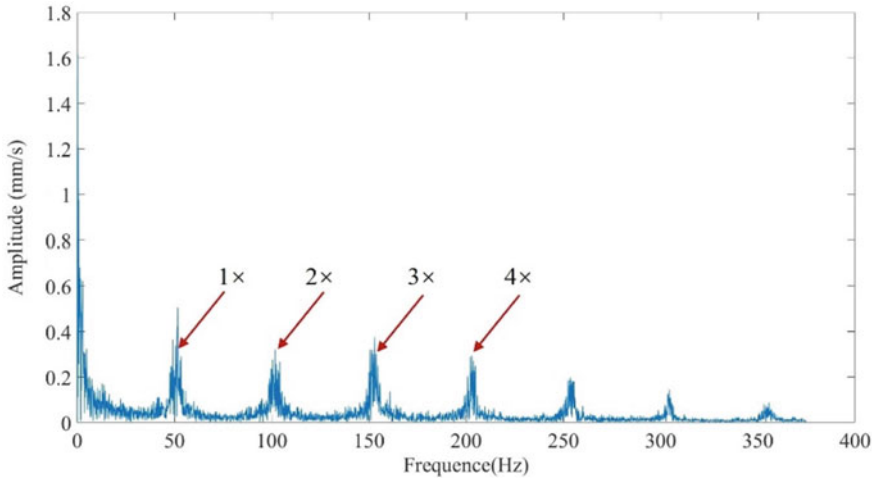


Fig. 9 Reconstructed signal envelope spectrum

4 Conclusion

A fault diagnosis method for gas turbine rotor is based on MOMEDA-VNCMD. Firstly perform effective value analysis on the original signal to extract the fault signal; then resample the fault signal and perform MOMEDA filtering; perform short-time Fourier transform from the filtered signal to extract the instantaneous with large energy Frequency, then VNCMD decomposition and reconstruction, and envelope analysis. VNCMD has good applicability for processing data with high speed, effectively avoiding the limitation of time–frequency resolution in traditional time–frequency analysis, and overcoming the narrow-band constraint of adaptive decomposition. When a gas turbine fails, this method can be used for analysis and verification.

Acknowledgements This research is supported by the National Natural Science Foundation of China (Grant No. 51975058).

References

1. Fabio, C.G., Gatta, N., Venturini, M.: Resistant statistical methodologies for anomaly detection in gas turbine dynamic time series: development and field validation. *J. Eng. Gas Turbines Power* **140**(5), 052401–052411 (2018)
2. Li, Y.B., Xu, M.Q., Liang, X.H., Huang, W.H.: Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings. *IEEE Trans. Industr. Electron.* **64**(8), 6506–6517 (2017)

3. Wang, X., Liu, C., Bi, F., Bi, X., Shao, K.: Fault diagnosis of diesel engine based on adaptive wavelet packets and EEMD-fractal dimension. *Mech. Syst. Sig. Process.* **41**, 581–597 (2013)
4. Chen, S., Dong, X., Peng, Z.: Nonlinear chirp mode decomposition: a variational method. *IEEE Trans. Sig. Process.* **65**(22), 6024–6037 (2017)
5. Li, Z., Chenn, J., Zi, Y., Pan, J.: Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive. *Mech. Syst. Sig. Process.* **85**, 512–529 (2017)
6. Guo, W.J., Jiang, X.X., Li, N., Shi, J.J., Zhu, Z.K.: A coarse TF ridge-guided multi-band feature extraction method for bearing fault diagnosis under varying speed conditions. *IEEE Access* **7**, 18293–18310 (2019)
7. Niu, J.C., Ning, G.C., Shen, Y.J., Yang, S.P.: Detection and identification of cutting chatter based on improved variational nonlinear chirp mode decomposition. *Int. J. Adv. Manuf. Technol.* **104**, 2567–2578 (2019)
8. Chen, S.Q., Yang, Y., Peng, Z.K.: Detection of rub-impact fault for rotor-stator systems: a novel method based on adaptive chirp mode decomposition. *J. Sound Vib.* **440**, 83–99 (2019)
9. Sun, W., Li, X., Jin, X., Huang, J., Zhang, X.: Feature extraction method based on MED and envelope cepstrum. *Zhendong Ceshi Yu Zhenduan/J. Vibr. Meas. Diagn.* **38**, 1057–1062 (2018)
10. McDonalda, G.L., Zhao, Q.: Multipoint optimal minimum entropy deconvolution and convolution fix: application to vibration fault detection. *Mech. Syst. Sig. Process.* **82**, 461–477 (2017)
11. Zhang, X., Zhao, J.M., Ni, X.L., Sun, F.C., Ge, H.Y.: Fault diagnosis for gearbox based on EMD-MOMEDA. *Int. J. Syst. Assur. Eng. Manag.* **10**, 836–847 (2019)

Classification and Recognition Method of Bearing Fault Based on SDP-CNN



Wang Xing-he, Wang Hong-jun, Cui Ying-jie, and Liu Ze-ru

Abstract In view of the problems that the signal features are difficult to extract when a fault occurs in a rolling bearing, and the time domain image of the original vibration signal cannot obviously show the feature differences of different faults, and the direct deep feature learning and recognition will have a large impact on the system performance, etc., a bearing fault classification and recognition method based on symmetry dot pattern-convolutional neural network (SDP-CNN) is proposed. First, the SDP method is used to analyze the vibration signals of different faults, and the signal SDP images obtained can clearly show the feature differences of different faults; then, the SDP images are input into the CNN network for feature learning and state recognition; finally, Validation was performed using the Case Western Reserve University (CWRU) bearing dataset. The results show that the recognition accuracy of this method is 97.5%, which further verifies that the deep learning algorithm can adaptively extract the features of the SDP image and effectively identify bearing faults.

Keywords Symmetry dot pattern (SDP) · Convolutional neural network (CNN) · Feature learning · State recognition

W. Xing-he · W. Hong-jun (✉) · C. Ying-jie · L. Ze-ru
Mechanical Electrical Engineering School, Beijing Information Science and Technology University, Beijing 100192, China
e-mail: wanghj86@163.com

W. Xing-he
e-mail: wXH1362055780@163.com

W. Hong-jun
Beijing International Science Cooperation Base of High-end Equipment Intelligent Perception and Control, Beijing 100192, China

MOE Key Laboratory of Modern Measurement and Control Technology, Beijing 100192, China

1 Introduction

Technology has been advancing in recent years, the industrial level has continued to improve, the integration of CNC machine tools has become better and better, and the manufacturing process has become more automated. However, in order to meet the requirements of industrial production, machine tools need to operate continuously. Long-term operation of the equipment will definitely cause damage or even malfunction [1]. Even a small failure will affect the working performance of the machine tool, and then affect the processing quality of the product, causing economic losses [2]. Rolling bearings are one of the most important parts in CNC machine tools, and its fault identification is of far-reaching significance for maintaining the working performance of the machine tools and improving the quality of product processing [3]. Fault identification of the rolling bearings of CNC machine tools can ensure the processing quality of the product and the work efficiency of the CNC machine tools, and avoid the development of small faults into major faults that affect the work of CNC machine tools due to failure to discover in time.

In [4, 5], an SDP method was proposed and applied to fault visual diagnosis. It transforms a one-dimensional time series into a snowflake pattern composed of mirrored symmetric points in polar coordinates, and reflects the changes in the amplitude and frequency of the sound signal or vibration signal through the difference of the patterns, which can more intuitively reflect the status of each fault.

The deep learning neural network simulates the brain nervous system with a rich hierarchical structure, and extracts the features of the network input step by step to form an abstract high-level feature representation [6]. It is called deep learning because the deep learning network can learn the essential characteristics of massive data through the layer-by-layer feature extraction method. At the same time, this method organically combines feature extraction and feature recognition, and adaptively extracts and learns features [7]. CNN is a true multilayer structural learning algorithm in deep learning, which consists of multiple convolutional layers alternating with downsampling layers, and then connects one or more fully connected layers and the output layer to obtain the classification result of (image) features in the output layer [8]. The core idea is to reduce the free parameters of the network, reduce the complexity of the network, and improve the execution efficiency of the network through the sharing of local receptive fields and neuron weights.

Since the original 1D vibration signals and time domain images cannot obviously show the feature differences between each fault, the direct deep feature learning and recognition of vibration signals will have a large impact on the system performance. In order to show the signal features, it is convenient for the CNN model to perform feature learning on the signal. Therefore, this paper proposes a bearing fault classification and recognition method based on SDP-CNN. First, the SDP method is used to analyze the vibration signals of different faults, and the signal SDP images obtained can clearly show the feature differences of different faults; then, the SDP images are input into the CNN network for feature learning and state recognition; finally, the bearing data of Case Western Reserve University (CWRU) was verified.

2 Symmetry Dot Pattern (SDP)

SDP analysis method is a new technology that combines signal processing and image analysis [9]. Compared with other methods, the SDP method is simple to calculate and can intuitively show the difference of different vibration states. At the same time, this method can convert the time-domain vibration signal into polar coordinates and turn it into an SDP graph, and then intuitively reflect the difference in the graph. The characteristics of the fault status information [10]. For time domain vibration signals: $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$, transform it into a point in polar coordinate space by SDP method $S[r(i), \theta(i), \phi(i)]$ (see Fig. 1).

Among them, $r(i)$ represents the radius of polar coordinates, $\theta(i)$ represents the angle of counterclockwise deflection along the mirror plane in polar coordinates, and $\phi(i)$ represents the angle of clockwise deflection along the mirror plane in polar coordinates. The calculation formulas for the three are as follows.

$$r(i) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

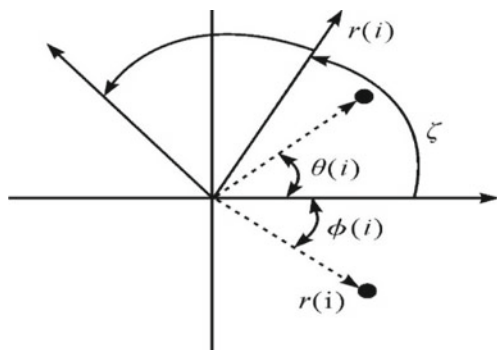
$$\theta(i) = \theta_s + \frac{x_{i+l} - x_{\min}}{x_{\max} - x_{\min}} \xi \quad (2)$$

$$\phi(i) = \theta_s - \frac{x_{i+l} - x_{\min}}{x_{\max} - x_{\min}} \xi \quad (3)$$

In the formula, x_{\max} is the maximum amplitude of signal X , x_{\min} is the minimum amplitude of signal X , l is the time interval parameter, θ is the rotation angle of the mirror symmetry plane, and ξ is the amplification factor ($\xi \leq \theta$) [11, 12]. $\theta_s = 360 s/n$, where $s = 1, 2, \dots, n$, where n is the number of mirror-symmetrical planes.

The SDP analysis method converts the time-domain waveform of the signal into an image in polar coordinates, and displays the characteristics of the vibration signal through the image. Compared with other image analysis methods, SDP can show the difference of different vibration forms more clearly, and at the same time has a better processing effect on noise signals.

Fig. 1 Schematic diagram of SDP algorithm analysis



3 Convolutional Neural Network (CNN)

The CNN model is a deep learning method with image recognition as the core of the application. The information transfer direction within the model flows unidirectionally from input to output. Therefore, it is also called a feedforward neural network and has been widely used in many fields [13]. The CNN model structure is composed of input layer, convolution layer, pooling layer, full link layer and output layer. The model is shown in Fig. 2. Among them, the model structure can be expanded by expanding the convolutional layer, the pooling layer, and the fully connected layer. The basic working principle of the CNN model is as follows: First, the input image is subjected to the convolution kernel sliding operation to extract the features, the convolutional layer and the pooling layer alternately extract the features, and then the features are integrated and input into the fully connected layer for feature fusion, and finally Discriminate the classification output result.

- (1) Input layer: It is to input the data signal into the network composed of two-dimensional plane, so that the convolutional layer can perform feature learning afterwards.
- (2) Convolutional layer: It is an important functional level of the CNN model, which mainly performs convolution operations on input data to achieve feature extraction. The image data feature extraction process requires the participation of convolution kernels, training weights, and activation functions. Among them, the convolution kernel can be understood as the feature extractor of the input image data. The convolution operation means the non-linear feature representation of the input data. The convolution kernel slides according to the set step size to realize the continuous convolution operation of the input data to obtain brand new Feature map. The formula for convolution operation is as follows.

$$y_i = f(c_i * x_i) \tag{4}$$

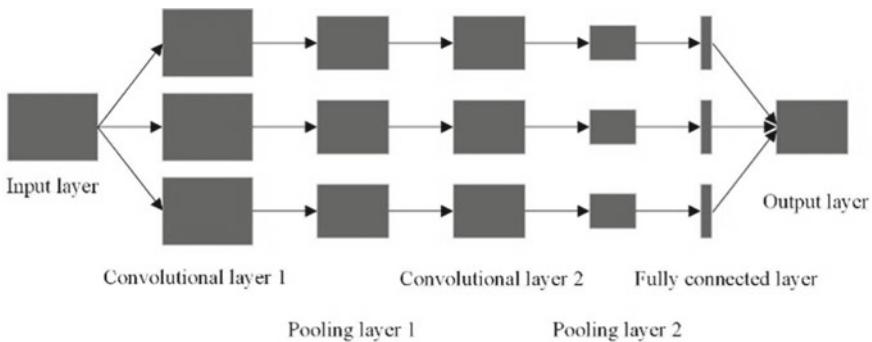


Fig. 2 Convolutional neural network model

where: y_i —the eigenvalues obtained by convolution operation; c_i —convolution kernel involved in the operation; x_i —input data involved in the operation; $f(*)$ —non-linear activation function.

- (3) Pooling layer: Also known as the down-sampling layer, the pooling layer reduces the size of the feature map obtained by the convolution operation to complete the secondary feature extraction of the feature map. The pooling operation needs to select a square small area in the feature map, and combine the feature values in the area to complete the pooling operation to obtain a brand new feature map.
- (4) Fully connected layer: Splice all the feature maps of two-dimensional planes into one-dimensional features as the input of the fully connected network, and obtain the output of this layer by weighted summation of the inputs and the response of the activation function.
- (5) Output layer: Classify and recognize the feature vectors finally extracted from the CNN model. Generally, the Softmax classifier is used to implement the classification method.

4 Fault Classification Model

The fault classification model based on SDP-CNN is shown in Fig. 3. First, the collected vibration signals are divided into multiple signals of a certain length; then the vibration signals are converted into SDP images by the SDP method, and they are randomly divided into training samples and test samples. From the SDP map of each type of signal, 80% is randomly selected as the training sample, and the rest are used as the test sample, which is input into the deep learning convolutional neural network for training. The deep learning convolutional neural network model is a two-layer convolution pooling: the number of convolutions in the first layer is 6, and the size of the convolution kernel is 5×5 ; the number of convolutions in the second layer is 12, and the size of the convolution kernel is 5×5 ; The size of the sampling pool core is 2×2 in each layer; the network output is a 4×1 feature vector.

5 Case Analysis

In order to verify the effect of the convolutional neural model on bearing fault diagnosis, this paper selects the rolling bearing public data of Case Western Reserve University (CWRU) for experimental verification. The CWRU bearing data acquisition system is shown in Fig. 4. The system uses acceleration sensors to collect the vibration signals of the motor drive end and the fan end. Since the signal collected by the drive end is more comprehensive and stable, this experiment selects the drive end signal with a sampling frequency of 12 kHz as the experimental data.

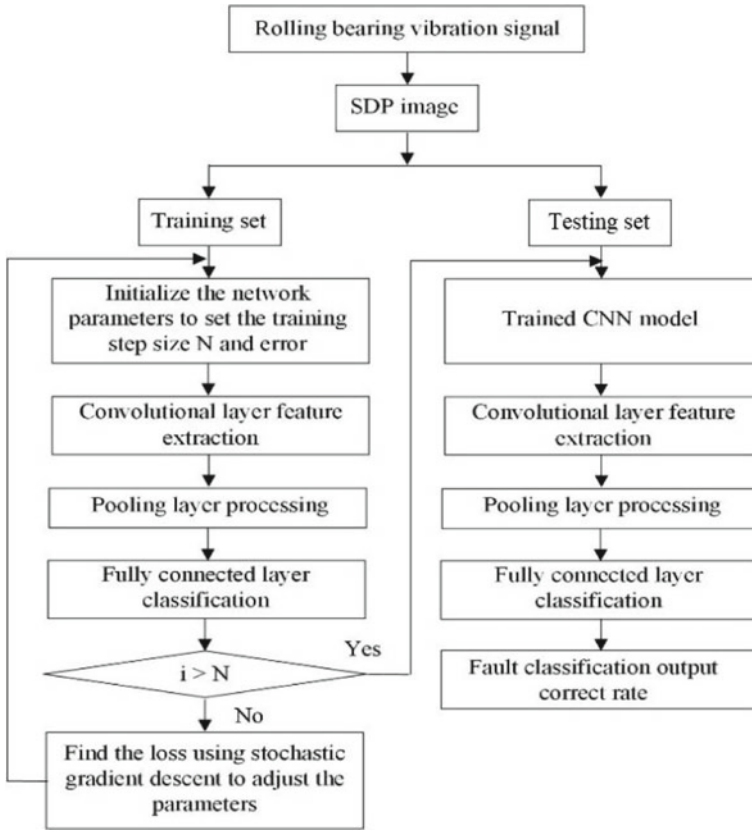


Fig. 3 Fault classification model

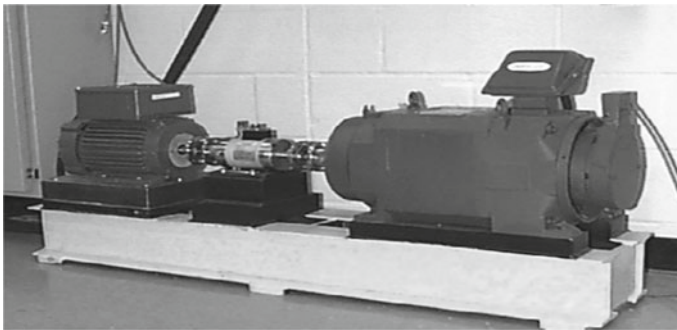


Fig. 4 CWRU bearing data acquisition system

In the CWRU data set, the fault types of bearings are divided into ball faults, inner ring faults and outer ring faults. The fault diameters are divided into four sizes: 0.007, 0.014, 0.021, and 0.028 inch. Therefore, in the experiment, for the three types of faults: ball fault, inner ring fault and outer ring fault, the fault data directly corresponding to the 0.007 inch fault is selected as the experimental data. Including normal bearing data, the data set constructed in the experiment contains a total of four categories. In addition, the CWRU data set contains four types of data collected at speeds of 1797, 1772, 1750, and 1730 rpm. The experimental data selected in this experiment are all bearing signal data collected at a speed of 1730 rpm. There are 200 fault types for each type. For the sample, the experiment selected a bearing signal with a length of 32×32 as the input of the neural network, and the training set and the test set were divided according to the ratio of 8:2.

The time domain diagram of the different states of the bearing is shown in Fig. 5.

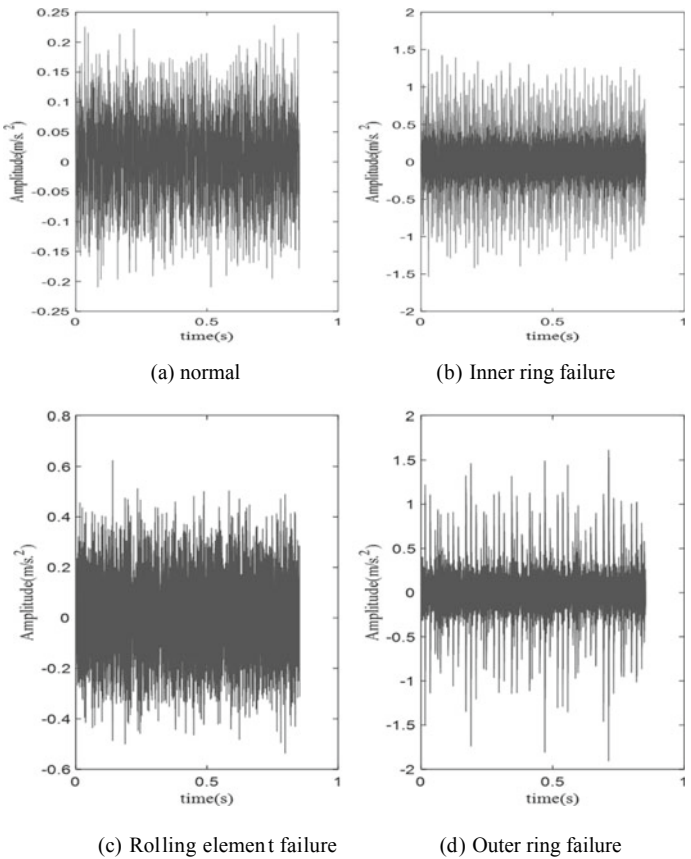


Fig. 5 Time domain diagram of different states

Since the SDP method has the same shape of the symmetry plane at each position, only one symmetry plane in one position needs to be extracted to display different signal characteristics. It can be seen from Fig. 6 that SDP analysis can fully display the bearing vibration signal characteristics under different fault conditions. The main difference between the SDP images of different fault conditions is the concentration and shape characteristics of the points around the mirror plane, which can be more obvious through the SDP diagram. The earth shows the difference of different vibration states, and then distinguishes them.

After the SDP image is generated, it is used as the input of the convolutional neural network. After the training is completed, the test set is input for analysis and diagnosis. The diagnosis results are shown in Table 1.

Using the traditional BP neural network and comparing with the method in this paper, introducing the fault diagnosis rate evaluation index, the formula is as follows, and the calculation result is shown in Table 2.

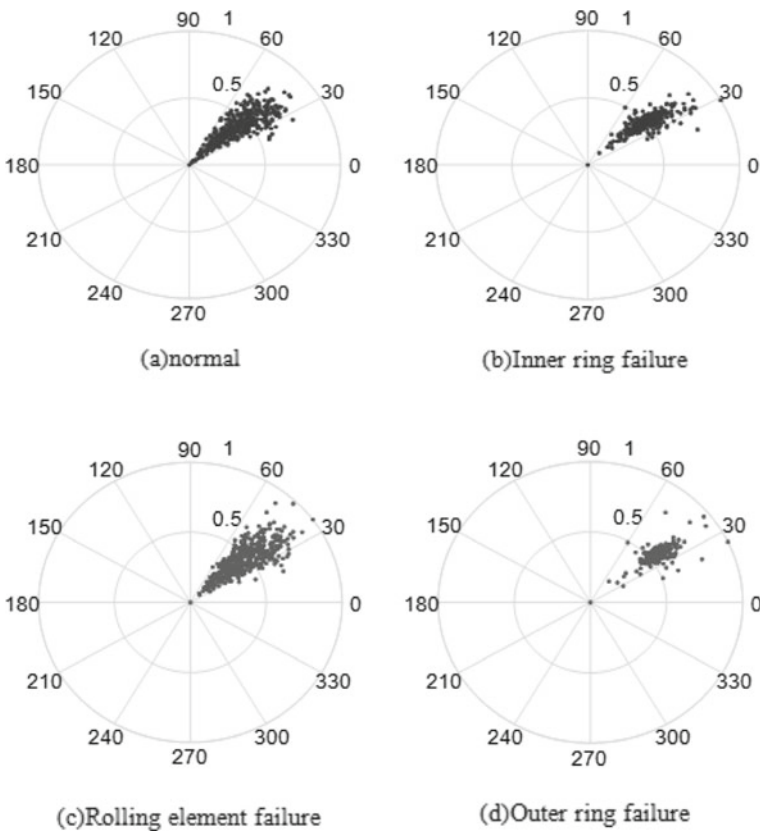


Fig. 6 SDP images in different states

Table 1 Diagnosis results

Actual/diagnostic	Normal	Inner circle	Rolling element	Outer ring
Normal	40	0	0	0
Inner circle	2	38	0	0
Rolling element	1	0	39	0
Outer ring	1	0	0	39

Table 2 Different methods of diagnosis

Fault diagnosis method	DR (%)
BP	88.5
CNN	97.5

$$DR = \frac{TN}{TN + FP} \tag{5}$$

In the formula, *TN* means that the fault state is diagnosed as a fault state, and *FP* means that the fault state is diagnosed as a normal state.

6 Conclusion

This research proposes a fault diagnosis method based on symmetry point pattern and convolutional neural network. Using SDP diagrams can more clearly show the difference between different vibration states, and then distinguish them; CNN can directly learn from low-level to high-level intelligent learning to get a good feature representation, avoiding the extraction and selection of artificial features, and enhances the intelligence of the recognition process. Finally, CNN is used for SDP image recognition, which finally realizes the recognition of bearing faults.

Acknowledgements This research’s support under the National Natural Science Foundation of China (Grant No. 51975058) and the Beijing Science and Technology Project (Grant No. Z201100008320004).

References

1. Zhang, S.: Research on fault diagnosis of rolling bearings of CNC machine tools based on web. Zhengzhou University (2020)
2. Wan, S., Peng, B., Wang, X.: Early fault diagnosis method of rolling bearing based on dual time domain transformation and sparse coding shrinkage. China Mech. Eng. **31**(23), 2829–2836 (2020)

3. Xiong, Z.: Bearing fault diagnosis method based on EEMD and adaptive redundant lifting scheme packet. *Vibroeng. Procedia* **34** (2020)
4. Bianchi, S., Corsini, A., Sheard, A.G., Tortora, C.: A critical review of stall control techniques in industrial fans. *Int. Sch. Res. Notices* **2013** (2013)
5. Shibata, K., Takahashi, A., Shirai, T.: Fault diagnosis of rotating machinery through visualisation of sound signals. *Mech. Syst. Sig. Process.* **14**(2) (2000)
6. Ren, B., Wang, L., Deng, X., Li, X.: An improved English character recognition of deep learning network structure. *J. Chengdu Univ. Inform. Technol.* **32**(03), 259–263 (2017)
7. Zhou, F., Yang, S., Fujita, H., Chen, H., Wen, C.: Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowl.-Based Syst.* **187** (2020)
8. Jing, A.N., Ping, A.I., Shen, X.U., et al.: An intelligent fault diagnosis method for rotating machinery based on one-dimensional convolutional neural network. *J. Nanjing University(Natural Science)* 55(01),133–142(2019).
9. Zhang, W., Zhou, J.: Research on fault diagnosis of rolling bearing based on dropout-CNN. *Light Ind. Mach.* **37**(02), 62–67 (2019)
10. Xu, X., Wang, S., Liu, H.: Real-time detection of centrifugal fan stall based on SDP and image matching. *J. Chin. Soc. Power Eng.* **35**(11), 906–911 (2015)
11. Bianchi, S., Corsini, A., Sheard, A.G.: Demonstration of a stall detection system for induced draft fans. *Proc. Inst. Mech. Eng. Part A J. Power Energy.* **227**(3) (2013)
12. Sheard, A.G., Corsini, A., Bianchi, S.: Stall warning in a low-speed axial fan by visualization of sound signals. *J. Eng. Gas Turbines Power* **133**(4), 041601–041610 (2011)
13. Yu, T., Xinyu, P., Zihan, W.: Fault diagnosis method of rolling bearing based on GADF-CNN. *J. Vibr. Shock* **40**(05), 247–253+260 (2021)

Spindle Health Assessment Based on Rotor Perception



Zhuangzhuang Zhang, Hongjun Wang, Jishou Xing, Fengshou Gu, and Xinghe Wang

Abstract Due to the complex structure of the spindle and many influencing factors, the failure analysis of the spindle has always been a focus. Use Solidworks to model the Dalian machine tool VDL600 machine tool spindle, analyze the modalities of the spindle and main components through the ANSYS Workbench platform, and analyze the possible resonance frequencies. Taking the data measured by Lion's gyration accuracy tester as a reference, the MEMS acceleration sensor is used to obtain the vibration signal of the spindle, and the modulation signal bispectrum (MSB) analysis method is used to obtain the characteristic frequency of the vibration signal. By comparing and analyzing the characteristic frequency of the vibration signal and the modal analysis result, it can be proved that the MSB method can extract the resonance frequency of the spindle well.

Keywords MSB · Rotation accuracy · Modal · ANSYS workbench

1 Preface

The machine tool plays an important role in the modernization of the national economy, and the spindle plays a very important role in the accuracy of the machine tool. Therefore, a large number of scholars at home and abroad analyze the failure of

Z. Zhang · H. Wang (✉) · J. Xing · F. Gu · X. Wang
Mechanical Electrical Engineering School, Beijing Information Science and Technology University, Beijing 100192, China
e-mail: wanghj86@163.com

Z. Zhang
e-mail: zzzhs1996@163.com

H. Wang · J. Xing · F. Gu
Beijing International Science Cooperation Base of High-end Equipment Intelligent Perception and Control, Beijing 100192, China

H. Wang · F. Gu
Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, Beijing 100192, China

the spindle from different aspects. Deng Sanpeng et al. [1] used acoustic sensors to collect the noise signal of the machine tool spindle, and obtained fault frequency information through wavelet packet and envelope spectrum analysis to monitor spindle faults. Sun Wei et al. [2] analyzed the dynamic characteristics of the spindle at high speed, low speed and static state by establishing a finite element model. Li Zhenyu et al. [3] proposed a method based on BP neural network and error calibration fitting to predict the axis trajectory error of the machine tool spindle to reflect and predict the operating state of the spindle. Li Mengmei et al. [4] proposed an algorithm for spindle mechanical identification and fault diagnosis to detect the shaft misalignment at the front end of the spindle. Memanshi et al. [5] used the holographic spectrum theory to analyze the vibration characteristics of the spindle, and found out the reasons for the deterioration of the rotation accuracy of the spindle as the speed increased. Hu Teng et al. [6] studied the influence of centrifugal force and gyro torque on the ‘spindle-bearing’ system at high speeds. This article attempts to explain the cause of the failure of the spindle through the dynamic analysis of the spindle and the vibration signal.

2 Spindle Dynamics Modeling

2.1 Build a Three-Dimensional Model

In the article, the main axis of Dalian machine tool VDL600 is taken as the analysis object. The main axis of the machine tool is shown in Fig. 1a, and the machine tool is a vertical machine tool. The structure of the main axis is shown in Fig. 1b. Use Soilworks to model the spindle as shown in Fig. 1c. The spindle is mainly composed of a housing, a spindle, a bearing, a broach mechanism and a BT40 tool holder. The

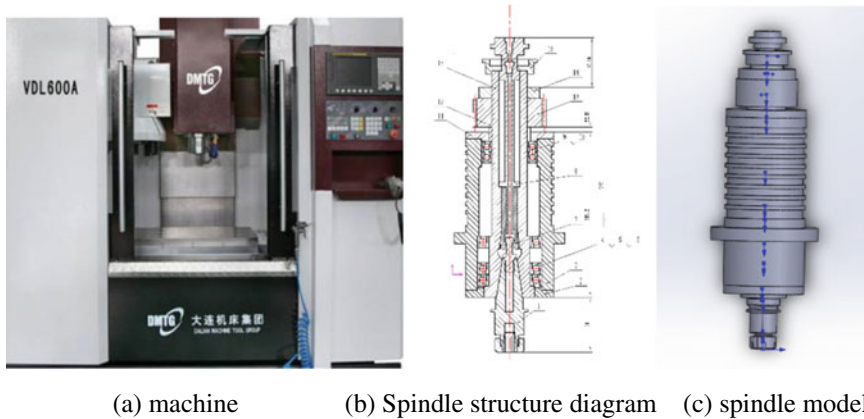


Fig. 1 Three-dimensional model of the spindle

bearings are all angular contact ball bearings, and the front bearing model of the spindle is 7014C TYN DBD LP4 (NSK), the spindle rear bearing model is 7012C TYN DBD LP4 (NSK).

2.2 ANSYS Workbench Modal Analysis

Modal analysis theory

Since the bearing is a continuous solid structure in the form of a component, it can be approximately equivalent to a discrete system with N degrees of freedom according to the dynamic modeling theory. The motion equation of the spindle system is:

$$M\ddot{u} + C\dot{u} + Ku = F \quad (1)$$

In the formula, M is the system mass matrix; the system damping matrix $C = C^s - \Omega G^b - \Omega G^d$, C^s is the structural damping; the system stiffness matrix $K = K^b + K_p^b - \Omega^2 M_c^b$ is the axial load The additional stiffness matrix caused by; G^d is the turntable rotation matrix.

When the system is freely vibrating, $F = 0$, and ignoring the influence of damping on the modal characteristics, the differential equation of undamped free vibration of the bearing system is established, namely

$$M\ddot{u} + ku = 0 \quad (2)$$

where: M is the mass matrix; \ddot{u} is the acceleration vector; K is the stiffness matrix; u is the displacement vector.

Spindle modal analysis

When the spindle system rotates at a high speed, the centrifugal force changes the deformation of the contact area between the inner raceway and the outer raceway of the angular contact ball bearing, so that the radial support stiffness of the bearing gradually decreases with the increase of the angular velocity, and the softening of the bearing stiffness occurs. The spindle will also produce shafting centrifugal force and gyro torque under high-speed operation. The dynamic characteristics of the spindle system under high-speed operation will be significantly different from those under static or low-speed conditions. The centrifugal force effect and the bearing softening effect will have a greater impact on the natural frequency of the system. According to the experimental rules provided by Cao Hongrui and others, analyze the natural frequency of the spindle under the first-order mode at static speed, 600, 1000, 5000, 10,000, and 15,000 r/min. As shown in Fig. 2, it can be seen from the figure that the first-order mode of the spindle only decreases slightly when the speed exceeds 1000 r/min, that is, the softening phenomenon of the bearing will affect the mode of the spindle at a higher speed. There is almost no drop in the spindle mode and static

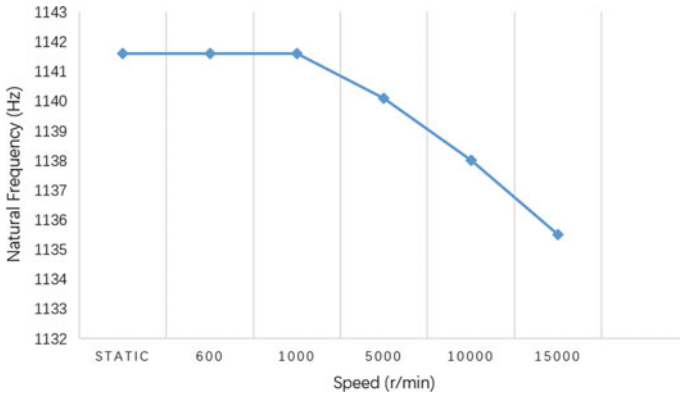


Fig. 2 Natural frequency change in the first-order mode

state at low speed. Since the spindle speed in this analysis is lower than 600 r/min, it is a low speed, so this study will ignore the influence of bearing softening on the spindle mode.

Due to the complex structure of the spindle, if the finite element analysis is directly performed on the spindle model, the existence of the bearing will greatly increase the amount of calculation of the model, so here we will simplify the bearing as follows [8]:

- (1) Replace the bearing with a spring unit, the axial position is the midpoint of the contact line of each pair of bearings, and it is considered that the simplified spring only has radial stiffness without considering its angular stiffness;
- (2) Regardless of the stiffness change caused by the bearing load, only the bearing stiffness is regarded as a fixed constant value. Each pair of bearings is simplified to 4 springs, and the angle between adjacent springs is 90°;

According to the bearing model, the axial preload of the bearing can be found. The calculation formula for the radial stiffness of a single angular contact ball bearing after preload is:

$$K_r = 17.7236 \sqrt[3]{Z^2 D_b} \frac{\cos^2 \alpha}{\sqrt[3]{\sin \alpha}} \sqrt[3]{F_{\alpha 0}}$$

In the formula, D_b —rolling element diameter; Z —bearing rolling element number; α —contact angle; $F_{\alpha 0}$ —axial preload. According to the above formula, the radial stiffness of the front and rear bearings can be calculated to be about 3.48×10^8 N/m.

According to the actual situation, a fixed constraint is imposed on the main shaft shell, and the mode of the main shaft can be obtained as shown in Fig. 3 and the values of the first six modes are shown in Table 1.

Fig. 3 Spindle mode

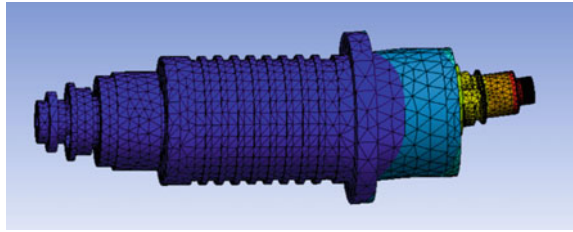


Table 1 The first six modes of the spindle

Mode	One	Two	Three	Four	Five	Six
Value (Hz)	1141.6	1150.3	1475.3	1495.7	1609.8	1630.8

Table 2 7–12 modes of each part of the spindle

Model		Seven	Eight	Nine	Ten	Eleven	Twelve
Value (Hz)	7012C bearing	1151	1831	1844	1938.8	2649.4	2879.7
	7014C bearing	1826.3	1834.1	2544.5	2545.2	4621.8	4639
	BT40 knife handle	4938.4	4954.5	7312.9	7360.3	9906.6	11,363
	Broach mechanism	504.18	504.19	1299.1	1299.2	2290.4	2308.7
	Spindle	1546.1	1546.1	3489.1	3783.8	3783.8	5765.6
	Spindle housing	3454.8	3454.9	3895.2	3895.3	4375.7	4376

In order to judge the relationship between the mode of each component and the mode of the spindle, the free mode value of each component is obtained separately. Since the first six-order mode value of the free mode is zero, the 7-of each component is listed. The 12-order modal values are shown in Table 2, and the seventh-order modal (fundamental frequency) vibration shape of each component is shown in Fig. 4.

From the data in Table 2, it can be seen that the seventh-order mode of the 7012C bearing is very similar to the ninth-order and tenth-order frequencies of the pneumatic device and is close to the first-order modal value of the spindle in Table 1, which is 1141.6 Hz, so at this frequency Resonance may affect the accuracy of the machine tool.

3 Vibration Signal Acquisition and Signal Processing

3.1 Signal Acquisition

Because the spindle tool holder is closely connected with the spindle and rotates with the spindle, the vibration state of the spindle can be well sensed at the spindle

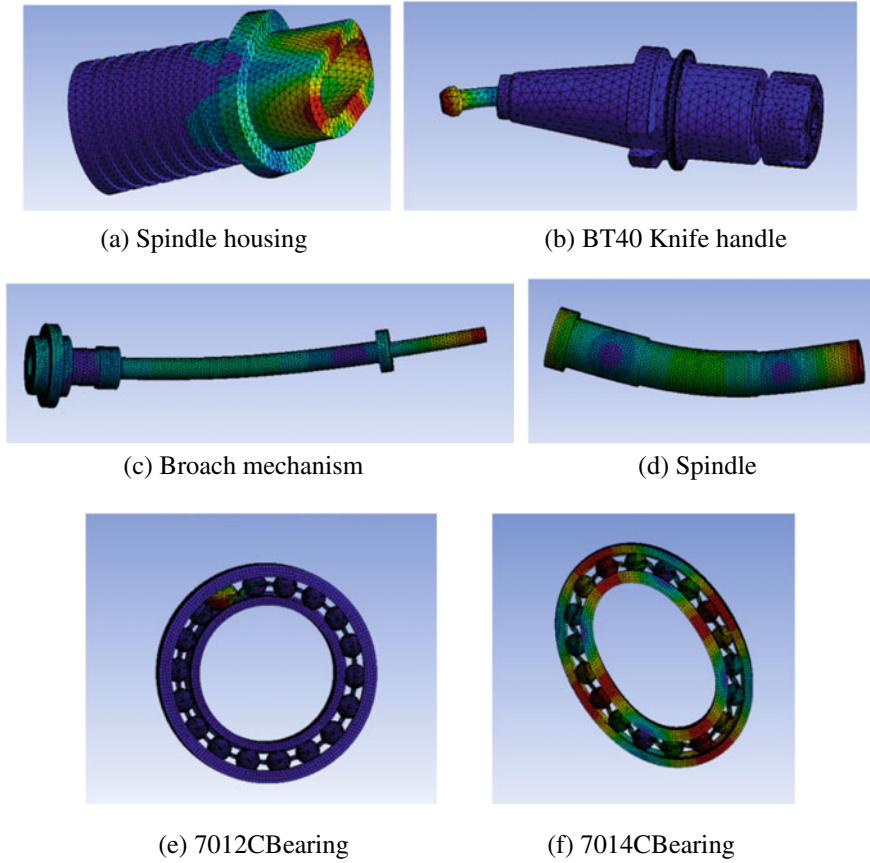


Fig. 4 The seventh-order formation of each part of the main axis

tool holder. Therefore, the MEMS acceleration sensor can be installed at the tool holder when collecting the vibration signal of the spindle. As shown in Fig. 6, the sensor installation method is shown in Fig. 5. The sensor is closely matched with the cylinder on the BT40 tool holder, and the upper end of the sensor is attached to the surface of the tool holder to ensure the parallelism between the sensor device and the spindle, and the sensor is evenly distributed around the circumference. There are 6 threaded holes, and 6 screws are used to fasten the sensor on the tool holder to ensure that the sensor and tool holder rotate with the spindle. MEMS sensors have the functions of self-powered and wireless data transmission. Axial, radial and torsional vibration signals with high signal-to-noise ratio can be collected to obtain the low-frequency operating characteristics and high-frequency characteristics of the spindle.

Due to the installation position of the sensor, the acceleration data measured by the sensor is not only caused by the vibration of the spindle but also the centrifugal

Fig. 5 Sensor installation method and test diagram

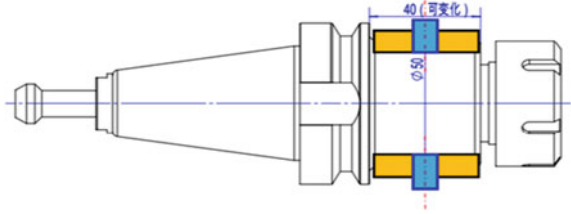
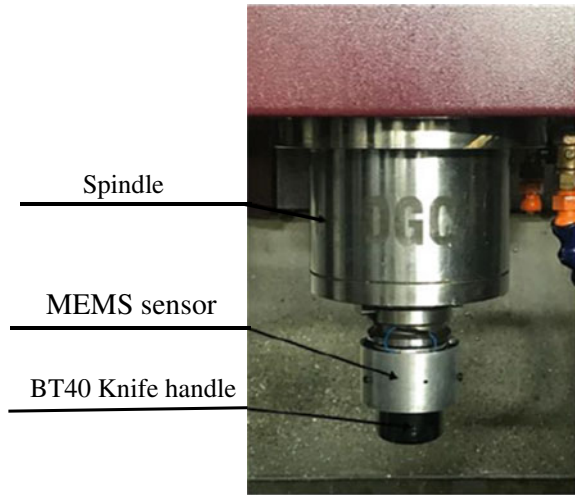


Fig. 6 Experimental layout



force generated when the spindle rotates. This will cause the sensor to exceed the range at high speed, so this experiment will collect the spindle. The vibration signal of the spindle idling at the speed of 100, 200, 300, 400, 500, and 600 r/min. The time domain diagram of the X, Y, and Z directions transmitted to the mobile terminal APP via Bluetooth is shown in the figure. Due to the sensor installation position, the Z axis of the sensor is the X axis direction of the machine tool, and we are concerned about the radial direction of the spindle Direction of the vibration signal, so we select the collected Z-axis signal for analysis (Fig. 7).

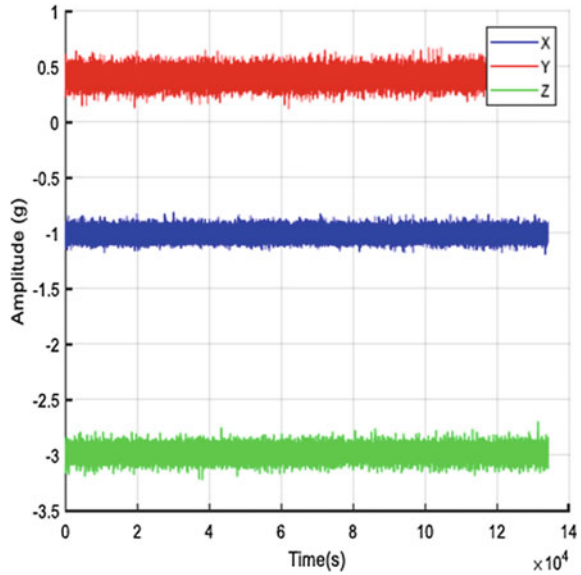
3.2 Using MSB Method for Signal Processing

Signal processing

After processing the signal collected in Sect. 3.1, import it into the MSB algorithm and analyze it to get the spectrogram as shown in Fig. 8.

In Fig. 8 we can see that when the spindle is working at speeds of 100, 200, 300, 500, and 600 r/min, the frequency of larger amplitudes is less than 700 Hz. But

Fig. 7 XYZ time domain diagram collected by MEMS sensor



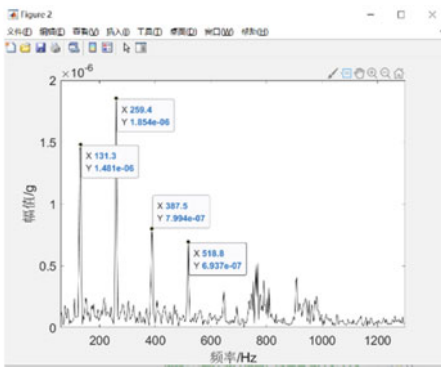
when the rotation speed is 400 r/min, there is a larger amplitude at a frequency of 1081 Hz, and the frequency here is very close to the first-order mode of the spindle we analyzed earlier, 1141.6 Hz, so it is likely that the spindle speed will be 400 here. The spindle resonates at r/min, which causes the accuracy of the spindle to decrease.

4 Spindle Rotation Accuracy Test

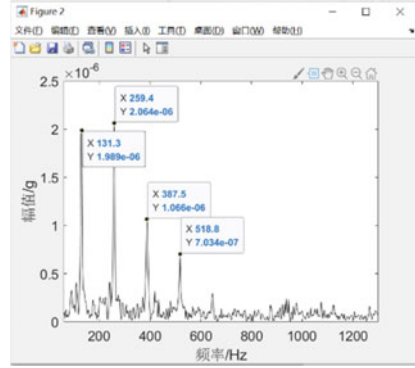
Rotation accuracy test is an important means to evaluate the accuracy of machine tools. Therefore, we will use the spindle rotation error analyzer of American Lion Company to test the rotation accuracy of the spindle. In order to compare with the analysis results above, we will also use 100, 200, 300, 400, 500 and 600 r/min are used in the experiment. In order to reduce the influence of other factors, MEMS sensors are used to collect data and test the rotation accuracy in Sect. 3.1. The data collected by the instrument is carried out at the same time, and the workbench layout is shown in Fig. 9.

The results of the rotation accuracy test are shown in Fig. 10, Due to limited space, we only show the results when the speed are 100 and 400 r/min. The error values at each speed are shown in Table 3.

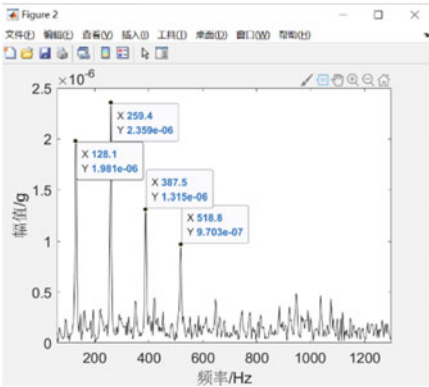
From the data in Table 3, it can be seen that the rotation error suddenly increases when the spindle speed is 400 r/min, and when the speed exceeds 400 r/min, the rotation error suddenly decreases. This also verifies our previous conclusions, when the spindle speed reaches 400 r/min. At min, the main shaft resonates and the rotation error of the main shaft increases suddenly.



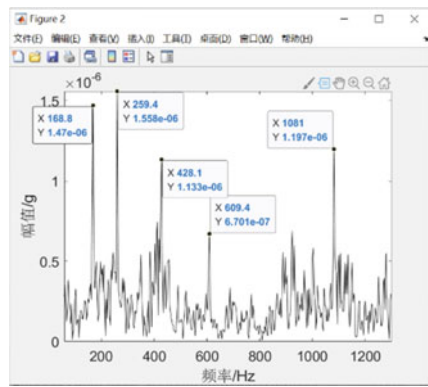
(a) 100 r/min



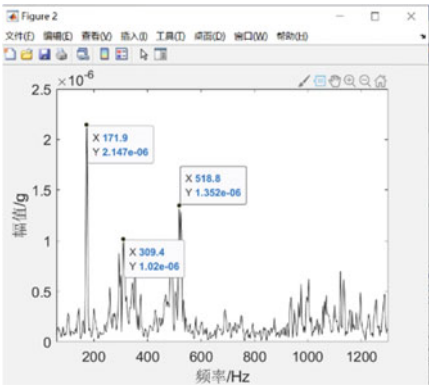
(b) 200 r/min



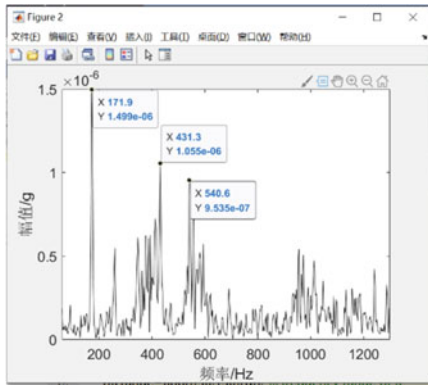
(c) 300 r/min



(d) 400 r/min



(e) 500 r/min



(f) 600 r/min

Fig. 8 Spectrum diagram of spindle 100–600 r/min

Fig. 9 Rotation accuracy sensor and MEMS sensor layout diagram

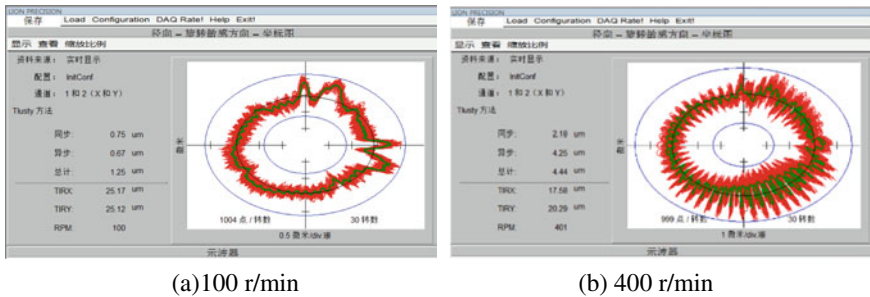
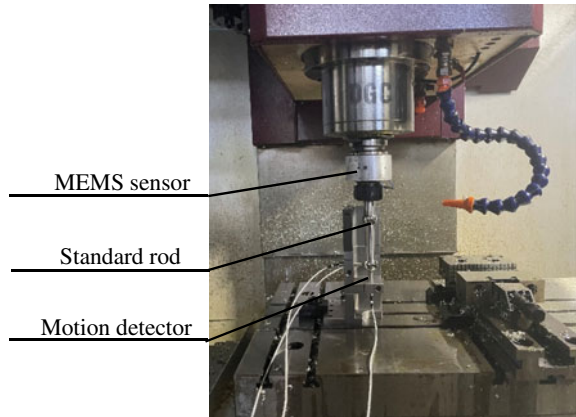


Fig. 10 Spindle 100–600 rad/min rotation accuracy test results

Table 3 Rotation accuracy error of each speed

Rotating speed (r/min)	100	200	300	400	500	600
Rotation error (μm)	1.25	1.36	0.98	4.44	1.34	1.51

5 Conclusion

This paper takes the main shaft of Dalian machine tool VDL600 as the research object, establishes a corresponding three-dimensional model, analyzes the mode of the main shaft and its parts, and judges the possible resonance frequency. A comparative analysis of the data collected by the MEMS sensor and the rotation accuracy test results of lion found that the spindle will excite the first-order mode of the spindle at a speed of 400 r/min and cause resonance, which leads to an increase in the rotation error of the spindle and decrease of spindle machining accuracy.

Acknowledgements This research is supported by Beijing Science and Technology Planning Project (Grant No. Z201100008320004).

References

1. Sanpeng, D., Xiaoli, X., Jianxin, Z., Xiangling, Z.: Research on fault diagnosis of CNC machine tool spindle based on noise wavelet envelope spectrum. *Mach. Tool Hydraulics* **37**(12), 219–221 (2009)
2. Wei, S., Bo, W., Bangchun, W.: Comparative analysis of dynamic characteristics of high-speed spindle system under static and operating conditions. *Chin. J. Mech. Eng.* **48**(11), 146–152 (2012)
3. Zhenyu, L., Haochen, W., Gongliang, W., Jiapeng, L.: Research on machine tool spindle fault diagnosis based on BP network. *Mech. Des. Manuf.* **4**(10), 130–133+139 (2019)
4. Mengmei, L., Xuezhen, C.: Spindle mechanical state recognition and fault diagnosis based on LabVIEW. *Manuf. Technol. Mach. Tool* **4**(12), 32–36 (2020)
5. Manshi, M., Hongjun, W., Yufei, Z.: Testing and deterioration analysis of spindle rotation accuracy of CNC machine tool. *J. Beijing Univ. Inform. Sci. Technol. (Natural Science Edition)* **30**(06), 43–46+53 (2015)
6. Teng, H., Guofu, Y., Mingnan, S.: Research on the dynamic characteristics of the “spindle-bearing” system based on the effects of centrifugal force and gyro torque. *Vibr. Shock* **33**(08), 100–108 (2014)
7. Hongrui, C., Bing, L., Zhengjia, H.: Dynamic modeling and high-speed effect analysis of high-speed spindle. *J. Vibr. Eng.* **25**(02), 103–109 (2012)
8. Yichuang, X., Huaicun, Z., Hongjun, W.: Modal analysis and dynamic characteristic experiment of high-speed electric spindle based on ANSYS workbench. *Mech. Eng.* **4**(01), 30–32 (2014)
9. Haifei, Q.: SKF6208 rolling bearing dynamics modeling and simulation research. *Mod. Manuf. Eng.* **4**(12), 112–116 (2019)

Fault Diagnosis of Rolling Bearing Based on Wavelet Packet Decomposition and SVM-LMNN Algorithm



Zhengbo Wang, Hongjun Wang, and Yingjie Cui

Abstract Aiming at the effective identification of failure modes of rolling bearings, a support vector machine (SVM) and Levenberg–Marquardt (LM algorithm) fault diagnosis method for rolling bearings is proposed. First, use wavelet packet decomposition to obtain sub-bands, reconstruct the decomposition coefficients, and expand the decomposed sub-band signals to the original signal length; then, use SVM to classify the fault state; finally, input the feature vector into LMNN (LM algorithm Neural network) to realize failure mode recognition. The method is verified by the rolling bearing fault diagnosis experiment. The results show that the SVM-LMNN based on wavelet packet decomposition has a rolling bearing fault diagnosis accuracy rate of up to 99.456%. The method proposed in the study is compared with the instantaneous energy method of the VMD component of the kurtosis criterion and the enveloping spectrum solution diagnosis method, and the higher accuracy is obviously obtained, which proves the feasibility and effectiveness of the proposed method.

Keywords Wavelet packet decomposition · SVM · LMNN algorithm · Fault diagnosis of rolling bearing diagnosis

Z. Wang · H. Wang (✉) · Y. Cui
School of Mechanical and Electrical Engineering, Beijing Information Science and Technology University, Beijing 100192, China
e-mail: wanghj86@163.com

H. Wang
Beijing International Science Cooperation Base of High-End Equipment Intelligent Perception and Control, Beijing 100192, China

MOE Key Laboratory of Modern Measurement and Control Technology, Beijing 100192, China

1 Introduction

As an important part of various types of equipment, rolling bearings must be reliable and stable. Therefore, the fault diagnosis of rolling bearings is highly valued by researchers [1]. The diagnosis of rolling bearing faults has become a research hotspot.

Wavelet can adjust the sampling of different frequencies in the time domain. Its multi-resolution analysis divides the signal according to a certain scale. The signals of different frequencies are divided into different frequency bands, and then each sub-frequency band is reconstructed. So as to separate the various harmonics [2]. Therefore, wavelet packet decomposition has better time–frequency window characteristics than other filters. In recent years, some scholars have done a lot of research on wavelet packets. Wang et al. [3] proposed an improved genetic algorithm based on wavelet packet decomposition to optimize the BP neural network, and established a more accurate rolling bearing IGA-BP state prediction model. The results show that the model converges faster and the prediction accuracy rate is higher. Xia et al. [4] based on the bearing fault diagnosis method based on the combination of wavelet packet and gradient boosting decision tree, and the results showed that the accuracy of the diagnosis using wavelet packet was improved by 11% compared with directly using the time series feature. Wen [5] integrated the early fault monitoring and diagnosis of rolling bearings. Research shows that the wavelet packet decomposition method is more superior than the comparison method in the early fault monitoring and early fault diagnosis of bearings.

For the fault diagnosis of rolling bearings, algorithms in the field of artificial intelligence represented by support vector machines (SVM) are mainly used. Support vector machine is a classification method based on statistics. It uses different methods to find multi-dimensional hyperplanes to achieve multi-dimensional nonlinear data classification and prediction tasks. In addition, support vector machines are easy to operate, self-learning, and relatively strong generalization capabilities. There are many applications in mechanical fault diagnosis, and a wealth of research results have also been accumulated. Fang et al. [6] proposed a particle swarm improved support vector machine algorithm for data classification and trend prediction. The classification experiment for rolling bearing failures achieved good results; Hu et al. [7] improved the support vector machine algorithm by genetic algorithm for rolling bearing analysis. Fault diagnosis is shorter and more accurate than SVM. But genetic algorithm is easy to fall into the problem of premature and low search efficiency when solving the scale calculation problem, particle swarm algorithm is easy to fall into the local optimum and prematurely converge so that it cannot find the global optimum solution [8–10].

2 Method

2.1 Wavelet Packet Decomposition

2.1.1 Deduction of Wavelet Decomposition Mechanism

In 1989 Mallat S proposed the concept of multi-resolution decomposition, unifying the previous methods of constructing wavelets. Multi-resolution decomposition uses wavelet function and scale function to decompose the low-frequency overview part of the signal, without considering the high-frequency details [9, 10]. Multi-resolution decomposition is in the function space $L^2(R)$, and the function f is represented as a series Approximate the limits of the function, each approximation is a smooth version of the function f , and is more and more refined approximation function, these approximations are obtained on different scales. Let N be the arbitrary scale to be decomposed, then the complete reconstruction formula of $f(t)$ at the decomposition level of N is

$$f(t) = \sum_{j=1}^N \sum_{k=-\infty}^{\infty} d_{j,k} \varphi_{j,k}(t) + \sum_{k=-\infty}^{\infty} c_{N,k} \varphi_{N,k}(t) \tag{1}$$

where: j is the scale; k is the translation factor; $d_{j,k} = \langle f(t), \varphi_{N,k}(t) \rangle$ is the wavelet expansion coefficient; $c_{N,k} = \langle f(t), \varphi_{N,k}(t) \rangle$ is the scale expansion coefficient; $\varphi_{j,k}(t)$ is the wavelet function; $\varphi_{N,k}(t)$ is the scaling function; the first term is the decomposition weight The detailed sequence obtained by the structure; the second item is the overview sequence. In order to improve the resolution [11, 12], the wavelet subspace frequency is further subdivided. The scale space V_j and the wavelet subspace W_j are unified to form a new space U_j^n as shown in Eq. (2).

$$\begin{cases} U_j^0 = V_j \\ U_j^1 = W_j \end{cases} \quad j \in Z \tag{2}$$

The orthogonal decomposition of the Hilbert space $V_{j+1} = V_j \oplus W_j$ is uniformly expressed by U_j^n as $U_{j+1}^0 = U_j^0 \oplus U_j^1, j \in Z$.

The definition subspace U_j^n is the closure space of the function $U_n(t)$, and U_j^{2n} is the closure space of the function $U_{2n}(t)$, which satisfies the following equations:

$$\begin{cases} u_{2n}(t) = \sqrt{2} \sum_{k \in Z} h(k) u_n(2t - k) \\ u_{2n+1}(t) = \sqrt{2} \sum_{k \in Z} g(k) u_n(2t - k) \end{cases} \tag{3}$$

In the formula: $g(k) = (-1)^k h(1 - k)$, when $n = 0$, the formula is

$$\begin{cases} u_0(t) = \sum_{k \in Z} h_k u_n(2t - k) \\ u_1(t) = \sum_{k \in Z} g_k u_n(2t - k) \end{cases} \tag{4}$$

Meet the two-scale equation in multi-resolution analysis:

$$\begin{cases} \phi(t) = \sum_{k \in Z} h_k \phi(2t - k), \{h_k\}_{k \in Z} \in l^2 \\ \psi(t) = \sum_{k \in Z} g_k \psi(2t - k), \{g_k\}_{k \in Z} \in l^2 \end{cases} \tag{5}$$

Comparing the above formulas, $u_0(t)$ and $u_1(t)$ are the degenerate scaling function $\phi(t)$ and wavelet basis function $\psi(t)$, respectively. The function set $\{u_n(t)\}$ is called an orthogonal wavelet packet determined by the basis function $u_0(t) = \phi(t)$.

2.1.2 Wavelet Decomposition Layer Selection

In wavelet decomposition, the larger the number of decomposition layers, the more obvious the different characteristics of noise and signal performance, which is more conducive to the separation of the two; but on the other hand, the greater the number of decomposition layers, the distortion of the reconstructed signal will also be The larger it is, it will affect the final denoising effect to a certain extent. By traversing the influence of the number of decomposition layers on the diagnosis result, the number of decomposition layers is weighed to 11 layers.

2.1.3 Wavelet Decomposition Threshold Selection

In the wavelet domain, the coefficient corresponding to the effective signal is large, and the coefficient corresponding to the noise is small. The coefficients corresponding to the noise in the wavelet domain still satisfy the Gaussian white noise distribution.

The threshold selection rule is based on the model

$$y = f(t) + e \tag{6}$$

where e is Gaussian white noise $N(0, 1)$. Therefore, the wavelet coefficients or the original signal can be used to evaluate the threshold that can eliminate noise in the wavelet domain.

The threshold selection method used in the study is fixed threshold estimation, and the function expression is shown in the formula:

$$\lambda = \sqrt{2 \log(N)} \tag{7}$$

Compared with extreme threshold estimation and unbiased likelihood estimation methods, fixed threshold estimation is more effective in denoising.

3 The Establishment of SVM Bearing Fault Classification Model

3.1 Rolling Bearing Fault Classifier Selection

SVM is a machine learning method based on statistical learning theory. It shows excellent performance in small sample situations. It is currently widely used in pattern recognition, function regression, fault diagnosis, etc. [13]. Since rolling bearings have to distinguish different types of faults, the classifier must be selected first. There are two typical methods for establishing SVM: The first is the “one-to-one” strategy, which, as the name suggests, is to compare classifications one to one. This is also the original idea of SVM. The second is the “one-to-many” strategy, because there are multiple categories, so multiple one-to-one processes need to be repeated. Therefore, this method is to complete the classification of multiple categories by designing an SVM model between every two types of samples, so it is necessary to construct multiple one-to-one SVM classifiers. Assuming that there are m sample categories, a total of $m(m - 1)/2$ SVM classifiers [14]. In the test, because the identified categories have different degrees of failures of the outer ring, inner ring and stick, this article chooses to use the “one-to-many” strategy to study the identification of different types of failures.

The specific application steps of the SVM model for rolling bearing fault identification are as follows:

1. The vibration signal of rolling bearing is normalized to eliminate the influence of dimension;
2. Select the appropriate kernel function for SVM according to the problem of bearing fault classification;
3. Solve high-dimensional optimization equations to obtain support vectors and corresponding Lagrange operators;
4. Calculate the optimal classification surface equation;
5. Using the obtained bearing fault classification model, the input unknown fault bearing samples can be used to identify different fault types.

3.2 Selection of Kernel Function for Bearing Fault Diagnosis Classification

The kernel function is a function that calculates the inner product of two vectors in the implicitly mapped space. The kernel function first does the inner product of the eigenvectors, and then uses the function to transform, which helps avoid direct calculation in high-dimensional space, greatly simplifying the problem solving, and this is equivalent to doing the kernel mapping on the vector first and then doing the inner

Product [15]. There are four commonly used kernel functions: linear kernel function, radial basis function kernel function, polynomial kernel function, and Sigmoid neuron activation kernel function. Their expressions are as follows [16]:

Linear kernel function:

$$K(x_i, x_j) = x_i^T x_j \quad (8)$$

Among them, x_i, x_j is a feature vector, and the inner product of two input feature vectors is obtained.

RBF kernel function:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (9)$$

Among them, x_i, x_j is an eigenvector, σ is the root variance, represents the width parameter of the function, and controls the radial range of the function.

Polynomial kernel function:

$$K(x_i, x_j) = [1 + (x_i, x_j)]^d \quad (10)$$

Among them, x_i, x_j as above, d represents the order of the polynomial.

Sigmoid kernel function:

$$K(x_i, x_j) = \tanh[y(x_i, x_j) + c] \quad (11)$$

Among them, x_i, x_j as above, the gamma displacement parameter in the formula c represents the control threshold range.

In order to compare the accuracy of the four kernel functions, this paper designs an optimal kernel function based on MATLAB calling the libSVM function model. Apply the c-svc algorithm to traverse each kernel function and the parameters under the kernel function.

3.2.1 L-M Algorithm Combined with SVM Method Construction

With the rise of artificial intelligence, many scholars directly predict the type of bearing failure based on artificial intelligence training on samples. This type of method is fast and the model is easy to build, but often requires a large number of samples, and it is also prone to problems of non-convergence and over-fitting, and it is lacking in robustness.

Therefore, this article considers the use of less sample size, more robust algorithm. The Levenberg–Marquardt (L-M) algorithm was proposed by K. Levenberg in the twentieth century and developed and improved by scholars such as D. Marquardt. It has become increasingly mature. Yan et al. [17] proved that at the beginning of the

iteration, the L-M algorithm has a large damping factor, has the characteristics of large initial decline, rapid iteration, and robustness, which reduces the dependence on the initial value. At the end of the iteration, the damping factor is close to 0, and the L-M algorithm has the second-order convergence of the Newton method, which avoids the zigzag oscillation of the steepest descent method. And by introducing an adaptive damping factor, the L-M algorithm can change the search direction and step length at the same time in the iterative process, which is easier to converge than the optimal multiplier method.

In 2003, Pan [18] proposed an L-M method based on trust region techniques, which takes iterative parameters $\mu_k = \alpha_k \|F_k\|$, among which the factors α_k are modified using trust region techniques.

In the k -th iteration, define the function

$$f(x) = \frac{1}{2} \|F(x)\|^2 \tag{12}$$

The actual reduction $Ared_k$ and the estimated reduction $Pred_k$ are:

$$Ared_k = \|F_k\|^2 - \|F(x_k + d_k)\|^2 \tag{13}$$

$$Pred_k = \|F_k\|^2 - \|F(x_k + J_k d_k)\|^2 \tag{14}$$

Its ratio is

$$r_k = \frac{Ared_k}{Pred_k} \tag{15}$$

r_k is used to decide whether to accept the step size d_k and adjust the size of the factor α_k in the iteration process. Generally speaking, the larger, the more the objective function drops, and the longer d_{k+1} the next time can be made, so it can be reduced α_k . On the contrary, the smaller r_k the value, it can be rejected d_k , increase α_k . The specific operations are as follows:

$$x_{k+1} = \begin{cases} x_k + d_k, & \text{if } r_k \geq h_0, \\ x_k, & \text{otherwise,} \end{cases} \tag{16}$$

$$\alpha_{k+1} = \begin{cases} a_1 \alpha_k, & \text{if } r_k < h_1, \\ \alpha_k, & \text{if } h_1 \leq r_k \leq h_2, \\ \max\{a_2 \alpha_k, \alpha_{\min}\}, & \text{otherwise,} \end{cases} \tag{17}$$

Among them, $0 < a_2 < 1 < a_1, 0 < h_0 < h_1 < h_2 < 1, 0 < \alpha_{\min} \leq \alpha_0$.

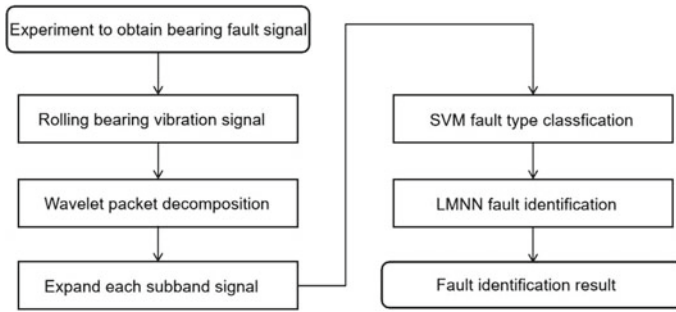


Fig. 1 SVM-LMNN algorithm bearing fault diagnosis network architecture

3.3 Fault Diagnosis Model of Rolling Bearing Based on SVM-LMNN Algorithm

The SVM-LMNN algorithm is used to construct the fault diagnosis model of rolling bearing, and the network architecture is shown in Fig. 1.

Specific steps are as follows:

1. First, perform wavelet packet decomposition on the bearing fault vibration signal, select the appropriate wavelet function and the number of decomposition layers, and obtain the decomposition coefficient of each node.
2. Reconstruct wavelet packet decomposition coefficients, expand each subband signal to the original signal length, and construct $10 \times 120 \times n$ fault feature vector matrices.
3. Input the fault feature vector of the training sample into the SVM, perform grid training, divide the fault into 10 types and test the accuracy.
4. Finally, input the test samples into the trained LMNN for further fault identification

4 Results

4.1 Experimental Verification and Simulation Analysis

4.1.1 Data Collection Experiment

In order to verify the effectiveness of the proposed method, this paper selects the bearing fault data published by Case Western Reserve University in the United States to analyze the bearing inner ring data plus background noise. The rolling bearing model is 6205-2RS JEM SKF, and the motor load is 0. The fault point is artificially manufactured with electric sparks. The diameter of the fault point is 0.007 in and the depth is 0.011 mm. Using the acceleration data of the bearing drive end, the

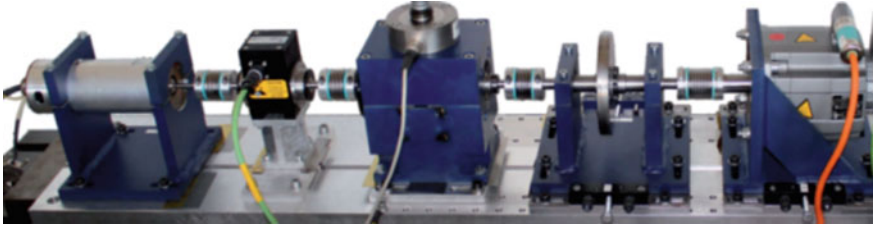


Fig. 2 Rolling bearing vibration signal acquisition test bench

signal sampling frequency is 12 kHz, and the spindle speed is 1772 r/min (rotation frequency is 29.53 Hz).

The test bench shown in Fig. 2 is a modular system that can generate the measurement data required for the corresponding characteristics and damage characteristics obtained by analyzing the motor current signal. The basic structure of the test bench is composed of a drive motor (permanent magnet synchronous motor) that acts as a sensor, a torque measuring shaft, a test module and a load motor (synchronous servo motor).

4.1.2 Comparison Results of Different Kernel Functions

The test is based on the bearing failure data published by Western Reserve University as the training sample, and 100 sets of test data are drawn, of which 80 sets of data are used for training and 20 sets of data are tested. The schematic diagram of the test process is shown in Fig. 3.

During the training process, for the same training set, different kernel functions are used in the SVM to compare. The test results show that the training degree of

Fig. 3 Flow chart of SVM kernel function selection

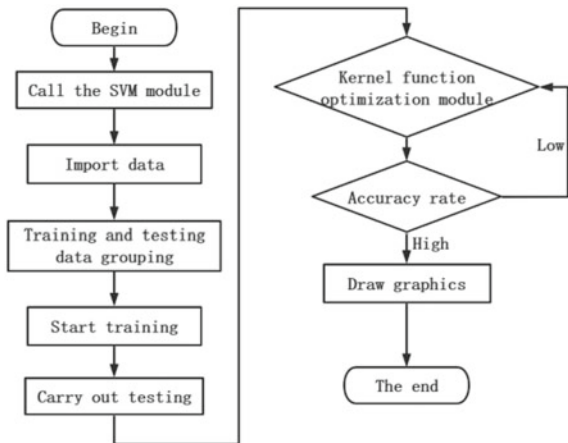


Table 1 Comparison of SVM fault classification results

Kernel function	Timing/s	Accuracy (training/testing)
Linear kernel function	0.306	1.0/0.97
RBF radial basis kernel function	0.295	1.0/1.0
Polynomial kernel function	0.353	1.0/0.98
Sigmoid function	0.333	1.0/0.93

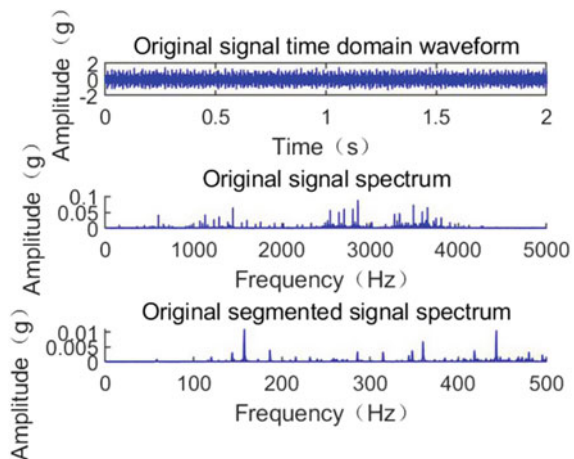
the RBF radial basis kernel function is closest to the actual result. The classification effect is the best, so the radial basis kernel function is used in the test. For the radial basis kernel function, the smaller the gamma value, the higher the training accuracy. The grid search method is used to select two parameters: the parameter that adjusts the penalty for the wrongly distinguished sample and the parameter gamma of the radial basis kernel function. Experiment selection, $\gamma = 0.01$, see Table 1.

This paper takes the inner ring of a rolling bearing as an example, and extracts part of the data as the fault data of the bearing. The extracted signal time domain diagram, frequency domain diagram and segmented frequency domain diagram are shown in Fig. 4.

Choose haar wavelet, decompose the number of layers to 11, set the threshold $k = 100$, and the wavelet detail coefficients and approximate coefficients obtained after Hilbert transform are shown in Fig. 5a, b.

The selection of the threshold in this step is the key to denoising, which directly affects the quality of the reconstructed signal. After decomposing the noisy signal, the approximation coefficient is simplified as the number of layers increases, and

Fig. 4 Time domain diagram, frequency domain diagram and segmented frequency domain diagram of the inner circle fault signal



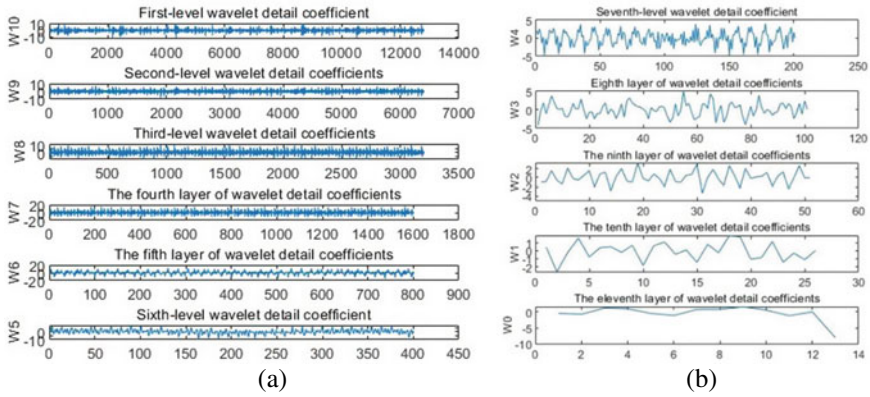


Fig. 5 1–11 layer wavelet packet decomposition fault signal

the smoothness of the reconstructed signal is improved, which can effectively overcome the “Gibbs” phenomenon, which proves that the currently selected threshold is effective.

After using SVM to classify bearing faults, the classification results are shown in Fig. 6a, and the classification accuracy rate reaches 95%. Compared with the 93.5% accuracy result obtained after envelope spectrum demodulation, as shown in Fig. 6b, it proves that the classification accuracy after wavelet decomposition is significantly higher than the envelope spectrum demodulation result.

The signal is input into the rolling bearing fault diagnosis framework of the SVM-LMNN algorithm. The diagnosis accuracy is shown in Fig. 7. The training accuracy is 99.938%, and the test accuracy is 99.456%. Compared with the instantaneous energy method that selects the VMD component through the kurtosis criterion, The accuracy rate is significantly improved.

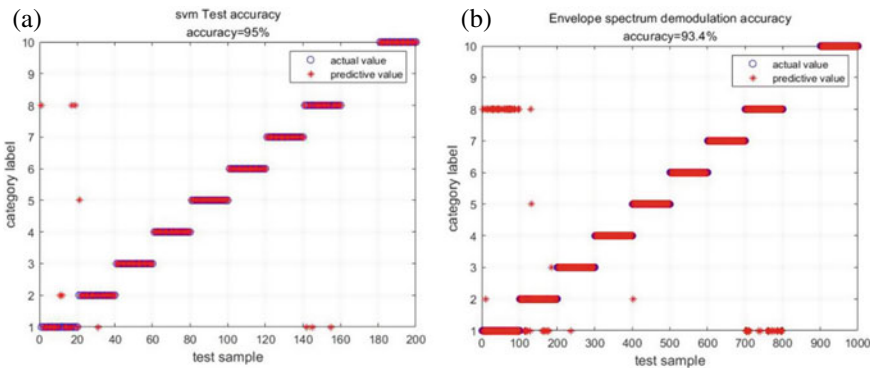


Fig. 6 Comparison of SVM classification accuracy and envelope spectrum demodulation accuracy

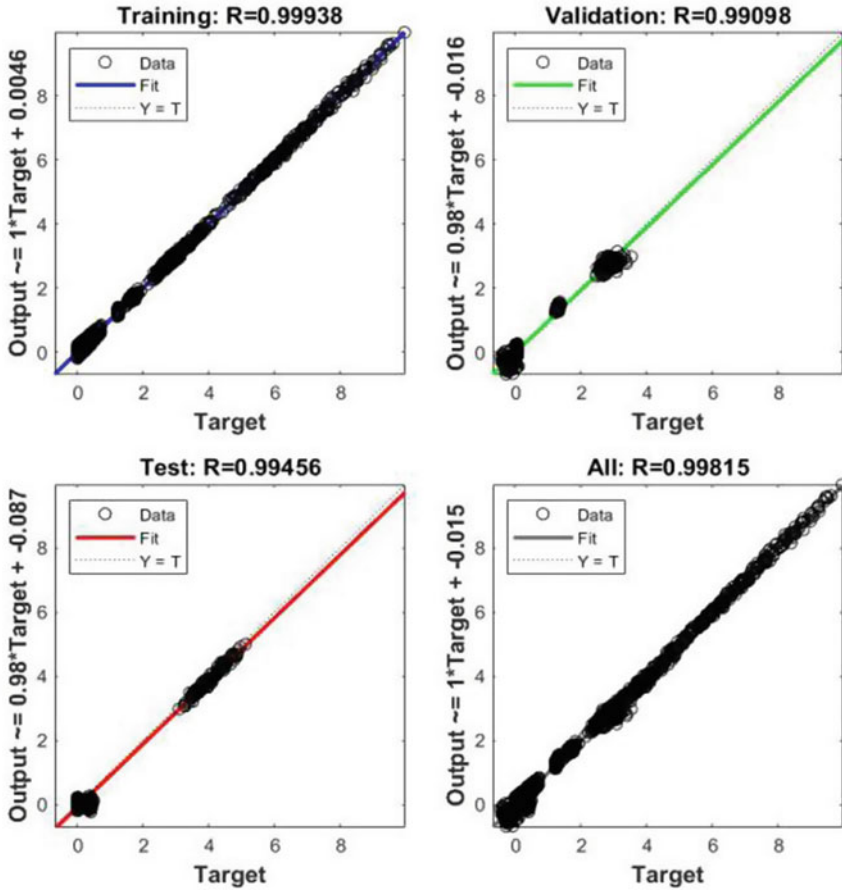


Fig. 7 SVM-LMNN algorithm diagnosis accuracy rate

5 Discussion

5.1 Conclusion

According to the reflection of the characteristic frequency of the rolling bearing fault on the fault type, a method of rolling bearing fault diagnosis based on the support vector machine (SVM) and Levenberg–Marquardt (LM algorithm) is proposed. First, perform wavelet packet decomposition on the rolling bearing measurement signal, adjust the appropriate filter threshold through the obtained detail coefficients and approximate coefficients; use SVM to multi-classify the fault state of the noise-reduced signal; finally, input the feature vector into LMNN (LM algorithm Neural network) to realize failure mode recognition and compare with existing methods. The

results show that this method can effectively extract the fault features of rolling bearings, and it is verified by comparison with existing methods. It has higher reliability and can be applied to fault diagnosis of rolling bearings.

Acknowledgements This research is supported by Beijing Science and Technology Planning Project (Grant No. Z201100008320004).

References

1. Peng, W., Yang, Y., Yu, L., Huang, H.: Frequency subdivision and fault characteristic period identification in bearing diagnosis. *Mech. Des.* **30**(11), 38–41 (2013)
2. Zhou, L., Xia, X., Wan, Y., Zhang, H., Lei, P.: Overview of harmonic measurement methods based on wavelet transform. *Trans. Chin. Soc. Electr. Eng.* **09**, 67–74 (2006)
3. Wang, Y., Wang, H., Han, Q., Li, L., Xiong, Q.: Rolling bearing fault identification method based on wavelet packet and IGA-BP neural network. *Mach. Tool Hydraul.* **48**(17), 184–187 (2020)
4. Xia, T., Zhan, Y., Guo, J.: Bearing fault diagnosis based on wavelet packet and gradient boosting decision tree. *J. Shaanxi Univ. Sci. Technol.* **38**(05), 144–149 (2020).
5. Wen, X.: Research on Early Fault Diagnosis Technology of Rolling Bearings Based on Wavelet Packet Decomposition and Graph Theory. Shandong University (2020)
6. Fang, Q., Liu, Q., Liu, T.: Rolling bearing fault diagnosis method based on hybrid domain feature set and PSO-SVM. *Mech. Eng. Autom.* **02**, 1–3 (2020)
7. Hu, Q., Zhu, H., Zhao, K., Qin, A.: Research on fault diagnosis method of rolling bearing based on GA optimized SVM. *J. Guangdong Inst. Petrochemical Technol.* **30**(01), 44–47+53 (2020)
8. Ren, B., Feng, Z.: Improved genetic algorithm and particle swarm optimization algorithm and their comparative analysis. *J. Nanjing Normal Univ. (Engineering Technology Edition)* (02), 14–20 (2002). Xiao, L., Zhang, J., Zeng, M., Dong, J.: Neural network short-term electricity price prediction based on wavelet decomposition and the same scale sequence. *Electr. Demand Side Manag.* **13**(04):19–22+29 (2011)
9. Ju, F.: Improvement and Application of Particle Swarm Optimization Algorithm. South China University of Technology (2014)
10. Peng, X., Gui, W., Huang, Z., Hu, Z., Li, Y.: GAPSO: an efficient genetic particle hybrid algorithm and its application. *J. Syst. Simul.* (18), 5025–5027+5031 (2008)
11. Zheng, W.: Power quality disturbance signal denoising algorithm based on improved wavelet threshold. *Electr. Switch* **59**(01), 28–33 (2021)
12. Sun, G., Liu, X., Ma, L.: Research on rolling bearing fault diagnosis method based on wavelet noise reduction and improved HHT. *Coal Mine Mach.* **36**(09), 323–326 (2015)
13. Li, Y., Li, H., Liu, Z.: Research on fault diagnosis of rolling bearing based on CEEMDAN multi-scale entropy and SSA-SVM. *Mech. Electr. Eng.* **38**(05), 599–604 (2021)
14. Li, Y., Liu, X., Han, L., Fu, C.: SVM-based phase recognition method of oil-water two-phase flow. *Pet. Pipes Instrum.* **7**(02), 91–95 (2021)
15. Dong, B.: Support vector technology and its application research. Dalian Maritime University (2016)
16. Fu, Y., Ren, D.: Research on kernel function and its parameter selection in support vector machine. *Sci. Technol. Innovation Herald* **09**, 6–7 (2010)
17. Yan, Z., Fan, X., Zhao, W., Xu, X., Fan, J., Wang, Y.: Adaptive Levenberg-Marquardt method improves power flow calculation convergence. In: *Proceedings of the Chinese Society of Electrical Engineering* **35**(08): 1909–1918 (2015)
18. Pan, F., Huang, S.: Research on visible light positioning based on improved LM algorithm. *World Electr. Prod.* **27**(11), 75–78 (2020)

Fine-Tuning and Efficient VGG16 Transfer Learning Fault Diagnosis Method for Rolling Bearing



Jinglei Su and Hongjun Wang

Abstract Nowadays, neural network become popular in modeling. However, the model training needs a lot of data, long training time and high hardware conditions. It is inefficient for ordinary computing devices to be used in training models. In this paper, VGG16 model was modified to fit ten labels and used as feature extractor. The default image size of model was 224×224 pixels. Then the images were reduced into low resolution as 112×112 , 75×75 , 56×56 , 45×45 , 32×32 pixels, which were 1/2, 1/3, 1/4, 1/5 of default side length and the minimum size. Next these images were sent to model for training. The training results illustrated that the images of 112×112 , 75×75 , 56×56 groups can still be adequate for modified VGG16 to classified and achieve high accuracy and meanwhile significantly reduce the training time. However, when the size dropped to 45×45 , 32×32 , overfitting appears and the training accuracy significantly dropped. Thus, it is recommended that set a target accuracy first and begin training from a small size. If the accuracy was not high enough, enlarge the size and train again.

Keywords Fault diagnosis · Bearings · Transfer learning · VGG16 · Fine-tuning

1 Introduction

Bearing is widely used in life, any bearing fault may affect production efficiency and life security. It could cause shutdown maintenance, safety accidents and even endanger personal safety. So timely and accurate bearing fault diagnosis can improve the safety and reliability of mechanical equipment, ensure personnel safety, timely maintenance and reduce costs.

J. Su · H. Wang (✉)

School of Mechanical Engineering, Beijing Information Science and Technology University, Beijing 100192, China

e-mail: wanghj86@163.com

Beijing International Science Cooperation Base of High-End Equipment Intelligent Perception and Control, Beijing, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

453

H. Zhang et al. (eds.), *Proceedings of IncoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_37

Generally, the method of bearing fault diagnosis is to collect bearing vibration signals, find the characteristics of bearing fault signals, and analyze and identify them. With the development of artificial intelligence technology, signal processing mode has gradually changed to intelligent analysis. Intelligent analysis methods include k-nearest neighbor algorithm, support vector machine, artificial neural network, etc. [1]. For example, in paper [2], the convolution neural network was used to predict the fault diagnosis of multi-sensor bearings; In paper [3], the multi-objective convolutional neural network was used to diagnose the fault and analyze the vibration signal; paper [4] proposed a stochastic convolutional neural network for the health detection of internal combustion engines; In paper [5] the data and model were combined to predict the remaining service life of bearings.

Furthermore, more scholars applied transfer learning method to the field of fault diagnosis, and there were many kinds of transfer methods. For example, the multi-scale convolution transfer neural network proposed in paper [6] was applied to variable condition bearing fault diagnosis, and good results have been achieved; In paper [7], rolling bearing migration fault diagnosis method based on manifold embedding and distribution alignment was proposed; paper [8] used deep adversarial transfer learning to generate data to solve the problem of insufficient data; paper [9] deals with multi-conditions problems based on feature transfer learning method; In paper [10], deep belief network transfer learning was used to predict planetary gearbox faults.

In the aspect of fault signal preprocessing, in paper [11], RGB images were generated by continuous wavelet analysis, wavelet packet time–frequency analysis, Wigner Ville distribution and spectral kurtosis, and then sent to VGG19 model; In paper [12], the pre-trained ResNet18 was used as the feature extractor. However, general methods require plenty data and time to train and distinguish the fault information, the training efficiency was low, the hardware equipment requires high conditions, and the ordinary computing equipment was inefficient to be used in the training model, so whether there was a way to improve the training efficiency and reduce the training time without losing too much accuracy. This paper has experimentally reduced the image size and according to the results, the features information carried by the image were enough for model to classified and significantly reduced time consuming.

2 VGGNet Basic Principles

VGG network was proposed by visual geometry group of Oxford University. The typical networks were VGG16 and VGG19. VGG16, as shown in Fig. 1 and Table 1, was the basic network in the first place of positioning task and the second place of classification task of ImageNet competition in 2014 and it has a total of 138,357,544 parameters. It was widely used in fine-tuning transfer learning. VGG19 was based on VGG16. It adds three layers of convolution layer based on VGG16. It has a total of 143,667,240 parameters.

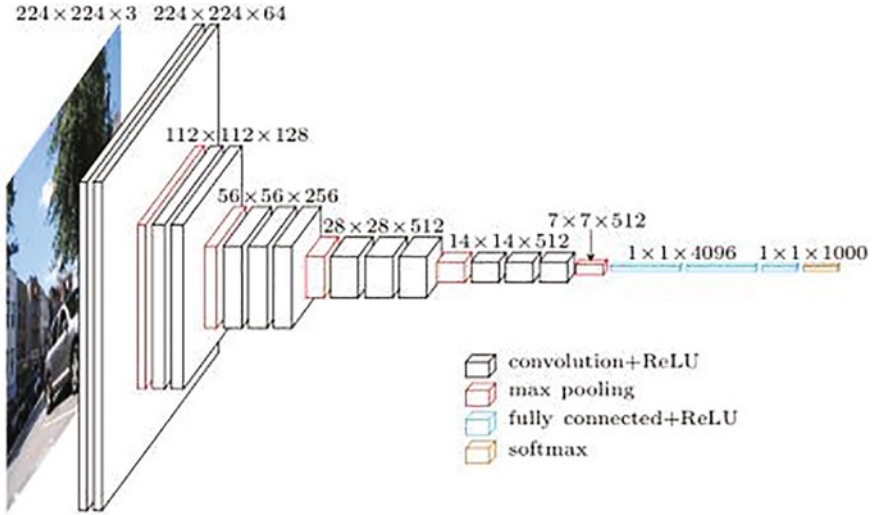


Fig. 1 VGG16 structure

Table 1 VGG16 and VGG19 structure

ConvNet Configuration	
VGG16	VGG19
16 weight layers	19 weight layers
input(224×224 RGB image)	
conv3-64	conv3-64
conv3-64	conv3-64
maxpool	
conv3-128	conv3-128
conv3-128	conv3-128
maxpool	
conv3-256	conv3-256
conv3-256	conv3-256
conv3-256	conv3-256

maxpool	
conv3-512	conv3-512
conv3-512	conv3-512
conv3-512	conv3-512
maxpool	
conv3-512	conv3-512
conv3-512	conv3-512
conv3-512	conv3-512
maxpool	
FC-4096	
FC-4096	
FC-1000	
soft-max	

3 Model and Preprocessing

3.1 Fine-Tuning Model

The modified VGG16 model is shown in Fig. 2. The weights of VGG16 network trained from ImageNet were retained. The weights of the first three sections of the network model were frozen. The weights of the fourth and fifth sections of the network were set as trainable. The full connection layer was reset in the last section. The source domain of transfer learning was the image trained on ImageNet, and the target domain was the gray image converted from the time domain signals.

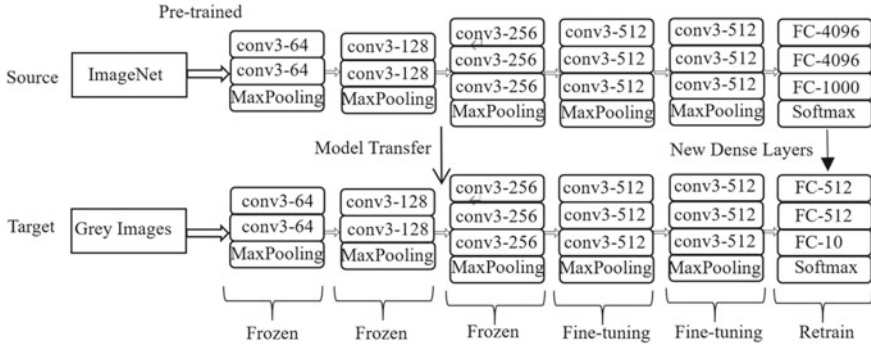
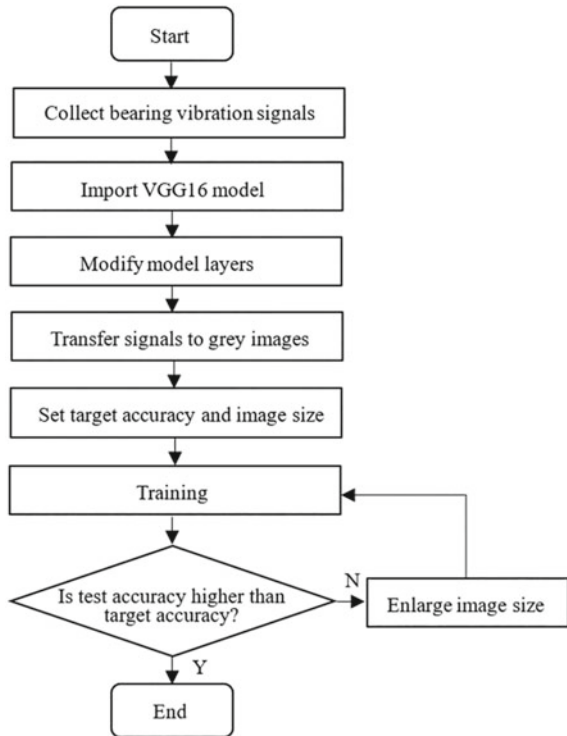


Fig. 2 VGG16 modified model

The training flow chart is shown in Fig. 3. First, the bearing fault vibration signal is obtained. Second, import the VGG16 model with the training weight of ImageNet. Third, modify model layers. Fourth, convert signals to gray images. Five, set a target accuracy and a small size of image (1/4 of default size is recommended). Then, train

Fig. 3 Flow chart



the model and check the test accuracy if it achieves the target. If negative, enlarge the size of images and train again and so on. If positive, the weight of model is enough to be used.

3.2 Bearing Database

The data sets used in the experiment was the bearing data sets of Kaiser Western Reserve University. The experimental platform is shown in Fig. 4.

Platform composition: a 1.5 kW (2 HP) motor (left side of the figure); One torque sensor/decoder (middle connection of Figure); A power tester (right side of the figure); Electronic controller.

According to the fault location and fault diameter, the data set was classified into 9 kinds of faults and 1 kind of normal, with a total of 10 kinds of data labels, as shown in Table 2. Each kind of fault data also includes acceleration data collected from multi-sensor at different positions of the drive end, fan end and base, and multi-conditions data of 0, 1, 2 and 3 HP, which are classified as one kind of fault. The training set, validation set and test set were selected from this data pool.

Fig. 4 Kaiser Western Reserve University experiment platform

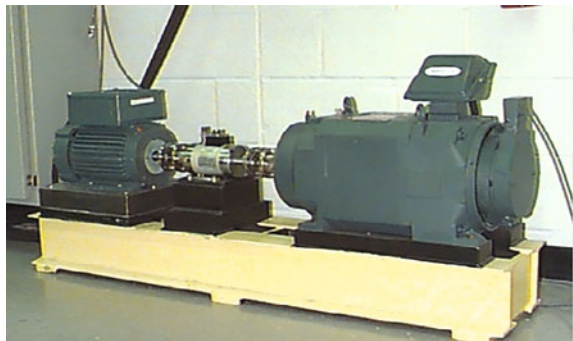


Table 2 Datasets labels

Fault diameter	Inner race	Ball	Outer race
0.007"	IR007	B007	OR007
0.014"	IR014	B014	OR014
0.021"	IR021	B021	OR021

3.3 Data Preprocessing

The preprocessing method proposed by Zhao et al. [13] was used as the preprocessing method of this experiment. Firstly, the vibration time-domain image was obtained from the acceleration sensor, and the maximum and minimum values of vibration amplitude were recorded. Each value was mapped into the (0, 255) interval and rounded

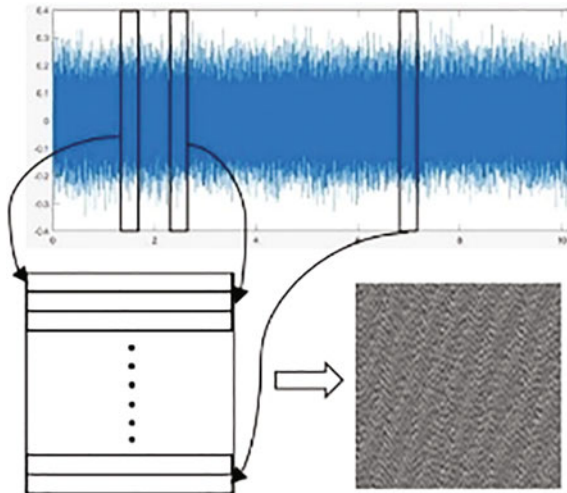
$$P_i = \text{round}\left(\frac{V_i - \min}{\max - \min} \times 255\right) \tag{1}$$

The default input image size of general model was no more than 224×224 , and the image size can be adjusted, so first set 224×224 as the initial image size, a total of 50,176 pixels, in order to obtain the most features.

As shown in Fig. 5, 224 continuous vibration signal segments with 224 values in length were randomly selected, and the 224 groups mapped values are arranged into a square matrix, and then converted into a gray image. According to this method, multi segment signals were randomly selected to obtain the second gray image. Put back sampling, and then randomly took 224 groups of fragments to form a matrix again, turn them into gray images, and so on to create data sets.

The default size of modified VGG16 is 224×224 which will be converted to 112×112 , 75×75 , 56×56 , 45×45 , 32×32 those are 1/2, 1/3, 1/4, 1/5 of the original side length and the minimum size of model. Original image and reduced images are shown in Fig. 6

Fig. 5 Data preprocessing



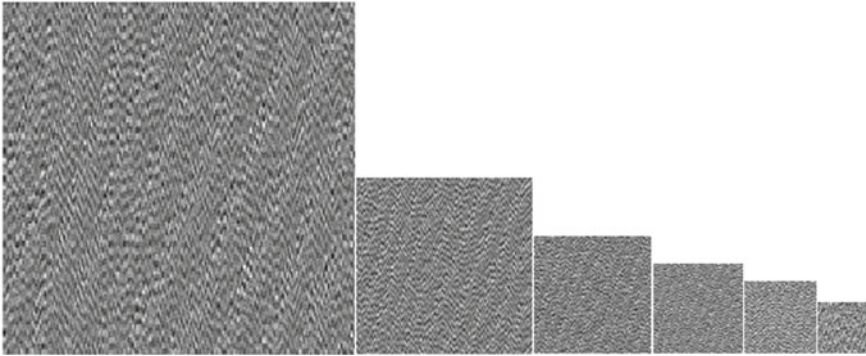


Fig. 6 Same grey picture with different size

4 Experiment and Results

The experiment used modified VGG16 network to test whether it is adequate for classifying images and count how much time it consumes.

4.1 Different Size Images Training

Firstly, the bearing vibration time domain image was transformed into square gray image dataset according to the above method, and then sent to modified VGG16 model for training.

As shown in Fig. 7a, b and c and Table 3, after decreasing the side length and resolution, the test accuracy did not decrease significantly at the beginning until the side length reaches 45×45 , Which was one fifth of the original side length, the accuracy of the validation set begins to decline in Fig. 7d, indicating that there was a certain overfitting phenomenon. With the further reduction of the size, the overfitting

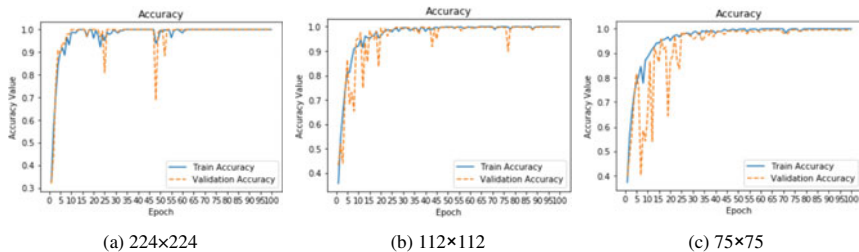


Fig. 7 Different size training graphs

Table 3 Test accuracy and time consuming of different size

Size (pixels)	Test accuracy (%)	Time consuming
224 × 224	100	100
112 × 112	99.9	31.0
75 × 75	99.1	20.8
56 × 56	97.5	14.0
45 × 45	91.6	12.8
32 × 32	88.1	10.2

phenomenon was more obvious in Fig. 7e and f, and the test accuracy of 32 × 32 group has dropped to 88.1%.

Test accuracy did not decrease linearly as sizes decrease as Table 3 shown. The grey images still held enough information for modified VGG16 to classify. It only slightly decreased until the side length was one fifth of the original length.

The default input of modified model is 224 × 224. The training time of group 224 is the longest, and with decrease of size, the training time is significantly shortened. The dataset 224 with the longest time consuming was set to be 100, the comparison and proportion of the time consuming of other groups were also shown in Table 3.

It can be seen that the time consuming of 112 data set was obviously reduced, only 31% of the 224 dataset and the time consuming of 45 dataset is only 12.8% of 224 dataset. Although the time consuming of size of 32 × 32 had shorten to 10.2% of the default size, the test accuracy had dropped to 88.1%.

5 Discussion

From results, when the grey images were reduced, the feature information the images carried was still adequate for modified VGG16 to extract and achieved a good prediction. The reduced image can still highly maintain the original data features. The accuracy can still reach 99.9% without serious overfitting and significantly shorten the time consuming even when if half the default size. Moreover, further reducing the image size can shorten the time consuming even more. Thus, when the training time was too long, it should set a target accuracy and train the one fourth of its default size dataset first. If the test accuracy was not enough, train size of one third dataset and so on. But beware that if the image size was too small, at one fifth or the minimum size of model, it would result in overfitting and test accuracy declining.

This is a trial that test if the modified VGG16 can extract the features from reduced images and check how much time it consumes. It has not tested every tiny decrease of pixels. It could draw a more smoothly line indicating the relation of extracting ability and image pixels in the future.

Acknowledgements This research is supported by The National Natural Science Foundation of China (Grant No. 51975058).

References

1. Li, S., Hou, Y., Li, X.: Review on time-frequency-domain analysis methods for vibration faults of rolling bearings. *J. Chongqing Univ. Technol.* (2021)
2. Xia, M., Li, T., Xu, L.: Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. *Trans. Mechatron.* (2017)
3. Xin, Y., Li, S., Wang, J.: Intelligent fault diagnosis method for rotating machinery based on vibration signal analysis and hybrid multi-object deep CNN. *Sci. Meas. Technol.* **14**(4), 407–415 (2020)
4. Wang, R., Chen, H., Guan, C.: Research on health monitoring method of internal combustion engine based on stochastic convolution neural network. *J. Vibr. Eng.* (2021)
5. Wang, B., Lei, Y., Li, N.: A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *Trans. Rel.* **69**(1), 401–412 (2020)
6. Zhao, B., Zhang, X., Zhan, Z.: Deep multi-scale convolutional transfer learning network: A novel method for intelligent fault diagnosis of rolling bearings under variable working conditions and domains. *Neurocomputing* (2020)
7. Wang, X., Tong, J., Zheng, J., Pan, H., Pan, Z.: Transfer fault diagnosis for rolling bearings based on manifold embedded distribution alignment. *J. Vibr. Shock* **40**(8), 110–116 (2021)
8. Yue, S., Lei, W., Xue, Y.: Research on fault diagnosis method of deep adversarial transfer learning (2021)
9. Kang, S., Hu, M., Wang, Y.: Fault diagnosis method of a rolling bearing under variable working conditions based on feature transfer learning. *Proc. CSEE* **39**(3), 764–772 (2019)
10. Chen, R., Yang, X., Hu, X.: Planetary gearbox fault diagnosis method based on deep belief network transfer learning. *J. Vibr. Shock* **40**(1), 127–150 (2021)
11. Li, C., Liu, Z.: Time frequency analysis and bearing fault detection of VGG19 transfer learning. *Electr. Measur. Technol.* **44**(5), 59–67 (2021)
12. Yang, B., Li, Q., Chen, L.: Bearing fault diagnosis based on ResNet and domain adaptation (2021)
13. Zhao, Y., Xu, G., Liu, M.: Method for fault diagnosis of bearing based on transfer learning with VGG16 model. *Spacecraft Environ. Eng.* **37**(5), 446–451 (2020)

An Investigation of Unsupervised Data-Driven Models for Internal Combustion Engine Condition Monitoring



Xiaoxia Liang, Chao Fu, Xiuquan Sun, Fang Duan, David Mba, Fengshou Gu, and Andrew D. Ball

Abstract Internal combustion (IC) engines are widely employed in power systems such as marine ships, small power stations and vehicles. However, due to its complex working conditions and sophisticated degradation mechanisms, IC engines commonly suffer various types of malfunctioning and faults, which affects their performance in power delivery. Therefore, it is important to monitor the condition of IC engines and detect faults occurred in time. In this paper, two unsupervised data-driven models using machine learning techniques are employed and investigated for the purpose of online condition monitoring and fault isolation of IC engines. A misfire and a lubrication system filter blocking faults are experimentally studied on a purposely built marine engine test rig. The performance of the two models and their contribution maps are discussed, which provides guidance for using such unsupervised models for the condition monitoring and fault detection of IC engines.

Keywords IC engine · Fault detection · Unsupervised machine learning · Misfire · Lubrication system filter blocking

1 Introduction

IC engines are crucial components of many reciprocating machinery and play a vital role in their performance, hence they affect seriously the reliable operation

X. Liang · F. Duan
School of Engineering, London South Bank University, London 1 0AA, UK

X. Liang · A. D. Ball
School of Engineering, Hebei University of Technology, Tianjin, China

C. Fu (✉) · X. Sun · F. Gu
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield H1 3DH, UK
e-mail: C.Fu@hud.ac.uk

D. Mba
Faculty of Technology, De Montfort University, Leicester L1 9BH, UK

of the power system equipped in ships, vehicles and generators. Upon this fact, effective condition monitoring and fault isolation of such power system are useful in practical engineering. However, due to their complicated nature and variable working conditions, it is a very clumsy work to do online diagnosis. Currently, data-driven machine learning methods emerge in the health monitoring of engines, which take advantages of modern development of facilities and provide an effective alternative to traditional approaches.

Modern industrial facilities are heavily automated and instrumented; consequently there is a lot of process data available which can be used to monitor the condition of the system.

For modern industrial facilities, there are a lot of process data available which can be used to monitor the condition of the system, but not all facilities are equipped with vibration sensors, as it usually has a high cost of sensor installation, data storage, and data analysis software/services. Therefore, we plan to use common multivariate process data, with no additional condition monitoring data assisted, to detect and isolate faults in IC engine systems.

Machine learning algorithms can either be supervised or unsupervised [1]. In industrial applications, supervised ones are not proper when there were missing labels in the historical data. Considering the unlabeled data are easier to obtain in practice, this paper applies two unsupervised machine learning models in condition monitoring and fault isolation of IC engine systems. Two case studies are carried out, and relevant results are discussed in detail. The performance and improvement required are highlighted as well.

The highlight of this paper is to applied the reconstruction based fault detection scheme from our previous work in reference [10] on an IC engine system. The training data in this paper is a combination of 13 different working conditions, which is more complex than that in [10], in which training data and test data were under the same working condition. Two-dimensional Q statistic contribution maps and residual maps are introduced for fault analysis.

The main contribution is to achieve automatic fault detection through the proposed method. In practical applications, the proposed method could give real-time alerts when anomalies detected via the online monitoring data. Meanwhile, the Q statistic contribution map and residual map can assist operators in inferring the possible fault type and planning the necessary inspection/maintenance activities.

2 Literature Review on Machine Learning Algorithms

Machine learning algorithms can be categorized into supervised and unsupervised methods [1]. The difference between these two main categories is whether there exist labels in the training dataset. Supervised algorithms use well-labelled data to perform analysis tasks, and then construct contingent functions to map new instances of attributes. The algorithms require pre-specifications (labeled datasets) for maximum settings to obtain the desired results or performance levels [1], which is often not

accessible in most condition monitoring scenarios. Wang et al. [2] utilized Bayesian networks for fault isolation and applied it on a diesel fuel injection system. As a supervised method, the prior probability of multiple faults was assigned according to experts' knowledge. Key parameters for fault isolation included injection starting pressure, peak injection pressure, injection duration, and peak factor. In contrast, unsupervised learning algorithms involves pattern recognition without the involvement of a target attribute. Due to the advantage of performing clustering without any labels, the unsupervised methods are more feasible for anomaly detection (namely one-class classification) area.

Among the unsupervised methods, the principal component analysis (PCA) and sparse autoencoder (SAE) are proved effective in anomaly detection. The PCA projects high dimensional data to a lower-dimensional subspace to better distinguish "normal" data from "abnormal" one. Jafarian et al. [3] applied PCA on vibration analysis based on Fourier transform and discrete wavelet transformation to detect and categorize faults that appeared due to poppet valve clearance and incomplete combustion (misfiring) in an internal combustion engine. Mathew and Zhang [4] developed a fault detection and classification method based on PCA and Bayesian optimization, using acoustic signals captured from cylinder heads of a Ford EcoBoost engine, to classify three faulty conditions, which are engine misfire, ignition timing variation and air-fuel ratio variation. After signal decomposition, relevant time-frequency information was extracted and used as inputs for fault classification. Wang et al. [5] developed a nonlinear PCA anomaly detection method for a Volkswagen 1.9-L turbocharged diesel engine. The autoencoder is a special type of neural networks, which outputs are reconstructions of inputs. The SAE is a variant from autoencoder by putting a sparsity constraint on the hidden units to deal with the overfitting problem [6]. The application of autoencoder fault detection methods can be found mainly in aircraft engines. Fu et al. [7] proposed a grouped convolutional denoising autoencoder model for aircraft engine fault detection. A supervised machine learning method, support vector machine (SVM), was used as a comparison. They found that the SVM achieved the best fault detection results on training data, but the worst results on test data; the hybrid method, which combines denoising autoencoder and SVM, obtained better results; and their proposed method had the best performance. The autoencoder method applied in aircraft engines using vibration data can be found in [8], and that using multivariate data can be found in [9].

3 Unsupervised Data-Driven Models for Online Fault Detection

A similar approach for fault detection, as proposed in [10], is employed in this work. Two unsupervised fault detection methods, PCA and SAE, are adopted. The Mahalanobis distance (MD) calculates the statistical difference between the monitored data and the reconstructed outputs estimated by PCA or SAE. Fault detection is based

on the comparison of MD calculated in the monitoring stage with the one calculated under healthy stage. Thereafter, the MD is used as a system-wide health indicator for fault detection.

3.1 Principle Component Analysis

Principal component analysis (PCA) is an effective unsupervised method for dimensionality reduction and feature extraction. By calculating the eigenvectors of the covariance matrix of the original input data, PCA linearly transforms a high-dimensional input vector into a low-dimensional with efficient features extracted [11]. The key equation is Eq. (1).

$$X = s_1 p_1^T + s_2 p_2^T + \dots + s_N p_N^T + \varepsilon \approx SP^T \quad (1)$$

where $P = [p_1, \dots, p_N]$ represents the loading or principal vector matrix, which is the eigenvector of the covariance matrix X . $S = [s_1, \dots, s_N]$ is the score matrix of the principal components and ε represents a residual error matrix.

3.2 Sparse Autoencoder

The sparse autoencoders (SAE) is a special unsupervised feedforward neural network. It converts the input data $x = (x_1, \dots, x_D)$ to efficient internal representations, and then decoding to a number of outputs $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_D)$ which looks very close to inputs. A representative structure of SAE is presented in Fig. 1.

As shown in Fig. 1, the SAE network includes two parts, i.e., encoder and decoder. The encoder connects the input layer to the hidden layer, with the weight matrix and the bias of this part being represented by $W^{(1)}$ and $b^{(1)}$, respectively. The decoder connects the hidden layer to the output layer, with the corresponding weight matrix $W^{(2)}$ and bias $b^{(2)}$.

The network tries to reconstruct the input vector in the output layer, as is shown in Eq. (2).

$$\tilde{x} = H_{W,b}(x) \approx x \quad (2)$$

where x is the input variables, and \tilde{x} represents the output variables. $H_{W,b}(x)$ is the nonlinear function of SAE, that predicts output \tilde{x} based on input x , using parameters W and b . More detailed theoretical background can be found in [10].

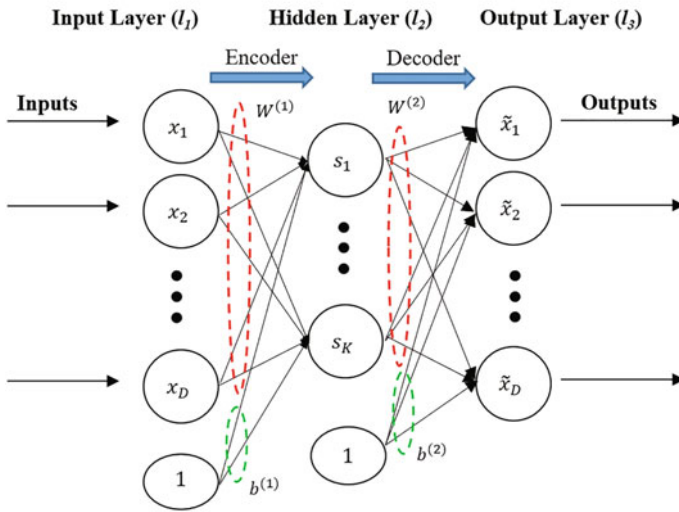


Fig. 1 Sparse autoencoder model structure [10]

3.3 Fault Detection Health Indicator Based on Mahalanobis Distance (MD)

The MD was proposed by Indian statistician Mahalanobis to represent the covariance distance of the data. Unlike the Euclidean distance, the MD is a unitless distance measurement, and takes into account the correlations among variables. The MD provides a univariate distance value for multivariate data, and therefore, was applied in the anomaly detection models.

MD for the training data in this paper was calculated by Eq. (3).

$$MD_i = \sqrt{((Y_i - \bar{Y}_i) - \hat{\mu})S^{-1}((Y_i - \bar{Y}_i) - \hat{\mu})^T} \tag{3}$$

where Y_i is the i -th features in a machine learning model \bar{Y}_i is the reconstructed values of Y_i . $(Y_i - \bar{Y}_i)$ denotes the residual between the reconstructed and the original values. $\hat{\mu}$ and S are the sample mean and covariance of the reference samples, respectively.

4 Experimental Setup

An engine test rig, as shown in Fig. 2, is developed. The engine is connected to a gearbox and provides power to the propeller. The load can be adjusted via a valve

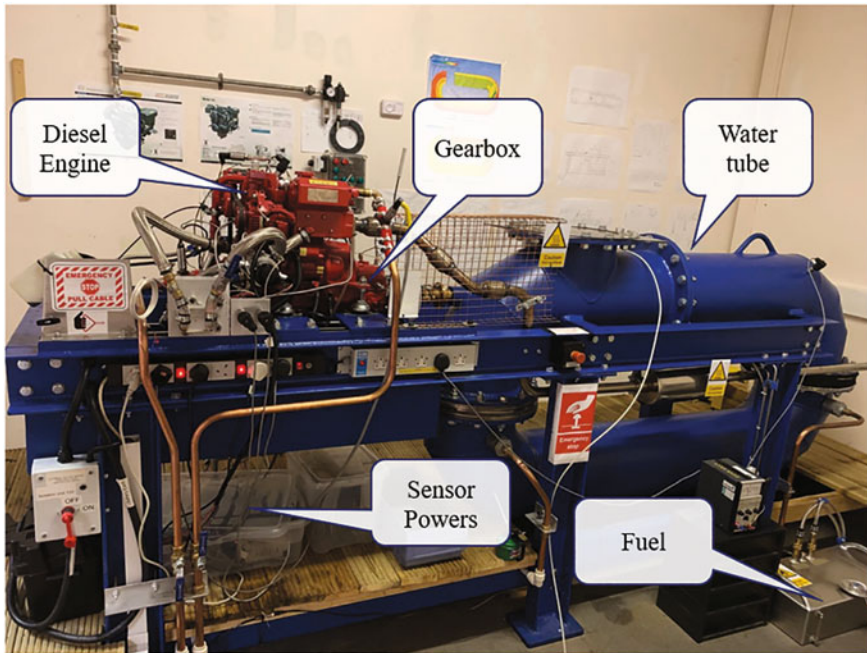


Fig. 2 Marine engine experiment test rig

installed in the middle of the pipe loop. The test rig can simulate faults occurred in the engine oil supply system, lubrication system, cooling circulation system, and power transmission system. In this paper, two types of faults are introduced, which are the misfire and the lubrication system philtre blocking fault. The misfire fault was simulated by only providing engine oil to one of the cylinders; and the lubrication system filter blocking fault was simulated by reducing the size of the filter inlet, which can result in the reduction of flow.

For process data acquisition, 20 variables are measured (see Table 1) to monitor the condition of lubrication, combustion and cooling systems. In this paper, we applied the proposed fault detection methods for detecting the misfire of a cylinder and lubrication system filter blocking. The multivariate process data set was obtained from an in-line two cylinders water-cooled marine engine. These multivariate process data were recorded every second.

Table 1 Measured variables of the in-line two cylinders water-cooled marine engine

ID	Variable names	Description
1	Pressure intake air	Engine intake pressure
2	Pressure coolant in	Inlet pressure of the internal circulating water
3	Pressure Cy1 exhaust	Engine cylinder 1 exhaust pressure
4	Pressure Cy2 exhaust	Engine cylinder 2 exhaust pressure
5	Pressure fuel supply	Pressure of engine oil supply system
6	Temperature gearbox	Temperature of the gearbox housing
7	Temperature ExWater out	Outlet temperature of the external circulating water
8	Temperature Cy2 exhaust	Engine cylinder 2 exhaust temperature
9	Temperature bushing	Bearing (connected to the propeller) oil temperature
10	Temperature coolant in	Inlet temperature of the internal circulating water
11	Temperature coolant out	Outlet temperature of the internal circulating water
12	Temperature Cy1 exhaust	Engine cylinder 1 exhaust temperature
13	Temperature ExWater in	Inlet temperature of the external circulating water
14	Pressure ExWater out	Outlet pressure of the external circulating water
15	Pressure ExWater in	Inlet pressure of the external circulating water
16	Pressure coolant out	Outlet pressure of the internal circulating water
17	Temperature lub oil	Lubrication oil temperature
18	Pressure lub oil	Lubrication oil pressure
19	Engine speed RPM	Engine shaft average speed
20	Pressure water tank	Load level

5 Case Study and Discussion

5.1 Misfire Detection and Diagnosis

Misfire, which is a common engine fault that occurs in an IC engine, may be caused by faulty spark plug, cracked distributor cap, blown head gasket, too high temperature resulting in engine pinging, lean fuel/air mixture, lack of compression or even exhaust gas recirculation issues like the valve sticking closed or open causing too much flow [12]. When the engine suffered from misfire, it will produce consequences like exhaust temperature change, reduced fuel efficiency, increased power loss, and create unique patterns in the vibration domain [12].

In this case study, the PCA and SAE anomaly detection models were trained on the data taken under various health operation conditions. The results of the training process using unsupervised machine learning methods, namely PCA and SAE, are presented in Fig. 3. In the figure, the calculated health indicator is MD; the blue points that below the fault detection threshold, represent normal data; and the red points, which exceeded the threshold, are regarded as anomalies. This is because

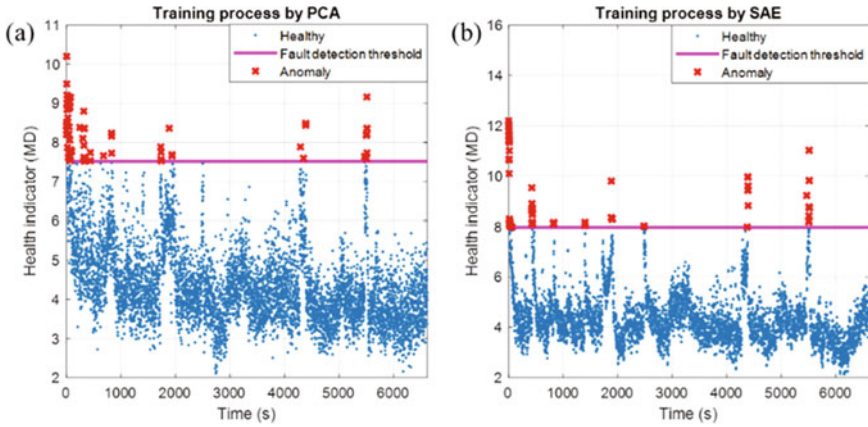


Fig. 3 Training process by two unsupervised machine learning methods, **a** PCA, **b** SAE. (blue dots: data points which were assumed as healthy in the training data, magenta line: fault detection threshold with confidence interval set at 99%, red points: anomalies in training data)

the training data may include some outliers even after data pre-processing. Here we assume 99% of the training data is normal and the left 1% of the training data is viewed as outliers or anomalies.

In this case, fault detection for Cylinder 1 misfire fault by PCA and SAE based methods are shown in Fig. 4. The fault detection thresholds were applied from the training process. The figure clearly shows that, the values of the health indicators are

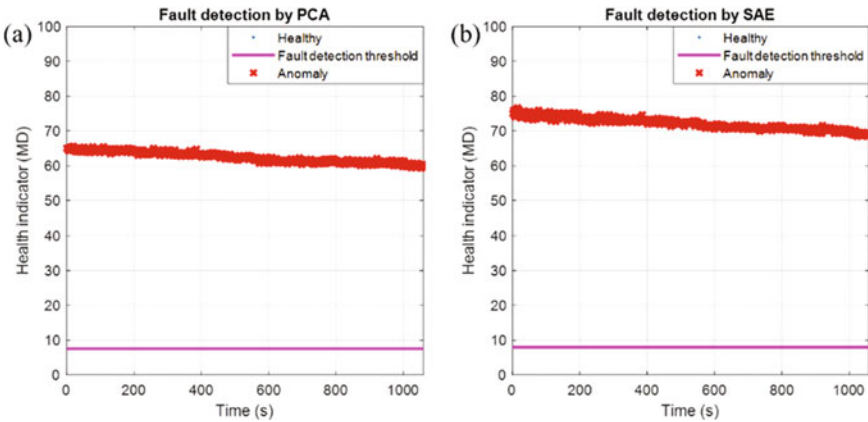


Fig. 4 Fault detection for cylinder 1 misfire fault by two unsupervised machine learning methods, **a** PCA, **b** SAE. (Blue dots: data points which were assumed as healthy in the training data, magenta line: fault detection threshold with confidence interval set at 99%, red points: anomalies in training data)

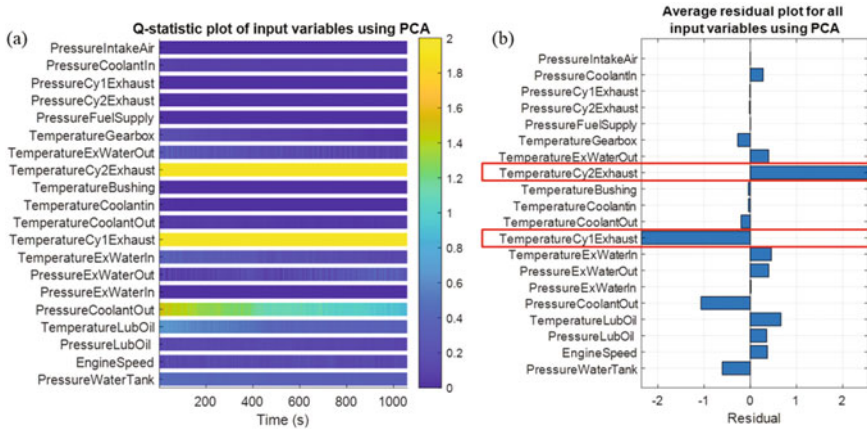


Fig. 5 Fault analysis for cylinder 1 misfire fault using PCA-based fault detection method **a** Q-statistic plot of all variables, **b** Average residuals

much higher than the defined thresholds, therefore, the misfire fault was successfully detected by both methods.

To have a clear understanding on which variables that contribute the most to the detection of the fault, the Q-statistic [10] is calculated in time series. The time series Q-statistic plot of all variables using PCA-based fault detection method is demonstrated in Fig. 5a. It clearly shows that, it is the exhaust cylinder 1 and 2’s temperatures that contribute the most to the misfire fault.

Furthermore, the average residuals for all input variables using PCA-based fault detection method can be found in Fig. 5b. As can be seen, under the misfire fault condition, the reconstructed exhaust cylinder 1’s temperature is lower than normal, while the reconstructed exhaust cylinder 2’s temperature is higher than normal. In theory, the cylinder 1’s temperatures should drop due to its misfire, and the cylinder 2’s temperatures should increase. The average residuals plot is as expected.

The time series Q-statistic plot of all variables using SAE-based fault detection method is demonstrated in Fig. 6a. As can be seen that, the top 2 variables that have the most contribution to the misfire fault, are the exhaust cylinder 1 and 2’s temperatures. Besides, the outlet pressure of the internal circulating cooling water in the engine and the temperature of the gearbox housing also show a high contribution to the fault.

The average residuals for all input variables using SAE-based fault detection method can be found in Fig. 6b. It shows that, under the cylinder 1 misfire condition, the cylinder 1’s exhaust temperature is lower than normal, while the cylinder 2’s exhaust temperature is higher than normal. This result is consistent with the failure mechanism of cylinder 1 misfire.

In this case, both the PCA and SAE based methods can detect and isolate the misfire fault.

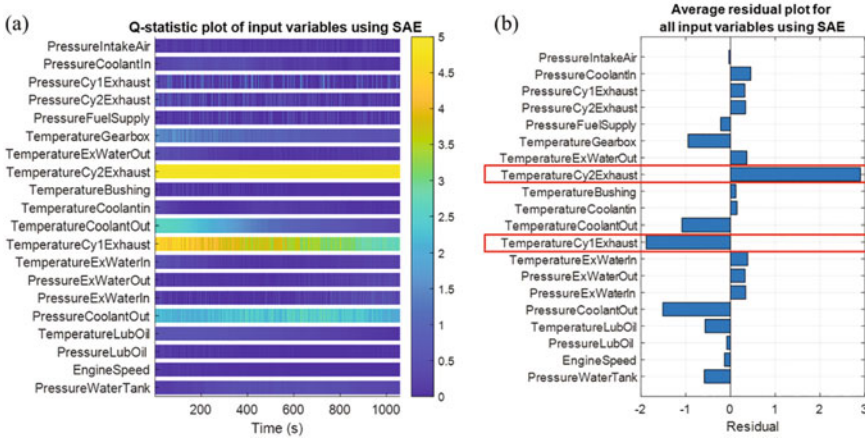


Fig. 6 Fault analysis for cylinder 1 misfire fault using SAE-based fault detection method **a** Q-statistic plot of all variables **b** Average residuals

5.2 Lubrication System Filter Blocking Detection and Diagnosis

In the diesel engine lubrication system, the function of the filter is to filter out the impurities in the lubricating oil, to keep the oil clean and therefore extend the service life of the diesel engine. Filter blocking can cause insufficient lubrication oil supply, and increase the failure rate of the diesel engine.

In this case study, the PCA and SAE were applied to detect the lubrication system filter blocking issue. The same training data and parameters were applied as in the case of detecting the misfire. Therefore, the results of the training process using PCA and SAE can be found in Fig. 3.

The fault detection process of detecting the diesel engine lubrication system filter blocking can be found in Fig. 7. It clearly shows that, this type of fault can be effectively detected as well, because the values of the health indicators (in MDs) are much higher than the defined thresholds.

The time series Q-statistic map and average residual plot of all variables using PCA-based fault detection method are demonstrated in Fig. 8a, b, respectively. As can be seen that, instead of measurements related to lubrication oil, the abnormality in shaft speed has the main contribution to the filter blocking fault. Therefore, the Q-statistic map and average residual plot are not as expected.

As a contrast, the Q-statistic plot of all variables using SAE-based fault detection method is shown in Fig. 9a, and the residual plot can be found in Fig. 9b. Both figures clearly show that it is the decreasing of the lubrication oil pressure that mainly caused the fault.

In this case, the SAE out performed PCA method in the isolation of lubrication oil system filter blocking fault.

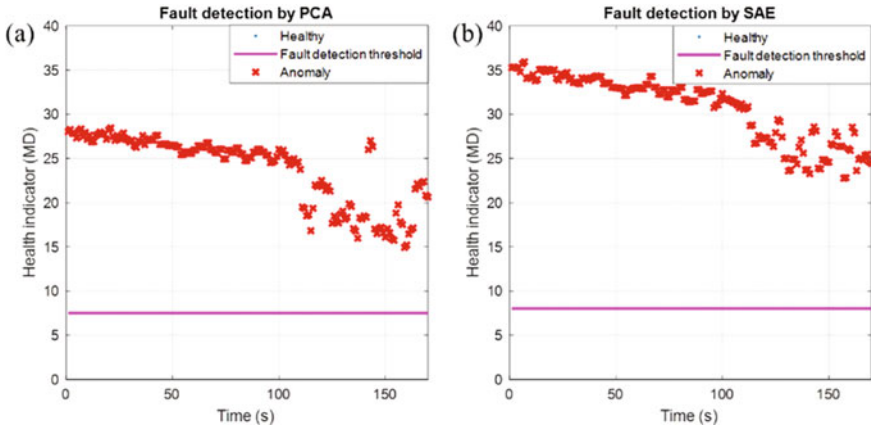


Fig. 7 Fault detection for lubrication system filter blocking by two unsupervised machine learning methods, **a** PCA, **b** SAE. (Blue dots: data points which were assumed as healthy in the training data, magenta line: fault detection threshold with confidence interval set at 99%, red points: anomalies in training data)

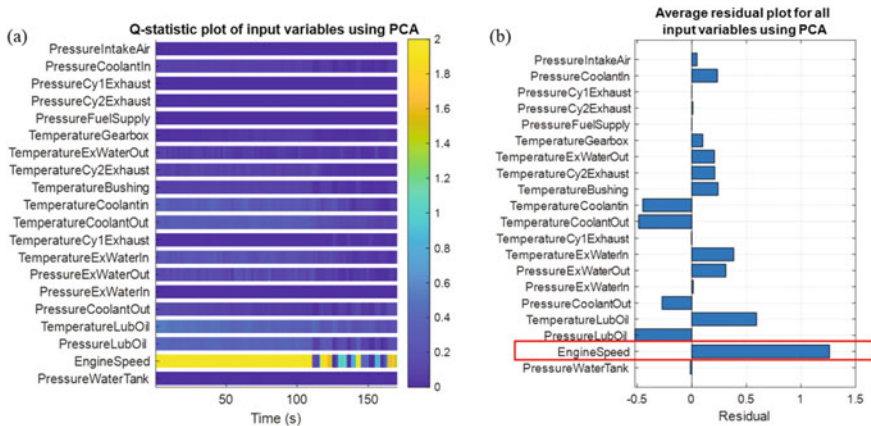


Fig. 8 Fault analysis for lubrication system filter blocking using PCA-based fault detection method **a** Q-statistic plot of all variables, **b** Average residuals

5.3 Discussion

As a comparison, the training data used in reference [10] is under similar operating condition to the test data. Different faults, including a pump system misalignment fault in the 1st case and a misalignment fault and a bearing fault in the 2nd case, were successfully detected via PCA and SAE methods. The contribution maps of SAE for the two cases were as expected. The performance of PCA was not as good as SAE, but some variables in the contribution maps were as expected.

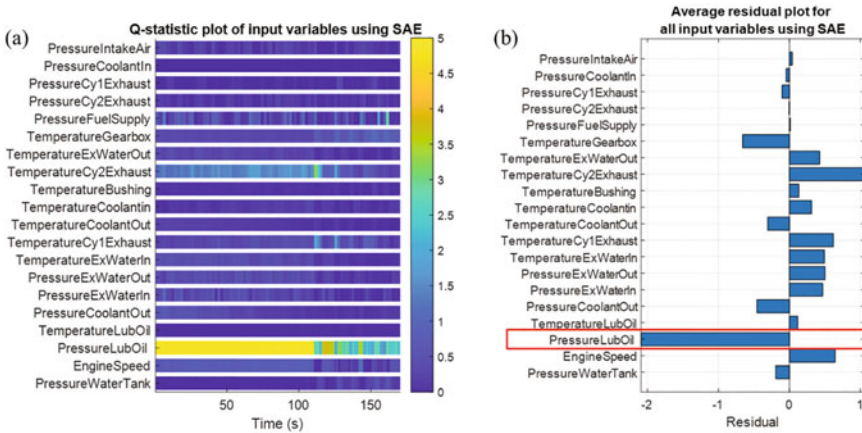


Fig. 9 Fault analysis for lubrication system filter blocking using SAE-based fault detection method **a** Q- statistic plot of all variables, **b** Average residuals

However, in this paper, the training data was collected under more than 13 operating conditions, which makes the training process for the unsupervised machine learning methods much more complicated than that in reference [10]. In this study, both the misfire fault and the lubrication system filter blocking fault are successfully detected by PCA and SAE, however, in the detection of the lubrication oil filter blocking fault, the Q statistic and residual figures of PCA are not as expected. In contrast, the contribution map and residual plot of SAE are as expected in both cases.

In industrial applications, the IC engine could have various types of faults. Many faults are difficult to distinguish, because some faults could have similar sensitive parameters and therefore have a similar contribution map.

Therefore, future work could be: (1) test the SAE fault detection method using more testing data with more types of faults; (2) apply additional parameters or features such as vibration features to achieve more accurate diagnostics.

6 Conclusions

In this paper, two unsupervised data-driven methods are employed for condition monitoring for an IC engine. Experimental results show that both of the unsupervised learning fault detection methods, PCA and SAE, can successfully detect the cylinder misfire fault and the lubrication system filter blocking fault. The Q statistic and residual plots of PCA in the detection of the lubrication system filter blocking fault are not consistent with the fault’s nature. In contrast, the SAE has better performance than PCA in both cases, with both the contribution maps and residual plots of SAE being as expected. In summary, the unsupervised machine learning methods, PCA

and SAE, are effective in fault detection, while the SAE has better performance than PCA in fault isolation.

References

1. Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., Aljaaf, A.J.: A systematic review on supervised and unsupervised machine learning algorithms for data science. In: Berry, M.W., Mohamed, A., Yap, B.W. (eds.) *Supervised and Unsupervised Learning for Data Science*, pp. 3–21. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-22475-2_1
2. Wang, J., Wang, Z., Stetsyuk, V., Ma, X., Gu, F., Li, W.: Exploiting Bayesian networks for fault isolation: a diagnostic case study of diesel fuel injection system. *ISA Trans.* **86**, 276–286 (2019). <https://doi.org/10.1016/j.isatra.2018.10.044>
3. Jafarian, K., Darjani, M., Honarkar, Z.: Vibration analysis for fault detection of automobile engine using PCA technique. In: 2016 4th International Conference on Control, Instrumentation, and Automation (ICCIA), pp. 372–376 (2016)
4. Mathew, S.K., Zhang, Y.: Acoustic-based engine fault diagnosis using WPT, PCA and Bayesian optimization. *Appl. Sci.* **10**(19), Art. no. 19 (2020). <https://doi.org/10.3390/app10196890>
5. Wang, X., Kruger, U., Irwin, G.W., McCullough, G., McDowell, N.: Nonlinear PCA with the local approach for diesel engine fault detection and diagnosis. *IEEE Trans. Control Syst. Technol.* **16**(1), 122–129 (2008). <https://doi.org/10.1109/TCST.2007.899744>
6. Géron, A.: *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O’Reilly Media, Inc. (2017)
7. Fu, X., Luo, H., Zhong, S., Lin, L.: Aircraft engine fault detection based on grouped convolutional denoising autoencoders. *Chin. J. Aeronaut.* **32**(2), 296–307 (2019). <https://doi.org/10.1016/j.cja.2018.12.011>
8. Reddy, K.K., Sarkar, S., Venugopalan, V., Giering, M.: Anomaly detection and fault disambiguation in large flight data: a multi-modal deep auto-encoder approach. In: *Annual Conference of the Prognostics and Health Management Society*, vol. 2016 (2016)
9. Tagawa, T., Tadokoro, Y., Yairi, T.: Structured denoising autoencoder for fault detection and analysis. In: *Asian Conference on Machine Learning*, pp. 96–111 (2015)
10. Liang, X., Duan, F., Bennett, I., Mba, D.: A sparse autoencoder-based unsupervised scheme for pump fault detection and isolation. *Appl. Sci.* **10**(19), 6789 (2020)
11. Jolliffe, I.: *Principal component analysis*. Springer (2011)
12. Sharma, A., Sugumaran, V., Devasenapati, S.B.: Misfire detection in an IC engine using vibration signal and decision tree algorithms. *Measurement* **50**, 370–380 (2014)

Online Pipe Leakage Detection Using the Vibration-Based Wireless Sensing System



Xiaoli Tang, Yu Jia, Guojin Feng, Yuandong Xu, Fengshou Gu,
and Andrew D. Ball

Abstract Piping systems are widely utilized in industry and home. Leakage of piping systems induced by prolonged corrosion, severe weather, or man-made damage will lead to serious consequences like explosion disasters, severe damage to industrial equipment, unforeseeable waste of resources and even threaten human life. WSNs significantly attract attentions in Industry 4.0 in recent years due to their advantages of wide distribution, remote controllability, convenient portability, easy programming, and economy. Meanwhile, as a non-intrusive measurement technique, vibration manifests a great potential for leakage detection of piping systems. In this paper, a vibration-based wireless sensing system is developed to remotely monitor the condition of piping systems in real time. According to the analytical results of vibration signals at two different positions on the piping system, the effective statistical features are extracted at the wireless sensor node to detect the leakage and its severity of the piping system. Furthermore, it can reduce the amount of data transmitted to reduce the power consumption then prolong the service life of the designed wireless sensing system. The diagnostic result can be conveniently observed on the mobile device in real time.

Keywords Piping system · Leakage detection · Wireless vibration sensing system · Condition monitoring

X. Tang · Y. Jia (✉)

School of Engineering and Technology, Aston University, Birmingham B4 7ET, UK

e-mail: y.jia1@aston.ac.uk

X. Tang

e-mail: x.tang4@aston.ac.uk

G. Feng · F. Gu (✉) · A. D. Ball

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield H1 3DH, UK

e-mail: f.gu@hud.ac.uk

G. Feng

e-mail: Guojin.Feng@outlook.com

Y. Xu

Department of Mechanical Engineering, Imperial College London, London SW7 2AZ, UK

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_39

1 Introduction

The leakage of the piping system induced by corrosion, weather, or man-made damage will bring potential safety hazard to human life in the industry and daily life, and it will also cause waste of resources and pollution of the environment. For example, 62 people were killed, and 136 people were injured after a leaking oil pipeline caught fire and explosion in Qingdao in November 2013, which also results in the economic loss of up to 80 million pounds. The traditional pipeline leakage monitoring technologies [1, 2] are primarily according to ultrasonic, acoustic emission, optic and fiber optic [3], flow and pressure detection [4], and soil monitoring. The equipment used for these technologies may be expensive or require intrusion and damage the pipeline. Vibration measurement is a non-intrusive detection method, which shows great advantages in detecting leakage in the piping system, especially the exposed pipes above the ground applied in industry or our daily life.

The piping system has complicated vibration sources due to its complex structures with various valves, corners, clamps, etc. at different locations. From the microscopic view, when the fluid passes through the pipeline, fluid molecules randomly collide against the pipe wall with most of them moving towards the same direction with the fluid flow. In this process, most of kinetic energy is converted into pressure oscillation energy which will provide dynamic forces from the inner wall of the pipe [5]. Simultaneously, little kinetic energy will be converted into heat energy and dissipates in the environment because of thermal effects. According to the kinetic analysis, the faster the fluid flows, the greater the force of the fluid molecules hitting the pipe wall, which results in a greater deformation of the pipe wall, as well as larger vibration amplitude. From a macro perspective, the laminar can be destroyed by the small whirlpool and generate turbulence [6] when the flow rate suddenly increases because of disordered and irregular fluid motion at the corner or around the valve. Randomness of fluid movements is the fundamental characteristic of turbulence. As a result, the vibration of the pipeline caused by turbulence is also complex and random. Another cause of pipe wall vibration is the internal friction [7] generated by the relative motion of the fluid to the pipe wall. The frictional effects of fluid motion are the results of momentum transfer and cohesion function between the molecules. All these behaviors are the source of pipe wall vibration. As the fluid flow generates a wide range of excitation, it will cause larger vibration amplitudes at the natural frequencies [8] of the piping system. Therefore, the excitations around natural frequencies will lead to enlarged vibration amplitude to show more distinct features for the leakage detection of the piping system.

Some researchers have investigated lots of leakage detection methods of piping systems based on vibration measurements. For example, Qu et al. [9] proposed a SVM-based pipeline leakage detection method with the distributed vibration sensors to recognize and locate the leakage in real time. Dinardo et al. [10] investigated that the first harmonic amplitude of the vibration signal has the linear relationship with the flow rate, which is very useful to improve the measurement of fluid flow rates according to the vibration. Shukla and Piratla [11] proposed to exploit a deep

learning algorithm to process the vibration signals to fill the gap due to the lack of a comprehensive understanding of the structural dynamics of the piping system and the correlation of leak detection. Additionally, some researchers achieved the leakage detection with the low-cost and low-power WSNs [12–14]. However, most of researchers developed the WSNs for leakage detection of pipelines according to the pressure, temperature, strain, acoustic signals, etc. [15, 16]. In this paper, a smart wireless vibration sensing system, including a vibration-based wireless sensor node and remote real-time monitoring application, was designed and validated to remotely detect the leakage severity of the piping system. Meanwhile, an optimal installation location was selected via comparison and analysis of the statistical vibration features. Furthermore, the design employs data compression which can effectively detect the leakage and its severity level and reduce the number of transmitted data to prolong the service life of the hardware.

The rest of this paper is arranged as follows. Section 2 develops a wireless vibration sensing system with the low-cost MEMS components, which includes the wireless vibration sensing node and remote monitoring application design. Then, a water leakage test rig was set up with the domestic pipeline in Sect. 3. Different leakage severity levels were simulated with the open of the tap to validate the effectiveness and efficiency of the designed wireless vibration sensing system for online pipe leakage detection. In Sect. 4, vibration measurements captured with the designed wireless vibration sensing system are analyzed through feature extraction for the selection of the optimal sensor installation position and leakage detection. Furthermore, the selected sensor installation position is validated with the online and remote water leakage monitoring in real time. Finally, a conclusion is drawn according to the previous analysis and validation.

2 Design of Wireless Vibration Sensing System

2.1 Wireless Vibration Sensing Node Design

Compared with the traditional data acquisition system, wireless sensor nodes have the characteristics of tiny dimension, low cost, integrability and flexibility. They commonly consist of four units: sensing unit, processing unit, communication unit and power unit. Figure 1 shows the architecture of a wireless sensor node. The wireless sensor node can be powered by an external battery. Moreover, the battery can be supported by the energy harvested from the ambient resources to achieve self-powered WSNs [17]. The collected raw data or extracted features will be directly transmitted to the portable devices or the Cloud via various wireless communication methods like Bluetooth and WiFi.

In our research, the sensing unit should be a 3-axis MEMS accelerator for capturing the vibration response of the pipes. ADXL345 from Analog Devices company is selected because of its characteristics of high data rate (3200 Hz), high

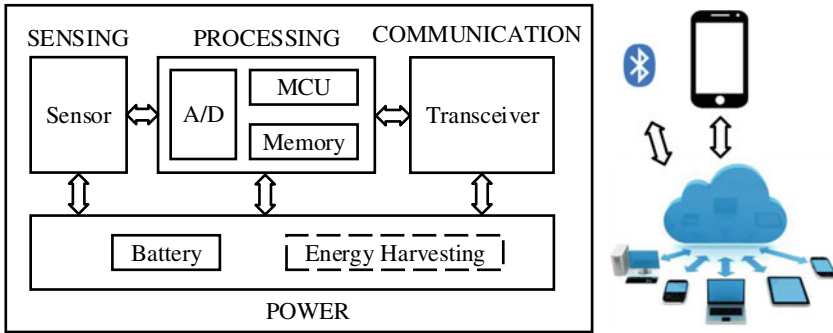


Fig. 1 Architecture of a wireless sensor node

resolution (13 bits), high sensitivity and wide range (± 16 g), as well as its low cost and low power cost.

According to the processing and communication units, the Adafruit Feather M0 Bluefruit LE board is very suitable for our research. It integrates ARM Cortex-M0 and the nRF51822 chipset from Nordic. This processing unit includes both the memory and the microcontroller to implement the primary functions of storage and computation. Additionally, the microcontroller is also responsible for communicating with other modules. Then, the collected vibration signal is transmitted to the designed mobile application through Bluetooth Low Energy (BLE4.2). The transmission speed can be up to 1 Mbps. To power these units, a charging module is integrated on the Adafruit Feather M0 Bluefruit LE board, which can recharge the Li-ion battery to extend the service life of the sensing node. As a result, a 3.7 V rechargeable Li-Ion battery is applied to provide all the power required by the sensing node. Moreover, the sensing node can be powered by the energy harvested from the ambient environment or monitored system [17].

Furthermore, to minimize the vibration loss in the process of transmission, the selected material of the sensor base is same as the object to be installed. Since the accelerometer was applied for the leakage condition of the standard copper water pipes, the sensor base was designed with copper which was tightly attached to the surface of the pipe by the clips within silicon between them to increase the contact area. The base thickness is only 2 mm to reduce the vibration damping. The whole vibration sensing node was covered with a white box as shown in Fig. 2.

2.2 Monitoring Application Design

To achieve remote and real-time condition monitoring with the designed wireless vibration sensing node, an Android mobile application named “Leak Detector” was designed with Android Studio as displayed in Fig. 3.

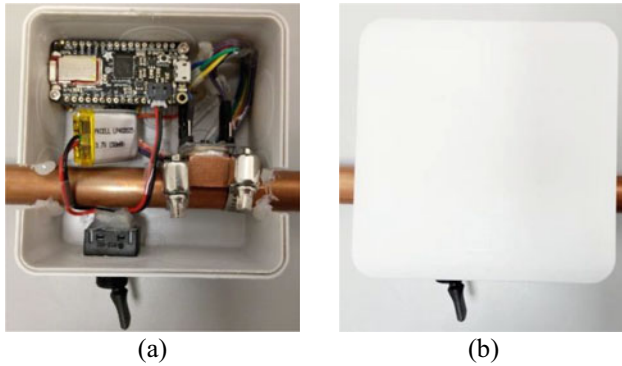


Fig. 2 Wireless vibration sensing node: **a** inside, and **b** outside

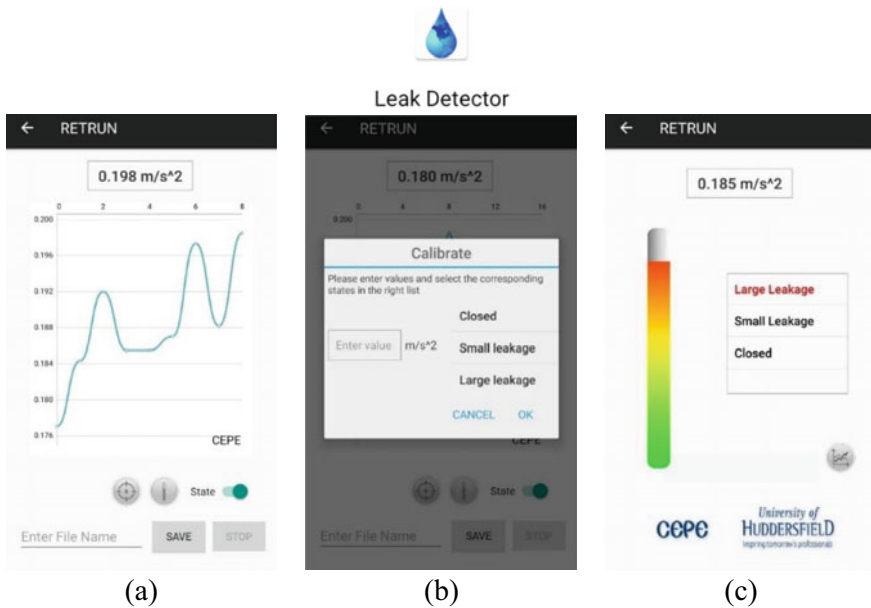





Fig. 3 Designed mobile application interfaces: **a** line chart display interface, **b** calibration interface, and **c** process bar display interface

The wireless vibration sensing node can be discovered by the mobile Bluetooth once the node is powered. After successful connection, the real-time vibration signal is transmitted and drawn with circles and blue lines as shown in Fig. 3a. At the same time, the real-time RMS value of vibration signals is displayed above the line chart. Additionally, there is a saving function to store the collected datasets in the mobile internal memory for further analysis. For accurate condition monitoring, the switch

button  can change the display interface from the line chart to the calibration panel, i.e., from Fig. 3a–b. The calibration panel can set thresholds of small, medium, and large water leakage. Another switch button  can convert the line chart to the process bar which aims to differentiate the level of the wake leakage amount according to the thresholds. Similarly, the switch button  can switch the interface from the line chart back to the process bar chart. With this mobile application, the condition of water pipes can be remotely monitored in real time.

3 Leakage Detection of Domestic Water Piping Systems

To validate the availability of the designed wireless vibration sensing system, some experiments were carried out on the domestic water piping systems with the copper pipe diameter of 15 mm in the laboratory. The pipes have a complicated structure at the corners and valves, and it is easier to generate turbulences leading to large vibration amplitudes of pipe walls. In addition, long straight pipes are more prone to resonance. As a result, two detection positions, P1 (closing to Tap 1 at the corner) and P2 (at the middle of the long straight pipe and with the distance of 3.6 m to Tap 2) were typically selected to compare the influence of the vibration mechanism and select the optimal position for remote online water leakage detection. The designed wireless vibration sensing node was installed on the surface of the pipe wall at these two locations. The different sizes of leaks were simulated by independently opening Tap 1 or Tap 2 with different angles, respectively. The structure of the domestic piping test rig is displayed in Fig. 4.

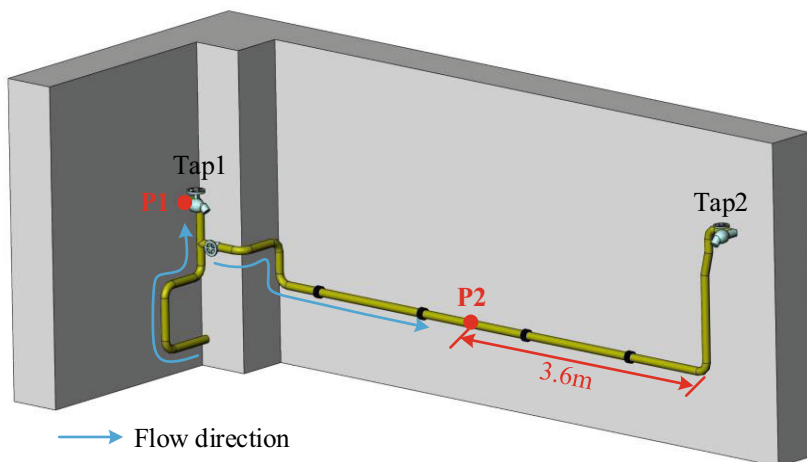


Fig. 4 Structure of the domestic piping test rig

Table 1 Flowrates at different leakage conditions

Conditions	Flow rate (L/min)
Zero flow	0
Small flow	0.51, 0.83, 0.94
Medium flow	1.07, 1.46, 1.57
Large flow	2.14, 2.68, 3.23, 4.63, 5.87

Table 2 Parameters of the wireless vibration sensing node

Parameters	Values
Range	±16 g
Sensitivity	256 LSB/g
Resolution	13 bits
Data rate	3200 Hz
Bandwidth	0.1–1600 Hz
Collection period	20 s

To diagnose the severity of the leak, four different leak conditions, including zero, small, medium and large flows, were simulated with the Taps 1 and 2 closed or opened with different rotating angles in this experiment. The flow rates caused by leakage were measured using a flow transducer and presented in Table 1.

To acquire a broad range of the vibration frequencies, the maximum sampling rate of 3200 Hz was set for the wireless vibration sensing node. Three repetitive tests for each sizes of leaks at Tap 1 and Tap 2 were carried out with 20 s of data recorded for further analysis. The other parameters are listed in Table 2 in detail.

4 Results and Discussion

In this section, 17 tests with different flow rates were carried out to simulate four different conditions, that is zero leakage, small leakage, medium leakage and large leakage, respectively. The experimental results at P1 and P2 will be compared and discussed to select the optimal position of the piping system to install the wireless vibration sensing node.

4.1 Results at Position 1 (P1)

P1 is the position close to the tap and the pipe corner. To eliminate the influence of ambient noise, a high-pass filter with 100 Hz as the low threshold was utilized in vibration waveforms. Figure 5 illustrates the three-axis RMS values of vibration

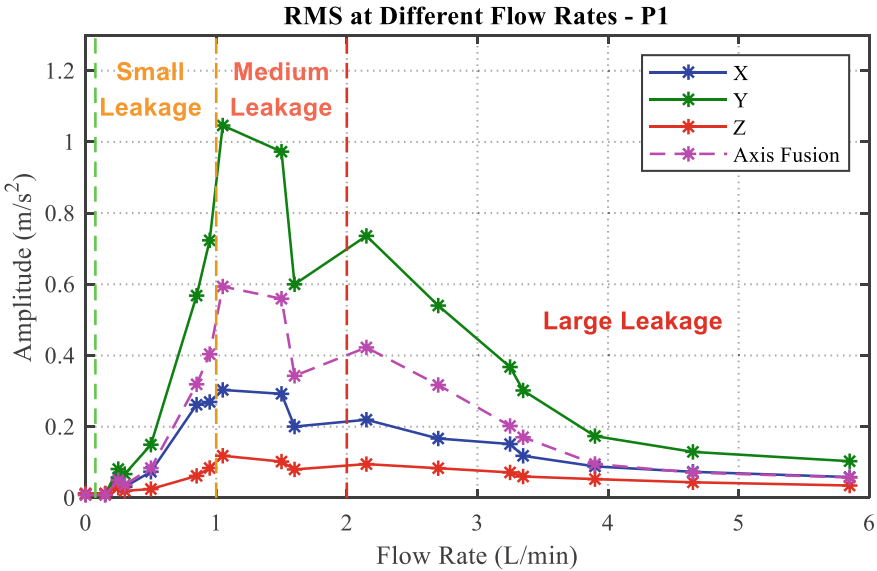


Fig. 5 RMS values of vibration at different flow rates at P1

waveforms at different flow rates (leaks) measured at P1. Additionally, a three-axis acceleration fused RMS values are calculated and plotted for different flow rates (leaks) with the magenta dotted line. From the perspective of change, it is clear that RMS value of zero leakage is the smallest. It increases as the rise of the leakage amount in the small leakage range, then goes to a significantly high value of 1.046 m/s² when the leakage is heavier to the medium level. The increasing trend of acceleration is caused by the resonance of the piping system widely excited by the more apparent force of water molecules acting on the pipe wall. However, it starts to descend rapidly after reaching its peak, no matter how severe the leak is, which manifests that vibrational RMS values have nonlinear correlations with the severity of the leak. This phenomenon may be because as the water amount increases and the water level rises in the water pipe, the turbulence and friction effects gradually calm down.

Furthermore, RMS values of 0.4 m/s² and 0.6 m/s² can be considered as the thresholds to distinguish between small and medium leakage, and medium and large leakage, respectively. Unfortunately, it is impossible to recognize the small and large leakage with these two thresholds. Moreover, considering the complexity of the piping system structure at P1, the vibration caused by turbulence will be unstable. Meanwhile, the installation of the wireless vibration sensing node is inconvenient and complicated at the corner and tap position. Consequently, P1 is not very suitable for leakage detection with the vibration-based wireless sensor node.

4.2 Results at Position 2 (P2)

P2 locates in the middle of the long straight pipes. The tests with same flow rates (leaks) as P1 were carried out and measured at P2. Figure 6 also displays the RMS values of the high-pass (100 Hz as the low threshold) filtered vibration at 17 different flow rates (leaks) at P2. The RMS values of vibration in X and Y axes dramatically increase approximately with the rising of the flow rates when the leakage quantity is at the small level. But it only gently rises at the Z direction, which represents the force of water molecules in this direction is relatively small. When the flow rates continue to increase, the RMS values decrease but not as significant as the results at P1, which illustrates that friction effects may contribute more than turbulence. From the analytical results of these three axes, RMS values at the Y or X directions are more appropriate as the efficient indicator for condition monitoring of piping systems than the Z direction because of their high amplitudes. However, the installation of the wireless sensor node requires high requirement on the installation direction and position. Therefore, the RMS value of the acceleration fused all three axes is more suitable for pipe leakage detection and diagnosis. As a result, the acceleration RMS values fused all three axes at the long straight pipe can be considered as the position to monitor condition of the piping system. As the water pipe structure in different positions is completely different, it may require the customer to perform non-professional calibration by controlling the tap after the wireless sensor node is installed on the surface of the pipes.

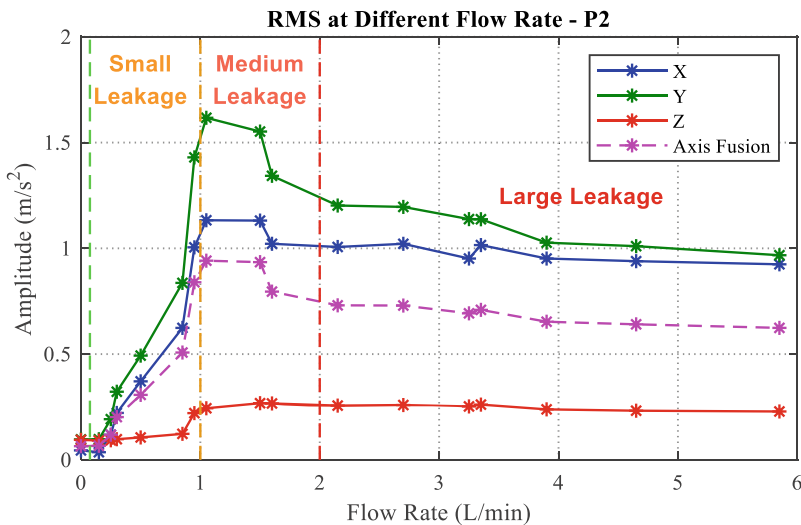


Fig. 6 RMS values of vibration at different flow rates at P2

4.3 Evaluation of the Designed Remote CM System

According to the experimental results analyzed in the previous sub-sections, the vibration-based wireless sensor node was installed on the pipe outer wall at a long straight pipe. The raw vibration signals were primarily captured and filtered by the sensor. Then, RMS values were calculated as the compressed and extracted statistical to transmit to the remote devices for further leakage diagnostic of the piping system in the real time. This pre-processing process can not only reduce the power consumption of the wireless sensor node, but also guarantee the effective transmission of characteristic signals [17].

An online field test was performed with the designed wireless sensor node located at about 6–7 m far from the leak position and the monitoring location. The test results can be remotely observed at the portable device in real time. Figures 7 and 8 display the monitoring interface with line graphs and bar process charts at four different leakage levels, respectively. In Fig. 7, the acceleration RMS values are about 0.194, 0.209, 0.307, and 1.902 m/s^2 for these four different cases, which can clearly distinguish whether there is a leak or not, and even can effectively measure the amount of the leak. Simultaneously, the condition of the pipe can be also intuitively observed with the bar process charts as shown in Fig. 8. When the leakage amount

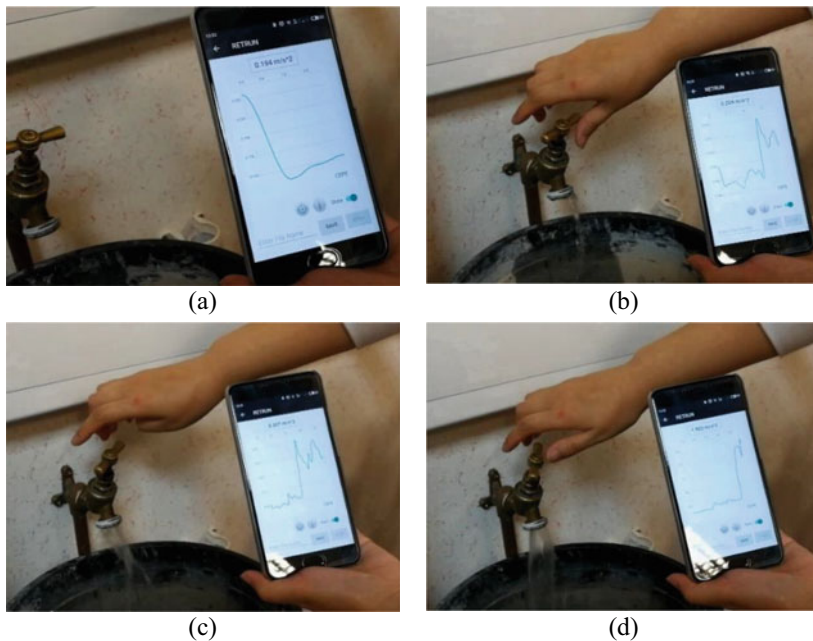


Fig. 7 Real-time leakage monitoring interface with line graphs: **a** no leakage, **b** small leakage, **c** medium leakage, and **d** large leakage

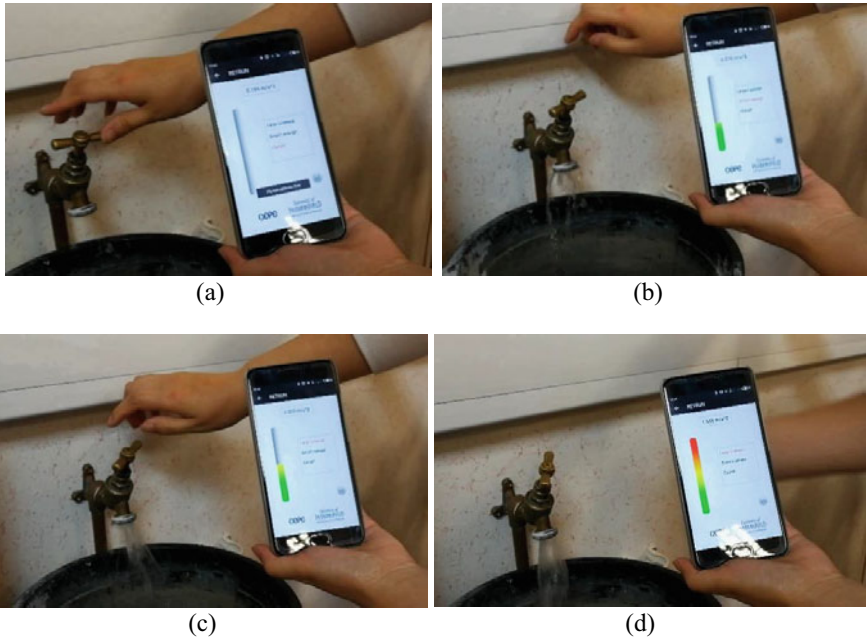


Fig. 8 Real-time leakage monitoring interface with bar process charts: **a** no leakage, **b** small leakage, **c** medium leakage, and **d** large leakage

is higher than 2 L/min, the vibration-based wireless sensing system will issue a red alert as displayed in Fig. 8d.

This real-time test demonstrates that the designed wireless vibration sensing system is effective and efficient for the leakage detection and severity diagnosis of piping systems with the statistical RMS values of the vibration of triaxial fusion.

5 Conclusions

A smart wireless sensing system was designed and validated to remotely detect the leakage severity of the piping system in real time according to its vibration mechanism in this paper. Two representative positions on the pipeline were selected to compare the influence of turbulence and friction effects of fluid. It seems that resonance of the pipe wall at the long straight pipe is more prone to the excitation. The diagnostic results can be explicitly observed with the portable device from both the line graphs and bar process charts. Moreover, the design of the wireless vibration sensing system relies on statistical features, which can not only effectively detect the leakage faults of piping systems, but also greatly reduce the amount of transmitted data. Therefore, it will reduce the power consumption of the node to prolong its service life, which

supplies the potential to achieve maintenance-free pipeline leakage detection system with the assist of energy harvesting in future.

References

1. Datta, S., Sarkar, S.: A review on different pipeline fault detection methods. *J. Loss Prev. Process Ind.* **41**, 97–106 (2016). <https://doi.org/10.1016/j.jlp.2016.03.010>
2. BenSaleh, M.S., Manzoor Qasim, S., Abid, M., Jmal, M.W., Karray, F., Obeid, A.M.: Towards realisation of wireless sensor network-based water pipeline monitoring systems: a comprehensive review of techniques and platforms. *IET Sci. Meas. Technol.* **10**(5), 420–426 (2016). <https://doi.org/10.1049/iet-smt.2015.0255>
3. Tejedor, J., Macias-Guarasa, J., Martins, H.F., Pastor-Graells, J., Corredera, P., Martin-Lopez, S.: Machine learning methods for pipeline surveillance systems based on distributed acoustic sensing: a review. *Appl. Sci.* **7**(8), 841 (2017)
4. Rashid, S., Akram, U., Khan, S.A.: WML: wireless sensor network based machine learning for leakage detection and size estimation. *Procedia Comput. Sci.* **63**, 171–176 (2015). <https://doi.org/10.1016/j.procs.2015.08.329>
5. Ismail, M., Dziyauddin, R.A., Salleh, N.A.A.: Performance evaluation of wireless accelerometer sensor for water pipeline leakage. In: 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS), pp. 120–125 (2015). <https://doi.org/10.1109/IRIS.2015.7451598>.
6. Ustinov, M.: Tollmien-Schlichting wave generation by flow turbulence. *Fluid Dyn.* **49**(4), 468–480 (2014)
7. Rudyak, V.Y., Belkin, A.: Fluid viscosity under confined conditions. *Dokl. Phys.* **59**, 604–606 (2014)
8. Wachel, J., Morton, S.J., Atkins, K.E.: *Piping Vibration Analysis*, pp. 119–134 (1990)
9. Qu, Z., Feng, H., Zeng, Z., Zhuge, J., Jin, S.: A SVM-based pipeline leakage detection and pre-warning system. *Measurement* **43**(4), 513–519 (2010). <https://doi.org/10.1016/j.measurement.2009.12.022>
10. Dinardo, G., Fabbiano, L., Vacca, G.: Fluid flow rate estimation using acceleration sensors. In: Seventh International Conference on Sensing Technology (ICST), pp. 221–225 (2013). <https://doi.org/10.1109/ICST.2013.6727646>
11. Shukla, H., Piratla, K.: Leakage detection in water pipelines using supervised classification of acceleration signals. *Autom. Constr.* **117**, 103256 (2020). <https://doi.org/10.1016/j.autcon.2020.103256>
12. Karray, F., Garcia-Ortiz, A., Jmal, M.W., Obeid, A.M., Abid, M.: EARNPIPE: a testbed for smart water pipeline monitoring using wireless sensor network. *Procedia Comput. Sci.* **96**, 285–294 (2016). <https://doi.org/10.1016/j.procs.2016.08.141>
13. Moulik, S., Majumdar, S., Pal, V., Thakran, Y.: Water leakage detection in hilly region PVC pipes using wireless sensors and machine learning. In: IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), pp. 1–2 (2020). <https://doi.org/10.1109/ICCE-Taiwan49838.2020.9258144>
14. Mohd Ismail, M.I., Dziyauddin, R.A., Mohd Salleh, N.A., Ahmad, R., Hadri Azmi, M., Mad Kaidi, H.: Analysis and procedures for water pipeline leakage using three-axis accelerometer sensors: ADXL335 and MMA7361. *IEEE Access* **6**, 71249–71261 (2018). <https://doi.org/10.1109/ACCESS.2018.2878862>
15. Liu, Y., Ma, X., Li, Y., Tie, Y., Zhang, Y., Gao, J.: Water pipeline leakage detection based on machine learning and wireless sensor networks'. *Sensors* **19**(23), 23 (2019). <https://doi.org/10.3390/s19235086>

16. Salameh, H., Dhainat, M., Benkhelifa, E.: A survey on wireless sensor network-based IoT designs for gas leakage detection and fire-fighting applications. *Jordanian J. Comput. Inf. Technol.* 1 (2019). <https://doi.org/10.5455/jjcit.71-1550235278>
17. Tang, X., Wang, X., Cattley, R., Gu, F., Ball, A.D.: Energy harvesting technologies for achieving self-powered wireless sensor networks in machine condition monitoring: a review. *Sensors* **18**(12), 4113 (2018). <https://doi.org/10.3390/s18124113>

Meta-Learning Guided Few-Shot Learning Method for Gearbox Fault Diagnosis Under Limited Data Conditions



Ming Zhang , Duo Wang , and Yuchun Xu 

Abstract Recently, intelligent fault diagnosis technology based on deep learning has been extensively researched and applied in large industrial equipment system for ensuring safe and stable production. However, these deep models only effective when enough data for each observed failure category are available in the training durations. Otherwise, the performance of these models will notably decrease. As the critical component in large machinery, the gearbox often changes the speed and load along with the production demand in the practical application, which caused few data samples to be collected at certain conditions. This phenomenon introduces the few-shot fault diagnosis, and its goal is to identify the fault types with extremely limited data samples. To address this problem, a Meta-learning guided Few-shot Fault Diagnosis method, named MFFD, is proposed for gearbox fault diagnosis under limited data conditions. The results verify the effectiveness of our MFFD method at one-shot and five-shot fault diagnosis tasks under different speed and load conditions.

Keywords Meta-learning · Few-shot learning · Fault diagnosis · Gearbox · Limited data samples

1 Introduction

As the critical technology of Prognostic and Health Management (PHM), fault diagnosis aims at identifying failure attribution accurately, thus decide on actions to prevent their occurrence in preventive maintenance. Since the machinery equipment is developed toward high-speed and complexity in recent years, meanwhile, it also needs to meet the needs of health, stable and long-term running, fault diagnosis has become an indispensable technology in most industrial complex systems. With

M. Zhang (✉) · Y. Xu
College of Engineering and Physical Sciences, Aston University, Birmingham B4 7ET, UK
e-mail: m.zhang21@aston.ac.uk

D. Wang
Department of Automation, Tsinghua University, Beijing 100084, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_40

491

the rapid development of information technology and sensor technology, the industry has become a data-rich environment, which provides opportunities for developing the data-driven fault diagnosis method to substitute time-consuming knowledge-based and rule-based methods, thus enhance efficiency and applicability of fault diagnosis. In recent years, the fault diagnosis methods based on deep learning have been well studied and applied in large industrial system [1–6].

However, the effectiveness of the advanced models can be guaranteed only existing enough data samples into the model during the training period. The capability of these deep models may be degraded dramatically when this assumption can not be satisfied. Unfortunately, in the real-world application, the gearbox needs to switch speed and load frequently according to the production demand. Usually, the machine would not run at high speed and load unless there is an emergency need, and it could not work with worrying fault. For this reason, the data coming from the heavy condition and sudden serious failure should be hard to collect, while the normal data are sufficient. This phenomenon is referred to as a few-shot fault diagnosis, in which training a model can generalize well only used one or a few data samples.

Recently, some researchers in the fault diagnosis field have paid attention to the few-shot fault diagnosis problem and proposed advanced methods to overcome the overfitting problem by training with limited data samples [7–9]. Although these methods are effective for the few-shot fault diagnosis tasks to a certain degree, they did not utilize the feature information of related diagnosis tasks as support for helping the current few-shot learning task. We try to deal with the few-shot fault diagnosis problem in the meta-learning perspective, which could with the help of the transferable knowledge learning from the source domain to address the few-shot diagnosis issue in the target domain.

In this paper, a Meta-learning guided Few-shot Fault Diagnosis framework, called MFFD, is presented for the gearbox fault diagnosis under different limited data conditions. The preliminaries is detailed in Sect. 2, and the theoretical framework is described in Sect. 3. Section 4 shows detailed experiments and analysis results. The conclusions has been made in Sect. 5.

2 Preliminaries

2.1 Problem Definition

Figure 1c shows the few-shot fault diagnosis, the training and test data draw from the different domain, the target domain is the few-shot fault diagnosis task which has limited labeled samples, while the source domain owns sufficient data which is annotated but totally different with the target domain. The meta-learning method is to exploit the transferable knowledge by utilizing the episodic training mechanism for facilitating the model in a few-shot condition [10–12]. The target domain of few-shot fault diagnosis is defined as \mathcal{T}^T which has a C^T -way, K -shot, M -test task. In this

task, there are an annotated support data set $\mathcal{S}^T : (x_{\mathcal{S}a}^T, y_{\mathcal{S}a}^T)_{a=1}^{N_{\mathcal{S}}^T}$ and an unannotated query data set $\mathcal{Q}^T : (x_{\mathcal{Q}a}^T)_{a=1}^{N_{\mathcal{Q}}^T}$. In our work, K is equal to 1 or 5, while M set to 25, x is the vibration signal sample and y is the label. Samples of auxiliary set in source domain are defined as $X^S = (x_a^S, y_a^S)_{a=1}^{N^S}$. The source domain has totally different failure categories C^S .

2.2 Metric-Based Meta-Learning

The metric-based meta-learning aims at addressing the few-shot learning problem with the set-to-set approach [11, 13], which can predict the labels for the unobservable categories without making any changes to the trained model. Specifically, the metric-based model is a probability distribution mapping $P(\hat{y}|\hat{x}, S)$ from the input samples \hat{x} to the output labels \hat{y} , while the $S = \{(x_i, y_i)\}_{i=1}^k$ is a support set of K labeled samples. When a new few-shot support set S' coming, the model will be directly use to output the label \hat{y} for each test sample \hat{x} : $P(\hat{y}|\hat{x}, S')$. The metric-based meta-learning model is defined in general as follows:

$$\hat{y} = \sum_{i=1}^k \alpha(\hat{x}, x_i) y_i, \tag{1}$$

where x_i, y_i draw from the support set $S = \{(x_i, y_i)\}_{i=1}^k$, and α is the metric-based attention kernel. Unlike the typical supervise learning model, the metric-based meta-learning is a non-parametric method and quickly adapts to any new support set.

Matching Networks As the initial method of metric-based meta-learning, Matching Networks [10] define a conditional classification model with the support set, which explicitly chooses the softmax over the cosine distance as the similarity function α for the few-shot test sample \hat{x} and the support sample x_i . The similarity kernel function is expressed as follows:

$$\alpha(\hat{x}, x_i; \theta) = \frac{\exp[d(f(\hat{x}), f(x_i))]}{\sum_{j=1}^k \exp[d(f(\hat{x}), f(x_j))]} \tag{2}$$

where the f is an appropriate neural network with parameter θ as the embedding function; d is a cosine distance function.

Prototypical Networks Prototypical Network [14] performance better than Matching Networks for solving the few-shot problem in the image classification field, which benefit from the idea that there should exist a single prototype representation for each class in the points cluster. The prototype representation is the mean vector of its belonging category, which can be calculated as follows:

$$c_k = \frac{1}{|S_k|} \sum_{x_j, y_i \in S_k} f(x_i), \quad (3)$$

where S_k denotes the set of examples labeled with class k in the given supporting set. With the changes in the Prototypical Networks, the similarity kernel will turn to

$$\alpha(\hat{x}, c_k; \theta) = \frac{\exp[d(f(\hat{x}), c_k)]}{\sum_{j=1}^k \exp[d(f(\hat{x}), c_k)]}, \quad (4)$$

where d denotes the Euclidean distance in such a method. Similar to the Matching Networks, the training process also aims at minimizing the negative log-probability $J(\theta) = -\log(P(y|x, S; \theta))$ through stochastic gradient descent (SGD).

3 Theoretical Framework

The framework of our proposed MFFD method is shown in Fig. 1a, which consists of three-part, including few-shot tasks sampling, embedding mapping, and matching classification. The data samples of the source domain are reformed to generate few-shot tasks for episode training. Then the support set S_i and the query set Q_i are input to embedding mapping, while the embedding mapping is optimized by the metric loss function L_m , which is defined in Eq. (9). After the training procedure, the embedding mapping is directly used to identify the test few-shot tasks from the unseen target domain.

3.1 Embedding Mapping

For the fault diagnosis tasks, the vibration wave is the main signal to be processed, which is a typical one-dimensional signal. The embedding mapping is the critical term of our proposed method, which consist of several one-dimensional convolution layers and a fully connected layer. More specifically, the one-dimensional convolutional layer is defined as follows:

$$C_{ij}^l = \phi(k_{n \times 1}^j * x_{i:i+n}^i + b_{ij}) \quad (5)$$

where $k_{n \times 1}^j$ is the j th kernel which belong to the kernels K_j^l size $n \times 1 \times j$ of the l -th convolution layer; $x_{i:i+n}^i$ is the i th input segment; b_{ij} is corresponding to the bias; ϕ is the activation function; C_{ij}^l is the i -th feature point of the j -th kernel in the l -th convolution layer.

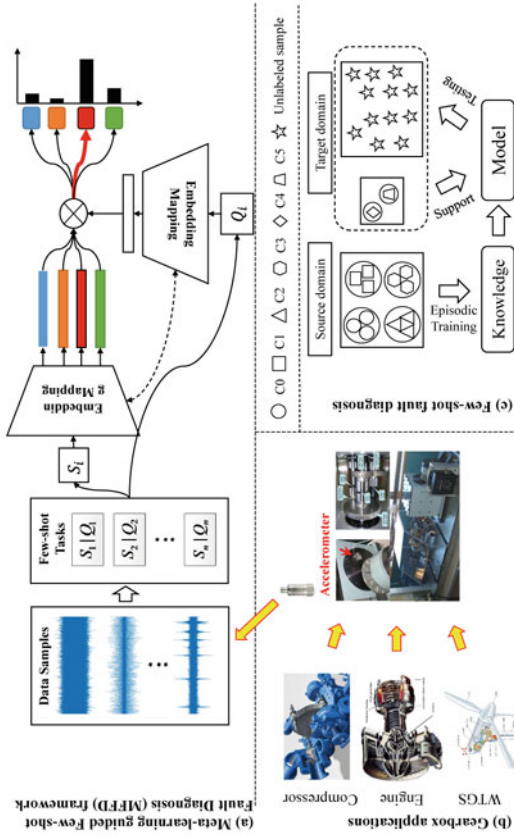


Fig. 1 a Meta-learning guided few-shot fault diagnosis (MIFFD) framework; b gearbox applications; c few-shot fault diagnosis

3.2 Episodic Learning Procedure

The metric embedding mapping is defined as $f_M(\cdot)$ with parameters θ_M . For implementing the episodic training, the data sampled of source domain will be randomly sampled to generate a range for few-shot learning tasks of fault diagnosis \mathcal{T}^S , in each of which there has C^S -way, K -shot, M -test and C^S don't need to equal C^T of \mathcal{T}^T in the target domain. The support set and the query set of source domain are defined as $\mathcal{S}^S = (x_{\mathcal{S}a}^S, y_{\mathcal{S}a}^S)_{a=1}^{N_{\mathcal{S}}^S}$ and $\mathcal{Q}^S = (x_{\mathcal{Q}a}^S, y_{\mathcal{Q}a}^S)_{a=1}^{N_{\mathcal{Q}}^S}$, respectively, where $N_{\mathcal{S}}^S$ equals to $C^T \times K$ and $N_{\mathcal{Q}}^S$ equals to M .

The core of our MFFD is to learn the transferable feature of embedding mapping from the source domain and directly use it to solve the few-shot fault diagnosis in the target domain. Therefore, after processing the data samples of the source domain, the classification performance of the query set by matching the feature of embedding mapping to the support set should be conducted. The one-dimensional raw fault signal x as the input for the embedding mapping, the prediction is calculated by the weighted sum of support labels, which is defined as follows (note that the superscripts S are omitted for simplicity):

$$\hat{y}_{\mathcal{Q}n} = \sum_{a=1}^{N_{\mathcal{S}}} w[f_M(x_{\mathcal{Q}n}), f_M(x_{\mathcal{S}a})] \cdot y_{\mathcal{S}a} \quad (6)$$

where the w is the softmax normalization of distance, which is calculated by the distance between the feature of the query set and each support set:

$$w[f_M(b_{\mathcal{Q}n}), f_M(b_{\mathcal{S}a})] = \frac{\exp(-\tau * d[f_M(b_{\mathcal{Q}n}), f_M(b_{\mathcal{S}a})])}{\sum_{j=1}^{N_{\mathcal{S}}} \exp(-\tau * d[f_M(b_{\mathcal{Q}n}), f_M(b_{\mathcal{S}j})])} \quad (7)$$

where τ is a hyper-parameter to control the convergence of training. The kernel function $d[\cdot]$ is defined as:

$$d[x_i, x_j] = \frac{x_i \cdot x_j^T}{|x_i| |x_j|} \quad (8)$$

Finally, the objective for optimizing metric embedding mapping with parameter θ_M is defined as follows:

$$L_M(\theta_M) = \sum_{\mathcal{T}^S} L_M(\mathcal{T}^S; \theta_M) = \sum_{\mathcal{T}^S} \left[-\frac{1}{N_{\mathcal{Q}}^S} \sum_{n=1}^{N_{\mathcal{Q}}^S} \sum_{i=1}^{C^S} \mathbb{I}[y_{\mathcal{Q}n} = i] \cdot \log(\hat{y}_{\mathcal{Q}n}(i)) \right] \quad (9)$$

The above learning process is described following the Matching Network (MN). As the alternative way of Matching Network, Prototypical Network (PN) in our MFFD framework is defined as follows:

$$\hat{y}_{\mathcal{Q}_n} = \sum_{a=1}^{C^S} w[f_M(b_{\mathcal{Q}_n}), P_a^S] \cdot y_a \quad (10)$$

where P_a^S is the mean vector of embedding features of the a th category:

$$P_a^S = \frac{1}{K} \sum_{j \in a} f_M(b_{S_j}) \quad (11)$$

The calculation of weights and distance and training objective are similar to Eqs. (7), (8) and (9). In our MFFD, MN and PN versions are compared and analyzed.

3.3 The Workflow

In this section, The workflow of training strategy for our method is detailed as follows:

1. Data Processing: Reform the data samples of source domain X^S as the few-shot learning tasks $\{\mathcal{S}_i^S, \mathcal{Q}_i^S\}_{i=1}^n$;
2. Initialization: Initialize parameters of embedding mapping θ_M ;
3. Training: Input the few-shot tasks of source domain $\{\mathcal{S}_i^S, \mathcal{Q}_i^S\}_{i=1}^n$ to train the embedding mapping with the objective L_M , which can be expressed as:

$$\theta_M \leftarrow \theta_M - \alpha L_M; \quad (12)$$

4. Test: Input the query set \mathcal{Q}^T of the target domain to embedding mapping with trained parameters θ_M and limited labeled support set S^T .

4 Experiment

4.1 Data Preparation

A gearbox fault diagnosis dataset [15, 16] is utilized to testify our proposed method. The test rig is shown in Fig. 1b, an accelerometer is installed on the gearbox body closing to the drive shaft for collecting the 1D vibration signal. As shown in Table 1, there are six categories of gear failure: Normal, Chipped Tooth (CT), and Missing Tooth (MT), Bent Shaft (BS), Rotor Imbalance (RI), and Mixed fault with BS and MT. Each type has 500 samples of different speeds (30, 35, 40, 45 and 50 Hz) and different loads (Low and High). The sampling frequency is 66.67 kHz and every input data has 6600 points.

Table 1 Description of gearbox dataset

Fault location	Normal	CT	MT	BS	RI	Mix	Speed (Hz)	Load
Category labels	0	1	2	3	4	5		
Dataset 30L no.	500	500	500	500	500	500	30	Low
Dataset 30H no.	500	500	500	500	500	500	30	High
Dataset 35L no.	500	500	500	500	500	500	35	Low
Dataset 35H no.	500	500	500	500	500	500	35	High
Dataset 40L no.	500	500	500	500	500	500	40	Low
Dataset 40H no.	500	500	500	500	500	500	40	High
Dataset 45L no.	500	500	500	500	500	500	45	Low
Dataset 45H no.	500	500	500	500	500	500	45	High
Dataset 50L no.	500	500	500	500	500	500	50	Low
Dataset 50H no.	500	500	500	500	500	500	50	High

4.2 Experimental Setup

Few-Shot Setup for Fault Diagnosis In the few-shot fault diagnosis experiments, the source domain has enough labeled dataset $D^s = \{x_i^s, y_i^s\}_{i=1}^{n^s}$, and the target domain is part of Support dataset $S^t = \{x_k^t, y_k^t\}_{k=1}^K$ and the test task dataset $D^t = \{x_j^t\}_{j=1}^{n^t}$. It is for sure that the fault types of the supporting dataset would not be seen in the source domain, and the task dataset is unlabeled. One-shot means only 1 sample of each category in the support set, while five-shot denote 5 samples in the support set.

In this work, we mainly evaluate our model in the following two scenarios: (1) The source and target domain draw from the different working conditions for 1-shot and 5-shot learning fault diagnosis situations; (2) The source and target domain from the different categories under the same working condition for 1-shot and 5-shot learning fault diagnosis situations.

Compared Methods All the few-shot tasks is compared with several few-shot learning methods for comparing with our proposed method, which is detailed as follow:

1. Finetune Last
2. Fintune Whole
3. Feature KNN
4. Feature KNN Proto
5. MFFD-MN
6. MFFD-PN

(1)–(3) are based on pre-training the embedding mapping and the last classifier with source domain data with supervised learning. (1) and (2) are designed to finetune the classifier or the whole model by utilizing the few-shot data from the target domain;

Table 2 Details of network architecture

Components	Layer type	Kernel	Stride	Channels	Padding
Embedding mapping	Convolution 1 ReLU 1	64×1	16×1	16	No
	Max Pooling 1	2×1	2×1	16	No
	Convolution 2 ReLU 2	3×1	1×1	32	Yes
	Max Pooling 2	2×1	2×1	32	No
	Convolution 3 ReLU 3	3×1	1×1	64	Yes
	Max Pooling 3	2×1	2×1	64	No
	Convolution 4 ReLU 4	3×1	1×1	64	Yes
	Max Pooling 4	2×1	2×1	64	No
	Convolution 5 ReLU 5	3×1	1×1	64	Yes

(3) and (4) compare the abstracted features with those of support samples by few-shot data of target domain; (5) and (6) are the Matching Network and Prototypical Network model with embedding mapping as the backbone in raw data space.

Implementation Details The detail of network architecture is shown in Table 2. Adam optimization is used to train, and all models are trained for the labeled source domain is close 100% before used for the target domain. The number of iterations is 100 and the learning rate is 0.001 in Finetune Last and Whole models. Cosine distance is utilized in the KNN model. In meta-learning methods, the number of query samples is 25, the maximum training epochs is 100. The query samples generally decided on the capacity of GPU memory, but it suggests to large than the number of support samples. All the experiments had been tested on the computer with GPU of Nvidia GeForce GTX 2060 6GB and CPU of Intel Core i7-10750H of 2.60GHz.

4.3 Result and Analysis

The few-shot fault diagnosis between different speed and load conditions including one-shot and five-shot learning tasks have been carried out, the experimental results are shown in Table 3. These results illustrate that our MFFD method performs best compared with other advanced baseline methods for all the few-shot tasks. Results of five-shot tasks generally better than the one-shot tasks mean that increasing the number of data samples will notably improve the performance of our model. MFFD with PN is better than MFFD with MN in most cases, but the number of samples for every category of support set is a more critical determinant. Limited by the PN mechanism, the PN based method only tested on five-shot situations.

Table 3 Accuracy (%) on one-shot and five-shot learning tasks for gearbox fault diagnosis with different working conditions

	30L→35H		30L→40H		30L→50L	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Finetune last	40.77	45.53	36.27	36.82	32.24	34.18
Fintune whole	44.43	51.87	39.33	46.34	37.50	43.29
Feature KNN	61.21	69.20	49.40	52.93	44.67	50.76
MFFD-MN	85.48	87.63	84.23	89.25	84.87	86.64
Feature KNN Proto	*	67.58	*	58.99	*	58.04
MFFD-PN	*	87.51	*	90.22	*	88.49
	40L→45L		40L→50L		40L→45H	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Finetune last	50.06	52.73	49.53	52.61	50.06	52.73
Fintune whole	50.08	61.05	51.03	62.65	50.08	61.05
Feature KNN	61.49	66.80	68.76	75.87	61.49	66.80
MFFD-MN	65.33	72.04	83.94	89.23	65.32	70.04
Feature KNN Proto	*	58.80	*	63.91	*	58.80
MFFD-PN	*	75.85	*	80.20	*	75.85

Then, we test different methods on one-shot and five-shot learning tasks under different categories of the same conditions. In this case, there are two kinds of fault type as the few-shot tasks of the target domain, CT and MT, respectively. In CT, we assume that the chipped tooth and normal data samples are limited, while the missing tooth and normal data are limited in MT. Therefore, in this part of gearbox experiments, all the few-shot fault diagnosis tasks have 5 categories in the source domain in terms of training and 2 categories in the target domain waiting for testing. The results are shown in Table 4, the MFFD performs best at the majority of cases, and an interesting phenomenon is revealed which is that all the accuracies of MT are higher than the CT. We believe that the training data of the source domain in MT is more difficult to identify than the task in CT, then the embedding network will be learned much better, therefore, it can offer more transferable information for supporting the target few-shot learning tasks.

In summary, comparing with other state-of-the-art methods, our MFFD model performs best on all one-shot and five-shot tasks for the gearbox fault diagnosis under multiple limited data conditions. The PN-based MFFD is verified better than the MN-based MFFD in certain circumstances.

Table 4 Accuracy (%) on one-shot and five-shot learning tasks for gearbox fault with different categories under same conditions

	30L		30H		40L	
	CT	MT	CT	MT	CT	MT
One-shot						
Finetune last	88.46	93.96	64.38	93.12	97.16	91.58
Finetune whole	77.98	90.96	64.22	95.82	76.02	93.54
Feature KNN	91.7	100.00	71.38	100.00	99.66	100.00
MFFD-MN	85.08	100.00	65.84	100.00	99.84	100.00
Five-shot						
Finetune last	87.80	99.36	75.60	93.88	98.98	92.18
Finetune whole	89.12	96.62	79.30	99.36	90.22	96.36
Feature KNN	90.52	100.00	82.74	100.00	99.70	100.00
MFFD-MN	90.64	100.00	90.22	100.00	99.80	100.00
Feature KNN Proto	84.18	100.00	78.02	100.00	99.78	100.00
MFFD-PN	86.90	100.00	92.06	100.00	99.94	100.00
	40H		50L		50H	
	CT	MT	CT	MT	CT	MT
One-shot						
Finetune last	89.04	95.78	94.92	80.3	87.68	89.68
Finetune whole	80.82	97.00	79.26	74.92	72.46	96.74
Feature KNN	93.68	100.00	99.76	100.00	94.10	100.00
MFFD-MN	99.22	100.00	100.00	100.00	97.80	100.00
Five-shot						
Finetune last	96.02	97.06	98.76	86.22	94.04	90.90
Finetune whole	93.38	99.60	93.08	90.30	84.86	99.72
Feature KNN	95.24	100.00	99.98	100.00	96.84	100.00
MFFD-MN	96.90	100.00	100.00	100.00	99.14	100.00
Feature KNN Proto	96.88	100.00	100.00	100.00	99.14	100.00
MFFD-PN	97.20	100.00	100.00	100.00	94.92	100.00

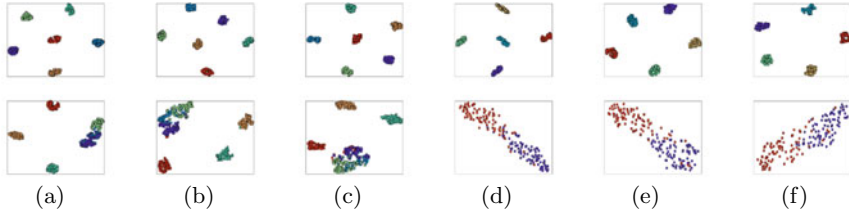


Fig. 2 t-SNE visualization of gearbox data features embedding derived from the MFFD. **a–c** denote results of MFFD-MN 1-shot, MFFD-MN 5-shot and MFFD-PN 5-shot respectively for the 30L→40H task. **d–f** denote results of MFFD-MN 1-shot, MFFD-MN 5-shot and MFFD-PN 5-shot respectively for the chipped tooth (CT) fault task under the condition 30 Hz low load. For each sub-figure, the up figure denotes the data feature from the source domain and the down figure denotes the data feature from the target domain

4.4 Visualization Analysis

The t-SNE [17] has been used to visualize the embedding mapping for analyzing the results of our MFFD method. The visualization of two-dimensional results are displayed in Fig. 2, different colour denotes different failure categories. The results of MFFD demonstrate that the embedding feature of the same category is placed at the closed area and the feature from different fault types is clearly distinguished in the source domain, which means that the model is trained well in the source domain. However, some fault types in the target domain are interacted with different type, while most of the fault is notably separable. These results verify that our MFFD method can really utilize the transferable knowledge learning from the source domain to deal with the few-shot learning tasks in the target domain. All in all, we can confirm that our MFFD method has very remarkably effective to overcome the few-shot fault diagnosis problem for gearbox under various limited data conditions.

5 Conclusion

In this work, a novel Meta-Learning guided Few-shot Fault Diagnosis framework, named MFFD, to address the few-shot learning problem for gearbox fault diagnosis under various limited data conditions has been proposed and verified. The essence of our work is to learn the transferable knowledge from the source domain, then making the trained model with a support set able to directly identify the fault types in the target domain. Experimental and Visualized results on one-shot and five-shot tasks of the gearbox fault diagnosis indicate that MFFD performs better than other advanced methods.

Acknowledgements The authors acknowledge the supported by the RECLAIM project “Remanufacturing and Refurbishment Large Industrial Equipment” and received funding from the European Commission Horizon 2020 research and innovation programme under grant agreement No. 869884.

References

1. Jiang, G., He, H., Yan, J., Xie, P.: Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Trans. Ind. Electron.* **66**(4), 3196–3207 (2018)
2. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., Nandi, A.K.: Applications of machine learning to machine fault diagnosis: a review and roadmap. *Mech. Syst. Signal Process.* **138**, 106587 (2020)
3. Lu, W., Liang, B., Cheng, Y., Meng, D., Yang, J., Zhang, T.: Deep model based domain adaptation for fault diagnosis. *IEEE Trans. Ind. Electron.* **64**(3), 2296–2305 (2016)
4. Wen, L., Li, X., Gao, L., Zhang, Y.: A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Ind. Electron.* **65**(7), 5990–5998 (2017)
5. Zhang, M., Wang, D., Lu, W., Yang, J., Li, Z., Liang, B.: A deep transfer model with Wasserstein distance guided multi-adversarial networks for bearing fault diagnosis under different working conditions. *IEEE Access* **7**, 65303–65318 (2019)
6. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X.: Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **115**, 213–237 (2019)
7. Hu, T., Tang, T., Lin, R., Chen, M., Han, S., Wu, J.: A simple data augmentation algorithm and a self-adaptive convolutional architecture for few-shot fault diagnosis under different working conditions. *Measurement* **156**, 107539 (2020)
8. Ren, Z., Zhu, Y., Yan, K., Chen, K., Kang, W., Yue, Y., Gao, D.: A novel model with the ability of few-shot learning and quick updating for intelligent fault diagnosis. *Mech. Syst. Signal Process.* **138**, 106608 (2020)
9. Zhang, A., Li, S., Cui, Y., Yang, W., Dong, R., Hu, J.: Limited data rolling bearing fault diagnosis with few-shot learning. *IEEE Access* **7**, 110895–110904 (2019)
10. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D.: Matching networks for one shot learning. In: *Advances in Neural Information Processing Systems*, pp. 3630–3638 (2016)
11. Wang, D., Cheng, Y., Yu, M., Guo, X., Zhang, T.: A hybrid approach with optimization-based and metric-based meta-learner for few-shot learning. *Neurocomputing* **349**, 202–211 (2019)
12. Wang, D., Zhang, M., Xu, Y., Lu, W., Yang, J., Zhang, T.: Metric-based meta-learning model for few-shot fault diagnosis under multiple limited data conditions. *Mech. Syst. Signal Process.* **155**, 107510 (2021)
13. Lu, J., Gong, P., Ye, J., Zhang, C.: Learning from very few samples: a survey (2020). arXiv preprint [arXiv:2009.02653](https://arxiv.org/abs/2009.02653)
14. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: *Advances in Neural Information Processing Systems*, pp. 4077–4087 (2017)
15. Xie, J., Zhang, L., Duan, L., Wang, J.: On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis. In: *2016 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–6. IEEE (2016)
16. Zhang, M., Lu, W., Yang, J., Wang, D., Bin, L.: Domain adaptation with multilayer adversarial learning for fault diagnosis of gearbox under multiple operating conditions. In: *2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, pp. 1–6. IEEE (2019)
17. Maaten, L.v.d., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(Nov), 2579–2605 (2008)

Investigation into LSTM Deep Learning for Induction Motor Fault Diagnosis



Xiaoyu Zhao, Ibrahim Alqatawneh, Mark Lane, Haiyang Li, Yongrui Qin, Fengshou Gu, and Andrew D. Ball

Abstract As motor faults could lead to unwanted loss in industry, it is important to find out the motor faults in time. Currently, with the popularity and mature application of deep learning, researchers in the field of electrical machine health assessment have begun to focus on deep learning methods. It is hoped that motor fault detection can be achieved with the help of deep learning methods. This paper presents to adopt deep learning methods represented by LSTM neural network for motor fault diagnosis and evaluates on our own experimental platform. Considering two typical motor faults with two different degrees of severity, the results show that the proposed LSTM approach has a high accuracy (98.81%) on motor fault classification. The results also confirm that: (1) adequate effort of preprocessing, including sample length selection in the time domain and frequency band selection in the frequency domain, can significantly improve accuracy and computational efficiency; (2) different faults can be separated through the information in frequency band of 100–1000 Hz, which has not been fully modelled analytically before.

Keywords Deep learning · LSTM · Motor fault diagnosis

1 Introduction

Owing to the importance of induction motors to the industry, research related to induction motors have received much attention, and fault diagnosis of induction motors is one of them. The increasing demand for high reliability of motors and people's increasing awareness of safe production, have promoted the development of motor fault diagnosis technique. In addition, the complex environment in which

X. Zhao · I. Alqatawneh · M. Lane · H. Li · Y. Qin · F. Gu (✉) · A. D. Ball
Centre for Efficiency and Performance Engineering (CEPE), School of Computing and Engineering, University of Huddersfield, Huddersfield, UK
e-mail: F.Gu@hud.ac.uk

X. Zhao
e-mail: Xiaoyu.zhao@hud.ac.uk

the motor works has also increased the possibility of multiple faults occurred, which makes motor fault diagnosis imperative.

As introduced in [1], current motor fault diagnosis methods can be categorized to three types: model-based, signal-based or knowledge-based methods. Model-based method builds an effective mathematical model using the parameters in real system to obtain a predicted value, and to compare the consistency between the prediction value and practical value from the real system [2, 3]. While the signal-based methods use the practical output signal to capture useful information and perform fault diagnosis on information without depending on a model. For a signal-based method, it knows the patterns of motor's healthy working condition, and does fault diagnosis through checking the consistency between the measured value and the healthy value [4]. As for the knowledge-based method, it requires a large amount of historical data. Through applying artificial intelligent approaches to implicitly extract features from historical data, it obtains the knowledge base. Then, through checking the consistency between the measured value in the real system and the knowledge base, a fault diagnosis decision can be made using classifier [4]. One of the well-known knowledge-based methods is deep learning (DL). DL methods are good at learning the patterns or features from raw data, and can find the correlation automatically. They have been considered as a good choice to deal with the fault diagnosis problems which usually have complex and variable data. The earliest DL applications in motor fault diagnosis can be traces back to 2016 [5, 6] by Sun et al., and have gradually become popular these years with the advent and development of multiple neural network toolkits (TensorFlow, 2015 [7]; Keras, 2015 [8]; MATLAB deep learning Toolbox, 2019 [9]), which brings convenience to researchers to apply DL neural network into their own data set without deeply studying and constructing its complex network structure. Several DL methods have been investigated for automated fault diagnosis of motor, including Convolutional Neural Network (CNN) [10], Recurrent Neural Network (RNN) [11, 12], and Deep Belief Network (DBN) [13]. The mentioned DL-based studies all train their models using acquired dataset, spontaneously learning features in data set, and draw diagnosis conclusions. However, some methods claim that it can feed their dataset into the DL based model without preprocessing and only use raw data can achieve good classification results. Also, some methods feed all the frequency data or time domain data directly, which brings large amount of calculation and increases requirements for computing equipment. Therefore, some models in this paper are proposed to address the mentioned issues, hoping to verify the necessity of performing preprocessing on motor data and to find a good solution in order to reduce the size of data set.

This paper adopts Long Short-Term Memory neural network (LSTM) [14], a widely used and improved variant of RNN, for building models in induction motor fault diagnosis. It takes time domain signal and frequency domain signal in different frequency distribution of motor current signal as data set for model training. As a result, the RNN method represented by LSTM is evaluated for its ability to process various types of data, and the amount of information contained in various data for motor fault diagnosis is verified. Also, it is found that some essential preprocessing

methods are critical to LSTM network, and it has huge impact on improving accuracy for motor current fault diagnosis.

We organize the rest of this paper as follows. Section 2 briefly reviews the theory of adopted LSTM methodology. Section 3 introduces the testing scenes setup procedure and shows the characteristics of motor dataset. Section 4 presents our data preprocessing techniques, which are applied before data fed into the LSTM network and provides detailed analysis of the results. Finally, the paper is concluded in Sect. 5 with discussions of future work.

2 Methodology

2.1 LSTM Theory

Long Short-Term memory (LSTM) network is a type of RNN developed by Hochreiter and Schmidhuber [14] to overcome the exploding and vanishing gradient problems in RNN. As shown in Fig. 1, the structure of an LSTM unit consists of a cell and three gates, the cell is the memory part of the LSTM unit; gates are used to control the flow of information inside the LSTM unit, including forget gate f_t , input gate i_t and output gate o_t . Each gate in the LSTM cell receives the current input data vector x_t , the previous hidden state h_{t-1} , and the previous cell state c_{t-1} .

As shown in Fig. 1, the **Forget gate** f_t determines the internal information needs to be removed from the previous cell state c_{t-1} , and it can be calculated as follows [15]:

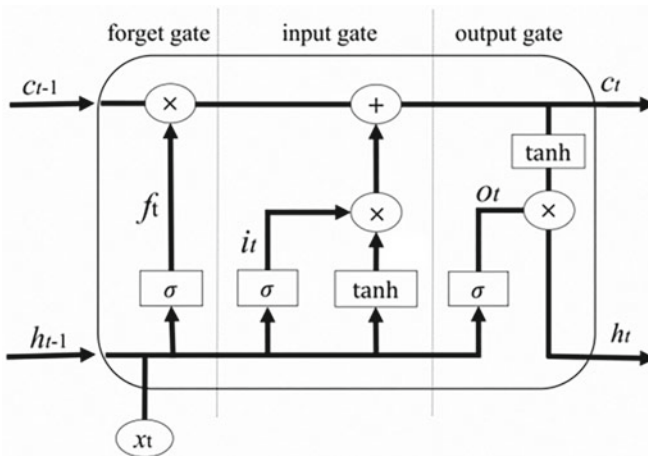


Fig. 1 LSTM structure

$$f_t = \text{sigmoid}(w_{xf}x_t + w_{hf}h_{t-i} + b_f) \quad (1)$$

Input gate i_t determines the new information will be stored in the current cell state c_t , and it can be calculated as follows [16]:

$$i_t = \text{sigmoid}(w_{xi}x_t + w_{hi}h_{t-i} + b_i) \quad (2)$$

$$\bar{c}_t = \tan h(w_{xc}x_t + w_{hc}h_{t-i} + b_c) \quad (3)$$

Output gate o_t determines which information will be passed through the gate and get into the rest of the network, and it can be calculated as follows [17]:

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \bar{c}_t \quad (4)$$

$$o_t = \text{sigmoid}(w_{xo}x_t + w_{ho}h_{t-i} + b_o) \quad (5)$$

$$h_t = o_t \otimes \tan h(c_t) \quad (6)$$

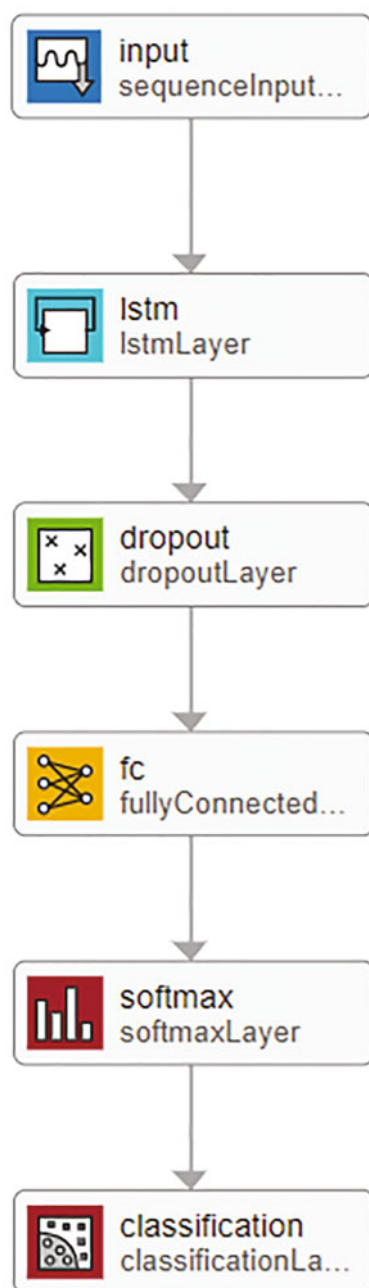
where: (1) x_t is the input of the memory cell, h_t is the current hidden state, (2) f_t , i_t , o_t are the output of forget gate, input gate and output gate, respectively, (3) c_t is the current cell state, (4) c_{t-1} is the previous cell state, (5) h_{t-i} is the previous hidden state, (6) w_{xf} , w_{hf} , w_{xi} , w_{hi} , w_{xc} , w_{hc} , w_{xo} , and w_{ho} represent the corresponding weight matrix, (7) b_f , b_i , b_c , and b_o are the bias vectors.

2.2 Adopted LSTM Structure

Figure 2 shows the LSTM network architecture used in this study. It consists of six layers. Input layer, LSTM layer, dropout layer, fully connected layer, softmax layer and finally the classification layer.

The first layer is a sequence input layer, inputting sequence data to adopted network. Followed by a LSTM layer, LSTM layer is used to learn the long-time dependences in the time series and sequence data as described in the above section. A dropout layer is placed after the LSTM layer, and is used to reduce the possibilities of overfitting by randomly setting input elements from LSTM layer to zero with a given probability of 0.4. A fully connected layer multiplies the input elements obtained from dropout layer by a weight matrix and then adds a bias vector, it often follows LSTM layers and is used for outputting a prediction. In this case, the output size in the fully connected layer is the number of motor working states. The softmax layer is used before the classification layer to normalize the output of the fully connected layer of K values into a probability distribution in the range of [0,1], with the target class having the highest probability among K probabilities. Cross-entropy

Fig. 2 Adopted LSTM
neural network structure



loss function is employed at the end of the architecture for performing multi-class classification task. It is used to measure the error between the estimated softmax output probability distribution and the target class.

3 Test Scenes Setup and Generated Datasets

3.1 Test Scenes Setup

The working mechanism of the motor test rig is shown in Fig. 3. It mainly includes an induction motor, a host PC for collecting signal through the data acquisition system and a generator which is mechanically coupled with the motor to provide various loads and speeds. In addition, a generator controller is used to control the generator to simulate multiple operating conditions. See Figs. 4 and 5 for the test rig layout

Fig. 3 Diagram of testing platform

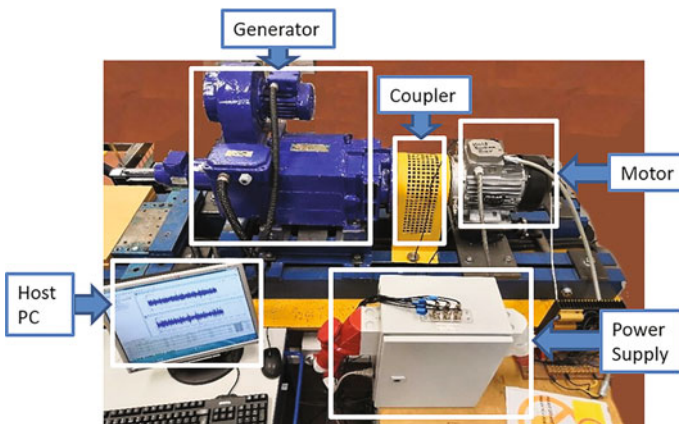
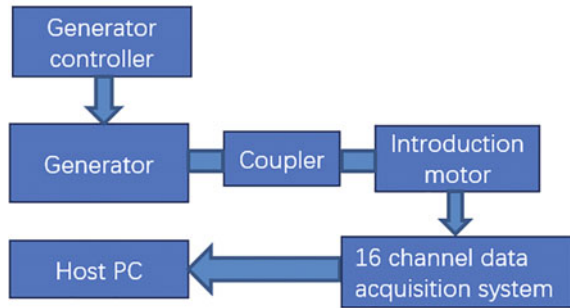


Fig. 4 Induction motor test rig



Fig. 5 Data acquisition system (left) and generator controller (Right)

and other equipment used in our motor data acquisition experiment, more detail can be found in [18].

3.2 Generated Datasets

The test rig is used in our experiment to produce the broken bar faults and horizontal misalignment faults with two different degrees of severity, see Table 1. Except for healthy working condition (healthy motor working in healthy working condition $\text{misalignment} \leq 0.07 \text{ mm}$), a total of five different motor working conditions have been generated. In order to better explain, we use N, B1, B2, M1, M2 to represent the 5 different conditions, in which N is used to indicate healthy condition, other representations can be found Table 1. The broken bar fault is achieved by installing three motors in different degree of fault. The horizontal fault misalignment is made by adjusting the position of the flexible coupler through moving a different imbalance degree.

In the motor fault data acquisition tests, we monitor and collect motor current signal from host PC in the data acquisition system. In order to obtain a reliable dataset, we keep the motor running for 5 min in each working condition, and begin to sample after running 3 min, sampling for 60 s with a sample rate of 96 kHz. Eventually, we obtain a time sequence in a total of 5,760,000 sampling points for

Table 1 Fault categories and groups

Fault category	Fault degree of severity	
	Broken bar fault	Motor with half broken bar (B1)
Horizontal misalignment fault	Misalignment = 0.7 mm (M1)	Misalignment = 1.1 mm (M2)

each motor working condition, obtaining a dataset in 5 different working conditions, which are healthy working condition (N), working in broken bar fault: motor with half broken bar (B1) and motor with one broken bar (B2), working in horizontal misalignment fault: misalignment = 0.7 mm (M1), misalignment = 1.1 mm (M2).

4 LSTM Data Modelling and Results Analysis

4.1 Dataset Preparation

As introduced in Sect. 3, we obtain a 60-s raw signal data in the time domain for each working state. Here, we divide the raw time sequence to 84 samples using partial overlap method and each sample has a time sequence of 131,073, for the purpose of making datasets. Therefore, the adopted LSTM neural network can have sufficient samples in the datasets for model training and test.

For Model_T, we take the separated samples as dataset. For the other models, written as Model_F, we convert samples from time domain to frequency domain through performing Fast Fourier Transform (FFT), and choose data in different frequency ranges as datasets in order to identify the most useful information in various frequency bands for fault diagnosis. Table 2 describes the different samples composition used in our models. Here, we not only selected the data near the power frequency (0–100 Hz) that is commonly used for fault analysis, but also creatively proposed and selected other higher frequency bands such as 100–1000 Hz and 2500–3500 Hz for modeling. In this way, we can test the importance of preprocessing methods in motor fault diagnosis.

Table 2 Model description using different datasets

Model	Data size in each sample	Description
Model_T	32,768	Samples in time domain
Model_F(0–100 Hz)	274	Samples in frequency domain with frequency range of 0–100 Hz
Model_F(0–100 Hz)_R	264	Same as Model_F (0–100 Hz) but excluding samples in the 48–52 Hz band
Model_F(100–1000 Hz)	2457	Samples in frequency domain with frequency range of 100–1000 Hz
Model_F(2500–3500 Hz)	2731	Samples in frequency domain with frequency range of 2500–3500 Hz
Model_F(0–100 Hz)_S	101	A non-overlapping supplementary model. It uses a time sequence of 10,000, and separates into 57 samples in frequency domain with frequency range of 0–100 Hz

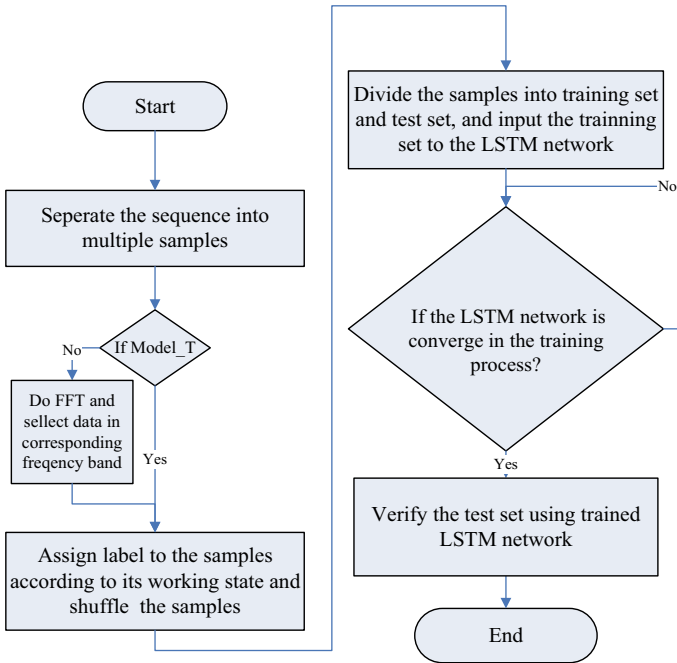


Fig. 6 Flowchart for the proposed method

We assign labels to datasets according to the corresponding working conditions for supervised learning and shuffle the datasets. In order to building our models, we divide the datasets to training sets and test sets according to the ratio of 8:2 in view of limited samples, so each time we use 80% data in data set for model training and use 20% data for model verifying. The flow chart in Fig. 6 displays the whole process of our LSTM Dataset Modelling.

4.2 Results and Analysis

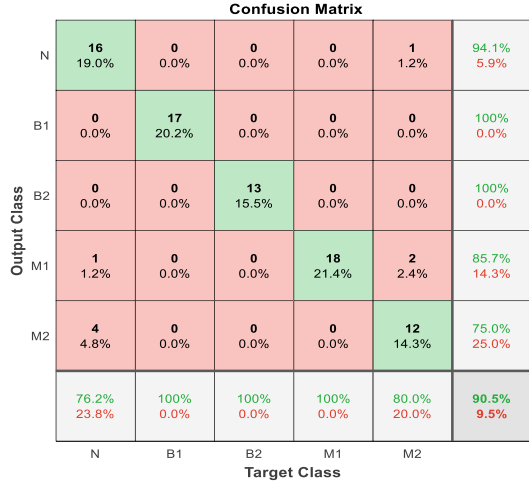
The fault diagnosis is performed in a laptop with Intel Core i5-9300H 2.4 GHz and 8 GB RAM. The operating system is 64bit. The LSTM neural network is built with the help of Deep Learning Toolbox in MATLAB R2021a. The model accuracy performance is listed in Table 3.

Model_T takes 66 min for the training, as the amount of raw data in time domain is quite large. However, the classification accuracy is low, due to the complex information components and inadequate percentage of useful information, which is only 43.66%. According to the fault diagnosis performance on Model_F(0–100 Hz) and Model_F(0–100 Hz)_R, we find the power frequency band around 50 Hz (48–52 Hz) has a huge impact on LSTM-based neuron network as after removing the data in this

Table 3 Model accuracy performance on N-B1-B2-M1-M2 five classes fault classification

Model	Accuracy performance (%)
Model_T	43.66
Model_F(0–100 Hz)	35.07
Model_F(0–100 Hz)_R	90.48
Model_F(100–1000 Hz)	98.81
Model_F(2500–3500 Hz)	21.43

Fig. 7 Confusion matrix for Model_F(0–100 Hz)_R



power frequency range, the accuracy has increased from 35.07 to 90.48%. Furthermore, from the confusion matrix for Model_F(0–100 Hz)_R in Fig. 7, we can see this model achieves better fault diagnosis performance in broken bar faults than misalignment faults. To be specific, this model obtains 100% accuracy in classifications of all broken bar faults, while it only achieves 85.19% accuracy in classifying misalignment faults. Model_F(100–1000 Hz) achieves a highest classification accuracy among all the models, which is 98.81%, with only 1 sample misclassified between fault M1 and M2 (See Fig. 8). Furthermore, Model_F(100–1000 Hz) shows the best overall classification performance on misalignment failure. Because of Model_F(100–1000) that performs well in higher frequency bands, we also consider the performance of higher frequency band. Nevertheless, Model_F(2500–3500 Hz) provides the worst performance, where the adopted LSTM network could not find much useful information from its input data.

In addition to the above models, we develop a supplementary Model_F(0–100 Hz)_S (the detail of this model is shown in Table 2), which is used for fast broken bar faults diagnosis, hoping to adopt a model with smaller size in samples to decrease model training time, while retaining good classification performance as shown in Model_F(0–100)_R. However, this model does not perform as good as we expected, having a good fault diagnosis performance though it uses frequency domain

Fig. 8 Confusion matrix for Model_F(100–1000 Hz)

		Confusion Matrix						
		N	B1	B2	M1	M2		
Output Class	N	19 22.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100%	0.0%
	B1	0 0.0%	17 20.2%	0 0.0%	0 0.0%	0 0.0%	100%	0.0%
	B2	0 0.0%	0 0.0%	12 14.3%	0 0.0%	0 0.0%	100%	0.0%
	M1	0 0.0%	0 0.0%	0 0.0%	20 23.8%	0 0.0%	100%	0.0%
	M2	0 0.0%	0 0.0%	0 0.0%	1 1.2%	15 17.9%	93.8%	6.2%
			100%	100%	100%	95.2%	100%	98.8%
		N	B1	B2	M1	M2	Target Class	

data in 0–100 Hz as well. The decreased size in each sample slightly does accelerate the training process of LSTM network, while performs poor in classification accuracy. This is because the non-overlapping way of making data into 57 samples deceases the sampling time in each sample, leading to reduction of resolution. Therefore, the sideband fault feature cannot be found due to the narrow bandwidth, which results in fault feature cannot be kept in data set and causes the bad diagnosis performance.

From the five Models, we can arrive at some findings.

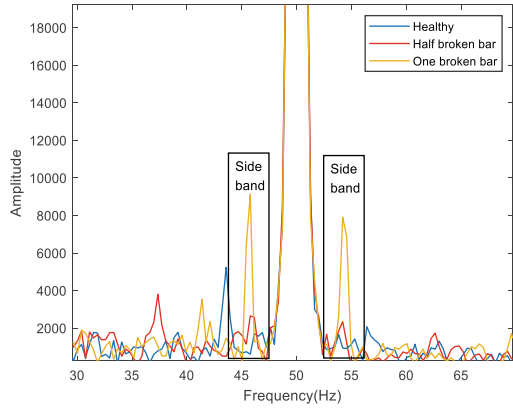
- **The processing capacity of neural network for raw data in time domain is limited.**

Model_T uses the unprocessed time domain current signal samples as input for the LSTM network. The data size for each sample in Model_T is quite large compared to other samples (see Table 3), which brings a huge computational burden both on extra memory footprint and increasing training time. Nevertheless, the Model_T obtains an undesirable result due to complex information components in time domain and the background noise caused by system disturbance.

- **The LSTM method can successfully learn sideband features near the power frequency in the broken bar faults scenarios, but it needs to remove the huge interference caused by the power amplitude.**

The performance of Model_F(0–100)_R is in line with prior knowledge that it has obvious features that can be used to detect the broken bar faults on both sides of the power frequency as shown in Fig. 9. The sidebands on both sides can be clearly seen, which leads to the occurrence of broken bar failure being quickly and accurately determined. However, comparing the performance on Model_F(0–100) and Model_F(0–100)_R, the motor power frequency has a lot of interference

Fig. 9 The amplitude spectrum for broken bar dataset



on fault diagnosis. This is because the huge amplitude in power frequency reduces the influence of sidebands on our LSTM model.

- **The information in the frequency range of 100–1000 Hz is found to be useful for motor fault diagnosis.**

Except for the frequency band information near the motor power frequency that is usually used for analyzing motor fault, the low frequency band of 100–1000 Hz has also been found carrying useful information for fault diagnosis through Model_F(100–1000). It is mainly because this frequency band contains important harmonic components of power frequency.

- **The higher frequency band of 2500–3500 Hz is not useful through the LSTM model for motor fault diagnosis.**

Model_F(2500–3500) verifies that LSTM approach does not find sufficient information that can be used to detect motor faults in the higher frequency band of 2500–3500 Hz. Therefore, the selection of frequency bands in frequency domain greatly influences the diagnosis accuracy.

- **The sample length selection in the time domain is significant, and inappropriate frequency resolution would lead to the loss of fault feature.**

The poor performance in Model_F(0–100)_S is because the non-overlapping way of data preprocessing decreases the sampling time in each sample, and leads to sideband fault feature cannot be collected in to the data set, which shows the significance of sample length selection in time domain.

5 Conclusion

Considering two typical motor faults, and faults in different degrees of severity, we demonstrate the superiority of the deep learning algorithm represented by LSTM neural network, and verify the feasibility of deep neural network in motor fault diagnosis. Through the models we adopted, it confirms that the LSTM algorithm can achieve a high accuracy rate for identifying the type of motor faults. In this process,

preprocessing for the raw data plays a critical role, which needs to be carefully planned and dealt with before feeding into the network. In addition, by using different frequency bands based on LSTM network modeling, the 100–1000 Hz band is found having useful information for fault classification, which has not been fully considered before.

In future, we plan to optimize the adopted LSTM model and add other features such as components of motor vibration signal into our work for further improving the accuracy of our fault detection models. In addition, we prepare to carry on experiments using different motors, then use collected data to evaluate the generalization of LSTM-based model.

References

1. Ince, T.: Real-time broken rotor bar fault detection and classification by shallow 1D convolutional neural networks. *Electr. Eng.* **101**(2), 599–608 (2019)
2. Gao, Z., Cecati, C., Ding, S.X.: A survey of fault diagnosis and fault-tolerant techniques—part I: fault diagnosis with model-based and signal-based approaches. *IEEE Trans. Industr. Electron.* **62**(6), 3757–3767 (2015)
3. Beard, R.V.: Failure accommodation in linear system through self reorganization (Ph.D. thesis). Massachusetts Institute of Technology, USA (1971)
4. Cecati, C.: A survey of fault diagnosis and fault-tolerant techniques—part II: fault diagnosis with knowledge-based and hybrid/active approaches (2015)
5. Sun, W., Shao, S.Y., Yan, R.: Induction motor fault diagnosis based on deep neural network of sparse auto-encoder. *J. Mech. Eng.* **52**(9), 65–71 (2016)
6. Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., Chen, X.: A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement* **89**, 171–178 (2016)
7. TensorFlow.: Retrieved 13 May 2021 (2021). From <https://www.tensorflow.org/>
8. Team, K.: Keras: the python deep learning API. Retrieved 13 May 2021 (2021). From <https://keras.io/>
9. Deep Learning Toolbox.: Retrieved 13 May 2021 (2021). From <https://uk.mathworks.com/products/deep-learning.html>
10. Choi, D.J., Han, J.H., Park, S.U., Hong, S.K.: Comparative study of CNN and RNN for motor fault diagnosis using deep learning. In: *IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, pp. 693–696. IEEE, (2020)
11. Xiao, D., Huang, Y., Zhang, X., Shi, H., Liu, C., Li, Y.: Fault diagnosis of asynchronous motors based on LSTM neural network. In: *Prognostics and system health management conference (PHM-Chongqing)*, pp. 540–545. IEEE, (2018)
12. Wang, L.H., Zhao, X.P., Wu, J.X., Xie, Y.Y., Zhang, Y.H.: Motor fault diagnosis based on short-time Fourier transform and convolutional neural network. *Chin. J. Mech. Eng.* **30**(6), 1357–1368 (2017)
13. Shao, S.Y., Sun, W.J., Yan, R.Q., Wang, P., Gao, R.X.: A deep learning approach for fault diagnosis of induction motors in manufacturing. *Chin. J. Mech. Eng.* **30**(6), 1347–1356 (2017)
14. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
15. Yin, A., Yan, Y., Zhang, Z., Li, C., Sánchez, R.V.: Fault diagnosis of wind turbine gearbox based on the optimized LSTM neural network with cosine loss. *Sensors* **20**(8), 2339 (2020)
16. Shi, J., Peng, D., Peng, Z., Zhang, Z., Goebel, K., Wu, D.: Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks. *Mech. Syst. Signal Process.* **162**, 107996 (2022)

17. Zou, P., Hou, B., Lei, J., Zhang, Z.: Bearing fault diagnosis method based on EEMD and LSTM. *Int. J. Comput. Commun. Control* **15**(1) (2020)
18. Li, H., Feng, G., Zhen, D., Gu, F., Ball, A.D.: A normalized frequency-domain energy operator for broken rotor bar fault diagnosis. *IEEE Trans. Instrum. Meas.* **70**, 1–10 (2020)

Degradation Trend Construction of Aircraft Engine Using Complex Network Model



Yongsheng Huang, Yongbo Li, Khandaker Noman, and Shun Wang

Abstract Health condition monitoring (HCM) of an aircraft engine is crucial to enhance its reliability running. In this paper, a novel method is proposed using the complex model to monitor its degradation trend. With the help of the data collected by multi-sensors, a complex dynamic model is built using the data-driven approach, which aims to achieve the purpose of HCM of aircraft engine. First, the Gath-Geva fuzzy clustering method is utilized for health condition division. Second, the network model based on correlation analysis is conducted. Finally, the dynamic improved logistic model is developed to describe the changes of sensors data of aircraft engine degradation trend. To verify the effectiveness of the proposed method, simulated aircraft gas turbofan engine data is utilized for validation. The results demonstrate that our method is effective to track its degradation process of aircraft gas turbofan engine.

Keywords Complex network · Health condition monitoring · Aircraft engine · Correlation analysis

1 Introduction

An aircraft engine produces thrust to propel an aeroplane, which is the main source of energy of an airplane. However, during the actual operations, the key components of engine such as gear, bearing and rotor are particularly prone to damage due to

Y. Huang · Y. Li (✉) · K. Noman · S. Wang
School of Aeronautics, Northwestern Polytechnical University, Xi'an, China
e-mail: yongbo@nwpu.edu.cn

Y. Huang
e-mail: hysheng@mail.nwpu.edu.cn

K. Noman
e-mail: khandakernoman@sjtu.edu.cn

S. Wang
e-mail: wangshun@mail.nwpu.edu.cn

heavy dynamic load, highly rough and harsh operating environments. Any abnormal behaviors in engine are very likely to decrease the overall performance of an aircraft and cause economic losses in real applications. Therefore, the health conditions of aircraft engines directly influence the safety and operation of the whole aircraft system [1].

In order to monitor the health condition of an aircraft engine, multiple sensors of varying types are embedded or deployed, which create a sensors network [2]. Methods using monitoring sensors data to predict the future health and estimates the remaining useful life has attracted increasing attention from academic researchers and industrial operators.

Sensor networks have been widely used to monitor the health condition of aircraft engines for predicting or isolating the serious abrupt faults [3]. In summary, aircraft engines health condition monitoring (HCM) methods can be classified into two major categories: the model-based method and the data-driven method [4]. For the model-based method, enough priori knowledge and stochastic model are required to identify the practical the accurate model, which is difficult. The data-driven method is based on modern signal processing techniques, which extracts the fault information by using the data collected from sensor networks, such as time domain features, frequency domain features, entropy. Moreover, classifiers such as neural network and support vector machine are employed to accomplish the fault pattern recognition and the monitoring of health conditions [5]. However, these machine learning-based methods mostly focus on specific components and do not analyze the mutual interaction of the parts of an aircraft engine. As a consequence, lots of information among the components of an aircraft engine is ignored.

In recent years, complex network theory is attracting increasing attentions in many interdisciplinary areas including sociology, biology, electricity, economics, communication, social network, etc. [6, 7]. Complex network can be used to describe the topological structure of various systems and reveal the internal law and evolution process of network, which can dig out the hidden information among the nodes of the network. Therefore the application of complex network methods for equipment health monitoring should be exploited. However, few scholars have done some research on complex network for machine HCM and fault diagnosis. Zhang proposed frequency complex network and build a new complex network structure feature named subnetwork average degree for rolling bearing fault diagnosis and degradation state recognition [8]. Zhang proposed an agglomerative cluster method and extracted diagnosis rules to fault diagnosis for the marine engine system [9]. In paper, a new HCM method for aircraft gas turbofan engine is proposed based on the complex network. Moreover, an experiment is conducted to prove its effectiveness.

The remaining part of the paper is organized as follows: Sect. 2 describes the HCM method for aircraft gas turbofan engine. An experiment base on the data from Prognostics and Health Management (PHM) challenge data in 2008 is used to prove the effectiveness of the proposed method. Finally, Sect. 4 concludes this paper.

2 Method

In this work we propose a method based on complex network to monitor the condition of an aircraft gas turbofan engine: (1) all the sensors of the aircraft gas turbofan engine as the nodes of the complex network, (2) the Pearson correlation coefficient among the sensor data are use as the weight of the network edge, (3) a connected weighted complex network is constructed using sensors data, (4) a dynamic model is built to help monitor the health condition of the aircraft gas turbofan engine. The flowchart for the proposed method is shown in Fig. 1.

2.1 Degradation Condition Division

The turbofan engine will undergo a gradual degeneration process from beginning operation to complete failure. Timely and accurately recognizing current performance degradation of turbofan engine can be important to improve the reliability of the whole mechanical turbofan engine. The data are normalized from zero to one.

Aiming at solving the problem of degradation condition recognition of bearings and other mechanical equipment, Wang Bing proposed the Gath-Geva (GG) fuzzy clustering method [10]. The method adds the original time series into the feature vector, considering the continuity of the bearing degradation state at the time scale. The dimensional eigenvector are constructed including all the sensors data (or the feature extracted) and the time series, and the GG fuzzy clustering method is used to divide the different states of the performance degradation conditions and

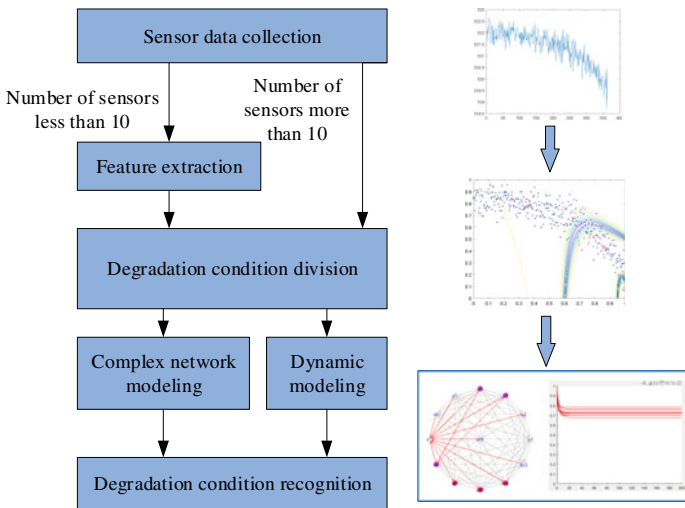


Fig. 1 The flowchart of the method for degradation condition recognition

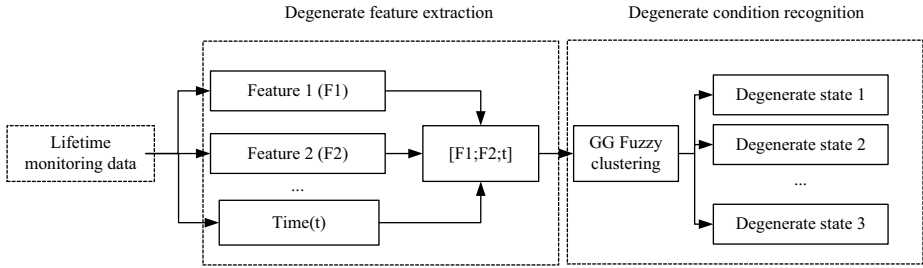


Fig. 2 Degradation condition recognition flow by GG fuzzy clustering

realize bearing degradation recognition. The Degradation condition recognition flow is shown in Fig. 2.

2.2 Complex Network Modeling

By using sensor network data, a complex network is built. The complex network should have a certain number of nodes and connected lines to present many different characteristics. Moreover with degradation of turbofan engines, the connected lines between the nodes appear or disappears evidently. The nodes of the sensor network can be the nodes of the complex network. The complex network model is based on Pearson correlation coefficient. The signs of the correlation coefficient indicate the positive and negative correlation between the two nodes: the positive sign indicates they are positively correlated; the negative sign indicates they are negatively correlated. The value of Pearson correlation coefficient can indicate the strong and weak relationship among the sensors [11]. Equation (1) defines the Pearson correlation coefficient.

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \tag{1}$$

Here, a threshold value is set to define whether the data collected from two different sensors should be correlated, in other words, whether the two nodes in the complex network should be connected with a line. The threshold should not be too low for the very weak correlations are considered. It should be neither high because few lines would be left and it is not conducive to complex network modeling. For example the threshold value is set at 0.3 in the experiment in Sect. 3.

There are many concepts and methods to describe the statistical characteristics of complex network structures [12]. Average degree is used in our work. Equation (2) defines average degree.

$$\bar{K} = \frac{1}{N} \sum_{i=1}^N k_i \tag{2}$$

Average degree means averaged number of other adjacent to graph nodes. In Eq. (2), N means the number of nodes; k_i represents the i th nodes.

2.3 Dynamic Modeling

Because of time varying nature of node states, node sets may belong to nonlinear dynamic systems. At present, the state of complex network is judged by observing the network structure during application, which lacks quantitative evaluation basis, and the result may lead to misjudgment, and the efficiency is low. So, a dynamic model based on data-driven method is built, and the model is interpretable. The dynamic model of every node can be expressed as follow:

$$\frac{dx_i}{dt} = F(x_i) + \alpha \sum_{j=1}^k ABx_i^m x_j^n - \beta \sum_{j=1}^k AB' x_i^m x_j^n \tag{3}$$

In Eq. (3), k is the number of the sensors, x_i is the sensor parameters to be solved, and x_j is the sensor parameters with predefined value. Equation (1) can be divided into three parts.

$$F(x_i) \tag{4}$$

$$\alpha \sum_{j=1}^k ABx_i^m x_j^n \tag{5}$$

$$\beta \sum_{j=1}^k AB' x_i^m x_j^n \tag{6}$$

Equation (4) shows the influence of the current value of x_i on the trend of its change; Eq. (5) shows the positive influence of the other sensor parameters (x_j) on the trend of its (x_i) change; Eq. (6) shows the negative influence of the other sensor parameters (x_j) on the trend of its (x_i) change.

To fit Eq. (4), the selected logic growth model and improve it. Equation (8) shows the differential equation of logic growth model.

$$\frac{dx_i}{dt} = rx_i \left(1 - \frac{x_i}{k}\right) \tag{7}$$

In Eq. (7), r is the maximum value that x can be achieve; k shows the growth potential [13]. We improve the logic growth model to fit the actual situation of condition monitoring as Eq. (8).

$$\frac{dx_i}{dt} = -x_i \left(1 - \frac{x_i}{2}\right) \quad (8)$$

In Eq. (8), assign r to one and k to two. The value 1 means the best condition, for that the data were normalized and r cannot be more than one. In the actual turbofan engine failure degradation it always starts slowly and then speeds up. So we assign k to two so that the value decreases slowly at one and then accelerate to fit the actual situation. Besides, in order to facilitate degradation condition recognition, we adjust the trend of change of all parameters with time to decrease.

3 Experiment

The sensor data sets that are utilized to carry out the evaluation are provided by National Aeronautics and Space Administration (NASA) Ames Research Center and have been used as Prognostics and Health Management (PHM) challenge data in 2008 [14]. We used the first group of the common test data set. The data for each cycle of each unit include the unit ID, cycle index, 3 values for the operational settings and 21 values for 21 sensor measurements. The sensor data are contaminated with noise. The data can accurately reflect the degradation characteristics of aircraft gas turbofan engines under different fault modes. Each group of the data is composed of 24 dimensional time series, of which 3 dimensional is the operation parameters and 21 dimensional is the measured value of the simulation sensor. There are one hundred data sets. The number of the sample for each simulation is around two hundred to four hundred. Each group of simulation data can be viewed as obtained from the sensor channel. The data contain complete aircraft gas turbofan engines life data. Some sensor channel data is sensitive to the fault of aircraft gas turbofan engine and has obvious degradation tendency, but some not. We remain the former and eliminate the latter.

We consider the process of degradation to be divided into six stages. After normalization of the data, we use the GG fuzzy clustering method to divide the process of degradation into six stages for each data set. A two-dimensional contour clustering result diagram is shown in Fig. 3.

In Fig. 3, the red dots mean the clustering centers, and the lines in different colors are the clustering contours, which can be regard as the boundary for different health condition. And we fused the data of the same health condition in one hundred clustering results to increase the number of samples. We set the threshold value at 0.3 to make the nodes connected have at least moderate correlation [15], and built the network model in six stages including the whole network, the positive network and the negative network.

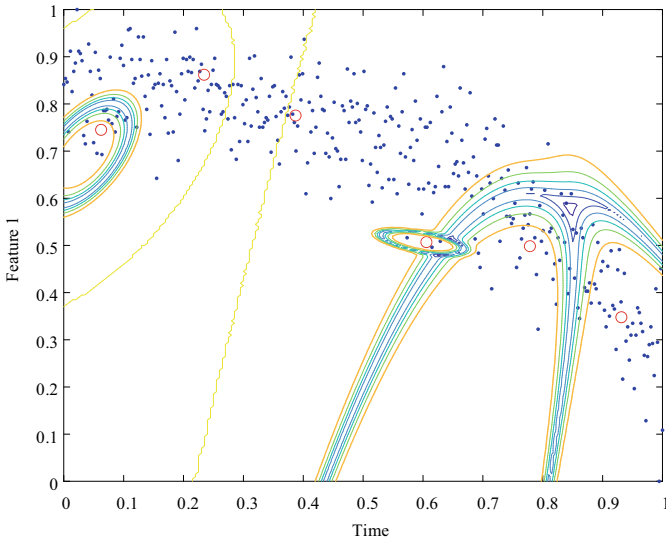


Fig. 3 Contour line effect for GG

Figure 4 shows the changes of the complex network in six degradation stages. The red lines mean the connected two nodes have negative correlation; the gray lines mean the connected two nodes have positive correlation Fig. 4a shows the change of the whole network; Fig. 4b shows the change of the positive network; Fig. 4c shows the change of the negative network. The two statistical characteristics are shown in Table 1.

Seen from Fig. 4 and Table 1, it can be observed as follows: (1) there are few connected lines in the network at first (low average degree); (2) the number the connected lines decreases in the second stage and the third stage, and there are even no lines left (the average degree is zero) in the third stage; (3) for the fourth stage, the number of lines increases, and finally there are many lines in the network (average degree increases). It is noticed that the more degenerate, the greater the number of lines. It can be concluded that, when the lines in the complex network increase, the fault have occurred in an aircraft gas turbofan engine, because with degradation of turbofan engines, the damage propagates from one component to others, which makes the number of nodes have correlation increases.

The genetic algorithm (GA) is used to find the best four parameters in Eq. (1), which helps to build the dynamic model of complex network. Here, the clustering center is considered as the most stable part in the degenerate stages, where the value hardly changes near it.

Figure 5 shows the built dynamic model based on improved logistic equation. It can approximate the variation of one sensor parameter in the process of degradation. If the current sensors parameters data are known, the health condition of engines can be easily judged. It provides decision basis for health condition monitoring.

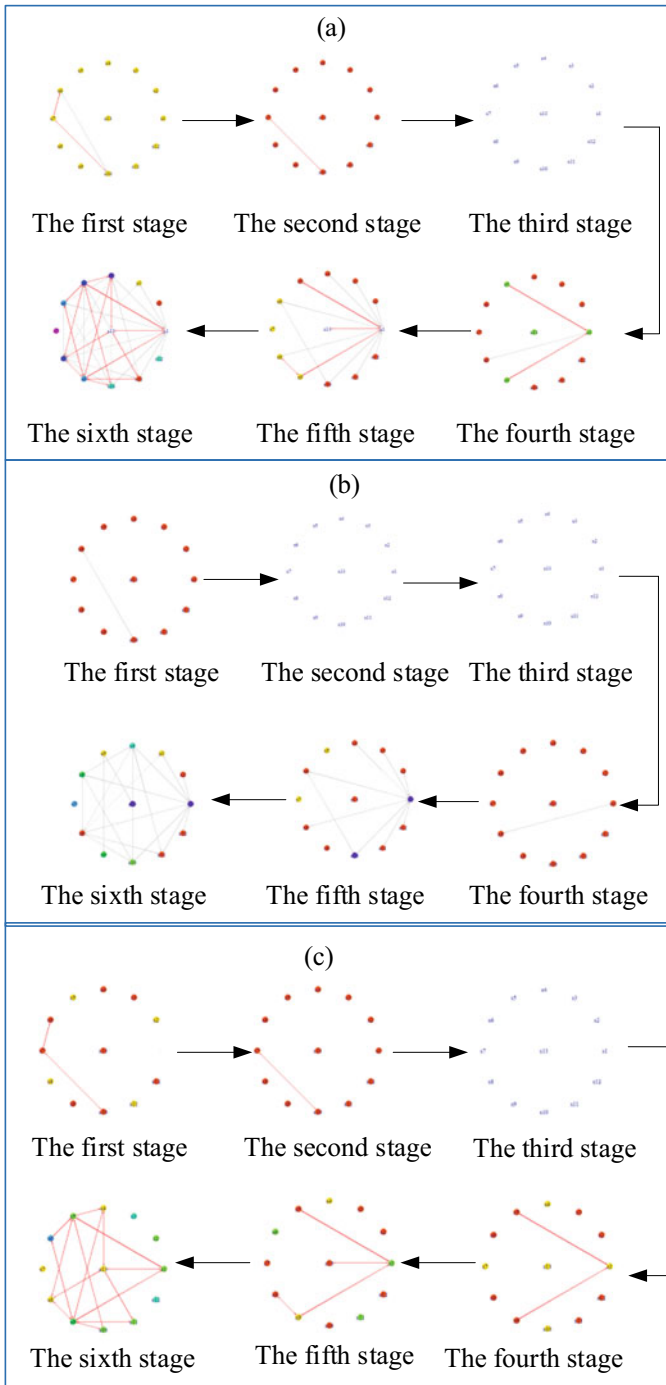


Fig. 4 The complex network model in the six stages: **a** the whole network, **b** the positive network, **c** the negative network

Table 1 Average degree in six stages

Stages number	The whole average degree	The positive average degree	The negative average degree
Stage 1	0.2308	0.0769	0.1538
Stage 2	0.0769	0	0.0769
Stage 3	0	0	0
Stage 4	0.2308	0.0769	0.1538
Stage 5	1	0.6923	0.3077
Stage 6	2.6154	1.4615	1.1538

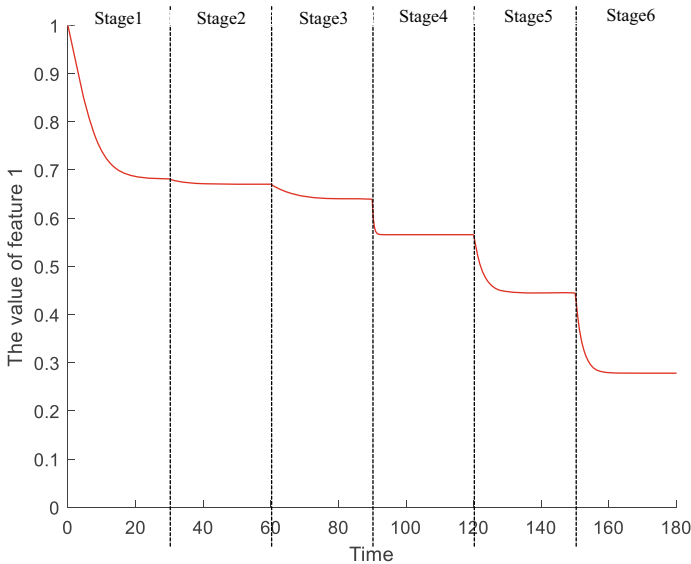


Fig. 5 Dynamic model of simulation

4 Conclusions

In the paper, we proposed a new health condition monitoring method based on the complex network and an experiment is used to prove its effectiveness. It can be concluded that the value of the average degree (No matter which kind of average degree) is related to the degree of degradation. With the increase of average degree, the health condition of aircraft turbofan engines deteriorates. Specially, when the average degree increases dramatically, the aircraft gas turbofan engine may fail. The multi-sensor network can provide a novel way to monitor the health condition monitoring of aircraft gas turbofan engine. The complex network-based approaches need nearly no prior knowledge about the aircraft gas turbofan engine, solve the problem of HCM for aircraft gas turbofan engines. Though it provides a new way of thinking for

aircraft gas turbofan engine health condition monitoring, the proposed method is not perfect. In future work, more highly interpretable indicators and dynamic model will be utilized to improve the complex network flowchart for engine health condition monitoring.

References

1. Zhang, B., Zheng, K., Huang, Q., et al.: Aircraft engine prognostics based on informative sensor selection and adaptive degradation modeling with functional principal component analysis. *Sensors* **20**(3), 920 (2020)
2. Yildirim, M.T., Kurt, B.: Aircraft gas turbine engine health monitoring system by real flight data. *Int. J. Aerosp. Eng.* (2018)
3. Wang, H., Jing, X.J.: A sensor network based virtual beam-like structure method for fault diagnosis and monitoring of complex structures with improved bacterial optimization. *Mech. Syst. Signal Process.* **84**, 15–38 (2017)
4. Liu, L., Liu, D., Zhang, Y., et al.: Effective sensor selection and data anomaly detection for condition monitoring of aircraft engines. *Sensors* **16**(5), 623 (2016)
5. Kong, C.: Review on advanced health monitoring methods for aero gas turbines using model based methods and artificial intelligent methods. *Int. J. Aeronaut. Space Sci.* **15**(2), 123–137 (2014)
6. Tang, Y., Qian, F., Gao, H., et al.: Synchronization in complex networks and its application—a survey of recent advances and challenges. *Annu. Rev. Control* **38**(2), 184–198 (2014)
7. Meng, C., Jiang, X.S., Wang, J., et al.: The complex network model for industrial data based on spearman correlation coefficient. In: *International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 28–33. IEEE (2019)
8. Zhang, Q., Tian, T., Wen, G., et al. A new modelling and feature extraction method based on complex network and its application in machine fault diagnosis. *Shock Vib.* **2018** (2018)
9. Zhang, A.P., Ren, G., Jia, B.Z.: Fault diagnosis for marine engine system based on complex network cluster method. In: *Advanced Materials Research*, vol. 655, pp. 801–805. Trans Tech Publications Ltd. (2013)
10. Wang, B., Wang, W., Hu, X., et al.: Degradation condition recognition method based on Gath-Geva fuzzy clustering. *Chin. J. Sci. Instrum.* **39**(03), 21–28 (2018)
11. Xu, H., Deng, Y.: Dependent evidence combination based on shearman coefficient and pearson coefficient. *IEEE Access* **6**, 11634–11640 (2017)
12. Watanabe, Y., Ishida, Y.: Performance evaluation of immunity-based diagnosis on complex networks. In: *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pp. 846–853. Springer, Berlin, Heidelberg (2007)
13. He, M., Chen, F., Li, Z.: Permanence and global attractivity of an impulsive delay logistic model. *Appl. Math. Lett.* **62**, 92–100 (2016)
14. Saxena, A., Goebel, K., Simon, D., et al.: Damage propagation modeling for aircraft engine run-to-failure simulation. In: *International Conference on Prognostics and Health Management*, pp. 1–9. IEEE (2008)
15. Munandar, T.A., Sumiati, S., Rosalina, V.: Pattern of symptom correlation on type of heart disease using approach of Pearson correlation coefficient. In: *IOP Conference Series: Materials Science and Engineering*, vol. 830, no. 2, pp. 022086. IOP Publishing (2020)

Research on the Influence of Mesh Stiffness of Fixed Gearbox with Chipping Fault



Jiacong Zhang, Yongbo Li, Shun Wang, and Khandaker Noman

Abstract Gearbox is a key component in rotating machinery and prone to chipping fault due to poor working environment. Hence, it is necessary to carry out research on chipping fault. Time-varying mesh stiffness is a periodic function caused by the change in the number of contact tooth pairs and the contact positions of the gear teeth. Time-varying mesh stiffness of is one of the main sources of vibration of a gear transmission system. Time-varying meshing stiffness provides the important information about the health status of the gear system. When a chip happens in one gear, mesh stiffness will decrease and consequently the vibration properties of the gear system will change. The vibration change can be characterized through gearbox dynamic modelling approach. In order to comprehensively understand the vibration properties of a gear set, it is essential to evaluate the time-varying mesh stiffness effectively. In this paper, the potential energy method is applied to analytically evaluate the time-varying mesh stiffness of a gear set. A modified cantilever beam model is used to represent the external gear tooth and analytical equations are derived without any modification of the gear tooth involute curve. A chip model is developed and the mesh stiffness reduction is quantified when chipping fault occurs in the pinion or the gear.

Keywords Mesh stiffness · Potential energy method · Gear set · Chip modeling

J. Zhang · Y. Li (✉) · S. Wang · K. Noman
School of Aeronautics, Northwestern Polytechnical University, Xi'an, China
e-mail: yongbo@nwpu.edu.cn

J. Zhang
e-mail: zhangjiacongz@163.com

S. Wang
e-mail: wangshun@mail.nwpu.edu.cn

K. Noman
e-mail: khandakernoman@sjtu.edu.cn

1 Introduction

Since the gearbox signal is faintness, its fault signal is difficult to identify when coupling with the gearbox signal in the transmission system. Together with large noise signal interference, fault diagnosis for gearbox is prone to make mistake judge. Therefore, it is necessary to study the coupled vibration characteristics of transmission systemic caused by fault in coupling signal. Time-varying mesh stiffness, caused by the change of tooth contact number and contact position, is one of the main sources of vibration of a gear transmission system. Hence, in order to comprehensively understand the vibration properties of a gear set, it is necessary to evaluate the mesh stiffness effectively.

When a pair of standard involute spur gear meshes, the tooth contact number and the tooth mesh position change during meshing. It leads to a periodic variation in the gear mesh stiffness. When a chip takes place, the gear mesh stiffness will reduce and consequently the vibration characteristics of the gear system will change. If the stiffness reduction can be quantified for different chip levels, the corresponding vibration signal can be obtained through dynamic simulation. The vibration signal can be processed further for chip detection and prognosis.

In this paper, external-external gears are used to denote a meshing gear pair which contains two external gears. Many researchers have applied the analytical method to evaluate the mesh stiffness of a pair of fixed shaft external-external gears. The potential energy method [1] is used to evaluate the mesh stiffness of a pair of fixed-shaft external-external gears with the consideration of Hertzian energy, bending energy, shear energy and axial compressive energy. Meanwhile, we model the gear teeth more rigorously by considering the gear tooth starting from the base circle when all teeth are perfect. The analytical equations for the Hertzian stiffness, the bending stiffness, shear stiffness and axial compressive stiffness are derived. The overall mesh stiffness is represented as a function of the rotation angular displacement of the driven gear named time-varying mesh stiffness.

The remaining parts of this paper are as follows.

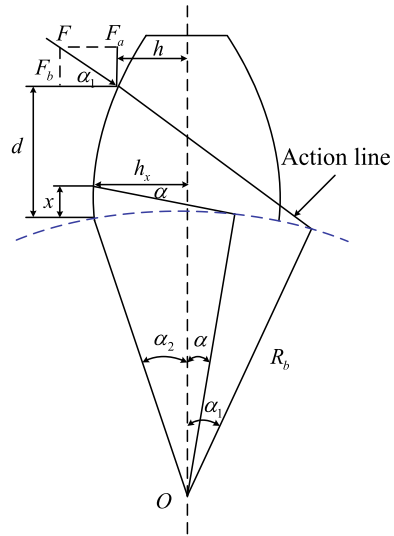
Calculation method of time-varying meshing stiffness of standard involute spur gear with both healthy tooth and chipping fault is presented in Sects. 2 and 3. Then simulation and discussion are carried out in Sect. 4. Finally, conclusion is drawn in Sect. 5.

2 Time-Varying Mesh Stiffness Calculation of Healthy Tooth

2.1 Mesh Stiffness Calculation

In this paper, it is assumed that the gear system is perfect since that the manufacturing and transmission error is ignored and the gear body is treated as solid. We will

Fig. 1 The geometry of a gear tooth



consider the lubrication condition is perfect such that the friction is ignored. For the purpose of simplification in calculating the mesh stiffness, only the stiffness of the meshing gears is considered. All other system components are considered rigid. The error in the mesh stiffness due to this simplification can be ignored since what we are concerned about is the difference between the faulty case and the perfect case.

In the potential energy method, we consider that the total potential energy stored in the meshing gear system included four components [2, 3]: the Hertzian energy, the bending energy, the shear energy and the axial compressive energy, which can be used for the calculations of the Hertzian contact stiffness, the bending stiffness, the shear stiffness and the axial compressive stiffness, respectively. According to material mechanics and elastic mechanical, the Hertzian energy, the bending energy, the shear energy and the axial compressive energy can be expressed (Fig. 1).

According to the properties of the involute curve, the action line of two meshing gears is always tangent to the gear base circle and normal to the tooth involute profile. The action force F which is along the action line, can be decomposed into two orthogonal forces F_a and F_b , the expression is as follows:

$$F_a = F \sin \alpha_1 \tag{1}$$

$$F_b = F \cos \alpha_1 \tag{2}$$

In this study, the tooth on the gear is modeled as a cantilevered beam. It is assumed that the base of the beam (at the root of the tooth) does not experience any deflection. According to potential energy method, the Hertzian energy, the bending energy, the shear energy and the axial compressive energy stored in a tooth can be expressed as

follows [2]:

$$U_h = \frac{F^2}{2k_h} \tag{3}$$

$$U_b = \frac{F^2}{2k_b} = \int_0^d \frac{[F_b(d-x) - F_a h]^2}{2EI_x} dx \tag{4}$$

$$U_s = \frac{F^2}{2k_s} = \int_0^d \frac{1.2F_b^2}{2GA_x} dx \tag{5}$$

$$U_a = \frac{F^2}{2k_a} = \int_0^d \frac{F_a^2}{2EA_x} dx \tag{6}$$

From the result derived by Yang and Sun, the Hertzian contact stiffness, k_h for a pair of external-external gears, is linearized to a constant along the entire line of action independent of both the position of contact and the depth of interpenetration [4].

$$k_h = \frac{\pi EL}{4(1 - \nu^2)} \tag{7}$$

Correspondingly, the bending energy, the shear energy and the axial compressive energy, there are the reciprocal of the bending stiffness, the shear stiffness [5] and the axial compressive stiffness, respectively as follows:

$$\frac{1}{k_b} = \int_0^d \frac{[\cos \alpha_1(d-x) - \sin \alpha_1 h]^2}{EI_x} dx \tag{8}$$

$$\frac{1}{k_s} = \int_0^d \frac{2.4(1 + \nu^2) \cos^2 \alpha_1}{EA_x} dx \tag{9}$$

$$\frac{1}{k_a} = \int_0^d \frac{\sin^2 \alpha_1}{EA_x} dx \tag{10}$$

where k_b , k_s and k_a denote the bending, shear and axial compressive stiffness, respectively, E and G represent Young’s modulus and shear modulus [6], respectively, h shows the distance between the gear contact point and the tooth central line, d is the distance from the contact point to the gear root, A_x and I_x indicate the area and the area moment of inertia of the section where the distance to the tooth root is x .

$$h = R_b[(\alpha_1 + \alpha_2) \cos \alpha_1 - \sin \alpha_1] \tag{11}$$

$$h_x = R_b[(\alpha_2 + \alpha) \cos \alpha - \sin \alpha] \tag{12}$$

$$d = R_b[\cos \alpha_1 + (\alpha_2 + \alpha_1) \sin \alpha_1 - \cos \alpha_2] \tag{13}$$

$$I_x = \frac{1}{12}(2h_x)^3 L = \frac{2}{3}h_x^3 L \tag{14}$$

$$A_x = 2h_x L \tag{15}$$

$$G = \frac{E}{2(1 + \nu)} \tag{16}$$

where R_b and L denote the base circle radius and tooth width of the external gear, respectively, h_x is the height of the section where the distance to the tooth root is x , α_2 represents the half tooth angle on the base circle [7].

$$\alpha_2 = \frac{\pi}{2N} + \tan \alpha_0 - \alpha_0 \tag{17}$$

where N is the tooth number of the external gear and α_0 is the pressure angle.

Substituting Eqs. (11)–(15) into Eq. (8), the bending stiffness of the external gear can be expressed as [8]:

$$\frac{1}{k_b} = \int_{-\alpha_1}^{\alpha_2} \frac{3\{1 + \cos \alpha_1[(\alpha_2 - \alpha) \sin \alpha - \cos \alpha]\}^2}{2E[(\alpha_2 - \alpha) \cos \alpha + \sin \alpha]^3 L} \{ (\alpha_2 - \alpha) \cos \alpha \} d\alpha \tag{18}$$

Substituting Eqs. (12) and (15) into Eq. (9), the shear stiffness of the external gear is given as:

$$\frac{1}{k_s} = \int_{-\alpha_1}^{\alpha_2} \frac{1.2(1 + \nu) \cos^2 \alpha_1}{EL[(\alpha_2 - \alpha) \cos \alpha + \sin \alpha]} \{ (\alpha_2 - \alpha) \cos \alpha \} d\alpha \tag{19}$$

Substituting Eqs. (12) and (15) into Eq. (10), the axial compressive stiffness of the external gear can be obtained:

$$\frac{1}{k_a} = \int_{-\alpha_1}^{\alpha_2} \frac{\sin^2 \alpha_1}{2EL[(\alpha_2 - \alpha) \cos \alpha + \sin \alpha]} \{ (\alpha_2 - \alpha) \cos \alpha \} d\alpha \tag{20}$$

2.2 The Angular Displacements Relation of the Meshing Gears

α_2 and α'_2 are the half tooth angles on the base circles of the pinion and the gear, respectively, which are constants for a given gear pair and can be calculated as Eqs. (21) and (22), respectively.

$$\alpha_2 = \frac{\pi}{2N_1} + inv\alpha_0 \tag{21}$$

$$\alpha'_2 = \frac{\pi}{2N_2} + inv\alpha_0 \tag{22}$$

where N_1 and N_2 are the numbers of teeth of the pinion and the gear, respectively.

Consider the initial meshing point of the first pair of meshing teeth as the reference point. At this point, the angular displacements of the meshing gears, θ_1 and θ_2 , are both zero. Equations are derived of the angles $\alpha_{1,1}^0$ (corresponding to the angle α_1 of the pinion) and $\alpha_{1,1}^{0'}$ (corresponding to the angle α_1 of the gear). The corresponding angle for the first pair of meshing teeth can be expressed as:

$$\alpha_{1,1}^0 = \tan(\arccos \frac{N_1 \cos \alpha_0}{\sqrt{(N_2 + 2)^2 + (N_1 + N_2)^2 - 2(N_2 + 2)(N_1 + N_2) \cos(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2} - \alpha_0)}}) - \frac{\pi}{2N_1} - inv\alpha_0 \tag{23}$$

$$\alpha_{1,1}^{0'} = \tan(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2}) - \frac{\pi}{2N_2} - inv\alpha_0 \tag{24}$$

When the angular displacement of the pinion is θ_1 , relation between the angular displacements of the pinion and the gear, θ_1 and θ_2 are consider. The initial reference displacement point is selected as the point where the double-tooth-pair mesh duration begins. The angular velocities of the pinion and the gear, ω_1 and ω_2 , have the following relationship:

$$\omega_2 = \frac{N_1}{N_2} \omega_1 \tag{25}$$

We can get the relationship of θ_1 and θ_2 :

$$\theta_2 = \frac{N_1}{N_2} \theta_1 \tag{26}$$

Therefore we can get the equations of the angles $\alpha_{1,1}$ (corresponding to the angle α_1 of the pinion) and $\alpha'_{1,1}$ (corresponding to the angle α_1 of the gear), which corresponding to the angular displacement of the pinion is θ_1 :

$$\alpha_{1,1} = \alpha_{1,1}^0 + \theta_1 = \theta_1 - \frac{\pi}{2N_1} - inv\alpha_0 + \tan(\arccos \frac{N_1 \cos \alpha_0}{\sqrt{(N_2 + 2)^2 + (N_1 + N_2)^2 - 2(N_2 + 2)(N_1 + N_2) \cos(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2} - \alpha_0)}}) \quad (27)$$

$$\begin{aligned} \alpha'_{1,1} &= \alpha'_{1,1} - \theta_2 = \tan(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2}) - \frac{\pi}{2N_2} - inv\alpha_0 - \theta_2 \\ &= \tan(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2}) - \frac{\pi}{2N_2} - inv\alpha_0 - \frac{N_1}{N_2} \theta_1 \end{aligned} \quad (28)$$

For the second pair of meshing teeth, there are corresponding angles $\alpha_{1,2}$ and $\alpha'_{1,2}$. The difference between the angle $\alpha_{1,2}$ and $\alpha_{1,1}$ is $\frac{2\pi}{N_1}$, and the difference between the angle $\alpha'_{1,1}$ and $\alpha'_{1,2}$ is $\frac{2\pi}{N_2}$ with negative. Thus, they can be expressed as:

$$\alpha_{1,2} = \alpha_{1,1} + \frac{2\pi}{N_1} = \theta_1 + \frac{3\pi}{2N_1} - inv\alpha_0 + \tan(\arccos \frac{N_1 \cos \alpha_0}{\sqrt{(N_2 + 2)^2 + (N_1 + N_2)^2 - 2(N_2 + 2)(N_1 + N_2) \cos(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2} - \alpha_0)}}) \quad (29)$$

$$\begin{aligned} \alpha'_{1,2} &= \alpha'_{1,1} - \frac{2\pi}{N_2} = \tan(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2}) - \frac{5\pi}{2N_2} - inv\alpha_0 - \theta_2 \\ &= \tan(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2}) - \frac{5\pi}{2N_2} - inv\alpha_0 - \frac{N_1}{N_2} \theta_1 \end{aligned} \quad (30)$$

2.3 Relation of Mesh Stiffness and Angular Displacement

For a pair of meshing gears whose contact ratio varies between 1 and 2, gear meshing includes single tooth meshing and double tooth meshing in one meshing period. The length of the duration of Double/Single-tooth-pair Meshing need calculated primarily, and then the calculation expressions of mesh stiffness under any angular displacement of the shaft can be calculated.

According to Ref. [8], the length of duration of double-tooth-pair meshing as following description:

$$\begin{aligned} \theta_d &= \tan(\arccos \frac{N_1 \cos \alpha_0}{N_1 + 2}) - \frac{2\pi}{N_1} \\ &\quad - \tan(\arccos \frac{N_1 \cos \alpha_0}{\sqrt{(N_2 + 2)^2 + (N_1 + N_2)^2 - 2(N_2 + 2)(N_1 + N_2) \cos(\arccos \frac{N_2 \cos \alpha_0}{N_2 + 2} - \alpha_0)}}) \end{aligned} \quad (31)$$

The length of duration of single-tooth-pair meshing as following description:

$$\theta_s = \frac{2\pi}{N_1} - \theta_d \quad (32)$$

Double-tooth-pair Meshing Range of Duration [9]:

$$\theta \in [(n-1)\frac{2\pi}{N_1}, \theta_d + (n-1)\frac{2\pi}{N_1}](n = 1, 2, \dots) \quad (33)$$

Single-tooth-pair Meshing Range of Duration:

$$\theta \in [(n-1)\frac{2\pi}{N_1} + \theta_d, n\frac{2\pi}{N_1}](n = 1, 2, \dots) \quad (34)$$

For the single-tooth-pair meshing duration, the total effective mesh stiffness can be calculated as:

$$k_t = \frac{1}{\frac{1}{k_{h,1}} + \frac{1}{k_{b1,1}} + \frac{1}{k_{s1,1}} + \frac{1}{k_{a1,1}} + \frac{1}{k_{b2,1}} + \frac{1}{k_{s2,1}} + \frac{1}{k_{a2,1}}} \quad (35)$$

where subscripts 1 and 2 represent the driving gear and the driven gear, respectively.

For the double-tooth-pair meshing duration, there are two pairs of gears meshing at the same time. The total effective mesh stiffness can be obtained as:

$$k_t = k_{t1} + k_{t2} = \sum_{i=1}^2 \frac{1}{\frac{1}{k_{h,i}} + \frac{1}{k_{b1,i}} + \frac{1}{k_{s1,i}} + \frac{1}{k_{a1,i}} + \frac{1}{k_{b2,i}} + \frac{1}{k_{s2,i}} + \frac{1}{k_{a2,i}}} \quad (36)$$

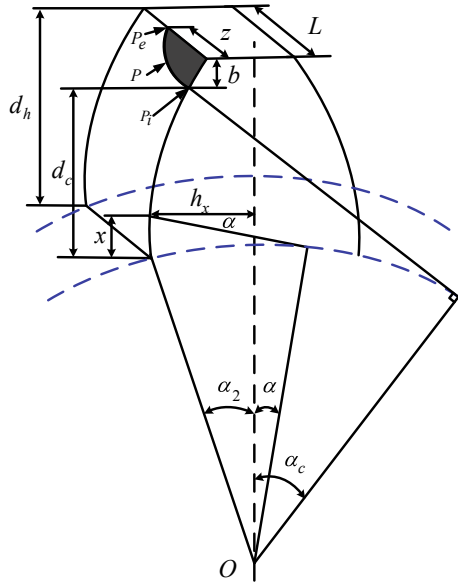
where $i = 1$ for the first pair and $i = 2$ for the second pair of meshing teeth.

3 Time-Varying Mesh Stiffness Calculation of Pinion with a Chipped Tooth

3.1 Mesh Stiffness Calculation of Pinion with a Chipped Tooth

It is assumed that there exists a chipped tooth in the pinion as shown in Fig. 2, where the shade area is the chipped part. It is assumed that the chip is so thin compared with the thickness of the tooth, that the change of the tooth thickness of the chipped part can be neglected when considering the bending stiffnesses, shear stiffnesses and axial compressive stiffnesses [10]. Therefore the bending stiffnesses, shear and axial compressive stiffnesses in this case can be approximately considered the same as those in perfect case. In the healthy condition, the width along the tooth curve is constant. In the chipped case as shown in Fig. 2, however, the width of the effective work surface will change as the contact position moves along the tooth curve. The effective work width at the roof of this tooth has the minimal value $L - z$. Then, the width will increase along the tooth curve from the tooth roof toward the root. The

Fig. 2 The geometry of a chipped tooth



width will change back to the normal width L at the position where the distance from the tooth roof is equal to b . It is assumed that the edge of the shade area satisfies the hyperbolic equation with an expression as $A = L(d_h - b) = Ld_c = d_h(L - z)$, Thereafter, the width will stay at the constant L . The formula for the change in width can be expressed as follows:

$$L_c = \begin{cases} L & 0 \leq d \leq d_c \\ \frac{A}{d} & d_c \leq d \leq d_h \end{cases} \tag{37}$$

where d_h is the tooth height (the distance between the roof and the root of the tooth). d_c and A can be expressed as:

$$d_c = R_{b1}[(\alpha_2 + \alpha_c) \sin \alpha_c + \cos \alpha_c - \cos \alpha_2] \tag{38}$$

Correspondingly, the Hertzian stiffness will proportionally vary as well. It can be calculated as:

$$k_{hc} = \frac{\pi E L_c}{4(1 - \nu^2)} \tag{39}$$

It is assumed that the chip occurs at the tooth that has a reference point and α_c (corresponding to d_c) small than $\alpha_{1,1}^0$, therefore the k_h just changes during the double-tooth-pair meshing range of duration. Substituting Eq. (37) into Eq. (39), the Hertzian contact stiffness of the external gear with a chipped tooth can be obtained:

$$k_{hc} = \begin{cases} \frac{\pi EL}{4(1-\nu^2)} & \text{other} \\ \frac{\pi E}{4(1-\nu^2)} \left(\frac{d_h(L-z)}{R_{b1}[(\alpha_2+\alpha)\sin\alpha + \cos\alpha - \cos\alpha_2]} \right) & \alpha_c \leq \alpha \leq \frac{2\pi}{N1} + \theta_d \end{cases} \quad (40)$$

3.2 Mesh Stiffness Calculation of Pinion with a Chipped Tooth

The relationship of the angular displacements of the meshing gears which contains a pinion with a chipped tooth is the same as in the healthy state. Similarly, mesh stiffness of double/single-tooth-pair meshing duration with a chipped tooth has the identical expression compared to in the healthy condition. Time-varying mesh stiffness of chipped condition changes because Mesh stiffness of pinion with a chipped tooth changes. Time-varying mesh stiffness of chipped tooth will be discussed as follows:

Similarly as healthy condition, for the single-tooth-pair meshing duration, the total effective mesh stiffness can be calculated as:

$$k_t = \frac{1}{\frac{1}{k_{h,1}} + \frac{1}{k_{b1,1}} + \frac{1}{k_{s1,1}} + \frac{1}{k_{a1,1}} + \frac{1}{k_{b2,1}} + \frac{1}{k_{s2,1}} + \frac{1}{k_{a2,1}}} \quad (41)$$

where subscripts 1 and 2 represent the driving gear and the driven gear, respectively.

Corresponding, for the double-tooth-pair meshing duration, there are two pairs of gears meshing at the same time. Assuming that the chipped tooth is the first pair mesh teeth, the total effective mesh stiffness can be obtained as:

$$k_t = k_{t1} + k_{t2} = \sum_{i=1}^2 \frac{1}{\frac{1}{k_{hc,i}} + \frac{1}{k_{b1,i}} + \frac{1}{k_{s1,i}} + \frac{1}{k_{a1,i}} + \frac{1}{k_{b2,i}} + \frac{1}{k_{s2,i}} + \frac{1}{k_{a2,i}}} \quad (42)$$

where k_{t1} for the first pair and k_{t2} for the second pair of meshing teeth.

4 Mesh Stiffness of Fixed-Shaft External-Internal Gears

Table 1 gives the parameters of a gear set. Both of the gears are standard involute spur teeth. The Time-varying mesh Stiffness of the external-external gear pair can be evaluated using the method developed in Ref. [8]. The result is shown in Fig. 3.

In this paper, assumed that the chip is $z = 8$ mm, $b = 2$ mm (as shown in Fig. 2). According to these parameters, the effective work contact width of the mating teeth can be obtained according to Eq. (37). The Hertzian contact stiffness k_h can be calculated by Eq. (40). The result of the Hertzian contact stiffness of a perfect gear and a gear with a chipped tooth is shown in Fig. 4. Point A corresponds to the initial contact position of the chipped part, which corresponds to the angular displacement

Table 1 Physical parameters of the gear set

Parameters	The driver gear	The driven gear
Gear type	Standard involute spur teeth	Standard involute spur teeth
Material	Steel	Steel
Number of teeth	19	31
Module (mm)	3.2	3.2
Pressure angle	20°	20°
Face width (m)	0.0381	0.0381
Young's modulus	2.068×10^{11}	2.068×10^{11}
Poisson's ratio	0.3	0.3

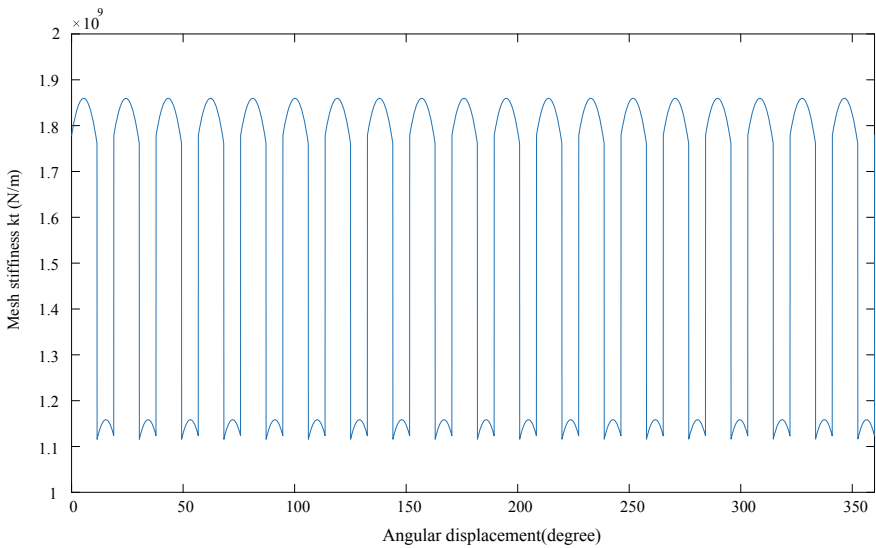


Fig. 3 The time-varying mesh stiffness of the external-external gear pair

of the driver α_c . From point A to point B, the Hertzian stiffness decreases with the decrease of the effective contact width L_c . Point C correspond to the transition point from double-tooth-pair mesh to single-tooth-pair mesh.

The corresponding change of time-varying mesh stiffness of a perfect gear and a gear with a chipped tooth is shown in Fig. 5. Instead of curve \widehat{DE} (which corresponding to the time-varying mesh stiffness of a perfect gear), the mesh stiffness will change as \widehat{DF} (which corresponding to the time-varying mesh stiffness of a gear with a chipped tooth) during teeth mating in the chipped part, which reflects the severity of the chipped part.

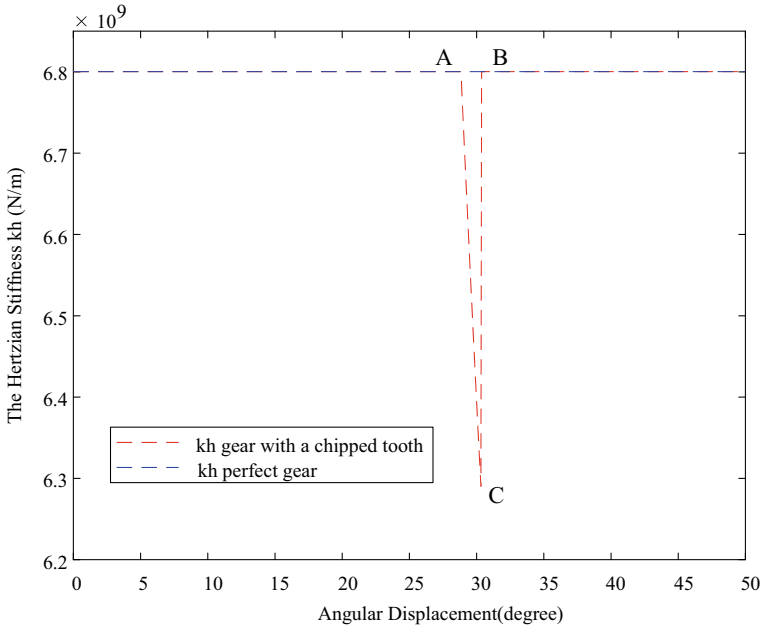


Fig. 4 The Hertzian contact stiffness of a perfect gear and with a chipped tooth

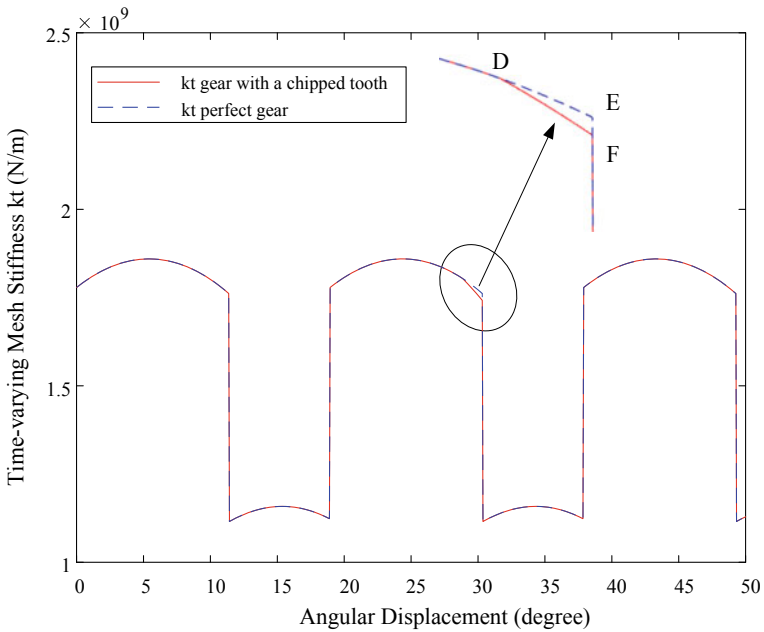


Fig. 5 Time-varying mesh stiffness of a perfect gear and a gear with a chipped tooth

5 Conclusions

This paper presents an analytical method for analyzing the effects of chipping fault on mesh stiffness of fixed gearbox. The time-varying mesh stiffness is represented as a function of the angular displacement of the gear, which is easy to use. In this study, Hertzian contact stiffness, bending stiffness, shear stiffness and axial compressive stiffness are considered. This study can be used to determine the quantitative relationship between mesh stiffness and gears with chipping fault. In future, the vibration signals of a gear set will be obtained for different chipping levels and chipping positions through dynamic simulation. Chipping severity and the chipping position indicators can be generated by analyzing the vibration signals.

References

1. Yang, D.C.H., Lin, J.Y.: Hertzian damping, tooth friction and bending elasticity in gear impact dynamics (1987)
2. Tian, X., Zuo, M.J., Fyfe, K.R.: Analysis of the vibration response of a gearbox with gear tooth faults. In: ASME International Mechanical Engineering Congress and Exposition 47063, pp. 785–793 (2004)
3. Liang, X., Zuo, M.J., Patel, T.H.: Evaluating the time-varying mesh stiffness of a planetary gear set using the potential energy method. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **228**(3), 535–547 (2014)
4. Yang, D.C.H., Sun, Z.S.: A rotary model for spur gear dynamics (1985)
5. Boresi, A.P., Schmidt, R.J., Sidebottom, O.M.: *Advanced Mechanics of Materials*. Wiley, New York (1985)
6. Shigley, J.E., Shigley.: *Applied Mechanics of Materials*. McGraw-Hill, New York (1976)
7. Tian, X.: *Dynamic Simulation for System Response of Gearbox Including Localized Gear Faults*. University of Alberta (2004)
8. Litvin, F.L.: *Gear geometry and applied theory*. Prentice-Hill, Englewood Cliffs, nJ (1994)
9. Liang, X.: *Dynamics based vibration signal modeling and fault detection of planetary gearboxes* (2016)
10. Li, G., Liang, X., Li, F.: Model-based analysis and fault diagnosis of a compound planetary gear set with damaged sun gear. *J. Mech. Sci. Technol.* **32**(7), 3081–3096 (2018)

Damage Diagnosis of Railway Vehicle Car Body Based on Strain Modes



Hui Cao, Gangjun Li, and Fengshou Gu

Abstract In the rail vehicle structure design, the key of the strength design and fatigue life estimation is to analyse the stress state of the structure under the dynamic load. As the stress cannot be measured directly, the displacement response model of the body structure is established by the displacement mode analysis method, and the strain response of the car body is obtained by the relationship between displacement and strain, thus the stress state is obtained. Since displacement to strain is a differential process, the variation of displacement will be magnified and the error will be generated. Strain mode theory and its property are derived from displacement mode theory. The results show that it is more sensitive for strain mode than displacement mode through the simulation analysis of the car body equivalent vertical model calculation. Strain mode difference curve can determine the structural damage location. The vehicle FE model verifies this result. The strain and stress versus time history of car body can be obtained by the mode superposition method, which provides basis for fatigue life prediction and load spectrum research.

Keywords Railway Vehicle · Car body · Damage diagnosis · Strain mode · Strength

1 Introduction

In order to ensure the safety and reliability of high-speed Electrical Multiple Units (EMU), the structure strength analysis and fatigue life prediction are the important contents in structural design and health monitoring. Strength design and fatigue life prediction are related to structure stress status. The structure stress cannot be obtained

H. Cao · G. Li

College of Intelligent Manufacturing, Chengdu Technological University, Chengdu 611830, China
e-mail: ch_hello@163.com

F. Gu (✉)

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK
e-mail: F.Gu@hud.ac.uk

directly through measurement but indirectly through displacement or strain measurement because there are the relationship between stress, strain and displacement. The displacement mode analysis method can be used to establish the displacement response model of EMU structure and analyze the structural vibration response. And then, the strain mode of vehicle response is obtained by displacement-strain calculation. However, it is a differential process from displacement to strain, and the error of displacement will be amplified [1].

Therefore, many scholars have studied to apply the displacement mode method to the strain mode, and obtain the structural strain mode. Hillary et al. measured the force-strain transfer function with resistance strain gauge and proposed the concept of strain mode [2]. Li [3], Bernasconi [4] and Yam et al. [5] used the differential operation method of displacement mode to deduce and discuss the strain mode theory. Tsang [6], Li et al. [7] used the finite element method to verify the strain mode theory, and compared it with the calculation simulation and experimental test. Li et al. [7–9] also demonstrated the orthogonality of the strain mode. These studies are focused on the construction of strain mode theory, but it is rarely applied to the damage identification and fault diagnosis of complex structures.

In this paper, the equivalent vertical model of a type of EMU is used to simulate and compare the displacement mode and strain mode. So the local damage or fatigue position of the structure is determined, and the fatigue crack size is determined according to the damage degree. With the strain mode of the car body structure, the strain of all structure, such as the strain near the window, can be studied, which provides the basis for the study of structural stress concentration, the influence of local structural improvement on the vicinity of the changing region, and the health detection of car body structure.

2 Strain Mode Theory

In the three-dimensional coordinate system, a displacement vector is set as,

$$\mathbf{x} = (\mathbf{u}, \mathbf{v}, \mathbf{w})^T \quad (1)$$

where \mathbf{x} is the displacement vector, $\mathbf{u}, \mathbf{v}, \mathbf{w}$ is the displace vector in u, v, w directions. The i th displacement mode is

$$\boldsymbol{\varphi}_i = (\varphi_i^u, \varphi_i^v, \varphi_i^w)^T \quad (2)$$

the displacement response can be expressed through the superposition of vibration modes and mode coordinates of each order, namely,

$$\mathbf{x} = (\mathbf{u}, \mathbf{v}, \mathbf{w})^T = \sum_{i=1}^n q_i (\varphi_i^u, \varphi_i^v, \varphi_i^w)^T = \boldsymbol{\Phi} \mathbf{q} \quad (3)$$

where Φ is the displacement mode shape, \mathbf{q} is the mode coordinates. Assuming the strain vector is

$$\mathbf{E} = (\boldsymbol{\varepsilon}_x, \boldsymbol{\varepsilon}_y, \boldsymbol{\varepsilon}_z)^T \tag{4}$$

The i th strain mode is

$$\boldsymbol{\Psi}_i^\varepsilon = (\psi_i^u, \psi_i^v, \psi_i^w)^T \tag{5}$$

Among them,

$$\boldsymbol{\varepsilon}_x = \frac{\partial u}{\partial x} = \frac{\partial}{\partial x} \sum_{i=1}^n \boldsymbol{\varphi}_i(x) q_i \tag{6}$$

Because q_i is a function related to time and its partial derivative to position x is 0, the above equation can be written as

$$\boldsymbol{\varepsilon}_x = \sum_{i=1}^n q_i \frac{\partial \boldsymbol{\varphi}_i(x)}{\partial x} = \sum_{i=1}^n q_i \boldsymbol{\Psi}_i^\varepsilon(x) \tag{7}$$

The formula $\boldsymbol{\Psi}_i^\varepsilon(x) = \frac{\partial \boldsymbol{\varphi}_i(x)}{\partial x}$ is called strain mode. Writing the above formula as a discrete form and getting the following result,

$$\boldsymbol{\varepsilon}_x = \sum_{i=1}^n q_i \boldsymbol{\Psi}_i^\varepsilon(x) = \sum_{i=1}^n \frac{\boldsymbol{\Psi}_i^\varepsilon \boldsymbol{\varphi}_i^T}{k_i - \omega^2 m_i + j\omega c_i} \mathbf{F} e^{j\omega t} \tag{8}$$

Similarly, the strain mode expression in other directions can be obtained, which is the same as formula (8). According to the superposition principle, the strain response can be written as

$$\mathbf{E} = (\boldsymbol{\varepsilon}_x, \boldsymbol{\varepsilon}_y, \boldsymbol{\varepsilon}_z)^T = \sum_{i=1}^n q_i (\psi_i^u, \psi_i^v, \psi_i^w)^T = \sum_{i=1}^n \boldsymbol{\Psi}_i^\varepsilon q_i \tag{9}$$

From formula (8), the strain mode response matrix is

$$\mathbf{H}^\varepsilon = \sum_{i=1}^n \frac{\boldsymbol{\Psi}_i^\varepsilon \boldsymbol{\varphi}_i^T}{k_i - \omega^2 m_i + j\omega c_i} \tag{10}$$

It can be seen from the above formula that the characteristics of the strain frequency response function matrix are as follows.

- (1) because of $H_{ij} \neq H_{ji}$, the strain frequency response function matrix is an asymmetric matrix.

- (2) Each element in the matrix contains c_i, m_i, k_i information.
- (3) Each row of the matrix contains all information of the displacement mode, and each column of the matrix contains all information of the strain mode. Therefore, for the strain mode analysis, if you want to get all the mode parameters of the displacement mode and the strain mode, you must measure one line and one column of H^ϵ .

The ultimate goal of obtaining the structural strain mode is to carry out stress analysis. After obtaining the strain mode, the mode stress can be obtained according to Hooke’s law, and the stress response can be calculated by mode superposition method. According to the stress-strain relationship of elastic mechanics

$$\begin{aligned}
 \sigma_x &= \lambda\theta + 2G\epsilon_x, \quad \tau_{xy} = G\gamma_{xy} \\
 \sigma_y &= \lambda\theta + 2G\epsilon_y, \quad \tau_{xy} = G\gamma_{yx} \\
 \sigma_z &= \lambda\theta + 2G\epsilon_z, \quad \tau_{zx} = G\gamma_{zx}
 \end{aligned}
 \tag{11}$$

where, θ —volume strain, λ —lame constant, G —Shear modulus.

Using the strain mode superposition method,

$$\theta = \epsilon_x + \epsilon_y + \epsilon_z = \sum_{i=1}^n q_i (\psi_i^{\epsilon_x} + \psi_i^{\epsilon_y} + \psi_i^{\epsilon_z}) = \sum_{i=1}^n q_i \Theta_i \tag{12}$$

The above formula is brought into formula (11), obtained

$$\sigma_x = \lambda \sum_{i=1}^n q_i \Theta_i + G \sum_{i=1}^n q_i \psi_i^{\epsilon_x} = \sum_{i=1}^n q_i (\lambda \Theta_i + G \psi_i^{\epsilon_x}) = \sum_{i=1}^n q_i \Xi_i^x \tag{13}$$

where, Ξ_i^x — x direction stress mode of the i th, q_i —the i th mode coordinates, Similarly, y, z direction stress modes can be obtained.

3 Strain Mode Analysis of EMU Car Body

The car body structure is welded by large and light hollow aluminum alloy. In order to meet the needs of the structural design, the windows, doors and other structures be installed. Therefore, these locations are prone to stress concentration and structural fatigue. Because the strain mode is more sensitive to the stress change and local damage caused by local structural changes than the displacement mode, the numerical simulation method is used to analyze this.

According to the structural parameters of an EMU car body, just considering the vertical equivalent section stiffness and mode, according to the principle of equivalent of vertical stiffness and mode frequency, the car body is equivalent to an equal

section free homogeneous Euler-Bernoulli beam with rectangular section. After the equivalent treatment of the car body, the section size is as follows, the width of the beam is 0.32 m, the height is 1.25 m, and the length is 24.5 m. The basic properties of car body material are as follows, Elastic modulus, Poisson's ratio and density are 69 GPa, 0.3, 2700 kg/m³, respectively.

The numerical simulation results of displacement mode and strain mode are obtained when the structure has no damage and middle positions of the equivalent beam has damage respectively. The damage treatment process of the model beam is as follows, assuming that defect groove on the upper surface of the model beam is equaled to the width of the model, the length of the defect groove is 5 mm, and the depth of the defect groove is 5%, 10%, 15%, 20% and 0% of the height of the model beam respectively, as shown in Fig. 1.

The 8-node solid, 185 element model is used to establish finite element by ANSYS software. The width mesh number of the beam is 4, the vertical mesh number is 13, and the longitudinal is divided into 245 elements. The finite element model of the beam is shown in Fig. 2, which is divided into 17,220 nodes and 12,740 elements. Through the finite element post-process, the mode displacement and mode strain of

Fig. 1 Diagram of the damage of car body structure

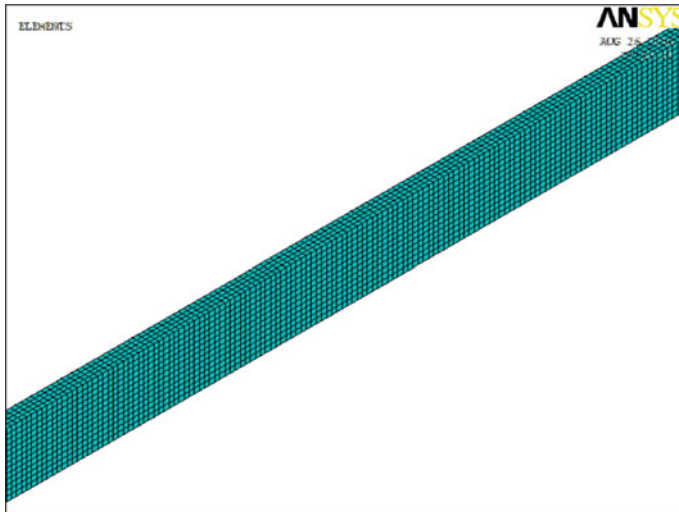
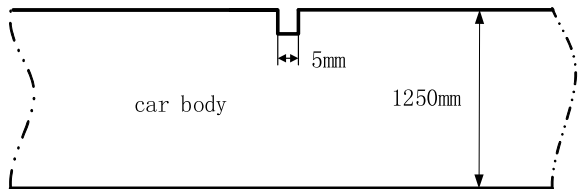


Fig. 2 Finite element analysis model

Table 1 The mode frequency comparison of car body model

order	Mode frequency without damaging	Mode frequency with 10% damage	Frequency ratio	Mode frequency with 30% damage	Frequency ratio (%)
1	10.755	10.695	99.44%	10.185	94.7
2	29.159	29.159	1	29.157	99.9
3	55.858	55.627	99.6%	53.786	96.8

each node of the model beam can be extracted, and the displacement mode and strain mode of the model beam can be obtained.

The frequency is the inherent characteristic of the structure. The first three mode frequency of the equivalent model of the car body is as shown in Table 1. From Table 1, the mode frequency changes little before and after the damage, so the frequency can't be used as the index of damage diagnosis, and the damage location can't be determined by the frequency.

Therefore, in the case of the first three displacement modes and strain modes under the damage amount of 0, 5, 10, 15, 20 and 30% are extracted, as shown in Figs. 3 and 4. It can be seen from the figure that the mode shapes of displacement modes are basically unchanged in the case of intact and damaged, that is, the sensitivity of displacement modes to damage is low. Therefore, the damage of location and degree cannot be determined by the displacement vibration mode. In the case of intact and damaged strain mode, the change of vibration mode shape is obvious, especially the first-order and third-order mode shape is significant. The second-order strain mode change is not obvious because the second-order mode node is in the middle position. So, the strain mode is sensitive to damage, but if the damage position is the mode node, the sensitivity will decrease.

In order to obtain the damage location, the variation curve of strain mode change rate can be obtained by applying the derivative, and the damage location can be determined from the strain mode change rate curve. Because the strain mode curve is not smooth, the difference method in numerical calculation is a method to calculate the rate of change of discrete value.

Using numerical difference calculation method of the strain mode curve, the strain mode difference curve is shown in Fig. 5. It can be seen from the figure that the strain mode curves of no damage structure change continuously, and the difference curve is also the same. But in damaged structure, the strain mode curve will have a sudden change at the damage position, and the difference curve will change drastically at the damage position. For the damage at the mode nodes, the strain mode difference curves will also change sharply. The difference curves for each damage amount almost intersect at the same point, which is the damage point. So the strain mode difference value can be used as a judgment index for damage location.

The simulation analysis results using the equivalent model of the car body show that the strain mode and its differential curve can be used as an effective method for car body structure strength calculation and damage location diagnosis. However,

Fig. 3 Mode shape of displacement mode

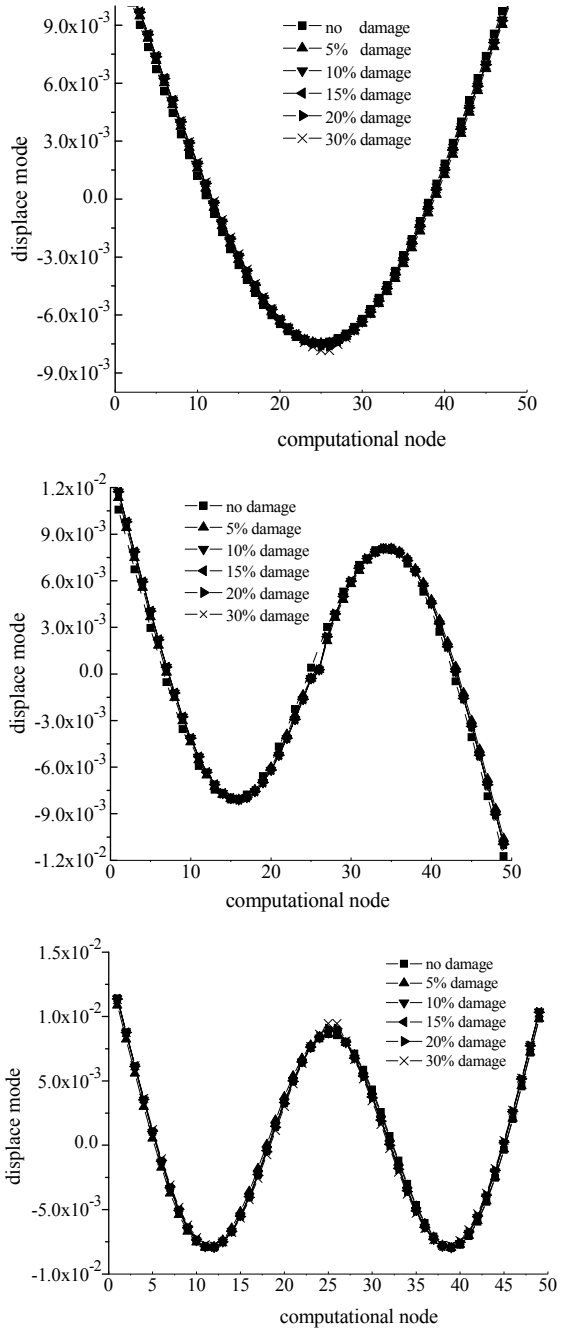
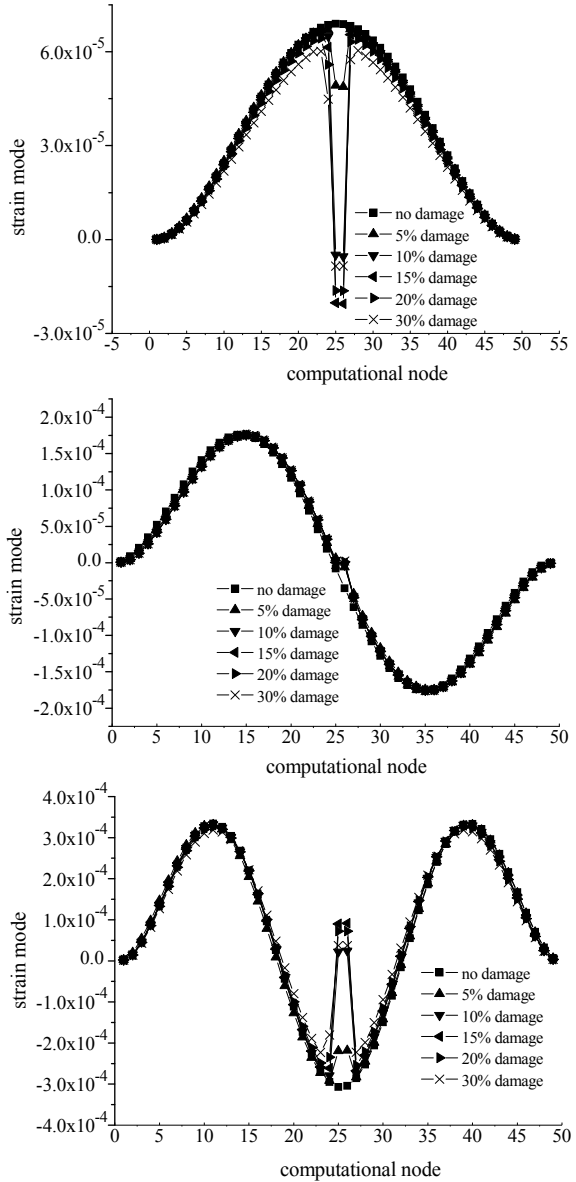
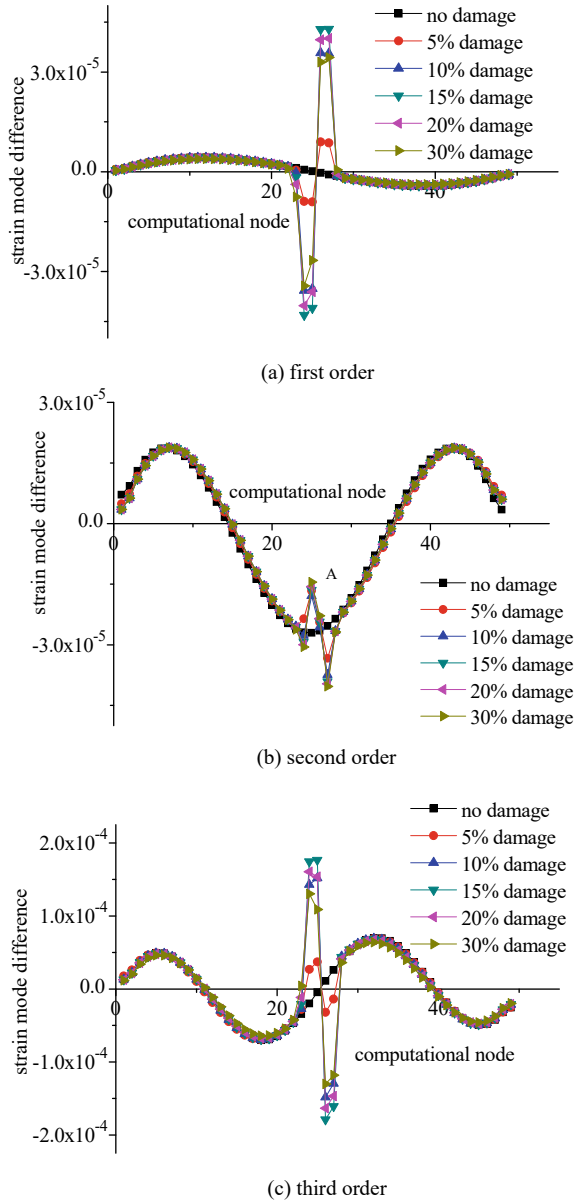


Fig. 4 Strain mode shape



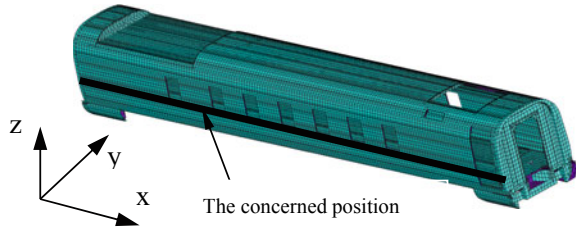
the car body is actually a complex welded structure, and its strength calculation and damage location diagnosis are more complicated than the equivalent model. Therefore, the finite element model of the car body structure is used to extract the displacement mode and strain mode, and the strain mode method is used to analyze the strength of the car body structure.

Fig. 5 Differential curve of the strain mode



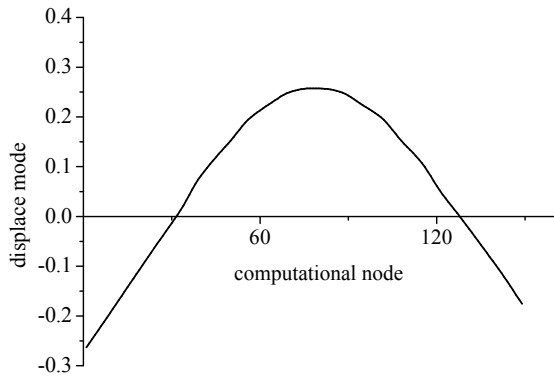
The car body of high-speed train is mainly welded by large hollow and wide aluminum alloy. It consists of the under frame, roof, end walls and side walls. Figure 6 is a finite element model of a certain type of EMU car body. The mesh of the model is divided into 49,503 nodes and 82,527 elements.

Fig. 6 Car body finite element model

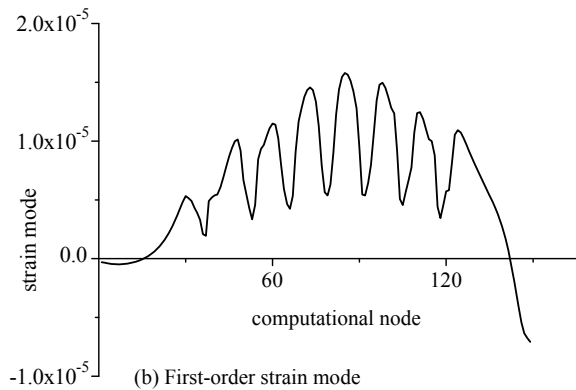


Extracting the first-order vertical bending displacement mode and strain mode is extracted, which is shown in Fig. 7. It can be seen from the figure that the displacement mode curve continuously changes without obvious change, indicating that the displacement mode of the car body structure is not sensitive to structural changes such as windows, and the structural strength calculated by the displacement mode is not reflect the influence of the window structure of the car body. However, the strain mode curve at the position of the window has changed, indicating that the window structure has a great influence on the strain of the car body structure. The

Fig. 7 The first-order vertical bending displacement mode and strain mode of car body



(a) First-order displacement mode



(b) First-order strain mode

strain changes alternately in the length direction of the car body. It can be seen from formula (13) that the car body will have stress changes at this position, and structural defects such as stress concentration and insufficient strength will easily occur, the attention should be paid to the structural design.

4 Conclusion

Using the equivalent model and the finite element model of the car body, the displacement mode and strain mode of the car body are analyzed, and the application of the strain mode in strain and stress acquisition is studied. The main conclusions of body strength design and damage diagnosis are as follows,

- (1) Strain mode and displacement mode are two manifestations of the same energy balance state. Strain mode can also be solved by superposition of strain mode shape and mode coordinate.
- (2) Compared with the displacement mode, the strain mode is more sensitive to structural damage. This conclusion is verified by the car body equivalent model and the finite element model. The strain mode difference curve can assist in determining the location of structural damage.
- (3) For complex structures such as car bodies, the strength design is an important part of the structural design. The strain and stress calculation methods of the car body based on the strain mode have certain applicability and feasibility, which can be used for car body fatigue strength analysis and load spectrum research.

References

1. Hui, C.: High-speed EMU car body vibration control and strain modal analysis. Southwest Jiaotong University, Chengdu (2016)
2. Hillary, B., Ewins, D.J.: The use of strain gauges in force determination and frequency response function measurements. In: Proceedings of the 2nd International Modal Analysis Conference and Exhibit, vol. 2, pp. 6–9 (1984)
3. Li, D., Zhuge, H., Wang, B.O.: The principle and techniques of experimental strain modal analysis. In: 7th International Modal Analysis Conference, pp. 1285–1289 (1989)
4. Bernasconi, O., Ewins, D.J.: Application of strain modal testing to real structures. In: 7th International Modal Analysis Conference, pp. 1453–1464 (1989)
5. Yam, L.Y., Leung, T.P., Li, D.B., et al.: Theoretical and experimental study of modal strain analysis. *J. Sound Vib.* **191**(2), 251–260 (1996)
6. Tsang, W.F.: Use of dynamic strain measurements for the modelling of structures. In: 8th International Modal Analysis Conference, pp. 1246–1251 (1990)
7. Li, D., Zhang, Y.: Modal method for dynamic strain/stress field analysis. *Vibr. Shock* **11**(4), 15–22 (1992)

8. Li, D.: Review on the progress of experimental strain/stress modal analysis. *Vibr. Shock* **15**(1), 13–17 (1996)
9. Lu, Q., Li, D.: Progress in modal theory. *Adv. Mech.* **26**(4), 464–472 (1996)

A Chip Defect Detection System Based on Machine Vision



Xindan Qiao, Ting Chen, Wanjing Zhuang, and Jinyi Wu

Abstract For chip testing, in the process of the actual chip manufacturing since most of the chip size is very small, so the artificial extremely difficult to discern defect signals, such as lack of pin, pin bending, surface defects such as scratches, lack of the shape signal, thus easy to cause the yield is not ideal, therefore in the process of actual production introduction of machine vision. Chip defect detection system based on machine vision is a kind of machine vision, chip bearing platform, automatic rotating disc, etc., on the basis of combining computer terminal to control the whole test system, in view of the chip pins, surface, shape features such as visual algorithm analysis, finally through the man–machine interface technology of motion control system and chip testing results show that the Finally, the system is made and the best detection state is debugged. It can improve the yield of products and improve the production efficiency in the actual manufacturing process.

Keywords Machine vision · HALCON · Defect detection · Chip detection · The man–machine interface

1 Introduction

Because the chip has been developed into the economic lifeline to maintain information security in our country, and chip detection is a process of chip design and manufacturing, occupies a pivotal important position, and because the image processing technology is increasingly perfect and the production process requires a more effective, more accurate and more high-speed detection means. Therefore, we will conduct research on chip defect detection based on machine vision.

X. Qiao · T. Chen (✉) · W. Zhuang · J. Wu
Beijing Institute of Technology, Zhuhai, China
e-mail: 1621474110@qq.com

X. Qiao
e-mail: 564961500@qq.com

College of Industrial Automation, Zhuhai 519088, Guangdong, China

There are three main methods for chip appearance detection: traditional artificial vision is one-to-one detection method; The method of detecting the appearance of the chip by laser measurement technology and returning the measurement results of the chip; Automatic detection method based on machine vision.

This study designed a machine vision technology to replace the chip production and application of human eye measurement and detection process, and has the advantages of high efficiency, rapidity, accuracy and high reliability; At the same time, the design of the system structure is simple and easy to maintain, practicability and usability is very strong; Do not directly contact the chip workpiece, with the advantages of non-contact and flexibility; Finally, a stable result is achieved after many experiments. It is a practical, mechanized and intelligent chip defect detection system with high layers.

2 Related Technology Introduction

2.1 Machine Vision Stand

The mechanical vision frame is divided into four parts: the base, the light source frame device, the CCD frame device and the connecting device. The bottom plate part is composed of a base cover, a bottom plate and a long rod. The long rod and the base shell is fixed, the long rod is connected by a hand screw-cross fixing clamp on the rod, the long rod and the base shell is fixed, the long rod is connected by a hand screw-cross fixing clamp on the rod; The light source frame device is composed of the light source frame, the light source frame side frame and the rod, the light source frame is connected with the rod, the light source frame side frame is connected by the fixing block and the light source frame long frame; The CCD frame device is composed of a CCD frame, a CCD frame rubber clip block, a CCD frame long bolt and a CCD frame nut. CCD frame long bolt is connected with the CCD cross frame, the bolt ends are connected with the nut and rubber clamp block, by adjusting the spacing to clamp the camera; The connecting device is composed of a hand twist cross fixing clip, a hand twist cross fixing clip short bolt and a hand twist cross fixing clip short bolt cap. The hand twist cross fixing clip is connected with a long rod and a rod, which can adjust the height and expansion length of the camera and the light source. The bolt cap and the short bolt can adjust the tightness of the hand twist cross fixing clamp, and play the role of moving and clamping the bar. The explosion diagram of the machine vision frame is shown in Fig. 1.

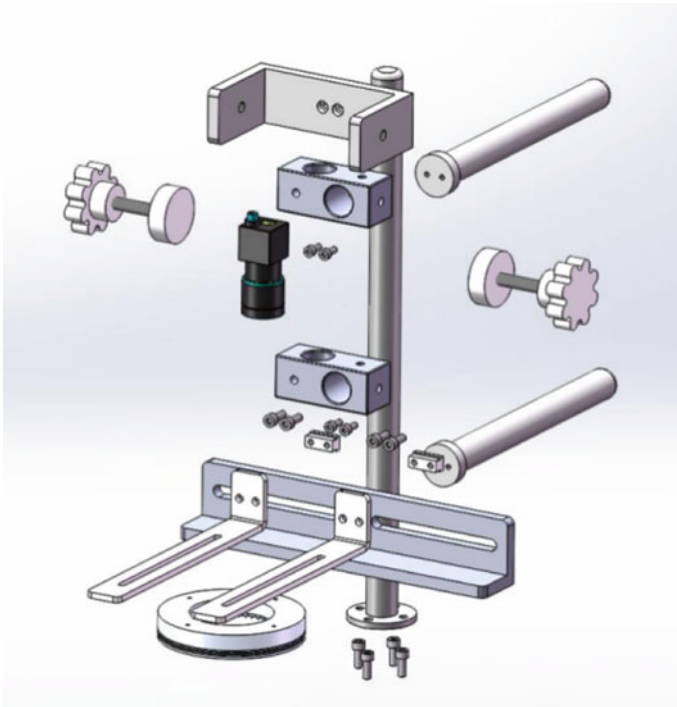


Fig. 1 Visual frame explosion

2.2 HALCON

HALCON is designed by a German company is very widely used in the world of a set of machine vision software, integrated with a very perfect development environment, is a standard comprehensive standard software library.

It has a set of image processing library, by more than one thousand independent operators, as well as the underlying data management core; It includes all kinds of filtering, color analysis and geometry, mathematical change, morphological calculation and analysis, correction, classification, identification, shape search and other basic geometry and image computing functions [1].

2.3 Template Matching

Template matching is a basic method, and the algorithm has its own technical limitations. It can only carry out a parallel rotation of the target itself. If the target in the original image cannot undergo these changes, the matching algorithm will be ineffective.

Since the camera is fixed in this design and the chip bearing platform drives the chip to move in the same plane, the shape defect of the chip can be detected by template matching.

2.4 Blob Analysis

The method of BLOB image analysis is to extract and analyze a connected domain of a two-dimensional numerical waveform image separated from the foreground/image and background region. A BLOB that marks each goal you want to accomplish can represent a new vision related goal [2].

2.5 Man–Machine Interface Development

The user interface or the user interface is called the man–machine interface, which is the medium of information exchange and transmission between human and computer [3, 4]. The interaction between human and a machine containing a computer is called human–computer interaction, and the part visible to users is called human–computer interaction interface [5]. Human–machine interaction interface is an important part of computer system often used in industrial production.

3 The System Design

3.1 Overview of Overall System Design

The whole system consists of three parts: image acquisition unit, optical system, image processing unit and motion control unit.

The image acquisition unit includes an optical system and a visual system. The visual system includes a light source, a light source controller, a lens and an industrial camera, etc. The optical system consists of different types of light sources, lenses and cameras. The image processing unit uses Halcon software for image processing and Visual Studio software for man–machine interface development to process the collected images. The motion control unit mainly includes the control computer (PC) and the transmission device automatic rotation plate and so on. The overall scheme design diagram of the system is shown in Fig. 2.

In this project, our group uses Haikang gigabit network port industrial array camera as the image acquisition tool. A chip bearing platform suitable for various types of chips is designed. The motor is used as the power unit to control the rotation of the chip bearing platform. The industrial camera collects the chip in the process of

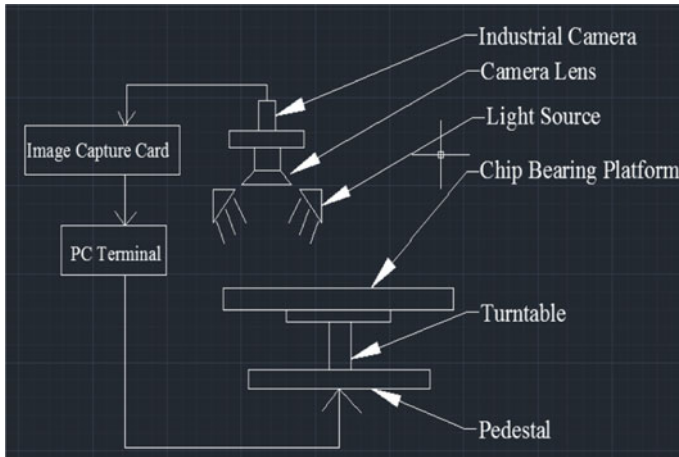


Fig. 2 Overall scheme design

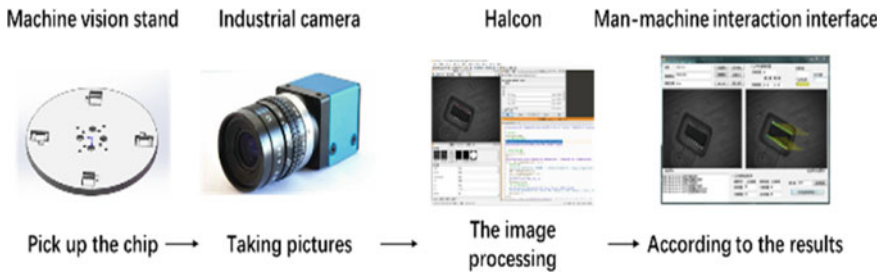


Fig. 3 System work flow

rotation. After the pictures are collected by the industrial camera, the machine vision algorithm is calculated by Halcon, and then displayed by the human-computer interaction interface. Finally, the chip measurement and detection effect are realized. The working process of the system is shown in Fig. 3.

3.2 The Principle and Flow Chart of Chip Defect Detection

Chip defect detection based on HALCON process, first to imaging of objects or areas of interest, then according to the image information with image processing software for processing, automatic judgment according to the results of the treatment location, size, appearance of object detection, and based on artificial pre-set standards for qualified or not judgment, output the judgment information to the actuator.

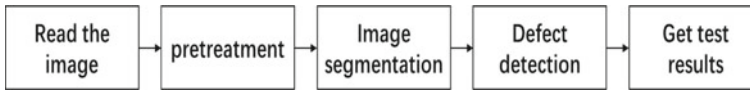


Fig. 4 Flow chart of chip defect detection

Machine vision inspection system USES the camera will be the object of information into image signal, and send it to the dedicated image processing system, according to the pixel distribution and brightness, color and other information, into a digital signal, image processing system then these signals are all kinds of operation to extract the characteristics of the object, such as size, number, location, length, Then output results according to preset values and other conditions, including size, Angle, number, qualified/unqualified, yes/no, etc., to achieve automatic identification function [6].

The chip defect detection flow chart is shown in Fig. 4. The chip defect detection is gradually realized. It can be divided into the following parts: image reading, preprocessing, image segmentation, defect detection and finally obtaining the detection results.

3.3 *The Realization of Chip Surface Scratch Defect Detection.*

Each step of the Halcon based chip surface scratch defect detection is also performed by a large number of operators provided by Halcon internally, which greatly simplifies the time required for the detection.

In the design process, Blob analysis method and feature method are used to detect the surface scratch defect of the chip. The detection process is shown in Fig. 5.

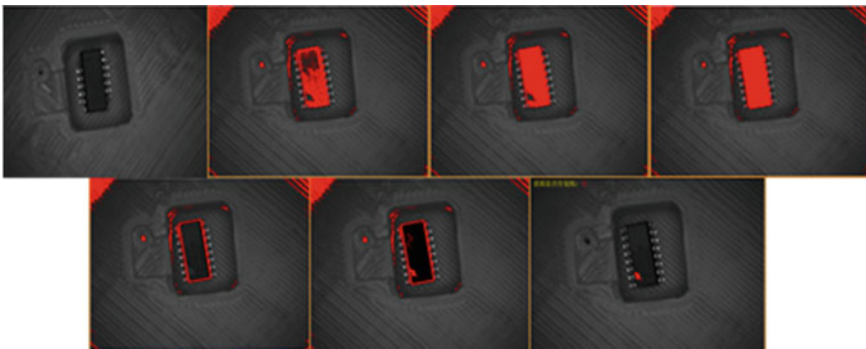


Fig. 5 Surface scratch test process

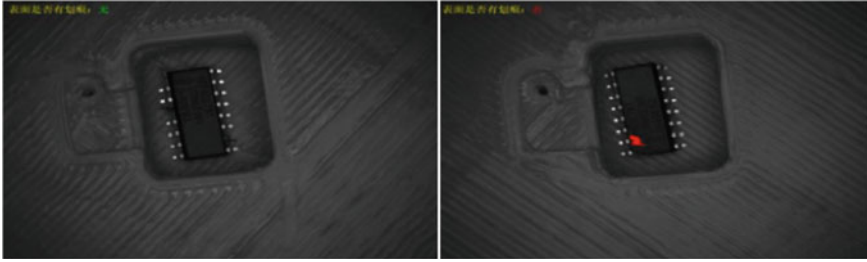


Fig. 6 Surface scratch results

After the image is read, the image is preprocessed. The preprocessing process includes four steps: (1) Threshold segmentation. By threshold segmentation of the image, the chip region is preliminarily selected; (2) The edge of the system is filled to fill the inside of the chip; (3) To calibrate the chip region, select the region with the largest area through the discrete region, that is, the chip region, and then calculate the smallest outer rectangle of the chip region to calibrate the chip region; (4) Select the chip area in the box.

On the basis of preprocessing, the average filtering of the image can smooth the image and remove the influence of characters, and then the original image and the filtered image are differentiated to segment the scratch area. Finally, after filling the scratch cavity area, the number of scratches is calculated to determine whether there are scratches. If there are scratches, the position of the scratches will be marked and the words “Pin defects: Yes” will be displayed. Otherwise, the words “Pin defects: No” will be displayed. Figure 6 shows the Surface scratch results.

3.4 The Realization of Chip Pin Defect Detection

The design process of chip pin defect detection algorithm based on Halcon adopts a combination of measurement and fitting method to detect the flaw of chip pin. The detection process is shown in Fig. 7.

Compared with the detection process of chip surface scratch defects, the detection of pin defects requires the calibration of the area where the chip pin is located, that is, after the calibration of the chip position is completed, an external rectangle including the chip pin is added, and the difference between the two rectangles is obtained to get the area of the chip pin, and the calibration of the pin area is carried out.

After the pin region was calibrated, the pin regions on both sides were calibrated as two regions (a) and (b) respectively. First, region (a) was selected and image segmentation was performed on this region to obtain the specific location of the chip pin. The same processing was performed on region (b). After finding the specific position of pins, the number of pins can be calculated, and then the threshold value is fixed. Whether the pins are bent or not can be determined by the difference between

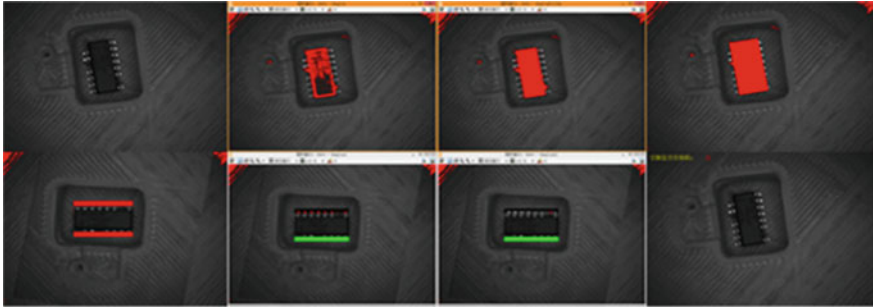


Fig. 7 Pin defect detection process

the threshold value of bent feet and that of normal feet. Finally, by judging whether the number of pins is qualified and whether the pins are bent to determine whether the chip has pin defects. If there are defects, the words “Pin defects: Yes” will be displayed; otherwise, the words “Pin defects: No” will be displayed.

3.5 The Realization of Chip Shape Defect Detection

Halcon based chip shape defect detection in the design process using the template matching method to detect the chip shape. The detection process is shown in Fig. 8.

On chip shape defect detection in the process of the need for image segmentation, just the chip area of the calibration and the actual chip area difference (i.e., the chip area filling), through the difference of value are to judge whether the shape of the chip



Fig. 8 Profile defect detection process

is missing, if there is a missing form is displayed “Shape defects: Yes” the words, the opposite shows “Shape defects: No”.

4 System Implementation

The realization of the whole system needs hardware and software. The hardware includes industrial camera, industrial lens, light source and light source controller, etc. The software includes SolidWorks software to build the three-dimensional model of the system, Visual Studio software to develop the man-machine display interface, MVS client software to debug the industrial camera and Halcon software algorithm processing, etc. The combination of software and hardware is able to detect the unqualified parts of the chip intelligently, automatically and efficiently. Figure 9 shows the structure diagram of the system.

4.1 Hardware Platform Implementation

The industrial camera and lens are fixed on the machine vision frame, and the computer (PC) end and display are installed. The industrial camera and computer communicate with each other using gigabit network cable and the MVS client of the industrial camera. See Fig. 10.

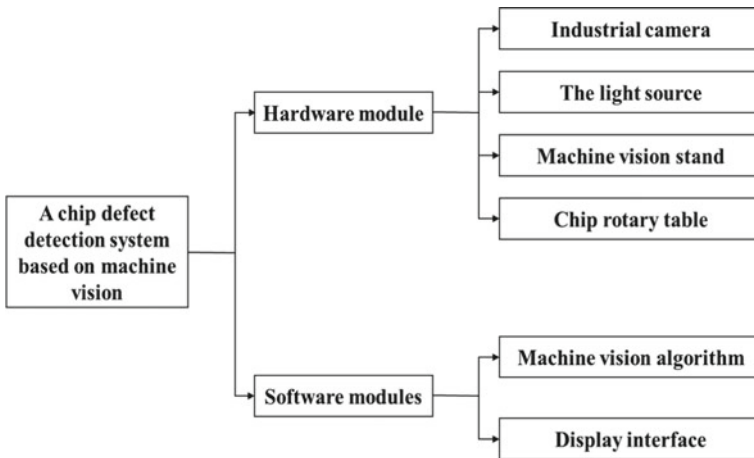


Fig. 9 Structure diagram of the system

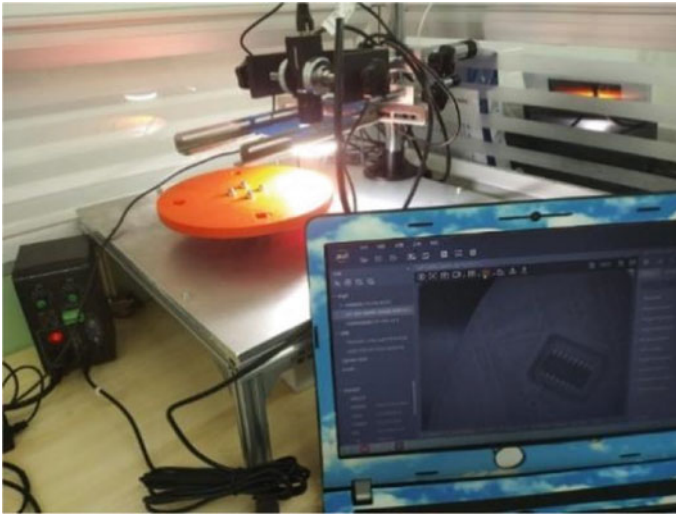


Fig. 10 System hardware platform

4.2 The Realization of Software System

The realization of human–computer interaction interface

Based on Visual Studio software, C# language and Winform function are used to design the overall system man–machine interface. The designed man–machine interface of the chip detection system is shown in Fig. 11. In the figure, the detection of a

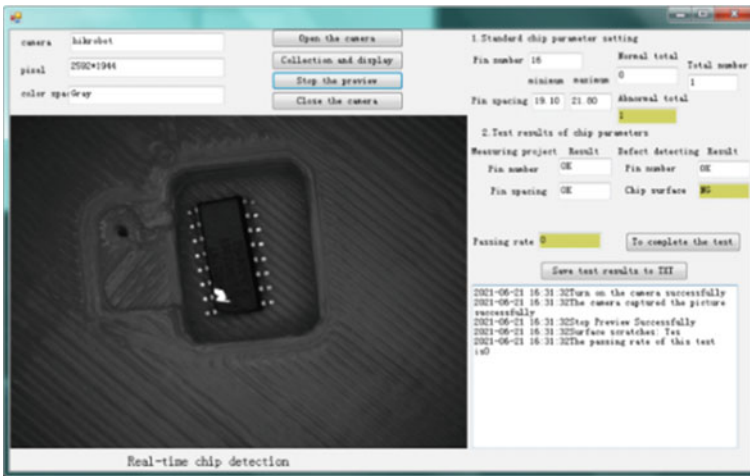


Fig. 11 System machine vision control

chip surface scratches unqualified chip, through the detection interface can be seen the detection of unqualified items, pass rate and log and other specific conditions.

The realization of image acquisition module

The image acquisition module uses the data interface provided by Halcon to open the camera, process the image and detect the chip in real time. Halcon provides the universal dynamic link library halcondotnet.dll. Based on this, the dynamic dynamic link library is called to open the camera in real time, process the pictures of the camera chip in real time and display the processing results. As shown in Fig. 12, it specifically includes turning on the camera, collecting and displaying, stopping the preview and closing the camera, etc., as well as the functions of collecting and processing images by directly calling the processing and displaying function of images on the PC end. The processed pictures are displayed in the window detected by HWindowControl chip in real time through button controls corresponding to different processing algorithm events.

The realization of defect detection module

The realization of the detection module of the man-machine interface is shown in Fig. 13, which is divided into three parts: setting the standard parameters of the chip, testing results of the chip parameters, and saving the log.



Fig. 12 System Image acquisition module

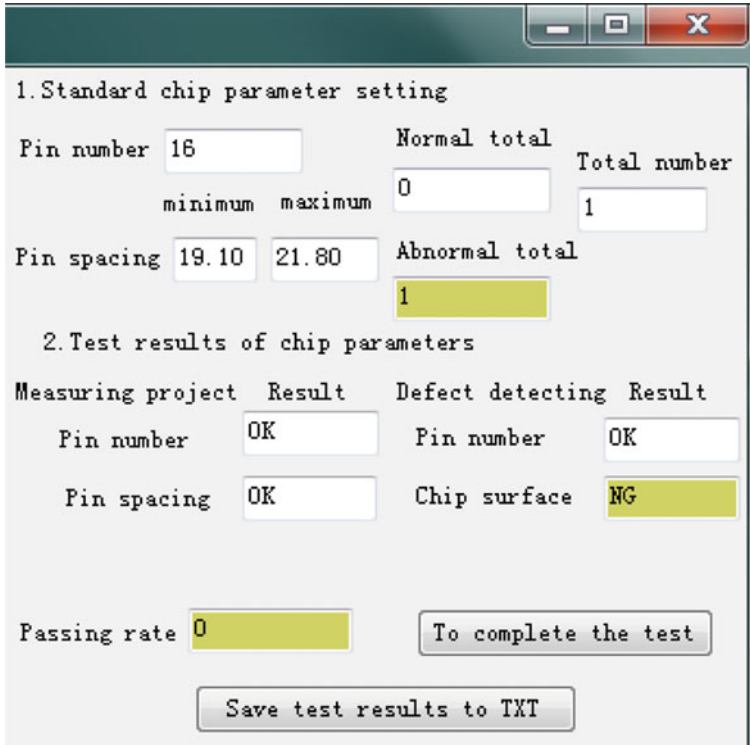


Fig. 13 System vision calibration module

The parameters of the chip standard in the first part are output variable results after processing by the Halcon algorithm, including the maximum and minimum values of the number of chips and the spacing of pins.

The second part of the test results, by measuring the number and spacing of pins in the project to determine whether there is a pin defect. At the same time, through the number of pins in defect detection and chip surface two data, to judge whether there are surface defects and shape defects. Finally through the normal total, abnormal total and pass rate to display the test results.

The third part of the log record, save the test results to TXT, you can pass the test rate and test results saved in TXT file saved to the computer.

5 System Application Display

The system overcomes the influence of uncertain factors such as the chip deformation caused by manual contact, reduces the interference and protects the integrity of the chip; The man-machine interface of the system is intelligent, and the data of the

system is easy to collect, which greatly improves the efficiency and reduces the labor cost; The system image photo processing real-time, staff can observe the status of the chip in real time, the software will process the algorithm and display the test results, in the human-computer interface to obtain the chip detection information, judge whether the chip is qualified. The system implementation display in Fig. 14

The system has been tested continuously for a long time and the test results are obtained. As shown in Fig. 15, the standard value is 0.995, and the system test results are all between 0.9965 and 0.9945; the measurement data is relatively stable.

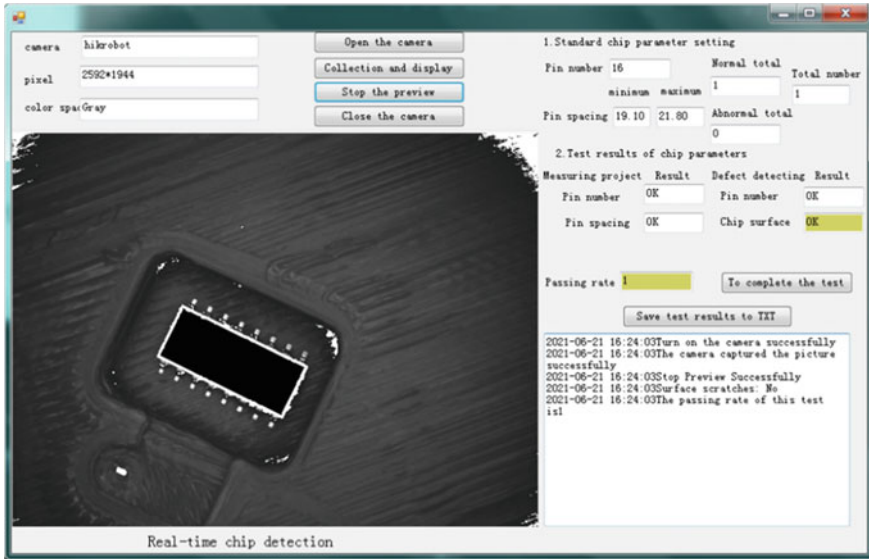


Fig. 14 System implementation display

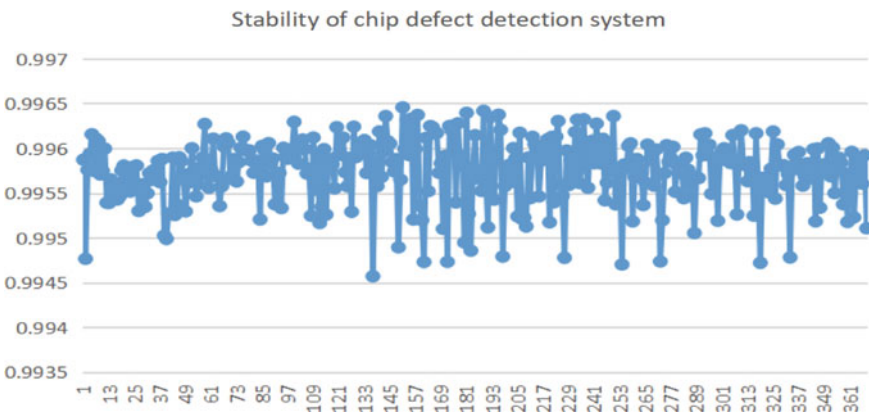


Fig. 15 Stability of chip defect detection system

6 Conclusion

The design team members respectively through the machine vision, image processing, mechanical structure design, circuit connection, human-machine interface construction and other knowledge of learning integration, combined with the current situation of chip defect detection, designed a can load different chips; Can accurately detect three kinds of chip defects; The hardware part can accurately coordinate the motor, industrial camera, LED light source, motor encoder, chip bearing table, sensor, automatic turntable parts work flexibly; Realized the mechanical hardware part and the software part can match smoothly.

This system has developed a set of simple structure, high precision, fast detection speed, strong practicability and high reliability of chip detection system. In daily industrial production, machine vision is introduced to detect the chip, and the qualified rate of the chip is analyzed. It greatly improves the working efficiency and is more practical in practical application.

References

1. Yuan, C.: Research on key technologies of on-line detection of semiconductor chip surface defects based on machine vision. Southeast University (2017)
2. Hui, Z.: Research on moving object detection and tracking algorithm based on blob. Harbin Institute of technology (2016)
3. Mechanical design manual compiled by Editorial Board of mechanical design manual. Beijing, China Machine Press (2005)
4. Hui, K.: Research on SMT pin detection based on machine vision. Guangdong University of technology (2006)
5. Min, X.: Research on wafer positioning system based on machine vision. Shanghai Jiaotong University (2013)
6. Yangfen, Z., Dongping, Y., Shunzhang, Z., Haiwen, W., Wei, Z.: Development and application of machine vision in automotive industry. *Automot Pract Technol* **2017000**(022):8–11

A Bayesian Probabilistic Score Matrix Based Collaborative Filtering Recommendation System for Rolling Bearing Fault Identification



Yinghang He, Guangbin Wang, Fengshou Gu, and Andrew D. Ball

Abstract As the amount of data generated by monitoring the condition of rolling bearings is increasing, matrix factorization-based collaborative filtering can effectively dig out valuable fault information from it. However, in practice, the amount of data generated by the normal state of the bearing is much larger than the amount of data of the bearing fault. As the total amount of data increases, this imbalance will become more and more and more severe, bearing fault information is often overwhelmed in it. In response to this problem, this paper starts from the perspective of mathematical statistics, a method of mean conjugate prior is proposed for the bearing normal condition data of bearing score matrix, from which the prior distribution of the probability distribution parameters of the bearing fault data is obtained. Then combined with the Bayesian method, we get the posterior distribution. According to the distribution, the random number is used to construct the Bayesian probabilistic scoring matrix (BPSM). Relying on BPSM, the collaborative filtering recommendation algorithm is used to identify different types of faults in rolling bearings. Under unbalanced data, comparing with the identification under a conventional joint score matrix (CJSM), the model built based on BPSM has a better identification effect on bearing fault state.

Keywords Fault identification · Bayesian method · Recommendation system · Collaborative filtering

Y. He · F. Gu · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK

e-mail: Yinghang.He@hud.ac.uk

G. Wang (✉)
School of Mechanical and Electrical Engineering, Lingnan Normal University, Zhanjiang 524048,
China

e-mail: jxxwgb@126.com

1 Introduction

The amount of data generated by monitoring the condition of rolling bearing conditions is increasing, and the information overload problem of these data is gradually highlighted. In the era of Internet, collaborative filtering recommendation technology is an effective solution to solve information overload problems [1, 2]. Therefore, in recent years, some scholars have gradually used the collaborative filtering recommendation technology to deal with the information overload problem of mechanical equipment monitoring data. Paper [3] applies collaborative filtering theory in the fault diagnosis field of civil aircrafts. Similarities between faults in the theory are calculated by the Pearson method and vector cosine method. By analysing the defects of the collaborative filtering method, the concept of met similarity and weight is applied to solve these problems. Paper [4] makes failure recommendation for online electric multiple units by using collaborative filtering algorithms based on the real-time status data and further provides the accessible scheme for failure by solution knowledge base. But the above methods all belong to memory-based collaborative filtering. This method relies on the calculation of the similarity of faults. Moreover, the sparseness of fault data often leads to inaccurate similarity calculation, and it is impossible to identify more universal conditions and more general faults. Matrix factorization-based collaborative filtering has excellent performance in dealing with the sparse problem of fault data. Paper [5] obtains the bearing feature matrix based on the wavelet frequency band energy and then design a scoring matrix that accurately describes the bearing state; finally, we design a joint scoring matrix for bearing state identification by combining the matrix of these two different characteristics. After that, a collaborative filtering recommendation system for bearing state identification is proposed based on matrix factorization-based collaborative filtering and gradient descent algorithm. This method has achieved a good identification effect on rolling bearing faults.

However, in the practice bearing state monitoring, a large number of bearing normal state data are often obtained, and in comparison, the bearing fault state data is much less. For supervised learning, this creates a class imbalance problem. In simple terms, the number of samples of a class is significantly larger than the other class [6–8]. From a learning perspective, a few classes often contain more important classification information, and the cost of misclassification of minority samples is higher [9]. The collaborative filtering algorithm proposed by paper [5] for rolling bearing fault identification only considers the balanced training set, and may generate an unsatisfactory classification model when faced with unbalanced data, some information of bearing fault state is often submerged in a large number of bearing normal state data. Aiming at the problem of unbalanced data that is widely used in practical applications, relevant researchers have proposed a series of solutions, which can be mainly divided into internal methods and external methods [7]. Among them, the external method reduces the influence of class imbalance on the classification by pre-processing the data [10, 11]. This method directly changes the unbalanced distribution of the data, and the effect is obvious and the adaptability is wide. Therefore,

the research results are more. In addition, Bayesian statistics has unique advantages in digging out data information. Bayesian schools attach importance to the collection, mining and processing of prior information, quantify them, form a priori distribution, participate in statistical inference, and improve statistical inference quality [12].

This paper starts from the perspective of mathematical statistics, a method of mean conjugate prior is proposed for the bearing normal operation data of bearing scoring matrix, from which the prior distribution of the probability distribution parameters of the bearing fault data is obtained. Then combined with the Bayesian method, the posterior distribution of the bearing feature score table is obtained. According to the distribution, the random number is used to form the feature score of Bayesian probabilistic scoring matrix. Relying on the feature scoring matrix, the collaborative filtering recommendation algorithm is used to identify different types of faults in rolling bearings. Compared with the identification with the normal scoring matrix, the new method allows for mining the fault information submerged in noise and achieving a higher identification rate.

2 Collaborative Filtering Recommendation Algorithm Based on Joint Score Matrix for Bearing Fault Identification

A conventional joint score matrix C is the basis of the collaborative filtering recommendation system for bearing state identification. It is constructed by two parts: the bearing feature score matrix A and the state score matrix B [5].

For u sets of vibration signals $S^{(1)}, \dots, S^{(h)}, S^{(h+1)}, \dots, S^{(u)}$ the training data of the former h subset signals can be taken as training data, and the latter $u - h + 1$ is the test data. To highlight useful information for detection and diagnosis, the signals usually prepressed with different methods such as wavelet transforms. In doing so, the i -th subset signal can be decomposed into levels by wavelet packet transform, resulting in decomposed inguinal in $b = 2^a - 1$ sub-bands. So, the total signal $S^{(i)}$ can be expressed as follows [13]:

$$S^{(i)} = S_{a0}^{(i)} + S_{a2}^{(i)} + \dots + S_{ab}^{(i)} \tag{1}$$

Let $S_{aw}^{(i)} (w = 0, 1, \dots, b)$ correspond to the energy $E_{aw}^{(i)}$, and then

$$E_{aw}^{(i)} = \int |S_{aw}^{(i)}(t)|^2 dt \tag{2}$$

The total energy of the signal is:

$$E^{(i)} = \sum_{w=0}^b E_{aw}^{(i)} \tag{3}$$

Normalized feature vector $T^{(i)}$ of energy is constructed as follow:

$$T^{(i)} = [e_{a0}^{(i)}, e_{a1}^{(i)}, \dots, e_{ab}^{(i)}] \tag{4}$$

where $e_{aw}^{(i)} = E_{aw}^{(i)}/E^{(i)}$ and $e_{aw}^{(i)} \in (0, 1)$.

The bearing feature score matrix $A \in R^{(b+1) \times u}$ is obtained according to the energy feature vector $T^{(i)}$:

$$A = (A_{tr} A_{te}) = \begin{pmatrix} e_{a0}^{(1)} \cdots e_{a0}^{(h)} & e_{a0}^{(h+1)} \cdots e_{a0}^{(u)} \\ e_{a1}^{(1)} \cdots e_{a1}^{(h)} & e_{a1}^{(h+1)} \cdots e_{a1}^{(u)} \\ \vdots & \vdots \\ e_{ab}^{(1)} \cdots e_{ab}^{(h)} & e_{ab}^{(h+1)} \cdots e_{ab}^{(u)} \end{pmatrix} \tag{5}$$

where $A_{tr} \in R^{b \times h}$ represents the training data set, $A_{te} \in R^{b \times (u-h)}$ represents the test data set.

Assume that there are v states $Z_1, Z_2 \dots Z_v$ to be identified for rolling bearing health conditions. For the training data $S^{(1)}, S^{(2)}, \dots, S^{(h)}$, its corresponding state score is defined to be unitary value 1, while the non-existent state score is given to be a small value $\varepsilon (\leq 1/10000)$, and it constructs as $B_\varepsilon \in R^{v \times h}$. For test data $S^{(i')}$ ($i' = h + 1, \dots, u$), its score for state Z_t ($t = 1, \dots, v$) is unknown but set with a value of 0, denoted as $P_t^{(i')}$, and it constructs as $B_p \in R^{v \times (u-h)}$. The bearing state score matrix $B \in R^{v \times u}$ is obtained:

$$B = (B_\varepsilon B_p) = \begin{pmatrix} 1 \cdots \varepsilon & p_1^{(h+1)} \cdots p_1^{(u)} \\ \varepsilon \cdots \varepsilon & p_2^{(h+1)} \cdots p_2^{(u)} \\ \vdots & \vdots \\ \varepsilon \cdots 1 & p_v^{(h+1)} \cdots p_v^{(u)} \end{pmatrix} \tag{6}$$

Finally, the joint score matrix C_{CJSM} is constructed as:

$$C_{CJSM} = \begin{pmatrix} A_{tr} & A_{te} \\ B_\varepsilon & B_p \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} \tag{7}$$

where $C_{CJSM} \in R^{d \times u}$ and $d = b + 1 + v$.

As joint score matrix C is low-rank sparse matrix, it can be factorized into the product of two characteristic matrices: $\Theta \in R^{u \times k}$ and $X \in R^{d \times k}$ to solve $p_t^{(i')}$ in the matrix, namely $C = X\Theta^T$:

$$\begin{pmatrix} e_{a0}^{(1)} & \cdots & e_{a0}^{(h)} & e_{a0}^{(h+1)} & \cdots & e_{a0}^{(u)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{ab}^{(1)} & \cdots & e_{ab}^{(1)} & e_{ab}^{(h+1)} & \cdots & e_{ab}^{(u)} \\ 1 & \cdots & \varepsilon & p_1^{(h+1)} & \cdots & p_1^{(1)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \varepsilon & \cdots & 1 & p_v^{(h+1)} & \cdots & p_v^{(u)} \end{pmatrix} \cong \begin{pmatrix} x_1^{(1)} & \cdots & x_k^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(d)} & \cdots & x_k^{(d)} \end{pmatrix} \cdot \begin{pmatrix} \tilde{\theta}_1^{(1)} & \cdots & \tilde{\theta}_k^{(1)} \\ \vdots & \ddots & \vdots \\ \tilde{\theta}_1^{(u)} & \cdots & \tilde{\theta}_k^{(u)} \end{pmatrix} \tag{8}$$

To find the optimal parameters Θ and X , a minimisation is implemented based on the overall cost function $L(\Theta, X, C)$:

$$(\theta, X) = \arg \min L((\Theta, X, C) + \lambda(\|\theta\|^2 + \|X\|^2)) \tag{9}$$

where λ is the regularization coefficient.

Finally, the gradient descent method is used to optimize the parameters, and then the predicted score $p_t^{(i')}$ of the test data $S^{(i')}$ for the state Z_t is obtained:

Then, the state Z_t corresponding to the highest score $\max p_t^{(i')}$, that is, the state of the identification test data $S^{(i')}$ is obtained.

3 Bayesian Probabilistic Scoring Matrix

Because of random noise influences, measured vibration datasets include a great degree of uncertainties and noise. Even with advanced noise suppression pre-processing, the feature parameters can still have certain degree of randomness. In this sense, the energy feature score $e_{aw}^{(i'')}$ ($i'' = 1, \dots, h$) in matrix A_{tr} can be regarded as a realization from a random variable.

Assume that in the fault state Z_t , there are α_t sets of training data: $e_{aw}^{(1)}, \dots, e_{aw}^{(\alpha_t)}$. For the very slow change in states during bearing service period, it can be assumed that scores obtained by different time instants for same bearing state satisfy the independent and identical distribution conditions. According to the central limit theorem, the uncertain measurements and environment effects usually follows a Gaussian distribution [16]. It means that that the randomness of energy feature score $\tilde{E}_{aw}^{(t)}$ of state Z_t can be analysed based on a Gaussian distribution which consists of two parameters: a mean of $\theta_{aw}^{(t)}$ and a variance of $(\theta_{aw}^{(t)})^2$, denoted as $\tilde{E}_{aw}^{(t)} \sim N_t(\theta_{aw}^{(t)}, (\theta_{aw}^{(t)})^2)$ for brevity.

Figure 1 shows typical probability density curves established according to [14, 15], which is for the experimental data in Sect. 4 (bearing outer ring crack, current

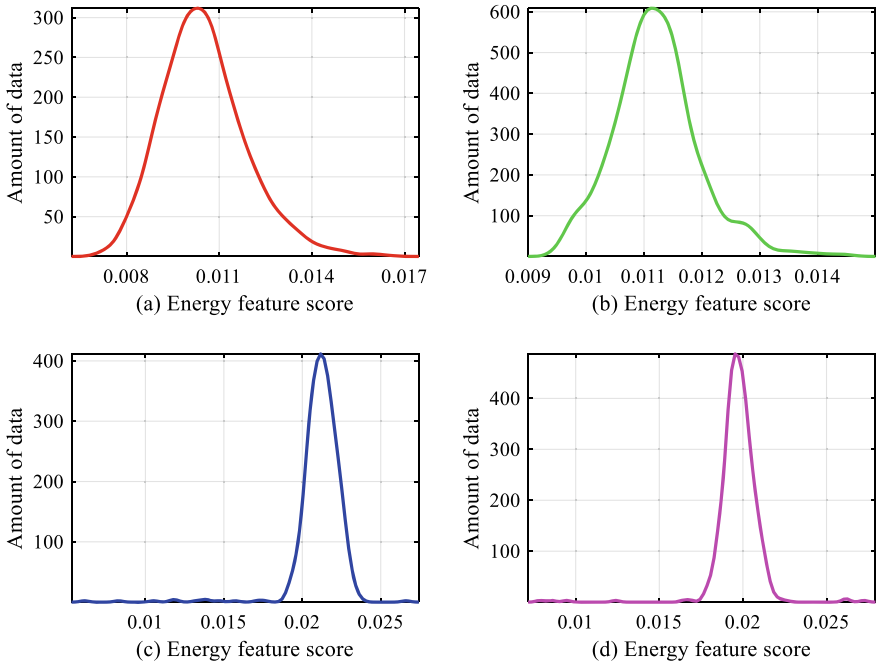


Fig. 1 Probability density curve based on the 9th energy feature score (e_{58}). **a** normal; **b** outer ring crack; **c** outer ring current' damage; **d** outer ring pitting

damage, pitting and normal based on the 9th energy feature score (e_{58}). It can be observed that these curves exhibit near Gaussian characteristics.

For $\tilde{E}_{aw}^{(t)} \sim N_t(\theta_{aw}^{(t)}, (\theta_{aw}^{(t)})^2)$, comparing the probability density curves of each energy segment's bearing normal state and fault state, we find that the influence of $(\theta_{aw}^{(t)})^2$ is relatively small and can be regarded as a known amount, estimated by the principle of maximum likelihood:

$$(\hat{\sigma}_{aw}^{(t)})^2 = \frac{1}{\alpha_t - 1} \sum_{i=1}^{\alpha_t} e_{aw}^{(i)} - \frac{1}{\alpha_t} \sum_{j=1}^{\alpha_t} e_{aw}^{(j)} \tag{10}$$

According to the Bayesian framework, we have a certain knowledge of $\theta_{aw}^{(t)}$ before we get the bearing fault status data, called prior knowledge, and the prior knowledge can be expressed by a certain probability distribution, called prior distribution [16]. In CJSM of the second section, we find that the energy feature scores of the different bearing fault state comparing with the bearing normal state tend to have a large difference in some sub-bands, while the other sub-bands have little difference relatively. Based on the experimental data in Sect. 3, Fig. 2 shows the probability density curves of the bearing normal state and outer ring pitting in different energy sub-band.

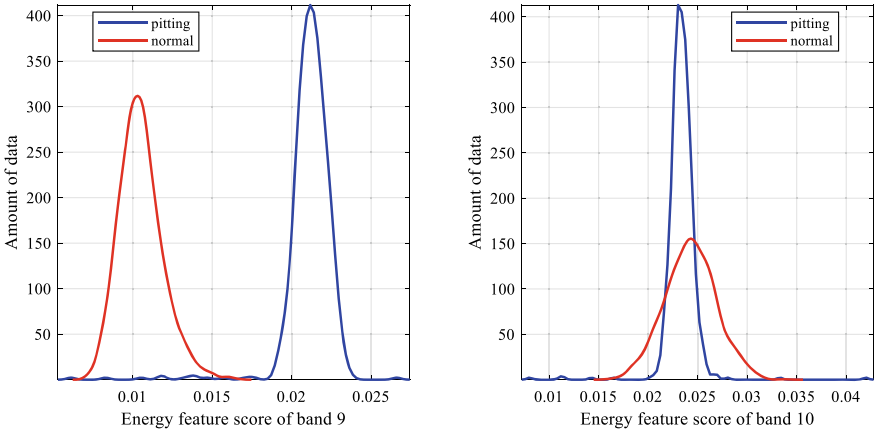


Fig. 2 Probability density curve based on different energy feature score in bearing normal state and outer ring pitting

Therefore, this paper believes that in the actual monitoring, a large number of bearing normal state data contain some “prior knowledge” of mean parameter $\theta_{aw}^{(t)}$ of bearing fault state data: some sub-band score of the bearing normal state which is not much different from the bearing fault data. How to obtain the prior distribution of the “prior knowledge” is a major problem. For the training set data, this paper proposes a method that can combine the prior information of the bearing normal state score to obtain the posterior distribution of the bearing fault state score, and then obtain the Bayesian probability feature score matrix. In this way, it is possible to reduce the influence of a large number of bearing normal state data and fully dig out the effective information therein, which is encountered in the common scenario for model-based diagnostic which involves model of training and learning.

According to the Bayesian statistical prior distribution characteristics, we believe that the prior distribution of the mean parameter $\theta_{aw}^{(t)}$ is given by a conjugate prior $N(\mu_{aw}, \mu_{aw}^2)$. In this paper, the conjugate prior is related to the bearing normal state data. It is assumed that the energy feature score of the normal bearing in the w sub-band is $Y_{aw} \in R^{1 \times (n \times m)}$:

$$Y_{aw} = \left(\overbrace{y_{aw}^{(1)} \dots y_{aw}^{(m)}}^m, \overbrace{y_{aw}^{(m+1)} \dots y_{aw}^{(2m)}}^m, \dots, \overbrace{y_{aw}^{(\Omega)} \dots y_{aw}^{(n \times m)}}^m \right) \quad (11)$$

where $\Omega = (n - 1) \times m + 1$.

Divide Y_{aw} into m groups and calculate the mean values to get \bar{Y}_{aw} :

$$\bar{Y}_{aw} = \left(\frac{1}{m} \sum_{i=1}^m y_{aw}^{(i)}, \frac{1}{m} \sum_{i=m+1}^{2m} y_{aw}^{(i)}, \dots, \frac{1}{m} \sum_{i=\Omega}^{n \times m} y_{aw}^{(i)} \right) = (\bar{Y}_{aw}^{(1)}, \bar{Y}_{aw}^{(2)}, \dots, \bar{Y}_{aw}^{(n)}) \tag{12}$$

Mean μ_{aw} and variance τ_{aw}^2 in the prior distribution $N(u_{aw}, \tau_{aw}^2)$ estimated using \bar{Y}_{aw} , it can be estimated by the principle of maximum likelihood:

$$\hat{\mu}_{aw} = \frac{1}{n} \sum_{i=1}^n \bar{Y}_{aw}^{(i)} \tag{13}$$

$$\hat{\tau}_{aw}^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{Y}_{aw}^{(i)} - \hat{\mu}_{aw})^2 \tag{14}$$

Therefore, the probability density function of the prior distribution:

$$\pi(\theta_{aw}^{(t)}) = \frac{1}{\tau_{aw} \sqrt{2\pi}} \exp \left\{ -\frac{(\theta_{aw}^{(t)} - \mu_{aw})^2}{2\tau_{aw}^2} \right\} \tag{15}$$

The energy feature score $e_{aw}^{(1)}, \dots, e_{aw}^{(\alpha_t)}$, is a set of sample observations from $N_t(\theta_{aw}^{(t)}, (\theta_{aw}^{(t)})^2)$, where $(\theta_{aw}^{(t)})^2$ is known, then the likelihood function of these samples is:

$$l(\mathbf{e}_{aw} | \theta_{aw}^{(t)}) = \left(\frac{1}{\theta_{aw}^{(t)} \sqrt{2\pi}} \right)^{\alpha_t} \exp \left\{ -\frac{1}{2(\theta_{aw}^{(t)})^2} \sum_{i=1}^{\alpha_t} (e_{aw}^{(i)} - \theta_{aw}^{(t)})^2 \right\} \tag{16}$$

Let $\bar{e}_{aw} = \frac{1}{\alpha_t} \sum_{i=1}^{\alpha_t} e_{aw}^{(i)}$, $\sigma_0^2 = \frac{(\sigma_{aw}^{(t)})^2}{\alpha_t}$, $A^* = \frac{1}{\sigma_0^2} + \frac{1}{\tau_{aw}^2}$, $B^* = \frac{\bar{e}_{aw}}{\sigma_0^2} + \frac{\mu_{aw}}{\tau_{aw}^2}$ and $C^* = \frac{1}{(\sigma_{aw}^{(t)})^2} \sum_{i=1}^{\alpha_t} (e_{aw}^{(i)})^2 + \frac{\mu_{aw}^2}{\tau_{aw}^2}$

The joint probability density function of sample \mathbf{e}_{aw} and the mean parameter $\theta_{aw}^{(t)}$ is

$$h(\mathbf{e}_{aw}, \theta_{aw}^{(t)}) = l(\mathbf{e}_{aw} | \theta_{aw}^{(t)}) \cdot \pi(\theta_{aw}^{(t)}) = k \exp \left\{ -\frac{(\theta_{aw}^{(t)} - B^*/A^*)^2}{2/A^*} \right\} \tag{17}$$

where $k = (2\pi)^{-(\alpha+1)/2} \cdot \tau_{aw}^{-1} \cdot (\sigma_{aw}^{(t)})^{-h} \cdot \exp\{- (C^* - (B^*)^2/A^*)/2\}$.

The edge probability density of \mathbf{e}_{aw} can be calculated:

$$m(\mathbf{e}_{aw}) = \int_{-\infty}^{+\infty} h(\mathbf{e}_{aw}, \theta_{aw}^{(t)}) d\theta_{aw}^{(t)} = k \left(\frac{2\pi}{A^*} \right)^{1/2} \tag{18}$$

Therefore, the posterior distribution of $\theta_{aw}^{(t)}$ is:

$$\begin{aligned} \pi(\theta_{aw}^{(t)} | e_{aw}) &= h(e_{aw}, \theta_{aw}^{(t)}) / m(e_{aw}) \\ &= \left(\frac{2\pi}{A}\right)^{-\frac{1}{2}} \cdot \exp\left\{-\frac{(\theta_{aw}^{(t)} - B^*/A^*)^2}{2/A^*}\right\} \end{aligned} \tag{19}$$

Namely, $\pi(\theta_{aw}^{(t)} | e_{aw}) \sim N(\mu'_{aw}, \tau_{aw}^2)$.

where $\mu'_{aw} = \frac{B^*}{A^*} = \frac{\bar{e}_{aw}\sigma_0^{-2} + \mu_{aw}\tau_{aw}^{-2}}{\sigma_0^{-2} + \tau_{aw}^{-2}}$, $\tau_{aw}^2 = \frac{\sigma_0^2 \cdot \tau_{aw}^2}{\sigma_0^2 + \tau_{aw}^2}$.

According to the generalized maximum likelihood estimation principle [17], take μ'_{aw} as the estimated value of $\theta_{aw}^{(t)}$:

$$\hat{\theta}_{aw}^{(t)} = \mu'_{aw} \tag{20}$$

Therefore, $\tilde{E}_{aw}^{(t)} \sim N_t(\theta_{aw}^{(t)}, (\sigma_{aw}^{(t)})^2)$.

For every $\tilde{E}_{aw}^{(t)} (t = 1, \dots, v; w = 0, \dots, b)$, we generate a random number based on the distribution $N_t(\theta_{aw}^{(t)}, (\sigma_{aw}^{(t)})^2)$ to replace the energy feature score of bearing fault state in A_{tr} , and construct a new matrix A_{tr}^{BPSM} . The Bayesian probabilistic scoring matrix is constructed as:

$$C_{BPSM} = \begin{pmatrix} A_{tr}^{BPSM} & A_{te} \\ B_\varepsilon & B_p \end{pmatrix} \tag{21}$$

4 Performance Verification

To verify the performance of BPSM proposed, the collaborative filtering recommendation technology is implemented for bearing fault identification. It identifies bearing health conditions with different types of bearing outer ring faults. As shown in Fig. 3, there are three common faults including outtrace pitting, outer ring crack, and outer ring current damage as depicted in Fig. 3a, b and c respectively which are on three popular 6205EKA deep groove ball bearings. By operating the bearing

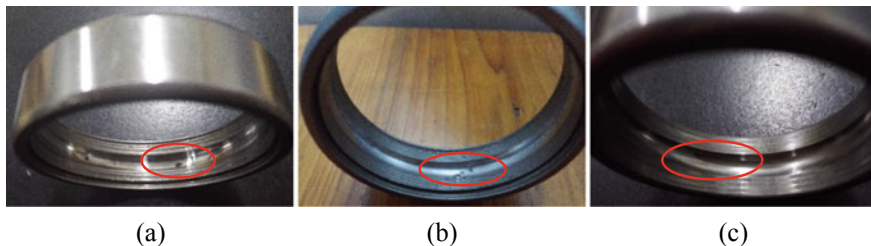


Fig. 3 a Outer ring pitting; b outer ring crack; c outer ring current damage

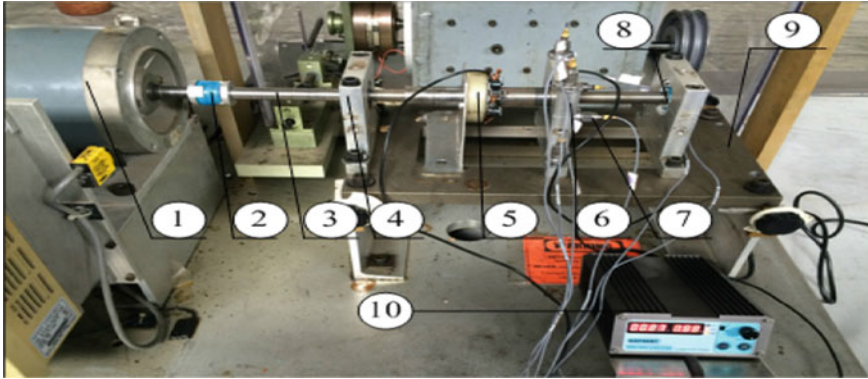


Fig. 4 a Bearing test bench ① electric motor; ② insulated coupling; ③ principal axis; ④ supporting bearing pedestal; ⑤ current loading device; ⑥ test bearing's bearing pedestal; ⑦ vibration acceleration sensor; ⑧ insulated bearing; ⑨ base; ⑩ current simulator

test bench shown in Fig. 4 at three motor speeds 600, 1200 and 1800 rpm, each under three radial loads of 500, 1000 and 1500 N, we have obtained several vibration accelerations signals of the bearings in the four states with the sampling frequency of 16,384 Hz. Then, we divide these signals into small segments, which are 512 sets of outer ring pitting corrosion, 540 sets of outer ring crack, 656 sets of current damage, and 2880 sets of normal, and total 4588 sets of samples.

For implementing the approach, 4588 data are divided into three subsets randomly: training set (2753 sets), cross-validation set (918 sets), and test set (917 sets).

The data of the training set includes 1736 bearing normal state data, 316 outer ring crack state data, 404 outer ring current damage state and 297 outer ring pitting state. With these datasets, the BPSM is constructed to be 36×3671 with cross-validation set. For comparison, a CJSM is also formulated to have the same size as BPSM but with difference of sub-matrix as detailed, and the same cross-validation set and test set are used to determine the model parameters and verify the generalization ability, as shown in Fig. 5.

Using different regularization coefficients λ and feature number k , relying on the training set learning model, the cross-validation set is used for state identification, and the identification rate is obtained, as shown in Fig. 6. It can be seen that the BPSM result in a higher identification rate for the cross-validation set, the highest rate being 91.61% for BPSM when $\lambda = 0.0016$ and $k = 10$, and 89.32% for CJSM when $\lambda = 0.002$ and $k = 10$.

Select $\lambda = 0.0016$, $k = 10$ for BPSM and $\lambda = 0.002$, $k = 10$ for CJSM to evaluate the performance of the model on the test set.

Figure 7 shows the specific identification result of the BPSM model for different bearing states. The overall identification rate is as high as 91.2%. Comparatively, CJSM can only have a correct rate of 85.9%. It is proved that the model has good generalization ability under this parameter, and the effect is better when using BPSM. Especially, CJSM model takes almost all the crack fault data (99.1% of crack fault

$$A_{tr} = \begin{pmatrix} 0.1236 & 0.1205 & \dots & 0.0677 \\ 0.075 & 0.0814 & \dots & 0.085 \\ \vdots & \vdots & \ddots & \vdots \\ 0.0146 & 0.0155 & \dots & 0.0148 \end{pmatrix}_{32 \times 2753}$$

$$A_{tr}^{BPSM} = \begin{pmatrix} 0.1148 & 0.1195 & \dots & 0.0677 \\ 0.0789 & 0.0796 & \dots & 0.085 \\ \vdots & \vdots & \ddots & \vdots \\ 0.0146 & 0.0157 & \dots & 0.0148 \end{pmatrix}_{32 \times 2753}$$

$$A_{te} = \begin{pmatrix} 0.0509 & 0.0343 & \dots & 0.1143 \\ 0.0534 & 0.0634 & \dots & 0.0762 \\ \vdots & \vdots & \ddots & \vdots \\ 0.0177 & 0.0164 & \dots & 0.0137 \end{pmatrix}_{32 \times 918}$$

$$B_e = \begin{pmatrix} 0.00001 & 0.00001 & \dots & 1 \\ 0.00001 & 0.00001 & \dots & 0.00001 \\ 1 & 0.00001 & \ddots & 0.00001 \\ 0.00001 & 1 & \dots & 0.00001 \end{pmatrix}_{4 \times 2753}$$

$$C_{CJSM} = \begin{pmatrix} A_{tr} & A_{te} \\ B_e & B_p \end{pmatrix}_{36 \times 3671}$$

$$B_p = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix}_{4 \times 918}$$

$$C_{BPSM} = \begin{pmatrix} A_{tr}^{BPSM} & A_{te} \\ B_e & B_p \end{pmatrix}_{36 \times 3671}$$

Fig. 5 CJSM and BPSM constructed by training set and cross-validation set

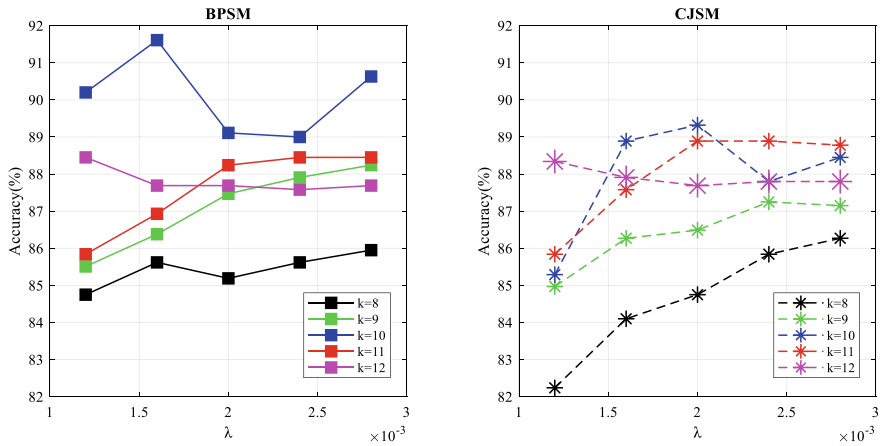


Fig. 6 Identification rate of cross-validation sets under two methods

data) into the normal state. This is mainly due to the influences of large amount normal datasets. In comparison, 71.3% of the crack fault data are identified correctly by BPSM.

5 Conclusions

Face the problem of unbalanced scoring matrix for the imbalance of data sets in practice, this paper starts from the perspective of mathematical statistics, a method of mean conjugate prior is proposed for the bearing normal operation data of bearing

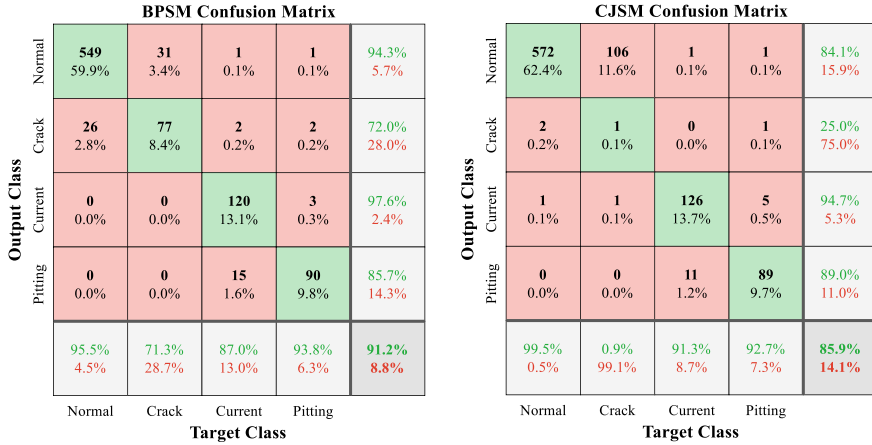


Fig. 7 Identification rate for test datasets

score matrix, from which the prior distribution of the probability distribution parameters of the bearing fault data is obtained. Then combined with the Bayesian method, the posterior distribution of the bearing feature score table is obtained. According to the distribution, the random number is used to form the feature score of Bayesian probabilistic score matrix. Relying on the feature score matrix, the collaborative filtering recommendation algorithm is used to identify different types of faults in rolling bearings.

Experiments were carried out on the normal bearings and bearings of pitting, crack, and current damage on the outer ring of the rolling bearing, and the vibration signal data were obtained. The data of the bearing normal state in the data accounts for more than half of the total data, which highlights the imbalance of the data set. We use the proposed method to obtain the BPSM of the training set to train the model. For comparison, CJSJM is also used as a training set to train the model, and the same cross-validation set and test set are used to determine the model parameters and verify the generalization ability, and the identification effect of the two is compared. It shows that the BPSM model has a higher identification rate than CJSJM overall when they are under the same model parameters. Moreover, crack fault information is submerged in the unbalanced data set. The model trained under the CJSJM has a poor effect on crack fault state identification, while the model trained by the BPSM can identify the crack fault state to a certain extent.

References

1. Resnick, P., Varian, H.R.: Recommender systems. *Commun. ACM* **40**, 56–58 (1997)
2. Zenebe, A., Norcio, A.F.: Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems. *Fuzzy Sets Syst.* **160**, 76–94 (2009)

3. Xu, P.: Research on fault diagnosis method of civil aircraft based on collaborative filtering theory. *J. Civil Aviation University of China* **32**, 23–26 (2014)
4. Guo, M.L.: Research and realization of EMU's operation and maintenance decision-making recommended techniques based on knowledge base. M.E. thesis, Beijing Jiaotong University, Beijing, China (2015)
5. Wang, G.B., He, Y.H., Peng, Y.F., Li, H.J.: Bearing fault identification method based on collaborative filtering recommendation technology. *Shock and Vibration* (2019)
6. Chawla, N.V., Japkowicz, N., Kotcz, A.: Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsl* **6**, 1–6 (2004)
7. López, V., Fernández, A., García, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics. *Inf. Sci.* **250**, 113–141 (2013)
8. He, H., Garcia, E.A.: Learning from imbalanced data. *Knowledge and Data Eng. IEEE Trans. on* **21**, 1263–1284 (2009)
9. Elkan, C.: The foundations of cost-sensitive learning. In: *Proceedings of the International Joint Conference on Artificial Intelligence* (2001)
10. Batista, G.E.A.P.A., Prati, R.C., Monard, M.C.A.: study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsl* **6**, 20–29 (2004)
11. Estabrooks, A., Jo, T., Japkowicz, N.: A multiple resampling method for learning from imbalanced data sets. *Comput. Intell.* **20**, 18–36 (2004)
12. Mao, S.S., Tang, Y.C.: *Bayesian Statistics*, 2nd edn. China Statistics Press, Beijing, China (2012)
13. Zhang, D.F.: *MATLAB Wavelet Analysis*. China Machine Press, Beijing, China (2010)
14. Hill, P.D.: Kernel estimation of a distribution function. *Commun. Statist.-Theory Methods* **14**, 605–620 (1985)
15. Jones, M.C.: Simple boundary correction for kernel density estimation. *Stat. Comput.* **3**, 135–146 (1993)
16. Chen, X.R.: *Probability Theory and Mathematical Statistics*, 1st edn. Press of University of Science and Technology of China, Hefei, China (2009)
17. Chen, X.R., Ni, G.X.: *Mathematical Statistics Course*, 1st edn. Press of University of Science and Technology of China, Hefei, China (2009)

On-Line Monitoring of the Dimensional Error in Turning of a Slender Shaft



Pengyu Lu, Kaibo Lu, Yipei Liu, Bing Li, Xin Wang, Meixia Tian,
and Fengshou Gu

Abstract Bending deformation is easy to occur when turning a long slender workpiece due to its low stiffness, which seriously affects the machining dimensional accuracy. Currently, the dimension of the part is generally measured off-line after the completion of the operation. The purpose of this paper is to explore an on-line monitoring method for the dimensional error of slender shaft in turning processes. First the deformation of the whole workpiece in the process of machining is analyzed. The deformation correlation at the measuring point and that at the cutting point is deduced. Then an on-line monitoring approach to radial dimension is proposed using a single fixed displacement sensor and Wavelet Transform. Finally, the reliability of the monitoring method is verified by machining experiments. The experimental results show that the presented on-line monitoring model enables to predict the dimensional error of the machined workpiece effectively.

Keywords Deformation · Turning · Slender shaft · Displacement signal

P. Lu · K. Lu (✉) · Y. Liu · X. Wang · M. Tian
College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan
030024, China
e-mail: lvkaibo@tyut.edu.cn

P. Lu
e-mail: lvpengyuu@163.com

B. Li · F. Gu
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, China

F. Gu
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK

1 Introduction

A shaft with the length to diameter ratio greater than 20 is usually called the slender beam. Turning slender shafts is prone to deformation as well as chatter vibrations under the cutting force due to their low stiffness, which seriously affects the machining accuracy and productivity [1, 2]. The adverse influence, therefore, makes the monitoring or prediction of the dimensional deviation an indispensable activity to pursue intelligent manufacturing processes [3–5].

The deformation of the machined workpiece is affected by cutting force and structural stiffness, so it is of great significance to derive accurate cutting force and stiffness for monitoring the dimensional error. The stiffness at each position along the workpiece is associated with the clamping conditions. Wang presented a finite element method to analyze the machining deformation under different clamping ways and obtained the optimal clamping ways to reduce the deformation [6]. Lu et al. analyzed the stiffness distribution at different positions of the slender shaft by analytical method and finite element method, discussed the influence of spindle support bearing and chuck on the workpiece chatter, and obtained an accurate chatter monitoring model [7, 8]. Gong used an improved artificial bee colony algorithm to optimize cutting parameters with cutting stability and surface roughness taken into account [9]. Guo carried out numerical simulation of the turning process based on Rayleigh beam theory, and determined the dynamic response of the rotating workpiece under the action of cutting force. The error compensation was carried out according to the deformation, which has guiding significance for the actual machining [10]. Wu used the finite element method to compare the two kinds of turning forms, and stated that the reverse turning was more beneficial to deformation reduction [11]. Yang developed a flexible hydraulic tool rest and elastic tail rest to lower the dimensional error during machining, and effectively improved the radial dimensional accuracy and surface roughness [12]. Also, a self-made following frame was designed to improve the stiffness of slender shaft and obtained better machining quality [13].

In summary, research on slender shaft machining mainly is focused on updating the clamping approaches of the flexible workpiece to be machined to improve its structural stiffness. And on-line monitoring of turning processes is mainly focused on machining chatter or surface roughness. Relatively, literature concerning dimensional deviation monitoring in cutting of the slender workpiece is sparse. In this paper, we introduce an analytical method for dimensional error prediction in turning of a slender shaft. The correlation between the deformation at the measuring point and that at the cutting point is deduced. Then an on-line monitoring approach to radial dimension is proposed using a single fixed displacement sensor. Finally, the reliability of the monitoring method is verified by machining experiments.

2 Dimensional Error Model

It is known that the cutting force can be decomposed into three components, that is, the tangential force, radial force, and axial force. Since the workpiece deformation in the radial direction mainly affects the depth of cut, resulting in radial dimension error after processing. Comparatively, the effect of the deformation caused by the tangential force and the axial force on the depth of cut can be ignored. Thus, the influence of the bending deformation generated by the radial cutting force on the final diameter of the machined workpiece is only considered in this study.

Figure 1 shows the deflection of the workpiece subjected to the radial cutting force. Obviously, the actual depth of cut will decrease due to deformation of the workpiece during turning. The actual depth of cut ap_e is

$$ap_e = ap - \omega_{(c)} \tag{1}$$

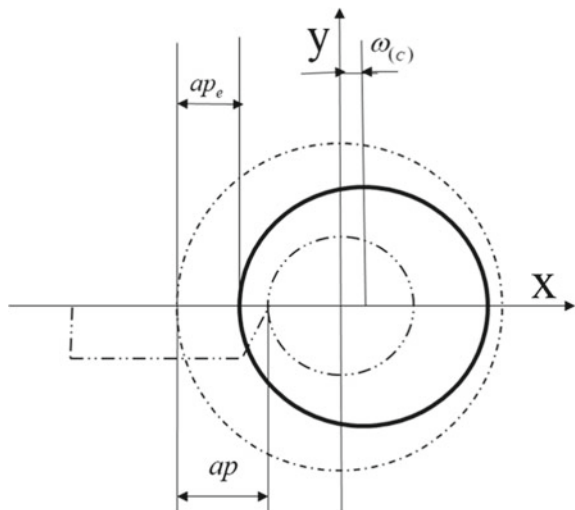
where ap is the nominal depth of cut and $\omega_{(c)}$ is the deflection of workpiece caused by the radial cutting force at the current cutting point. Then, the actual diameter d_e of workpiece after cutting is

$$d_e = d - 2ap_e \tag{2}$$

where d is the diameter of workpiece before a turning operation.

Therefore, the key issue for the dimensional deviation monitoring is to determine the deflection of workpiece along its length. To this end, the natural thought is to measure the deflection at the cutting point directly. But it could be impractical mainly

Fig.1 Radial dimensional error model of a slender workpiece in turning



due to the interference of cutting chips and fluid in operation. As a result, a strategy for indirect measurement of $\omega_{(c)}$ is needed. An analytical method for dimensional error prediction in turning of a slender shaft will be explored in the following.

3 Method for the Radial Deformation In-Process Monitoring

In the present study, we are focused on the case of a uniform slender shaft, which is clamped by a chuck at one end and simply supported by a live center at the other end on a horizontal turning lathe. For simplicity, the workpiece here is modeled as propped cantilever beam model. Obviously, this model is a statically indeterminate beam, as shown in Fig. 2. The support of the chuck and tailstock is generally idealized or simplified as fixed and joint constraints, respectively. The point C denotes the current cutting position along the workpiece, and the point M denotes the measurement point of a displacement sensor.

3.1 Equation of the Workpiece Deflection

According to the basic theory of material mechanics, the reaction force at the simply supported point A can be known as:

$$F_A = \frac{Fb^2}{2l^3}(3l - b) \tag{3}$$

Furthermore, the bending moment equation on the section at a distance x ($0 < x < a$) from the A end is

$$M(x) = F_A x \tag{4}$$

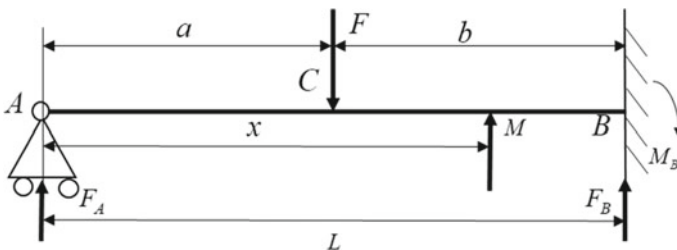


Fig. 2 Simplified model of a slender beam in turning

Then, the approximate differential equation of the torsion curve $y(x)$ is expressed as

$$y''(x) = \frac{d^2y}{d^2x} = \frac{M(x)}{EI} \quad (5)$$

where E is Young's modulus and I is the moment of inertia of the beam cross section.

$$EIy''(x) = \frac{Fb^2}{2l^3}(3l - b)x \quad (6)$$

The general solution of Eq. (6) is obtained as

$$y(x) = \frac{1}{EI} \left[\frac{1}{6} F_A x^3 + c_1 x + d_1 \right] \quad (7)$$

For $a < x < l$, there is the following solution,

$$M(x) = F_A x - F(x - a) \quad (8)$$

$$y(x) = \frac{1}{EI} \left[\frac{1}{6} F_A x^3 - F \frac{(x - a)^3}{6} + c_2 x + d_2 \right] \quad (9)$$

With boundary conditions idealization, for a fixed end, the deflection and slope are zero; for a simply supported end, the bending moment and deflection are zero. Substituting the boundary conditions in Eqs. (7) and (9) yields:

$$c_1 = c_2 = -1/2(F_A l^2 - Fb^2), \quad d_1 = d_2 = 0 \quad (10)$$

Finally, the equation of the deflection of the workpiece can be obtained as:

$$\begin{aligned} \omega(x) &= \frac{F}{6EI} \left(\frac{b^2}{2l^3} (3l - b)(x^3 - 3l^2x) + 3xb^2 \right) \quad 0 < x < a \\ \omega(x) &= \frac{F}{6EI} \left(\frac{b^2}{2l^3} (3l - b)(x^3 - 3l^2x) + 3xb^2 - (x - a)^3 \right) \quad a < x < l \end{aligned} \quad (11)$$

The deformation at the cutting position is:

$$\omega_{(c)} = \frac{F}{6EI} \left(\frac{b^2}{2l^3} (3l - b)(a^3 - 3l^2a) + 3ab^2 \right) \quad (12)$$

3.2 Correlation Between Deformation at Measuring Point and Cutting Point

By dividing Eqs. (11) by (12), the relationship between the deformation at the measuring point and the cutting point can be obtained as:

$$\omega_{(c)} = \omega(x) \frac{\left(\frac{b^2}{2l^3}(3l-b)(a^3-3l^2a)+3ab^2\right)}{\frac{b^2}{2l^3}(3l-b)(x^3-3l^2x)+3xb^2} \quad 0 < x < a$$

$$\omega_{(c)} = \omega(x) \frac{\left(\frac{b^2}{2l^3}(3l-b)(a^3-3l^2a)+3ab^2\right)}{\frac{b^2}{2l^3}(3l-b)(x^3-3l^2x)+3xb^2-(x-a)^3} \quad a < x < l$$
(13)

4 Experimentation and Discussion

4.1 Experimental Setup

This section presents the verification of the proposed monitoring approach through turning tests. The experimental setup is shown in Fig. 3. Experiments were performed on a CA6140 conventional horizontal lathe with the power of 7.5 kW. All the test bars with the length 400 mm and the initial diameter 25 mm were made of 1045 steel. The material has the Young modulus 2.06×10^5 MPa and density 7850 kg/m^3 . The rhombic insert was used for dry cutting, which has the nose angle 55° , nose

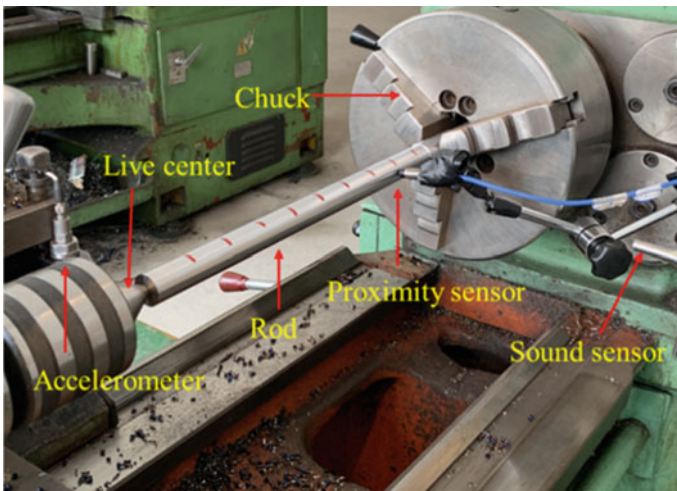


Fig. 3 Experimental setup

Table 1 Cutting parameters

Cutting pass	Spindle rotation speed/rpm	Depth of cut /mm	Feed rate /(mm/rev)
1st	1120	0.5	0.1
2nd	1120	1.2	0.1

radius 0.4 mm, and clearance angle 7°, respectively. The data acquisition system of CoCo-80 was applied to record the measured signals.

In the machining tests, one proximity sensor was fixed on the slide-way to acquire the displacement of the workpiece. The signal was sampled at the rate of 10.24 kHz. Table 1 shows the machining parameters in the experiments.

4.2 Results and Discussion

Wavelet transform is to project the signal onto the orthogonal wavelet base and expand the signal on different scales, so that the relevant features can be extracted within different frequency bands. In addition, the time-domain features of the signal in each frequency band can be retained. Wavelet transform can only decompose the approximate signal feature, which can deal with the low frequency signal well. Wavelet decomposition technology has powerful denoising functions. Thus, the processing method for displacement signals based on the wavelet technology is applied for feature extraction.

Figure 4 shows the detail coefficients of each order after 12th order decomposition of the original displacement signal using the Symlet4 wavelet, in which the reconstruction of low-frequency signal is selected as the desired deformation data. The original signal and the reconstructed signal related to workpiece deflection are compared in Fig. 5.

After turning, the workpiece was measured by a micrometer with an accuracy of 0.001 mm. Figure 6 shows the radial dimension of the workpiece along its axis. It can be seen that the maximum deformation normally occurs near the midpoint of the workpiece during turning operations because the position holds the lowest stiffness.

Through inputting the reconstructed deformation data at the measuring point into the on-line monitoring model of Eq. (13), the radial dimension of the workpiece can be predicted. Figure 7 shows the comparison of the on-line monitoring results with the off-line measurements. Overall, the predicted and measured values of the dimensional deviation distribution along the workpiece have the similar trends, which basically verified the proposed monitoring method. Also, it can be seen that the predictive geometric dimensions are generally lower than the practical measurements, indicating that the monitoring model is stiffer than the physical machining system.

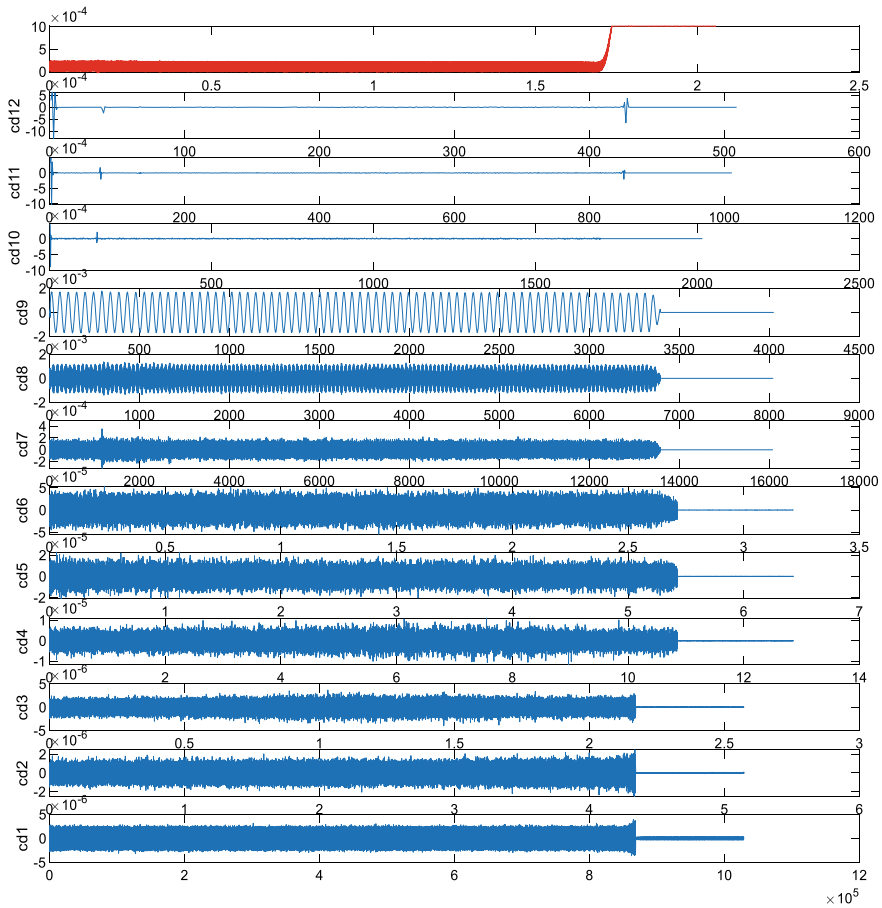


Fig. 4 Wavelet Transform of the displacement signal

5 Conclusions

This paper introduces an online monitoring method for the dimensional deviation of the compliant clamped-simply supported workpiece in turning operations. The radial dimension of the workpiece both in the on-line monitoring models and the off-line test results increases firstly and then decreases, like a Chinese ‘waist drum’ shape. It can be found that the maximum deformation position is about $0.414 l$, located around the mid-point of the length of workpiece. The monitoring results are relatively lower than the practical measurements, since the boundary conditions are simplified in modelling which leads to a certain error. The range of error is between 3 and $40 \mu\text{m}$, indicating that this monitoring method is feasible. The future work will focus on updating boundary conditions to explore more accurate monitoring models.

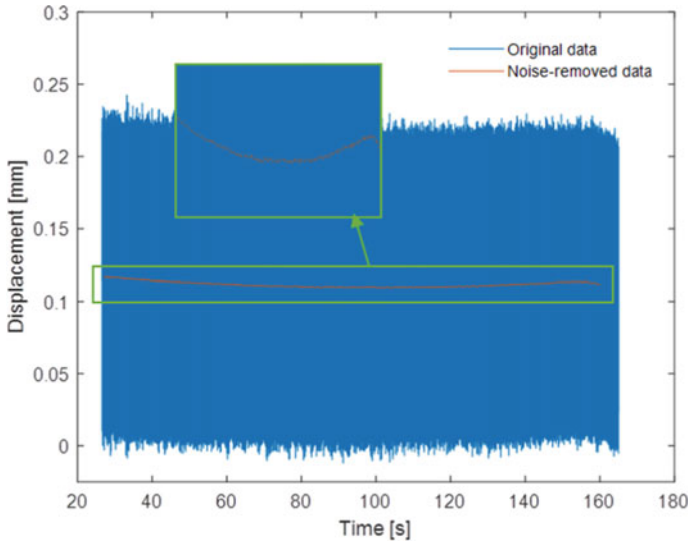


Fig. 5 Workpiece displacement signals during a cutting pass

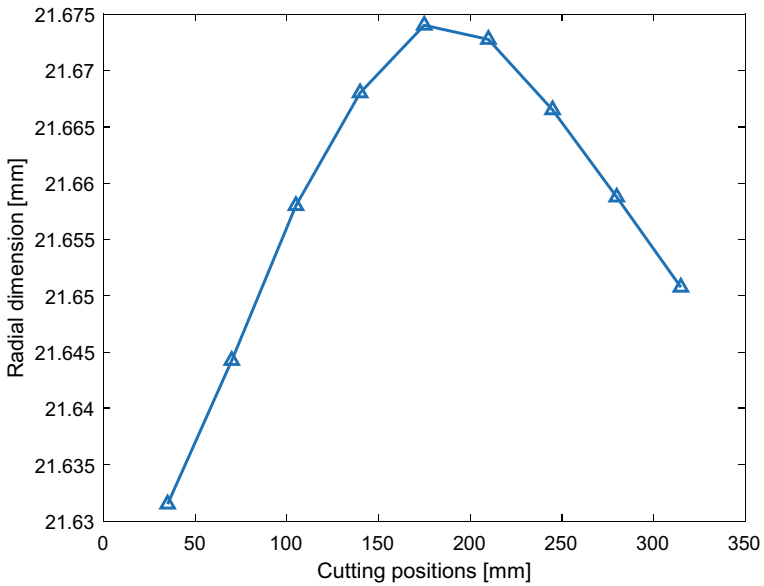


Fig. 6 Radial dimension along the length of the workpiece

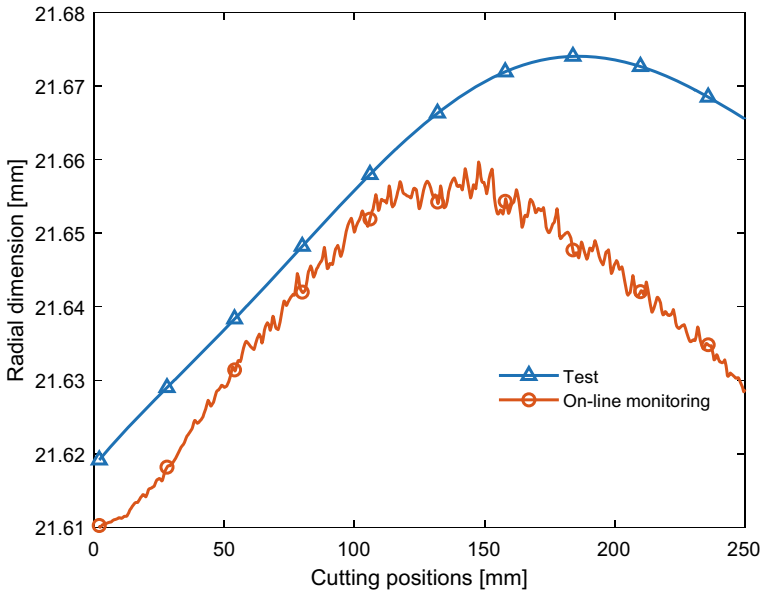


Fig. 7 Comparison of on-line monitoring model and experimental results

Acknowledgements The authors are grateful for the support of National Natural Science Foundation of China under grant No. 51805352, and Shaanxi Key Laboratory of Mine Electromechanical Equipment Intelligent Monitoring under grant No. SKL-MEEIM201904.

References

1. Wang, H., Men, S., Zhang, Y.: Mechanical analysis of machining error in turning slender shafts by reversible turning technique. *Adv. Mater. Res.* 430–433 (2011)
2. Yan, Y., Xu, J., Wiercigroch, M.: Modeling of regenerative and frictional cutting dynamics. *Int. J. Mech. Sci.* **156**, 86–93 (2019)
3. Risbood, K.A., Dixit, U.S., Sahasrabudhe, A.D.: Prediction of surface roughness and dimensional deviation by measuring cutting forces and vibrations in turning process. *J. Mater. Process. Technol.* **132**, 203–214 (2003)
4. Guo, J., Han, R.: A united model of diametral error in slender bar turning with a follower rest. *Int. J. Mach. Tools Manuf.* **46**, 1002–1012 (2006)
5. Manikandan, H., Bera, T.: Modelling of dimensional and geometric error prediction in turning of thin-walled components. *Precis. Eng.* **72**, 382–396 (2021)
6. Wang, T., Zhao, X.: FEA and contrast test of slender shaft turning. *Adv. Mater. Res.* 140–143 (2014)
7. Lu, K., Wang, Y., Gu, F., Pang, X., Andrew, B.: Dynamic modeling and chatter analysis of a spindle-workpiece-tailstock system for the turning of flexible parts. *The Int. J. Adv. Manuf. Technol.* **5**, 3007–3015 (2019)
8. Lu, K., Lian, Z., Gu, F., Liu, H.: Model-based chatter stability prediction and detection for the turning of a flexible workpiece. *Mech. Syst. Signal Process.* **100**, 814–826 (2018)

9. Gong, L.: Optimization of cutting parameters of slender bar based on improved artificial bee colony algorithm. *Appl Mech Mater* 156–162 (2015)
10. Guo, J.: Simulating the diameter error due to the dynamic response of a spinning slender shaft in turning operation. *Simulation* 227–233 (2006)
11. Wu, M., Zhou, T.: Analysis of turning process of slim shaft based on finite element method. *J. Phys.: Conf. Ser. Conf.* **1550**, 042022
12. Yang, L., Wang, P., Zhong, W.: The research on turning operation of varied diameter slender shaft. *Appl. Mech. Mater.* 2621–2624 (2012)
13. He, X., Guo, G., Li, K.: Ultrasonic vibration turning of stainless slender shaft: investigation on run-out of two clamping methods. *Key Eng. Mater.* 990–995 (2016)

Research on Feature Extraction and Recognition of Dongba Hieroglyphs



Hao Huang, Guoxin Wu, and Xiaoli Xu

Abstract The Naxi people in Lijiang, China, have created a pictograph that represents the Dongba culture. The ancient books written with this script are one of the three world heritages in Lijiang, and are known as “the only living ancient script in the world” (Likun in Dongba ancient books and documents of Naxi nationality 2021 [1]). Dongba hieroglyphs are written by different Dongba elders. Different writing habits lead to the phenomenon of variant characters in multiple versions of the same character, as well as the complex structure, different forms, complex background and image noise of Dongba text, this paper puts forward two parts to realize feature extraction and image recognition, topological characteristics and characteristics of grid, Input of neural network training, and combining the neural network using multi-level identification model, template matching and through experiment verification, this algorithm is 9.92% higher than that of the recognition rate of template matching algorithm, and the algorithm of recognition is got improved significantly, the results show that the method to achieve accurate and efficient implementation of Dongba hieroglyphics identification purposes.

Keywords Topological characteristics · Network characteristics · Template matching · The neural network

1 Introduction

It is written by a unique hieroglyphic created in ancient times by the ancestors of Naxi people in China. Dongba script is a kind of hieroglyphic which is very original and expressed in the form of pattern. The local Naxi people call it “senzhuolujo”, which means “the imprint left on the stone and wood” [2]. This hieroglyphic is the only hieroglyphic still in use in the world, It is a cultural treasure of great academic value [3, 4]. In 2003, Dongba classic ancient books of Naxi nationality in China

H. Huang · G. Wu (✉) · X. Xu

Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, Beijing Information Science and Technology University, Beijing 100912, China

e-mail: wgx1977@bistu.edu.cn

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

595

H. Zhang et al. (eds.), *Proceedings of IncoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_48

were listed as “World Memory Heritage” by UNESCO [5]. The book written by this hieroglyph is called Dongba Scripture. Its valuable information records the development and changes of typical community culture in the history of human culture, and accumulates the changes of ancient Chinese culture. There are more than 1800 Dongba hieroglyphs. It is a primitive hieroglyphic symbol system which belongs to the middle development stage of picture and ideograph. Its morphological structure is complex. It is a kind of character symbol between picture and ideograph. In order to recognize Dongba characters, the first step is to extract the characters’ features. The extracted features should be comprehensive and different from other characters. However, Dongba hieroglyphs have many pictorial features, such as complex structure and different strokes, which makes feature extraction more important. In the past few decades, a large number of scientific research has formed a variety of character feature extraction and recognition methods. It includes statistical method, structural method and neural network method [6, 7]. This paper aims to study the character symbol feature extraction and character recognition according to the characteristics of Naxi Dongba characters on the basis of other character feature extraction methods.

2 Recognition System of Dongba Pictograph

Described in this article, Dongba hieroglyphics recognition system is first in Dongba ancient Dongba script for the digital collection, then given the Dongba script itself exists di phenomenon, complicated structure, and the records of the Dongba script Dongba paper itself exists texture feature may cause interference for identifying the Dongba script, so the image information preprocessing operations, Then the noise information in the image is removed and the image is more conducive to the extraction of feature information. Finally, the Dongba text in the image is recognized by the combination of template matching and neural network.

2.1 *The Principle of Dongba Pictograph Recognition*

The recognition of Dongba hieroglyphs is mainly divided into two parts: feature extraction and image recognition. Feature extraction is to obtain digital image information through CCD image acquisition system [8], and then store and convert the JPEG format image collected by CCD into bmp format file that can be processed under Windows system for image preprocessing operation, including de drying, binarization, edge detection, text segmentation, thinning and normalization, which is convenient for feature extraction of Dongba characters [9–11], As shown in Fig. 1, the block diagram of Dongba character recognition system preprocesses the preprocessed Dongba character data by extracting the topological features [12–14] and grid features of Dongba character, and then inputting them into neural network for training to get the weight file of Dongba character recognition. Aiming at the phenomenon that

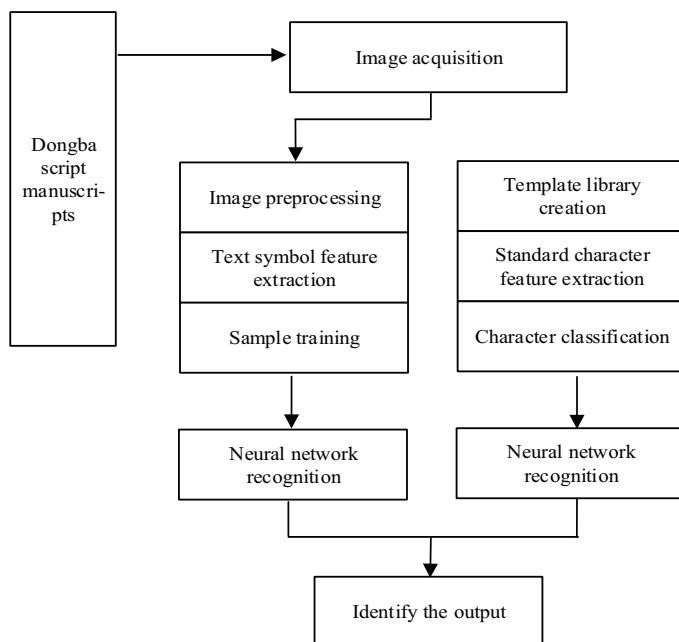


Fig. 1 Block diagram of Dongba character recognition system

the recognition speed of neural network is slower than that of template matching, this paper combines the advantages of the two methods and adopts a multi-level recognition method. The first level uses template matching with high recognition speed [15–17], and the second level uses neural network with higher recognition efficiency to recognize [18, 19], so as to realize the purpose of recognizing Dongba characters.

3 Feature Extraction of Dongba Pictograph

After preprocessing, the image information has a very clear Dongba text individual, which improves the accuracy of the combination of topological features and grid features. In view of the influence of different features, deformation and artificial writing on the same Dongba text, this paper proposes the topology and grid feature extraction to unify the features and strictly classify them. To achieve the text feature extraction and more comprehensive distinction from other types of text.

3.1 Topological Feature Extraction

In view of the influence of different changing features, deformation and artificial writing on the characters of the same Dongba document, a method of topological and grid feature extraction is proposed to achieve the acquisition of feature information by unifying features (1) Endpoint: in graph theory, the vertex with degree 1 refers to the position where strokes disappear in Dongba hieroglyphs, that is, there is a black pixel in the pixel 8 near point P, which is the endpoint.

(2) Cross point: the vertex with degree greater than 2, that is, the intersection of lines, including Trident point and Quad point.

(3) Block: the connected part of graph theory;

(4) Hole: internal hole in graph theory.

As shown in Fig. 2, it is 3 with P as the center pixel × According to this neighborhood, the variable n is defined as:

$$n = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 \tag{1}$$

If the following conditions are satisfied: ① num = 1, ② P = 1, then the point is the endpoint; If the following conditions are satisfied: (1) n = 3 or 4, and (2) P = 1, then the point is a trigeminal point or a quad point.

In the notation in Fig. 3, ○ represents the endpoint, ⇔ represents the block, and □ represents the fork, × Representing holes, the extraction result in the figure is divided into 16 cross points, 2 blocks, 7 holes and 8 endpoints.

Fig. 2 3'3 neighborhood

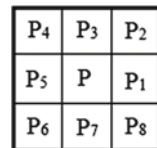
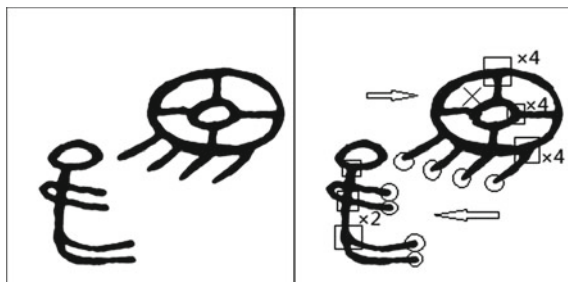


Fig. 3 Schematic diagram of topological structure



3.2 Grid Feature Extraction

Considering the characteristics of Dongba characters expressed in picture form, it is still necessary to extract grid features after topological feature extraction. According to whether the size and position of the mesh are fixed, it can be divided into static mesh feature extraction and dynamic mesh feature extraction.

(1) Static mesh feature extraction

Static grid feature extraction is mainly used to distinguish and extract the features of Dongba characters which have the same composition but have black pixels.

Some Dongba hieroglyphs have the same structural features, but the filling degree of black pixels in some areas is different, as shown in the three groups of Dongba hieroglyphs in Fig. 4.

The process of feature extraction is as follows.

- ① Let the character lattice be $n \times n$. Divide the text image into N parts N grids;
- ② The effective pixels in each grid are calculated and represented by $P_{11}, p_{12}, \dots, p_{1n}, p_{21}, P_{22}, \dots, P_{nn}$;
- ③ Calculate the total effective pixels of text: $P = P_{11} + p_{12} + \dots + p_{1n} + p_{21} + P_{22} + \dots + P_{nn}$;
- ④ Calculate the proportion of the number of black pixels in each grid to the black pixels of the whole text: $PIJ = PIJ * 100 / p$;

The feature vector $p = (P_{11}, p_{12}, \dots, p_{1n}, p_{21}, P_{22}, \dots, P_{nn})$ is the grid feature of the text.

(2) Dynamic grid feature is to distribute the size, stroke form and partially deformed characters of the same type of characters written by old Dongba hands with different writing habits. It plays a good role in solving this kind of phenomenon. For the comprehensive consideration of feature extraction, the black pixels of Dongba hieroglyphic are extracted from four directions [14]

- ① Horizontal component: P1 and P5 are black pixels;
- ② Vertical component: P3 and P7 are black pixels;
- ③ Skimming component: P2 and P6 are black pixels;
- ④ Na component: P4 and P8 are black pixels.

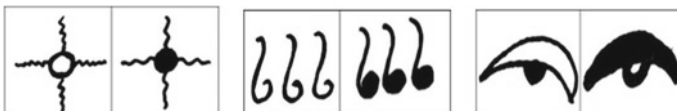


Fig. 4 Dongba characters distinguishable by black pixels

Fig. 5 Decomposition result of Dongba script

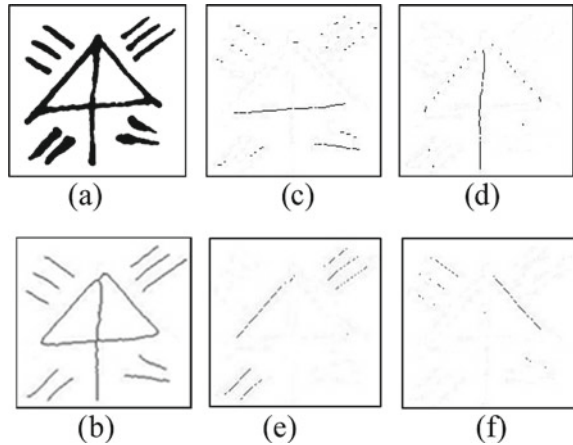


Figure 5 shows the decomposition result of Dongba hieroglyphic.

Figure 5a shows the original Dongba image; Fig. 5b is a refined Dongba image; Fig. 5c is a transverse component image; Figure 5d is a vertical component image; Fig. 5e is a skim component image; Fig. 5f is a component image.

From the following two formulas, the horizontal and vertical non-uniform grid lines II and ij can be obtained respectively, where $N1$ is the number of grids in the horizontal direction and $N2$ is the number of grids in the vertical direction.

4 Recognition of Character Symbol

After feature extraction of character symbols by the above method, the combination of template matching and neural network recognition is used to recognize, so as to increase the accuracy and speed of recognition.

4.1 Template Matching

Template matching needs to establish a text template library, and import all the digital information of Dongba characters in Naxi hieroglyphic spectrum into font creator program through true type technology to build Dongba character template. As shown in Fig. 6, it is the schematic diagram of template matching design.

For the character strokes that are easy to recognize, the template matching is used to recognize the character symbols, which can be classified and recognized by the features of endpoint, cross point, block and hole. The character to be recognized is matched with the part of the template symbol library that has the same topological features [15]. This is a rough classification. The error value of the character to be

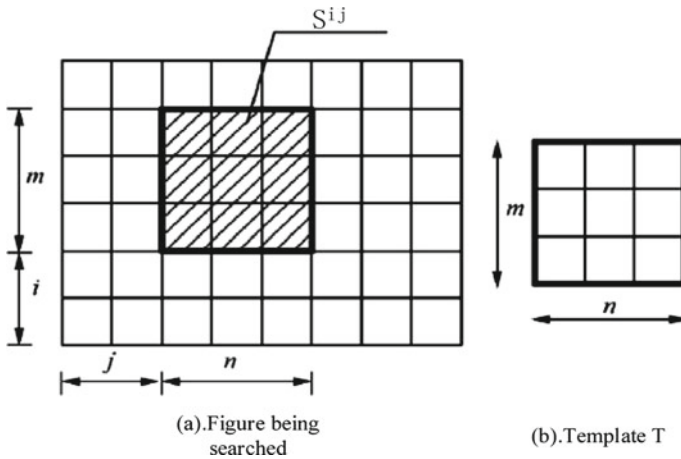


Fig. 6 Template matching design principle

tested and the template character features is calculated by the error method, Set the error to ± 1.5 , That is to say, In other words the threshold of $E(i, j)$. If it is less than 3, we further discuss the uniqueness of the matched Dongba characters. If $E(i, j)$ is greater than 3, the next stage of template matching will be started.

In view of the complex characteristics of Dongba characters, such as diverse structure, different strokes and so on, after rough classification, the remaining parts of the characters can achieve relatively high accuracy through template matching and feature region recognition.

In this step, the similarity value between the above grid features and the sample features is calculated by the similarity method to realize the further recognition of Dongba characters

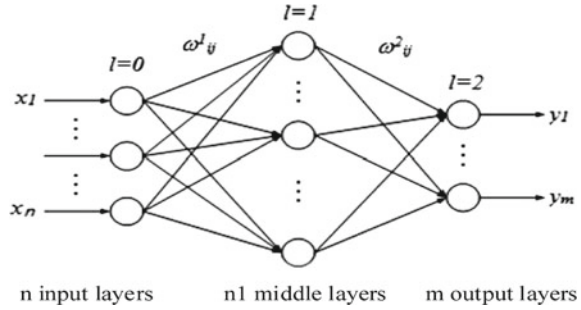
$$R = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [S(i, j) - \bar{S}][T(i, j) - \bar{T}]}{\sqrt{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [S(i, j) - \bar{S}]^2 \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [T(i, j) - \bar{T}]^2}} \quad (2)$$

Among them: $S(m, n)$ is the sample eigenvector, $T(m, n)$ is the feature vector to be identified, \bar{S} , \bar{T} are the average value of the corresponding response vector of the module, As shown in Eqs. 3 and 4

$$\bar{S} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} S(i, j), \quad (3)$$

$$\bar{T} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} T(i, j), \quad (4)$$

Fig. 7 Neural network structure and symbol conventions



R is the correlation value. When the R calculated by the traversal search algorithm is the maximum, the recognition is completed.

Through the above two-step recognition method, the unrecognized characters are verified by experiments to be Dongba characters with different structural features. In view of this kind of similar problem, the application of feature region matching method can achieve the purpose of recognition.

4.2 Neural Network Identification

Neural network recognition is to reduce the mean square error of the maximum prediction output and the expected output, input the sample characters to the hidden layer output through the forward step, and then reverse verify the error of the two, modify the weight of the output layer by the error and replace it to the node of the second layer, and so on to improve the weight of each layer. The structure of the neural network is shown in Fig. 7.

- (1) The minimum random number is used to initialize the weights, and the training time $t = 0$.
- (2) A total of 1000 iterations are used to achieve stability. First, a training sample is obtained in the training set $x = [x_1, x_2, \dots, x_n]^T \in R^n$. The expected output is $D = [D_1, D_2, \dots, D_m]^T \in R^m$.
- (3) Output of neural network with X as input

$$y_r = f \left(\sum_{s=1}^{n_{L-2}} \omega_{sr}^{l=L-1} \dots f \left(\sum_{j=1}^{n_1} \omega_{jk}^{l=2} f \left(\sum_{i=1}^n \omega_{ij}^{l=1} x_i \right) \right) \right), r = 1, \dots, m \quad (5)$$

$f(\cdot)$ is a sigmoid function, ω_{ij}^l This figure represents the weight of node i of layer $l-1$ connected to node j of layer l , It is shown in formula 6

$$f(a) = \frac{1}{1 + e^{-a}} \quad (6)$$

- (4) The weights are adjusted from the output layer (summation layer) and corrected by formula 7

$$\omega_{ij}^l(t + 1) = \omega_{ij}^l(t) + \Delta\omega_{ij}^l(t), \quad j = 1, \dots, n_l, i = 1, \dots, n_{l-1} \quad (7)$$

$$\Delta\omega_{ij}^l(t) = \eta\delta_j^l x_i^{l-1} \quad (8)$$

$\Delta\omega_{ij}^l(t)$ is the weight correction term.

η is learning step size.

- (5) After correcting the weights of all layers, calculate and output again, sort out the error to be less than the set threshold, otherwise set $t = t + 1$, and return to step (2).

5 Experimental Results

In this experiment, 1000 commonly used Dongba hieroglyphs are selected as samples to train the neural network until the parameters of the neural network are stable and all samples can be identified accurately. A total of 2056 Dongba characters in Dongba classic books are randomly selected for the experiment. In the experiment, template matching, neural network and the combination of template matching and neural network are used to recognize the Dongba characters. The recognition effect is shown in Fig. 8.

It can be clearly seen from Fig. 8 that the algorithm in this paper can clearly recognize Dongba characters from Dongba Scripture images. At the same time, the algorithm also accurately corresponds the corresponding Dongba characters with the Unicode codes in the template library and marks them below Dongba characters.

Through the data collection and analysis of the experimental results.

From the experimental results of different recognition algorithms, the number of experimental words, the number of recognition words, recognition rate, recognition time for data collection, get Table 1, as shown in the figure below.

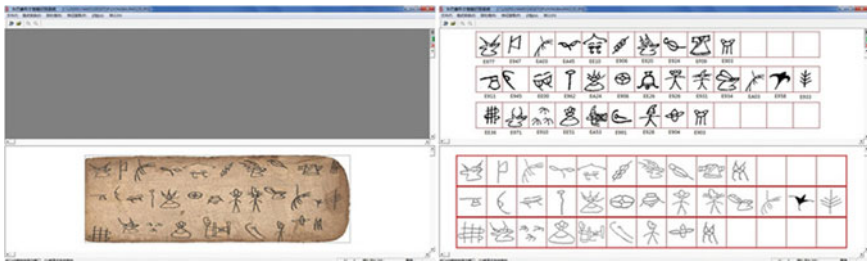


Fig. 8 Recognition effect of Dongba characters

Table 1 Comparison experiment of recognition methods

Recognition methods	Number of experimental words	Number of recognized words	Recognition rate (%)	10 Dongba time (s)
Template matching	2056	1689	82.15	0.3
Neural network	2056	1897	92.27	0.8
Combination of template matching and neural network	2056	1893	92.07	Generally 0.3 ~ 0.5 Maximum 0.8

6 Conclusion

In view of the problems of low accuracy and slow speed of Dongba character recognition caused by the background noise and different forms of Dongba characters in the images of Dongba ancient books, this paper can comprehensively collect the features of Dongba character symbols and distinguish them from other characters by extracting the topological features and grid features of Dongba characters. The recognition of Dongba characters is realized by combining template matching with neural network. Compared with the previous research method, it has the advantages of short recognition time, high accuracy, higher stability, and through a large number of Dongba character recognition training and experimental results, it is found that the system has achieved the practical application purpose of Dongba character recognition, which lays the foundation for the next Dongba culture research.

References

1. Likun, H.: Dongba ancient books and documents of Naxi nationality. *Yunnan Archives* **01**, 46 (2021)
2. Jingting, W.: Research on Dongba pictograph. East China Normal University (2017)
3. Narruo.: Contemporary publication and innovation of Dongba literature. *China Soc. Sci. J.* 2021–02–24 (012)
4. Qunxing, H.: Cultural confidence in the protection and inheritance of Naxi's mother tongue. *Folk Music* **04**, 51–53 (2020)
5. Zhihong, Z.: Research on the protection and inheritance of national intangible cultural heritage—Taking Naxi Dongba painting as an example. *Popular Literature and Art* **22**, 9–10 (2013)
6. Yunmu, W.: Research on Dongba text detection and recognition algorithm based on deep learning. Nanjing University (2020)
7. Houliang, K., Yuting, Y.: Automatic calculation and extraction algorithm of Dongba hieroglyphic feature components. *J. Hubei Univer. Nationalities (Natural Science Edition)* **38**(02), 158–164 (2020)
8. Runhao, Z.: Application analysis of CCD and camera technology in industry. *Appl. Integrated Circuit* **38**(04), 154–155 (2021)

9. Yaohui, G.: Review of preprocessing technology in image character recognition. *Inform. Commun.* **9**, 291–292 (2017). <https://doi.org/10.3969/j.issn.1673-1131.2017.09.155>
10. Zili, W., Hefeng, Z.: Application of wavelet threshold denoising in handwritten numeral recognition image preprocessing. *Comput. Sci. Appl.* **11**, 1969–1976 (2019)
11. Rongxiu, W.: Application of wavelet threshold denoising in handwritten numeral recognition image preprocessing. *Sci. Technol. Econom. Market* **11**, 1–2 (2020)
12. Node proximity preserved dynamic network embedding via matrix perturbation. *Continental Shelf Res.: A Companion J. Deep-Sea Res. Progress in Oceanography* **196** (2020). <https://doi.org/10.1016/j.knosys.2020.105822>
13. Duncan, C.W., Manna, S., Nielsen, A.E.B.. Topological models in rotationally symmetric quasicrystals. *American Family Phys.* **101**(11), (2020). <https://doi.org/10.1103/PhysRevB.101.115413>
14. Robustness on topology reconfiguration of complex networks: An entropic approach. *Mathem. Comput. Simul.* 170379–409 (2020). <https://doi.org/10.1016/j.matcom.2019.11.013>
15. Malek, A.S., Elnahrawy, A., Anwar, H. et al. :Automated detection of premature ventricular contraction in ECG signals using enhanced template matching algorithm. *Biomed. Phys. Eng. Expr.* **6**(1), 015024, 11 (2020). <https://doi.org/10.1088/2057-1976/ab6995>
16. Alessandro Paolo, D., Luigi, G.: GA-adaptive template matching for offline shape motion tracking based on edge detection: IAS estimation from the SURVISHNO 2019 challenge video for machine diagnostics purposes. *Algorithms* **13**(2), 33 (2020). <https://doi.org/10.3390/a13020033>
17. Zhang, L., Zhao, L., Liu, G.: Accelerated generation of spatiotemporal atlas through fast template matching. *J. Comput. Commun.* **8**(01), 16–27 (2020). <https://doi.org/10.4236/jcc.2020.81002>
18. Xinyi, P., Lei, L.: Image recognition system based on deep learning neuralnetwork. *Fujian Comput.* **36**(1), 65–67 (2020). <https://doi.org/10.16707/j.cnki.fjpc.2020.01.022>
19. Fei, W.: Recognition method of station state information based on imageprocessing and neural network. *Railway Commun. Signal* **57**(3), 60–63 (2021). <https://doi.org/10.13879/j.issn.1000-7458.2021-03.20513>

Study on Feature Extraction of Gearbox Vibration Signal for Wind Turbines



Jinang Guo, Guoxin Wu, Xiwei Zhao, Hao Huang, and Xiaoli Xu

Abstract As a clean energy, the development of wind power has attracted wide attention. In view of the characteristics of non-linear and non-stationary mixed signals in the vibration state of wind turbines, the separation of noise is the key problem of information feature extraction. In this study, sensors are utilized to collect blind source signals and mixed matrix information in order to retrieve source signals and extract features from information. This paper integrates EMD (Empirical Model Decomposition) with ICA (Independent Component Analysis) with the aim of extracting feature signals from the wind turbine generator system (WTGS). By analyzing signals with obvious fault characteristics, this approach considerably increases the accuracy in extracting feature signals from the WTGS transmission system.

Keywords Wind turbine · Fault diagnosis · ICA · EMD

1 Introduction

In recent years, the rapid consumption of fossil fuels forces the world's energy needs to turn to the direction of low-carbon consumption, circular development, green environmental protection and renewable new energy. In China, wind power equipment is developing rapidly, and plays an important role in the global wind power industry [1].

Wind turbines often need to be built in areas with sufficient wind energy, such as mountains, grasslands, hills, coastal areas, where there are low temperature, sand-storm, lightning strike, vibration and other adverse weather conditions. Under the direct impact of the working environment, wind turbines are prone to failure, which will lead to increased maintenance costs and operating costs, and seriously affect the production efficiency of wind farms, and the maintenance is difficult, even causing unnecessary safety accidents such as cabin oil fire, endangering personal safety [2].

J. Guo · G. Wu (✉) · X. Zhao · H. Huang · X. Xu
Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, Beijing Information Science and Technology University, Beijing 100912, China
e-mail: wgx1977@bistu.edu.cn

Fault diagnosis of wind turbine is an important means to ensure its safe service. ICA signal processing has the function of recovering the source signal from the mixed signal [3], which can separate the interference and noise signal from the original vibration detection information, and achieve the purpose of feature extraction.

2 Principle of ICA Analysis

ICA analysis is called independent component analysis (ICA) in signal processing, which is a calculation method for separation of multiple components [4]. In the practical application of wind turbine fault diagnosis feature extraction engineering, ICA method, neural network, wavelet analysis and other methods are generally used to separate the source signal from the acquired signal, so as to solve the problem of mixing other interference and noise signals in the signal [5–7].

Firstly, the mathematical formula of obtaining blind source signal by using independent component is as follows

$$x(t) = F(A^\circ s(t)) + n(t) \quad (1)$$

One of them is the observation signal obtained by the channel sensor, that is, the mixed signal, $s(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$. It needs to decompose the mixed signal to obtain the vibration source signal. In order to simulate the actual situation more realistically and satisfy the conditions, the mixed matrix represented by matrix and the nonlinear function represented by function are both unknown quantities.

Basic assumptions required [8]:

- (1) If the probability distribution function of the signal is; Then it satisfies the following probability distribution:

$$p(s) = \prod_{i=1}^n p_i(s_i) \quad (2)$$

- (2) The number of mechanical equipment sources that generate signals m should be less than or equal to the number of sensor channels, that is, the number of observed signals n ;
- (3) If the mixed matrix has full column rank, there is an inverse matrix;
- (4) Only one Gaussian signal can exist in the source signal; And the noise is statistically independent from each source signal, and it is additive white Gaussian noise.

$$x(t) = As(t) + \sum_{j=1}^N \frac{a_j}{c_j} c_j s_j \quad (3)$$

According to the above formula, it is easy to find that the difference between the amplitude of the separated signal and the source signal will not affect the time domain characteristics.

3 Application of Fault Feature Extraction Method Based on EMD and ICA

At present, the common fault feature extraction methods are ICA method, neural network [5–7], empirical mode decomposition (EMD), wavelet analysis and many other signal processing methods (Fig. 1).

3.1 The Principle of the Method Combined with ICA and EMD

To make up for the defects of both, we propose a signal processing method combining EMD-ICA [10]. The process is as follows:

- (1) EMD method is used to analyze the signal from the sensor channel, and the IMF component is obtained;
- (2) The virtual noise channel signal is composed of IMF components with larger correlation coefficients;
- (3) Fast ICA process signal and noise channels, separate the source signal components, and then analyze the diagnosis. The process is shown in Fig. 2

3.2 Simulation Analysis and Experimental Verification of EMD and ICA

- (1) Simulation application and analysis of joint method

Three simple sine base signals are set. The frequencies are respectively: 10, 20, 50 Hz, and 0 phase difference. Set the amplitudes of the three signals to be 1, 2, and 0.5 mm respectively. The time domain and frequency domain waveforms of the mixed signal generated by the source signal and the mixing matrix are shown in Fig. 3.

EMD algorithm is used to decompose the mixed signal to get IMF component, as shown in Fig. 4.

Therefore, we use ICA method to recover the feature information lost after EMD decomposition and restore the inherent characteristics of the signal.

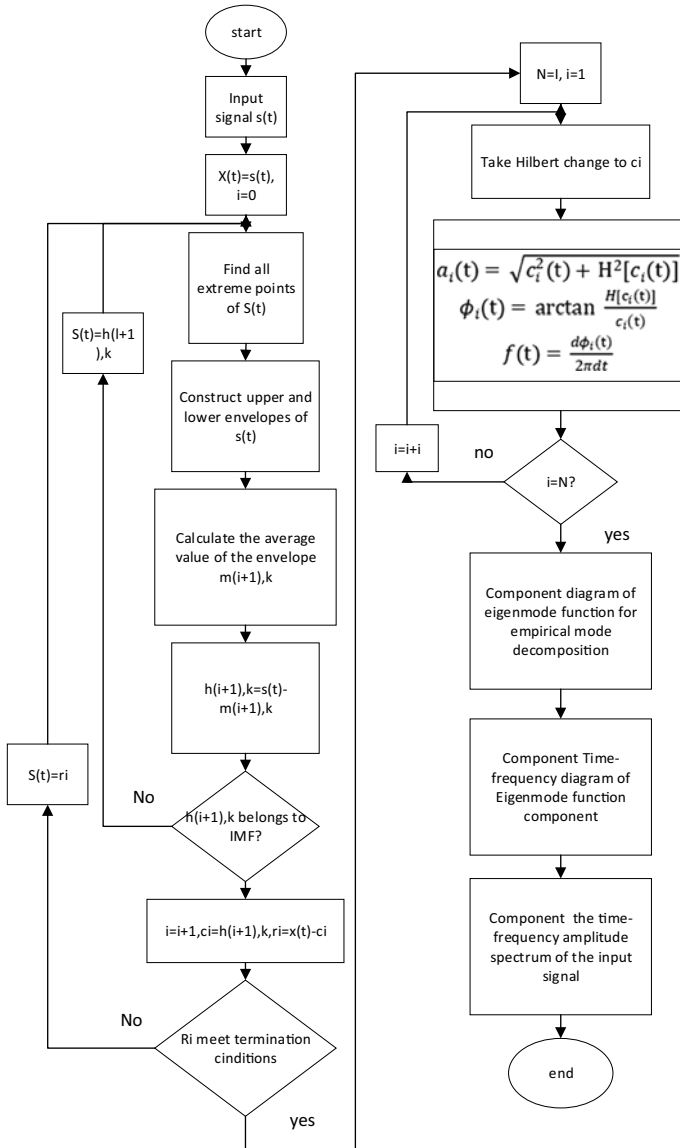


Fig. 1 EMD algorithm flowchart

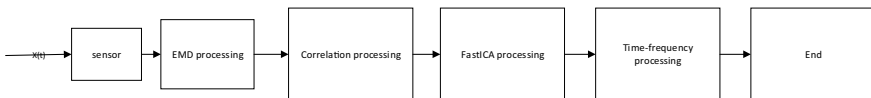


Fig. 2 EMD-ICA analysis signal flow chart

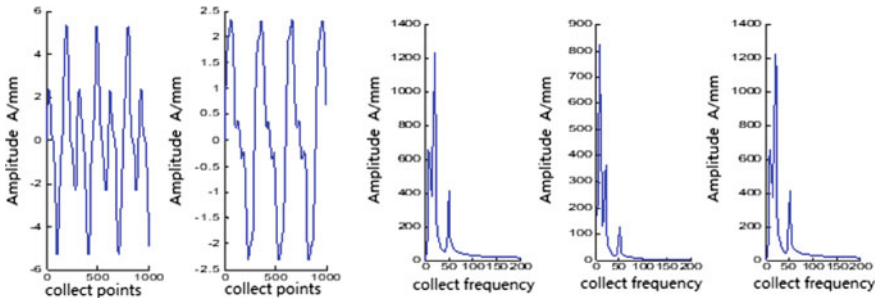


Fig. 3 Time domain diagram and frequency domain diagram

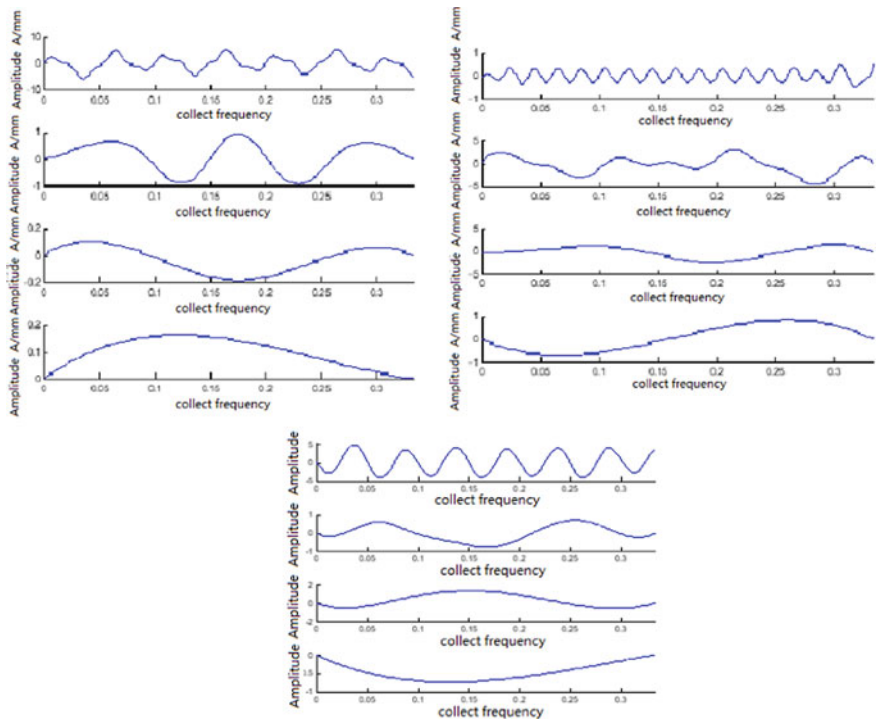


Fig. 4 IFM component mixed signal

Through the correlation coefficient processing, the correlation value between each IMF component and each mixed signal is obtained. As shown in Table 1.

Because the threshold is fixed and the removal coefficient is small, the components with large coefficients are added together with their respective mixed signals as the input matrix of ICA, and the fast ICA method is used to obtain the time domain and frequency domain diagrams, as shown in Fig. 5.

Table 1 Correlation coefficient of simulation data

	Z1(1)	Z1(2)	Z1(3)	Z1(4)
X(1,:)	0.9748	0.3314	0.0044	0.0072
X(2,:)	0.8365	0.0018	0.0713	-0.0520
X(3,:)	0.9930	0.0092	0.0315	-0.0716

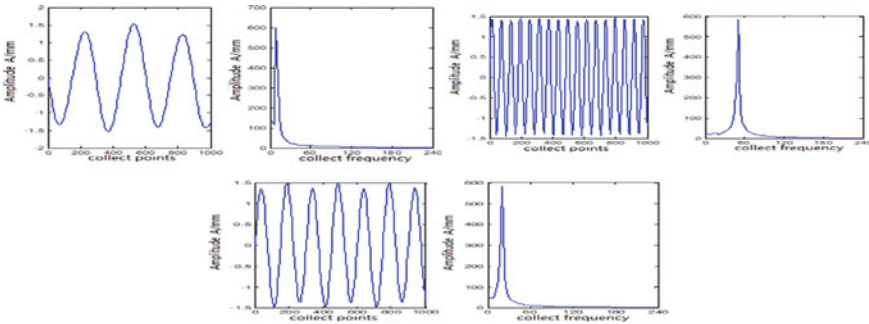


Fig. 5 Time–frequency diagram of Fast ICA decomposition frequency of 10, 20, 50 Hz

The results show that: after EMD-ICA separation, the time-domain and frequency-domain diagrams are basically consistent with the base signal, and the lost feature information and the inherent characteristics of each signal after EMD decomposition are restored.

(2) Experimental verification

Through the vibration test-bed of wind turbine drive system, the actual bearing fault is simulated, and the normal operation and fault signal of the equipment are analyzed. The bearing model is nsk6025zz. The running speed of the faulty rotor is 1750r / min, the sampling frequency is 12 kHz, and 2048 sampling points are selected. The feature frequency of each part, the fault frequencies of outer ring, inner ring and rolling element are 104.57, 157.94 and 137.48 Hz respectively. Real data are detected in the experiment, and their time domain data and IMF components separated from EMD are shown in Fig. 6.

The correlation coefficient is obtained by operation, as shown in Table 2. At the same time, the time domain and frequency domain are transformed by fast ICA, as shown in Fig. 7.

After calculation, the characteristic frequency of rolling bearing fault is $f_s = 157.2$ Hz, which corresponds to the results obtained by EMD-ICA combined method. Through the data simulation and the actual verification of the experimental platform, it is found that this method can extract the characteristics of each IMF component, solve the aliasing and loss of EMD components, and overcome the condition that the

Fig. 6 Time-domain plot and decomposition IMF component plot

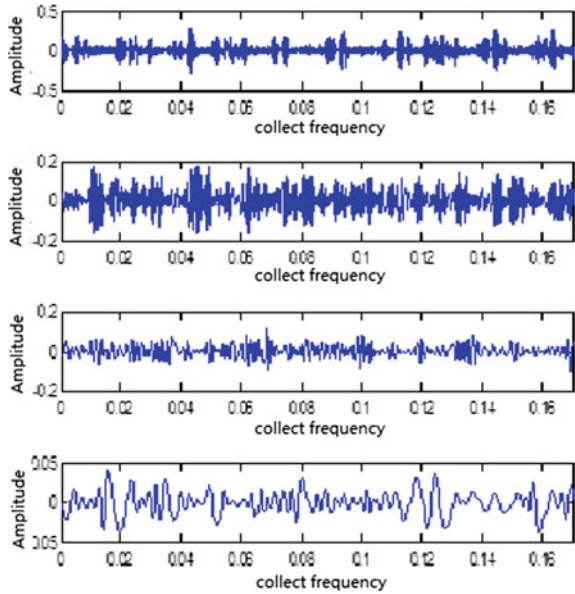


Table 2 Correlation coefficients of experimental data

Z1(1)	Z1(2)	Z1(3)	Z1(4)	Z1(5)	Z1(6)
0.7348	0.6082	0.3013	0.0919	0.0344	0.0000

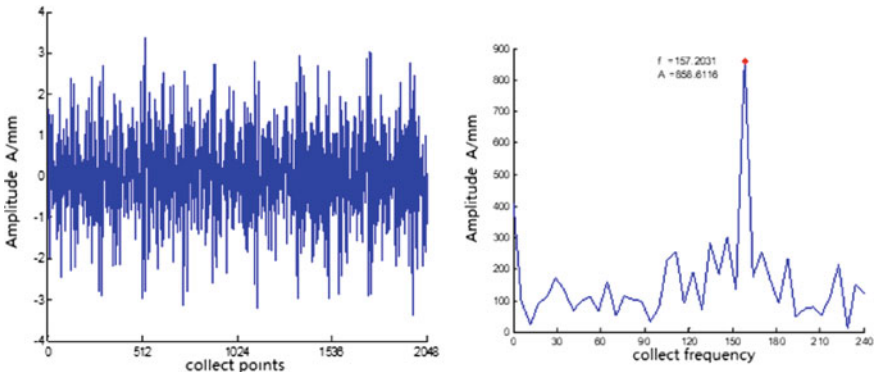


Fig. 7 Time and frequency domain diagram of fast ICA transform

detected signal of ICA is larger than the source signal, and can complete the analysis and extraction of fault characteristics of wind turbine bearing vibration signal.

4 Conclusions

By studying the basic principle, advantages and disadvantages of EMD method, combined with the processing of mixed simulation signal, the ability of EMD method to decompose nonlinear and non-stationary signal is verified. Taking advantage of the advantages and disadvantages of ICA and EMD signal processing methods, a method combining EMD-ICA is adopted to extract fault features of vibration signal of wind turbine gearbox. Data analysis shows that the joint processing method can achieve the same signal characteristics.

References

1. Global Wind Report: Annual Market Update (2014). <http://www.gwec.net/publications/global-wind-report-1,2014>
2. Dajun, X., Jin, Z., Yunsheng, G., et al.: Status quo and trend analysis of wind power fire protection technology. *Fire science and technology* **12**, 1407–1410 (2013)
3. Bin, L., Xieting, J.: Linear distortion image correction based on blind separation. *J. Fudan Univer.* **34**(2), 185–190 (1995); Dajun, X., Jin, Z., Yunsheng, G. et al.: Analysis of current situation and trend of wind power fire protection technology. *Fire Protection Sci. Technol.* **12**, 1407–1410 (2013)
4. Gang, L., Hang, W., Wenjun, M.: Bit vibration signal recognition method based on independent component analysis. *Sci. Technol. Eng.* **18**(16), 33–37 (2018)
5. Yan, C.: Application of ICA and adaptive hybrid intelligent algorithm. *J. Zhejiang Univer. Water Resour. Hydropower* **31**(1), 68–72 (2019)
6. Yuegang, L., Yangyang, H.: Application of EWT and ICA in bearing fault diagnosis. *Vibration and Shock* **38**(16), 42–48, 70 (2019)
7. Lizheng, P., Dashui, Z., Shigang, S.: Research on gear fault diagnosis based on wp-ica and SVM. *Comput. Simul.* **37**(1), 411–416, 483 (2020)
8. Shunmo, L.: Blind source separation technology and application of vibration signa. pp. 24–181. Beijing, Aviation Industry Press (2011)
9. Xi, Y.S., Song, H.J., Tong, W.Z. et al.: Comparison of time-frequency analysis of vibration signals of rotating machinery based on Hilbert transform and wavelet transform. *Chinese J. Electri. Eng.* **06**, 102–107 (2003)
10. Yubiao, S., Baojie, X., Guoxin, W. et al.: Application of EMD and ICA in fault diagnosis. *Manuf. Autom.* **13**. 89–92+118 (2014)

Condition Monitoring of a Reciprocating Air Compressor Using Vibro-Acoustic Measurements



Debanjan Mondal, Fengshou Gu, and Andrew D. Ball

Abstract Fault diagnosis in reciprocating compressor (RC) requires time-consuming feature-extraction processes due to the complexity of the compressor operation and fluid–solid interaction. This causes the useful information to be corrupted and difficulty in accurately diagnosing the faults with traditional methods. The aerodynamic phenomenon has a large impact on acoustics signal compared to the vibration. Thus, this paper presents analytical modelling of compressor sound highlighting the important sound sources and their generation. The additional contribution of this paper is the application of a state-of-the-art signal processing technique: Modulation Signal Bispectrum (MSB) which overcomes the challenges by showing good noise suppression capability and characterising the modulating components present in the signal, thereby resulting in stable modulation components for accurate diagnostics. The result reveals that the fault diagnosis based on airborne acoustics using MSB method outperformed the vibration-based method.

Keywords Vibro-acoustic analysis · Sound source analysis · MSB analysis · Reciprocating compressor · Fault diagnosis

1 Introduction

The condition monitoring (CM) of the reciprocating compressor is not an easy task because of its complicated dynamic coupling of its operational physics. Different strategies have been developed to detect and diagnose the compressor components based on different sensing methods that include visual inspection, vibration analysis [1–7], motor current signature analysis [8–11], acoustic emission measurement [12–14], airborne acoustic analysis (noise analysis) [15–19], temperature and thermal

D. Mondal · F. Gu (✉) · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK
e-mail: F.Gu@hud.ac.uk

D. Mondal
e-mail: Debanjan.Mondal@hud.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of IncoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_50

615

image [20] etc. These strategies are important to implement so that the maintenance actions can be taken in a more accurate and reliable way before a fatal damage occurred to its components. An extensive literature review has shown that majority of the study is focused on the vibration-based diagnosis method with a vast application of various signal processing techniques often combined with artificial intelligence techniques such as artificial neural network [5, 6, 21, 22], support vector machine [23], decision tree [24], etc.

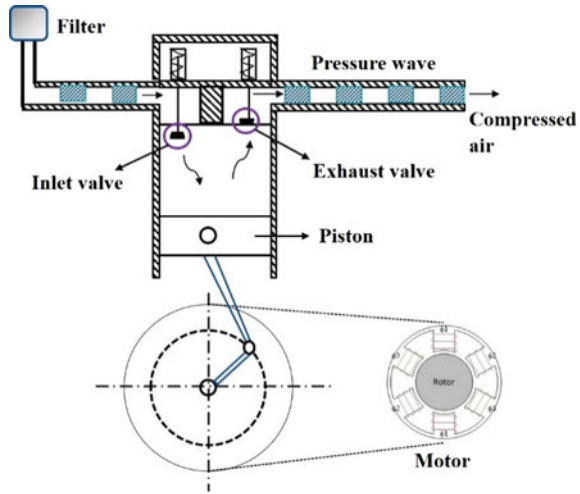
Acoustic signals can be advantageous in the context due to its non-intrusive [25] measurements. Moreover, its capability of gathering information from multiple sources with fewer sensors make the CM method easier and less expensive to implement [25]. Furthermore, technological advancement in the form of MEMS microphones [26] makes it possible to install those in on-board system. The typical advantage of using acoustic method is that it can cover a wide range of frequencies using a single microphone. A drawback of using acoustic signal is that it can be contaminated with huge background noise [27] from unwanted noise sources and thus making the post processing methods very much important. Due to that reason, a limited number of studies based on acoustics have been performed on RC. The main difficulty of implementation of this strategy comes in two points: (1) choosing a proper signal processing method based on the types of the signals and types of the machine components, (2) developing a feature extraction method that will allow to detect the abnormalities present and their possible causes.

The novel contribution of this paper lies on two aspects: (1) Analytical modelling of RC vibro-acoustic sources and (2) Effectiveness of using MSB for airborne acoustic signal in diagnosis of compressor faults for characterising the typical modulating components and suppressing the noise. Section 2 discusses about the vibro-acoustic of RC highlighting the possible sound sources and their relationship with the vibration for normal operating condition and with the presence of faults. It also addresses the challenges of using traditional signal processing methods and specify the need for using MSB. Section 3 highlights the capability of MSB in extracting the features from airborne sound signal of the compressor with the presence of faults. The concluding remark can be found in Sect. 4 which also paves the path for future study.

2 Key Properties of Vibroacoustic in RC

According to the working process and structure of a typical RC, illustrated in Fig. 1, its vibroacoustic stems from of many different sources, but they can be largely categorised into mechanical, aerodynamic, and electromechanical sources. In this way, various sources and their interactions can be examined with sufficient details so that their amplitude, phase, and frequency behaviours can be understood with great details, providing fundamentals for processing measured signals, and developing accurate and reliable diagnostic tools.

Fig. 1 Basic vibroacoustic sources of a reciprocating compressor operation



2.1 Vibration Characteristics of Healthy RC

According to research works [28] vibration signal from a RC cylinder head is composed of flow-induced and impact induced excitations. The flow-induced excitations are caused by air interactions with valves resulting in periodic flow oscillations, while impact induced excitations are caused by the effects of the valve plate hitting the seat when opening and closing. The working process of RC is illustrated in Fig. 1.

The reciprocating motions of piston, torsional oscillation of cranks, belt radial and longitudinal oscillations and various impacts between cylinders and pistons, valve plate landing, bearing clearances contributes to the high-level vibration. Moreover, high pressure oscillation inside pipe and cylinders also excites the structure vibrations. In electric motor, the electromagnetic forces create air gap between stator and rotor which leads to the generation of radial forces. These radial forces cause unbalance which contributes to the vibration of rotor–stator module. The flow of gas through the discharge chamber and piping system of the RC are unsteady and contains time varying pulses superimposed on the steady (average) flow. This causes the pressure pulsation through the valves of the RC cylinder head during the compression process within its cycles. These pulses are made-up of the geometrical, physical and mechanical characteristics of the compressor [29].

This pressure pulsation is generally too weak to cause any problem. However, it may be amplified due to coincidence with an acoustic natural frequency of the piping. The resulting resonance can induce severe piping vibration and/or malfunction of instruments, compromising plant operation. Pressure pulsations in piping induce excitation forces. Large vibration occurs when the pressure pulsation frequency matches the mechanical frequency of the piping. Figure 2 shows the typical one compression cycle of time domain waveform for vibration and two cycles for acoustics at the discharge pressure of 120 psi under healthy condition.

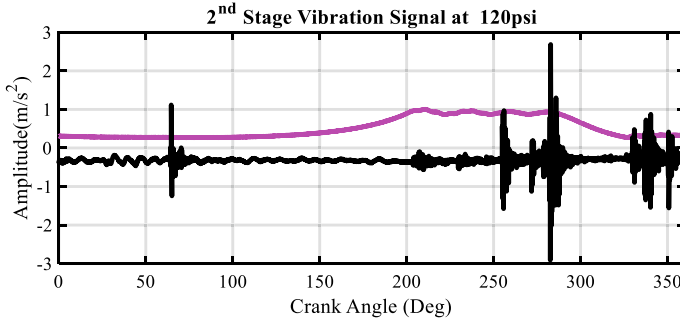


Fig. 2 Vibration signal of healthy RC at 120 psi

Figure 2 shows the Significant impacts are observed and they occur at the discharge valve closing (DVC) time for the second-stage cylinder. The black colour curve represents the vibration signal and the curve in purple colour represents the in-cylinder pressure. The high amplitude impacts result from high pressure air acting as a resistant force in the discharge plenum of the cylinder causing the valve to close harshly. Furthermore, the overall vibration amplitudes from the second-stage cylinder head are greater than those from the first-stage RC cylinder head, because the second-stage cylinder compresses gas at a higher pressure. From the figure, the transient response in the vibration signal signifies the opening and closing of the valves with impacts, more details can be found in the previous study [16].

If Z denotes the input impedance of piston of mass M_p with velocity V_0 then after impact, the piston moves with a velocity v_1 formulated as [32]:

$$v_1 = V_0 e^{-t/\tau} \tag{1}$$

where $\tau = M_p/Z$

Therefore, after applying Fourier transform to Eq. 1, the vibration velocity spectrum can be written as,

$$|V(\omega_v)| = (V_0\tau/\pi)[1 + (\omega_v \tau)^2]^{-1/2} \tag{2}$$

where symbol ω_v denotes the circular vibration frequency.

2.2 Acoustics Characteristics of Healthy RC

The sound can come from (1) vibrating surface radiation, (2) the aero-acoustic response of the system to air motions and (3) electromechanical force from the AC motor. An aerodynamic source could be defined as a source of sound caused by air

perturbation: phenomena within the discharge and intake valves, pipe can be considered as aerodynamic sources. A mechanical source generates sound by vibrating structures mentioned earlier. The interactions between these physical sources are complicated but can be largely examined according to the flow diagram in Fig. 3.

This vibration due to electromechanical force from three phase induction motor contributes to the pressure variations in air and perceived as sound of periodical characteristics. The motor sound can also be classified in three categories: (a) electromagnetic sound: magnetic force and variation of flux density, (b) mechanical sound: bearings, unbalance, and mechanical resonances, (c) aerodynamic sound: rotor slots. The sound generated from electrical motor is mainly of non-stationary, broadband noise.

The mechanical sound of the compressor has three significant sources: (1) sound from belt drive, (2) valve impact sound and (3) sound due to impacts on piston kinematic chain. If the belt is worn, it may cause slippage while accelerating, which in terms contributes to the squealing sound. Moreover, the valve impacts on their seats make a periodic impulse noise during their usual opening and closing. It is possible that the amplitude of this sound gets further amplified with the mechanical resonance. Apart from that, the non-linearity effect of the valve fluttering can also be seen in the sound signal as a form of harmonics. In broad sense, the characteristics of sound sources is harmonic due to the periodic nature of the compression process. The impacts on the piston kinematic chain are also a source of compressor body vibration which eventually leads to the generation of mechanical sound.

The radiated sound power W [31] is equal to:

$$W = \frac{Sp^2}{\rho c} \cos \theta = \rho c S \sigma \langle v \rangle^2 \tag{3}$$

where

ρ = density of the air, σ = radiation ratio (0 ~ 1), c = speed of sound.

S = surface area, p = sound pressure, $\langle v \rangle$ = surface-averaged vibration velocity,

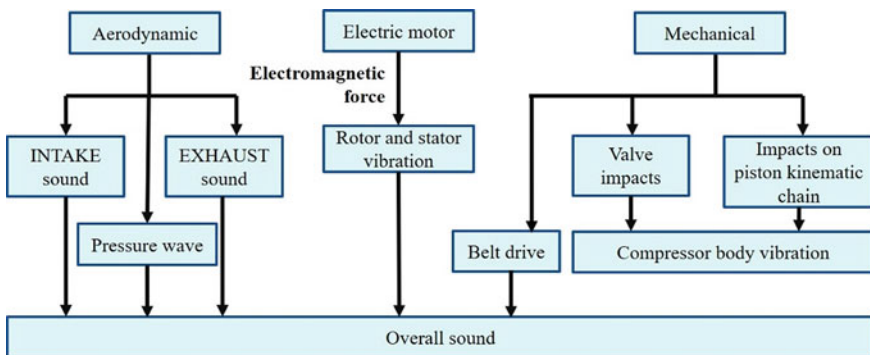


Fig. 3 Generation of vibration and noise

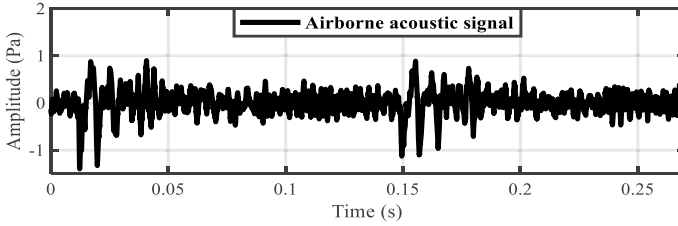


Fig. 4 Acoustic signal of healthy RC at 120 psi

θ = angle between the direction of propagation of the sound and the normal to the surface.

ρ , c are constant and the value of σ is very small, hence the sound pressure P is equivalent to surface-averaged vibration velocity v .

The vibration velocity can be estimated from the vibration acceleration as follows:

$$|v| = \frac{|a|}{i\omega} \quad (4)$$

$|a|$ is the vibration acceleration, $i = \sqrt{-1}$, and ω is the circular frequency.

The air flow inside the cylinder also generates sound in terms of pressure wave. The aerodynamic sound has a large impact on acoustic signals obtained from RC shown in Fig. 4.

The sound pressure p for an air discharge as a sound source can be written as [30]:

$$p = \frac{\rho c}{A} q(t) \quad (5)$$

where ρ is the density of air, c is the speed of sound, A is the cross-sectional area and $q(t)$ is the difference between the total flow volume velocity and the flow DC-component.

From Fig. 4 it can be observed that the acoustic signal not only shows the valve impacts or mechanical sound source, but it also reflects the air flow sound, aerodynamic sound sources which can be useful for determining the compressor properties under different RC conditions considering the physical phenomenon of the compression process.

2.3 Vibro-Acoustics of RC for Faulty Conditions

Three common faults: intercooler leakage (ICL), discharge valve leakage (DVL) and filter blockage (FB) [16] have been simulated which largely affect the vibro-acoustic signature of RC. The intercooler pipe is a section where pressure pulsation signal

is very much dominated. The leakage in the pipe not only reduces the compressor efficiency but also effect the vibroacoustic phenomenon contributing to the airborne sound. The leakage in discharge valve is very common in industries. With the presence of this fault irregular intake-exhaust takes place that also affects the opening and closing of the cylinder valves. Hence the valve-sitting impact and aerodynamic property are very much affected. Similarly, a clogged filter reduces the quantity of airflow to the RC, creating a vacuum condition inside which also greatly affects the vibro-acoustic phenomenon of RC. Figure 5 shows the vibroacoustic signals recorded for different compressor conditions under a particular discharge pressure.

When the intercooler leakage occurs the pressure inside the first stage cylinder is lower than the baseline one. This causes a reduction of force to keep the first stage discharge valve closed. Thus, the discharge valve opens earlier in first stage cylinder with intercooler leakage fault, whereas in second stage, the intercooler leakage causes a small change in the cylinder pressure; the pressure inside the cylinder is lower than in healthy operation.

This will result in reduction in discharge efficiency. The drop in the cylinder pressure causes a delay in the opening of both suction and discharge valves. The discharge valve leakage leads to a slight but significant increase in the pressure of the first stage cylinder. This increase in the cylinder pressure causes the first stage suction and discharge valves to open early due to the higher pressure in the cylinder.

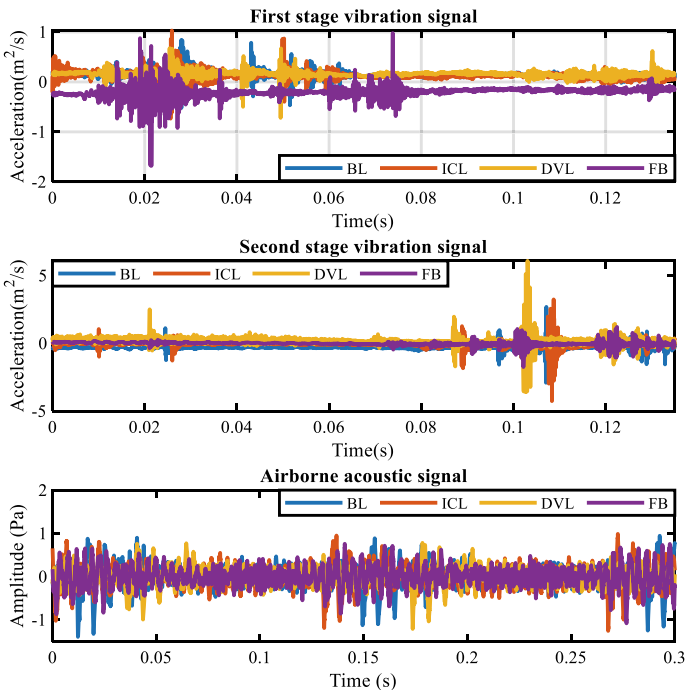


Fig. 5 Vibration and acoustic signals of RC for different conditions at 120 psi

The second stage discharge valve leakage causes both valves open and close earlier than for the healthy operation. The discharge valve leakage fault causes the second stage cylinder pressure to build-up earlier than the healthy operation; thus, the process of discharge occurs earlier.

Moreover, due to the presence of faults, the aerodynamic phenomenon has a large impact on the acoustic signals adding some more components to the signal due to the turbulence thus making it more difficult to extract the submerged information.

Malfunctional of reciprocating machine components results in excessive vibration and noise subsequently. The multiphysics application consists of heat transfer, fluid flow, mechanical motions, electrical to mechanical input, fluid-flow interaction make the reciprocating compressor a complex machine and thus condition monitoring of such machines are highly challenging. The non-stationary behaviour, non-linearity of the compressor valve motions, structural/acoustic resonance, pulsation make it even more difficult for applying a proper signal processing technique of interest. In a nutshell, the vibro-acoustic signal of RC is periodic, impulsive, non-linear, non-stationary, noisy and has a modulating characteristic due to the presence of structural resonance. Because of the abovementioned reason, state of the art signal processing technique like MSB [10] is required for the analysis of these type of signals as it uses quadratic phase coupling (QPC) to handle the non-linearity, highlights the effective modulating components considering the both sidebands around center frequency and has ability to suppress the noise shown in previous study [33].

3 Result and Discussion

Placement of the sensors in right place is very important to get the useful signals that reflect the fault characteristics of the respected machine. The accelerometers should be put on a surface where it can generate signal output that is characteristic of major events that are responsible for compressor vibration such as, opening and closing of valve, fluid–solid interaction, flow through valve etc. As present of any faults in reciprocating compressor affect the compression process and the valve movements, the vibration signals will have different signatures. For this reason, accelerometers were put on the two cylinders heads of high-pressure cylinder and low-pressure cylinder to acquire vibration signals that can monitor the performances of the valves, piston-cylinder arrangement. A microphone was placed at 60 cm from the compressor body at the side opposite to the flywheel minimising the effect of reverberation property of the sound [34]. The vibration signal was acquired simultaneously with the acoustic signal under a broad range of discharge pressure between 40 and 120 psi.

3.1 Vibration Based Diagnosis

Figure 6 shows the MSB amplitude of the vibration signal for different compressor conditions at two different discharge pressures of 60 and 110 psi within which the compressor normally operates.

The reason for choosing higher frequency range for the MSB calculation of vibration signal can be explained as valve excitation can be seen clearly in the time domain waveform, these impulses are of high frequency. The impulsive nature of the vibration signal has the characteristic frequencies in high frequency zone. Apart from that higher MSB coherence values help to identify the signal components very effectively from the signal in particular frequency range that implies the frequency zone is very much influenced by the modulating components and side bands. Hence the particular band 3200–5600 Hz is selected for the MSB magnitude calculation [16].

From Fig. 6, we can see there are three frequency regions where MSB amplitudes are dominant typically in the range of 3200–4000, 4000–4800 and 4800–5600 Hz. It is also noticeable that the MSB peaks are higher for faulty conditions and depends on the discharge pressure as the vibration energy is also high for higher discharge pressure.

Subsequently, a new three-dimensional feature parameter can be presented dependent on the averaged peak amplitudes of the MSB for certain frequency bands under various pressure conditions to differentiate diverse machine conditions. To get a characteristic feature vector, three frequency bands are selected: $F1 = 3200\text{--}4000$ Hz,

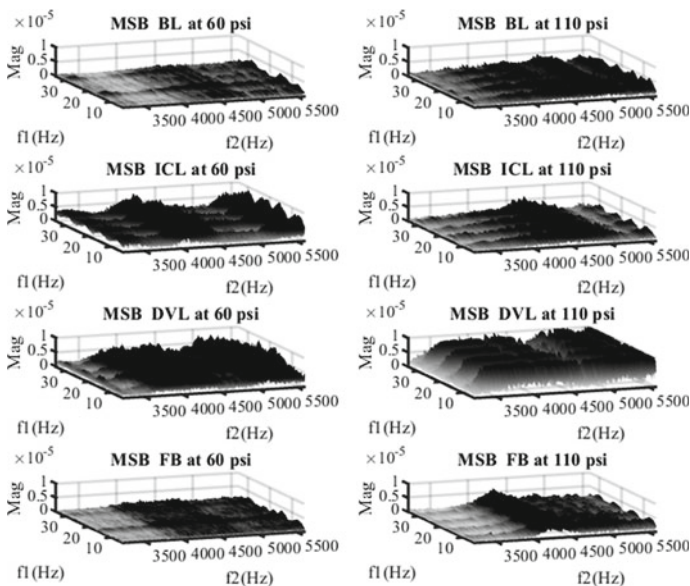


Fig. 6 MSB amplitudes of the vibration signals for different conditions.at 60 and 80 psi

F2 = 4000–4800, F3 = 4800–5600 Hz. Figure shows the averaged MSB amplitudes for different frequency bands under a broad range of discharge pressures. Finally, a three-dimensional feature vector is obtained combining F1, F2 and F3 and has been shown in Fig. 7.

Each figure was obtained by sufficient amount of averaging (187 times) to guarantee the steady results. First the frequencies along with f_1 are averaged. Then they are further averaged along f_2 for a particular frequency band.

In high frequency zone the MSB analysis can face some challenges as the higher order harmonics and their sidebands can be far away from each other making it difficult for identification and easily can be mixed with the other harmonic components. Therefore, choosing a low frequency range is always better while analysing MSB. In the case of compressor vibration signals the characteristic properties are visible in higher frequency range and hence high frequency range was selected for the MSB analysis. The analysis results show that process based on MSB cannot differentiate the compressor conditions accurately since carrier frequencies of the higher order harmonics are mixed with the other frequency components and hence difficult to represent them.

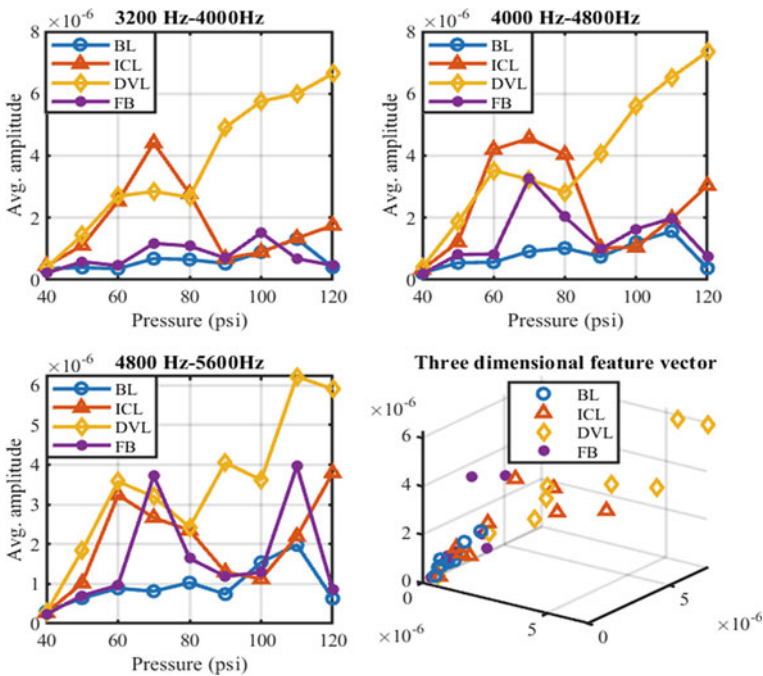


Fig. 7 Average MSB (vibration signals) for different compressor conditions under a broad range of pressures

3.2 Acoustics Based Diagnosis

Figure 8 shows the MSB magnitudes of the acoustic signals for different conditions at the discharge pressure of 60 psi and 110 psi for frequency range of 100–530 Hz as the structure borne sound has low frequency components.

From the analysis result shown in Fig. 8, the higher MSB amplitudes are distributed over a broad range of frequency. From the analysis results, it can be found that the MSB presents the distinctive peak patterns in presence of faults across the carrier frequency, and this is often due to the abrupt change of the air flow, fluttering of valves during the time of suction and discharge processes.

To get a characteristic feature vector, three frequency bands are selected: F1 = 100–200 Hz, F2 = 240–300, F3 = 400–530 Hz. Figure shows the averaged MSB amplitudes for different frequency bands under a broad range of discharge pressures. Finally, a three-dimensional feature vector is obtained combining F1, F2 and F3 and has been shown in Fig. 9.

In the case of compressor acoustic signal, the characteristic properties are visible in lower frequency range and hence low frequency range was selected for the MSB

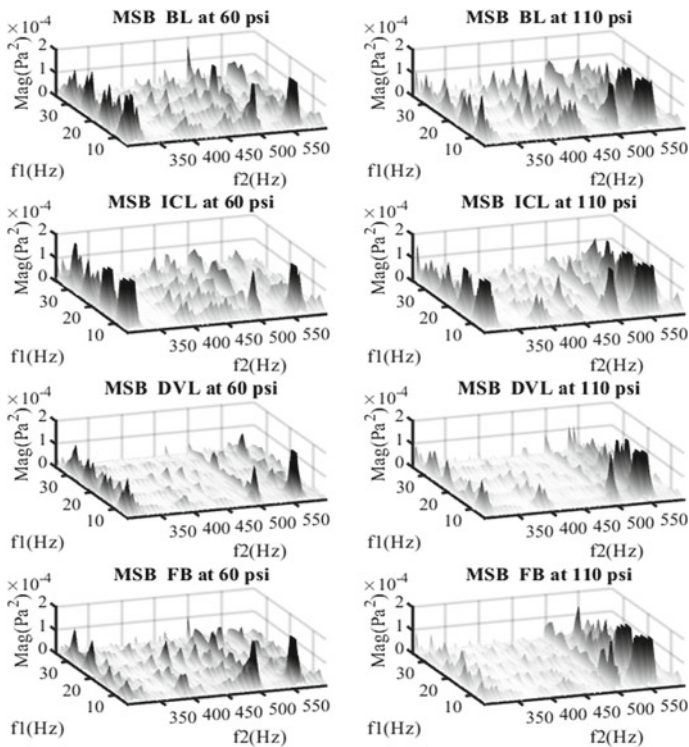


Fig. 8 MSB amplitudes of the acoustic signals for different conditions.at 60 and 80 psi

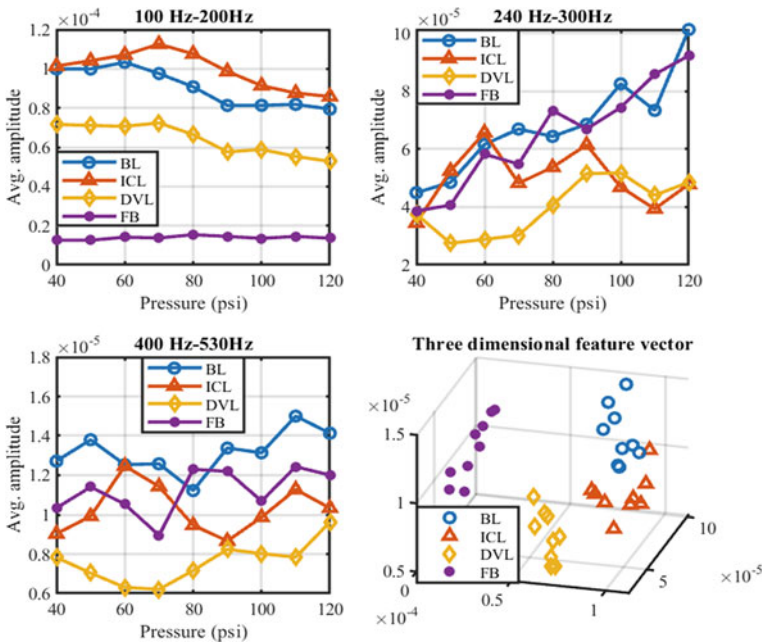


Fig. 9 Average MSB (acoustic signals) for different compressor conditions under a broad range of pressures

analysis which gives excellent result in terms of separating different compressor conditions from each other as a form of clusters.

The effective determination of the compressor condition based on the acoustic analysis was accomplished by utilizing the MSB method. The proposed method can separate the compressor conditions under a broad range of discharge pressure.

4 Conclusion

The vibro-acoustic measurement of the compressor shows the comparative study of the compressor sound signal with the vibration. The aerodynamic phenomenon and fluid–solid interaction have major effect on the acoustic signals compared to the vibration. The traditional signal processing methods are not enough to identify faults as the nature of the compressor signal is very complex. MSB has been proved to be an effective tool to differentiate the compressor conditions. The key contributions of the study include the successful identification of vibro-acoustic sources by analytical modelling and their generation which will pave the path for dynamic modelling of RC in the next study and most importantly MSB shows the promising result in terms

of noise suppression by multiple averaging, highlighting the relevant modulating components and differentiating the various RC conditions.

References

1. Rao, A.R., Dutta, B.: Vibration analysis for detecting failure of compressor blade. *Eng. Fail. Anal.* **25**, 211–218 (2012)
2. Ahmed, M. et al.: Fault detection and diagnosis using principal component analysis of vibration data from a reciprocating compressor. In: *Proceedings of 2012 UKACC International Conference on Control*. IEEE (2012)
3. Bardou, O., Sidahmed, M.: Early detection of leakages in the exhaust and discharge systems of reciprocating machines by vibration analysis. *Mech. Syst. Signal Process.* **8**(5), 551–570 (1994)
4. Haiyang, Z., et al.: A compound interpolation envelope local mean decomposition and its application for fault diagnosis of reciprocating compressors. *Mech. Syst. Signal Process.* **110**, 273–295 (2018)
5. Tran, V.T., et al.: Single and combined fault diagnosis of reciprocating compressor valves using a hybrid deep belief network. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **232**(20), 3767–3780 (2018)
6. Yang, H.-B. et al.: Fault diagnosis of reciprocating compressor based on convolutional neural networks with multisource raw vibration signals. *Mathem. Problems Eng.* (2019)
7. Zhang, Y., Sheng, H., Hu, G.: Study on compressor fault diagnosis based on space reconstruction and kI Transform. *Compressor Technol.* **4**, 19–21 (2011)
8. Gu, F., et al.: Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment. *Mech. Syst. Signal Process.* **25**(1), 360–372 (2011)
9. Zhen, D., et al.: Fault diagnosis of motor drives using stator current signal analysis based on dynamic time warping. *Mech. Syst. Signal Process.* **34**(1–2), 191–202 (2013)
10. Gu, F. et al.: Motor current signal analysis using a modified bispectrum for machine fault diagnosis. In: *2009 ICCAS-SICE*. IEEE (2009)
11. Haba, U. et al.: Modelling and simulation of a two stage reciprocating compressor for condition monitoring based on motor current signature analysis. *Condition Monitor. Diagnostic Eng. Managem.* **2** (2018)
12. Wang, Y., et al.: Fault diagnosis of reciprocating compressor valve with the method integrating acoustic emission signal and simulated valve motion. *Mech. Syst. Signal Process.* **56**, 197–212 (2015)
13. Ali, S.M., et al.: Automated valve fault detection based on acoustic emission parameters and support vector machine. *Alex. Eng. J.* **57**(1), 491–498 (2018)
14. Sim, H.Y., et al.: Detection and estimation of valve leakage losses in reciprocating compressor using acoustic emission technique. *Measurement* **152**, 107315 (2020)
15. Lin, J.: Feature extraction of machine sound using wavelet and its application in fault diagnosis. *NDT and E Int.* **34**(1), 25–30 (2001)
16. Mondal, D. et al.: Airborne acoustic signature analysis for fault diagnosis of reciprocating compressors using modulation signal bi-spectrum. In: *25th International Conference on Automation and Computing (ICAC)*. IEEE (2019)
17. Mondal, D., Gu, F., Ball, A.: Application of minimum entropy deconvolution in diagnosis of reciprocating compressor faults based on airborne acoustic analysis. In: *16th International Conference on Condition Monitoring and Asset Management (CM 2019)* (2019)
18. Mondal, D. et al.: Fault diagnosis of reciprocating compressor using empirical mode decomposition-based teager energy spectrum of airborne acoustic signal. In: *Advances in Asset Management and Condition Monitoring* pp. 939–952. Springer (2020)

19. Verma, N.K., et al.: Intelligent condition based monitoring using acoustic signals for air compressors. *IEEE Trans. Reliab.* **65**(1), 291–309 (2015)
20. Deng, R., et al.: Object-based thermal image segmentation for fault diagnosis of reciprocating compressors. *Sensors* **20**(12), 3436 (2020)
21. Zhao, D. et al.: Intelligent fault diagnosis of reciprocating compressor based on attention mechanism assisted convolutional neural network via vibration signal rearrangement. *Arabian J. Sci. Eng.* 1–14 (2021)
22. Ahmed, M., Gu, F., Ball, A.: Feature selection and fault classification of reciprocating compressors using a genetic algorithm and a probabilistic neural network. *J. Phys.: Conf. Ser. IOP Publishing* (2011)
23. Ren, Q., Ma, X., Miao, G.: Application of support vector machines in reciprocating compressor valve fault diagnosis. In: *International Conference on Natural Computation*. Springer (2005)
24. Prashanth, K., Elangovan, M.: Vibration based fault monitoring of a compressor using tree-based algorithms. In: *IOP Conference Series: Materials Science and Engineering*. IOP Publishing (2019)
25. Delvecchio, S., Bonfiglio, P., Pompoli, F.: Vibro-acoustic condition monitoring of Internal combustion engines: a critical review of existing techniques. *Mech. Syst. Signal Process.* **99**, 661–683 (2018)
26. Rinaldi, G., et al.: Dynamic pressure as a measure of gas turbine engine (GTE) performance. *Meas. Sci. Technol.* **21**(4), 045201 (2010)
27. Gu, F. et al.: The condition monitoring of diesel engines using acoustic measurements part 1: acoustic characteristics of the engine and representation of the acoustic signals. *SAE Technical Paper* (2000)
28. Nakamura, T. et al.: In: *Flow-induced vibrations: classifications and lessons from practical experiences*. Butterworth-Heinemann (2013)
29. Shejal, P.P., Desai, A.: Pulsation and vibration study of reciprocating compressor according to API 618 5 th edition. *Int. J. Modern Eng. Res.* **4**(7)
30. Tanttari, J.: On twin-screw compressor gas pulsation noise. In: *The 29th International Congress and Exhibition on Noise Control Engineering* (2000)
31. Harrison, M.: In: *Vehicle refinement: controlling noise and vibration in road vehicles*. Elsevier (2004)
32. Ungar, E.E., Ross, D.: Vibrations and noise due to piston-slap in reciprocating machinery. *J. Sound Vib.* **2**(2), 132–146 (1965)
33. Mondal, D. et al.: Reciprocating compressor fault diagnosis based on airborne acoustic signal: comparison study between modulation signal bi-spectrum and teager energy spectrum method. In: *International Conference on Maintenance Engineering*. Springer (2020)
34. Mondal, D. et al.: A study of diagnosing reciprocating compressor faults using EMD-entropy of the airborne acoustic signals. In: *International Conference on Maintenance Engineering* pp. 293–307. ShieldCrest Publishing (2018)

Application of Combined Normalized Least Mean Square and Ensemble Empirical Mode Decomposition Denoising Method in Fault Diagnosis of Rolling Bearings



Changsheng Xi, Jie Yang, Dong Zhen, Xiaohao Liao, Wei Hu, and Fengshou Gu

Abstract Rolling bearings are widely used in modern machinery and equipment, and the tough working environment is easy to cause their failure. To solve the problem of extracting fault signals of rolling bearings in a strong noise environment, a method based on Normalized Least Mean Square (NLMS) adaptive filtering and Ensemble Empirical Mode Decomposition (EEMD) noise reduction method is proposed. Firstly, NLMS is used to filter the signal, which is used for primary noise reduction. Then the signal is decomposed into a series of Intrinsic Mode Functions (IMFs) by EEMD, and the kurtosis value, root mean square value and sample entropy value of each IMF are calculated respectively. The appropriate one is selected according to the comprehensive index. Finally, the signal is reconstructed and the Hilbert transform is performed on the reconstructed signal to obtain the envelope spectrum, and the fault characteristic frequency is extracted. Simulation and experimental results show that the method can effectively reduce noise and successfully extract fault features.

Keywords Normalized least mean square · Ensemble empirical mode decomposition · Rolling bearings · Fault diagnosis · Denoising

C. Xi · J. Yang (✉) · D. Zhen
School of Mechanical Engineering, Hebei University of Technology, Tianjin 300131, China
e-mail: Yangjie871@126.com

C. Xi
e-mail: x757256720@163.com

F. Gu
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

X. Liao · W. Hu
WORLDTECH Transmission Technology Ltd, Tianjin 300401, China

1 Introduction

Rolling bearings are an important part of rotating machinery and equipment, which has the functions of bearing load, damping, reducing friction and so on. In complex operating conditions and tough working environments, rolling bearings are particularly vulnerable to damage, which affects the normal operation of the equipment and causes economic losses [1]. Therefore, timely and accurate detection of rolling bearing faults is of great significance to the normal operation of the equipment and safe production.

Periodic impact signals will be generated after rolling bearing failure, but these signals are easily submerged in strong background noise and difficult to extract [2, 4]. Nowadays, mechanical equipment fault diagnosis has been widely concerned by researchers, and many advanced signal processing methods have been developed, such as wavelet transform (WT), empirical mode decomposition (EMD), local mean decomposition (LMD), intrinsic time-scale decomposition (ITD) and variational modal decomposition (VMD). These methods have achieved good diagnostic results, but they also have many limitations. For example, WT can process non-stationary signals, but it lacks the adaptability of local features [3]; EMD can decompose signals adaptively, but its decomposition results have some defects such as boundary effects and modes mixing [1]; LMD is a new adaptive time–frequency analysis method, but its calculation process is complicated and the calculation efficiency is low [5]; The ITD improves the computational efficiency, but its decomposition results are distorted [6]; VMD has a high decomposition accuracy, but it is difficult to determine its parameters [7]. When in a strong noise environment, the above methods can achieve a certain noise reduction effect, but the stability is poor, and the fault signal cannot be effectively extracted. Therefore, a method for extracting the fault characteristics of rolling bearings in a strong noise environment needs to be developed.

Normalized least mean square (NLMS) is an adaptive filtering method. It is the result of step size normalization based on the Least mean square algorithm. It has the advantages of simple calculation and strong stability [8, 9]. NLMS can automatically adjust the filter parameters to filter out noise from the actual signal to achieve optimal filtering. Ensemble empirical mode decomposition (EEMD) is an improvement of EMD, which can effectively solve the defect of EMD modal mixing [11]. EEMD can decompose signals into a series of intrinsic mode functions (IMFs), and each IMF may contain fault signals. Therefore, the selection of IMFs is the key to fault diagnosis. At present, there are many methods for selecting IMFs, such as correlation coefficient [13], permutation entropy [14] and kurtosis. These methods are selected based on a single index and are easily affected by various factors. To improve the reliability of the selected IMFs, this paper uses a comprehensive index composed of kurtosis, root mean square and sample entropy to select IMFs.

Therefore, this paper proposes a new method which is used to extract the characteristic frequency of the fault signal in a strong noise environment. Firstly, NLMS is used to filter the signal. Then, EEMD is used to decompose the filtered signal.

The appropriate IMFs are selected to reconstruct the signal through the comprehensive index. Finally, the Hilbert transform is used to analyze the envelope of the reconstructed signal and extract fault features. The rest of the article is organized as follows: Sect. 2 introduces the related theories; Sect. 3 introduces the process of the proposed method; Sects. 4 and 5 carry out simulation and experimental verification respectively, and Sect. 6 gives some conclusions.

2 Basic Theories

2.1 Normalized Least Mean Square

Normalized least mean square adaptive filtering is an improvement of the Least mean square adaptive filtering. The basic principle of NLMS is to normalize the step size factor of the LMS algorithm, and it se a variable step-size method to shorten the adaptive convergence process, thereby solving the problem of the gradient noise amplification caused by the LMS algorithm when the input signal is large and improve the convergence speed of the algorithm. The specific steps of the NLMS algorithm are as follows [10]:

- (1) Initialization parameters. Determine the step size factor μ and the order M of the adaptive filter.
- (2) Determine the initial condition of weight $w(0)$ based on prior knowledge.
- (3) Calculate the output value. The input signal of the filter at time n is $x(n)$, and the output signal is calculated:

$$y(n) = w^T(n)x(n) \quad (1)$$

- (4) Calculate the error. The ideal signal is $d(n)$, then the output error $e(n)$ of the filter is:

$$e(n) = d(n) - y(n) \quad (2)$$

- (5) Normalize the step size factor.

$$\mu(n) = \frac{\mu}{\|X(n)\|^2} \quad (3)$$

- (6) Weight coefficient update:

$$w(n+1) = w(n) + 2\frac{\mu}{\|X(n)\|^2 + \gamma}e(n)x(n) \quad (4)$$

where γ is a constant($0 < \gamma < 1$).

- (7) Add 1 to time n , return to step (2), repeat the above steps, and get the final filtered signal.

2.2 Ensemble Empirical Mode Decomposition

Ensemble empirical mode decomposition is an improvement of EMD, Its essence is to add the Gaussian white noise to the original signal, and use the characteristic of zero white noise to eliminate the modal aliasing phenomenon that appears in EMD. The decomposition results of the original signal are averaged many times to cancel the noise, so as to eliminate the influence of noise on the original signal. The specific decomposition steps of EEMD are as follows[12]:

- (1) Set the average processing times m . Initial $I = 1, 2, \dots, m$.
 (2) Add Gaussian white noise $n_i(t)$ to the original signal $x(t)$ to obtain a set of new signals $y_i(t)$.

$$y_i(t) = x(t) + n_i(t) \quad (5)$$

- (3) Perform EMD on the new signal $y_i(t)$ to obtain a series of IMFs.

$$y_i(t) = \sum_{j=1}^n c_{i,j}(t) + r_{i,n}(t) \quad (6)$$

where n is the number of IMFs, $c_{i,j}(t)$ are the IMFs, and $r_{i,n}(t)$ are the residual components.

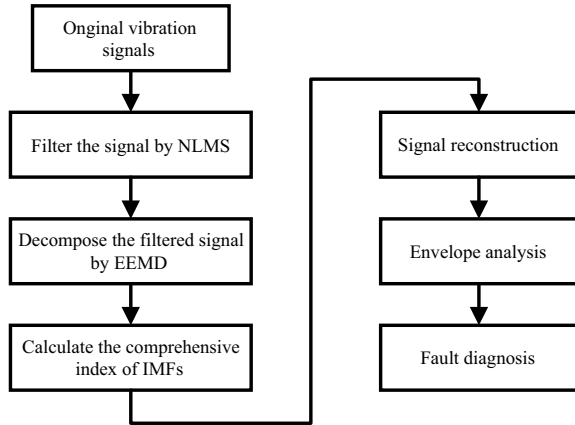
- (4) Repeat steps (2) and (3) m times, adding Gaussian white noise of different amplitude each time, and get a series of IMFs through EMD decomposition.

$$c_{1,j}(t), c_{2,j}(t), \dots, c_{m,j}(t) \quad (7)$$

- (5) All the obtained IMFs are averaged to obtain the IMF component $c_n(t)$ decomposed by EEMD.

$$c_n(t) = \frac{1}{m} \sum_{i=1}^m c_{i,n}(t) \quad (8)$$

Fig. 1 The algorithm flow chart of NLMS-EEMD method



3 Application of Combined NLMS and EEMD Denoising Method

To solve the problem of extracting fault signals of rolling bearings in a strong noise environment, a method based on NLMS adaptive filtering and EEMD noise reduction method is proposed. The process of this method is shown in Fig. 1, and the specific steps are as follows:

- (1) NLMS adaptive filtering is used to denoise the original signal for the first time, and the filtered signal with a high signal-to-noise ratio is obtained.
- (2) The signal is decomposed into a series of IMFs by EEMD.
- (3) Calculate the kurtosis, mean square error and sample entropy of each IMF, normalize these indicators and sum them to obtain comprehensive indicators, and select appropriate IMFs through the comprehensive indicators to reconstruct the signal to obtain the reconstructed signal.
- (4) The reconstructed signal is transformed into Hilbert transform and its envelope is solved, the characteristic frequency of the fault is extracted, and the fault diagnosis is performed.

4 The Simulation Analysis

In order to verify the effectiveness of the proposed method in the fault diagnosis of rolling bearings, the simulation signal is defined as[4]:

$$\begin{cases} x(t) = e^{-\alpha t_1} \sin(2\pi f_g t_1) \times (\sin(2\pi f_r t) + 3) + n(t) \\ t_1 = \text{mod}(t, 0.98/f_m) \end{cases} \quad (9)$$

where $\alpha = 1000$ represents the decay factor, $f_r=30$ Hz is the rotating frequency, $f_g=2000$ Hz is the resonant frequency, and $f_m=76$ Hz indicates the fault frequency. $n(t)$ is Gaussian white noise with a SNR of -15.44 dB. The sampling point is 6000 and the sampling frequency is 12,000 Hz.

Figure 2a shows the original waveform of the impact signal, and Fig. 2b shows the time waveform of mixed-signal $x(t)$. Figure 2c shows the spectrum of signal $x(t)$, and Fig. 2d is the Hilbert envelope spectrum of signal $x(t)$. The red arrow in the figure indicates the characteristic frequency of the fault. It can be clearly seen from the figure that the periodic impact signal generated by the rolling bearing fault is completely submerged in the noise, and the fault characteristic frequency cannot be extracted from its spectrum and the Hilbert envelope spectrum.

To extract the fault characteristics of rolling bearings in a strong noise environment, the NLMS-EEMD method proposed in this paper is used to process the simulation signal. For the NLMS adaptive filtering algorithm, the step factor μ is set to 0.05, the order M of the filter is set to 2, and the constant γ is set to 10^{-4} . Figure 3 shows the signal with noise filtered by NLMS. It can be seen that noise and other interference components have been significantly reduced, but there is still a lot of

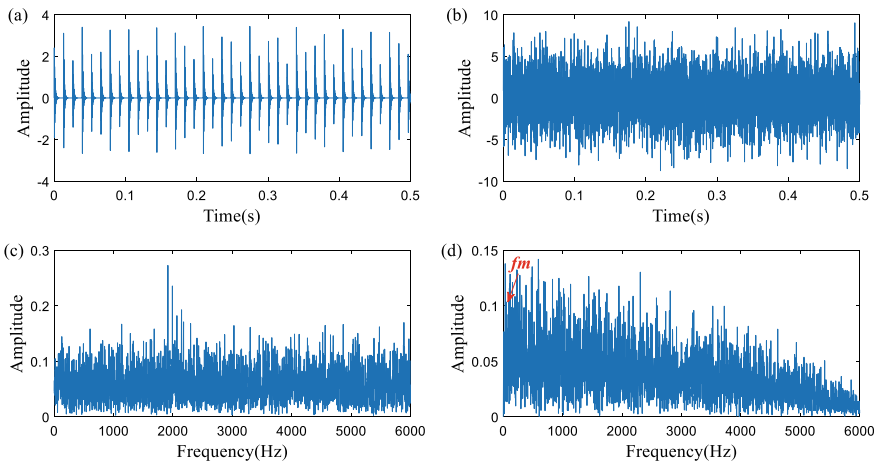
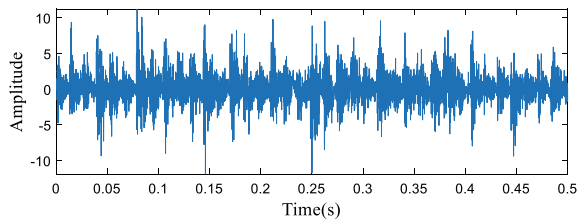


Fig. 2 The simulation signal: **a** the original waveform of impact signal; **b** the time waveform of mixed signal $x(t)$; **c** the spectrum of signal $x(t)$; **d** Hilbert envelop spectrum of signal $x(t)$

Fig. 3 The signal with noise filtered by NLMS



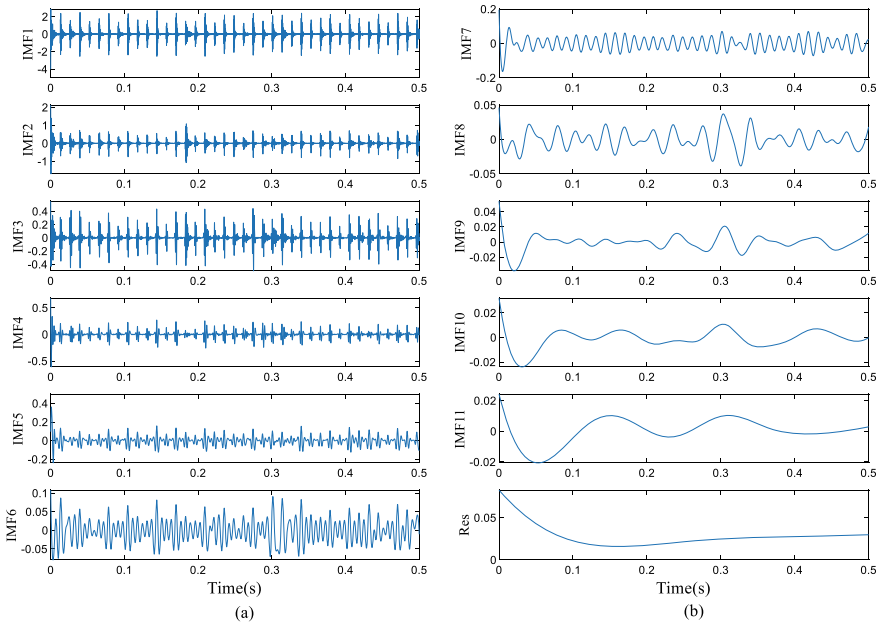


Fig. 4 The IMFs decomposed by EEMD: **a** IMF1–IMF6; **b** IMF7–IMF11 and the residual components

noise. Then the filtered signal is decomposed by EEMD to obtain 11 IMFs and a residual component. The result of EEMD decomposition is shown in Fig. 4.

Wang et al. [15] proposed a comprehensive evaluation method for IMFs using multiple indexes, and good results were obtained through verification. Inspired by it, this paper uses a comprehensive index method to select appropriate IMFs. Kurtosis is very sensitive to impact signals, and it is suitable for detecting fault signals. The root mean square can reflect the degree of dispersion of the signal. The sample entropy represents the complexity of the signal, the larger the sample entropy, the more information it contains. Therefore, this paper chooses kurtosis, root mean square and sample entropy as comprehensive index to improve the accuracy of selecting IMFs. To avoid the differences caused by the different dimensions of every index, each index is normalized and summed to get the comprehensive index f . The calculation formula is as follows:

$$f = \sum_{t=1}^k \frac{f_{te}}{\max(f_i)} \tag{10}$$

where, f_{te} is the index value of the e -th IMF under the t -th index, $t = 1, 2, 3$ respectively represents the kurtosis value, root mean square value and sample entropy value (The calculation formula of the three indexes can refer to [14] and [15]). Calculate the

kurtosis, mean square error, and sample entropy of each IMF, and normalize and sum these single indicators to obtain a comprehensive index. The calculation result of the comprehensive index is shown in Fig. 5.

It can be concluded from Fig. 5 that the comprehensive index of IMF1 and IMF2 is relatively large, which can well reflect the state and information of the signal. Therefore, IMF1 and IMF2 are selected to reconstruct the signal to obtain a reconstructed signal. Finally, the reconstructed signal is transformed by Hilbert transform and analyzed by envelope analysis. Figure 6a shows the Hilbert envelope spectrum of the reconstructed signal. From Fig. 6a, the fault characteristic frequency f_m and its

Fig. 5 The comprehensive index of each IMF

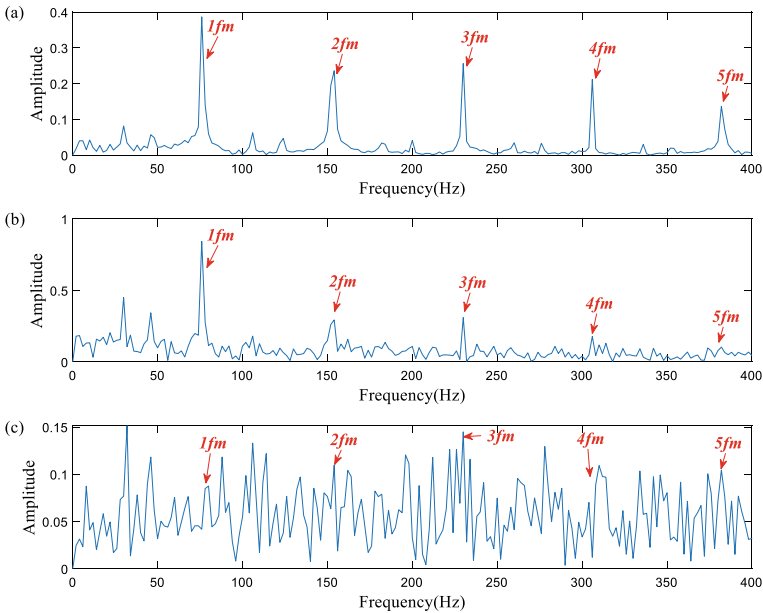
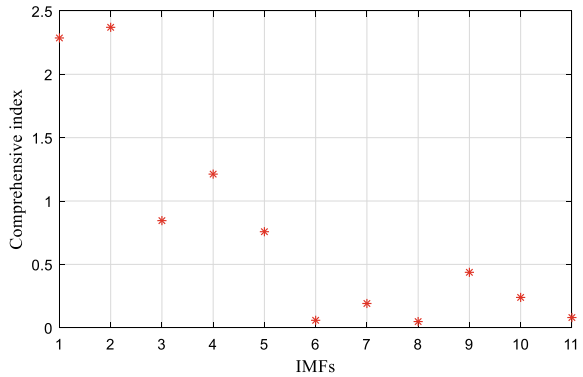


Fig. 6 Envelope analysis results of three methods

Table 1 The SNR of NLMS-EEMD, NLMS and EEMD after noise reduction

Signal	SNR/dB
NLMS noise reduction signal	2.1569
EEMD noise reduction signal	0.4187
NLMS-EEMD noise reduction signal	11.9401

Table 2 The main parameters of the test bearings and fault frequency

Bearing type	Ball diameter d (mm)	Pitch diameter D_m (mm)	Ball number z	Contact angle α	Fault frequency f_o (Hz)
6206ZZ	9.53	46.4	9	0°	89.33

harmonics can be clearly seen. In order to illustrate the noise reduction advantages of NLMS-EEMD, NLMS noise reduction and EEMD noise reduction are performed on the noisy signal respectively. Figure 6b shows the envelope spectrum of the NLMS filtered signal, in which the fault frequency is not obvious and there is a strong background noise and interference frequencies. Figure 6c shows the envelope spectrum of the EEMD filtered signal. The frequency in the figure is very chaotic, indicating that EEMD cannot effectively reduce background noise in a strong noise environment.

To evaluate the effect of noise reduction, the signal-to-noise ratios of NLMS-EEMD, NLMS and EEMD after noise reduction were calculated, and the results are shown in Table 1. It can be seen from Table 2 that the signal-to-noise ratio of NLMS-EEMD is the largest, indicating the superiority of the method proposed in this paper.

5 Experiment Verification

For verify the effectiveness of the NLMS-EEMD noise reduction method proposed in this paper, the outer ring fault of the induction motor bearing was analyzed. The rolling bearing test bench is shown in Fig. 7. The test platform consists of an AC motor, a flexible coupling, a dynamic brake, a shaft generator, a piezoelectric accelerometer sensor and two supporting bearings. The accelerometer is installed in the vertical direction of the bearing seat of the drive end of the motor to collect the vibration signal of the supporting bearing. The fault type of the supporting bearing is set as outer ring fault. The main parameters and fault characteristic frequency of the faulty bearing are shown in Table 2.

Figure 8 shows the time waveform, frequency spectrum and Hilbert envelope spectrum of the fault signal of the supporting bearing outer ring. It can be seen from the figure that the waveform of the fault signal is very chaotic and has been completely overwhelmed by background noise and interference components. From the envelope spectrum of Fig. 8c, the fault characteristic frequency and its multiplier

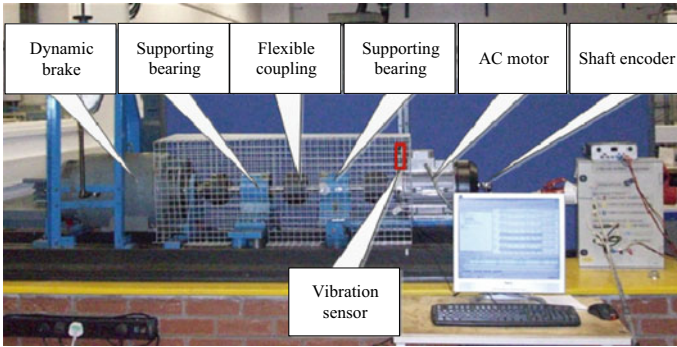


Fig. 7 Rolling element bearing test platform

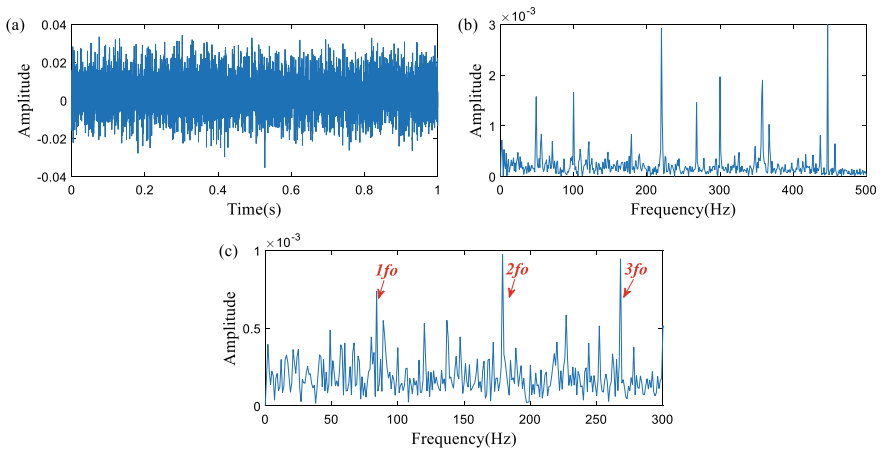
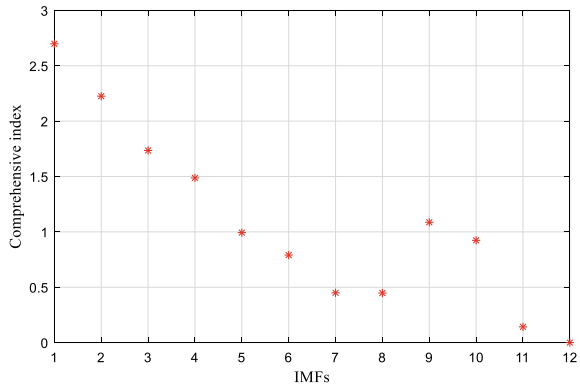


Fig. 8 The supporting bearing outer ring signal: a the time waveform of the signal; b the spectrum of the signal; c Hilbert envelop spectrum of the signal

of the bearing outer ring can be roughly found, but it contains a lot of noise and interference frequency, which can easily be masked and the fault characteristic frequency cannot be extracted.

To extract the fault characteristic frequency from the noise accurately, the proposed NLMS-EEMD denoising method is used to process the vibration signal of the supporting bearing outer ring fault. Firstly, NLMS adaptive filtering is used to filter the fault signal for the first time. Set the initial parameters according to the actual working environment, the step factor μ is set to 0.05, the order M of the filter is set to 2, and the constant γ is set to 10^{-4} . Then, the filtered signal is decomposed into 12 IMFs and one residual component by EEMD. Calculate the kurtosis, root mean square and sample entropy of each IMF and calculate the comprehensive index

Fig. 9 The comprehensive index of each IMF



according to Eq. (10). The comprehensive index calculation results of each IMFs are shown in Fig. 9.

It can be seen from Fig. 9 that the comprehensive index of IMF1 and IMF2 are larger than other components, indicating that these two components contain the complete information of the original signal, so IMF1 and IMF2 are selected for signal reconstruction. Finally, perform the Hilbert transform on the reconstructed signal and find its envelope. The analysis result is shown in Fig. 10a. From the figure, it can clearly see the characteristic frequency f_o of the supporting bearing outer ring fault and its multiplier. The method proposed in this paper can accurately diagnose bearing faults. To illustrate the effectiveness and advantages of the method, the analysis of NLMS noise reduction and EEMD noise reduction was carried out on the supporting bearing fault signal, and the envelope analysis of the results of the two methods was carried out. The results are shown in Fig. 10b and c, it is obvious that the envelope spectrum of these two noise reduction methods contains a lot of noise and harmonic interference, and the amplitude of the fault frequency is not prominent, and the fault characteristic frequency cannot be accurately identified.

To evaluate the noise reduction effect of the method proposed in this paper, the signal-to-noise ratio of the signal after noise reduction by three methods is calculated, and the results are shown in Table 3. The NLMS-EEMD noise reduction method has the highest signal-to-noise ratio, indicating that its noise reduction effect is the best.

6 Conclusion

This paper proposes a joint noise reduction method of NLMS and EEMD, which is used to extract the fault characteristics of rolling bearings in a strong noise environment. The main conclusions are as follows:

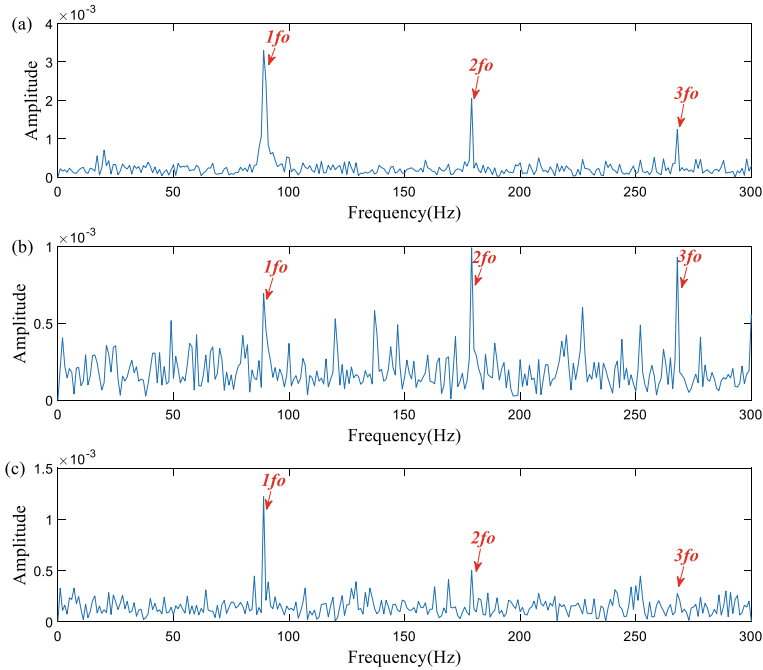


Fig. 10 Envelope analysis results of three methods

Table 3 The characteristic parameters of each IMF

Signal	SNR/dB
NLMS noise reduction signal	0.0331
EEMD noise reduction signal	1.1308
NLMS-EEMD noise reduction signal	4.5945

- (1) The Combined NLMS and EEMD denoising method can effectively reduce the background noise and provide a good condition for rolling bearing fault diagnosis.
- (2) Using the comprehensive index to select the appropriate IMFs can ensure the integrity of the information, so that the reconstructed signal contains the complete information of the original signal.
- (3) The simulation and experimental results show that the proposed method has certain advantages, and this method can effectively extract the fault features of rolling bearing in strong noise environment and carry out fault diagnosis.

References

1. Lv, Y., Yuan, Song, G.B.: Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing. *Mech. Syst. Signal Process.* **81**, 219–234 (2016)
2. Xue, X.M., Zhou, J.Z.: A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery. *ISA Trans.* **66**, 284–295 (2017)
3. Li, N., Zhou, R.: Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine. *Mech. Syst. Signal Process.* **28**, 608–621 (2012)
4. Li, J.M., Li, M.: Rolling bearing fault diagnosis based on time-delayed feedback monostable stochastic resonance and adaptive minimum entropy deconvolution. *J. Sound Vib.* **401**, 139–151 (2017)
5. Han, M.H., Pan, J.L.: A fault diagnosis method combined with LMD, sample entropy and energy ratio for roller bearings. *Measurement* **76**, 7–19 (2015)
6. Yang, Y., Pan, H.Y.: A roller bearing fault diagnosis method based on the improved ITD and RRVPMCD. *Measurement* **55**, 255–264 (2014)
7. Li, J.M., Yao, X.F.: Periodic impulses extraction based on improved adaptive VMD and sparse code shrinkage denoising and its application in rotating machinery fault diagnosis. *Mech. Syst. Signal Process.* **126**, 568–589 (2019)
8. Wang, Y.S., Feng, T.P.: An improved LMS algorithm for active sound-quality control of vehicle interior noise based on auditory masking effect. *Mech. Syst. Signal Process.* **108**, 292–303 (2018)
9. Bershad, N.J., Bermudez, J.C.M.: A switched variable step size NLMS adaptive filter. *Digital Signal Process.* **101**, 1–9 (2020)
10. Sofiane, F., Mohamed, D.: Hybrid PSO-NLMS (HPSO-NLMS) algorithm for blind speech quality enhancement in time domain. *Appl. Acoustics* **177**, 1–11 (2021)
11. Jiang, H.K., Li, C.L.: An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis. *Mechan. Syst. Signal Process.* **36**(2), 225–239 (2013)
12. Cheng, Y., Wang, Z.W.: An improved complementary ensemble empirical mode decomposition with adaptive noise and its application. *Inf. Sci.* **563**, 269–289 (2021)
13. Na, Q., Peng, J.: Fault diagnosis of high speed train bogie based on EEMD and permutation entropy. *J. Vibration Measurem. Diagnosis* **35**(5), 885–897 (2015)
14. Li, Y.B., Feng, K.: A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy. *J. Sound Vib.* **439**, 271–286 (2019)
15. Wang, S.J., Peng, Z.R.: Fault diagnosis of gearbox bearing based on multi-sensor signal processing. *Modular Mach. Tool Autom. Manuf. Technique* **11**, 5–10 (2020)

Rolling Bearing Fault Diagnosis Based on Weighted Variational Mode Decomposition and Cyclic Spectrum Slice Energy



Dongkai Li, Xiaoang Liu, Yue You, Dong Zhen, Wei Hu, Kuihua Lu, and Fengshou Gu

Abstract As the main parts of rotating machinery, rolling bearing is prone to failure due to its harsh working environment. Aiming at the problem that the early fault features of a rolling bearing are easily submerged by noise and difficult to extract, a fault diagnosis method based on weighted variational mode decomposition (WVMD) and cyclic spectrum slice energy (CSSE) is proposed. Firstly, the signal is decomposed into intrinsic mode functions (IMFs) by VMD and the sparsity is used to measure the amount of information contained in each IMF, and all IMFs are weighted and reconstructed to suppress the noise interference components in the signal. Secondly, the advantage of the CSSE which can accurately mediate the fault information is used to analyze the reconstructed signal, and then the fault characteristic frequency of the reconstructed signal is extracted. Finally, the bearing simulation signal and outer ring fault signal are used to verify that the proposed diagnosis method can effectively extract the early fault features of rolling bearing.

Keywords Variational mode decomposition · Weight coefficient · Cyclic spectrum slice energy · Rolling bearing · Fault diagnosis

D. Li · X. Liu · D. Zhen (✉)

School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China

e-mail: d.zhen@hud.ac.uk

D. Li

e-mail: li_d_k@163.com

F. Gu

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

W. Hu · K. Lu

WORLDTECH Transmission Technology Ltd, Tianjin 300401, China

Y. You

China Machinery Engineering Corporation, Beijing 100073, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_52

1 Introduction

Rolling bearing is widely used in kinds of large rotating machinery and electrical equipment. The running state of the rolling bearing directly affects the normal operation of the machine [1]. Monitoring its running state, especially early fault diagnosis, is of great significance for the operation and maintenance of mechanical equipment. Therefore, the early fault diagnosis of rolling bearing has become a research hotspot.

The early fault signal of rolling bearing is nonlinear and nonstationary. The fault signal is weak and easy to be submerged in the background noise [2]. These factors make it difficult to extract the early fault features of rolling bearing. In recent years, many scholars use time–frequency analysis methods to deal with fault signals. The main time–frequency analysis methods include short-time Fourier transform (STFT) [3], wavelet transform (WT) [4], Wigner–Ville distribution (WVD), empirical mode decomposition (EMD) [5] and so on. These methods have been successfully applied to fault signal analysis, but they also have limitations.

Variational mode decomposition (VMD) is a new adaptive time–frequency analysis method proposed by Dragomiretskiy [6]. At present, most VMD algorithms use some indexes to select the optimal mode for reconstruction, so as to reduce the influence of Gaussian noise. For example, Ali et al. [7] proposed the envelope spectrum weighted kurtosis index to select the sensitive mode, Xu et al. [8] used the average kurtosis to select some IMF whose kurtosis value is greater than the average kurtosis value to reconstruct the new signal. Yan et al. [9] selected the IMFs containing the main fault feature information through the single failure feature amplitude energy ratio, and used it for subsequent analysis. However, the information contained in other modal components is ignored when selecting the optimal mode reconstruction, which makes the reconstructed signal lack some effective information contained in the original signal. Chen et al. [10] used sparsity as an index to select effective components of EEMD, and explained the relationship between signal sparsity and energy. In view of the advantage of sparsity, this paper constructs a weight coefficient by sparsity to measure the amount of useful information contained in different modal components and then reconstructs all modal components by weighting.

Because of the periodicity of the motion, the statistics of the early fault signals of rolling bearings are periodic time-varying and show the stability of the cycle. The cyclic spectrum (CS) analysis is a powerful tool to deal with the cyclic stationary signal [11, 12]. And the cyclic spectrum slice energy (CSSE) is a unary function of cyclic frequency, which is more conducive to display fault characteristics, Wang et al. [13] used the CSSE to diagnose the bearing fault, and pointed out that it can make up the defect of envelope demodulation analysis. So CSSE is used to reveal the fault information contained in WVMD filtered signal in this study. To sum up, this paper proposes a rolling bearing fault diagnosis method based on the combination of weighted variational mode decomposition and cyclic spectrum slice energy. The rest of the article is organized as follows: Sect. 2 investigates basic theories, Sect. 3 introduces the process of the proposed method, Sects. 4 and 5 verify the proposed method by simulation and experiment and the conclusions are drawn in Sect. 6.

2 Basic Theories

2.1 Variational Mode Decomposition

VMD can adaptively decompose a complex signal into a series of IMFs. The center frequency and bandwidth of each IMF are determined by iteratively searching the optimal solution of the variational model. In the frequency domain, the decomposed IMFs are sparse, which can separate the intrinsic mode functions effectively. The basic idea of VMD algorithm is to construct and solve the variational problem to decompose the complex signal into k IMFs [9].

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t [(\delta(t) + j/\pi t) * u_k(t)] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_k u_k = f \end{cases} \tag{1}$$

In order to solve the above variational problem, VMD transforms the constrained variational problem into an unconstrained variational problem by introducing the quadratic penalty factor and Lagrange operator. The augmented Lagrange function is:

$$\begin{aligned} L(\{u_k(t)\}, \{\omega_k(t)\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t [(\delta(t) + j/\pi t) u_k(t)] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda, f(t) - \sum_{k=1}^K u_k(t) \right\rangle \end{aligned} \tag{2}$$

where α is the penalty factor, λ is Lagrange operator; $\langle \cdot \rangle$ is convolution. The alternating direction method of multipliers (ADMM) algorithm is used to solve the saddle point of the augmented Lagrange function.

2.2 Wighted Reconstruction of IMFs

Most of the early faults of a rolling bearing are localized pitting defects of the outer ring, inner ring and rolling element [14]. In the fault signal of rolling bearing, the local pitting often appears as a periodic pulse, which makes the fault signal sparse. Sparsity is a statistic that characterizes the sparsity of time-domain signals, For signal, $x(n)$ the expression formula of sparsity [10] is as Eq. (3). this paper uses sparsity to construct the weight coefficients of each mode.

$$S = \sqrt{\frac{1}{N} \sum_{n=1}^N x_i(n)^2} / \frac{1}{N} \sum_{n=1}^N |x_i(n)| \tag{3}$$

When there are fewer Gaussian noise components and more impact components in the signal, the energy of the signal is concentrated in the pulse, and the amplitude of the signal has obvious multiple bulges. The signal shows strong sparsity, and the sparsity of the signal is larger. On the contrary, when the signal contains more Gaussian noise components, the energy of the signal is dispersed into the interference components, the amplitude distribution of the signal is more uniform, the sparsity of the signal is weaker, and its sparsity is smaller. Therefore, this paper uses sparsity to construct the weight coefficients of each mode. The signal is decomposed by VMD to get k IMFs, The weight coefficient of IMF is defined as Eq. (4).

$$c(k) = S(k) / \sum_k S(k) \tag{4}$$

where $S(k)$ represents the sparsity of the k th IMF. The expression of weighted reconstructed signal is:

$$\hat{x} = \sum_k c(k)u_k \tag{5}$$

2.3 Spectral Correlation Density Slice Energy

As a special case of non-stationary signal, the statistical characteristics of cyclostationary signals change periodically. According to the order of periodic statistical characteristics, cyclostationary signals can be divided into first-order cyclostationary signals, second-order cyclostationary signals and high-order cyclostationary signals. The second-order cyclostationary signal is a non-stationary signal whose autocorrelation function changes periodically. The early fault signals of rolling bearings show amplitude modulation characteristics, its autocorrelation function is periodically time-varying and belongs to second-order cyclostationary signal. For the second-order cyclostationary signal, the Eq. (6) of time-varying autocorrelation function is obtained by the statistical average of the signal's time-delay quadratic transformation.

$$R_x(t, \tau) = E \left\{ x^* \left(t - \frac{\tau}{2} \right) x \left(t + \frac{\tau}{2} \right) \right\} \tag{6}$$

where E is mathematical expectation; (*) represents a conjugate operation. Since the time-varying autocorrelation function $R_x(t, \tau)$ is a periodic function, it can be

expanded into Eq. (7) by Fourier series.

$$R_x(t, \tau) = \sum_{\alpha} R_x^{\alpha}(\tau) e^{j2\pi\alpha t} \quad (7)$$

where $\alpha = n/T_0$ is the cycle frequency. The cyclic spectral density (CSD) $R_x^{\alpha}(\tau)$ can be obtained by Fourier transform (FT) of cyclic autocorrelation function (CAF) as follows:

$$S_x^{\alpha}(f) = \int_{-\infty}^{\infty} R_x^{\alpha}(\tau) e^{-j2\pi f\tau} d\tau \quad (8)$$

The CSD is a double spectrum diagram composed of cyclic frequency and spectrum frequency. In order to analyze the fault types of rolling bearings by using the cyclic frequency, the CSD is sliced along the direction perpendicular to the cycle frequency axis at each cycle frequency. The cyclic spectrum slice energy (CSSE) is defined as:

$$E(\alpha) = \int_{-\infty}^{\infty} |S_x^{\alpha}(f)|^2 df \quad (9)$$

3 Algorithm Flow of Weighted Variational Mode Decomposition and Cyclic Spectrum Slice Energy

In this paper, a rolling bearing fault diagnosis method based on weighted variational mode decomposition and cyclic spectrum slice energy is proposed. The specific diagnosis process is as follows: firstly, the number of IMFs in the variational mode decomposition is determined, and the original signal is decomposed to obtain IMFs; Then, the sparsity of each modal component is calculated, and IMFs are weighted by the sparsity; Finally, the cyclic spectrum density of the reconstructed signal is calculated, and the cyclic spectrum slice energy is obtained. The cyclic spectrum slice energy is used to analyze the reconstructed signal and extract the characteristic frequency of the fault signal. The diagnosis flow chart of the proposed algorithm is shown in Fig. 1.

4 Simulation Verification

When the rolling bearing has a local pitting fault, there will be an impact on other parts in contact with the local pitting. With the rotation of the rolling bearing, the periodic pulse will be generated in the bearing fault signal, and there will be small

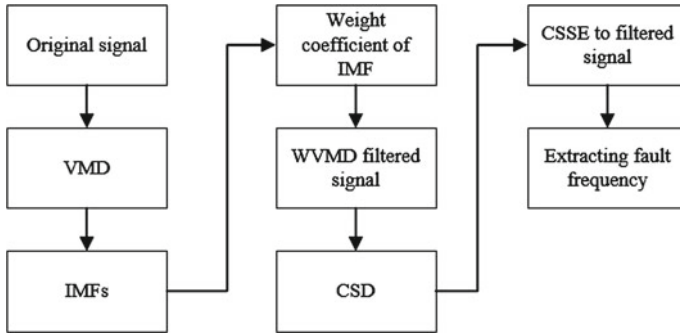


Fig. 1 The flowchart of the WVMD-CSSE method

sliding between the rolling element and the raceway. Therefore, the fault simulation signal of rolling bearing is as follows:

$$\begin{aligned}
 x(t) &= \sum_i A_0 h(t - iT - \tau_i) + n(t) \\
 h(t) &= e^{-\beta t} \sin(2\pi f_n t)
 \end{aligned}
 \tag{10}$$

where A_0 represents the amplitude and is set to 1, T denotes the interval of fault pulse, set to 1/120 s, τ_i is the random slip caused by the i -th impact, it's related to T , β represents the attenuation coefficient, equal to 800, f_s stands the sampling period, set to 10 000 Hz, t is the sampling time and is set to 1 s, f_o denotes the bearing failure frequency, equal to 120 Hz, f_n stands the resonant frequency, set to 1 000 Hz. Considering that the rolling bearing works in harsh environment, Gaussian noise is added to the simulation signal and $SNR = -7$ dB.

The time domain waveform and frequency spectrum of the simulation signals are shown in Fig. 2b and c. The fault signal is submerged by noise, so it is difficult to find fault feature information in Fig. 2b and c can not effectively reflect the fault frequency. In order to extract the fault characteristic frequency, WVMD-CSSE is applied to the simulation signal. Firstly, the simulation signal is decomposed into seven IMFs by using VMD, and the number of IMFs is determined by center frequency observation method. In this method, the number of IMFs is preset as 2, 3, ..., 8, and decomposed respectively to observe the center frequency of adjacent IMF. When the number of modes is 8, the center frequency distance of adjacent modes is too small, resulting in over decomposition; When the number of modes is 6, under decomposition occurs, so the number of modes is 7. The result of VMD is shown in Fig. 3. Considering the different amounts of useful information contained in each IMF and the suppression of noise components, the weighted IMFs are formed by calculating the modal weight coefficients composed of sparsity, and the reconstructed signal is obtained by adding all the weighted IMFs. The weight coefficient of IMF and the spectrum of weighted reconstruction signal are shown in Figs. 4 and 5. It can be found that the

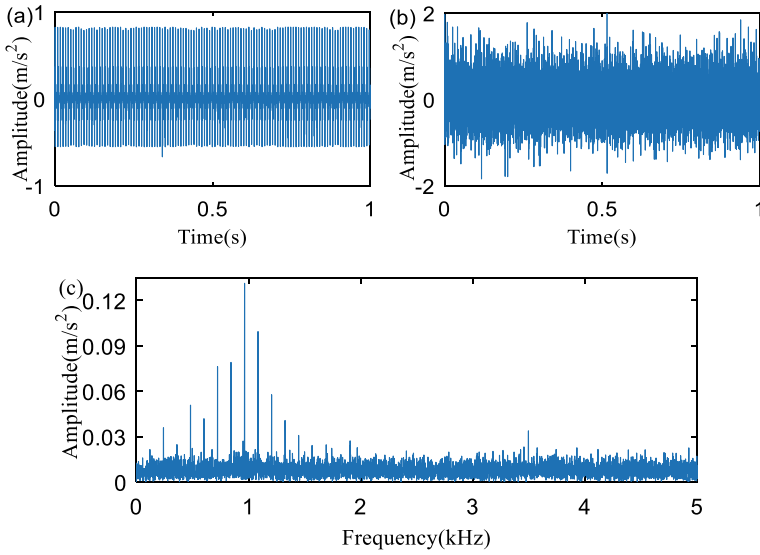


Fig. 2 Rolling bearing simulation signal. a Pure signal. b Noisy signal. c Spectrum

Fig. 3 Decomposed modes

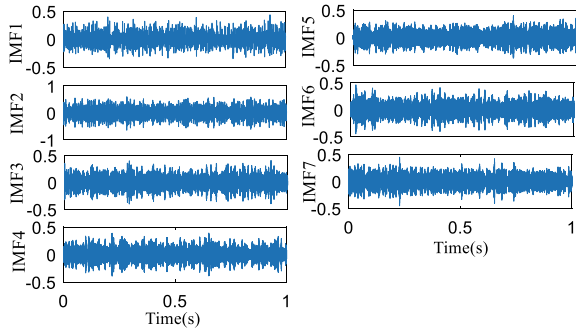


Fig. 4 IMF weighting coefficient

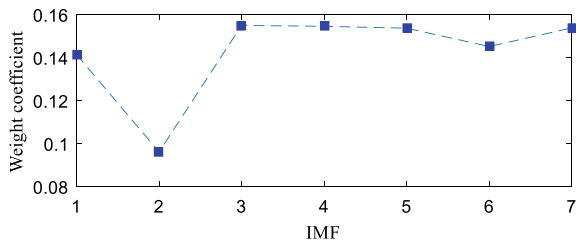


Fig. 5 Weighted reconstruction signal spectrum

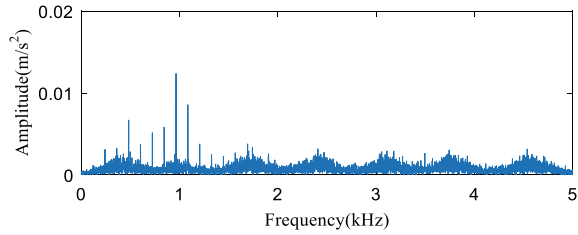
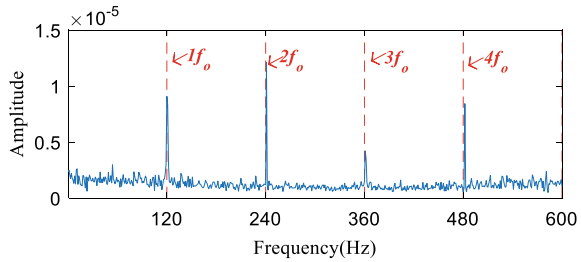


Fig. 6 Applying WVMD-CSSE to simulation signal



noise component of the original signal is obviously suppressed in Fig. 5, however, the fault characteristics are still not obvious. Then CSSE is applied to analyze the reconstructed signal and identify the fault components in the signal. The CSSE of the reconstructed signal is shown in Fig. 6, which can clearly display the fault characteristic frequency f_o and its frequency doubling components ($2f_o$, $3f_o$ and $4f_o$).

5 Experimental Analysis

In order to further prove that the WVMD-CSSE algorithm can effectively carry out the early failure of rolling bearing, experimental research of rolling bearing is carried out. The rolling bearing test bench mainly includes an AC motor, dynamometer, two rolling bearings of 6206ZZ, flexible coupling, vibration sensor and shaft encode. The bearing test bench is shown in Fig. 7. The vibration sensor is installed on the bearing seat of the drive end of the AC motor, and the speed of the motor is 1500 r/min. The length of bearing data of outer ring fault collected in rolling bearing test is 1,920,000, and the sampling frequency is 96 kHz. f_c is the theoretical characteristic frequency of outer ring fault, $f_o = 89.33$ Hz. The specific parameters of bearing in the test are shown in Table 1.

The time domain waveform and frequency spectrum of the outer ring fault rolling bearing obtained from the test are shown in Fig. 8. Affected by the background noise, the periodic fault impact of the signal is submerged by the noise, and it is difficult to extract the fault features from the frequency spectrum. Applying the WVMD-CSSE diagnosis algorithm proposed in this paper, firstly, VMD is performed on

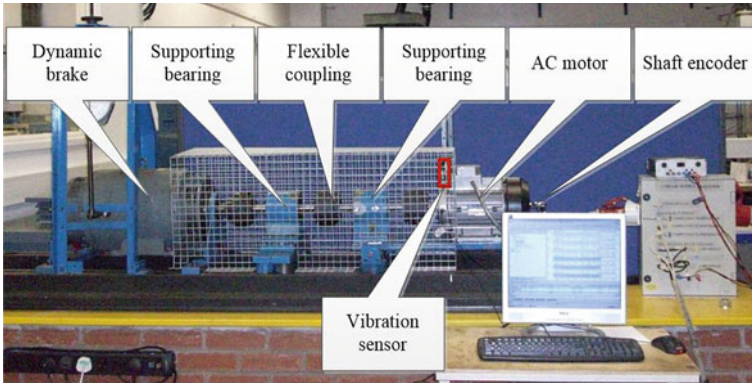
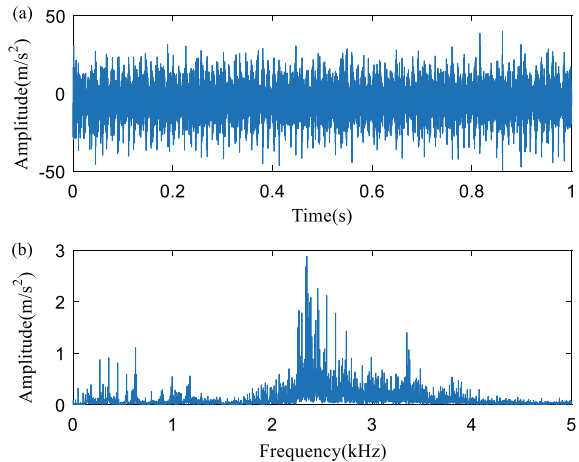


Fig. 7 Rolling bearing test bench

Table 1 Structural parameters of test bearing

Bearing model	Rolling body diameter (mm)	Pitch diameter (mm)	Ball number	Contact angle
6206ZZ	9.53	46.4	9	0°

Fig. 8 Outer ring fault rolling bearing test signal. **a** time domain waveform. **b** frequency spectrum



the experimental signal, and the number of IMFs is determined as $k = 5$ by center frequency observation method. The modes of VMD are shown in Fig. 9. Then, the sparsity and weight coefficient of IMF are calculated, and all IMFs are weighted and reconstructed. The results of weight coefficient and reconstructed signal are shown in Figs. 10 and 11, compared with Figs. 8b and 11, it can be seen that the noise interference on the reconstructed signal spectrum is obviously reduced in Fig. 11.

Fig. 9 Applying WVMD-CSSE to test signal

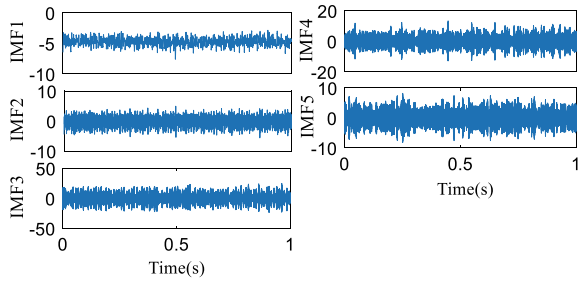


Fig. 10 IMF weighting coefficient

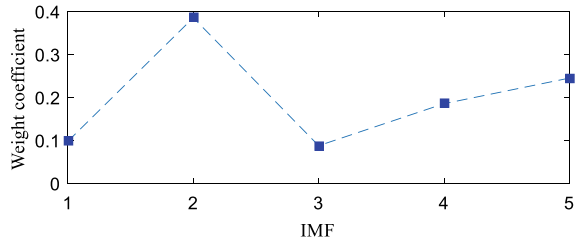
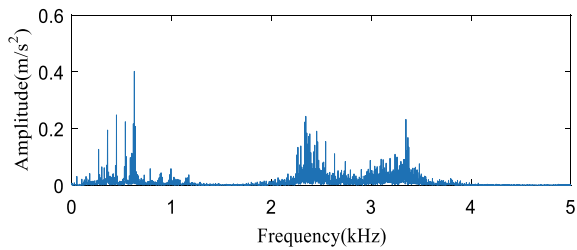
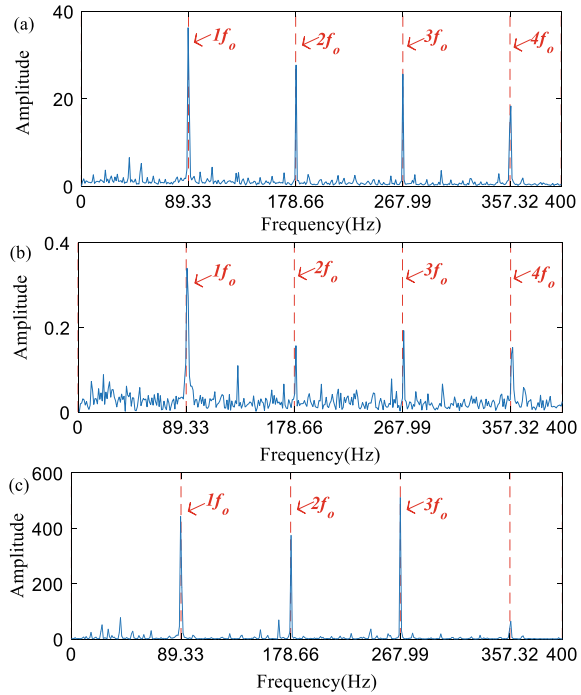


Fig. 11 Weighted reconstructed signal spectrum



Finally, CSSE is applied to demodulate the reconstructed signal, as shown in Fig. 12a. In order to compare the advantages of CSSE in extracting bearing fault characteristic frequency, envelope demodulation is performed on the reconstructed signal, as shown in Fig. 12b. Not only the fault characteristic frequency of the bearing is proposed successfully, but also the frequency doubling component has a very obvious peak value in Fig. 12a, but there are many interference frequency components in Fig. 12b. In order to further illustrate that using the weight coefficient composed of sparsity to reconstruct all IMFs can retain complete fault information and suppress noise, VMD-CSSE is applied to the original signal, as shown in Fig. 12c, it can be found that WVMD-CSSE has better signal processing effect than VMD-CSSE.

Fig. 12 Analysis results. **a** WVMD-CSSE. **b** Envelope demodulation. **c** VMD-CSSE



6 Conclusion

Aiming at the problem that the early fault features of a rolling bearing are easily disturbed by noise and difficult to extract, the WVMD-CSSE method is proposed in this paper, the main conclusions are as follows:

- (1) The weight coefficient constructed by sparsity can measure the fault information contained in each mode, suppress the noise interference component of the original signal and improve the SNR.
- (2) CSSE is used to process WVMD filtering signal, so as to remove periodic fault impulse contained in filtering signal and identify characteristic frequency.
- (3) CSSE has a stronger demodulation ability than envelope analysis and is easier to extract bearing fault information. By comparing WVMD-CSSE with VMD-CSSE, it is proved that WVMD-CSSE can get more accurate and clear characteristic frequency.

References

1. Wei, L., Jian, C., Jiazhu, L., Kang, X.: Derivative and enhanced discrete analytic wavelet algorithm for rolling bearing fault diagnosis. *Microprocessors and Microsystems* **82**, (2021)
2. An, X., Jiang, D., Chen, J., Liu, C.: Application of the intrinsic time-scale decomposition method to fault diagnosis of wind turbine bearing. *J. Vibration Control* **18**(2), (2012)
3. Kun, Z., Chaoyong, M., Yonggang, X., Peng, C., Jianxi, D.: Feature extraction method based on adaptive and concise empirical wavelet transform and its applications in bearing fault diagnosis. *Measurement* **172**, (2021)
4. Li, Z., Chen, J., Zi, Y., Pan, J.: Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive. *Mechan. Syst. Signal Process.* **85**, (2017)
5. Mantang, H., Guofeng, W., Kaile, M., Zenghuan, C., Shuai, Y.: Bearing performance degradation assessment based on optimized EWT and CNN. *Measurement* **172**, (2020)
6. Dragomiretskiy, K., Zosso, D.: Variational mode decomposition. *IEEE Trans Signal Process* **62**, 531–544 (2014)
7. Ali, D., Reza, H., Mohammad, E.M. et al.: Incipient fault diagnosis of bearings based on parameter-optimized VMD and envelope spectrum weighted kurtosis index with a new sensitivity assessment threshold. *ISA Trans* **114**, (2021)
8. Xu, B., Zhou, F., Li, H., Yan, B., Liu, Y.: Early fault feature extraction of bearings based on Teager energy operator and optimal VMD. *ISA Trans.* **86**, (2018)
9. Yan, X., Jia, M., Xiang, L.: Compound fault diagnosis of rotating machinery based on OVMD and a 1.5-dimension envelope spectrum. *Measur. Sci. Technol.* **27** (2016)
10. Liang, C., Xing, L., Xunbo, L., Zuoying, H.: Signal extraction using ensemble empirical mode decomposition and sparsity in pipeline magnetic flux leakage nondestructive evaluation. *Rev. Scientif. Instrum.* **80**(2), (2009)
11. Antoni, J.: Cyclic spectral analysis of rolling-element bearing signals: facts and fictions. *J. Sound Vibration* **304**(3), (2007)
12. Yang, M.: Rolling bearing fault diagnosis based on spectral correlation density slice analysis and support vector machine. *J. Vibration Shock* **29**(1), 196–199+246–247 (2010)
13. Wang, H., Chen, J., Dong, G.: Fault diagnosis of rolling bearing based on spectral correlation density combined slice energy. *J. Vibration Shock* **34**(03), 114–117 (2015)
14. Zhang, Y.Y., Shi, L.S.: Establishment and application of mathematical model for local fault of rolling bearing. *J. Vibration and Shock* **29**(04), 73–76 + 232 (2010)

Nonlinear Dynamic Analysis of a Planetary Gear System with Sun Gear Fault



Yinghui Liu, Zhanqun Shi, Dong Zhen, Xiaoang Liu, Wei Hu, and Fengshou Gu

Abstract Planetary gear systems often work under severe working conditions, causing gear failures to occur frequently. When the gear fails, the dynamic characteristics of the system will be altered with the excitation of nonlinear parameters. To study the dynamic response of the planetary gear system with gear fault, a nonlinear dynamic model for both health system and faulty system containing the backlash, time-varying mesh stiffness and manufacturing error is put forward. Then, the backlash and rotation frequency are taken as the excitation parameters to study the nonlinear characteristics of the faulty and healthy system through the global bifurcation diagram. At the same time, the local characteristics of the two systems are analyzed via the Poincaré maps and phase diagrams. The analysis results show that the faulty system has a more complex movement as the excitation parameters variation. This research can provide a reference for the dynamic design of planetary gear systems.

Keywords Planetary gear system · Tooth backlash · Bifurcation · Phase diagram · Poincaré map

Y. Liu · Z. Shi · D. Zhen (✉) · X. Liu
School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China
e-mail: d.zhen@hud.ac.uk

Y. Liu
e-mail: victoryhui@163.com

F. Gu
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

W. Hu
WORLDTECH Transmission Technology Ltd, Tianjin 300401, China

1 Introduction

The planetary gear system has many advantages, such as compact structure, large transmission ratio, etc. Hence, they have various applications in mechanical transmission equipment [1]. For the convenience of lubrication, tooth backlash is necessary for the gear transmission system. However, as a strong nonlinear factor, the backlash will affect the dynamic characteristics of the system, especially the faulty system. Therefore, for benefitting the fault diagnosis and dynamic design of gear systems, there needs to be a good master of the dynamic characteristics.

For the modeling and analysis of the gear system, a large number of scholars have conducted research. In 1994, Kahraman [2] put up with a dynamic model of planetary gear system. Later, the modes of the system are analyzed [3]. And the influence of the installation error of the carrier on the dynamic characteristics of the planetary gear system are analyzed [4]. Mo [5] proposed a calculation model of load distribution coefficient with considering various errors. Later, Li [6] analyzed the effects of static errors on transmission error and load sharing coefficient. By introducing the geometric eccentricity error, Zhao [7] analyzed the variation law of the contact state of the gear teeth. Then, Park [8] analyzed the influence of the gear initial position within backlash on the dynamic characteristics. Wang [9] established a three-degree-of-freedom torsional vibration model, revealing the nonlinear phenomena of the system. Considering the factors of backlash, mesh stiffness and transmission error, Wang [10] proposed a bending-torsion coupling dynamic model and analyzed the periodic and chaotic characteristics of the system. Yang [11] established a nonlinear dynamic model to analyze the frequency response characteristics of the spur gear transmission system. And then the influence of bearing clearance [12] and gear surface modification [13] on the dynamic characteristics of the planetary gear system is analyzed. Huang [14] established the torsional dynamics model considering the fractal tooth backlash and analyzed the chaos and bifurcation of the spur gear system with the rotation speed and initial backlash as variables. Combining bearing clearance and tooth backlash, Wang [15] analyzed the vibration response of the gear system with gear crack failure.

At present, there are limited researches on the chaos and bifurcation characteristics of the faulty planetary gear systems. Therefore, this study proposed a new pure-torsional dynamic model of planetary gear system considering time-varying mesh stiffness, backlash and manufacturing error. Meanwhile, the backlash and rotation frequency are selected as the excitation parameters to reveal the response characteristics of the system. The other sections are organized as follows. In Sect. 2, the planetary gear dynamics model is proposed. Section 3 analyzes the dynamic characteristics of the system. At last, Sect. 4 gives some conclusions.

2 Dynamic Model of the Planetary Gear System

In this model, the planetary gear system includes a sun gear, a ring gear, a planet carrier and three planet gears, as shown in Fig. 1. Among them, the carrier is selected as the input element, the sun gear as the output element, and the ring gear is fixed. All components are regarded as rigid and have only rotational degrees of freedom.

2.1 Parameters Definition

In the gear system, the meshing frequency is related to the rotation speed and the number of gear teeth, which can be expressed as [16]:

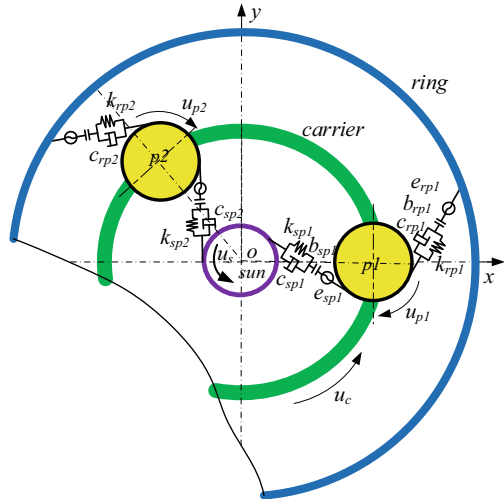
$$\omega_m = \omega_c z_r \tag{1}$$

where ω_c is the angular frequency of the carrier, and z_r is the tooth number of the ring gear.

The time-varying mesh stiffness k_{sp} and k_{rp} of the gear pairs can be simplified into rectangular waveform, which can be expanded into Fourier series with meshing frequency ω_m as the fundamental frequency. Taking the fundamental frequency part, the mesh stiffness can be expressed as [16]:

$$k_{sp} = k_{msp} + k_{asp} \cos(\omega_m t + \phi_{spn}) \tag{2a}$$

Fig. 1 The dynamic model of planetary gear system



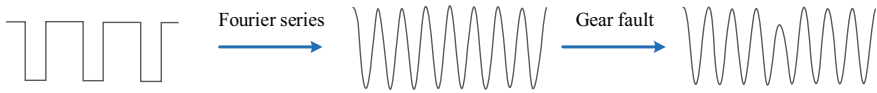


Fig. 2 Time varying mesh stiffness

$$k_{rp} = k_{mrp} + k_{arp} \cos(\omega_m t + \phi_{rpn}) \tag{2b}$$

herein k_{msp} and k_{mrp} are the average mesh stiffness of external-external mesh and external-internal mesh respectively; k_{asp} and k_{arp} are the variation amplitude of the mesh stiffness. ϕ_{spn} and ϕ_{rpn} are the initial phases.

In a planetary gear system, the sun gear rotates at a higher speed. Therefore, sun gear failures often occur. When the gear failure occurs, the mesh stiffness will attenuate, just like Fig. 2, which can be expressed as:

$$k_{sp} = \begin{cases} k_{mfsp} + k_{afsp} \cos(\omega_m t + \phi_{fspn}), & 2\pi m - \varphi_0 < \text{mod}(\omega_m t, 2\pi z_p) < 2\pi m \\ k_{msp} + k_{asp} \cos(\omega_m t + \phi_{spn}), & \text{others} \end{cases} \tag{3}$$

where k_{mfsp} represents the average mesh stiffness of the fault tooth pair, k_{afsp} is the variation amplitude of the mesh stiffness, and m is the sequence of the fault tooth.

For the simplification of the model, the gear teeth are assumed to be the same. At this moment, the static transmission error is a period, which can be written in the form of a triangular series with the meshing frequency as the fundamental frequency, as shown below [16]:

$$e_{sp} = e_{asp} \sin(\omega_m t + \psi_{spn}) \tag{4a}$$

$$e_{rp} = e_{rsp} \sin(\omega_m t + \psi_{rpn}) \tag{4b}$$

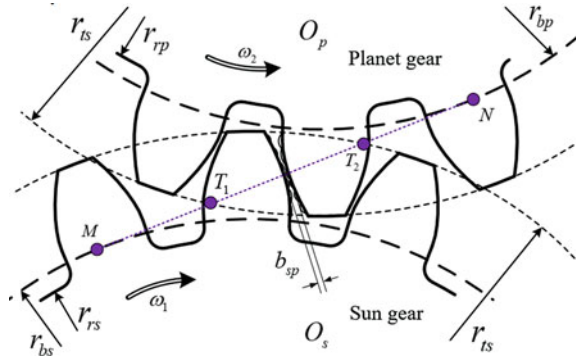
where e_{asp} and e_{rsp} are the amplitudes of the external and internal mesh errors, respectively. ψ_{spn} and ψ_{rpn} are the initial phases.

The meshing damping between the internal and external meshes can be expressed as [16]:

$$c_{sp} = 2\xi_{sp} \sqrt{\frac{k_{msp}}{\frac{1}{m_s} + \frac{1}{m_p}}} \tag{5a}$$

$$c_{rp} = 2\xi_{rp} \sqrt{\frac{k_{mrp}}{\frac{1}{m_r} + \frac{1}{m_p}}} \tag{5b}$$

Fig. 3 Diagram of sun-planet mesh



wherein ξ_{sp} and ξ_{rp} are the damping ratios of external and internal mesh respectively. I_j ($j = s, ce, r, p$) is the moment of inertia and m_j ($j = s, c, r, p$) is the equivalent mass that can be calculated by follows [17]:

$$I_{ce} = I_c + Nm_p r_c^2, m_s = \frac{I_s}{r_s^2}, m_c = \frac{I_{ce}}{r_c^2}, m_p = \frac{I_p}{r_p^2} \tag{5b}$$

As shown in Fig. 3, the backlash can be expressed as a piecewise linear function [17]:

$$f(\delta_{sp}) = \begin{cases} \delta_{sp} - b_{sp}, & \delta_{sp} > b_{sp} \\ 0, & |\delta_{sp}| \leq b_{sp} \\ \delta_{sp} + b_{sp}, & \delta_{sp} < -b_{sp} \end{cases} \tag{7a}$$

$$f(\delta_{rp}) = \begin{cases} \delta_{rp} - b_{rp}, & \delta_{rp} > b_{rp} \\ 0, & |\delta_{rp}| \leq b_{rp} \\ \delta_{rp} + b_{rp}, & \delta_{rp} < -b_{rp} \end{cases} \tag{7b}$$

where δ_{sp} and δ_{rp} are dynamic transmission errors, and b_{sp} and b_{rp} are the initial backlash.

2.2 Dynamic Differential Equations

The dynamic differential equations of torsional vibration can be listed out according to the Lagrangian equation:

$$\begin{cases} I_s \ddot{\theta}_s + \sum_{n=1}^N r_s (c_{sp} \dot{\delta}_{sp} + k_{sp} f(\delta_{sp})) = T_{out} \\ I_p \ddot{\theta}_p - r_p (c_{sp} \dot{\delta}_{sp} + k_{sp} f(\delta_{sp})) + r_p (c_{rp} \dot{\delta}_{rp} + k_{rp} f(\delta_{rp})) = 0 \\ I_c \ddot{\theta}_c - \sum_{n=1}^N r_c (c_{sp} \dot{\delta}_{sp} + k_{sp} f(\delta_{sp})) - \sum_{n=1}^N r_c (c_{rp} \dot{\delta}_{rp} + k_{rp} f(\delta_{rp})) = -T_{in} \end{cases} \quad (8)$$

where T_{out} and T_{in} represent the output torque and the input torque respectively.

During operation, considering the manufacturing error and the elastic deformation of the gear teeth, the dynamic transmission error can be expressed as [17]:

$$\delta_{sp} = r_s \theta_s - r_p \theta_p - r_c \theta_c - e_{sp} \quad (9a)$$

$$\delta_{rp} = r_p \theta_p - r_r \theta_r - r_c \theta_c - e_{rp} \quad (9b)$$

To facilitate the solution, a dimensionless treatment is carried out. By introducing the time scale ω_n and displacement scale b_c , the dimensionless parameters can be obtained. The dimensionless time displacement, velocity and acceleration can be expressed as:

$$\begin{aligned} \tau = \omega_n * t, \quad \bar{\delta}_{sp} = \frac{\delta_{sp}}{b_c}, \quad \dot{\delta}_{sp} = \frac{\delta_{sp}}{b_c \omega_n}, \quad \ddot{\delta}_{sp} = \frac{\ddot{\delta}_{sp}}{b_c \omega_n^2}, \quad \bar{\delta}_{rp} = \frac{\delta_{rp}}{b_c} \\ \dot{\delta}_{rp} = \frac{\delta_{rp}}{b_c \omega_n}, \quad \ddot{\delta}_{rp} = \frac{\ddot{\delta}_{rp}}{b_c \omega_n^2}, \quad \omega_n = \sqrt{\frac{k_{mrp}}{\frac{1}{m_s} + \frac{1}{m_c}}} \end{aligned} \quad (10)$$

$$\bar{k}_{sp} = \frac{k_{msp}}{k_{mrp}} + \frac{k_{asp}}{k_{mrp}} \cos\left(\frac{\omega_m}{\omega_n} t + \phi_{spn}\right) \quad (11a)$$

$$\bar{k}_{rp} = 1 + \frac{k_{arp}}{k_{mrp}} \cos\left(\frac{\omega_m}{\omega_n} t + \phi_{rpm}\right) \quad (11b)$$

$$\bar{e}_{sp} = \frac{e_{asp}}{b_c} \sin\left(\frac{\omega_m}{\omega_n} t + \psi_{spn}\right) \quad (12a)$$

$$\bar{e}_{rp} = \frac{e_{rsp}}{b_c} \sin\left(\frac{\omega_m}{\omega_n} t + \psi_{rpm}\right) \quad (12b)$$

$$f(\bar{\delta}_{sp}) = \begin{cases} \bar{\delta}_{sp} - b_{sp}/b_c, & \bar{\delta}_{sp} > b_{sp}/b_c \\ 0, & |\bar{\delta}_{sp}| \leq b_{sp}/b_c \\ \bar{\delta}_{sp} + b_{sp}/b_c, & \bar{\delta}_{sp} < -b_{sp}/b_c \end{cases} \quad (13a)$$

$$f(\bar{\delta}_{rp}) = \begin{cases} \bar{\delta}_{rp} - b_{rp}/b_c, & \bar{\delta}_{rp} > b_{rp}/b_c \\ 0, & |\bar{\delta}_{rp}| \leq b_{rp}/b_c \\ \bar{\delta}_{rp} + b_{rp}/b_c, & \bar{\delta}_{rp} < -b_{rp}/b_c \end{cases} \quad (13b)$$

Then, substituting the dimensionless quantity into the dynamic Eq. (8) and simplifying it can be obtained:

$$\left\{ \begin{aligned}
 & \ddot{\delta}_{sp} + \left(\frac{1}{m_s} + \frac{1}{m_c}\right) \frac{1}{\omega_n} \sum_{n=1}^N c_{sp} \dot{\delta}_{sp} + \left(\frac{1}{m_s} + \frac{1}{m_c}\right) \frac{1}{\omega_n^2} \sum_{n=1}^N k_{sp} f(\bar{\delta}_{sp}) \\
 & + \frac{1}{m_p} \frac{1}{\omega_n^2} k_{spn} f(\bar{\delta}_{sp}) + \frac{1}{m_p} \frac{1}{\omega_n} c_{sp} \dot{\delta}_{sp} - \frac{1}{m_p} \frac{1}{\omega_n} c_{rp} \dot{\delta}_{rp} \\
 & - \frac{1}{m_p} \frac{1}{\omega_n^2} k_{rp} f(\bar{\delta}_{rp}) + \frac{1}{m_c} \frac{1}{\omega_n} \sum_{n=1}^N c_{rp} \dot{\delta}_{rp} \\
 & + \frac{1}{m_c} \frac{1}{\omega_n^2} \sum_{n=1}^N k_{rp} f(\bar{\delta}_{rp}) = \frac{T_{out} r_c}{I_s b_c \omega_n^2} + \frac{T_{in} r_c}{I_c b_c \omega_n^2} + \frac{e_{msp}}{b_c} \left(\frac{\omega_m}{\omega_n}\right)^2 \sin\left(\frac{\omega_m}{\omega_n} t + \psi_{spn}\right) \\
 & \ddot{\delta}_{rp} - \frac{1}{m_p} \frac{1}{\omega_n} c_{sp} \dot{\delta}_{sp} - \frac{1}{m_p} \frac{1}{\omega_n^2} k_{sp} f(\bar{\delta}_{sp}) + \frac{1}{m_p} \frac{1}{\omega_n} c_{rp} \dot{\delta}_{rp} + \frac{1}{m_p} \frac{1}{\omega_n^2} k_{rp} f(\bar{\delta}_{rp}) \\
 & + \frac{1}{m_c} \frac{1}{\omega_n} \sum_{n=1}^N c_{sp} \dot{\delta}_{sp} + \frac{1}{m_c} \frac{1}{\omega_n^2} \sum_{n=1}^N k_{sp} f(\bar{\delta}_{sp}) + \frac{1}{m_c} \frac{1}{\omega_n} c_{rp} \dot{\delta}_{rp} + \\
 & \frac{1}{m_c} \frac{1}{\omega_n^2} \sum_{n=1}^N k_{rp} f(\bar{\delta}_{rp}) = \frac{T_{in} r_c}{I_c b_c \omega_n^2} - \frac{e_{mrp}}{b_c} \left(\frac{\omega_m}{\omega_n}\right)^2 \sin\left(\frac{\omega_m}{\omega_n} t + \psi_{spn}\right)
 \end{aligned} \right. \quad (14)$$

3 Numerical Simulation and Result Analysis

In this model, the meshing between the three planetary gears and the sun gear has the same dynamic characteristics. Hence, the dimensionless dynamic transmission error between the first planetary gear and the sun gear is selected as the example to analyze the dynamic characteristics of the system. The dynamic equations are solved by the fourth-order Runge–Kutta method. The basic parameters of components and meshing are displayed in Tables 1 and 2, respectively. The scale of displacement b_c is assigned 1e-5 m. The input torque is set as 100 Nm. In order to eliminate the influence of transient response, the first 500 response cycles are discarded.

Table 1 Parameters of components

Parameter name	Sun gear	Ring gear	Planet gear	The carrier
Teeth number	10	62	26	–
Mass (kg)	0.28	3.64	0.34	2
Inertia moment (kg·m ²)	2.10E-04	5.05E-04	2.90E-04	4.93E-03
Module (mm)	2.25	2.25	2.25	–
Pressure angle (°)	20	20	20	–

Table 2 Parameters of meshing

Parameter name	Sun-planet	Ring-planet
Average mesh stiffness(N/m)	4.5e8	5e8
Stiffness variation amplitude(N/m)	3.6e8	4e8
Damping ratio	0.07	0.07
Error amplitude(mm)	1e-5	1e-5

3.1 Response Characteristics of the Systems with Excitation Frequency

As an excitation parameter, the system motion state will change with the variation of the rotation frequency. Hence, the dimensionless rotation frequency Ω is selected as the excitation parameter to study the dynamic characteristics of the systems. Meanwhile, the dimensionless backlash b is assigned 2. The bifurcation diagrams of the two systems are shown in Figs. 4 and 5. It can be seen from the figures that as the excitation frequency increases, both systems exhibit complex bifurcation

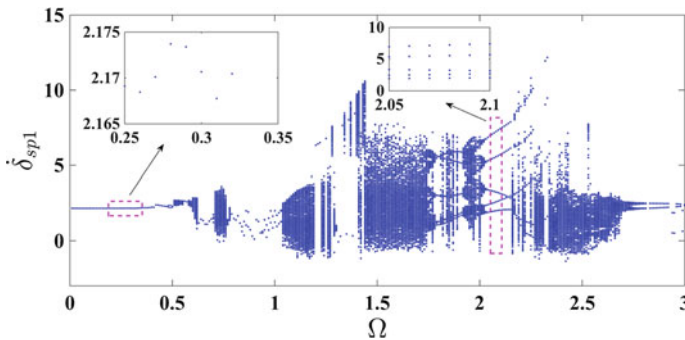


Fig. 4 Bifurcation diagram for the healthy system with dimensionless rotation frequency

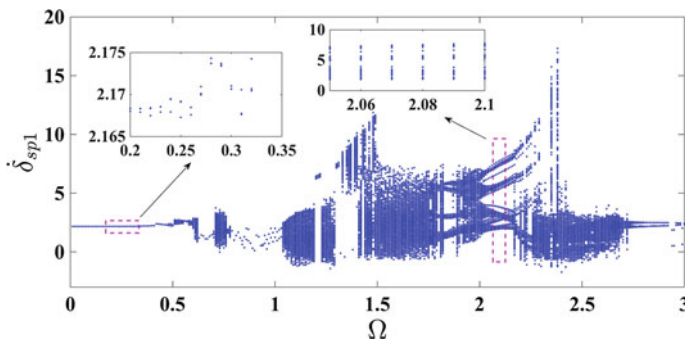


Fig. 5 Bifurcation diagram for faulty system with dimensionless rotation frequency

and chaotic characteristics. For the healthy system, as Ω varies in the range of 0–0.47, the system exhibits a single-period motion; And when it varies in 0.48–0.78, the system changes from paroxysmal chaos to periodic motion at beginning, and then as the excitation frequency increases, it undergoes a brief period-doubling bifurcation to evolve into chaotic motion with a small amplitude; In the interval of 0.79–1.03, the movement state of the system is single-period motion and 2-period motion alternately; When Ω changes in the range of 1.04–1.81, the motion of the system is more complicated, mainly showing chaotic motion, but there are multiple periodic windows alternately interspersed in the middle, and then it enters a 4-period motion through inverse period-doubling bifurcation; In the 1.82–2.16 interval, the system exhibits periodic motion and chaotic motion alternately and finally enters a 5-period motion; Afterwards, it undergoes inverse period-doubling bifurcation and enters chaotic motion with the periodic motion interspersed in the middle in the range of 2.17–2.68; In the interval of 2.69–3, the system finally enters a 2-period motion from a small amplitude chaotic motion. The dynamic characteristics of the faulty system is similar to the healthy system, but when the healthy system presents the periodic movement, the faulty system shows quasi-periodic motion. And in the excitation frequency range of the chaotic motion of the healthy system, the chaotic amplitude of the faulty system becomes larger, and the excitation frequency range corresponding to the chaotic motion of the faulty system is wider than that of the healthy system.

The bifurcation diagrams show the global characteristics of the system. Then, the local characteristics are depicted through the phase diagrams and the Poincaré maps. With various dimensionless excitation frequencies, the Poincaré maps and phase diagrams of the systems are displayed in Figs. 6, 7, 8 and 9. As can be seen from

Fig. 6 **a** Healthy system at $\Omega = 0.3$. **b** Faulty system at $\Omega = 0.3$

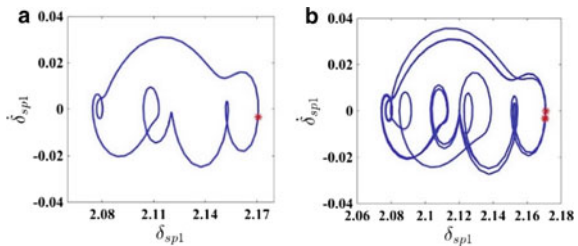


Fig. 7 **a** Healthy system at $\Omega = 1.2$. **b** Faulty system at $\Omega = 1.2$

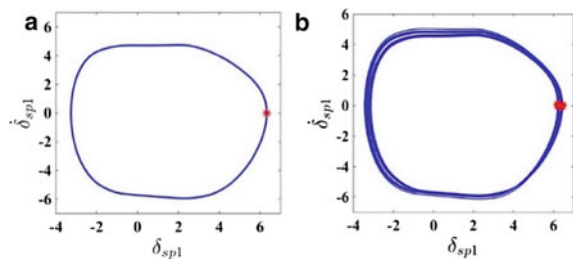


Fig. 8 **a** Healthy system at $\Omega = 2.1$. **b** Faulty system at $\Omega = 2.1$

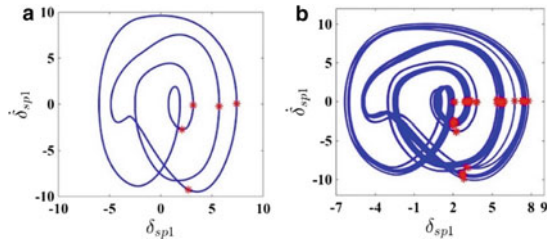


Fig. 9 **a** Healthy system at $\Omega = 2.38$. **b** Faulty system at $\Omega = 2.38$

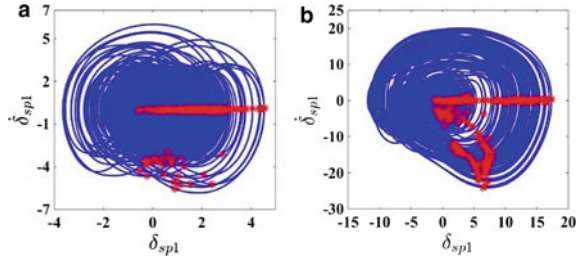


Fig. 6 that the healthy system presents a single-period motion at $\Omega = 0.3$, while the faulty system exhibits a two-period motion; When Ω is 1.2, as shown Fig. 7, the healthy system enters a single-period motion from chaotic motion, the faulty system enters a quasi-periodic motion from chaotic motion nevertheless; When Ω is 2.1, as shown in Fig. 8, the healthy system exhibits five-period motions respectively, while the faulty system presents quasi-periodic motions. At $\Omega = 2.38$, as shown in Fig. 9, both systems are in chaotic motion, but the amplitude of the faulty system is larger. Hence, when the healthy system exhibits periodic motion, the faulty system exhibits multi-period or pseudo-periodic motion at the corresponding excitation frequency and has a greater amplitude of motion.

3.2 Response Characteristics of the Systems with Backlash

As a nonlinear excitation, the tooth backlash has an important influence on the response characteristics of the system. To analyze the impact of backlash, Ω is assigned 0.6 and the dynamic responses with the initial backlash are shown in Figs. 10 and 11. As can be seen that the amplitude of the system motion expands with the increase of the backlash. For the healthy system, when the dimensionless backlash b varies between 0–0.6, the system exhibits a single period motion; When the backlash is between 0.48–0.97, the system enters the paroxysmal chaotic state after a short period-doubling bifurcation, and then enters the chaotic motion after the period of motion; When the backlash is in the interval of 0.98–3, the system enters chaotic motion from periodic motion, and finally enters chaotic motion through inverse

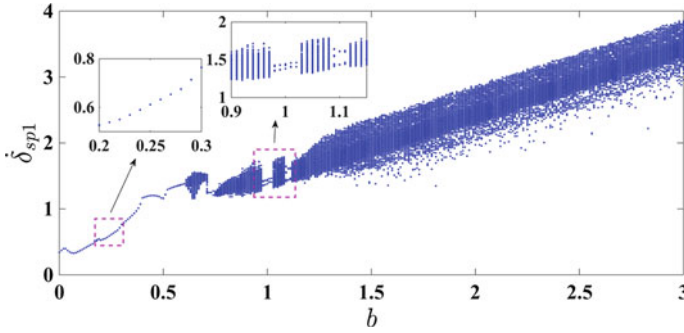


Fig. 10 Bifurcation diagram of healthy system with dimensionless backlash

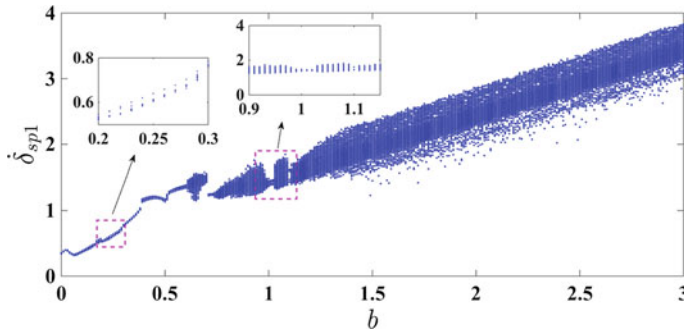


Fig. 11 Bifurcation diagram of faulty system with dimensionless backlash

period-doubling bifurcation and period-doubling bifurcation, and as the backlash increases, the amplitude of chaotic motion also increases. There is a similarity of the motion state between the faulty system and healthy system, but when the healthy system exhibits a single-period motion, the faulty system presents an obvious quasi-periodic motion. And when the backlash is 0.39 and 0.71, the system motion state occurs jump variation. The backlash interval corresponding to the chaotic motion of the faulty system is wider than that of the healthy system.

Then the Poincaré maps and phase diagrams are adopted to reveal local characteristics of the system. From Figs. 12, 13, 14 and 15, it can be seen that the healthy system exhibits a single-period motion at $b = 0.6$ and a quasi-periodic motion at $b = 0.61$, while the faulty system directly enters chaotic motion from quasi-periodic motion. When the dimensionless backlash is 0.72, the healthy system enters a single-period motion from chaotic motion, as shown in Fig. 14a. Nevertheless, the faulty system maintains a quasi-periodic movement consistently, as shown in Fig. 14b. When the backlash is 0.98, as shown in Fig. 15, the healthy system exhibits quasi-periodic motion, while the faulty system exhibits chaotic motion, indicating that the backlash

Fig. 12 **a** Healthy system at $b = 0.6$. **b** Faulty system at $b = 0.6$

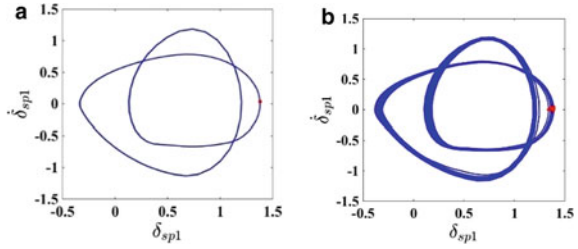


Fig. 13 **a** Healthy system at $b = 0.61$. **b** Faulty system at $b = 0.61$

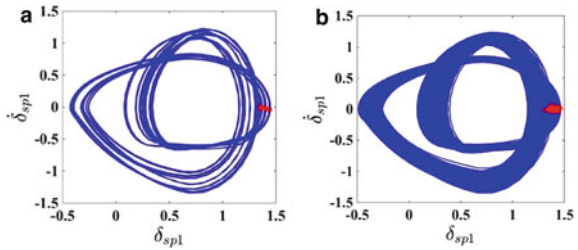


Fig. 14 **a** Healthy system at $b = 0.72$. **b** Faulty system at $b = 0.72$

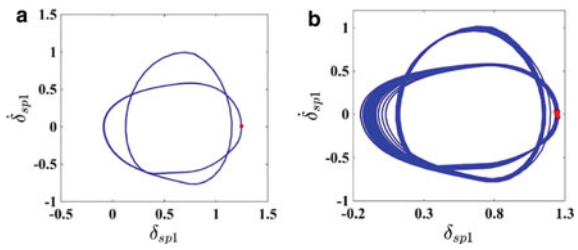
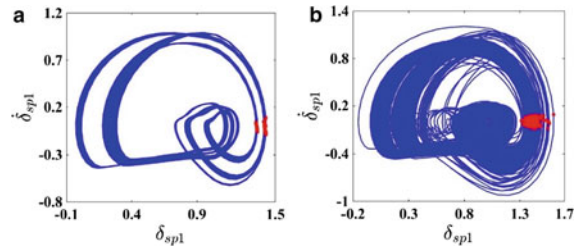


Fig. 15 **a** Healthy system at $b = 0.98$. **b** Faulty system at $b = 0.98$



bandwidth corresponding to the chaotic motion of the faulty system is wider than that of the healthy system.

4 Conclusions

In this study, a pure torsional dynamic model of planetary gear system with sun gear failure was established, which contains time-varying meshing stiffness, manufacturing error, and tooth backlash. Meanwhile, the dynamic characteristics of the healthy and faulty system with the variation of different nonlinear excitation are compared. By analyzing the nonlinear characteristics of the system, some conclusions can be obtained as follows:

- (1) Both the healthy and faulty system exhibit complicated nonlinear characteristics with the change of excitation frequency, and the multiple periods and quasi-period windows mix in the chaotic motion;
- (2) The motion state experienced by the faulty system is more complex than that of the healthy system. In the excitation frequency interval corresponding to the periodic motion state of the healthy system, the faulty system shows quasi-periodic motion, while in the chaotic motion interval of the healthy system, the faulty system has a larger chaotic motion amplitude and wider excitation frequency interval.
- (3) With the increase of the tooth backlash, the motion amplitude of both healthy and faulty systems increases.
- (4) In the dimensionless tooth backlash interval corresponding to the periodic motion of the healthy system, the faulty system shows quasi-periodic motion and appears to jump variation. The backlash interval corresponding to the chaotic motion of the faulty system is wider than that of the healthy system.

Acknowledgements The work was supported by the National Natural Science Foundation of China under Grant nos. 51875166. The support is gratefully acknowledged.

References

1. Kumar A., Gandhi C.P., Zhou Y., Kumar R., Xiang J.: Latest developments in gear defect diagnosis and prognosis: a review. *Measurement* **158**(107735) (2020)
2. Kahraman, A.: Load sharing characteristics of planetary transmissions. *Mech. Mach. Theory* **29**, 1151–1165 (1994)
3. Parker, R.G., Wu, X.: Vibration modes of planetary gears with unequally spaced planets and an elastic ring gear. *J. Sound Vib.* **329**, 2265–2275 (2010)
4. Zhai, H., Zhu, C., Song, C., Liu, H., Bai, H.: Influences of carrier assembly errors on the dynamic characteristics for wind turbine gearbox. *Mech. Mach. Theory* **103**, 138–147 (2016)
5. Mo, S., Zhang, Y., Wu, Q., Matsumura, S., Houjoh, H.: Load sharing behavior analysis method of wind turbine gearbox in consideration of multiple-errors. *Renew. Energy* **97**, 481–491 (2016)
6. Li, S.: Effects of machining errors, assembly errors and tooth modifications on loading capacity, load-sharing ratio and transmission error of a pair of spur gears. *Mech. Mach. Theory* **42**, 698–726 (2007)
7. Zhao B., Huangfu Y., Ma H., Zhao Z., Wang K.: The influence of the geometric eccentricity on the dynamic behaviors of helical gear systems. *Eng. Fail. Anal.* **118**(104907) (2020)

8. Park, C.I.: Dynamic behavior of the spur gear system with time varying stiffness by gear positions in the backlash. *J. Mech. Sci. Technol.* **34**, 565–572 (2020)
9. Wang, J., He, G., Zhang, J., Zhao, Y., Yao, Y.: Nonlinear dynamics analysis of the spur gear system for railway locomotive. *Mech. Syst. Signal Process.* **85**, 41–55 (2017)
10. Wang, J., Zhang, J., Yao, Z., Yang, X., Sun, R., Zhao, Y.: Nonlinear characteristics of a multi-degree-of-freedom spur gear system with bending-torsional coupling vibration. *Mech. Syst. Signal Process.* **121**, 810–827 (2019)
11. Yang, Y., Cao, L., Li, H., Dai, Y.: Nonlinear dynamic response of a spur gear pair based on the modeling of periodic mesh stiffness and static transmission error. *Appl. Math. Model.* **72**, 444–469 (2019)
12. Parker, R.G., Guo Y.: Dynamic analysis of planetary gears with bearing clearance. *J. Comput. Nonlinear Dyn.* **7**(041002) (2012)
13. Eritenel, T., Parker, R.G.: Nonlinear vibration of gears with tooth surface modifications. *J. Vib. Acoust. Trans. ASME* **135**(5), 051005 (2013)
14. Huang, K., Cheng, Z., Xiong, Y., Han, G., Li, L.: Bifurcation and chaos analysis of a spur gear pair system with fractal gear backlash. *Chaos, Solitons Fractals Interdiscip. J. Nonlinear Sci. Nonequilibrium Complex Phenom.* **142**, 110387 (2020)
15. Yang, Y., Xia, W., Han, J., Song, Y., Wang, J., Dai, Y.: Vibration analysis for tooth crack detection in a spur gear system with clearance nonlinearity. *Int. J. Mech. Sci.* **158**, 648–661 (2019)
16. Xiang, L., Gao, N., Hu, A.: Dynamic analysis of a planetary gear system with multiple nonlinear parameters. *J. Comput. Appl. Math.* **327**, 325–340 (2018)
17. Wei, S., Han, Q., Dong, X., Peng, Z., Chu, F.: Dynamic response of a single-mesh gear system with periodic mesh stiffness and backlash nonlinearity under uncertainty. *Nonlinear Dyn.* **89**, 49–46 (2017)

A Fault Diagnosis Method for Rolling Bearings Based on Improved EEMD and Resonance Demodulation Analysis



Wei Zhang, Xiange Tian, Guohai Liu, and Hui Liu

Abstract Rolling bearing is a kind of easily damaged mechanical equipment. The quality of rolling bearing is related to the normal operation of the equipment. Because the resonance demodulation method is susceptible to noise interference, and the band-pass filter parameters are largely dependent on personal experience selection. This paper proposes an analysis method based on the combination of Ensemble Empirical Mode Decomposition (EEMD) and the selection criterion of kurtosis-cross-correlation coefficient. Firstly, the vibration signal is decomposed by EEMD to get intrinsic mode functions (IMFs); Secondly, since the decomposed IMF components will produce mode aliasing, two criteria of cross-correlation coefficient and kurtosis are introduced to extract effective IMF components for signal reconstruction; Finally, the reconstructed signal is subjected to Hilbert transform and envelope analysis. Compared with the resonance demodulation analysis method, the EEMD decomposition method is selected to replace the band-pass filter to reduce the noise of the signal, which enhances signal to noise ratio and makes the fault characteristics more obvious. The experimental signal analysis results of rolling bearing faults show that a refinement of methodology presented in this article can effectively extract the fault characteristics of rolling bearing, and has more advantages than traditional envelope analysis methods.

Keywords Ensemble empirical mode · Resonance demodulation · Envelope spectrum · Rolling bearing · Fault diagnosis

W. Zhang · X. Tian (✉) · G. Liu · H. Liu
School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China
e-mail: tianxiange@ujs.edu.cn

W. Zhang
e-mail: zw18852893705@163.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_54

669

1 Introduction

The rolling bearing is in each kind of revolving machinery applies one of most widespread general machine parts. The normal running state of rolling bearing often directly affects the performance of the whole device, so fault diagnosis is very important [1].

The fault diagnosis of rolling bearing includes four steps: vibration signal acquisition, signal preprocessing, fault feature extraction and pattern recognition. Because there will be interference factors such as noise when the signal is collected, signal preprocessing is an indispensable part of the fault diagnosis process [2]. The signal processing method can be performed in the time domain, frequency domain, and time–frequency domain. Time domain analysis is widely applied in the breakdown diagnosis domain due to their intuitive, easy-to-understand, easy-to-calculate, and high-efficiency advantages [3]. In the time domain index, kurtosis is extremely sensitive to impact characteristics and plays the vital role in the bearing partial expiration [4]. Compared with the time domain analysis, the frequency range analysis mainly separates or the strengthened breakdown characteristic frequency component. When the rolling bearing fails, a modulation component will appear in the collected signal. Envelope spectrum analysis is a powerful tool for processing modulated signals [5]. In these methods, envelope spectrum analysis based on Hilbert transform is the most widely applied. However, the frequency domain analysis method is based on Fast Fourier Transform (FFT). FFT lacks local information for the analysis of non-smooth signals, and is not appropriate for analyzing non-smooth signals [6, 7].

The actual vibration signal is usually unstable. The analysis of nonstationary signal has been studied by many people in the field of signal processing. Empirical mode decomposition (EMD) is a method of analyzing non-stationary signals proposed by Huang [8], and its essence is to process non-stationary signals. EMD is equivalent to an adaptive filter. It can decompose non-smooth signals into a series of Intrinsic Mode Function (IMF). Each IMF component has its physical meaning, and the adaptive and noise reduction characteristics of EMD make it more and more widely used in rolling bearings [9–11]. Literature [12] proposed EMD to reduce the noise of rolling bearing vibration signals, and realizes fault diagnosis of rolling bearing with envelope spectrum analysis. However, the EMD method still has many defects such as modal aliasing, end effect, over-envelope and under-envelope phenomena. In order to improve the modal aliasing phenomenon of EMD, Ref. [13] proposed the total integrated empirical mode decomposition method (EEMD). EEMD through many times joins the white noise to the primary signal in, may suppress the modality aliasing effectively. In Ref. [14], the author used the decomposition method of EEMD and compared it with the EMD decomposition results. Finally, it is concluded that EEMD decomposition has more advantages.

Because the traditional resonance demodulation method needs to select a resonance high frequency band to design a bandpass filter based on personal experience and knowledge reserves. In order to improve this shortcoming, this paper uses EEMD decomposition to replace the filtering method of the band-pass filter, and selects the

IMF component of the reconstructed signal through the cross-correlation coefficient and kurtosis. Then Hilbert transform and envelope analysis are performed on the reconstructed signal, and finally the fault features are extracted. The method proposed in this paper can effectively avoid the interference of artificial selection of bandwidth factors on fault analysis, and the correctness of fault diagnosis and analysis is enormously enhanced.

2 Methodology

2.1 Resonance Demodulation Algorithm

Resonance demodulation technology uses band-pass filter to extract the high frequency resonance signal of low frequency fault pulse modulation, and then obtains the low frequency signal spectrum through envelope demodulation. The fault type is determined by comparing the actual characteristic frequency with the theoretical fault characteristic frequency. The diagnosis effect mainly depends on the parameter selection of band-pass filter: center frequency and bandwidth. Appropriate band-pass filter can effectively filter noise and other interference factors, and improve the accuracy of fault feature frequency extraction [15]. The specific algorithm is as follows:

- (1) The signal is transformed by fast Fourier transform to get the spectrum;
- (2) Observe the frequency band where the modulation phenomenon is more obvious from the spectrogram, and select this frequency band as the bandwidth of the band-pass filter;
- (3) After determining the bandwidth, design a band-pass filter to filter the signal;
- (4) Perform Hilbert transform and envelope analysis on the filtered signal to obtain its envelope spectrum;
- (5) Observe the envelope spectrum, extract the characteristic frequency of the fault, and compare it with the theoretical value to judge the fault type.

In the fault diagnosis of rolling bearing, resonance demodulation is the most widely used method, but because the information generated by early small faults of bearing is often disturbed by background noise, the application of resonance demodulation method in improving signal-to-noise ratio is limited, and the diagnosis effect is not obvious. In recent years, some new denoising methods have been developed rapidly. Wavelet denoising has the advantage of multi-resolution. However, the effect of wavelet denoising largely depends on the selection of basis function and threshold, so designers need to have rich experience. EMD is a new signal processing method, which is very suitable for processing nonlinear and non-stationary signals. However, the EMD decomposition is prone to mode aliasing. In order to solve this problem, some scholars put forward the EEMD method based on the research of white noise

EMD decomposition. This method effectively overcomes the shortcoming of mode aliasing in EMD method.

2.2 Ensemble Empirical Model Decomposition

EMD can adaptively process nonlinear and non-stationary signals, but this method has problems and shortcomings, mainly the phenomenon of modal aliasing. The EEMD method adds multiple groups of different white noises to the original signal and then performs EMD decomposition, and then uses the random characteristic of zero white noise to average the IMF components obtained from all EMD decompositions as the components of the EEMD decomposition IMF to eliminate the white noise [16]. At the same time, the problem of modal aliasing is solved. EEMD decomposition steps are as follows:

- (1) Select the total average number of decomposition M ;
- (2) A white noise $n_i(t)$ with normal distribution is added to the original vibration signal $x(t)$ to form a new signal:

$$x_i(t) = x(t) + n_i(t) \tag{1}$$

where $n_i(t)$ represents the i th additive white noise sequence, and $x_i(t)$ represents the additional noise signal of the i th experiment, $i = 1, 2, \dots, M$;

- (3) The new signal $x_i(t)$ is decomposed by EMD to get the respective IMF:

$$x_i(t) = \sum_{j=1}^J c_{i,j}(t) + r_{i,j}(t) \tag{2}$$

where $c_{i,j}(t)$ is the j th IMF decomposed after adding white noise for the i th time, $r_{i,j}(t)$ is the residual function, which represents the average trend of the signal, and J is the number of IMF;

- (4) Repeat steps (2) and (3) for M times, and add white noise signals with different amplitudes each time to get the set of IMF:

$$c_{1,j}(t) c_{2,j}(t) \dots c_{M,j}(t), j = 1, 2, 3, \dots, J$$

- (5) Based on the principle that the statistical mean value of uncorrelated sequence is 0. The final IMF component can be obtained by calculating the above IMF components, namely:

$$c_j(t) = \frac{1}{M} \sum_{i=1}^M c_{i,j}(t) \tag{3}$$

where $c_j(t)$ is the j th IMF decomposed by EEMD, $i = 1, 2, \dots, M$; $j = 1, 2, \dots, J$.

In the EEMD decomposition method, two parameters are needed: the number of average M and the amplitude of white noise. The amplitude of white noise is usually characterized by the ratio of the standard deviation of white noise amplitude to the standard deviation of original signal amplitude [17].

The EEMD algorithm is an effective method to deal with non-linear and non-stationary signals. It solves the mode aliasing in the process of signal decomposition, but it also has some disadvantages, such as residual white noise in the process of signal decomposition. The choice of an effective IMF depends entirely on experience. All these affect the accuracy of EEMD decomposition and reconstruction. For this reason, two criteria, cross-correlation coefficient and kurtosis, are introduced to select and reconstruct IMF components.

2.3 Kurtosis and Cross-Correlation Coefficient

Kurtosis is a measure of how much the distribution of a set of random variables deviates from the Gaussian distribution. The signal of normal rolling bearing is close to Gaussian distribution, and its kurtosis value is about 0. When the rolling bearing fails, its kurtosis value is greater than 0, and the impact component of the fault signal is prominent. The magnitude of the kurtosis value reflects the degree of impact of the impact component, and a value between 3 and 8 has a significant effect on the extraction of weak faults.

The cross-correlation coefficient indicates the degree of correlation between two signals. The greater the correlation coefficient between two random signals, the stronger the correlation degree [18]. Generally, the correlation coefficient should be greater than 0.1. Equation (4) is the definition of the correlation coefficient in this article:

$$R(x, imf) = \frac{\sum_{i=1}^N [x(t) - \bar{x}][imf(t) - \overline{imf}]}{\sqrt{\sum_{i=1}^N [x(t) - \bar{x}]^2} \sqrt{\sum_{i=1}^N [imf(t) - \overline{imf}]^2}} \tag{4}$$

where N is the number of sampling points; $x(t)$ is the original vibration signal; $imf_i(t)$ is the i th IMF component, and $\bar{x} = \frac{1}{N} \sum_{i=1}^N x(t)$.

From the cross-correlation coefficient between each IMF component and the original signal, we can find the first imf_k with the local minimum value of the cross-correlation coefficient and the imf_{k+1} is considered to be the modal aliasing component. Then the first k IMF components are highly correlated with the original signal and contain more fault information. In addition, since the IMF component is from high frequency to low frequency, the high frequency part contains more fault information, so we give priority to the high frequency part. The remaining components can be directly eliminated, and then the selected components can be accumulated and reconstructed to obtain the denoised signal [19, 20].

3 Improved EEMD Decomposition Algorithm

Due to the modal aliasing phenomenon in the IMF components decomposed by the EMD method, and the end effect affects the decomposition effect. In order to avoid such problems in the experiment, this paper chooses to improve the EMD method, that is, the EEMD method, which effectively solves the above problems. Because the spectrum of white noise is evenly distributed, when we add white noise to the signal to be analyzed, it will be automatically distributed to the appropriate location. Because the mean value of noise is 0, the influence of white noise can be eliminated after several average calculations. The final result can be obtained by integrating and averaging each IMF. Therefore, the shortcomings of EMD decomposition are improved. The signal is decomposed by EEMD to get IMF component. Generally, the first component will be selected as the next signal to be studied, which will lose some fault information. Here, we make a little improvement: by calculating the kurtosis of IMF component and the cross-correlation coefficient between IMF component and original signal, we select the component reconstruction signal according to the selection criteria in Sect. 2.3. The steps of the improved EEMD algorithm are as follows:

- (1) EEMD decomposition of the vibration signal will result in a number of IMF components;
- (2) Compute the kurtosis of IMF and the cross-correlation coefficient between IMF and signal;
- (3) According to the selection criteria proposed earlier in the thesis, compare the correlation values and kurtosis values, and select the appropriate IMF component to reconstitute the signal;
- (4) Perform Hilbert transform on the reconstructed signal, and perform envelope demodulation analysis to obtain an envelope spectrogram;
- (5) Observe the envelope spectrum, look for the characteristic frequency of the fault, and compare it with the theoretical value to judge the fault type.

EEMD decomposition can effectively denoise the signal. The selection component is improved slightly. In the next analysis, it is used to replace the band-pass filter, which avoids the shortcomings of choosing the band-pass filter parameters according to personal experience. This paper considers the shortcomings of traditional methods, and uses an improved EEMD method instead of a band-pass filter to analyze the signal. Figure 1 shows the algorithm flow of the traditional Hilbert envelope spectrum, and Fig. 2 shows the algorithm flow of the Hilbert envelope spectrum based on the improved EEMD method.

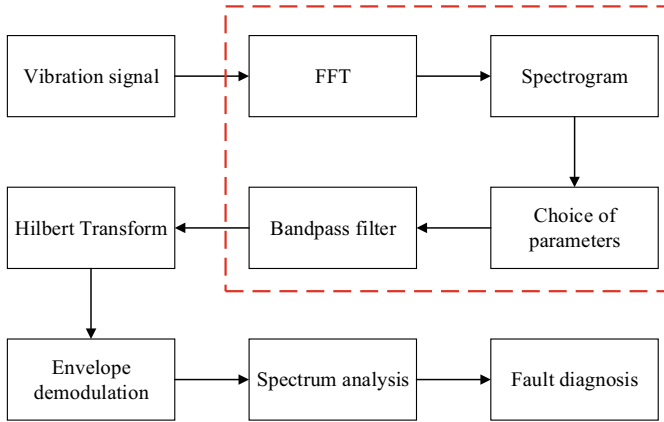


Fig. 1 Flow chart of resonance demodulation method based on fixed bandwidth

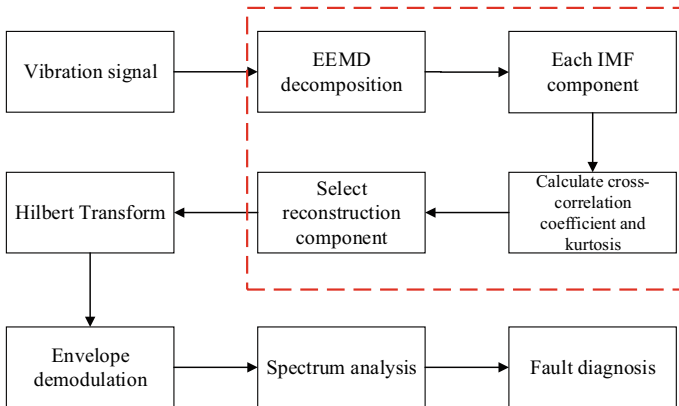


Fig. 2 Flow chart of resonance demodulation method based on improved EEMD

4 Experiment and Analysis

4.1 Data Sources

The data used in this paper are from the life cycle bearing data provided by the University of Cincinnati [21]. Figure 3 shows the physical picture of the bearing in the experiment and the simulation picture which is easy to watch. The experimental equipment consists of an AC motor, four bearings (Rexnord za-2115 double row bearings) and a vibration sensor. In this experiment, the number of rolling elements is 16 ($z = 16$), and the pitch diameter of bearing raceway is 2.815 inches ($D = 2.815$ in); The diameter of rolling element is 0.331 inch ($d = 0.331$ in); The contact angle

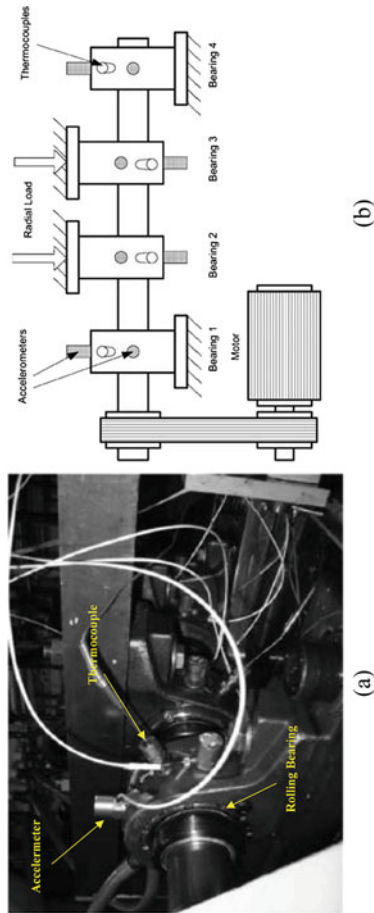


Fig. 3 The experimental platform: **a** Experimental platform of bearing. **b** Simulation experimental platform of bearing

is $15.17^\circ (\alpha = 15.17^\circ)$. The rotation speed of the bearing is 2000 rpm ($f_r = 33.33$ Hz) and the sampling rate is 20 kHz. The vibration signal is collected every ten minutes. Each file in the dataset consists of 20,480 points. NI DAQ 6062E was used to collect data in the experiment.

The formula for outer ring fault frequency [22] is

$$f_o = \frac{N}{2} \left(1 - \frac{D}{d} \cos \alpha \right) f \tag{5}$$

According to Eq. (5), the outer ring fault frequency is 236.4 Hz.

4.2 Data Analysis Based on Fixed Bandwidth

Firstly, we select four time-domain indicators of RMS, absolute average, variance and kurtosis to make a preliminary analysis and judgment on the signal. Figure 4 shows the waveforms of the four indicators of this signal. RMS is the reflection of signal impulse characteristics. The absolute average reflects the energy of the signal. Variance reflects the degree of signal dispersion. From the change trend of the four indicators in Fig. 4, we can determine that the bearing must have a fault in the later stage.

According to the changing trends of the four indicators, the failure of the entire cycle initially occurred near Document 500. Next, we use the content of Sect. 2.1 to analyze the vibration signal. Because the traditional resonance demodulation method needs to rely on experience to select the bandwidth of the band-pass filter. In order to better select the bandwidth, here we select the data collected without failure (file number 30), the data collected at the initial stage of the failure (file number 533), and the data collected after the failure (file number 800) for analysis. Figure 5 shows

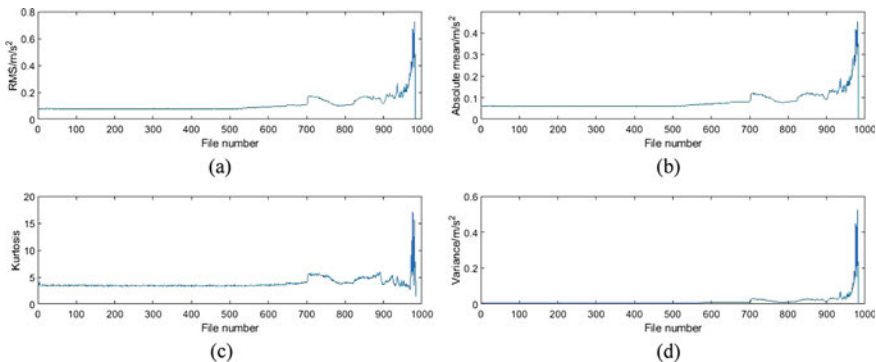


Fig. 4 Four time-domain indicators of vibration signals: **a** RMS. **b** Absolute mean. **c** Variance. **d** Kurtosis

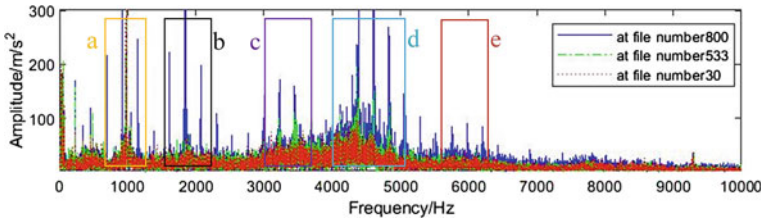


Fig. 5 Resonance frequency band: **a** [600, 1400 Hz]; **b** [1600, 2400 Hz]; **c** [3000, 3800 Hz]; **d** [4000, 5000 Hz]; **e** [5500, 6500 Hz]

the selection of resonance frequency band in the spectrum of the three files. It can be found from the figure that when the center frequency is 1000, 2000, 3400, 4500 and 6000, the resonance frequency band is more prominent. Based on this, the bandwidth of the band-pass filter is designed to be [600, 1400 Hz], [1600, 2400 Hz], [3000, 3800 Hz], [4000, 5000 Hz], [5500, 6500 Hz] to band-pass filter the signal. Choose different bandwidths to design the band-pass filter, and then analyze the signals separately according to Sect. 2.1. Finally, it is found that the envelope spectrum obtained by [4000, 5000 Hz] has the most obvious fault characteristics. Figure 6 shows the envelope spectrum at file numbers 30, 500, and 800, with a bandwidth of [4000, 5000 Hz].

It is obvious from the Fig. 6 that the fault has occurred in the later stage of the signal, and the characteristic frequency is 230.5 Hz, which is very close to the result

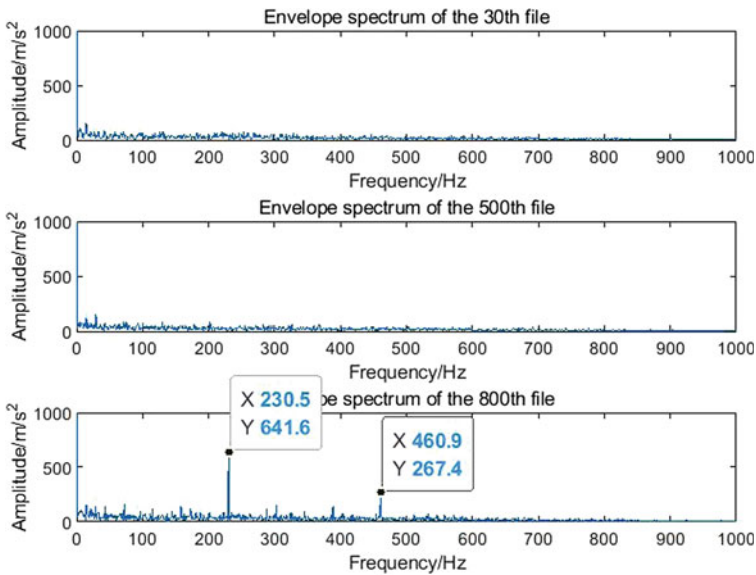


Fig. 6 Envelope spectrum obtained with a bandwidth of [4000, 5000 Hz] at file number 30 500 and 800

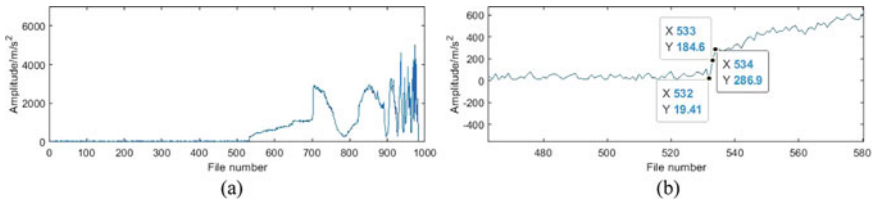


Fig. 7 Trend chart of fault characteristic frequency: **a** Whole cycle; **b** Partially enlarged trend graph

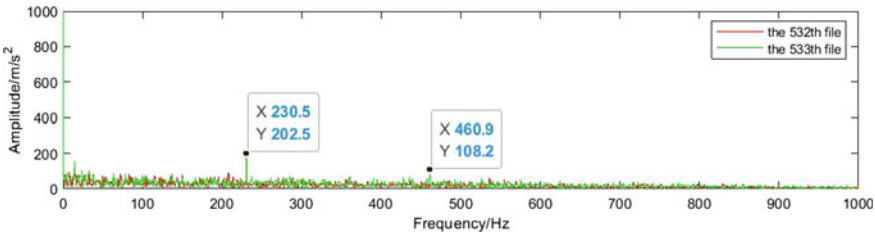


Fig. 8 Envelope spectrogram at file number 532 and 533

calculated by Eq. (5) of 236.4 Hz, and it also can find that the twice fault characteristic frequency. Based on this, it can be judged that the actual fault characteristic frequency is 230.5 Hz.

In order to know the time of the initial failure of the fault more accurately, we know the fault characteristic frequency of the bearing through the above analysis. Figure 7 shows the trend diagram of the extracted fault feature frequency and its partial enlarged diagram.

It can be found from Fig. 7b that the rolling bearing failure at file number 533 began to occur. Then the file number 532 and 533 are analyzed by the resonance demodulation method to obtain the envelope spectrum. Figure 8 shows the envelope spectrogram obtained after the resonance demodulation method at file number 532 and 533. We can find that no clear fault characteristic frequency can be found in the envelope spectrum at file number 532. The fault characteristic frequency and its twice fault characteristic frequency can be found at file number 533. Therefore, it can be known that the bearing started to fail at file number 533.

4.3 Data Analysis Based on Improved EEMD

Through the analysis of the traditional resonance demodulation method, we know that the initial fault location is at file number 533, and then use the improved method proposed in this paper for this set of data. The signal is decomposed by EEMD, and then each IMF component can be obtained. The kurtosis of each component and the cross-correlation coefficient between each IMF component and the original signal

Table 1 Cross-correlation coefficients and Kurtosis

IMF	Cross-correlation coefficients	Kurtosis
IMF1	0.7614	3.7983
IMF2	0.4543	3.6152
IMF3	0.4630	2.4166
IMF4	0.2171	3.1758
IMF5	0.1393	3.3171
IMF6	0.1061	2.8522
IMF7	0.1643	3.0198
IMF8	0.1287	2.3135
IMF9	0.0256	2.7608
IMF10	0.0070	3.7430
IMF11	0.0020	3.2776

are calculated. Table 1. shows the cross-correlation number and kurtosis of each IMF component.

Because the resonance caused by the fault is mostly obvious in the high frequency part, and the IMF component is from high frequency to low frequency. Therefore, we prefer to choose the kurtosis and cross-correlation coefficients of the first few IMF components to observe and compare. According to the selection rules described in Sect. 2.3, from Table 1, we can find that the kurtosis of IMF1 and IMF2 components are between 3 and 8, while the kurtosis of IMF3 is not in this range, so IMF1 and IMF2 contain more fault characteristics. Similarly, it is found that the cross-correlation coefficients of IMF1 and IMF2 components are both greater than 0.1, and the first local minimum cross-correlation coefficient is found at IMF2. Therefore, the kurtosis and cross-correlation coefficients of IMF1 and IMF2 components conform to the selection rules, so we choose IMF1 and IMF2 to reconstruct the signal. Figure 9 shows the envelope spectrum at file number 533 obtained using the method in Part 3.

Comparing Fig. 9. with Fig. 8, in the analysis of the same initial fault, the fault frequency obtained by the method proposed in this paper is more obvious, and the twice frequency, three times frequency, and four times frequency can all be clearly found. The fault diagnosis has a very intuitive judgment function.

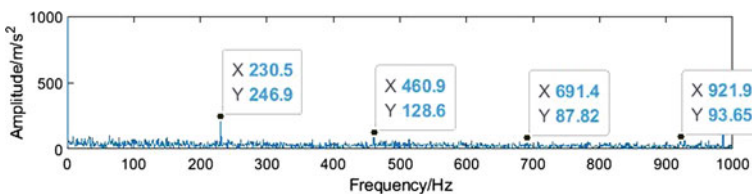


Fig. 9 The envelope spectrum at file number 533 based on the improved EEMD

Table 2 SNR of the two methods

File number	The SNR of traditional method	The SNR of the method in this paper
533th file	-1.1795	5.7100
540th file	-1.3481	6.2237
600th file	-1.8977	6.5063
700th file	-2.7630	7.9383
800th file	-2.0808	6.9007
850th file	-6.1887	10.5551

The above analysis shows that the method proposed in this paper has a better filtering and denoising effect than the band-pass filter, and has more advantages in the extraction of fault signals. Here we use the signal-to-noise ratio (SNR) [23] to quantitatively analyze the denoising effect. The calculation expression of SNR is

$$SNR = 10 \lg \left[\frac{\sum_{n=0}^{N-1} S_n^2}{\sum_{n=0}^{N-1} (S_N - S'_n)} \right] \tag{6}$$

where S_n is the original signal; S'_n is the signal after noise reduction; N is the number of sampling points.

Table 2 shows the SNR calculated by the resonance demodulation method using a fixed frequency band and the method proposed in this paper after denoising the signal. Select several sets of data from the early stage to the later stage of the failure for analysis and calculation. It can be found from the table that the SNR after denoising the signal using this method has been greatly improved.

5 Conclusion

In this paper, the resonant demodulation method is vulnerable to noise interference and the parameters of band-pass filter are difficult to determine. In most cases, the design of band-pass filter depends on experience to select the bandwidth, which has a great impact on the signal diagnosis. Therefore, this paper proposes an improved EEMD method instead of band-pass filter to remove the noise in the vibration signal. The measured vibration signal is decomposed by EEMD, and the appropriate IMF component is selected for reconstruction by combining the cross-correlation coefficient criterion and kurtosis criterion. Finally, the reconstructed signal is processed by Hilbert transform and envelope demodulation to get the envelope spectrum. By observing the envelope spectrum obtained by the two methods and comparing the SRN of the two methods, it can be seen that the proposed method has greater advantages than the traditional envelope analysis method. This study is supported by Jiangsu University Senior Talents Research Start-up Fund 4111140012.

References

1. Yang, J.H., Li, M., Ding, F.Y.: On-site practical technology of rolling bearing diagnosis. Mechanical Industry Press (2015)
2. Zhang, F.B., Huang, J.F., Chu, F.L.: Mechanism and method for outer raceway defect localization of ball bearings. *IEEE Access* **08**, 4351–4360 (2020)
3. Bian, J., Wang, P., Mei, Q.: Bearing fault diagnosis based on EEMD combining energy features and wavelet denoising. *J. Guangxi Univer. (Nat. Sci. Ed.)* **39**(06), 1206–1211 (2014)
4. Ju, P.H.: Research on time-frequency analysis method for early fault feature extraction of rotating machinery. Chongqing University (2010)
5. Yang, Y., Yu, D., Cheng, J.: A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Measurement* **40**(09), 943–950 (2007)
6. Randall, R.B., Antoni, J., Chobsaard, S.: The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mech. Syst. Signal Process.* **15**(05), 945–962 (2001)
7. Hahn S.L.: *Hilbert Transforms in Signal Processing*, pp. 7–15 Artech House Publish (1996)
8. Zhang, S., Liu, D.P.: Bearing fault diagnosis based on BFA optimization of VMD parameters. *Modular Mach. Tool Automated Proc. Technol.* **51**(05), 45–47 (2020)
9. Leng, J.F., Jing, S.X., Yu, J.G.: Application of EMD and energy operator demodulation in fault diagnosis of hoist gearbox. *J. China Coal Soc.* **38**, 530–535 (2013)
10. Jing, S.X., Dong, S.C., Hua, W.: Research on vibration fault diagnosis of shearer cutting part based on EMD and energy operator demodulation. *J. Henan Polytech. Univer.* **33**(06), 766–769 (2014)
11. Cheng, J.S., Yu, D.J., Yang, Y.: Energy operator demodulation method based on EMD and its application in mechanical fault diagnosis. *J. Mech. Eng.* **40**(08), 115–118 (2004)
12. Wang, Z.C., Wang, S.L., Ren, K.S.: EEMD-based method for diagnosing the failure of the hoist wheel bearing. *J. Coal Sci.* **37**(04), 689–694 (2014)
13. Wu, Z., Huang, N.E.: Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **1**(01), 1–41 (2011)
14. Peng, Z.K., Chu, F.L.: A comparison study of improved Hilbert-Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. *Mech. Syst. Sig. Process.* **19**(05), 974–988 (2005)
15. Lei, Y.G., Han, D., Lin, J.: New adaptive stochastic resonance method and its application in fault diagnosis. *Mech. Eng.* **48**(07), 62–67 (2012)
16. Chen, R.X., Tang, B.P., Ma, J.H.: Adaptive noise reduction method of vibration signal based on EEMD. *Vib. Shock* **31**(15), 82–86 (2012)
17. Wang, J., Gao, R.X., Yan, R.: Integration of EEMD and ICA for wind turbine gearbox diagnosis. *Wind Energy* **17**(5), 757–773 (2014)
18. Cai, Y.P., Li, A.H., Shi, L.S.: Rolling bearing fault detection based on EMD and spectral kurtosis is improved into envelope spectrum analysis. *Vib. Impact* **30**(2), 167–172 (2011)
19. Liu, B., Dong, H., Qian, S.Y.: Ultrasonic signal noise reduction method based on empirical mode decomposition and wavelet analysis. *Test. Technol.* **32**(05), 422–428 (2018)
20. Li, Y., Peng, J.L., Ma, H.T., Lin, H.B.: Research on the influence of transitional intrinsic modal function on the denoising result of empirical mode decomposition and its improved algorithm. *Chin. J. Geophys.* **56**(02), 626–634 (2013)
21. Hai, Q., Jay, L., Jing, L.: Wavelet Filter-based weak signature detection method and its application on roller bearing prognostics. *J. Sound Vib.* **289**, 1066–1090 (2006)
22. Tian, R.: Research on de-trend analysis and fault feature extraction methods of bearing vibration signals. *Mach. Des. Manuf.* **12**, 100–104 (2018)
23. Ji, Z.X., Ma, C.W.: Fiber optic gyroscope's EMD filtering method based on SNR detection. *Piezoelectric Acousto-Optic* **34**(06), 831–833 (2012)

A Study on the Contact Characteristics of a Planetary Centrifugal Vari-Speed Drive



Jin Li, Jing Liu, Chaojie Zhong, Wujun Zou, and Ruikun Pang

Abstract The vari-speed drive, which transmits motion and torque through planetary gear and centrifugal rotor, plays an important role in mechanical transmission system. It is mainly used in ships, rolling mills, automobiles and other fields. The vari-speed drive can improve the dynamic performance and economy of the vehicle through the continuous change of transmission ratio. In this paper, an innovative planetary centrifugal vari-speed drive is presented. The structure and working principle of the device are introduced. The multi-body dynamics modeling of the vari-speed drive is established to obtain the working characteristics and vibration characteristics of the mechanism under different working conditions. A compression spring is installed between the centrifugal rotors, and the influence of spring on the system performance is analyzed. This paper can provide a theoretical basis for the design of planetary stepless transmission.

Keywords Vari-speed drive · Spring stiffness · Vibration · Contact force · Stepless transmission

J. Li · C. Zhong · W. Zou
China North Vehicle Research Institute, Beijing 100081, P. R. China
e-mail: lijin929@126.com

J. Liu (✉)
School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, P. R. China
e-mail: jliu0922@nwpu.edu.cn

Key Laboratory for Unmanned Underwater Vehicle, Northwestern Polytechnical University, Xi'an 710072, P. R. China

R. Pang
College of Mechanical Engineering and Vehicle, Chongqing University, Chongqing, China

1 Introduction

The vari-speed drive is a form of transmission system. Compared with the step speed change, vari-speed drive can realize the speed of the stepless adjustment, to achieve the input and output in any range of changes. In recent years, many researchers have investigated mechanical vari-speed drive. Lagonravov et al. [1] considered the vibration of the internal components of the stepless mechanical transmission. They described the principle of automatic control of the transmission according to the vibration amplitude of the internal components. Dai et al. [2] introduced a new satellite star type gear (SSG) stepless transmission system based on the pulse stepless transmission. They described the basic mechanical structure and kinematics principle of the whole mechanism. They also carried out kinematics simulation, and obtained the kinematics characteristic curve. Milazzo et al. [3] proposed a new self-adjusting annular variable speed transmission, which has the function of passive control of speed ratio. By deducing the mathematical model of the system including dry friction dissipation, Lazarek et al. [4] proposed a new type of constantly variable transmission (CVT) designed for a new tuned mass damper. The motion of the CVT is oscillating, and the test is carried out under actual working conditions. Li et al. [5] proposed a non-circular gear design method for CVT based on a modified high-order elliptical pitch curve. They studied the power loss assessment of non-circular gear pairs. Ferguson et al. [6] discussed several methods for improving the efficiency of CVT by combining non-circular gears. The test results of a prototype transmission are reported. Liu et al. [7, 8] established a planetary gear system considered the flexibility of ring gear. They studied the impact of planet bearing fault and the support stiffness on the planetary gear system. Li et al. [9] established a multi-body dynamic (MBD) model of a healthy planetary gear system considering the flexible foundation of the ring gear. Hu and Wei et al. [10, 11] adopted the principle of least beat system to design the speed ratio tracking control algorithm. They achieved a better control effect of stepless transmission. Zhu et al. [12] designed an N-class infinitesimal transmission (IVT) that could provide a continuous output-input ratio from zero to a certain value by using a pair of non-circular gears and a directional control system using planetary gear sets. Hoeijmakers et al. [13] proposed an electromechanical converter with two mechanical ports and one electrical port that could be used as a continuously variable transmission between the mechanical ports. The device replaces the clutch, gearbox, generator and starter motor in the automobile. Xiong et al. [14] proposed a planetary bevel gear CVT considering the state of elasto-hydrodynamic lubrication. They analyzed the oil film thickness of each elliptical contact surface.

In this paper, a planetary gear stepless transmission device is presented. The roller contacting with the orbit is made of rubber material. The input torque of the input shaft is transferred between the input shaft and the output shaft through the centrifugal rotor and the roller. The transfer efficiency is related to the contact force between the roller and the track disc, so the contact force of the roller is analyzed. The influences of springs installed between the centrifugal rotors on the roller contact force are analyzed in time domain and frequency domain.

2 Dynamic Modeling

Figure 1 plots the structure diagram of the planetary centrifugal stepless transmission. The model includes four planetary gears and a sun gear. The output shaft is fixedly connected to the frame. The planetary gear is composed of a planetary gear and a planetary gear shaft. The planetary gear shaft is connected with the prismatic guide groove drive disk through the special design. The rotate of the planetary wheel drives the rotate of the centrifugal rotor. The rollers are attached to and rotate around the rotors. The outer end of the planetary gear shaft is provided with a prismatic guide groove drive disk for driving the centrifugal rotor to rotate. The centrifugal rotor, under the constraint of planetary gear drive and orbit, makes a compound curve motion of rotating and moving. The centrifugal force of the centrifugal rotor contacts the curve orbit in the frame and produces a force to push the frame to rotate and produces the output torque. The dynamic performance and accuracy of the vari-speed drive are affected by the contact force between the centrifugal rotors and frame. The geometric parameters of the gears of Planetary centrifugal stepless transmission are listed in Table 1.

Although the vari-speed drive does not include the ring gear, the motion state of the main parts is still consistent with the characteristics of the planetary gear system in the whole motion process. Thus, it satisfies the following equation

$$\omega_p = (\omega_s - \omega_c) \frac{Z_s}{Z_p} \tag{1}$$

where ω_p represents the angular velocity of the planetary gears, which is the angular velocity of the rotors; ω_s represents the angular velocity of the sun gear; ω_c represents the angular velocity of the carrier; Z_s represents the tooth number of the sun gear, and Z_p represents the tooth number of the planetary gears.

Fig. 1 Planetary centrifugal stepless transmission structure diagram

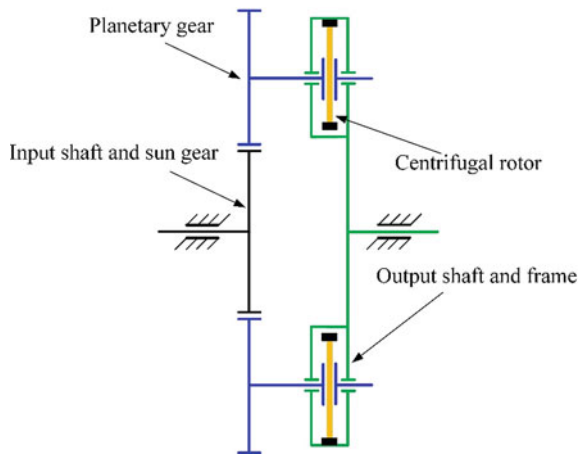


Table 1 Gear parameters of the vari-speed drive

Parameters	Sun gear	Planetary gear
Gear type	Spur gear	Spur gear
Material	Steel	Steel
Modulus of elasticity, E	210 Gpa	210 Gpa
Poisson's ratio, ν	0.32	0.32
Face width	0.015 m	0.020 m
Module	2 mm	2 mm
Number of teeth	19	48
Pressure angle	20°	20°
Theoretical contact ratio	1.59	1.59
Theoretical angle of meshing cycle	24.912	24.912
Addendum	1.00 m	1.00 m
Dedendum	1.25 m	1.25 m

The governing equation of each parts of the vari-speed drive model can be written as

$$m\ddot{q} + (c_s + c_c)\dot{q} + (k_s + k_c)q = T + F \tag{2}$$

where m is the mass of the parts of the vari-speed drive. k_s and k_c are the structure stiffness and the contact stiffness for the vari-speed drive. c_s and c_c are the structure damping and the contact damping for parts of the vari-speed drive, respectively. q is the displacement. T is the internal and external. F is the internal excitation force including the contact force of the rollers. These parameters are integrated over the individual vectors of each parts of the vari-speed model.

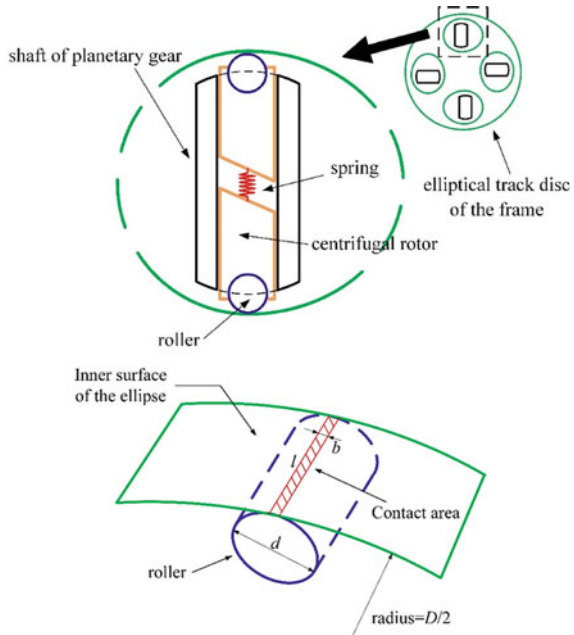
The centrifugal rotor is composed of two centrifugal rotor bodies and rollers. There is a line contact between the roller and the frame. While the vari-speed drive is working, the roller and the inner curve track contact, which is used to achieve force conduction. The contact between the roller and the elliptical track disc of the frame is shown in Fig. 2. The contact stiffness between the cylindrical roller and the elliptical track disk is the Hertz contact stiffness. The contact form is Hertz elastic contact with a very small contact area. In order to simplify the model, the influence of the contact friction between the roller and the frame is ignored. The algorithm of contact stiffness is shown in Eqs. (1)–(9) [15–18].

b is the width size of the contact surface, which is given by [8]:

$$b = \left[\frac{4Q}{\pi l \sum \rho} \left(\frac{(1 - 1/m_1^2)}{E_1} + \frac{(1 - 1/m_2^2)}{E_2} \right) \right]^{\frac{1}{3}} \tag{3}$$

$$\sum \rho = \rho_1 + \rho_2$$

Fig. 2 The contact between the roller and the elliptical track disc of the frame



ρ_1 and ρ_2 are the densities of rollers and orbitals respectively.

The maximum contact stress σ_{max} is [8]:

$$\sigma_{max} = \frac{2Q}{\pi lb} \tag{4}$$

where Q represents the radial load acting on the roller. E_1 and E_2 represent the elastic modulus (MPa) of roller and frame. $1/m_1$ and $1/m_2$ represent the Poisson's ratio of roller and track disc. l is the equivalent contact length of rollers.

The contact deformation of ideal line contact δ is [8]:

$$\delta = \frac{2Q(1 - 1/m^2)}{\pi El} \ln \left[\frac{\pi El^2}{Q(1 - 1/m^2)(1 \mp \gamma)} \right] \tag{5}$$

where γ represents the ratio of rotor diameter d to track diameter D , where the inner contact is— and the outer contact is +. The empirical formula of contact deformation is as follows [8]:

$$\delta = 1.36 \frac{(\eta Q)^{0.9}}{l^{0.8}} \tag{6}$$

According to Eq. (5), Q can be expressed as

$$Q=0.71 \frac{l^{8/9}}{\eta} \delta^{10/9} \tag{7}$$

According to Hertz contact theory, two contact objects will bear the load, and δ is [8]:

$$\delta = \frac{3Q(\Theta_1 + \Theta_2)}{8\pi a} K \tag{8}$$

$$\Theta = \frac{4(1 - 1/m^2)}{E} \tag{9}$$

Based on the above two formulas, it can be known that the contact stiffness between the roller and the elliptical orbit can be expressed as [8]:

$$K = 0.71 \frac{l^{8/9}}{\eta} (\text{N/mm}^{10/9}) \tag{10}$$

$$\eta = \frac{1 - 1/m_1^2}{E_1} + \frac{1 - 1/m_2^2}{E_2} \tag{11}$$

The multi-body dynamics model of vari-speed drive was established by ADAMS, which is shown in Fig. 3.

Fig. 3 The multi-body dynamics model of the vari-speed drive

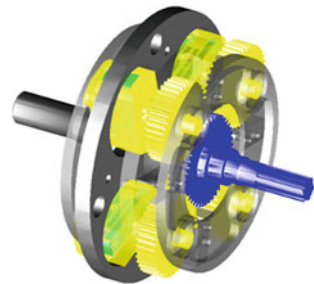
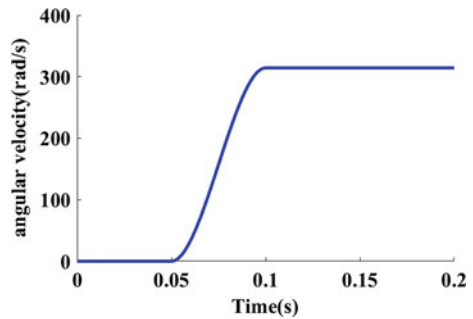


Fig. 4 The Step function curve of input rotate speed



The input shaft is driven. The output shaft rotates at a uniform speed. Using the STEP function to prevent excessive instantaneous contact force caused by sudden change of input speed. The expression of the STEP function is as follows

$$\text{STEP}(\text{time}, t_0, n_0, t_1, n_1) \quad (12)$$

where t_0 is the moment when the initial time of the shaft rotating speed. n_0 is the initial velocity, which is usually zero. t_1 is the moment when the shaft reaches the final rotate speed and rotates at the final rotate speed. n_1 is the final rotate speed (Fig. 4).

In this paper, t_0 is set to be 0.05 s, t_1 is 0.1 s. n_1 is from 0 to 4000 r/min. The input shaft and output shaft are respectively applied in the same direction of rotate speed. The final speed of the input shaft and output shaft are 3000 and 500 r/min. The two cases of adding spring and not adding spring between the rotors are simulated. The stiffness of the spring is 15 N/mm. The initial length is 20 mm. The spring is subjected to the axial pressure, which has an axial preload of 150 N.

As can be seen from the Fig. 5, when the vari-speed drive is working, the two rotors will move in the opposite directions due to the influence of centrifugal force, the rollers contact with the elliptical orbit of the frame. The peak contact force occurs at different times due to the different angles of the rotor's initial position. However, due to the influence of the elasticity of the rollers and the frame, the contact force will appear zero value in the process of motion. At this time, the rollers will be separated from the frame and the output torque cannot be generated. If a compression spring is installed between the two rotors of the same planetary gear shaft, under the action of spring pressure, the contact force between the eight rollers and the frame always exists during the operation of the vari-speed drive. The separation between the rollers and the frame does not occur and the torque is continuously transmitted. When the compression spring is installed, the maximum contact force and the fluctuation range of contact force is reduced, and the transfer efficiency is improved.

3 Results and Discussions

Based on the model of the stepless transmission device and the dynamic analysis method, the model was simulated and analyzed. In order to obtain the relationship between the input rotate speed, the input torque and the output torque, the input rotate speed of the input shaft is set from 500, 1000, 1500, 2000, 2500, 3000, 3500 and 4000 r/min. The output speed is 500 r/min. The spring with the stiffness of 15 N/mm and the preload of 150 N are installed between the centrifugal rotors. The influence of the input rotate speed on the output torque is compared, the maximum (MAX), minimum (MIN), average (MEAN) and mean square values (RMS) of the output torque are calculated at different speeds.

Figure 6 shows the time-domain variation curve of input torque and output torque when the input rotate speed of vari-speed drive is from 500 to 4000 r/min. The

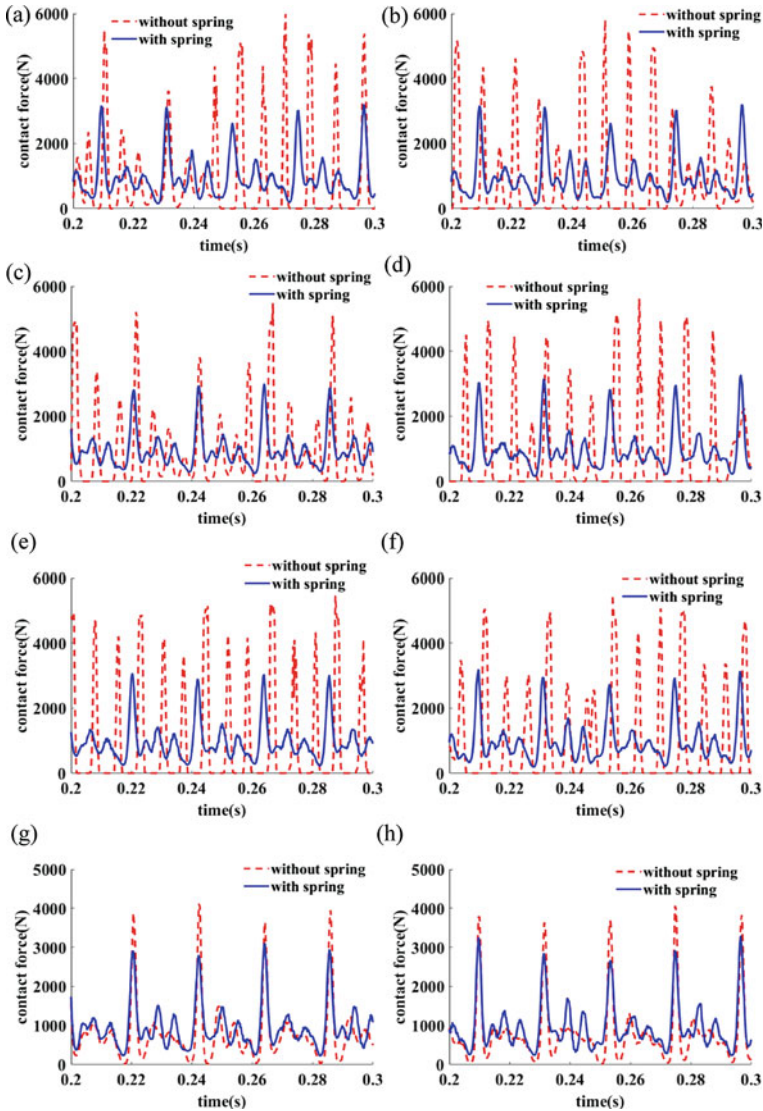


Fig. 5 Contact force between the roller and the elliptical orbit of the frame when Install compression springs between rotors or not. **a–h** are the contact forces for the 8 rollers

changes of torque present periodicity, and when the input speed is larger, the torque period is shorter, the frequency of change is faster, and the impact is more severe. The peak values of output torque and input torque increase with the increase of input speed.

As shown in Figs. 7 and 8, with the increase of the input speed, the MAX, MIN, MEAN and RMS of the input torque and the output torque all show an upward trend.

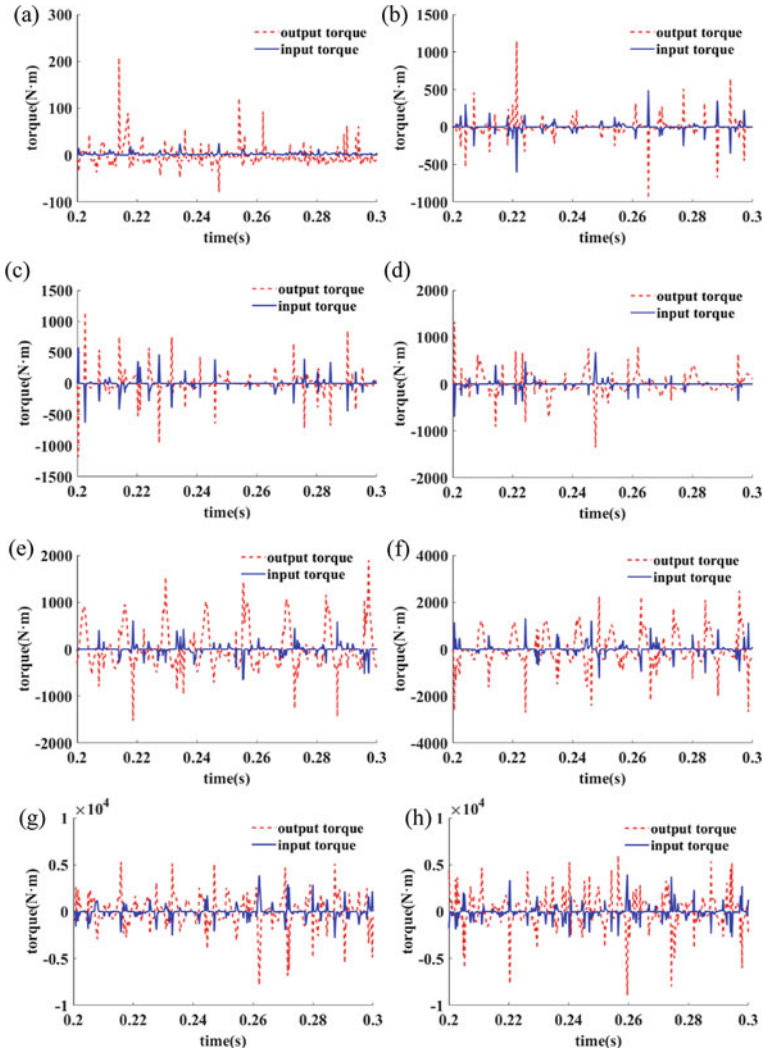


Fig. 6 Time domain variation of input torque and output torque when input rotate speed is 500–4000 r/min and output rotate speed is 500 r/min. **a** 500 r/min, **b** 1000 r/min, **c** 1500 r/min, **d** 2000 r/min, **e** 2500 r/min, **f** 3000 r/min, **g** 3500 r/min, **h** 4000 r/min

The increases of the MAX and MIN values are obviously greater than the MEAN and RMS, which indicates that the higher the input speed can cause the greater the increase of the contact force between the rollers and the frame; With the increase of the input speed, under the condition that the overall mass remains unchanged, the kinetic energy of the system increases, and the work done by the torque per unit time increases, so the input torque increases and a more obvious impact will be caused too. It is not conducive to the improvement of the transfer efficiency. The compression

Fig. 7 The statistical parameters of the input torque change with the input speed

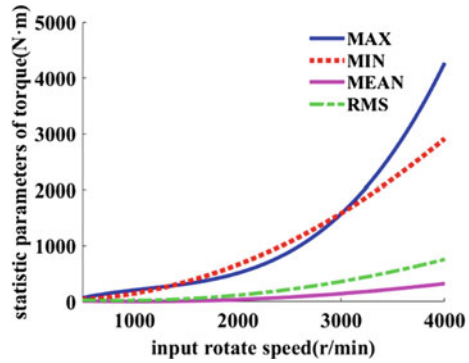
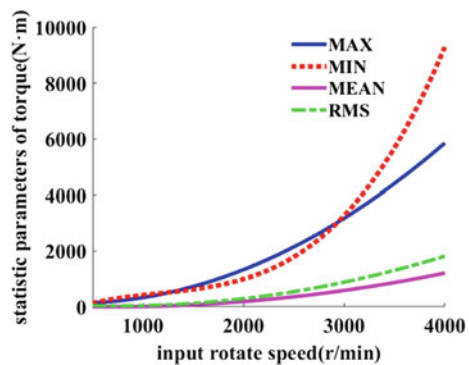


Fig. 8 The statistical parameters of the output torque change with the input speed



spring with the greater stiffness can be replaced when the input rotating speed is higher. At the same input rotate speed, the input torque is less than the output torque.

4 Conclusions

This paper presents a new type of stepless transmission device. According to the structure of the device, the output speed and torque of the planetary mechanism can be controlled; and the self-adaptive stepless change of the transmission device can be realized in principle. A spring is added into the stepless transmission device to make the rollers and the track always contact. Through the simulation analysis of the model, the following conclusions are obtained:

- (1) During the working process of the vari-speed drive, the roller contacts with the frame due to the influence of the centrifugal force. When the contact force is 0, the roller is separated from the frame. An appropriate spring installed between the centrifugal rotor can reduce the impact between the frame and the roller, which makes it constantly be in contact.

- (2) With the increase of input rotate speed, both the output torque and input torque have a tendency to increase. The increases of maximum and minimum values are greater than those of average and mean square values. It indicates that the higher the speed can cause the greater impact.

Funding This study was funded by the National Natural Science Foundation of China (No. 51975068 and 52175120) and the Fundamental Research Funds for the Central Universities (No. 3102020HHZY030001).

References

1. Blagonravov, A.A., Yurkevich, A.A., Yurkevich, A.V.: Automatic control of a stepless mechanical transmission with an internal force function. *Russ. Eng. Res.* **37**(3), 185–188 (2017)
2. Zhenkun, D., et al.: Mechanical structure analysis and kinematic simulation of the satellite star gear continuously variable transmission system. SAE International, Shanghai, China (2008)
3. Milazzo, M., Moretti, G., Burchianti, A., Mazzini, D., Oddo, C.M., Stefanini, C., Fontana, M.: A passively regulated full-toroidal continuously variable transmission. *Meccanica* **55**(1), 211–226 (2020)
4. Lazarek, M., Brzeski, P., Perlikowski, P.: Design and modeling of the CVT for adjustable inerter. *J. Franklin Inst.* **356**(14), 7611–7625 (2019)
5. Gang, L., Weidong, Z.: Design and power loss evaluation of a noncircular gear pair for an infinitely variable transmission. *Mech. Mach. Theor.* **156**, 104137 (2021)
6. Ferguson, R., Daws, L., Kerr, J.: The design of a stepless transmission using noncircular gears. *Mech. Mach. Theory* **10**(6), 467–478 (1975)
7. Liu, J., Pang, R., Ding, S., et al.: Vibration analysis of a planetary gear with the flexible ring and planet bearing fault. *Measurement* **165**, 108100 (2020)
8. Liu, J., Pang, R., Li, H., et al.: Influence of support stiffness on vibrations of a planet gear system considering ring with flexible support. *J. Central South Univer.* **20**(8), 2280–2290 (2020)
9. Li, H., Liu, J., Ma, J.J., et al.: Effect of the radial support stiffness of the ring gear on the vibrations for a planetary gear system. *J. Low Freq. Noise, Vib. Active Control* 1461348419844642 (2019)
10. Hu, J.B., Wei, C., Dual, J.Y.: A study on the speed ratio follow-up control system of hydro-mechanical transmission. *Trans. Beijing Inst. Technol.* **28**(6), 15–19 (2008)
11. Wei, C., Yuan, S.H., Hu, J.B., et al.: Theoretical and experimental investigation of speed ratio follow-up control system on geometric type hydro-mechanical transmission. *J. Mech. Eng.* **47**(16), 101–105 (2011)
12. Zhu, W.D., Wang, X.F.: Design, modeling, and simulation of a geared infinitely variable transmission. *J. Mech. Des.* **136**(7), 071011 (2014)
13. Hoeijmakers, M.J., Ferreira, J.A.: The electric variable transmission. *IEEE Trans. Ind. Appl* **42**(4), 1092–1100 (2006)
14. Xiong, B.S., Chen, J.Y., Wu, X.L., et al.: Research of planetary stepless transmission with ring cones for exciter of vibrational fire grate. Science Press Beijing, 16 Donghuangchenggen North St, Beijing 100707, Peoples R China **26**(30), 319–324 (2006)
15. Liu, J.: A dynamic modelling method of a rotor-roller bearing-housing system with a localized fault including the additional excitation zone. *J. Sound Vib.* **469**, 115144 (2020)
16. Liu, J., Xu, Y., Pan, G.: A combined acoustic and dynamic model of a defective ball bearing. *J. Sound Vib.* **501**, 116029 (2021)

17. Liu, J., Tang, C., Wu, H., Xu, Z., Wang, L.: An analytical calculation method of the load distribution and stiffness of an angular contact ball bearing. *Mech. Mach. Theory* **142**, 103597 (2019)
18. Liu, J., Wang, L., Shi, Z.: Dynamic modelling of the defect extension and appearance in a cylindrical roller bearing. *Mech. Syst. Signal. Pr.* **173**, 109040 (2022)

The Fatigue Failure Prediction of a Vari-Speed Drive with Different Rollers



Jing Liu, Ximing Zhang, Jingtao Shang, Jin Li, and Shizhao Ding

Abstract The track is the main failure part of vari-speed drive. The failure damage of the track disc will generate huge vibrations. The fatigue fracture became a problem during the normal working process of these machines. However, the steel roller has a higher deformation resistance capacity. The large contact stress of the steel roller will generate larger amplitude of cyclic stress and decrease the life of vari-speed drive. In order to improve the reliability of the vari-speed drive and increase the fatigue life, the rubber roller is used to replace the steel roller. A 3D model of the vari-speed drive is established by using Solidworks. The contact analysis of track is calculated by using Adams. The dynamic stress is analyzed by using the rain-flow counting method, which can determine the amplitudes and mean values of counted cycles. According to the assumption of a linear Palmgren–Miner hypothesis of damage accumulation and typical fatigue characteristics of the material, the fatigue failure life of the roller is calculated. The results show that the contact force of the steel roller is more than that of the rubber roller. The contact stress of the steel roller is much larger than the rubber roller. The fatigue life of the steel roller is less than that of the rubber roller. This paper presents a new method for solving the fatigue failure of the vari-speed drive.

Keywords Fatigue failure · Rain-flow counting method · Vari-speed drive · Life prediction

J. Liu

School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, P.R. China

e-mail: jliu0922@nwpu.edu.cn

Key Laboratory for Unmanned Underwater Vehicle, Northwestern Polytechnical University, Xi'an 710072, P. R. China

X. Zhang · J. Shang · J. Li (✉)

China North Vehicle Research Institute, Beijing 100081, P. R. China

e-mail: lijin929@126.com

S. Ding

College of Mechanical Engineering and Vehicle, Chongqing University, Chongqing, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

H. Zhang et al. (eds.), *Proceedings of IncoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_56

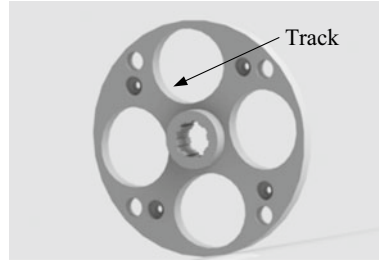
1 Introduction

The track is the main failure part of vari-speed drive. The time-varying contact force will generate the time-varying contact stress, which can result in the fatigue failure of the track during the working processing of vari-speed drive. Although the vari-speed drive is generally designed with a high safety margin in order to avoid the unacceptable fatigue strength of the material, the large number and high amplitude cyclic loading and local stress concentrations allow cracks to grow even when the fatigue strength exceeds the average load value.

A significant amount of research works has been published on the vari-speed drive. Guo et al. [1] et al. established a simulation model of a hydraulic machinery step-less transmission based on the AMESim. They studied the effects of major parameters on the hydraulic machinery step-less transmission. Li and Wei [2] studied the power machine step-less transmission on the electric vehicle. Xia and Sun [3] proposed a new continuously variable transmission system. Li and Zhu [4] proposed a novel noncircular gear design method based on a modified high-order elliptical pitch curve for an infinitely variable transmission and power loss evaluation of a noncircular gear pair is investigated. Li et al. [5] proposed a novel configuration factor table method for the configuration synthesis of basic no-spin traction continuously variable transmission mechanisms. Olyaei et al. [6] presented a new ratcheting continuously variable transmission mechanism. Qian and Wang [7] analyzed the fatigue life of the vehicle body according to the Miner damage theory. Shinde et al. [8] et al. proposed a new Microsoft excel-based algorithm for the modified graphical rain-flow counting method. Ding and Chen et al. [9] presented nonlinear fatigue damage for existing lifting equipment.

Most research works on the fatigue failure of vari-speed drives were limited to simulation analysis and structural design. The deformation of rubber roller is larger than steel roller, which can decrease the value of contact stress. The steel roller with large contact stress can be replaced by rubber roller. This paper is focused on using the rubber roller instead of the steel roller to decrease the time-varying contact stress and increase the fatigue life of vari-speed drive. A 3D model of the vari-speed drive is established by using Solidworks. The contact analysis of track is calculated by using Adams. The dynamic stress is analyzed by using the rain-flow counting method, which can determine the amplitudes and mean values of counted cycles. According to the assumption of a linear Palmgren–Miner hypothesis of damage accumulation and typical fatigue characteristics of the material, the fatigue failure life of the rollers is calculated. This paper presents a new method for increasing the fatigue failure of vari-speed drive.

Fig. 1 Structure diagram of the track



2 Failure Description of Vari-Speed Drive

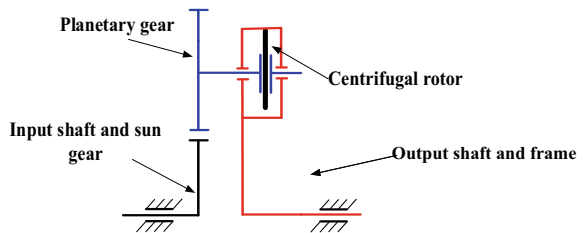
The motion and torque of vari-speed drive are transmitted by the planetary axle and rotor. The vari-speed drive is important for the machine-driven system, which is used in rolling mills, ships, etc. The gearing inputs the torque by the input axis. The torque between the input axis and output axis is transmitted by the rotor and roller. The transmission efficiency relates to the time-varying contact force between the roller and track. Hence, the changes of time-varying contact force greatly affect the fatigue life of vari-speed drive. In reality, the fracture position of vari-speed drive is given in Fig. 1, which can be considered as the failure type of high cyclic stress.

3 Simulation Method Descriptions

3.1 A 3D Model

The structural diagram of stepless transmission is given in Fig. 2. According to the structural diagram, A 3D model of the vari-speed drive is established by using Solidworks and the contact analysis of track is calculated by using the Adams.

Fig. 2 Structural diagram of vari-speed drive



3.2 Contact Stress

The non-idea Hertzian contact calculation program is given in Fig. 3, in which a_{0j} is initial the half contact width; n is the total number of elements in the contact area as given in Fig. 3; Q_1 is the applied load; and ϵ is the calculation accuracy.

In this method, the maximal stress of the contact area is given as

$$p_{\max} = \sqrt{E'Q(L\pi R_z)} \tag{1}$$

The equivalent radius is described as

$$R_z = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}} \tag{2}$$

The equivalent elastic module is defined as

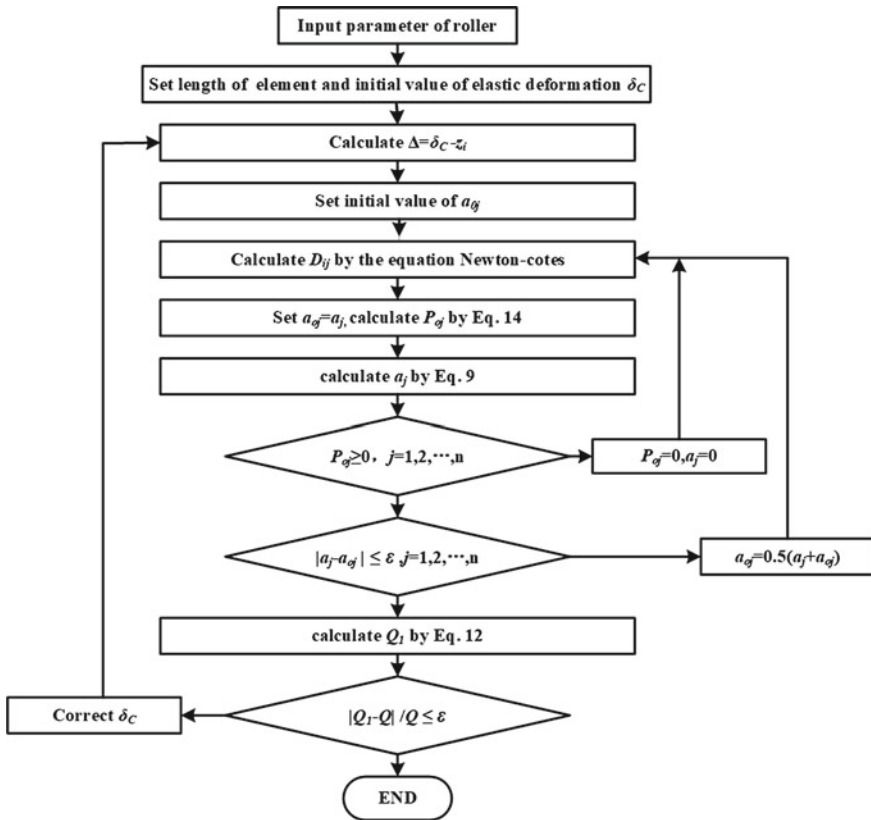


Fig. 3 The non-idea Hertzian contact calculation program

$$E' = 1 / \left(\frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \right) \tag{3}$$

where R_1 and R_2 is the contact body radius; E_1 and E_2 is the contact body elastic modulus; ν_1 and ν_2 is the Poisson's ratio; and the half contact width is described as

$$a = \sqrt{4R_z Q (L\pi E')} \tag{4}$$

where L is length of roller. The contact stress of an arbitrary point in the width of the contact area ($-b < x < b$) is defined as

$$p = p_{\max} \sqrt{1 - \left(\frac{x}{a}\right)^2} \tag{5}$$

Then, the calculated load Q is given as

$$4 \sum_{j=1}^{m \times n} a_j b_j p_j = Q \tag{6}$$

$$\frac{1}{\pi E'} \sum_{j=1}^{m \times n} D_{ij} p_j = \delta_C - z_i(x_i, y_i) \quad i = 1, 2, \dots, n \tag{7}$$

where, D_{ij} is integral of flexibility coefficient. The contact stress distributed in the x direction is given as

$$p_j = p_{0j} \sqrt{1 - (x'/a_j)^2} \tag{8}$$

where p_{0j} is the maximal contact stress. Moreover, its half contact width is described as

$$a_j = 2R_z p_{0j} / E' \tag{9}$$

The displacement generated by the stress of element j along the center of element i is defined as

$$w_{ij} = \frac{p_{0j}}{\pi E'} \int_{-a_j}^{a_j} \int_{x_j - h_j}^{x_j + h_j} \frac{\sqrt{1 - (x'/a_j)^2}}{\sqrt{x'^2 + (y_i - x_j - y')^2}} = \frac{p_{0j}}{\pi E'} D_{ij} \tag{10}$$

where the flexibility coefficient D_{ij} is given as

$$D_{ij} = \int_{-a_j}^{a_j} \int_{x_j-h_j}^{x_j+h_j} \frac{\sqrt{1 - (x'/a_j)^2}}{\sqrt{x'^2 + (y_i - x_j - y')^2}} dx' dy' \tag{11}$$

The integral of flexibility coefficient D_{ij} in the y' direction is described as

$$D_{ij} = \int_{-a_j}^{a_j} \sqrt{1 - (x'/a_j)^2} \ln \left(\frac{|y_i - y_j| + h_j + \sqrt{x'^2 + (y_i - y_j + h_j)^2}}{|y_i - y_j| - h_j + \sqrt{x'^2 + (y_i - y_j - h_j)^2}} \right) dx' \tag{12}$$

where h is width of slice. Then, more simplifications for Eqs. (6) and (7) are given as

$$4 \sum_{j=1}^{m \times n} a_j b_j p_{oj} = Q \tag{13}$$

$$\frac{1}{\pi E'} \sum_{j=1}^{m \times n} D_{ij} p_{oj} = \delta_C - z_i(y_i) \quad i = 1, 2, \dots, n \tag{14}$$

Because the contact stress cannot be negative, the nonnegative condition should be

$$p_{oj} \geq 0 \quad j = 1, 2, \dots, n \tag{15}$$

3.3 Result Analysis

According to the 3D model in Sect. 3.1, the initial load is 50 N. The contact force between the steel roller/rubber roller and the rollaway nest is calculated. Then, the contact stress between the steel roller/rubber roller and the rollaway nest is calculated. Figures 4 and 5 are the contact force/stress of the steel/rubber roller distribution diagram. The results show that when the operation is stable, the contact force of the rubber roller has a lot of volatility, which is periodic. The contact force of steel has a lot of volatility. The maximum value of the contact force of the steel roller is greater than that of the rubber roller. The deformation of the rubber roller is more than that of the steel roller. Thus, the contact stress of steel roller is more than rubber roller.

Fig. 4 The contact force and contact stress of the rubber roller. **a** Contact force, and **b** Contact stress

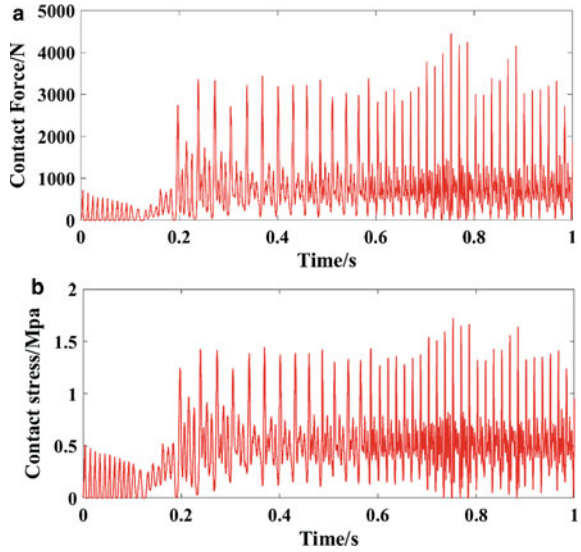
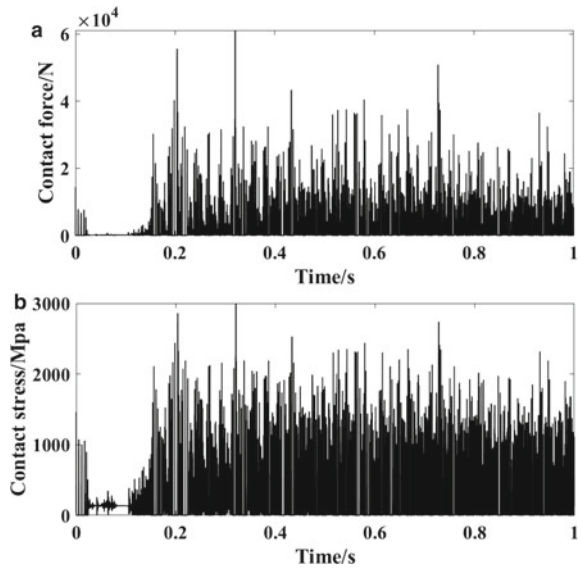


Fig. 5 The contact force and contact stress of steel roller. **a** Contact force, and **b** Contact stress



4 Fatigue Life Prediction

Based on the assumption of a linear Palmgren–Miner hypothesis of damage accumulation, by using typical fatigue characteristics of the material and contact force, the fatigue life of the roller will be calculated.

4.1 Rain Flow Counting Method

The time-varying contact stress of track of vari-speed drive is analyzed by using the rain flow counting method [10], which can be used to determine the amplitudes and mean values of counted cycles. The rain-flow matrices determined for the stress histories of the track are shown in Figs. 6 and 7. The results show that the stress distribution of the rubber roller likes normal distribution. Most of the stress is distributed between 0 and 200 Mpa.

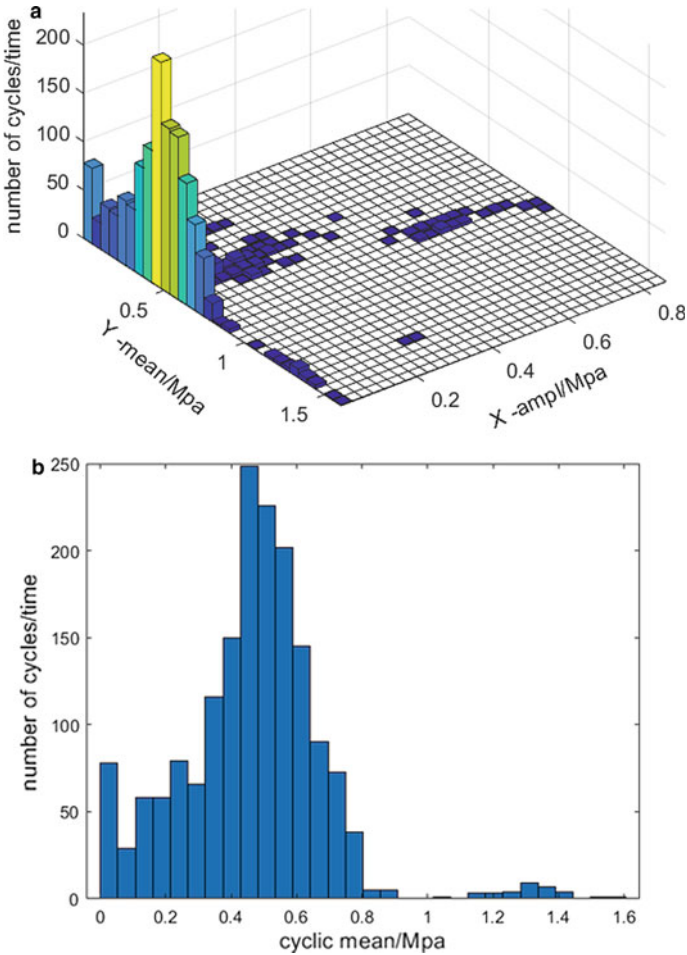


Fig. 6 The rain-flow matrices of the rubber roller. **a** The rain-flow stress matrices, and **b** The stress cycle

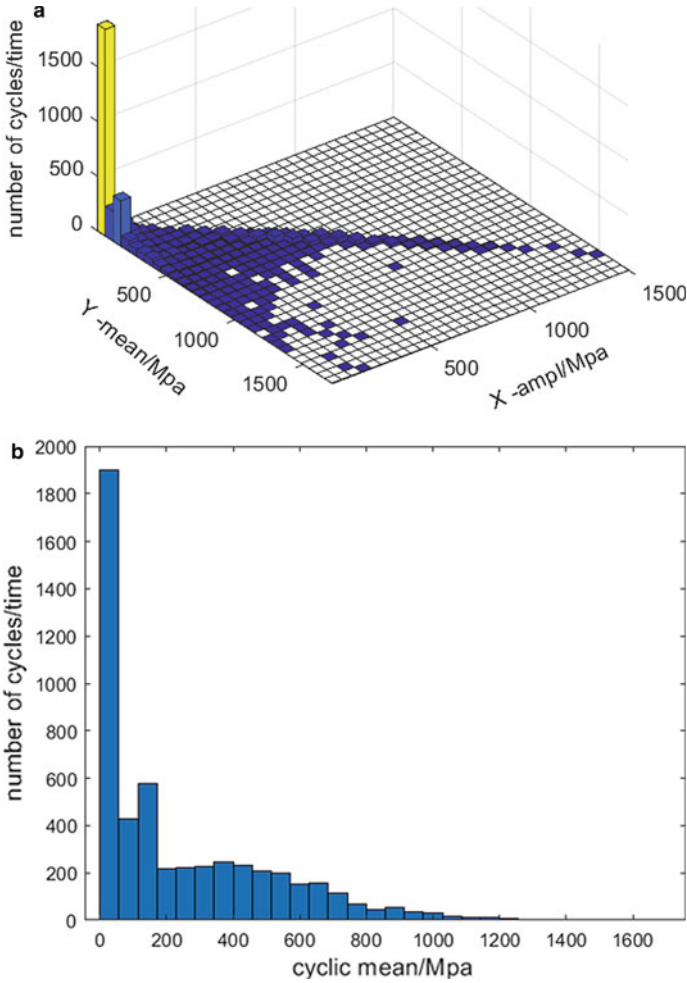


Fig. 7 The rain-flow matrices of rubber roller. **a** The rain-flow stress matrices, and **b** The stress cycle

4.2 Fatigue Life

Using Miner’s theory, the particular values of damage D_i for each cycle or half-cycle are defined as [11–14]

$$D_i = \frac{n_i}{N_i} \text{ for } i = 1, 2, 3, \dots, k \tag{16}$$

where n_i is equal to 1 for a cycle and 0.5 for a half-cycle, N_i the number of cycles to material failure computed from a uniaxial fatigue characteristic, and k is the number

of cycles and half-cycles determined from the history using cycle a counting algorithm. When the mean value is considered, the fatigue failure life can be calculated using the Morrow stress-life characteristic.

$$N_i = 0.5 * \left(\frac{\sigma_{ai}}{\sigma'_f - \sigma_{mi}} \right)^{\frac{1}{b}} \quad (17)$$

where σ_{ai} the cycle amplitude determined from the stress history on the housing using the rain-flow algorithm, σ_{mi} the cycle means value determined from the stress history of the housing using the rain-flow algorithm, σ'_f is the fatigue strength coefficient, and b is the fatigue strength exponent. According to the Palmgen–Miner hypothesis, the total damage can be expressed as the sum of all the particular damage caused by distinct individual cycles [15]. It is given as

$$D = \sum_{i=1}^k D_i \quad (18)$$

where k is the number of cycles determined from the stress history using the cycle-counting algorithm. Then the expected fatigue life of the housing, T can be calculated from the accumulated damage, D caused by each distinct cycle in the time interval T_0 of the stress history.

$$T = \frac{T_0}{D} \quad (19)$$

The fatigue life of the track is calculated according to the Morrow stress-life characteristic and the hypothesis of fatigue damage accumulation. The results show that the fatigue life of the rubber roller is more than that of the steel roller (Tables 1 and 2).

5 Conclusions

In order to improve the reliability of vari-speed drive and increase life, the rubber roller is used to replace the steel roller. A 3D model of the vari-speed drive is established by using Solidworks. The contact analysis of track is calculated by using Adams. The dynamic stress is analyzed by using the rain-flow counting method. According to the assumption of a linear Palmgren–Miner hypothesis of damage accumulation, the fatigue failure life was calculated. When the operation is stable, the contact force of rubber roller has a lot of volatility, which is periodic. The contact force of steel has a lot of volatility. The maximum value of contact force of steel roller is greater than that of the rubber roller. The deformation of rubber roller is more than that of the steel roller. Thus, the contact stress of steel roller is more than rubber roller. The stress distribution of rubber roller likes normal distribution. Most

Table 1 The major parameter of steel roller and rubber roller

Type of roller		
Parameter	Steel roller	Rubber roller
Fatigue strength coefficient (Mpa)	450	100
Fatigue strength exponent	-0.14	-0.32

Table 2 The comparison of life between steel roller and rubber roller

Type of roller		
Life	Steel roller	Rubber roller
Time(h)	161	488

of the stress is distributed between 0 and 200 Mpa. The fatigue life of rubber roller is more than that of the steel roller.

Funding This study was funded by the National Natural Science Foundation of China (No. 51975068) and the Fundamental Research Funds for the Central Universities (No. 3102020HHZY030001).

References

- Guo, Z., Yuan, S., Jing, C., et al.: Modeling and simulation of shifting process in hydraulic machinery step-less transmission based on AMESim. *Trans. Chin. Soc. Agric. Eng.* **25**(10), 86–91 (2009)
- Li, K., Wen, R.: Control strategy of electric machinery step-less transmission system. *Int. J. Mechatron. Appl. Mech.* **4**, 158–165 (2018)
- Xia, Y., Sun, D.: Characteristic analysis on a new hydro-mechanical continuously variable transmission system. *Mech. Mach. Theory* **126**, 457–467 (2018)
- Li, G., Zhu, W.: Design and power loss evaluation of a noncircular gear pair for an infinitely variable transmission. *Mech. Mach. Theory* **156**, 104137 (2021)
- Li, C., Yao, J., Li, H., et al.: Research on a novel configuration synthesis method of no-spin traction continuously variable transmission. *Mech. Mach. Theory* **152**, 103963 (2020)
- Olyaei, A.: Novel continuously variable transmission mechanism. *SN Appl. Sci.* **1**(9), 1–7 (2019)
- Qian, K., Wang, Y.: Research on analysis technology for fatigue life of subway car body based on actual working conditions. *Mach. Build. Autom.* **49**(06), 98–100 (2020)
- Shinde, V., Jha, J., Tewari, A., et al.: Modified rainflow counting algorithm for fatigue life calculation. In: *Proceedings of Fatigue, Durability and Fracture Mechanics*, pp. 381–387. Springer, Singapore (2018)
- Ding, K., Chen, L.: Research on fatigue life of lifting equipment based on nonlinear cumulative damage theory. *Mater. Res. Proc.* **18**
- Niesłony, A.: Determination of fragments of multiaxial service loading strongly influencing the fatigue of machine components. *Mech. Syst. Signal Process.* **23**(8), 2712–2721 (2009)
- Shao, Y., Liu, J., Mechefske, C.K.: Drive axle housing failure analysis of a mining dump truck based on the load spectrum. *Eng. Fail. Anal.* **18**(3), 1049–1057 (2011)
- Liu, J.: A dynamic modelling method of a rotor-roller bearing-housing system with a localized fault including the additional excitation zone. *J. Sound Vib.* **469**, 115144 (2020)

13. Liu, J., Xu, Y., Pan, G.: A combined acoustic and dynamic model of a defective ball bearing. *J. Sound Vib.* **501**, 116029 (2021)
14. Liu, J., Tang, C., Wu, H., Xu, Z., Wang, L.: An analytical calculation method of the load distribution and stiffness of an angular contact ball bearing. *Mech. Mach. Theory* **142**, 103597 (2019)
15. Hu, Y.: Basic theory and CAE analysis technology of vehicle structural strength. Chongqing University Press (2009)

State of Health Estimation of Lithium-Ion Batteries from Charging Data: A Machine Learning Method



Zuolu Wang, Guojin Feng, Dong Zhen, Fengshou Gu, and Andrew D. Ball

Abstract Accurate state of health (SOH) estimation of the lithium-ion battery plays an important role in ensuring the reliability and safety of the battery management system (BMS). The data-driven method based on the selection of degradation features can be effectively applied to SOH estimation. In practice, lithium batteries often work in complex discharge conditions, but they are charged under constant current (CC) conditions. Therefore, the suitable degradation features of the battery are extracted in this work for accurate SOH estimation. First, the degradation features are summarized and extracted from the CC charging data. Second, the Pearson correlation coefficient is utilized to quantify the relationship between the extracted degradation features and the battery SOH, thus determining the most influential degradation feature. Finally, the long short term memory (LSTM) is used for model training and SOH estimation based on the selected feature. The results show that LSTM model can give reliable and accurate SOH estimation with R^2 of 1 and lower mean absolute error (MAE) and maximum error (MAX).

Keywords Lithium-ion battery · State of health · Feature extraction · Long short term memory

Z. Wang · G. Feng (✉) · F. Gu · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK
e-mail: Guojin.Feng@outlook.com

Z. Wang
e-mail: Zuolu.Wang@hud.ac.uk

D. Zhen
Tianjin Key Laboratory of Power Transmission and Safety Technology for New Energy Vehicles,
School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, P. R. China

1 Introduction

Lithium-ion batteries have been the important energy source of electric vehicles (EVs) and energy storage systems owing to their advantages in terms of high efficiency, high energy density, and long-life cycle [1]. However, their performance gradually deteriorates with the increase of cycling. This degradation is a gradual process, which can reduce the operating efficiency of the electrical systems and potentially cause catastrophic failures [2]. Moreover, the battery is a dynamic and nonlinear electrochemical system, which makes the state of health (SOH) estimation more difficult in practice. Therefore, it is a demanding and challenging task to develop a reliable SOH estimation technique for the battery management system (BMS) to monitor the real-time performance of the battery during the operation.

Many methods have been proposed for the SOH estimation of the lithium-ion battery and they can be generally divided into three categories, including direct measurement methods, model-based methods, and data-driven methods [3]. The direct measurement method can be easily applied to practice based on the measurement of some degradation features such as degraded capacity and internal resistance [4]. However, this kind of method is more suitable for off-line monitoring to check the remaining useful life. And the errors from the measurement systems can lead to the inaccurate SOH estimation. The model-based methods include empirical model [5], electrochemical model [6], and electrical circuit model [7]. The empirical model is developed based on the mathematical statistics so that it only gives a rough estimation of the battery life. And the methods based on the electrochemical model and the electrical circuit model need to consider the electrochemical mechanism of the battery. Although these models can be applied to the real-time SOH estimation of the battery, they present higher complexity, and the estimation accuracy is prone to the model accuracy. In contrast, the data-driven method has high data processing capabilities and therefore has been a research hotspot for battery SOH estimation [8]. In addition, it does not need to consider the complex degradation mechanism of the battery and can effectively learn and map the nonlinear relationship between the degradation features and the battery SOH. In recent years, many data-driven methods have been proposed for the lithium-ion battery SOH estimation and the degradation features are mainly extracted from two processes, including the discharging process and charging process.

Based on the discharging process, Pan et al. [9] used the extreme learning machine (ELM) to map the relationship between the battery SOH and a series of health features including terminal voltage, current, temperature, ohmic resistance and polarization resistance. The maximum error and the mean absolute error (MAE) of the SOH estimation are 0.0581 and 0.0238, respectively. In [10], a novel metabolic extreme learning machine (MELM) was proposed for the battery SOH estimation using the incremental values of the ohmic internal resistance, polarized internal resistance, and degradation capacity. The results show that the MAE of the SOH estimation ranges from 0.38 to 0.6%. On the other hand, the degradation features can be extracted in the process of the charging phase. Zhang et al. [11] conducted the battery SOH

prediction based on the ANN model. In this method, five training features were extracted from different voltage regions in dQ/dV curves. However, the obtained MAE of the predicted results are much larger, ranging from 2.79 to 3.52. In [4], a new degradation feature was reconstructed based on the model parameter identification of the selected first-order RC model in the constant voltage charging stage. The MAE of the method ranges from 0.0088 to 0.0146. Based on the Oxford Battery Degradation dataset, Fan et al. [12] proposed the gate recurrent unit-convolutional neural network (GRU-CNN) for battery SOH estimation and the charging voltages in the CC charging mode are selected as the training features. However, it is revealed that the prediction error is high, for example, the overall MAE value of this method is 0.62.

Although the above data-driven methods have achieved excellent SOH estimation of the battery, a large number of features are used for model training and SOH prediction can increase the computational burden. It is still a challenging task to select the effective and few degradation features for reducing the computational complexity and ensuing the estimation accuracy. On the one hand, the degradation feature extraction based on the model parameter identification can increase the computational complexity of BMS. Also, the estimation results are easily affected by the inaccurate model parameter identification. Furthermore, the lithium-ion battery often works under the complex discharging conditions, while it is charged under CC charging mode. Therefore, the battery SOH estimation based on the selection of the degradation features from the charging data can facilitate the practical applications.

This paper summarizes and investigates the degradation features from the CC charging voltage profiles based on the Oxford Degradation dataset [13]. Subsequently, the Pearson correlation coefficient is used to evaluate the correlation between battery capacity and the degradation features, thus selecting the most effective feature for model training. Finally, long short-term memory (LSTM) is used for model training and SOH prediction based on the selected degradation features as it has the advantage of processing the sequence data. It is showed that the estimation values present a good agreement with the measured SOH with the R^2 close to 1.

This paper is organized as follows. Section 2 presents the dataset and the degradation feature extraction. In Sect. 3, the LSTM-based model and three benchmarks are illustrated. The predicted results and discussions are given in Sect. 4. Finally, the conclusions of the whole work are drawn in Sect. 5.

2 Dataset and Degradation Feature Extraction

2.1 Dataset

The data used in this study is the Oxford Degradation dataset [13] that is supported by the battery intelligence lab at the University of Oxford. It involves the degradation tests of 8 Kokam 740 mAh lithium-ion pouch cells (SLPB533459H4). In the process

of the degradation test, all the batteries are placed in a thermal chamber of 40 °C and charged in 1C CC mode until the terminal voltage reaches the nominal voltage of 4.2 V. To be specific, the charging data including voltage, current, cell temperature, and capacity are collected every 100 cycles. Here the 8 tested cells are named Cell1, Cell2, Cell3, Cell4, Cell5, Cell6, Cell7, and Cell8, in which the dataset from Cell1, Cell2, Cell3, Cell5, Cell6, and Cell7 are used for model training and the other two cells (Cell4 and Cell8) are used for model validation.

2.2 Degradation Feature Extraction

In this section, the degradation features are summarized and extracted from the Oxford Degradation dataset. The charging voltages are specifically studied for degradation feature extraction as the charging current and temperature are constant during the test. Figure 1 illustrates the charging voltages and SOH profiles of Cell1 under different cycles. It can be found the charging voltage exhibits differences at different life stages. As the battery degrades, the terminal voltage of the battery increases more slowly, therefore, it requires a longer charging time. Take Cell1 as an example, several degradation features are studied and compared based on the Pearson correlation coefficient to select the most effective one for model training.

The Mean Voltage. The first type of feature extracted from the charging data denotes the calculated mean voltage when the battery charges at the start voltage V_1 and stops charging at the end voltage V_2 . As can be seen from Fig. 2, the charging voltage rises slower with the battery degrades. In different degradation stages, the mean voltage calculated in the range of V_1 - V_2 is different. To extract this feature, the end voltage V_2 is set as 2.8 V and the start voltage V_1 is set as the different values. Table 1 illustrates the correlation coefficients between the mean voltage calculated using the different V_1 and the battery SOH according to the Pearson correlation coefficient. It is noted that the mean voltage has a high correlation with the battery SOH, especially when the V_1 is 2.8 V.

The Reached Voltage When Charging for the Same Duration. The second type of degradation feature denotes the reached voltage when the battery charges for

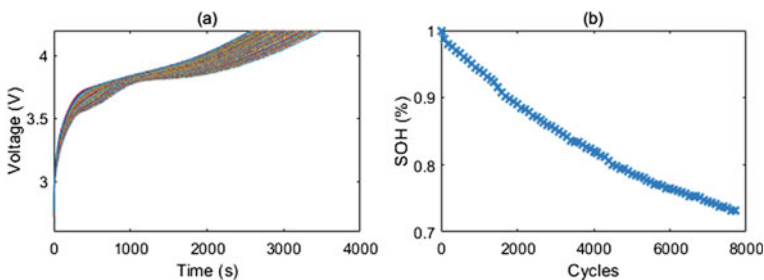


Fig. 1 a Charging voltage. b SOH profiles with time

Fig. 2 The charging voltage under different levels of degradation for the extraction of the first feature

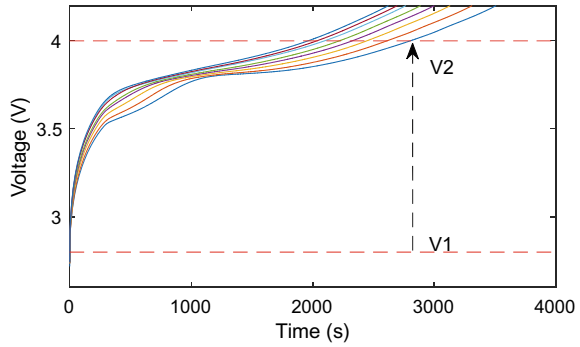


Table 1 The correlation coefficient between the mean voltage from V1 to V2 and SOH

Range for discharge stage (V1–V2)	Correlation coefficient
2.8–4.2	–0.9950
3.0–4.2	–0.9880
3.2–4.2	–0.9880
3.4–4.2	–0.9882
3.6–4.2	–0.9520
3.8–4.2	–0.9898
4.0–4.2	0.0254

the same duration T from the start voltage V1. Figure 3 shows the charging voltage profiles under the different levels of degradation. It can be seen the battery reaches the different terminal voltage such as V3 and V4 when charging for the same time from the start voltage V1. Here the start voltage V1 is set as 2.8 V considering the small time differences under different charging cycles. According to the Pearson correlation coefficient, Table 2 shows the correlation coefficients between the charging duration and the battery SOH. It is found that this type of feature also presents a high correlation with the battery SOH.

Fig. 3 The charging voltage under different levels of degradation for the extraction of the second feature

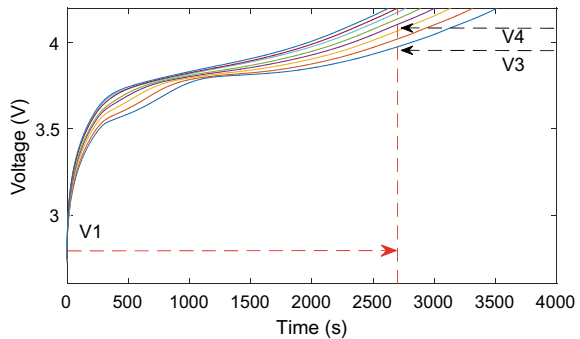


Table 2 The correlation coefficient between the reached voltage when battery charging for same duration from 2.8 V and SOH

Charge duration $T(s)$	Correlation coefficient
130	-0.9863
260	-0.9835
520	-0.9844
780	-0.9693
1040	-0.9870
1300	-0.9799
1560	-0.9899
1820	-0.9932
2080	-0.9925

The Required Charging Time. The third type of degradation feature is the required charging time from the start voltage V_1 to the end voltage V_2 . As shown in Fig. 4, the battery requires different charging times (e.g., $0-t_1$ and $0-t_2$) to reach from V_1 to V_2 under different SOHs. Here, the start voltage V_1 is respectively set as 2.8, 3.0, 3.2, 3.4, 3.6, 3.8 and 4.0 V and V_2 is set as the cut-off voltage (4.2 V) of the battery. Similarly, Table 3 lists the calculated correlation coefficients between the required charging time in different ranges of charging voltages and the battery SOH.

Fig. 4 The charging voltage under different levels of degradation for the extraction of the third feature

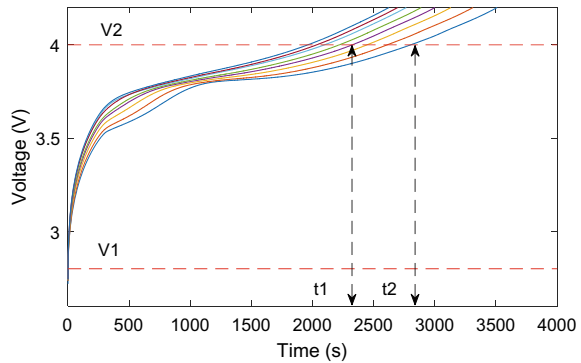


Table 3 The correlation coefficient between the required charging time from V_1 to V_2 and SOH

Range for charge phase (V_1 - V_2)	Correlation coefficient
2.8-4.2	-0.9999
3.0-4.2	-0.9999
3.2-4.2	-0.9999
3.4-4.2	-0.9998
3.6-4.2	-0.9881
3.8-4.2	-0.9971
4.0-4.2	-0.9084

Compared to the previous two types of degradation features mentioned above, this degradation feature presents a higher correlation with the battery SOH. The correlation coefficients are close to -1 when the V1 ranges from 2.8 to 3.2 V. Specifically, these degradation features extracted from the charging stages of 2.8–4.2 V, 3.0–4.2 V, and 3.2–4.2 V are described as f_1 , f_2 , and f_3 .

Considering the higher correlation of these three indicators extracted from the third type of degradation feature, every health feature is separately used for model learning to compare their performance for the battery SOH prediction.

3 Methodology

3.1 Long Short Term Memory

LSTM is an improved version of the traditional recurrent neural network (RNN) model, and it addresses the long-term dependence problem caused by the gradient vanishing and has the advantage of dealing with the time sequence learning task [14]. As result, it is used for SOH estimation in this work. As shown in Fig. 5, the LSTM neural network consists of an output gate, a forget gate, an input gate and some connections among these gates. These gates work together to retain useful information and ignore unimportant information [15]. The learning process can be expressed as the following equations:

$$i_k = \sigma(W_{i,x}x_k + W_{i,h}h_{k-1} + b_i) \tag{1}$$

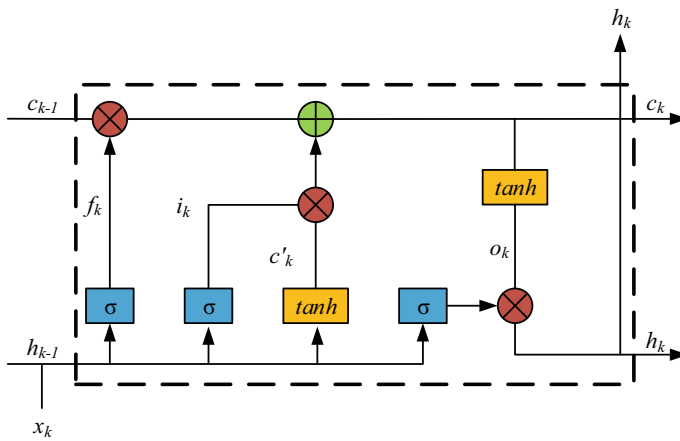


Fig. 5 The structure of the LSTM network

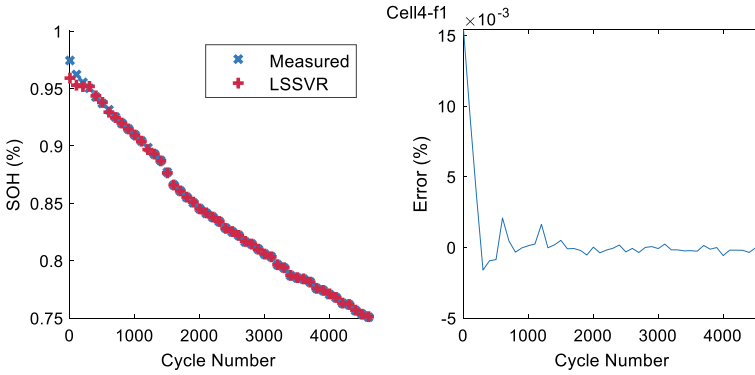


Fig. 6 SOH estimation results and errors of Cell4 based on feature f_1 and LSSVR model

$$f_k = \sigma(W_{f,x}x_k + W_{f,h}h_{k-1} + b_f) \tag{2}$$

$$o_k = \sigma(W_{o,x}x_k + W_{o,h}h_{k-1} + b_o) \tag{3}$$

$$c'_k = \tanh(W_{c,x}x_k + W_{c,h}h_{k-1} + b_c) \tag{4}$$

$$c_k = f_k \otimes c_{k-1} + i_k \otimes c'_k \tag{5}$$

$$h_k = o_k \otimes \tanh(c_k) \tag{6}$$

where h_k and i_k present the output and the input at the time k , respectively. i_k , f_k , o_k , and c_k denote the input gate, forget gate, output gate, and memory cell of the LSTM neural network. c'_k is the intermediate calculation. And W and b correspondingly represents the weights and biases of the gates. During the training stage, the weights and bias can be automatically adjusted based on the backpropagation and gradient descent optimization method [16]. The sigmoid activation function σ and hyperbolic tangent activation function \tanh are expressed as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{8}$$

where the output value of the sigmoid activation function is located between 0 and 1, and the output range of the hyperbolic tangent activation function is between -1 and 1.

3.2 Evaluation Metrics

In this paper, the Pearson correlation coefficient is used to evaluate the correlation between the extracted features and the battery SOH, and it is calculated as:

$$\rho_{x_i} = \frac{\sum(x_i - \bar{x}_i)(y - \bar{y})}{\sqrt{\sum(x_i - \bar{x}_i)^2 \sum(y - \bar{y})^2}} \quad (9)$$

where x_i and \bar{x}_i stand for the measured value and the mean of the measured value, respectively. y and \bar{y} are the output value and the mean of the output value. Additionally, three evaluation metrics, including R^2 , mean absolute error (MAE), and maximum error (MAX), are applied to assess the performance of the LSTM model for SOH estimation of the battery based on the selected degradation features.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (11)$$

$$MAX = \max |y_i - x_i| \quad (12)$$

where n and y_i denote the total number of the input data and the estimated value, respectively.

4 Results and Discussions

In this section, the data of six cells from the dataset is used for model training and Cell4 and Cell8 are used to validate the effectiveness of the extracted degradation features for the battery SOH estimation. Based on the extracted degradation feature $f1$, Figs. 6, 7, 8, 9, 10 and 11 compare the SOH estimation results of two cells using least squares support vector regression (LSSVR), feedforward neural network (FNN), and LSTM models. The absolute error is utilized in the evaluation to the accuracy of the prediction models. By contrast, the LSTM model provides more accurate SOH estimation according to the absolute errors. Moreover, Table 4 exhibits the estimation results of three models using different degradation features. It can be seen that the degradation features $f2$ and $f3$ can achieve a more accurate SOH estimation of the battery, especially for the feature $f3$. Compared to FNN and LSSVR models, the R^2 obtained by LSTM model is closer to 1 and the MAE and MAX of the SOH estimation are lower. For example, it achieved 1.4×10^{-5} in MAE, 5.1×10^{-3} in MAX, and 0.9997 in R^2 for the Cell4 when using the degradation feature $f3$.

Table 4 MAE, MAX, and R^2 of SOH estimation

Model	Metrics	Test cell					
		Cell4			Cell8		
		$f1$	$f2$	$f3$	$f1$	$f2$	$f3$
LSSVR	MAE	5.2×10^{-4}	8.4×10^{-5}	6.9×10^{-4}	1.8×10^{-3}	1.8×10^{-3}	1.6×10^{-3}
	MAX	1.5×10^{-2}	8.9×10^{-3}	2.3×10^{-3}	1.6×10^{-2}	1.2×10^{-2}	8.0×10^{-3}
	R^2	0.9986	0.9995	0.9996	0.9740	0.9748	0.9795
FNN	MAE	2.4×10^{-4}	3.6×10^{-4}	2.6×10^{-3}	1.0×10^{-3}	6.1×10^{-3}	1.4×10^{-3}
	MAX	3.7×10^{-2}	9.0×10^{-3}	9.6×10^{-3}	2.6×10^{-2}	1.0×10^{-2}	2.9×10^{-2}
	R^2	0.9594	0.9956	0.9697	0.9881	0.9813	0.9902
LSTM	MAE	4.0×10^{-5}	1.5×10^{-5}	1.4×10^{-5}	2.2×10^{-5}	3.3×10^{-6}	1.6×10^{-6}
	MAX	1.2×10^{-2}	5.5×10^{-3}	5.1×10^{-3}	1.2×10^{-2}	4.4×10^{-3}	3.50×10^{-3}
	R^2	0.9990	0.9996	0.9997	0.9995	0.9998	0.9998

Table 5 MAE and MAX of SOH estimation based on the training features of charge voltages [12]

Test cell	Feature	GPR		GRU-CNN	
		MAE	MAX	MAE	MAX
Cell4	Charge voltages	2.12	5.83	0.61	1.60
Cell8		2.65	4.13	0.65	1.62

Furthermore, Table 5 presents the SOH estimation results obtained by GPR and GRU-CNN models based on this dataset, while the selected training features are the charging voltages [12]. It does not consider the extraction of the degradation features, thus obtaining the relatively tough estimation with large MAE and MAX. By contrast, the LSTM model combined with the extracted degradation features in this work can offer higher SOH estimation accuracy, which also validates the effectiveness of the extracted degradation features for SOH estimation. Meanwhile, the single training feature can effectively reduce the computational complexity of the SOH estimation than using multiple features.

5 Conclusions

In this paper, we extracted and summarized the effective degradation features from the CC charging voltage data for the battery SOH estimation based on the machine

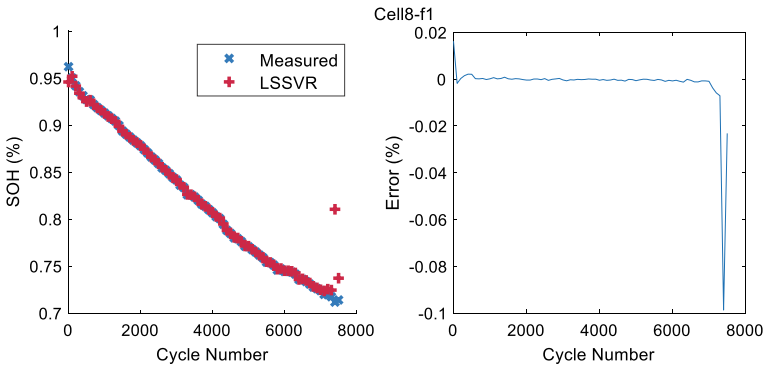


Fig. 7 SOH estimation results and errors of Cell8 based on feature f_1 and LSSVR model

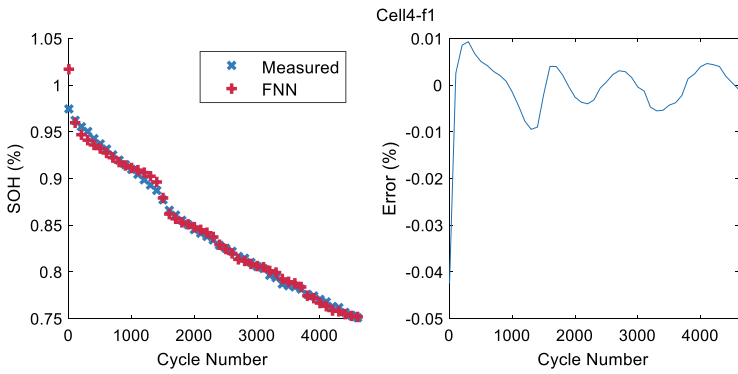


Fig. 8 SOH estimation results and errors of Cell4 based on feature f_1 and FNN model

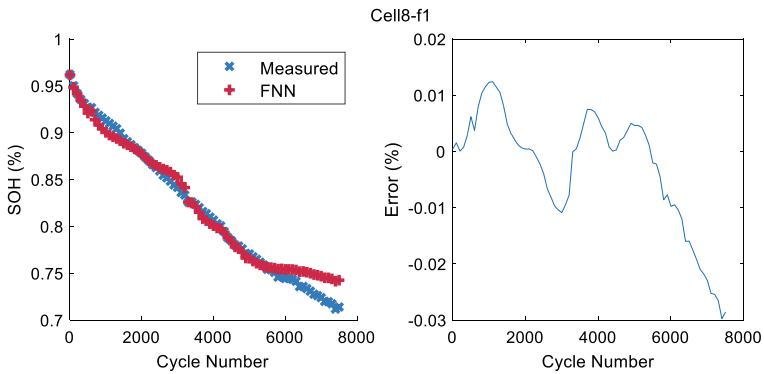


Fig. 9 SOH estimation results and errors of Cell8 based on feature f_1 and FNN model

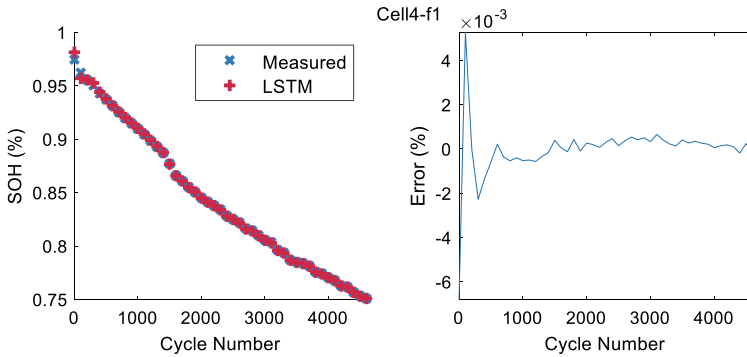


Fig. 10 SOH estimation results and errors of Cell4 based on feature $f1$ and LSTM model

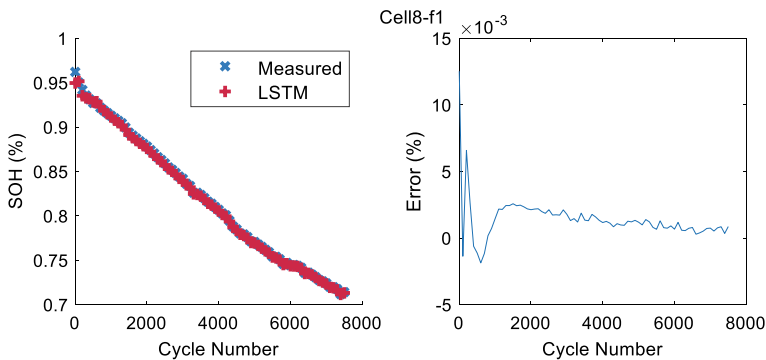


Fig. 11 SOH estimation results and errors of Cell8 based on feature $f1$ and LSTM model

learning methods. Three types of features are extracted and compared based on the Pearson correlation coefficient. It is found that the required charging time in the charging ranges of 2.8–4.2 V, 3.0–4.2 V, and 3.2–4.2 V are strongly correlated with the battery SOH. Especially for the range of 3.2–4.2 V, the extracted degradation feature can provide a more accurate SOH estimation. Compared to LSSVR and FNN models, the LSTM model can achieve higher estimation accuracy with R^2 close to 1 and lower MAE and MAX. It is important to mention that the estimated accuracy is higher than the previous study based on the training features of charge voltages.

References

1. Bian, X., Liu, L., Yan, J., Zou, Z., Zhao, R.: An open circuit voltage-based model for state-of-health estimation of lithium-ion batteries: model development and validation. *J. Power Sources* **448**, 227401 (2020)
2. Li, Y., Stroe, D.I., Cheng, Y., Sheng, H., Sui, X., Teodorescu, R.: On the feature selection for

- battery state of health estimation based on charging–discharging profiles. *J. Energy Storage* **33**, 102122 (2021)
3. Dai, H., Zhao, G., Lin, M., Wu, J., Zheng, G.: A novel estimation method for the state of health of lithium-ion battery using prior knowledge-based neural network and Markov chain. *IEEE Trans. Industr. Electron.* **66**(10), 7706–7716 (2018)
 4. Wang, Z., Zeng, S., Guo, J., Qin, T.: State of health estimation of lithium-ion batteries based on the constant voltage charging curve. *Energy* **167**, 661–669 (2019)
 5. Bian, X., Liu, L., Yan, J., Zou, Z., Zhao, R.: An open circuit voltage-based model for state-of-health estimation of lithium-ion batteries: model development and validation. *J. Power Sources*, 227401 (2019)
 6. Bi, Y., Yin, Y., Choe, S.Y.: Online state of health and aging parameter estimation using a physics-based life model with a particle filter. *J. Power Sources* **476**, 228655 (2020)
 7. Downey, A., Lui, Y.H., Hu, C., Laflamme, S., Hu, S.: Physics-based prognostics of lithium-ion battery using non-linear least squares with dynamic bounds. *Reliab. Eng. Syst. Saf.* **182**, 1–12 (2019)
 8. Ng, M.F., Zhao, J., Yan, Q., Conduit, G.J., Seh, Z.W.: Predicting the state of charge and health of batteries using data-driven machine learning. *Nat. Mach. Intell.* **2**(3), 161–170 (2020)
 9. Pan, H., Lü, Z., Wang, H., Wei, H., Chen, L.: Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine. *Energy* **160**, 466–477 (2018)
 10. Chen, L., Wang, H., Liu, B., Wang, Y., Ding, Y., Pan, H.: Battery state-of-health estimation based on a metabolic extreme learning machine combining degradation state model and error compensation. *Energy* **215**, 119078 (2020)
 11. Zhang, S., Zhai, B., Guo, X., Wang, K., Peng, N., Zhang, X.: Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks. *J. Energy Storage* **26**, 100951 (2019)
 12. Fan, Y., Xiao, F., Li, C., Yang, G., Tang, X.: A novel deep learning framework for state of health estimation of lithium-ion battery. *J. Energy Storage* **32**, 101741 (2020)
 13. Christoph, R.B.: *Diagnosis and Prognosis of Degradation in Lithium-Ion Batteries* Doctoral dissertation, Ph. D. Thesis, Department of Engineering Science, University of Oxford, Oxford, UK (2017)
 14. Liu, Y., Shu, X., Yu, H., Shen, J., Zhang, Y., Liu, Y., Chen, Z.: State of charge prediction framework for lithium-ion batteries incorporating long short-term memory network and transfer learning. *J. Energy Storage* **37**, 102494 (2021)
 15. Chen, Z., Zhao, H., Shu, X., Zhang, Y., Shen, J., Liu, Y.: Synthetic state of charge estimation for lithium-ion batteries based on long short-term memory network modeling and adaptive H-Infinity filter. *Energy* **228**, 120630 (2021)
 16. Ma, L., Hu, C., Cheng, F.: State of charge and state of energy estimation for lithium-ion batteries based on a long short-term memory neural network. *J. Energy Storage* **37**, 102440 (2021)

Harmonic Response Analysis of a Dual-Rotor System with Mass Unbalance



Yubin Yue and Hongjun Wang

Abstract The finite element model of a dual-rotor system (coaxial structure) was established by Ansys entity model. The inner rotor is supported by three bearings while the outer rotor is supported by two bearings. The outer rotor connects the inner rotor by an intermediary bearing. The critical speed Characteristic of the dual-rotor system was calculated by selecting the inner rotor and outer rotor separately as the main excitation source. The harmonic response analysis of the dual-rotor system with mass unbalance were analyzed in order to study the influence of mass unbalance of the rotor system. The results indicate that the outer rotor with mass unbalance is more likely to cause resonance than the inner rotor with mass unbalance, but the resonance amplitude is relatively small. It is possible to reduce the harmonic resonance frequency in the dual-rotor system by improving the design of the outer rotor. Harmonic resonance amplitude gradually increases with the increase of excitation frequency.

Keywords Dual-rotor system · Mass unbalance · Vibration · Harmonic response

Y. Yue

China Academy of Machinery Science and Technology Group Co., Ltd, Beijing 100044, China
e-mail: 18810608234@126.com

Y. Yue · H. Wang (✉)

School of Mechanical and Electrical Engineering, Beijing Information Science and Technology University, Beijing 100192, China
e-mail: wanghongjun@bistu.edu.cn

H. Wang

Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, BISTU, Beijing 100192, China

Beijing International Science Cooperation Base of High-End Equipment Intelligent Perception and Control, BISTU, Beijing 100192, China

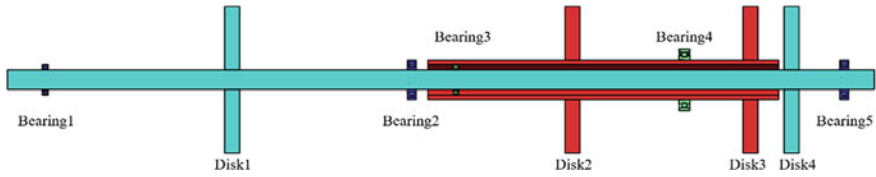


Fig. 1 Structural diagram of a dual-rotor system

1 Introduction

Modern aero-engine and gas turbine usually adopt the structure of dual-rotor suit. The rotor includes the Disk, blade and shaft of compressor and turbine, and is connected to the casing through rolling bearing and support system. Due to the double frequency excitation, airflow disturbance, structural nonlinearity and other factors of the high- and low-pressure rotors, the vibration of the high- and low-pressure rotors will be transmitted through the support and the intermediate bearing, and then the vibration between the rotors will be coupled. At the same time, the intermediate bearing has clearance and Hertzian contact nonlinearity, which increases the difficulty of rotor dynamics analysis and design [1–3].

The critical speed of the dual-rotor system with intermediate bearing is much more complex than that of the single rotor system. The main reason is that the critical speed needs to be divided into the critical speed mainly excited by the low-pressure shaft and the critical speed mainly excited by the high-pressure shaft. The high and low rotors can rotate in the same direction and in the opposite direction, and the critical speed under the same and opposite rotation is also different, which increases the complexity of the solution.

Scholars at home and abroad have carried out a lot of research work on unbalance fault of rotor system, including critical speed solution of dual-rotor system, unbalance force fault mechanism, numerical simulation analysis, experimental verification and unbalance fault identification and diagnosis [4–10].

Based on the general finite element software ANSYS, with the help of its natural mode solution module, the critical speed solution method of dual-rotor system under unbalanced excitation is given, and the influence of unbalanced excitation on the spectrum characteristics of dual-rotor system is studied.

2 Finite Element Model Construction of Dual-Rotor System

2.1 Physical Model

As shown in Fig. 1, the model consists of an inner (low-pressure) rotor and an outer (high-pressure) rotor system. The shaft of the inner rotor is solid and the shaft of the outer rotor is hollow. The inner rotor is supported by bearing1, bearing2 and bearing5, and the outer rotor is supported by bearing3 intermediate bearing and bearing4 support bearing. The two rotors are connected by bearing3 intermediate bearings. Disk1, Disk2, Disk3 and Disk4 represent low pressure compressor, high pressure compressor, high pressure turbine and low-pressure turbine respectively. The dimension parameters of inner and outer rotor shafts are shown in Tables 1 and 2. The position and setting parameters of the bearing are shown in Tables 3 and 4. The speed curves of inner and outer rotors are shown in Table 5.

Table 1 Size of the shafts (length/diameter)

Parameters	L1/D1	L2/D2 (in), D2 (out)
Values (m)	1.73/0.04	0.7/0.062, 0.08

Table 2 Disk properties for the rotor system

Parameters	Diameter (m)	Thickness (m)	Density (kg/m ³)
Values (m)	0.3	0.03	7850

Table 3 Component location for the rotor system

	Component	Position
1	Bearing1	0.0755
2	Disk1	0.4885
3	Bearing2	0.807
4	Bearing3	0.895
5	Disk2	1.1275
6	Bearing4	1.351
7	Disk3	1.483
8	Disk4	1.565
9	Bearing5	1.670

Table 4 Bearing properties for the rotor system

	K ₁₁ (MN/m)	K ₂₂ (MN/m)
Bearing1	8.1	8.1
Bearing2	10.5	10.5
Bearing3	8.1	8.0
Bearing4	10.5	10.5
Bearing5	10.5	10.5

Table 5 Speed curve of the rotor system

	Inner rotor (rad/s)	Outer rotor (rad/s)
1	0	0
2	105	314
3	209	419
4	314	523
5	419	628
6	523	733
7	628	837
8	733	942
9	837	1047

2.2 Finite Element Model of the Dual-Rotor System

The dynamic model of the dual rotor system was established by using finite element method. The shaft end is modeled by a beam, and the disk is regarded as a rigid body. The rotational inertia of the shaft end and the disk are considered in the model [11–13].

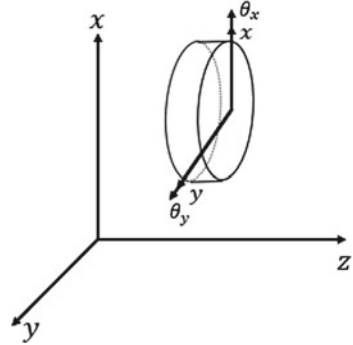
2.2.1 Rigid Disk Model

Take the displacement x, y and the angle of the transverse vibration of the rigid disk θ_x, θ_y is the generalized coordinate, which is shown in Fig. 2. After obtaining the kinetic energy of the disk, the Lagrange equation of motion is substituted to obtain the differential equation of motion of the rigid disk:

$$\begin{cases} M_d \ddot{q}_{1d} + \Omega J \dot{q}_{2d} = F_{1d} \\ M_d \ddot{q}_{2d} - \Omega J \dot{q}_{1d} = F_{2d} \end{cases} \quad (1)$$

where $M_d = \begin{bmatrix} m & \\ & J_d \end{bmatrix}$, $q_{1d} = \begin{bmatrix} x \\ \theta_x \end{bmatrix}$, $q_{2d} = \begin{bmatrix} y \\ \theta_y \end{bmatrix}$, $J = \begin{bmatrix} m & \\ & J_p \end{bmatrix}$. M_d represents the mass matrix of the disk. m represents the mass of the disk. J_d represents the diameter moment of inertia. Ω represents the angular velocity of disk rotation. G_d

Fig. 2 Schematic diagram of the coordinate system of disk element



$= \Omega \mathbf{J}$ represents the rotation matrix of the disk, and \mathbf{J}_p represents the polar moment of inertia. \mathbf{F}_{1d} and \mathbf{F}_{2d} represent the corresponding generalized forces, including unbalance force, rubbing force, binding force at the support, forces between adjacent elements and moments, etc.

2.2.2 Flexible Shaft Element Model

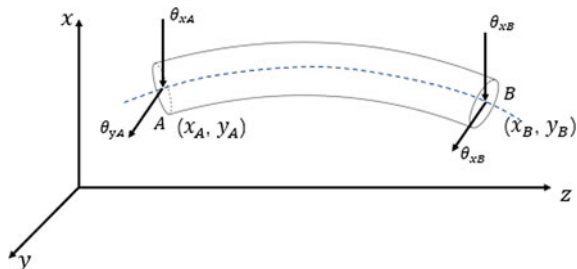
Take the generalized displacements at points A and B of the two ends of the flexible shaft element \mathbf{q}_{1s} and \mathbf{q}_{2s} , which is shown in Fig. 3, where $\mathbf{q}_{1s} = [x_A \ \theta_{xA} \ x_B \ \theta_{xB}]^T$, $\mathbf{q}_{2s} = [y_A \ \theta_{yA} \ y_B \ \theta_{yB}]^T$.

The differential equation of motion of flexible shaft element can be obtained by substituting Lagrange equation of motion:

$$\begin{cases} M_S \ddot{q}_{1s} + \Omega J_S \dot{q}_{2s} + K_S^e q_{1s} = F_{1s} \\ M_S \ddot{q}_{2s} - \Omega J_S \dot{q}_{1s} + K_S^e q_{2s} = F_{2s} \end{cases} \quad (2)$$

Timoshenko beam-shaft model is adopted. Where \mathbf{K}_S , \mathbf{M}_S , \mathbf{J}_S represent the stiffness matrix, mass matrix and polar moment of inertia matrix of the shaft element respectively. \mathbf{F}_{1s} and \mathbf{F}_{2s} represent the corresponding generalized forces.

Fig. 3 Schematic diagram of the coordinate system of shaft element



2.2.3 Intermediary Bearing Model

The motion differential equation of the intermediate bearing and the supporting assembly can be expressed as:

$$-C_b \dot{q}_b - K_b q_b = F_b \tag{3}$$

where C_b represents the bearing damping matrix. K_b represents the bearing stiffness matrix. q_b represents the bearing displacement, and F_b marks the bearing external force. C_b and K_b can be written as $C_b = \begin{bmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{bmatrix}$, $K_b = \begin{bmatrix} k_{xx} & k_{xy} \\ k_{yx} & k_{yy} \end{bmatrix}$, $c_{xx} = c_{yy}$, $k_{xx} = k_{yy}$, $c_{xy} = c_{yx} = 0$, $k_{xy} = k_{yx} = 0$.

2.2.4 Dynamic Modeling of the Dual-Rotor System

The independent differential equations of the double rotor system are divided into the differential equations of the low-pressure rotor and the high-pressure rotor.

Differential equation of motion of low-pressure rotor system:

$$\begin{cases} M_L \ddot{q}_{1L} + \Omega_L J_L \dot{q}_{2L} + K_L q_{1L} = F_{1L} \\ M_L \ddot{q}_{2L} - \Omega_L J_L \dot{q}_{1L} + K_L q_{2L} = F_{2L} \end{cases} \tag{4}$$

where $q_{1L} = [x_1, \theta_{x1}, x_2, \theta_{x2}, \dots, x_M, \theta_{xM}]^T$, $q_{2L} = [y_1, \theta_{y1}, y_2, \theta_{y2}, \dots, y_M, \theta_{yM}]^T$.

Differential equation of motion of high-pressure rotor system:

$$\begin{cases} M_H \ddot{q}_{1H} + \Omega_H J_H \dot{q}_{2H} + K_H q_{1H} = F_{1H} \\ M_H \ddot{q}_{2H} - \Omega_H J_H \dot{q}_{1H} + K_H q_{2H} = F_{2H} \end{cases} \tag{5}$$

where $q_{1H} = [x_{M+1}, \theta_{x(M+1)}, x_{M+2}, \theta_{x(M+2)}, \dots, x_{M+N}, \theta_{x(M+N)}]^T$,

$q_{2H} = [y_{M+1}, \theta_{y(M+1)}, y_{M+2}, \theta_{y(M+2)}, \dots, y_{M+N}, \theta_{y(M+N)}]^T$.

M_L and M_H are the integrated mass matrices of the low-pressure rotor system and the high-pressure rotor system. $\Omega_L J_L$, and $\Omega_H J_H$ are the integrated rotation matrices of the low-pressure rotor system and the high-pressure rotor system. K_L and K_H are the integrated stiffness matrices of the low-pressure rotor system and the high-pressure rotor system. F_{1L} , F_{2L} , F_{1H} and F_{2H} are external excitation. M and N are the total number of nodes of the low-pressure rotor system and the high-pressure rotor system respectively.

Considering all the damping, stiffness and mass matrices of the shaft, disc and bearing, the general dynamic equation of the double-rotor system is:

$$M \ddot{q} + (C + \Omega_L G_L - \Omega_H G_H) \dot{q} + K q = F_u + F_g \tag{6}$$

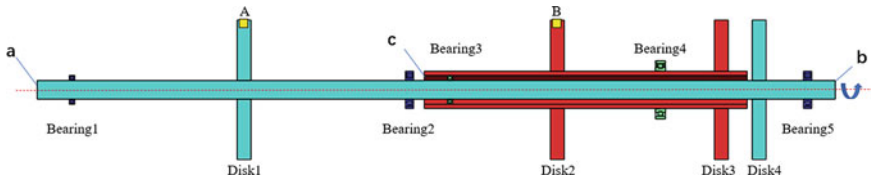


Fig. 4 Unbalance excitation of a dual-rotor system

where \mathbf{M} , \mathbf{C} and \mathbf{K} are the mass matrix, damping matrix and stiffness matrix of the double-rotor system respectively. \mathbf{q} is the displacement vector. Ω_L and Ω_H are respectively the speed of the low-pressure rotor and the high-pressure rotor. \mathbf{G}_L and \mathbf{G}_H are the rotation matrices of the low rotor and the high rotor respectively. \mathbf{F}_u is the unbalance excitation vector of the rotor system. \mathbf{F}_g is the gravity vector.

2.3 Unbalanced Excitation

As shown in Fig. 4, the rotation direction of the inner and outer rotors is positive while the unbalance occurs on Disk1 of the inner rotor and Disk2 of the outer rotor. The unbalance excitation is added to the outer diameter of the Disk (point A and point B), and the unbalance is 0.15 kg mm. By simulating the actual monitoring, the vibration monitoring points are arranged at the front face **a** of low-pressure compressor Disk1, the rear face **b** of low-pressure turbine Disk4 and the front face **c** of high-pressure compressor Disk2. We studied the influence of the single unbalance and coupling unbalance of Disk1 and Disk2 on the spectrum characteristics of the dual rotor system.

3 Simulation

3.1 Calculation of Critical Speed Characteristic of Dual-Rotor System

By solving the critical speed of the outer rotor when the inner rotor is the main excitation and the critical speed of the inner rotor when the outer rotor is the main excitation respectively, the critical speed spectrum of the double-rotor system is drawn according to the obtained results, as shown in Fig. 5. The abscissa is the inner rotor speed, while the vertical axis is the outer rotor speed. The oblique line L represents the speed curve of the rotor system, and the intersection point of the modified line and the critical speed spectrum is the critical speed point of the dual-rotor system. Specific critical speed values are shown in Tables 6 and 7.

Fig. 5 Critical speed spectrum of a dual-rotor system

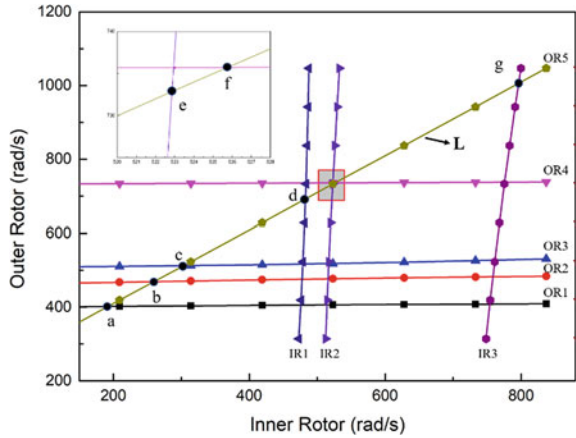


Table 6 The critical speed of the inner rotor (main excitation)

	d	e	g
Critical speed (rad/s)	482.6	522.9	796.4

Table 7 The critical speed of the outer rotor (main excitation)

	a	b	c	f
Critical speed (rad/s)	402.0	469.3	512.7	735.7

3.2 Harmonic Resonance

As shown in Figs. 6, 7 and 8, when Disk1 alone is unbalanced, resonance occurs at 64, 79 and 120 Hz. When Disk2 is unbalanced alone, resonance occurs at 64, 79,

Fig. 6 Harmonic response of Disk1 alone with mass unbalance

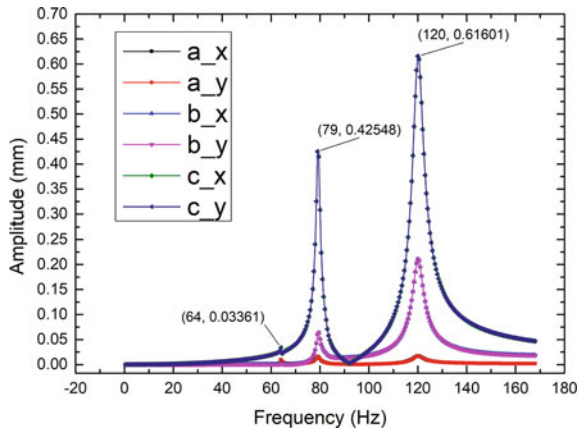


Fig. 7 Harmonic response of Disk2 alone with mass unbalance

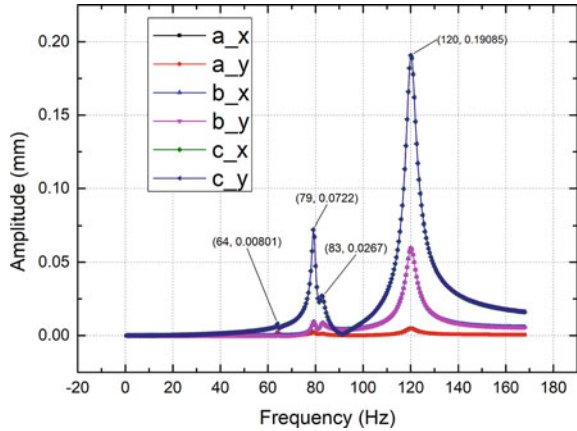
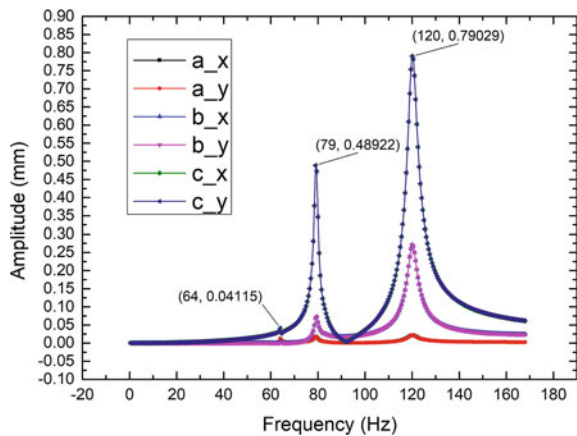


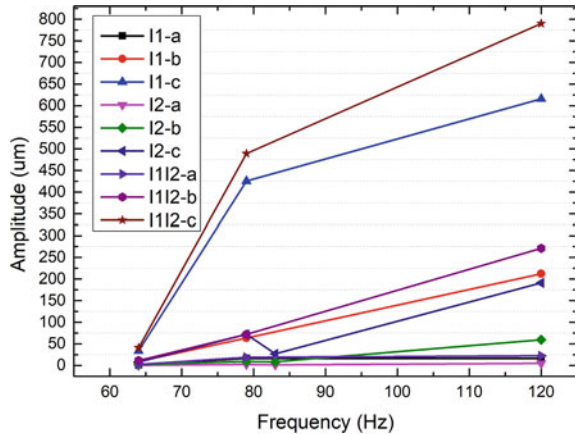
Fig. 8 Harmonic response of Disk1 and Disk2 with mass unbalance



83 and 120 Hz, and an 83 Hz resonance point is added relative to Disk when Disk is unbalanced alone. It can be judged here that due to the existence of intermediate bearings, Disk2 is more likely to cause resonance, but the amplitude is much smaller than the resonance induced by Disk1 unbalanced fault. Therefore, the resonance at this time is also easier to suppress. When Disk1 and Disk2 add unbalanced excitation at the same time, the 83 Hz resonance point that appears before disappears, which may be because the introduction of Disk1 unbalance inhibitions the generation of resonance, but also leads to the increase of resonance amplitude.

As shown in Fig. 9, comparative analysis of the harmonic response amplitude of the unbalanced excitation was made. Obviously, when Disk1 and Disk2 are simultaneously unbalanced excitation, the vibration amplitude of face a and face c of the two end faces of the inner and outer rotor is larger, especially the face c section of the outer rotor. With the increase of the excitation frequency, the vibration amplitude gradually increases.

Fig. 9 The amplitude of the harmonic response varies with mass unbalance



4 Conclusions and Discussions

A finite element model of unbalance fault of a double-rotor system is established, the critical speed Characteristic of the dual-rotor system was calculated and the harmonic response of the system under unbalance excitation is studied. Some conclusions drawn from the study can be summarized as follows.

- (1) Under the excitation of unbalanced force, the outer rotor is more likely to cause resonance than the inner rotor, but the resonance amplitude is relatively small;
- (2) It is possible to reduce the resonant frequency in the dual-rotor system by improving the design of the outer rotor;
- (3) Under the excitation of unbalanced force, the resonance amplitude gradually increases with the increase of excitation frequency;

In the future, we hope to make further study about the dynamics of unbalance force in the double-rotor system. The transfer matrix method will be used for analysis, the unbalance force will be loaded to other Disks for further analysis to verify the above conclusion.

Acknowledgements This research is supported by the National Natural Science Foundation of China (Grant No. 51975058).

References

1. Mingfu, L.: Aero-Engine Rotor Dynamics. Northwestern Polytechnical University Press, Xian (2015)
2. Zheng, W.: Rotor Dynamics Design for Rotating Machinery. Tsinghua University Press, Beijing (2015)

3. Yie, Z., Yanzong, H., Zheng, W., et al.: Rotor Dynamics. Tsinghua University Press, Beijing (1987)
4. Zhang, H., Huang, L., Li, X., Jiang, L., Yang, D., Zhang, F., Miao, J., Sopanen, J.: Spectrum analysis of a coaxial dual-rotor system with coupling misalignment. *Shock Vib.* (2020)
5. Dayi, Z., Yehui, L., Zhichao, L., Yanhong, M., Jie, H.: Prediction for critical speed of double spools system in aero engines. *J. Propul. Technol.* **36**(02), 292–298 (2015)
6. Yi, J.D.: A control strategy of actively actuated eccentric mass system for imbalance rotor vibration. *Actuators* **9**(3) (2020)
7. Xiaoling, S., Shuting, W., Yonggang, L., Lifeng, C.: Characteristics analysis of doubly-fed wind turbine under impeller mass unbalance fault. *Acta Energiæ Solaris Sinica* **38**(05), 1324–1332 (2017)
8. Peng, G., Lei, H., Yushu, C.: Nonlinear vibration characteristics of a dual-rotor system with inter-shaft bearing. *J. Vib. Shock* **38**(15), 1–10 (2019)
9. Haiqin, Q., Yaotao, Z., Kejun, X.: Rubbing vibration response theoretical analysis and experimental verification for a dual-rotor support casing system. *J. Mech. Eng.* **55**(19), 75–83 (2019)
10. Li Changyou, X., Minqiang, G.S., Yüewu, W., Rixin, W., Jingbo, G.: Estimation of the unbalance magnitude of rotor system based on model. *J. Aerosp. Power* **24**(07), 1530–1536 (2009)
11. Dexin, L.: Dual-Rotor System with Marine Gas Turbine Vibration Condition Monitoring and Evaluation Method Study. Harbin Institute of Technology (2013)
12. Jiang, C.: Dynamic Characteristic Research of Dual Rotor-Rolling Bearing Systems. Harbin Institute of Technology (2017)
13. Zhenyong, L.: Dynamical Modeling and Nonlinear Vibration Study of Aero-Engine Rotor System. Harbin Institute of Technology (2017)

A Novel Method for Stacking Optimization of Aeroengine Multi-stage Rotors Based on 3D Deviation Prediction Model



Jia Kang, Jun He, Zhisheng Peng, Haizhou Huang, and Shixi Yang

Abstract The assembly precision of high pressure compressor (HPC) multi-stage rotors is critical to the healthy and high-performance operation of aeroengine. The traditional trial assembly method usually requires multiple disassembly and assembly of rotors to ensure assembly precision, which is cumbersome and may cause damage to parts. It is important to predict and optimize the stacking precision of multi-stage rotors to improve the assembly quality and reduce aeroengine operational failure. This paper proposes a novel method for stacking optimization of aeroengine multi-stage rotors based on 3D deviation prediction model. First, the three-dimensional deviation propagation and accumulation process in the stacking process of a four-stage rotor is deduced using the coordinate transformation method, and the 3D deviation prediction model is established to derive the concentricity and perpendicularity of rotors under different bolt hole phase combinations; Second, the Gaussian distribution is used to simulate the radial and axial runout data of the upper end surface of each stage of rotor, and the center and unit normal vector of the upper end surface of rotors are obtained by the least squares method, then the predicted values of concentricity and perpendicularity of rotors under different bolt hole phase combinations are calculated through the 3D deviation prediction model; Third, a dual-objective integer optimization function is established, and the optimal installation phase and the corresponding predicted values of concentricity and perpendicularity of rotors are

J. Kang · J. He (✉) · S. Yang (✉)

The State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, China

e-mail: hejunzju@zju.edu.cn

S. Yang

e-mail: yangsx@zju.edu.cn

J. Kang

e-mail: jiakang@zju.edu.cn

Z. Peng

Hangzhou Steam Turbine & Power Group Co., Ltd, Hangzhou, China

H. Huang

Huadian Electric Power Research Institute Co., Ltd, Hangzhou, China

obtained. The proposed method can predict the concentricity and perpendicularity of rotors and optimize the stacking phase accurately, so as to achieve the target stacking precision requirements through a single assembly, which can provide theoretical guidance for the actual stacking process of aeroengine multi-stage rotors.

Keywords Stacking optimization · Multi-stage rotors · 3D deviation prediction model · Concentricity · Perpendicularity

1 Introduction

Assembly is one of the most crucial technique links in the field of aeroengine manufacturing. The stacking precision of core-engine HPC has a huge influence on the aeroengine performance [1]. Due to the manufacture error and leverage effect, the geometry deviation of aeroengine multi-stage rotors will propagate and accumulate, which may lead to severe vibration and even failure of aeroengine [2]. Concentricity and perpendicularity are set as the key parameters to evaluate and control the stacking quality of aeroengine multi-stage rotors, which should satisfy the requirements simultaneously [3]. Therefore, reasonable deviation propagation prediction and stacking optimization method of rotors should be developed to ensure the stacking precision and simplify the assembly process.

Lots of studies have been carried out to analyze the deviation propagation process and optimize the assembly precision. Hussian et al. [4] proposed the “straight-build assembly” and minimize eccentricity of two-dimensional rotating machines stage-by-stage. Lafond et al. [5] introduced Jacobian matrix into mechanical assembly deviation propagation, and use virtual joints to simulate the three degrees of freedom of translation and three rotation directions. Desrochers et al. proposed the Torsor model to describe the assembly of multi-stage rotors and combined the Jacobian matrix and Torsor model to analyze and calculate the geometric variations in three-dimension space [6–8]. Chen et al. [9] introduced four major 3D tolerance analysis model, T-Map, Matrix, Unified Jacobian-Torsor and DLM, and compared the characteristics in different aspects.

Many geometry deviation modeling methods are used to analyze the deviation propagation of aeroengine multi-stage rotors assembly, and some research also developed methods to control the geometry deviation and imbalance of assembly [10, 11]. Yang et al. [12] proposed a variation propagation control method for straight-build assembly by controlling the rotor eccentricity and calculated the probability of exceeding a particular value with Monte Carlo simulation. Kang et al. [13] analyzed the transmission of deviations of the rigid-compliant bolted connected aeroengine shell during the assembly process with Jacobian-Torsor matrix. Chen et al. [14] used the genetic algorithm to optimize the coaxiality and unbalance to improve the assembly quality, and proposed a new assembly optimization strategy. Zhang et al. [15] predicted the eccentricity of aeroengine rotors with the method of image identification and machine learning.

The stacking optimization of multi-stage rotors is actually integer and multi-objective optimization, and the establishment of reasonable deviation prediction model needs to consider dimensions, objectives and computational complexity, etc. Researches above ignored some of listed aspects or assumed the phase adjustment as continuous, which may affect the accuracy of stacking prediction and optimization of rotors. Therefore, the deviation propagation modeling and the multi-objective integer optimization of aeroengine multi-stage rotors awaits further study.

The paper proposed a novel optimization method for aeroengine multi-stage rotors stacking based on 3D deviation prediction model, which was established with the coordinate transformation method under the rigid body assumption and can predict the concentricity and perpendicularity of rotors in different bolt hole phase combinations. Furthermore, a dual-objective integer optimization method was presented to optimize the stacking phase.

The rest of the paper is organized as follows: Sect. 2 used the coordinate transformation to analyze the deviation propagation process and established the 3D deviation prediction model. Section 3 analyzed the predicted values with simulated runout data of rotors and optimized the stacking phase. The conclusion is presented in Sect. 4.

2 3D Deviation Prediction Model

2.1 Establishment of Coordinate Systems

The structure of multi-stage rotors assembly of aeroengine HPC is shown in Fig. 1, which is stacked by single-stage rotors. Each stage of rotor is positioned radially and axially in the way of interference fit, and connected through bolt holes evenly distributed on the rabbet joint surface.

The establishment of the coordinate system is the premise for the aeroengine multi-stage rotors stacking prediction by the coordinate transformation method. This paper takes the stacking process of four-stage aeroengine rotors as an example, each stage of which has 32 bolt holes in the rabbet joint surface.

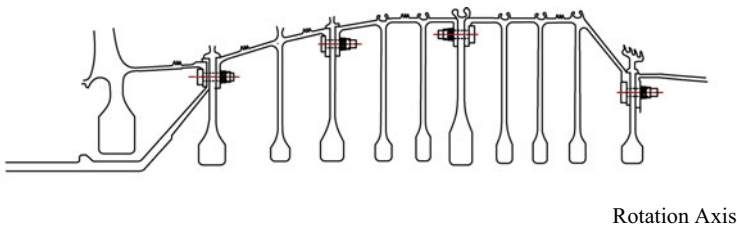
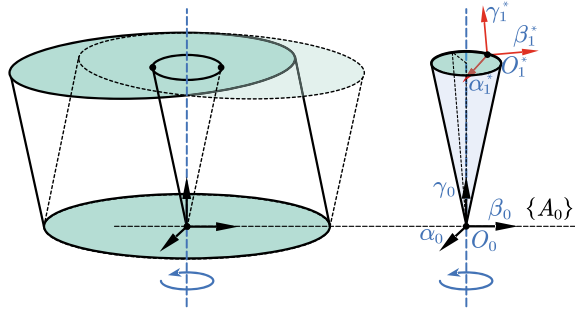


Fig. 1 Structure of aeroengine HPC multi-stage rotors

Fig. 2 Single-stage rotor and the coordinate system



First, the radial and axial runout data of the upper and lower surfaces of each stage of the rotor by measuring instruments or simulation method. Then the least square method is used to fit the radial and axial runout data to obtain the center and unit normal vector of the joint surface of the rotor.

As shown in Fig. 2, take the lower end surface of the first-stage rotor as the reference plane and establish a global coordinate system $\{A_0\} = O_0\alpha_0\beta_0\gamma_0$, in which γ_0 is the unit normal vector of the lower end surface of the first-stage rotor. The axis that passes through the center of the lower end surface O_0 and coincides with γ_0 is set as the reference axis during the rotors stacking process. α_0 and β_0 are orthogonal unit vectors in the reference plane that perpendicular to γ_0 .

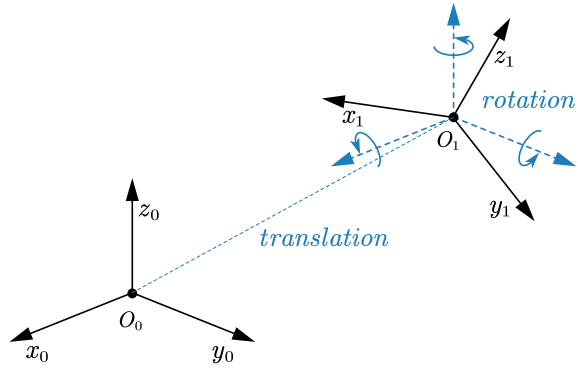
In the global coordinate system $\{A_0\}$, O_1^* is the center of the upper end surface of the first-stage rotor and γ_1^* is the upward normal vector of the surface that pass through O_1^* . The unit vector α_1^* is defined along the direction of the connection of O_1^* and one of the bolt holes centers, then the unit vector β_1^* can also be determined, thereby established the local coordinate system $\{A_1^*\} = O_1^*\alpha_1^*\beta_1^*\gamma_1^*$. Since each stage of rotor have similar characteristics as the rotating body, the local coordinate system of the upper end surface of the second and third stage rotor, i.e. $\{A_2^*\} = O_2^*\alpha_2^*\beta_2^*\gamma_2^*$ and $\{A_3^*\} = O_3^*\alpha_3^*\beta_3^*\gamma_3^*$, are established in the same way.

This paper assumed rotor as a rigid body, thus the deformation of rotors during the stacking process is ignored, and it should be noted that the upper end surface of the lower stage rotor is aligned with the bolt holes in the lower end surface of the higher stage rotor, and the center of the two surfaces coincides with the normal vector over the center, so the coordinate systems corresponding to the two surfaces can be regarded as the same coordinate system. Besides, rotor1 is used to represent the first stage of rotor, and rotor2 represents the second stage of rotor, and so forth.

2.2 Coordinate Transformation of Rotors

In order to describe the deviation propagation in the stacking process and predict the concentricity and perpendicularity of the rotors at each stage, the transformation relationship between the coordinate systems needs to be established, and local

Fig. 3 Coordinate transformation



coordinates of the center and the normal vector of the upper end surface of rotors need to be transformed into the coordinates in the global coordinate system $\{A_0\}$. As shown in Fig. 3, the transformation between coordinate systems is realized by translation and rotation. Translation is represented by the movement of the origin of the coordinate system from O_0 to O_1 , and rotation is represented by the rotation of the coordinate axis.

The rotation transformation of the orthogonal bases of the two coordinate systems is completed by the transition matrix. As shown in Eqs. (1) and (2), The orthonormal basis of coordinate system $\{A_0\}$ can be transformed into orthonormal basis coordinate system $\{A_1^*\}$ uniquely based on the transition matrix C_0^1 , and other orthonormal basis coordinate system are transformed in the same way.

$$\begin{cases} \alpha_1^* = c_{11} \cdot \alpha_0 + c_{21} \cdot \beta_0 + c_{31} \cdot \gamma_0 \\ \beta_1^* = c_{12} \cdot \alpha_0 + c_{22} \cdot \beta_0 + c_{32} \cdot \gamma_0 \\ \gamma_1^* = c_{13} \cdot \alpha_0 + c_{23} \cdot \beta_0 + c_{33} \cdot \gamma_0 \end{cases} \quad (1)$$

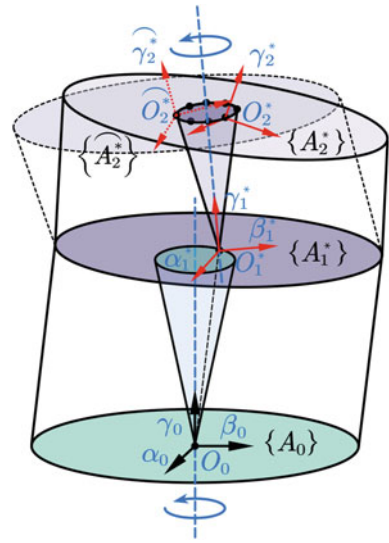
$$[\alpha_1^*, \beta_1^*, \gamma_1^*] = [\alpha_0, \beta_0, \gamma_0] \cdot C_0^1 \quad (2)$$

As shown in Eq. (3) and Fig. 4, according to the correspondence between the coordinates of the points and the coordinate system in coordinate transformation, the coordinates of the center of the upper end surface of the rotor2 in the local coordinate system $\{A_1^*\}$, $O_2^* = (x_2^*, y_2^*, z_2^*)$, can be transform into the coordinates in the global coordinate system $\{A_0\}$, $O_2 = (x_2, y_2, z_2)$.

$$O_2 = C_0^1 \cdot O_2^* + O_1^* \quad (3)$$

As shown in Eqs. (4) and (5), through the coordinate transformation method and transition matrix, the global coordinate of centers and unit normal vectors of upper surfaces of rotors can be obtained.

Fig. 4 Local and global coordinate systems of stacking rotors



$$\begin{cases} O_1 = O_1^* \\ O_2 = C_0^1 \cdot O_2^* + O_1 = C_0^1 \cdot O_2^* + O_1^* \\ O_3 = C_0^2 \cdot O_3^* + O_2 = C_0^1 \cdot C_1^2 \cdot O_3^* + C_0^1 \cdot O_2^* + O_1^* \\ O_4 = C_0^3 \cdot O_4^* + O_3 = C_0^1 \cdot C_1^2 \cdot C_2^3 \cdot O_4^* + C_0^1 \cdot C_1^2 \cdot O_3^* + C_0^1 \cdot O_2^* + O_1^* \end{cases} \quad (4)$$

$$\begin{cases} \gamma_1 = \gamma_1^* \\ \gamma_2 = C_0^1 \cdot \gamma_2^* \\ \gamma_3 = C_0^1 \cdot C_1^2 \cdot \gamma_3^* \\ \gamma_4 = C_0^1 \cdot C_1^2 \cdot C_2^3 \cdot \gamma_4^* \end{cases} \quad (5)$$

2.3 3D Deviation Prediction Model with Phase Adjustment

The stacking optimization process of the rotor is actually a process of adjusting the bolt holes installation phase of rotors at all stages to obtain optimal stacking precision, the deviation propagation and phase adjustment process of two-stage rotors are shown in Fig. 4. First, rotor1 is placed on the installation platform, and the global coordinate system and reference axis of the rotor during the stacking process are determined by the lower end surface of rotor1. When rotor1 and rotor2 are stacked, the center and unit normal vector of the joint surfaces of two rotors coincide. During the phase adjustment, rotor2 rotates along the unit normal vector which cross the center in the joint surface until bolt holes are aligned. As shown in Fig. 4, along with rotation during phase adjustment, the center and unit normal vector of upper end surface of rotor2 formed center trajectory and vector trajectory, and are scattered in

corresponding bolt holes phase. The inverted cone in Fig. 4 is used to represent the center trajectory and unit normal vector trajectory of the upper end surface of the rotor when the installation phase changes.

The phase adjustment process between rotors of stage i and $i + 1$ can be represented by the rotation matrix $\mathbf{R}(\theta_i)$, and θ_i is the rotation angle relative to initial phase. The rotor of stage $i + 1$ rotates along the $\boldsymbol{\gamma}_i^*$ in the local coordinate system $\{A_i^*\}$, so the form of $\mathbf{R}(\theta_i)$ is shown in Eq. (6).

$$\mathbf{R}(\theta_i) = \begin{bmatrix} \cos \theta_i & -\sin \theta_i & 0 \\ \sin \theta_i & \cos \theta_i & 0 \\ 0 & 0 & 1 \end{bmatrix}; i = 1, 2, 3, 4 \tag{6}$$

Considering the influence of the rotation matrix $\mathbf{R}(\theta_i)$, the transition matrix \mathbf{C}_i^{i+1} is transformed into $\mathbf{C}_i^{i+1} \cdot \mathbf{R}(\theta_i)$, thereby the global coordinates of O_i and $\boldsymbol{\gamma}_i$ of rotors are shown in Eqs. (7) and (8).

$$\begin{cases} O_1 = O_1^* \\ O_2 = \mathbf{C}_0^1 \cdot \mathbf{R}(\theta_1) \cdot O_2^* + O_1^* \\ O_3 = \mathbf{C}_0^1 \cdot \mathbf{R}(\theta_1) \cdot \mathbf{C}_1^2 \cdot \mathbf{R}(\theta_2) \cdot O_3^* + \mathbf{C}_0^1 \cdot \mathbf{R}(\theta_1) \cdot O_2^* + O_1^* \\ O_4 = \mathbf{C}_0^1 \cdot \mathbf{R}(\theta_1) \cdot \mathbf{C}_1^2 \cdot \mathbf{R}(\theta_2) \cdot \mathbf{C}_2^3 \cdot \mathbf{R}(\theta_3) \cdot O_4^* + \mathbf{C}_0^1 \cdot \mathbf{R}(\theta_1) \cdot \mathbf{C}_1^2 \cdot \mathbf{R}(\theta_2) \cdot O_3^* \end{cases} \tag{7}$$

$$\begin{cases} \boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_1^* \\ \boldsymbol{\gamma}_2 = \mathbf{C}_0^1 \cdot \mathbf{R}(\theta_1) \cdot \boldsymbol{\gamma}_2^* \\ \boldsymbol{\gamma}_3 = \mathbf{C}_0^1 \cdot \mathbf{R}(\theta_1) \cdot \mathbf{C}_1^2 \cdot \mathbf{R}(\theta_2) \cdot \boldsymbol{\gamma}_3^* \\ \boldsymbol{\gamma}_4 = \mathbf{C}_0^1 \cdot \mathbf{R}(\theta_1) \cdot \mathbf{C}_1^2 \cdot \mathbf{R}(\theta_2) \cdot \mathbf{C}_2^3 \cdot \mathbf{R}(\theta_3) \cdot \boldsymbol{\gamma}_4^* \end{cases} \tag{8}$$

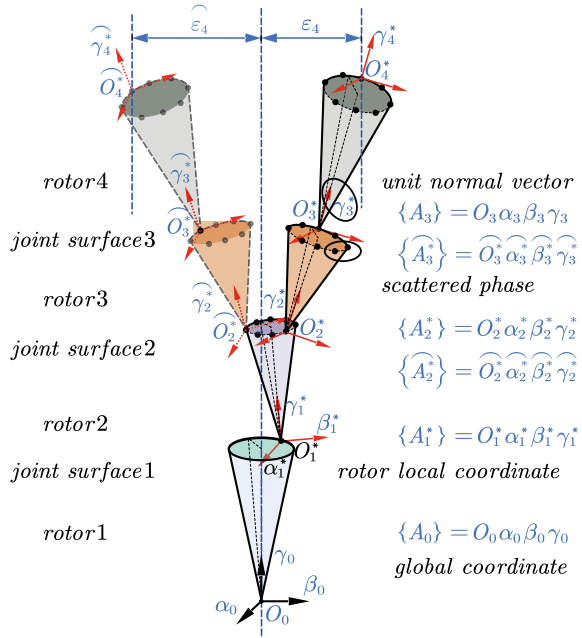
According to the coordinates of O_i of each stage after phase adjustment, the concentricity of the rotors relative to the reference axis can be calculated by Eq. (9).

$$e_i = \sqrt{O_{ix}^2 + O_{iy}^2}; i = 1, 2, 3, 4 \tag{9}$$

The global coordinates of $\boldsymbol{\gamma}_i$ is $\{\gamma_{ix}, \gamma_{iy}, \gamma_{iz}\}$, while coordinates of $\boldsymbol{\gamma}_0$ that coincides with the reference axis is $\{0, 0, 1\}$. The angle between $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_0$ satisfy the Eq. (10), and R_i is radius of the upper surface of rotors, so the perpendicularity p_i of the upper end surface of rotors relative to the reference axis can be calculated through Eq. (11).

$$\cos(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_0) = \frac{\boldsymbol{\gamma}_i \cdot \boldsymbol{\gamma}_0}{|\boldsymbol{\gamma}_i| \cdot |\boldsymbol{\gamma}_0|} = \frac{\gamma_{iz}}{\sqrt{\gamma_{ix}^2 + \gamma_{iy}^2 + \gamma_{iz}^2}}; i = 1, 2, 3, 4 \tag{10}$$

Fig. 5 Rotors and coordinate systems under different bolt hole phases



$$p_i = 2 \cdot R_i \cdot \tan(\gamma_i, \gamma_0) = 2 \cdot R_i \cdot \frac{\sqrt{\gamma_{ix}^2 + \gamma_{iy}^2}}{\gamma_{iz}}; i = 1, 2, 3, 4 \quad (11)$$

Through the above process, the 3D deviation prediction model of concentricity and perpendicularity of rotors with phase adjustment is established. As shown in Fig. 5, different local coordinate systems are formed when the bolt holes installation phase is adjusted, and the global coordinates of the center and unit normal vector of the upper end face of rotors in the global coordinate system change accordingly, which will affect the concentricity and perpendicularity of stacking rotors. Therefore, the bolt hole phase combination can be optimized to improve the stacking precision.

3 Stacking Optimization Method

3.1 Simulation and Fitting of Runout Data

Considering confidentiality requirements, the Gaussian distribution function is adapted to simulate the radial and axial runout data based on the structure size and the runout range of rabbet joint of rotors, and the least square method is used to fit the simulated runout data. The stacking results are obtained through the 3D deviation prediction model in Sect. 2 and the bolt holes phases are optimized.

Table 1 Four-stage rotors structure size and rabbet joint runout zone

Rotor	Height (mm)	R of radial (mm)	run-out zone (mm)	R of axial (mm)	Run-out zone (mm)
1	200	113	0.05	90	0.05
2	180	131	0.05	98	0.05
3	200	152	0.05	114	0.05
4	150	146	0.05	108	0.05

Refer to the structure size and runout requirements of the HPC rotors of a certain type of aeroengine, the radial size, axial size, height size and runout range of the four-stage rotors are listed in Table 1.

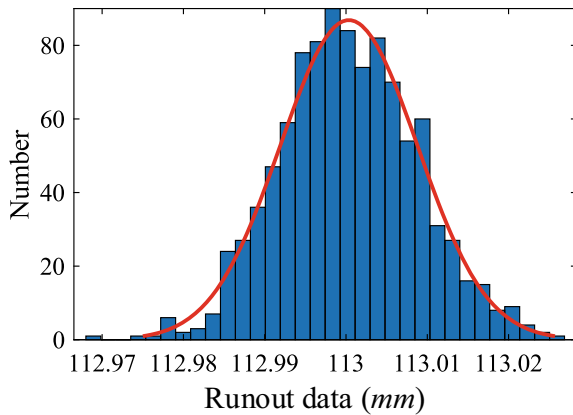
As shown in Eq. (12), the Gaussian distribution function is used to simulate the runout data of rotors. For the radial runout data of rotors, the radial radius of the upper rabbet of each stage of rotor is used as the mean value μ , and 1/6 of the runout range is used as the standard deviation σ through the 3σ method. The 1000 simulated runout points of rotor1 are shown in Fig. 6.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \tag{12}$$

The least square method is used to fit the center and radius of the radial runout circle, and the fitted circle center is used as the center of the upper surface of rotor. The radial runout fitting circle and the center of four rotors are shown in the Fig. 7, where the deviation between the fitting center of rotors and the theoretical center exists and will propagate and cumulative during the stacking process.

The axial runout data are simulated and fitted in similar way, while the height of each rotor is used as the mean value of Gaussian distribution function. As shown in Fig. 8, the fitting plane passes through the axial runout data and the upward unit

Fig. 6 Distribution of simulated radial runout data of rotor1



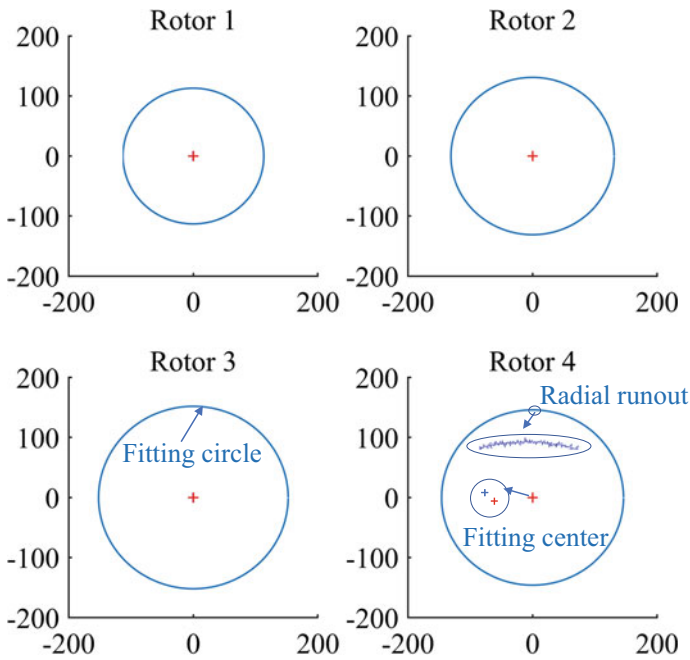
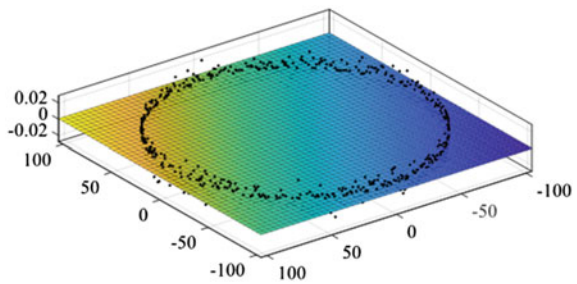


Fig. 7 Fitting centers and circles of four rotors

Fig. 8 The axial runout data and fitting plane of rotor4



normal vector of which is set as the unit normal vector of the upper surface of each stage of rotor. The local coordinates of fitted center and unit normal vector of the upper surface of each stage of rotor are shown in Table 2. The parameters listed represent the geometric characteristics of the upper end surface of rotors in the local coordinate system and can be used for deviation prediction and phase optimization of rotors.

Table 2 Local coordinates of fitted geometric parameters of upper end surface of rotors

Rotor	Center (O_i^*) (mm)	Unit normal vector (\mathbf{y}_i^*) (mm)
1	$(-1.1768 \times 10^{-4}, -3.8595 \times 10^{-4}, 200)$	$\{1.4058 \times 10^{-6}, -3.2838 \times 10^{-6}, 1.0000\}$
2	$(-6.9692 \times 10^{-4}, 1.2300 \times 10^{-4}, 180)$	$\{-2.6770 \times 10^{-6}, -1.4257 \times 10^{-6}, 1.0000\}$
3	$(-2.6772 \times 10^{-4}, -3.6326 \times 10^{-4}, 200)$	$\{3.6291 \times 10^{-6}, -1.7286 \times 10^{-6}, 1.0000\}$
4	$(-5.4805 \times 10^{-4}, -6.0428 \times 10^{-5}, 150)$	$\{1.5865 \times 10^{-6}, -1.4076 \times 10^{-6}, 1.0000\}$

3.2 Stacking Prediction of a Four-Stage Rotor

According to the aeroengine multi-stage rotors 3D deviation prediction model in Sect. 2 and the simulated and fitted data in Sect. 3.1, the concentricity and perpendicularity of each stage of rotors under different bolt hole phase combinations are predicted. The predicted concentricity and perpendicularity of rotor1 is 0.4035 and $0.6787 \mu m$, and the result will not change with phase adjustment.

There are 32 adjustable bolt hole phases between rotor1 and rotor2, and the concentricity and perpendicularity of rotor2 under different phases are obtained through the stacking prediction model. As shown in Fig. 9, the concentricity of rotor2 is scattered in range $[2.8664 \times 10^{-4}, 0.0017]$ mm, and presents an approximate sinusoidal change trend with phase, which corresponds to the inclined center trajectory of upper surface of rotor2 in Fig. 4. The perpendicularity of rotor2 is in range $[5.8233 \times 10^{-4}, 0.0048]$ mm.

Accordingly, rotor3 has 1024 adjustable bolt hole phases, and the concentricity and perpendicularity of rotor3 are show in Fig. 10. Different from the change trend of rotor2, the concentricity of rotor3 is approximately a cluster of sine curves with

Fig. 9 The stacking prediction values of rotor2

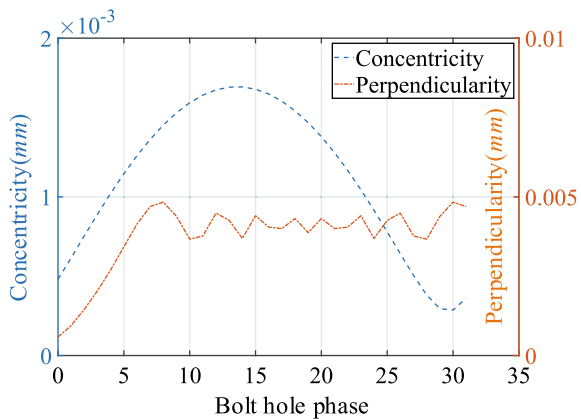
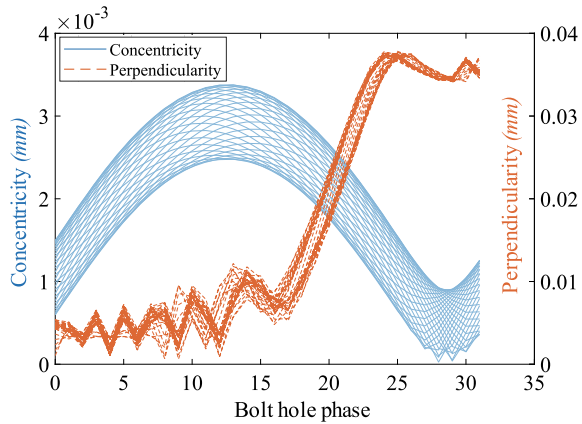


Fig. 10 The stacking prediction values of rotor3



different initial phases and amplitudes but the same period, and the change range of concentricity is $[2.6057 \times 10^{-5}, 0.0034]$ mm. As shown in Eq. (11), when calculating the perpendicularity of rotor, it needs to be multiplied by the diameter of the upper end surface of the rotor. Therefore, due to the leverage effect and vector composition, the perpendicularity of rotor3 changes more greatly with the phase of the bolt hole, and the change range of perpendicularity is $[6.0895 \times 10^{-4}, 0.0378]$ mm.

The statistical results of predicted concentricity and perpendicularity of rotor4 are shown in Fig. 11, of which have 32,768 kinds of phase combination. And the change range of concentricity and perpendicularity is $[1.5982 \times 10^{-5}, 0.0047]$ and $[8.4276 \times 10^{-5}, 0.0664]$ mm. The stacking prediction results of the concentricity and verticality of the four-stage rotors indicate that if the aeroengine rotor is assembled by the trial assembly method, the concentricity and perpendicularity of the intermediate or final stage of rotor are likely to be out of tolerance.

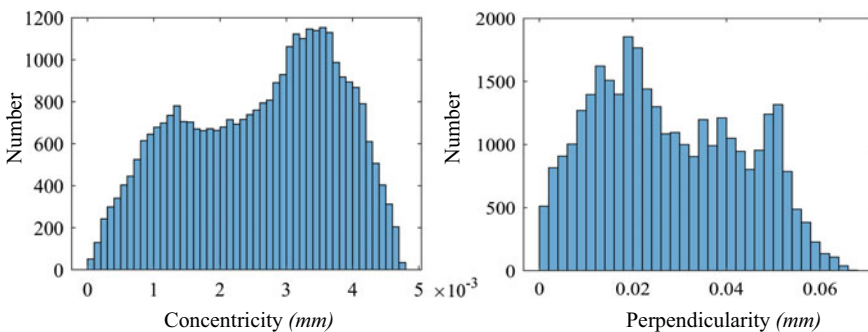


Fig. 11 Statistical results of predicted concentricity and perpendicularity of rotor4

3.3 Dual-Objective Integer Stacking Optimization

During the stacking process, if the measured concentricity and perpendicularity of the rotor are out of tolerance, it requires multiple disassembly and assembly to achieve the required stacking accuracy. This process is cumbersome and may affect the stacking quality of the rotor and the aeroengine, so it is important to predict and optimize the stacking precision of multi-stage rotors before assembly.

Since the bolt holes are uniformly distributed on the rabbet joint surface of the aeroengine rotor, the multi-stage rotor stacking optimization belongs to integer optimization. Additionally, the multi-stage rotor stacking optimization is also dual-objective optimization that contain concentricity and perpendicularity. Given that rotor4 is most likely to be out of tolerance due to the influence of the lever effect, the optimization target T is set as the weighted sum of the concentricity and perpendicularity of rotor4, meanwhile the intermediate stage of rotor should not be out of tolerance. The optimization objective function is established as shown in Eq. (13), and the weighting factor u is set to 2 considering the assembly requirements and the importance of two values.

$$\left\{ \begin{array}{l} T = \min\{u \cdot e_4(b_1, b_2, b_3) + p_4(b_1, b_2, b_3)\} \\ \text{s.t. } e_s(b_1, \dots, b_{s-1}) \leq \varepsilon_0 \\ \quad p_s(b_1, \dots, b_{s-1}) \leq t_0 \\ 0 \leq b_m \leq 31 \\ 1 \leq m \leq 3 \\ 1 \leq s \leq 3 \\ b_m, m, s \in I \end{array} \right. \quad (13)$$

The total phase combination of aeroengine multi-stage rotor stacking increases exponentially with the number of rotor stages. Therefore, the branch and bound method is used to prune the bolt hole phase combinations that do not meet the restriction conditions, thereby reducing the amount of calculation for stacking prediction, which is conducive to focus on the feasible phases and search for the optimal solution. The restriction conditions ε_0 and t_0 are set to 1.5 and 20 μm .

After pruning the bolt hole phase combinations, the predicted concentricity and perpendicularity prediction results of rotor4 are shown as Fig. 12. Compared with the results in Fig. 11, the search bolt hole phase combination is reduced from 32,768 to 2400 according to the calculation result and the ultimate feasible phase combination is 514, and the optimal results of the dual-objective integer optimization function and corresponding bolt hole phases are shown in Table 3.

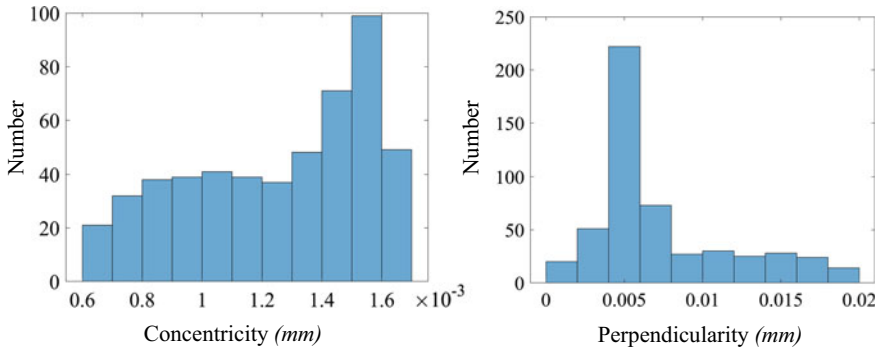


Fig. 12 Statistical results of predicted values of rotor4 after pruning

Table 3 The optimization phase and corresponding prediction values

	Concentricity (μm)	Perpendicularity (μm)
Rotor 1	0.4035	0.6787
Rotor 2	0.4806	0.5823
Rotor 3	1.0808	4.7621
Rotor 4	0.7155	0.2321
T	1.6631	
Bolt hole phase	(1, 14, 31)	

4 Conclusion

This paper proposes a novel method for stacking optimization of aeroengine multi-stage rotors assembly based on 3D deviation prediction model. The prediction model is established by analyzing the 3D deviation propagation process and predicting the concentricity and perpendicularity of rotors at each stage under different bolt hole phase combinations. The predicted values of rotors are calculated through the prediction model and show different distribution characteristic with the change of phase. Finally, the dual-objective integer optimization function of a four-stage rotor is established and the bolt hole phases are optimized. The proposed 3D deviation prediction model and stacking optimization method can provide theoretical guidance for the actual stacking process of aeroengine multi-stage rotors.

Acknowledgements The authors acknowledge the support from the National Natural Science Foundation of China (Grant No. U1809219) and the Key Research and Development Project of Zhejiang Province (Grant No. 2020C01088).

References

1. Ding, S.: An improved Jacobian-Torsor model for statistical variation solution in aero-engine rotors assembly. *Proc. Instit. Mech. Eng. Part B J. Eng. Manuf.* **235**(3), 466–483 (2021)
2. Liu, Y.Q.: Analysis of whole aeroengine vibration control technology. *Aeroengine* **39**(05), 1–8+13 (in Chinese, 2013)
3. Sun, C.: Prediction method of concentricity and perpendicularity of aero engine multistage rotors based on pso-bp neural network. *Defense & Aerospace Week* 332 (2019)
4. Hussain, T., Yang, Z.: Straight-build assembly optimization: a method to minimize stage-by-stage eccentricity error in the assembly of axisymmetric rigid components (two-dimensional case study). *J. Manuf. Sci. Eng. J. Manuf. Sci. Eng.* **133**(3) (2011)
5. Lafond, P., Luc: Jacobian-based modeling of dispersions affecting pre-defined functional requirements of mechanical assemblies. In: *IEEE International Symposium on Assembly & Task Planning*, pp. 20–25. IEEE (1999).
6. Desrochers, A.: Geometrical Variations Management in a Multi-Disciplinary Environment with the Jacobian-Torsor Model, *Models for Computer Aided Tolerancing in Design and Manufacturing*, pp. 75–84 (2007)
7. Ghie, W., Laperriere, L.: A unified jacobian-torsor model for analysis in computer aided tolerancing. In: *4th International Conference on Integrated Design and Manufacturing in Mechanical Engineering*, pp. 63–72. Clermont-Ferrand, France (2002)
8. Li, H.: Tolerance analysis of mechanical assemblies based on small displacement torsor and deviation propagation theories. *Int. J. Adv. Manuf. Technol.* **72**(1–4), 89–99 (2014)
9. Chen, H.: A comprehensive study of three dimensional tolerance analysis methods. *Comput. Aided Des.* **53**, 1–13 (2014)
10. Yang, Z.: A comparison of different optimization techniques for variation propagation control in mechanical assembly. *Top Conf.* **26**(1), 1–11 (2011)
11. Sun, C.: A method to control the amount of unbalance propagation in precise cylindrical components assembly. *Proc. Instit. Mech. Eng. Part B J. Eng. Manuf.* **233**(13), 2458–2468 (2019)
12. Yang, Z.: A probabilistic approach to variation propagation control for straight build in mechanical assembly. *Int. J. Adv. Manuf. Technol.* **64**(5–8), 1029–1047 (2013)
13. Kang, H.: Assembly research of aero-engine casing involving bolted connection based on rigid-compliant coupling assembly deviation modeling. *ARCHIVE Proc. Instit. Mech. Eng. Part C J. Mech. Eng. Sci.* **234**(14), 2803–2820 (2020)
14. Chen, Y.: Research on multistage rotor assembly optimization methods for aeroengine based on the genetic algorithm. *Complexity* **2021**(4), 1–14 (2021)
15. Zhang, Z.: Eccentricity of rotor prediction of aero-engine rotor based on image identification and machine learning. In: *Proceedings of the ASME 2019 International Mechanical Engineering Congress and Exposition*. 2(B), ASME, Salt Lake City, Utah, USA (2019).

A Sensor Fault Identification Method Based on Adaptive Particle Swarm Optimization Support Vector Machine



Xuezhen Cheng, Dafei Wang, Chuannuo Xu, and Jiming Li

Abstract Accurate identification of fault types is an important part of sensor fault diagnosis. Therefore, a sensor faults identification method based on Adaptive Particle Swarm Optimization Support Vector Machine (APSO-SVM) is proposed in this paper. Firstly, the appropriate Time-domain parameters are extracted from the fault data to realize feature extraction and dimension reduction. Then the Particle Swarm Optimization (PSO) algorithm is improved by adjusting the particle velocity with weight and introducing mutated particles, so as to improve the optimization ability of the algorithm and to optimize the parameters of Support Vector Machine (SVM). Finally, the optimized model is used to identify the sensor faults, and compared with other advanced algorithms, the results show that the proposed method can identify the sensor faults more accurately.

Keywords Sensor fault diagnosis · Time-DOMAIN parameters · Particle swarm optimization (PSO) · Support vector machine (SVM) · APSO-SVM

1 Introduction

As a basic data acquisition device, the sensor is pivotal to ensure the safe and reliable operation of the monitoring system [1]. Because of the working environment of the sensor is often bad, it is easy to fail in the process of operation, so the fault diagnosis of the sensor is very necessary. As an important part of sensor fault diagnosis, fault pattern recognition is pivotal to improve the reliability of diagnosis, so it is necessary to improve the accuracy of fault mode classification.

There are many methods of sensor fault diagnosis at domestic and overseas. Traditional diagnosis methods mainly include model-based method, knowledge-based

X. Cheng · D. Wang · C. Xu · J. Li (✉)
College of Electrical and Automation Engineering, Shandong University of Science and Technology, No.579 Qianwangang Road, Qingdao 266590, China
e-mail: wugetongxue@163.com

X. Cheng
e-mail: zhenxc6411@163.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_60

method and data-driven method [2]. Model-based method requires the establishment of accurate mathematical model for the research object, but modeling of complex system is usually more difficult. The knowledge-based method relies on expert experience and prior knowledge which lacks certain adaptability. With the development of artificial intelligence, data-driven intelligent diagnosis methods have attracted extensive attention to researchers [3]. The data of sensor is relatively easy to obtain, Due to fault data onto the sensor contains a lot of miscellany information, and similar characteristics from part of the fault data, it is difficult to directly use the fault data, so it is necessary to choose an appropriate method for feature extraction of the fault data. Time-domain parameters can reflect different characteristics of faults and are often used as reference indexes in the field of fault diagnosis [4].

Machine learning algorithms, such as Neural Network (NN) and Support Vector Machine (SVM), usually show excellent performance while process the data. However, NN needs a large number of training samples, otherwise its performance will be limited. SVM is suitable for processing small samples and nonlinear data, and the selection of appropriate hyper parameters can get better performance. PSO algorithm as a heuristic optimization algorithm, its basic algorithm and improved algorithm have been extended to many fields, which is suitable for parameter optimization of SVM. However, the basic PSO has some shortcomings in the search step size, global optimization and local optimization mechanisms, which affect its search ability. Zheng [5] proposes an improved adaptive particle swarm optimization (PSO) algorithm by adding weight to the initial velocity of particles, improved the adaptability of search step size in the iterative process, which is used to search the maximum target at the minimum cost. Wang [6] proposes a PSO algorithm with inertia factor to solve the local optimal solution problem, and for inversion of fault parameters. Cheng [7] proposes comprehensive learning particle swarm optimization (CPSO) algorithm, establishing the connection between the global optimal solution and the local optimal solution, so the search ability of the algorithm is improved. However, the ability of the CPSO algorithm still needs to be improved.

To solve these problems, a sensor faults identification method based on APSO-SVM is proposed to this paper. Firstly, several Time-domain parameters are extracted from the sensor fault data, and the combination of Time-domain parameters which could accurately reflect the fault characteristics is obtained through many experiments. Then, the PSO algorithm is improved. On the basis of adjusting the initial velocity of particles by weight, individual mutation particles are introduced to improve the optimization ability of the algorithm, and the improved algorithm is used to optimize the parameters of SVM. Finally, the samples composed of Time-domain parameters, are input to the classifiers such as Probabilistic Neural Network (PNN), Decision Tree (DT), CPSO-SVM and APSO-SVM, for the purpose of training and testing. The effectiveness of the proposed method is verified by compare the classification results and analyzing the performance.

2 An Improved Particle Swarm Optimization Algorithm

Particle Swarm Optimization (PSO) algorithm has been widely used for its advantages of simple calculation and high calculation accuracy. In the PSO, inertia weight ω of initial velocity determines the initial searching speed of the particle, $gbest$ reflects the global searching ability of the particle, while $pbest$ determines the local searching ability of the particle. However, the initial velocity of the traditional particle swarm optimization algorithm is fixed, so it is difficult to guarantee the search step size is appropriate in the whole iteration process. Besides, with the reduction of search space, individual particles are easy to fall into the optimal solution, which will also affect the search ability of the algorithm. So this paper makes the following improvements:

Suppose a space n with a D particle dimension, the position of the i particle in space is $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$, and the velocity of the particle, which x_i will be substituted into the fitness function $f(x_i)$ to obtain the fitness value. At this time, the optimal position passed by the individual i is $pbest_i = (p_{i1}, p_{i2}, \dots, p_{iD})$, and the optimal position experienced by the population is $gbest = (g_1, g_2, \dots, g_D)$. On the basis of these two optimal values, the speed and position are continuously updated based on the Eq. (1):

$$\begin{aligned} v_i^k &= \omega v_i^{k-1} + c \times r(gbest_i - pbest) \\ x_i^k &= x_i^{k-1} + v_i^{k-1} \end{aligned} \quad (1)$$

where the current position of particles is x_i , v_i^k is the velocity vector of particles i after k iterations; c is the weight factor, which is used to adjust the step size of updating. r is the random number in $[0, 1]$, through which the randomness of the search can be adjusted appropriately, ω is the inertia weight, which can be used weighted the initial velocity as following in Eq. (2):

$$\omega = 0.5 + 0.5 \cdot (T_{\max} - T_i) / T_{\max} \quad (2)$$

And then introduce mutated individual particles to solve the problem of local optimal solution, assuming that $K \in [0, 2]$, when $r > 0.5$, if $K = 1$, as following in Eq. (3):

$$pop(j, k) = (N - 1) \cdot rand + 1 \quad (3)$$

where $pop(j, k)$ is the current particle size, N is the maximum particle size. If $K = 2$, as following in Eq. (4):

$$pop(j, k) = (P_{g \max} - P_{g \min}) \cdot rand + P_{g \min} \quad (4)$$

where $P_{g \max}$ is the maximum g parameter, $P_{g \min}$ is the minimum g parameter.

3 Fault Identification Method Based on APSO-SVM

3.1 Establishment of APSO-SVM Model

The establishment method of mathematical model of multi-classification support vector machine is actually a convex quadratic programming problem. It is of great significance to select appropriate hyper parameters to improve classification accuracy.

The objective function of convex quadratic programming is constructed in Eq. (5):

$$\begin{aligned} \min Q(\alpha) &= \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(X_i, X_j) - \sum_{i=1}^N \alpha_i \\ \text{S.t. } \sum_{i,j=1}^N \alpha_i y_i &= 0 \\ 0 \leq \alpha_i &\leq C, i = 1, 2, \dots, N \end{aligned} \tag{5}$$

where α_i is the Lagrange multiplier, X_i, X_j is the input vector, y_i is the category label, and $K(X_i, X_j)$ is the kernel function. In fact, not all data is completely linearly separable, hence the Hinge loss is adopted in Eq. (6):

$$\begin{aligned} \min_{\omega, b, \xi_i} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t. } y_i(\omega \times x_i + b) &\geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, 2, \dots, N \end{aligned} \tag{6}$$

where ω is the normal plane vector, ξ_i is the relaxation variable, and each sample corresponds to one ξ_i , represents the degree to which the sample does not satisfy the constraint, and C is the penalty factor. The corresponding classification function is shown in Eq. (7):

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^N \alpha_i^* y_i K(x, x_i) + b^* \right\} \tag{7}$$

where b^* is the bias constant. Introduce kernel function can enhance the processing ability of support vector machine for nonlinear problems. In this paper, Gaussian kernel function with superior performance is selected in Eq. (8):

$$K(x, z) = \exp(\gamma \|x - z\|) \tag{8}$$

where γ is the kernel function coefficient. It can be seen from the above equations, the key to establishing SVM model is the selection of C and γ parameters.

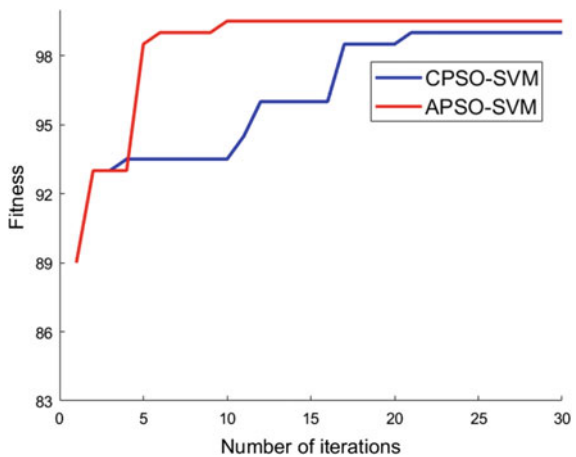
3.2 APSO Algorithm Optimization of SVM

When APSO algorithm is used to optimize SVM parameters, the search range of kernel function is set as $[0, 0.01]$, and the search range of penalty factor is set as $[1, 100]$. The specific optimization is as follows:

- Step 1: Set particle size, maximum number of iterations, search dimension, and initialize particle position;
- Step 2: Initialize the parameters (C, γ) of support vector machine and search range;
- Step 3: Calculate the classification error of SVM under the current parameters to update the position of particles;
- Step 4: Take the minimum classification error of SVM as the fitness value, and update the position of particles according to the Eq. (1);
- Step 5: Compare with the fitness value of the last iteration. If it is lower than the original fitness value, no update will be made; otherwise, update the fitness value;
- Step 6: Calculate the cycle until it reaches the maximum number of cycles, output the optimal parameters (C, γ) , take it as the optimal parameter and according to the Eqs. (5–8) establish the SVM model;

Figure 1 shows the fitness curve of CPSO and APSO. After 20 iterations of CPSO algorithm, the fitness reaches 99 while APSO reach 99.5 after 10 iterations. Besides, it can be seen from the fitness that the APSO algorithm performs better

Fig. 1 Optimized flow chart



in each iteration, indicating that the improved algorithm has a better capability of optimization.

4 Experimental Analysis

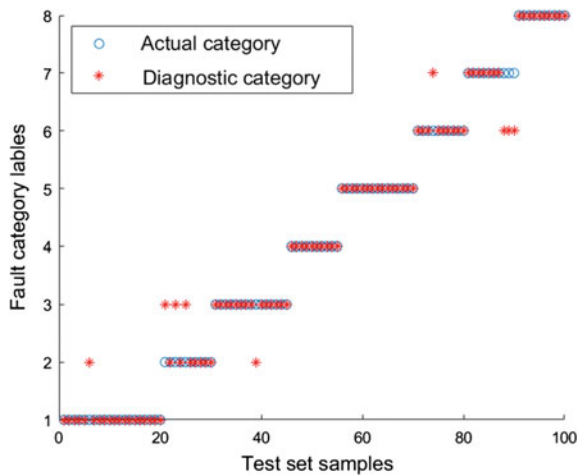
4.1 Comparison of Classification Results of Different Classifiers

The data in this paper are from the sensor data set [8] published by Intel Lab, and the required experimental data samples are obtained by fault injection [9] according to the existing methods. 300 groups of samples are selected, of which 200 groups are used as the training set and 100 groups as the test set. The four classifiers all used the same samples for training and testing. Samples distribution is shown in Table 1.

Table 1 Distribution of samples data

Fault type	Training set	Testing set
Spike fault	46 groups	20 groups
Drift fault	13 groups	10 groups
Bias fault	23 groups	15 groups
Random fault	12 groups	10 groups
Stuck fault	38 groups	15 groups
Erratic fault	20 groups	10 groups
Data loss fault	18 groups	10 groups

Fig. 2 Diagnosis results for the CPSO-SVM



It can be seen from the Figs. 2 and 3, the CPSO-SVM has misclassified multiple types of faults, suggesting the lowest ability to identify the faults. The APSO-SVM makes lower errors, and the performance is better than the CPSO-SVM.

More detailed diagnostic results can be seen from Table 2 and Fig. 4, PNN has the largest number of wrong classification groups, with an accuracy rate of only 87%. DT and CPSO-SVM have the same number of error classification groups, and have high classification accuracy rate. APSO-SVM has the least number of error classification groups, and the classification accuracy rate is 94%, which fully demonstrates the effectiveness of the proposed method in sensor fault identification. However, it can also be seen from the table that the performance of the classifier cannot be comprehensively evaluated by the accuracy rate. Therefore, some complete performance evaluation indexes Precision and Kappa coefficient are added.

4.2 Performance Comparison of Classifiers

In dealing with unbalanced samples of multiple classifications, the Precision and Kappa coefficient can be used to evaluate the performance of the classifier more comprehensively [10]. Precision reflects the classifier’s ability to correctly classify

Fig. 3 Diagnosis results for the APSO-SVM

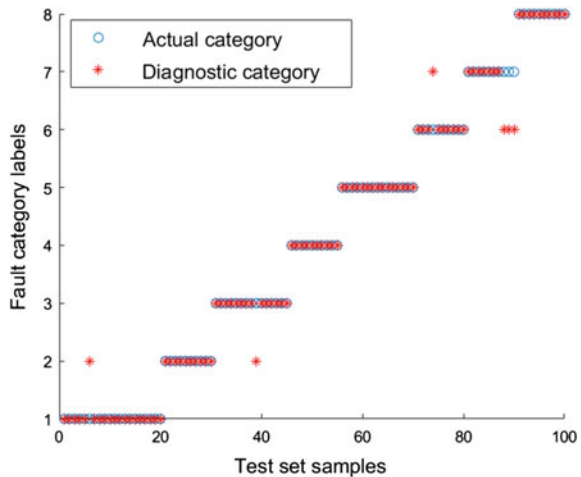
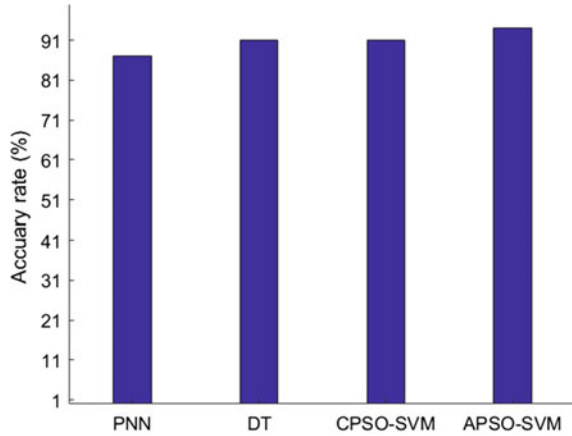


Table 2 Comparison of classification results

Classifier	Misclassification	Accuracy (%)
PNN	13 groups	87
DT	9 groups	91
CPSO-SVM	9 groups	91
APSO-SVM	6 groups	94

Fig. 4 Comparison of accuracy rate



each fault. The greater value shows the stronger ability to identify different faults. The Kappa coefficient reflects the classifier’s overall ability to classify faults. The larger value shows the better classification performance. The mathematical equations for the Precision and Kappa coefficient are as follows:

Precision: Calculate the accuracy for each fault label separately, and then take the unweighted average.

$$P = \frac{T_P}{T_P + F_P} \tag{9}$$

where, T_P is the number of positive classes predicted to be positive, and F_P is the number of negative classes predicted to be positive.

Kappa coefficient:

$$K = \frac{P_0 - P_e}{1 - P_e}$$

$$P_e = \frac{\sum (a_1 \cdot b_1 + a_2 \cdot b_2 + \dots + a_c \cdot b_c)}{n^2} \tag{10}$$

where P_0 is the classification Precision of all samples, a_c is the actual number of class samples, b_c is the number of class c samples recognized by the classifier, and n is the total number of all samples.

As shown in Table 3 and Figs. 5 and 6, Precision of PNN is 85.63%, and Kappa coefficient is 85.43%. The Precision of DT is 91.11% and Kappa coefficient is 88.79%. Precision of CPSO-SVM is 89.79%, and the Kappa coefficient is 89.91%. The Precision of APSO-SVM is 93.54%, and the Kappa coefficient is 93.27%. analyzing the performance of the classifier from the Precision and Kappa coefficient, it can be seen that DT is better than CPSO-SVM in identifying different faults,

Table 3 Comparison of precision and kappa coefficient

Classifier	Precision (%)	Kappa coefficient (%)
PNN	85.63	85.43
DT	91.11	88.79
CPSO-SVM	89.79	89.91
APSO-SVM	93.54	93.27

Fig. 5 Comparison of precision

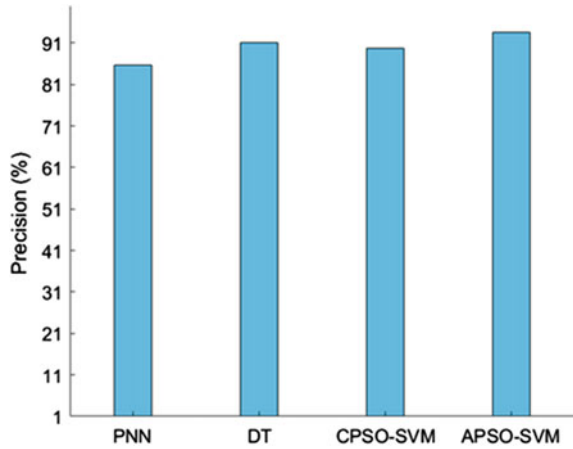
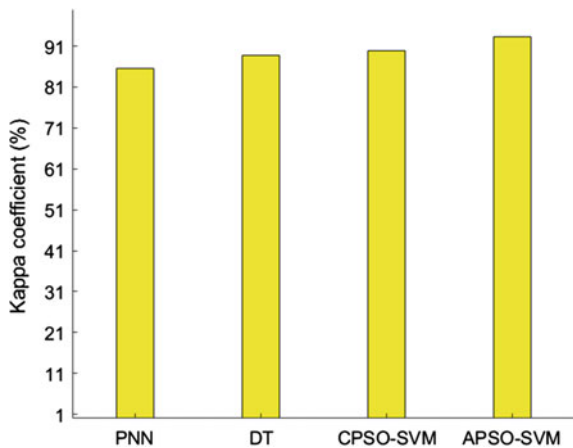


Fig. 6 Comparison of kappa coefficient



while CPSO-SVM is better than DT in overall classification performance. From the above performance analysis, it can be seen that APSO-SVM’s ability to identify different faults and overall classification performance have been greatly improved, which fully proves the effectiveness of the method proposed in this paper.

5 Conclusion

Accurate identification of fault types is an important link to improve the reliability of sensor fault diagnosis. The main contributions of sensor fault identification methods proposed in this paper are summarized as follows:

- In order to improve the accuracy of sensor fault pattern recognition, a data-driven intelligent diagnosis algorithm is proposed in this paper. The experimental results show that this method can accurately identify multiple fault types of sensors and improve the reliability of diagnosis.
- In order to obtain the effective information from the sensor fault data, several time-domain parameters are selected. The experimental results show that the time-domain parameter method can effectively extract the useful information from the fault data, so as to improve the accuracy of fault classification.
- An APSO algorithm is proposed to optimize the parameters of SVM. This algorithm sets the weight of the initial velocity of particles and introduces individual mutation particles. Compared with the novel particle swarm optimization algorithm, the improved algorithm has better optimization ability. The performance analysis of several classifiers shows that the proposed method has better ability of fault identification.

References

1. Liu, X.L., Zhao, Y., Wang, W.J., Ma, S.X., Zhuang, J.: Photovoltaic self-powered gas sensing: a review. *IEEE Sens. J.* **21**(5), 5628–5644 (2021)
2. Zhou, D.H., Zhao, Y.H., Wang, Z.D., He, X., Gao, M.: Review on diagnosis techniques for intermittent faults in dynamic systems. *IEEE Trans. Industr. Electron.* **67**(3), 2337–2347 (2020)
3. Lu, N.N., Xiao, H.H., Sun, Y.J., Han, M., Wang, Y.F.: A new method for intelligent fault diagnosis of machines based on unsupervised domain adaptation. *Neurocomputing* **427**, 96–109 (2021)
4. Qian, J.G.: Diagnosis of machinery fault by time-domain parameters. *Coal Mine Mach.* **27**(09), 192–193 (2006)
5. Cheng, Z., Zhang, J., Lv, Y.L., Xu, H.W.: Wafer yield improvement based on enhanced particle swarm optimization. *Comput. Integrat. Manuf. Syst.* 1–17 (2021)
6. Wang, L.Y., Jin, B.X., Xu, G.Y.: Particle swarm optimization algorithm with dynamic inertia factors for inversion of fault parameters. *Geomatics Inf. Sci. Wuhan Univer.* **46**(04), 510–519 (2021)
7. Cheng, X.Z., Wang, C.A., Li, J.M., Bai, X.Z.: Adaptive fault diagnosis of motors using comprehensive learning particle swarm optimizer with fuzzy petri net. *Comput. Inf.* **39**(1–2), 246–263 (2020)
8. Intel Lab Data. Available online: <http://db.csail.mit.edu/labdata/labdata.html> Accessed on 13 June 2019
9. Benchmark datasets for fault detection and classification in sensor data. In Proceedings of the 5th International Conference on Sensor Networks, Orlando, FL, USA, 30 October–2 November 2016, pp. 185–195

10. Das, S., Mukherjee, H., Roy, K., Saha, C.K.: Shortcoming of visual interpretation of cardiocography: a comparative study with automated method and established guideline using statistical analysis. *SN Comput Sci* **1**(3), 180–190 (2020)

Correlation Analysis of Sensor Fault Based on Fuzzy Petri Net and Apriori Algorithm



Chuannuo Xu, Shenglei Zhao, Haitao Hao, Yandong Zhang, Jiming Li, and Xuezhen Cheng

Abstract Due to the complex internal structure of the sensor, the corresponding fault causes are also diverse. Once a fault occurs, the cause of the fault is difficult to determine. This paper proposes a sensor fault correlation analysis method combining fuzzy Petri net (FPN) and Apriori algorithm. First, obtain the typical fault type waveform of the sensor according to the method of fault simulation, calculate its fault waveform characteristics, find out the residual between it and the normal waveform characteristics, and normalize the residual; then, use the modeling method of FPN to establish the correlation analysis model between fault types, fault characteristic indicators and fault modes; finally, the establishment of model weights and transition threshold parameters is achieved through the Apriori algorithm based on association rules. The maintainer can analyze the fault correlation of the sensor through the abnormal waveform of the sensor to preliminarily judge the fault cause, to achieve the purpose of improving the efficiency of maintenance.

Keywords Sensor · Fault causes · Fuzzy petri net · Apriori algorithm · Correlation analysis

1 Introduction

As the degree of integration of sensors becomes higher and higher, the potential factors that cause malfunctions during sensor operation also increase. According to statistics, 60% of industrial accidents are caused by sensor failures [1]. If the sensor's abnormal signal can be used to preliminarily determine the cause of the sensor's failure, this will play an important role in the stable operation of the sensor [2].

C. Xu · S. Zhao · J. Li · X. Cheng (✉)
College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China
e-mail: zhenxc6411@163.com

H. Hao · Y. Zhang
Anju Coal Mine of Shandong Jining Mining Group Co., Ltd, Jining 272000, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of IncoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_61

The Petri net model not only contains rigorous mathematical formulas, but also symbolic and visual expressions. It can be used as a modeling method for distributed concurrent systems and is widely used in formal modeling and system decision-making and fault diagnosis. Xie [3] et al. gave a model based on Petri net and visualized the results of generalized association rule mining. To effectively identify incomplete and uncertain protection and circuit breaker operation information, the article [4] studies an improved fuzzy Petri net (FPN) fault diagnosis model that takes into account the logical correlation of alarm information. The article [5] establishes a complex system fault diagnosis method through the mapping and fusion of maintenance metadata, fault-related itemsets, and system structure relationships, and combining with the dynamic fault reasoning of FPN. Li [6] et al. proposed a conditional state FPN based on association rules. The association rule method of data mining was used to extract the fuzzy rules and confidence of the FPN and used it for industrial process fault reasoning and diagnosis.

However, the confidence, credibility, and other parameters of the FPN model rely on prior experience, which may lead to inaccurate reasoning and misjudgment of faults [7, 8]. Apriori algorithm is one of the most influential data mining algorithms for mining frequent itemsets of Boolean association rules. It is widely used in various fields such as business, network security, and mobile communications [9]. Zhang [10] et al. uses the Apriori association rule algorithm to mine the big data of high-speed EMUs and obtain information related to failures. The article [11] takes the fault characteristic quantity as the front item and the fault type as the back item sets the minimum support and minimum confidence and uses the classic algorithm of Apriori data mining to dig out the association rules between transformer faults and key state variables. Jin [12] et al. established a motor fault correlation analysis model based on the Apriori algorithm and improved the search mode in the algorithm, which effectively improved the computational efficiency of mining strong association rules between fault elements.

The parameter settings of traditional FPN mostly rely on prior experience, there are relatively few studies on sensor fault correlation analysis. This paper proposes a sensor fault correlation analysis method that combines the FPN with the Apriori algorithm. Firstly, the five typical fault types of the sensor are simulated, and the fault correlation analysis model is established by combining the characteristic indicators and the cause of the fault. Then use the Apriori algorithm to mine the confidence parameters of the model, and finally build an experimental platform to verify the validity of the proposed method.

2 Fuzzy Petri Net

2.1 Definition of Fuzzy Petri Net

Fuzzy Petri Net (FPN) is proposed by combining the original Petri Net and fuzzy theory. It can effectively express and analyze the inaccurate fault information in the model. According to different application backgrounds, the FPN model can set corresponding transition thresholds, weights, and transition trigger conditions, thereby constraining the state transition of the network to achieve the purpose of precise deduction, reasoning, and analysis.

Definition 1 The FPN is an 8-tuple:

$$\sum = (P, T, D, I, O, \alpha, f, \beta) \tag{1}$$

- (1) $P = \{P_1, P_2, P_3, \dots, P_n\}$ represents a finite set of places;
- (2) $T = \{T_1, T_2, T_3, \dots, T_m\}$ represents a finite set of transitions;
- (3) $D = \{d_1, d_2, d_3, \dots, d_n\}$ is the set of propositions. $P \cap T \cap D = \emptyset, |P| = |D|$;
- (4) $I : P \times T \rightarrow N$ is the input function. $N = \{0, 1, 2, \dots\}$ is a set of non-negative integers, representing the mapping from the place to the transition. If $I(P, T) = 1$, it means there is a connection between P and T , at this time, P is the input place of T ; if $I(P, T) = 0$, it means there is no connection between P and T , at this time, P is not the input place of T ;
- (5) $O : T \times P \rightarrow N$ is the output function. $N = \{0, 1, 2, \dots\}$ is a set of non-negative integers, representing the mapping from the place to the transition. If $O(P, T) = 1$, it means there is a connection between P and T , at this time, P is the output place of T ; if $O(P, T) = 0$, it means there is no connection between P and T , at this time, P is not the output place of T ;
- (6) $\alpha : P \rightarrow [0, 1]$ assigns a credibility CF to each place, denoted by θ_i ;
- (7) $f : T \rightarrow [0, 1]$ assigns a confidence CF to each place, denoted by μ_i ;
- (8) $\beta : T \rightarrow D$ is the mapping between the place of each node and the proposition.

Definition 2 The rules of occurrence of FPN:

Under the system identification m_1 , for the transition T_i : if $\forall P_j \in I(T_i) : m_1(P_j) = 1 \wedge \alpha(P_j) = \theta_i$ is the number of tokens in P_j under the system identification m_1 , it is called T_i enabled, and the enabled T_i is fired, and the new system identification m_2 is obtained;

- (1) $\forall P_j \in I(T_i) : m_2(P_j) = m_1(P_j) - 1$
- (2) $\forall P_k \in O(T_i) : m_2(P_k) = m_1(P_k) + 1 \wedge \alpha(P_j) = \theta_i \times \mu_i$

2.2 Fuzzy Petri Net Representation of Production Rules

The fault diagnosis method based on FPN has strong parallel reasoning ability. However, there is not only a simple mapping relationship between sensor fault signal characteristic value indicators, fault types, and fault causes but may involve multiple, non-linear and uncertain mapping relationships. This article combines fuzzy production rules with FPN, which can more accurately describe the relationship between fault signal characteristic value indicators, fault types and fault causes, and the transmitted fault information. As shown in Fig. 1, this paper defines fuzzy production rules for the following four forms, which are used to express the conditions for judging the fault type.

- (1) Correspondence rules between single-cause and single-effect form and fuzzy production:

$$if\ p_i\ then\ p_i\ (C = \mu_R)\ i, j = 1, 2, \dots, n$$

According to Table 1, under the current mark M , after the transition t occurs,

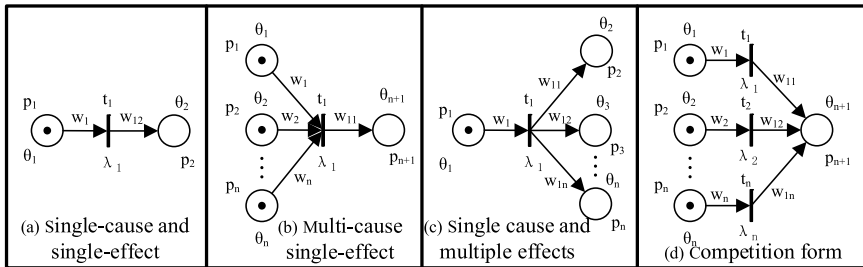


Fig. 1 Four production rules of fuzzy Petri net

Table 1 The mapping relationship between fuzzy Petri net and fuzzy production rules

FPN	Fuzzy production rules
Place	Premise and conclusion of the rule
Transition	Rule
Place mark	Truth degree of premise and conclusion
Transition threshold	Rule credibility
Transition fired	The credibility of the rules is higher than the support
Fuzzy Petri net model	Rule base
Weights between places and transitions	Proposition support

$$\theta_2 = \begin{cases} w_1 \times \theta_1 \times w_1/\lambda, & p \in t. \\ \theta_1, & p \notin t. \end{cases} \tag{2}$$

- (2) Correspondence rules between multi-cause single-effect form and fuzzy production:

$$if\ p_1 \wedge p_2 \wedge \dots \wedge p_{n-1}\ then\ p_n\ (C = \mu_R)$$

According to Table 1, under the current mark M , after the transition t occurs,

$$\theta_{n-1} = \begin{cases} w_{11} \times \min(\theta_1 \times w_1, \theta_2 \times w_2, \dots, \theta_n \times w_n)/\lambda, & p \in t. \\ \theta_i, & (i = 1, 2, \dots, n),\ p \notin t. \end{cases} \tag{3}$$

- (3) Correspondence rules of fuzzy production with single cause and multiple effects:

$$if\ p_1\ then\ p_2 \wedge p_3 \wedge \dots \wedge p_n\ (C = \mu_R)$$

According to Table 1, under the current mark M , after the transition t occurs,

$$\theta_i\ (i = 2, 3, \dots, n) = \begin{cases} w_1 \times \theta_1 \times w_{11}/\lambda_1, & p_2 \in t. \\ w_1 \times \theta_1 \times w_{12}/\lambda_1, & p_3 \in t. \\ \dots \\ w_1 \times \theta_1 \times w_{1n}/\lambda, & p_n \in t. \\ \theta_1, & p_n \notin t.\ (i = 2, 3, \dots, n) \end{cases} \tag{4}$$

- (4) Correspondence rules of competition form and fuzzy production:

$$if\ p_1 \vee p_2 \vee \dots \vee p_{n-1}\ then\ p_n\ (C = \mu_R)$$

According to Table 1, under the current mark M , after the transition t occurs,

$$\theta_{n+1} = \begin{cases} \max(w_1 \times \theta_1 \times w_{11}/\lambda_1, w_2 \times \theta_2 \times w_{21}/\lambda_2, \dots, w_n \times \theta_n \times w_{n1})/\lambda_n, \\ p_{n+1} \in t, & (i = 1, 2, \dots, n) \\ \theta_i, & p \notin t, p_{n+1} \in t, (i = 1, 2, \dots, n) \end{cases} \tag{5}$$

In Eqs. (2)–(5), p_1, p_2, \dots, p_n is the place, which is also the name of the fuzzy rule, t_1, t_2, \dots, t_n is the transition, which is also the fuzzy rule, w_1, w_2, \dots, w_n is the weight and the support of the fuzzy rule, $\theta_1, \theta_2, \dots, \theta_n$ is the mark, and $\lambda_1, \lambda_2, \dots, \lambda_n$ is the transition corresponds to the threshold.

As shown in Table 1, the mapping relationship between fuzzy production and FPN. Assuming that the precondition of the fuzzy production rule reaches the prescribed lower limit, that is, the transition threshold, the transition in the FPN model is fired,

and the truth degree of the conclusion place can be obtained. Based on the article [2] and expert experience, this paper establishes the FPN sensor fault correlation analysis model, as shown in Fig. 4 in the appendix, and the semantic explanation is shown in Table 5.

3 Data Acquisition

3.1 Raw Data Acquisition

This paper selects digital temperature sensors for research. By collecting and recording the raw time-domain data of 54 temperature sensors deployed in the Intel Berkeley Research lab from February 28 to April 5, 2004, this paper selects the data collected for five consecutive days. research. The fault waveforms of offset faults, drift faults, spike faults, precision drop faults, and stuck faults obtained by the method of fault simulation are shown in Fig. 2.

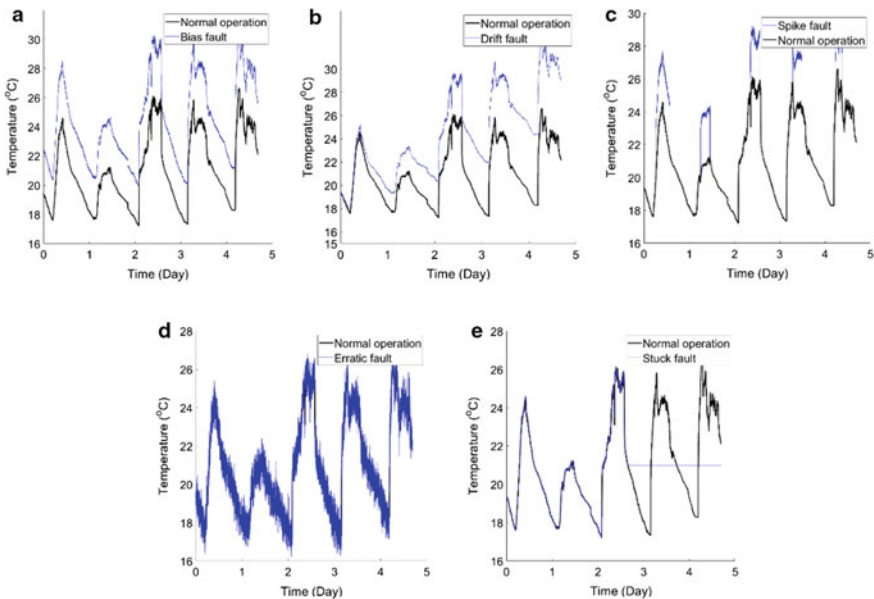


Fig. 2 Simulated normal and fault waveforms (bias fault, drift fault, spike fault, erratic fault, and stuck fault)

Table 2 Residuals of time-domain characteristic indicators after normalization

Fault type	Peak value	Crest factor	Skewness coefficient	Kurtosis coefficient
Bias fault	0.9365	0	0	0
Drift fault	0.9889	0.0503	0.0336	0.0718
Spike fault	0.8819	0.0656	0.1748	0.1937
Erratic fault	0.3803	0.0196	0.0105	0.0587
Stuck fault	0.3246	0.0028	0.1536	0.8019
Normal	0	0	0	0

3.2 Feature Extraction and Residual Calculation

This paper extracts the characteristic values of each group of fault data and normal data, in turn, calculates the characteristic residuals and normalizes them. The characteristic indicators include peak value, crest factor, skewness coefficient, and kurtosis coefficient. The calculation results are shown in Table 2.

The residual calculation of fault and normal waveform characteristic indicators:

$$\delta = d - \Delta d \tag{6}$$

Normalize the residuals of the fault characteristic indicators:

$$d' = 0.7 \cdot \arctan |\delta| \tag{7}$$

4 Fault Correlation Analysis

4.1 Association Rules

Sensing systems usually have a high degree of integration, and the network connectivity of structure and function is complex. There may be multiple, non-linear, and uncertain associations between fault characteristics and fault causes, which often involve more than simple mapping relationships.

Definition 3 The association rule R can express the constructive formula, that is, the implicit formula,

$$R : A \Rightarrow B, \text{ where, } A \subset I, B \subset I, \text{ and } A \cap B = \emptyset.$$

Definition 4 The degree of support of the association rule R ,

$$\text{support}(A \Rightarrow B) = \frac{\text{count}(A \cup B)}{|D|} = P(A \cup B) \tag{8}$$

Definition 5 The confidence level in the association rule R ,

$$confidence(A \Rightarrow B) = \frac{support(A \cup B)}{support(A)} = P(B|A) \tag{9}$$

The confidence in the association rule R is the situation that B occurs under the criterion of A , that is, the conditional probability, and it is also the confidence in the fuzzy rule. As shown in Fig. 1 and Table 1, according to the corresponding relationship between fuzzy production rules and FPN, the characteristic indicators of a large number of sensor data collected can be associated with the cause of the fault.

4.2 Apriori Data Mining Algorithm

The Apriori algorithm belongs to the support-confidence framework standard system. As a traditional algorithm for mining frequent itemsets of association rules from big data, it is suitable for mining association rules of discrete data. The core of the algorithm is to carry out traversal search and layer-by-layer iteration for the two-stage frequent itemsets. First, all frequent itemsets in the database are traversed, and then the frequent itemsets are obtained and the association relationship between each item set is mined, as shown in Tables 3 and 4.

Table 3 Mined rules and corresponding confidence (a)

Rule	Confidence	Rule	Confidence	Rule	Confidence
$P_a \rightarrow P_e$	0.1622	$P_a \rightarrow P_g$	0.1081	$P_d \rightarrow P_i$	0.3793
$P_b \rightarrow P_e$	0.3529	$P_c \rightarrow P_g$	0.1111	$P_a, P_b, P_c, P_d \rightarrow P_e$	1
$P_c \rightarrow P_e$	0.1667	$P_a \rightarrow P_h$	0.3243	$P_a, P_c \rightarrow P_f$	0.6
$P_d \rightarrow P_e$	0.2069	$P_d \rightarrow P_h$	0.4138	$P_a, P_c \rightarrow P_g$	0.16
$P_a \rightarrow P_f$	0.4054	$P_b \rightarrow P_i$	0.6471	$P_a, P_d \rightarrow P_h$	1
$P_c \rightarrow P_f$	0.4167	$P_c \rightarrow P_i$	0.3056	$P_b, P_c, P_d \rightarrow P_i$	0.6471

Table 4 Mined rules and corresponding confidence (b)

Rule	Confidence
$P_a, P_b, P_c, P_d, P_e \rightarrow P_1/P_2/P_3/P_4/P_{21}/P_{22}$	0.1667
$P_a, P_c, P_f \rightarrow P_1/P_4/P_5/P_6/P_7/P_8/P_9/P_{10}/P_{11}/P_{12}/P_{16}/P_{17}/P_{18}/P_{21}/P_{22}$	0.0667
$P_a, P_c, P_g \rightarrow P_2/P_3/P_{21}/P_{22}$	0.25
$P_a, P_d, P_h \rightarrow P_2/P_3/P_{11}/P_{12}/P_{13}/P_{14}/P_{15}/P_{16}/P_{17}/P_{18}/P_{19}/P_{20}$	0.0833
$P_b, P_c, P_d, P_i \rightarrow P_2/P_3/P_7/P_8/P_9/P_{10}/P_{12}/P_{17}/P_{18}/P_{19}/P_{20}$	0.0909

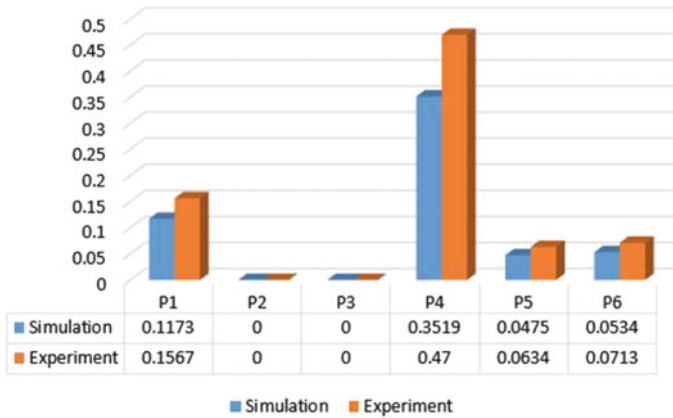


Fig. 3 Confidence comparison between simulation and experiment

4.3 Experimental Platform Verifications

In this paper, the Jingchuang brand RC-4 digital temperature sensor is used for experimental research. According to the characteristics of five typical fault types, this paper simulates the waveforms of the fault types and collects, and the waveform image is drawn. The sampling time of each fault sample is set to 24 h, and the sampling time of indoor temperature and outdoor temperature each account for 50%. The normal indoor temperature and outdoor temperature are combined as the standard normal temperature, and the fault type of the sensor is simulated during indoor data collection.

This paper takes the drift fault when the induction circuit fails as an example. As shown in Fig. 3, the experiment and simulation data are basically consistent. Covariance is used as a method to measure the total error of two variables in probability theory and statistics, as shown in Eq. (10), The calculated covariance of the two sets of vectors at this time is 0.0227, which is in line with the expected result.

$$\text{cov}(\sum_{i=1}^n X_i \sum_{j=1}^m Y_j) = \sum_{i=1}^n \sum_{j=1}^m \text{cov}(X_i, Y_j) \tag{10}$$

5 Conclusion

Due to the traditional FPN parameter setting mostly relies on expert experience, this article takes the temperature sensor as an example, and introduces the Apriori

algorithm based on association rules, establishes an FPN model including fault characteristic indicators, fault types and causes, and obtains the early fault reason of the sensor through data mining algorithm. Experimental simulations and covariance calculations of the data demonstrate the feasibility and effectiveness of the proposed method.

Appendix

See (Fig. 4; Table 5).

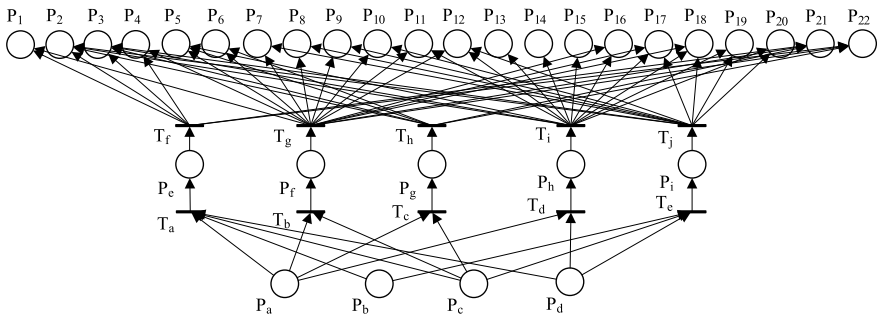


Fig. 4 FPN sensor fault correlation analysis model

Table 5 Semantics of FPN sensor fault correlation analysis model

Place	Semantics	Place	Semantics	Place	Semantics
P_a	Peak value	P_3	Open circuit	P_{14}	Insufficient high level
P_b	Crest factor	P_4	Electrical performance failure	P_{15}	Insufficient low level
P_c	Skewness coefficient	P_5	Parameter hyperparameter	P_{16}	Mechanical failure
P_d	Kurtosis coefficient	P_8	Overvoltage stress	P_{17}	Lightly doped drain is too high
P_e	Bias fault	P_7	Improper rise time	P_{18}	Function failure
P_f	Drift fault	P_8	Switching time failure	P_{19}	Metal defect
P_g	Spike fault	P_9	No output	P_{20}	Out of rating
P_h	Erratic fault	P_{10}	Energy loss	P_{21}	Decreased output capacity
P_i	Stuck fault	P_{11}	Assembly defect	P_{22}	Discontinuous work
P_1	Low gain	P_{12}	Leakage		
P_2	Short circuit	P_{13}	Superposition failure		

References

1. Qiaoning, Y.: Research on Sensor Fault Diagnosis Method Based on Entropy. Beijing University of Chemical Technology (2016).
2. Daoliang, L., Ying, W., Jinxing, W., et al.: Recent advances in sensor fault diagnosis: a review. *Sens. Actuators, A* **309**, 111990 (2020)
3. Zhenhua, X., Fengchao, Z., Dan, T.: Research on visual data mining based on Petri net. *J. Liaocheng Univer. (Nat. Sci. Ed.)* **23**(03), 96–99 (2010)
4. Zhan, B., Shihong, M., Yanbin, S., et al.: An improved fuzzy Petri net fault diagnosis model considering the logical correlation of alarm information. *Electric. Appl.* **34**(16) (2015)
5. Peijie, L., Bo, Y., Hongguang, L.: Conditional state fuzzy petri Net based on association rules and its application in fault diagnosis. *J. Chem. Ind.* **69**(08), 3517–3527 (2018)
6. Rui, L., Linfu, S.: Complex system fault diagnosis method based on multi-source information fusion fault tree and fuzzy petri net. *Comput. Integrat. Manuf. Syst.* **23**(08), 1817–1831 (2017)
7. Manuel, S.: Half a century after Carl Adam Petri's Ph.D. thesis: A perspective on the field. *Annual Reviews in Control*, **37**(2), 191–219 (2013)
8. Hu-Chen, L., Jian-Xin, Y., Zhiwu, L., et al.: Fuzzy petri nets for knowledge representation and reasoning: a literature review. *Eng. Appl. Artif. Intell.* **60**, 45–56 (2017)
9. Xiuxiu, C., Kai, L., Shuangtao, M., et al.: Correlation analysis of aviation equipment failure based on Apriori algorithm. *J. Shandong Normal Univer. (Nat. Sci. Ed.)*, **34**(01) (2019)
10. Chun, Z., Jing, Z.: Research and application of optimization algorithm for mining fault association rules of EMU. *Comput. Modern.* **09**, 74–78 (2017)
11. Xiang, D., Bi, Z., Ruicheng, D., et al.: Analysis of association rules of power transformer fault types and key state quantities. *Zhejiang Electric Power* **39**(02), 23–27 (2020)
12. Qing, J., Yongmei, Y., Lingwen, M., et al.: Correlation analysis of motor faults based on improved Apriori algorithm. *Software* **41**(11) (2020)

Review on Simulation and Optimization of Vehicle Ride Comfort Based on Suspension Model



Tang Jianghu, Xiong Qing, Zhu Yingmou, and He Zhuoyu

Abstract At present, commercial software of the multi-body dynamics is widely used in the research of vehicle ride comfort simulation and optimization. This paper reviews some literatures on vehicle ride comfort optimization based on ADAMS, and focuses on the research based on suspension models. The suspension rigid-flexible coupling models and the simulation research about optimizing suspension model parameters to achieve multi-objective optimization are the main areas of concern. Finally, the paper is summarized and the future trend of ADAMS applied to the simulation and optimization of vehicle ride comfort is prospected.

Keywords Ride comfort · Suspension models · Multi-body dynamics · ADAMS

1 Introduction

Ride comfort is mainly to keep the vibration and impact on the occupant's comfort within a certain range while the car is moving [1]. Suspension is one of the important parts of modern automobile, and plays a key role in ensuring vehicle ride comfort [2]. It connects the vehicle frame and the axle, transfer the force and torque between the wheel and the frame, and buffer the impact force transferred from the uneven road to the frame or body. Analyzing the dynamic characteristics of suspension system can improve its transmission characteristics and reduce vibration and impact. The effect of suspension characteristics on vehicle ride comfort can be studied quickly and effectively by using mature multi-body dynamics software, because it can shorten

T. Jianghu · X. Qing (✉) · H. Zhuoyu
School of Automobile and Transportation, Xihua University, Chengdu 610039, China
e-mail: xiong_qiming@163.com

T. Jianghu · X. Qing · Z. Yingmou
State Key Laboratory of Heavy Duty AC Drive Electric Locomotive Systems Integration,
Zhuzhou 412001, China

X. Qing
Key Laboratory of Automotive Measurement, Control and Safety, Xihua University, Chengdu
610039, China

the research and development cycle and save related costs [3]. Many studies have been carried out on vehicle ride comfort simulation at home and abroad, and multi-body dynamics simulation has been proved to be an effective method for vehicle ride comfort research [4–6].

The main structure of the automobile suspension is a multi-bar system, which is connected with the body by various hinges. This kind of structure can be modeled by means of computational multi-body dynamics. With the maturity of multi-body dynamics theory, a series of multi-body commercialized software have been successively developed, which provides an effective way to suspension kinematics and dynamics calculation and design [7]. In all the commercial software of the multi-body dynamics, ADAMS is the most authoritative one with the highest share, the largest number of applied industries in the field of CAE [8]. It can quickly establish complex suspension and vehicle virtual prototype models, and accurately simulate the driving conditions of the car in different road conditions according to the real working conditions. Moreover, it can analyze and compare various design strategies to ultimately improve the vehicle ride comfort. In recent years, in order to simulate the ride comfort of the whole vehicle, many scholars used ADAMS to model and solve the suspension system [9–11].

2 Rigid and Rigid-Flexible Coupling Suspension Models

In the early stage, multi-body suspension models were mainly rigid models. The dynamic frequency range expressed by this method is relatively low, and only the low frequency mode can be studied. With the development of multi-flexible body dynamics, the theories and methods based on rigid-flexible coupling analysis are becoming more and more mature [12]. Suspension modeling is changed to more complex rigid-flexible coupling modeling. Generally, it is necessary to build a finite element model for some elastic parts, generate the elastic characteristic files, and then insert them into the multi-body software to build the rigid-flexible coupling multi-body model [13].

2.1 Rigid Suspension Models

The components of the rigid suspension models are all rigid bodies, and the influence of flexible deformation of key parts of the suspension is not considered.

Jiang [14] used ADAMS/Car to build a car model with double wishbone suspension as front suspension and a car model with McPherson independent suspension as front suspension, respectively. Ride comfort simulations were carried out at 60 and 100 km/h on B-class road surface, respectively, and root mean square (RMS) of total weighted acceleration was taken as the evaluation index. The results showed that the model with double wishbone suspension has good ride comfort at low speed

and high speed, RMS of total weighted acceleration both were less than 0.315 m/s^2 . While the model with McPherson suspension has poor ride comfort, because the RMS of weighted acceleration was greater than 0.315 m/s^2 at high speed. Jiang [15] established the equal arm balance suspension model and vehicle dynamics model of the rear axle of a heavy-duty van in ADAMS/View. The RMS obtained from simulation and experiments are compared when driving on B-class road at different speeds, and the deviation between simulation and the experimental value was less than 11%. Chen [16] also established the ADAMS model based on balanced suspension. The RMS values of the vehicle's vertical acceleration when the balanced suspension on was 15.75% lower than that of the vehicle with the balanced suspension off. It has been proved that the balanced suspension can improve the vehicle ride comfort by as much as 15.9%. Li [17] established the simplified 1/4 vehicle model in ADAMS, and set up the passive suspension model and air suspension model, respectively. The results indicated that compared to the traditional passive suspension, air suspension working space was reduced 17.14%, vehicle ride comfort and road friendliness of comprehensive performance improvement of 9.26%.

However, the influence of flexible bodies were not considered in the above research, the simulation accuracy of the rigid suspension models may not reflect the real working conditions of the suspension, and the accuracy of vehicle ride comfort simulation need to be improved.

2.2 Rigid-Flexible Coupling Suspension Models

To solve the above problems, many scholars introduced more flexible bodies into the dynamic simulation of suspension to form the rigid-flexible coupling model of vehicle, so as to improve the accuracy of ride comfort simulation.

Wu [18] used Adams/Car to establish a multi-rigid body model of the vehicle and a rigid-flexible coupling model, respectively. Then, the total values of the weighted RMS acceleration for the two models at different speeds under the B-class road were compared. It found that the rigid-flexible coupling model had smaller total weighted RMS values which ranged from 0.315 to 0.63 m/s^2 , relative difference at 2.09 to 8.00%. However, the model only considered the flexibility of the rear suspension. Therefore, Ou [19] further increased the flexibility of key components in the suspension, considered the lower arm of the front suspension, the lateral stabilizer bar and the torsion beam of the rear suspension as flexible bodies, and established the rigid-flexible coupling vehicle model. ADAMS/Car was used to compare the vertical, lateral and longitudinal vibrations of the driver's position simulated under the B-class road, and the total weighted acceleration RMS of vibration was taken as the ride comfort evaluation parameter. Under the same working conditions, the total weighted RMS values of rigid-flexible coupling model and rigid coupling model were 0.342 and 0.653 m/s^2 , respectively, and the difference between them increased as speed increased. Duan [20] took advantage of ADAMS/Car to flexibility the lower arm and lateral stabilizer bar of the front and rear suspension, and combined with

other subsystems to build a rigid-flexible coupling model of a micro electric vehicle. By comparing the vertical acceleration curves obtained from the simulation and the real vehicle experiment, it can be seen that the trend of the two curves is basically the same, which reflects the correctness of the model and further proves that the horizontal stabilizer bar can prevent the problem of low vertical stiffness caused by improving the vehicle ride comfort. Zhang [21] implemented flexible operation of the front suspension, rear suspension and frame of the vehicle, established the rigid body, rigid-flexible coupling model and the simulation road surface of an off-road vehicle, respectively, and carried out the ride comfort simulation in ADAMS/Car. Compared with the real vehicle road tests, it was found that the deformation of suspension and frame in the rigid-flexible coupling model reduced the fluctuation between the peaks of the vibration acceleration curve on pulse road, while the attenuation of the rigid-body model was greater. On the C-class road, the weighted acceleration RMS of the rigid-flexible coupling model was larger than that of the rigid model, which was closer to the road test value. Wang [22] compared the rigid model, the rigid-flexible coupling model and the real vehicle experiment, and found that the acceleration curve amplitude and peak power spectral density (PSD) of the rigid-flexible coupling model were reduced, as well as the simulation accuracy was higher. Moreover, the Insight module of ADAMS was used to optimize and match the spring stiffness and shock absorber damping of the front and rear suspension, and the weighted acceleration RMS was taken as the ride comfort optimization objective to compare the results before and after optimization. The results show that the evaluation index after optimization was obviously reduced, from 0.216 to 0.201 m/s^2 at 40 and 0.341 m/s^2 to 0.316 m/s^2 at 80 km/h. Obviously, the ride comfort of the car was improved.

These studies prove that the rigid-flexible coupling multi-body model based on the suspension model can improve the simulation accuracy of ride comfort, and provide a direction for the refined study of ride comfort. However, it is far from sufficient to improve vehicle ride comfort through rigid model or rigid-flexible coupling model, and appropriate optimization algorithm should be combined to optimize model parameters in order to obtain better vehicle ride comfort (Table 1).

3 Multi-objective Optimization Based on Suspension Models

Suspension optimization is a typical multi-objective problem. The main problem is the contradiction between handling stability and ride comfort. In order to reconcile this contradiction, many scholars have paid key attention to this aspect.

Table 1 Ride comfort simulation based on rigid-flexible coupling models

References	Flexible component	Simulation of condition	Evaluating indicator	Conclusion
[18]	Rear suspension	ISO 2631-1, Class B road	Acceleration RMS at seat and floor	Comparing the RMSs of the total acceleration at 40-80 km/h speed, the percentage decrease is 2.09, 2.73, 4.11, 5.19 and 8.00% respectively
[19]	Lower arm of front suspension, horizontal stabilizer bar, rear suspension torsion beam	GB/T4970-2009, Class B road	RMS of total weighted acceleration	Comparing the RMSs of the total weighted acceleration at 40-70 km/h, the decreasing percentages are 2.8, 4.4, 4.11, 5.0 and 6.8% respectively
[20]	Lower arm, horizontal stabilizer bar of front and rear suspensions	Sinusoidal excitation experiment of single excitation for front and rear wheels	Vertical acceleration at driver's seat	The acceleration curve is in good agreement with the real vehicle test. The lateral stabilizer bar can effectively improve the vehicle roll performance and ensure good ride comfort
[21]	Upper and lower swing arm and torsion bar spring of front suspension, leaf spring of rear suspension	GB/T4970-2009, Class C road	RMS of center weighted acceleration of seat floor	The vertical acceleration response curve of the rigid flexible coupling model is smoother than that of the rigid model when passing through the pulse road at 60 km/h, and the peak value of RMS of weighted acceleration is smaller

(continued)

Table 1 (continued)

References	Flexible component	Simulation of condition	Evaluating indicator	Conclusion
[22]	Lower arm of front suspension, horizontal stabilizer bar, rear suspension torsion beam, body	GB/T4970-2009, Class B road	RMS of total weighted acceleration	Compared with the RMS of total weighted acceleration, it decreased by 6.95% at 40 km/h, 17.6% at 60 km / h and 13.45% at 80 km / h

3.1 Multi-objective Optimization Using NSGA-II Algorithm

The fast non-dominated sorting genetic algorithm based on elite strategy (NSGA-II) was first proposed by Deb [23]. With the advantages of low computational complexity and no need to specify the shared radius, it is one of the commonly used algorithms to realize multi-objective optimization of suspensions [24–26].

Wang [27] established a vehicle model in ADAMS, and selected stiffness of front and rear suspensions, damping coefficient of shock absorber and torsional stiffness of stabilizer bar as optimization variables. In order to optimize the vehicle handling stability and ride comfort, a Kriging model was established which responded to the change of weighted RMS of body acceleration with design variables. NSGA-II algorithm was used for multi-objective optimization of suspension system to obtain Pareto optimal solution set, and an optimal solution was selected for ADAMS simulation verification. As a result, the maximum roll angle of the vehicle was reduced by 12.9% and the weighted RMS value of acceleration was decreased by 5.4%. Therefore, the vehicle ride comfort and handling stability had been improved. Similarly, Chen [28] established a Kriging model to describe the relationship between the evaluation indexes of vehicle ride comfort and suspension parameters. NSGA-II algorithm was used for optimization in the model, weighted acceleration, wheel dynamic load and suspension deflection were taken as the evaluation indexes of ride comfort. At 60 km/h, the largest reduction value for the optimized RMS of the overall body acceleration was 0.017 m/s² and corresponding relative change was 8.15%, the largest reduction value for the optimized RMS of the suspension deflection was 0.80 mm and corresponding relative change was 16.14%. Du [29] set up a vehicle model based on ADAMS/Car and took suspension stiffness and shock absorber damping as design variables. NSGA-II genetic algorithm was used to optimize the characteristic parameters of spring and shock absorber. After optimization, the deficient turning increased, the RMS of vertical acceleration decreased from 0.322 to 0.308 m/s², and pitch angle and roll angle decreased from 2.13 to 1.84°, 1.04 to 0.98°, respectively. Zhang [30] took a heavy-duty vehicle as an example to optimize the suspension parameters that affect its dynamic performance, and took the

weighted RMS of acceleration as the evaluation index. By using NSGA-II algorithm, the multi-objective optimization of vehicle ride comfort, stability and road friendliness were realized, and the RMS value of vertical acceleration decreased by about 7%, which provided a reference for parameters design of heavy-duty vehicle. These studies verified the effectiveness of NSGA-II algorithm in solving multi-objective optimization problems of suspension system.

3.2 Multi-objective Optimization Using Other Algorithms

In addition to the NSGA-II genetic algorithm, some new algorithms were also applied to the suspension optimization to improve the ride comfort and other performance of the vehicles. Tey [31] used ADAMS/Car to model a car, the distributed multi-objective estimation algorithm based on regular model (RM-MEDA) was used to optimize the design variables of the vehicle suspension system, and the optimization objectives were RMS of weighted acceleration and average suspension working space. Five groups of optimized vehicle characteristics were selected and compared with the target values of the original vehicle design. It proved that the algorithm could provide a compromise between ride comfort and handling performance.

And beyond that, He [32] built commercial vehicle's a dynamic model in ADAMS/Car and a nonlinear damping model based on MATLAB/Simulink, respectively. Moreover, the particle swarm optimization (PSO) algorithm was used to optimize the ride comfort model based on nonlinear damping. Results showed that the frequency-weighted RMS of driver seat acceleration, RMS values of the suspension working space of the front and rear axles, and RMS values of the dynamic tyre load of front and rear wheels were decreased by an average of 27.4, 21.6, 25.0, 19.3 and 22.3%, respectively. Mahmoodi [33] used ADAMS to establish a double wishbone suspension, and used genetic algorithm (GA) in geometric design to improve maneuverability, stability and ride comfort. Pang [34] used ADAMS/View to establish a 8×4 truck model, the acceleration response characteristics of B-class road spectrum as excitation under full load were simulated and analyzed. Sequential quadratic programming algorithm (OPTDES-SQP) was used to optimize the maximum value of power spectrum density curve of vertical vibration. The optimization results showed that the power spectral density of the vertical acceleration at the center of mass was lower 5.28% than before optimization, and the ride comfort was improved.

On the other hand, the optimization effect comparison between algorithms is also a key point. Shi [35] used ADAMS/Car to establish an A0 vehicle model with a Macpherson suspension as the front suspension, and used the double-loop multi-objective particle swarm optimization particle swarm optimization algorithm (DL-MOPSO), PSO and GA for comparison. The results showed that despite an increased in the variation range of the caster angle by 9.24–10.69% using the DL-MOPSO algorithm, the variation ranges of the toe angle, camber angle, and kingpin inclination angle were observed to be reduced by 50.45–79.39%, 1.84–4.24%, and

Table 2 Ride comfort simulation after the parameter optimization

References	Optimization algorithm	Vehicle simulation test	Evaluation parameters	Conclusion
[27]	NSGA-II	80 km/h hunting test, and random road test of Class B road	Weighted RMS of body vertical acceleration	The maximum roll angle of the car body is reduced by 12.9%, the weighted RMS of the vertical acceleration of the car body is reduced by 5.4%, and the handling stability and ride comfort of the whole car are improved
[28]	NSGA-II	Driving at different speeds on class B roads	RMS of acceleration, Wheel dynamic load, Suspension working space	The maximum reduction of wheel dynamic load is 191.54 N, the maximum reduction of body acceleration RMS is 0.017 m/s ² , and the maximum reduction of suspension working space is 0.8 mm, which improves the ride comfort
[29]	NSGA-II	Steady static circular, Emergency braking input, Ride comfort pulse input	Vertical acceleration at driver's seat	The vertical acceleration decreases by 4.25%, the vehicle pitch angle decreases by 13.62%, and the body roll angle decreases by 5.77%. The ride comfort and handling stability of the vehicle are improved

(continued)

2.11–2.96%, respectively. Accordingly, the DL-MOPSO optimized model had the best ride comfort simulation and improved handling stability (Table 2).

4 Conclusions and Prospect

The above research summarize about simulation and optimization of vehicle ride comfort based on ADAMS suspension model. Mainly in that established the rigid

Table 2 (continued)

References	Optimization algorithm	Vehicle simulation test	Evaluation parameters	Conclusion
[30]	NSGA-II	speed of 30 km/h B-class curved road with a turning radius of 50 m	RMS of vertical weighted acceleration at the seat	RMS of vertical weighted acceleration at the seat is reduced by about 7%, the stability factors of the inner and side tires are increased by about 4.78% and 4.31%, respectively, and the combined force of the fourth power of 95 percentile of the road friendliness index is reduced by about 12.6%
[31]	RM-MEDA	ISO8608: 1995, Class C road, speed 80 km/h	RMS of weighted acceleration, RMS of dynamic travel of front and rear suspension	Compared with the results of five groups of optimization, the fourth group of data is obtained, which can improve the ride comfort and handling stability at the same time
[32]	PSO	GB/T4970-2009, Class C road, speed 30–100 km/h	Weighted RMS of vibration frequency at the seat	The frequency weighted RMS of the driver’s seat acceleration, front and rear suspension dynamic travel and front and rear wheel dynamic load are reduced by 27.4%, 21.6%, 25.0%, 19.3% and 22.3% respectively, and the ride comfort is improved
[34]	OPTDES-SQP	Class B road, Speed 30–60 km/h	RMS of weighted acceleration	The maximum value of the power spectral density curve of vertical vibration of the center of mass of the container decreases by 5.28%, RMS of acceleration is 0.6338 m/s ² , the ride comfort is improved

(continued)

Table 2 (continued)

References	Optimization algorithm	Vehicle simulation test	Evaluation parameters	Conclusion
[35]	DL-MOPSO, PSO, GA	GB/T4970-2009, Class A road, speed 60/80/100 km/h	RMS of vertical acceleration at seat bottom	DL-MOPSO outperforms PSO and GA in handling stability and ride comfort

model and rigid-flexible coupling model of different suspension, and realized the multi-objective optimization of the key parameters of vehicle ride comfort by using NSGA-II algorithm, RM-MEDA, PSO algorithm and so on. It can be seen that ADAMS can easily model and accurately solve the suspension system, and is an important tool for scholars to study the vehicle ride comfort.

It is still a key attention to accuracy of the models for scholars. Both rigid-flexible coupling and parameter optimization are aimed at making the simulation model reflect the real working conditions more accurately. With the development of various modeling theory and the improvement of computational efficiency, ADAMS will more comprehensively consider the system structure factors, working conditions factors and environmental factors, as well as the system modeling method in the case of multi-factor coupling. In the future, vehicle ride comfort modeling should consider various uncertain factors. In practical engineering, there are many uncertain factors such as system parameters, boundary conditions and external loads due to manufacturing error and measurement error. The traditional deterministic model will not be able to reflect the influence of the randomness of the system on the ride comfort. How to consider various uncertain factors and evaluate the influence of system uncertainty on vehicle ride comfort has become an unavoidable problem in vehicle ride comfort research.

Acknowledgements The authors gratefully acknowledge the support of the Centre for Efficiency and Performance Engineering (CEPE), University of Huddersfield, United Kingdom.

Funding This work was supported in part by the Opening Project of The State Key Laboratory of Heavy Duty AC Drive Electric Locomotive Systems Integration (2020ZJKF05), and in part by Natural Science Foundation of Sichuan Province (2022JY0400).

References

1. Yu, Z.: *Automotive Theory*, 6th edn. Machinery Industry Press, Beijing, China (2009)
2. Yang, R., Yuan, Z., Huang, X., et al.: Collaborative optimization of vehicle handling stability and ride comfort. *Automotive Eng.* **31**(11), 1053–1055+1059 (2009)
3. Yang, X., Chen, R.: History and development of vehicle ride comfort simulation modeling. *J. Nanjing Instit. Technol. (Nat. Sci. Ed.)* **5**(3), 66–70 (2007)

4. Yi, F., Zhou, C.: Vehicle ride comfort simulation under pulse road surface based on multi-body dynamics. *Appl. Mech. Mater.* **1151**(97), 1341–1344 (2011)
5. Xiong, J., Huang, J., Liao, Q.: Research and implementation of vibration based on analysis of vehicle ride comfort. *Appl. Mech. Mater.* **1918**(378), 281–284 (2012)
6. LI, C., Zhang, W., Pan, X., et al.: Modeling and simulation analysis of vehicle ride comfort based on ADAMS. *J. Zhengzhou Univer. (Eng. Sci.)* **31**(5), 99–102 (2010)
7. Lin, Y., Ma, T., Yao, W., et al.: Review on vehicle VNH characteristics. *Automot. Eng.* **24**(3), 177–186 (2002)
8. Wang, G., Zhang, J., Ma, R., et al.: *Virtual Prototyping Technology and its Practice*. Northwestern Polytechnical University Press, Xi'an, China (2002)
9. Wang, L., Wei, P., Liang, Y.: Parameter optimization design of double-wishbone front suspension based on ADAMS. *Noise Vib. Control* **39**(4), 120–124 (2019)
10. Zhang, Y., Sun, L.: Study on control technology of active suspension based on ADAMS and MATLAB. *Appl. Mech. Mater.* **3365**(602–605), 1372–1377 (2014)
11. Yang, J., Zhang, G., Zhang, M.: Ride comfort analysis of car suspension parameters on the random road. *Appl. Mech. Mater.* **2141**(248), 185–189 (2013)
12. Sun, J., Tian, Q., Hu, H.: Research progress on dynamic modeling and optimization of flexible multi-body system. *J. Mech.* **51**(6), 1565–1586 (2019)
13. Editorial department of China journal of highway.: review of academic research on automotive engineering in china. *China J. Highway Transp.* **30**(6), 1–197 (2017)
14. Jiang, B., Pan, Y.: Simulation analysis of vehicle ride comfort based on ADAMS_Car_ride. *Mach. Des. Res.* **30**(6), 166–169 (2014)
15. Jiang, H., Zhang, Z., LI, L., et al.: Modeling of equal arm balanced suspension and vehicle ride comfort simulation based on ADAMS. *J. Chongqing Jiaotong Univer. (Nat. Sci. Edition)* **34**(3), 171–174 (2015)
16. Chen, Q., Bai, X., Zhu, A., et al.: Influence of balanced suspension on handling stability and ride comfort of off-road vehicle. *Proc. Instit. Mech. Eng* **235**(6), 1602–1616 (2021)
17. Li, S., Meng, Z., Jiang, W.: Simulation analysis of automobile air suspension dynamics based on ADAMS. In: *Proceedings of the 2015 International Conference on Intelligent Systems Research and Mechatronics Engineering*. Atlantis Press, Paris, France (2015)
18. Wu, G., Fan, G., Guo, J.: Ride comfort evaluation for road vehicle based on rigid-flexible coupling multi-body dynamics. *Theor. Appl. Mech. Lett.* **3**(1), 39–43 (2013)
19. Ou, J., Zhang, Q., Yang, E., et al.: Research on rigid-flexible coupling vehicle ride competitiveness considering suspension flexibility. *Mach. Des. Manuf.* **4**(2), 132–134+138 (2015)
20. Duan, M., Shi, J.: Research on the influence of lateral stabilizer bar on vehicle performance based on rigid-flexible coupling model. *Mach. Des. Manuf.* **4**(4), 110–113 (2016)
21. Zhang, Y., Wang, W., Wang, T., et al.: Rigid-flexible coupling vehicle ride comfort simulation and experimental research. *Automobile Technol.* **4**(5), 20–25 (2014)
22. Wang, T., Ou, J., Zhang, Y., et al.: Simulation analysis and optimization of car ride comfort based on rigid-flexible coupling model. *J. Chongqing Univer. Technol. (Nat. Sci.)* **29**(12), 25–31 (2015)
23. Deb, K., Pratap, A., Agarwal, S., et al.: A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**(2), 182–197 (2002)
24. Li, W., Wang, S., Zhou, B., et al.: McPherson suspension optimization based on response surface methodology and NSGA-II algorithm. *J. Hunan Univer. (Nat. Sci. Ed.)* **38**(6), 27–32 (2011)
25. Gadhvi, B., Savsani, V., Patel, V.: Multi-objective optimization of vehicle passive suspension system using NSGA-II, SPEA2 and PESA-II. *Procedia Technol.* **23**(5), 361–368 (2016)
26. Zheng, R., Zhu, A.: Collaborative optimization of ride comfort and handling stability of a heavy off-road vehicle. *Agric. Equip. Vehicle Eng.* **56**(12), 52–55 (2018)
27. Wang, Z., Wang, L., Yuan, L., et al.: Multi-objective optimization of suspension system of a light truck based on Kriging model. *Mac. Des. Manuf. (S2)*, 31–34+39 (2018)
28. Chen, S., Shi, T., Wang, D., et al.: Multi-objective optimization of the vehicle ride comfort based on Kriging approximate model and NSGA-II. *J. Mech. Sci. Technol.* **29**(3), 1007–1018 (2015)

29. Du, X., Xiong, R., Wu, J., et al.: Multi-objective optimization of chassis based on vehicle handling stability and ride comfort. *Mod. Manuf. Eng.* **5**, 98–101 (2018)
30. Zhang, J., Cui, S.: Multi-objective optimization analysis of dynamic performance of heavy haul vehicles. *J. Beijing Jiaotong Univer.* **42**(3), 120–126 (2018)
31. Tey, J.Y., Ramli, R., Abdullah, A.S.: A new multi-objective optimization method for full-vehicle suspension systems. *Proc. Instit. Mech. Eng. Part D J. Automobile Eng.* **230**(11), 1443–1458 (2016)
32. He, S., Chen, K., Xu, E., et al.: Commercial vehicle ride comfort optimization based on intelligent algorithms and nonlinear damping. *Shock. Vib.* **2019**, 1–16 (2019)
33. Mahmoodi-Kaleibar, M., Javanshir, I., Asadi, K., et al.: Optimization of suspension system of off-road vehicle for vehicle performance improvement. *J. Central South Univer.* **20**(4), 902–910 (2013)
34. Pang, H., Fang, Z., Li, H., et al.: Optimization and experimental study of suspension parameters of a heavy truck. *J. Vib. Shock* **31**(08), 92–95+106 (2012)
35. Shi, Q., Peng, C., Chen, Y., et al.: Robust kinematics design of MacPherson suspension based on a double-loop multi-objective particle swarm optimization algorithm. *Proc. Instit. Mech. Eng. Part D: J. Automobile Eng.* **233**(12), 3263–3278 (2019)

The Application of ADAMS Software to Vehicle Handling Stability: A Review



Li Yixuan, Xiong Qing, Zhu Yingmou, and He Zhuoyu

Abstract Vehicle handling stability is one of the most important performance of automobile. At present, ADAMS software is widely used in the study of vehicle handling stability. Taking suspension as research objects, this paper reviews the progress of ADAMS software in the study of vehicle handling stability, introduces what evaluation indexes scholars have selected and what simulation tests have been carried out to study the vehicle handling stability. The results of simulation experiments of handling stability under different working conditions are summarized, and the contribution of these methods to improving handling stability is analyzed, which provides a useful reference for scholars in related research. Finally, the paper is summarized, and the future trend of ADAMS software applied to the simulation of vehicle handling and stability is prospected.

Keywords Handling stability · Vehicle · Suspension · ADAMS software

1 Introduction

Under the condition that the driver does not feel nervous or tired, the vehicle can follow the direction given by the driver, and can maintain stable driving in the case of external interference, which is called the vehicle handling stability [1]. The quality of the handling and stability performance not only directly affects the driver's driving experience, but also relates to the safety and stability of the vehicle at high speed [2], so it is particularly important to study it. The direct method to evaluate vehicle

L. Yixuan · X. Qing (✉) · H. Zhuoyu

School of Automobile and Transportation, Xihua University, Chengdu 610039, China
e-mail: xiong_qiming@163.com

L. Yixuan · X. Qing · Z. Yingmou

State Key Laboratory of Heavy Duty AC Drive Electric Locomotive Systems Integration, Zhuzhou 412001, China

X. Qing

Key Laboratory of Automotive Measurement, Control and Safety, Xihua University, Chengdu 610039, China

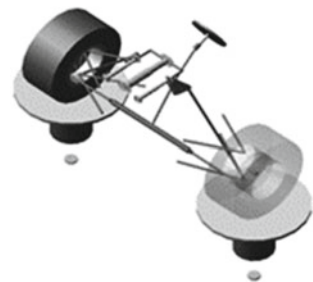
handling stability is actual vehicle test, but repeated test work is costly and exhausting. Moreover, it is difficult to unify the evaluation results, which leads to errors and prolongs the development cycle of automobiles [3]. ADAMS software can automatically establish and quickly solve the equation of motion of vehicle system and its components, so it has been widely used in the study of vehicle handling stability [4].

Suspension is the general term for all the connecting devices between the wheel and the body. Its main role is to transfer the force and torque used between the wheel and the body, so as to ensure the vehicle has good handling stability [5]. So optimization of suspension is the main method for many scholars to improve vehicle handling stability. In this paper, we take suspension as the research objects, review and summarize the literature on vehicle handling stability research using ADAMS software, mainly focusing on the structural design and parameter optimization of suspension.

2 Research on Handling Stability Based on Suspension Structure Design

Spring, as an important part in the design of suspension, has a certain influence on handling stability. Literature [6–9] studied the influence of suspension spring stiffness on the handling stability. Zhang [6] established the vehicle multi-rigid body system dynamics model by using ADAMS/Car. The simulation tests of steering wheel angle step input, steering angle pulse input and single shifter are carried out. The results show that appropriately increasing the front suspension spring stiffness can reduce the amplitude of yaw rate, lateral acceleration and roll angle, so as to improve the vehicle handling stability. Sert [7] and Dong [8] established the whole vehicle dynamics model of a bus in ADAMS/Car, studied the influence of suspension stiffness on handling stability of medium and mini buses, respectively. Liu [9] took FSAE (Formula Society of Automotive Engineers) racing car as the research object, built an interconnected double-wattarm independent suspension in ADAMS, the suspension model is shown in Fig. 1. And studied the influence of adding the third spring on the vehicle performance, took the roll angle as the evaluation index. The step steering and fish-hook simulation tests proved that the third spring can effectively

Fig. 1 Front double wishbone independent suspensions model with the third springs [9]



restrain acceleration head and braking nod in the process of acceleration and braking, and improve the handling stability of the racing car.

The antiroll bar can improve the roll stiffness of the suspension, reduce the roll angle of the vehicle body, improve the understeering characteristics and the roll characteristics of the vehicle body, so as to improve the vehicle handling stability [1]. The influence of antiroll bar on handling stability was studied in literature [10–12]. Xia [10] used ADAMS software to carry out virtual prototype model of suspension guide mechanism and steering bar system for FSAE racing car. Taking side acceleration and lateral acceleration as evaluation indexes, he carried out circular dynamic simulation in ADAMS, and analyzed the influence of antiroll bar with different moment arm lengths on the turning characteristics of racing car, and finally improved the handling stability in the process of car running. Qiu [11] carried out a complete vehicle modeling for a passenger car in ADAMS. Taking the roll angle as the evaluation index, he carried out the simulation test of middle position steering and steady rotation. The simulation results show that with the increase of the stiffness of the antiroll bar, the angle of the frame decreases non-linearly, thus improving the handling stability of the vehicle. Wang [12] established a complete vehicle model of Roewe750FCV in ADAMS. Taking roll angle, lateral acceleration and yaw rate as evaluation indexes, he analyzed the influence of two lightweight antiroll bar schemes on the vehicle performance through angle pulse and angle step input steering conditions. As a result, the antiroll bar of the hollow pipe can reduce the weight and ensure the handling stability of the original vehicle.

Torsion beam is the most important component of torsional suspension, mainly to meet the vehicle roll, vertical and longitudinal movement of the suspension can get the expected stiffness value, to ensure the vehicle handling stability [13]. Wu [14] and Gao [15] established the torsional beam suspension with ADAMS and carried out simulation. Figure 2 shows the suspension model they built, with the torsional beam in red. It turns out that when the opening of the torsion beam is downward about 75° , the understeering characteristic of the vehicle is larger, the roll angle of the center of mass is smaller, and the vehicle's handling stability is better.

The above literature improves the handling stability of the vehicle by changing the suspension structure, and Table 1 summarizes them. They built the virtual vehicle model by using ADAMS, carried out the simulation tests of steering reorientation, steady rotation, steering wheel step input, steering wheel pulse input and single

Fig. 2 Torsional beam suspension model [15]



Table 1 Study on handling stability based on suspension structure design

Literature	Models	Structure design	Evaluation index	Test method	Conclusion
[6]		Suspension spring stiffness	Lateral acceleration, yaw rate, roll angle	Steering wheel angle step input, pulse input	The lateral acceleration, yaw rate and roll angle amplitude can be reduced by increasing the front suspension spring stiffness
[7]	Medium sized bus	Leaf spring stiffness	Roll angle	Fishing-hook and turning condition	The critical value of roll angle can be increased by using larger leaf spring stiffness
[8]	FSAE car	The third spring	Roll angle	Step steering and fishing hooks	The third spring can increase the vertical line stiffness, can effectively suppress the acceleration of head and brake nod
[10]	FSAE car	Antiroll bar	Side acceleration, lateral acceleration	Loop dynamic simulation	The influence of different moment arm lengths on the steering characteristics of the racing car was analyzed, which improved the steering stability of the racing car
[11]	A passenger car	Antiroll bar	Roll angle	The middle position turns, steady turn	With the increase of the stiffness of the antiroll bar, the frame roll angle decreases non-linearly
[12]	Roewe 750 FCV sedan	Antiroll bar	Roll angle, lateral acceleration, yaw rate	Angle pulse, angle step input steering	The antiroll bar of the hollow pipe can reduce the weight and ensure the handling stability of the original car
[14]		Torsion beam	Roll angle gradient, yaw rate	angle step, angle pulse of vehicle	When the opening direction of the torsion beam is downward, the yaw rate response time and peak response time decrease
[15]		Torsion beam	Under steering, sideslip angle of center of mass, yaw rate, lateral acceleration	Steady-state radius rotation, steering wheel sweep input	When the opening of the torsion beam is downward about 75°, the understeering characteristic of the whole vehicle is larger, and the sideslip angle, transient yaw rate and lateral acceleration of the center of mass are smaller

shifter vehicle, and evaluated the handling stability of the vehicle with the yaw rate, roll angle, side acceleration or lateral acceleration as evaluation indexes, and finally obtained the expected simulation results. However, it is not enough to comprehensively improve the vehicle's handling stability only by increasing the stiffness of the suspension spring, installing the transverse antiroll bar and changing the opening direction of the torsion beam. If the geometric parameters of the suspension are optimized with the appropriate control algorithm, it is expected that better handling stability can be obtained.

3 Research on Handling Stability Based on Optimization of Suspension Parameters

Wheel toe-in angle can eliminate tire sideslip caused by camber angle and ensure vehicle handling stability [16]. In literature [17, 18], the suspension was optimized with the toe-in angle as the optimization objective. Wu [17] and Mao [18] established the front McPherson suspension model and the multi-link rear suspension model by using ADAMS/Car, respectively, and optimized the angle of toe-in with ADAMS/Insight. After optimization, the yaw rate of the vehicle decreases obviously, the response time of the diagonal step input decreases, and the vehicle handling stability improved obviously. In addition, the inclination angle can reduce steering control force, improve rebound and deviation, and ensure the stability of the vehicle in a straight line [16]. In ADAMS/Car, Huang [19] established an accurate dual-widarm front suspension model, and optimized the kingpin inclination angle of the suspension with a certain pickup truck as the research object. With the yaw rate, roll angle of front and rear axles, lateral acceleration as the evaluation indexes, the steering wheel angle step input and steady turning simulation experiment were conducted. The results show that the optimization improves the response characteristics of the steering wheel's angular step input and steady-state rotation, and improves the vehicle handling stability and steady-state response characteristics. However, there are many geometrical parameters that affect the performance of suspension, and only one of them is optimized in the above literature. If a suitable method is used for multi-objective optimization of several key parameters, it is expected that better handling stability can be obtained.

At present, multi-objective optimization design of suspension is mainly based on optimization method, and computer simulation and optimization software are used as tools [20]. ADAMS/Insight (Test Design and Analysis Module) is widely used in the optimization of vehicle handling and stability. Many scholars have used it to carry out multi-objective optimization of the geometric parameters of suspension. In literature [21–23], the dual-wheel coaxial excitation condition was used to analyze the suspension, Chen [21] took a mini pure electric vehicle as the research object and established a simulation model of McPherson suspension in ADAMS. Taking the root mean square value of kingpin inclination angle as the optimization objective, they

optimized the front wheel positioning parameters of the designed McPherson suspension based on ADAMS/Insight. The optimization results show that the proper reduction of the kingpin inclination can make the vehicle have better handling stability. Yu [22] and Zhang [23] optimized the hard point coordinates of the suspension with the full factor and response surface method in the ADAMS/Insight. The optimized results are shown in Figs. 3 and 4, It can be seen that the yaw rate is reduced after optimization, and the vehicle handling stability is significantly improved. In literature [24–27], scholars established McPherson suspension simulation model through ADAMS/Car, carried out multi-objective optimization on the wheel camber angle, front wheel toe-in angle, kingpin caster angle and kingpin inclination angle, respectively. The results show that the steady-state rotation performance of the vehicle model is stable after the suspension optimization, which is beneficial to improve the

Fig. 3 The changing curve of average yaw rate with the speed [23]

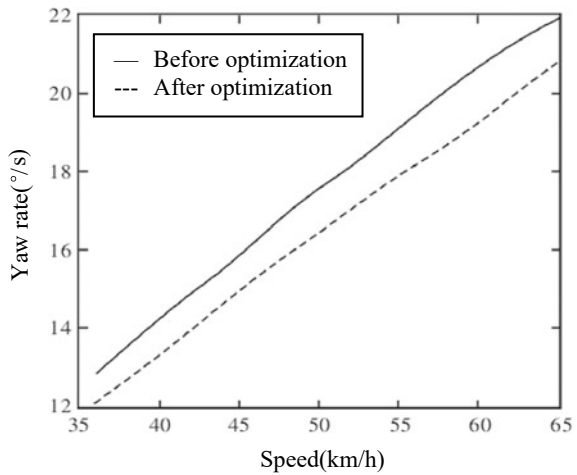
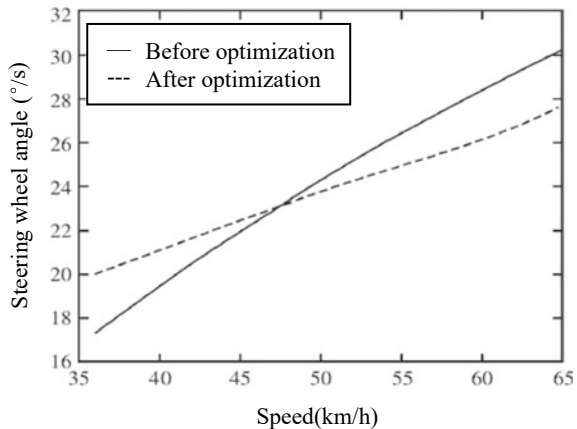


Fig. 4 The changing curve of average steering wheel angle with the speed [23]



handling stability. Dhurai [28] took an off-road vehicle as the research object, established a virtual simulation model of front suspension in ADAMS/Car, and carried out multi-objective optimization through the ADAMS/Insight, such as the camber angle, toe-in angle, kingpin inclination angle and kingpin caster angle, and finally improved the vehicle handling stability.

To sum up, the optimization based on ADAMS/Insight can comprehensively analyze the parameters of the suspension, accurately select the parameters that have a greater impact on the vehicle handling stability, and make the vehicle handling stability optimization more convenient and efficient.

In literature [29–35], genetic algorithm was used to carry out multi-objective optimization of geometric parameters of suspension. Zhang [29] established a simulation model of the McPherson front suspension of a car in ADAMS, and optimized the design of the minimum change of the toe-in angle and the minimum change of the camber angle based on the multi-island genetic algorithm. After optimization, the changes of toe-in angle and camber angle are reduced, the steering portability is improved, and the vehicle handling stability is enhanced. Shi [30] established a virtual model of a light bus in ADAMS/Car, and optimized the roll angle stiffness relationship of the front and rear antiroll bars by using the improved genetic algorithm NSGA-II. Using maximum lateral acceleration, average yaw rate and average roll angle as evaluation indexes, the simulation of central steering test and steady static circular test were carried out, and the results show that the optimization improves the steering stability of light bus when they are running at high speed. Mahmoodi [31] built a model for the double-wattarm suspension system in ADAMS, and optimized the geometric parameters of the suspension affecting the change of camber angle by using genetic algorithm. Taking yaw rate and lateral acceleration as evaluation indexes, the simulation test of J-shaped turn and lane change was carried out. The optimized suspension system reduces the change of camber angle, so that the change of tire force is reduced and the vehicle handling stability is improved. In reference [32–34], hubmotor driven vehicles were taken as the research object, and NSGA-II algorithm was used to conduct multi-objective optimization on the variation of front beam Angle, CAM Angle and wheel base. Through optimization, tire wear and lateral deviation were reduced and the handling stability of the vehicle was improved. Li [35] established the multi-body dynamics model of suspension in ADAMS/Car, and used NSGA-II algorithm to carry out multi-objective optimization on the gradient of wheel spacing variation, kingpin caster angle and kingpin inclination angle. The optimization results show that the variation range of kingpin caster angle, kingpin inclination angle and the tire wear were reduced, while the steering performance and the vehicle handling stability were improved.

To sum up, genetic algorithm and its improved algorithm are not only widely used in the study of vehicle handling stability, but also have good effects. In particular, the improved non-dominated sorting genetic algorithm can avoid the loss of individuals, and improves the search speed of the optimal solution and the robustness of the algorithm.

Literature [17–35] improves vehicle handling stability through optimization of suspension parameters, which are summarized in Table 2. They took the kingpin

Table 2 Research on handling stability based on suspension parameter optimization

Literature	Optimization goal	Optimization method	Evaluation index	Test method	Conclusion
[17]	Wheel toe-in angle	ADAMS/Insight	Yaw rate	Steering wheel angular step input	The yaw rate of the vehicle decreases obviously, and the response time of the diagonal step input decreases
[18]	Toe-in angle	ADAMS/Insight	Toe-in angle, camber angle	Parallel wheel jump, reverse wheel jump	After optimization, the change gradient of the front toe-in angle decreased significantly
[19]	Kingpin inclination angle	ADAMS/Insight	Yaw rate, side acceleration, sideslip angle	Steering wheel angular step input, steady rotation	Optimized steering wheel angular step input response characteristics and steady-state rotation response characteristics
[22]	Toe-in angle, camber angle, wheel lateral displacement	ADAMS/Insight	Yaw rate, vehicle roll angle	Steering back, steady turning, light steering	After optimization, yaw rate, steering wheel maximum rotation force and average friction force are all reduced
[23]	Camber angle, kingpin caster angle, inclination angle	ADAMS/Insight	Yaw rate, steering wheel angle	Steering wheel angle step input	The optimized vehicle yaw rate and steering wheel angle are reduced
[24]	Camber angle, toe-in angle, kingpin caster angle, inclination angle	ADAMS/Insight	sideslip angle	The steady state turning	After suspension optimization, the steady-state rotation performance of the vehicle model is stable, and the vehicle model has understeering characteristics

(continued)

Table 2 (continued)

Literature	Optimization goal	Optimization method	Evaluation index	Test method	Conclusion
[30]	Roll angle stiffness	NSGA-II	Lateral acceleration, yaw rate, sideslip angle	Central steering test, steady static circular test	Optimized and improved the handling stability of light passenger cars at high speed
[31]		Genetic algorithm	Yaw rate, lateral acceleration	J-turns, lane changes	When the camber angle is reduced, the tire force changes are reduced
[32]	toe-in angle, camber angle	NSGA-II	Lateral acceleration, roll angle	Serpentine test and two-lane change test	Both the maximum lateral acceleration and the maximum roll angle are reduced

positioning parameters, camber angle, toe-in angle and wheelbase variation as the optimization objectives, and yaw rate and lateral acceleration as evaluation indexes to carry out simulation tests such as steering angle step test, steering angle pulse test, snaking test and steering reorientation test. Finally, ADAMS was used to evaluate the handling stability of the optimized suspension model, and relatively accurate results were obtained.

4 Summary and Outlook

In this paper, the suspension is taken as the research object, and the evaluation and research on handling stability by ADAMS software at home and abroad in recent years are summarized. The vehicle handling stability is improved mainly through structural design of suspension (increase of suspension spring stiffness, install of lateral stabilizer bar, change of torsional beam structure) and parameter optimization (kingpin positioning parameters, wheel camber angle, toe-in angle and wheelbase variation). The main optimization methods are Adams/Insight and genetic algorithm.

With the continuous development of computer technology and virtual prototyping technology, ADAMS software is bound to develop in the direction of high precision, high stability and high efficiency. In particular, in recent years, more and more new composite materials have been used in automobiles to ensure the automobile lightweight. However, the use of composite materials can lead to geometric and material nonlinearity of the system. How to establish geometric and material nonlinearity models in ADAMS software and carry out efficient and accurate simulation of handling stability has become a difficult problem in the future automotive field.

Acknowledgements The authors gratefully acknowledge the support of the Centre for Efficiency and Performance Engineering (CEPE), University of Huddersfield, United Kingdom.

Funding This work was supported in part by the Opening Project of The State Key Laboratory of Heavy Duty AC Drive Electric Locomotive Systems Integration (2020ZJKF05), and in part by Natural Science Foundation of Sichuan Province (2022JY0400).

References

1. Yu, Z.: Automotive theory, 6th edn. Machinery Industry Press, Beijing, China (2006)
2. Cao, J.: Research on vehicle operation stability test and evaluation method based on human-vehicle-road closed-loop system. Shanghai Jiao Tong University (2015)
3. Wang, W., Bei, S., Zhao, Y., et al.: Development and prospect of vehicle handling and stability research. *Mach. Design Manuf. Eng.* **43**(10), 6–12 (2014)
4. Willamette, H.P.: Vehicle dynamics simulation and its method. Beijing Institute of Technology Press, Beijing, China (1998)
5. Guan, W.: Automobile structure. China Machine Press, Beijing, China (2016)
6. Zhang, L., Ge, X., Pan, F., et al.: Simulation research on the influence of suspension spring stiffness on vehicle handling and stability. *Agric. Equip. Veh. Eng.* **58**(1), 6–11 (2020)
7. Sert, E., Boyraz, P.: Optimization of suspension system and sensitivity analysis for improvement of stability in a midsize heavy vehicle. *Eng. Sci. Technol. Int. J.* **20**(3), 997–1012 (2017)
8. Dong, J., Cheng, A.: Optimization of suspension parameters matching of a minibus based on experimental design. *Sci. Technol. Rev.* **14**, 77–81 (2010)
9. Liu, K., Wang, Q., Li, Q.: Simulation analysis of formula racing vehicle with interconnected double wandarm independent suspension. *J. Zhejiang Univ. Sci. Technol.* **31**(3), 77–84+90 (2019)
10. Xia, Z., Chu, M., Zhou, J., et al.: Analysis of influence of transverse stabilizer bar on suspension performance of FSAE racing car based on ADAMS. *Agric. Equip. Veh. Eng.* **6**(273), 29–32 (2017)
11. Qiu, W., Zhu, S.: Influence of stiffness of lateral stabilizer bar on handling stability of passenger car. *Mech. Electr. Eng.* **36**(7), 744–748 (2019)
12. Wang, B., Ba, T., Ling, T., et al.: Lightweight design of rear transverse stabilizer rod of fuel cell car. *J. Mach. Design* **1**, 57–61 (2012)
13. Zhang, D., Liu, H., Zhang, L., et al.: DOE optimization of torsional beam suspension based on handling stability. In: The Fifth China CAE Engineering Analysis Technology Annual Conference Proceedings, pp. 48–53 (2009)
14. Wu, L., Li, G., Jing, L., et al.: Research on influencing factors of handling stability based on torsion beam structure. *Automot. Parts* **149**(11), 20–24 (2020)
15. Gao, J., Yang, X., Niu, Z., et al.: Analysis on influencing factors of suspension performance of torsional beam. *J. Jiangsu Univ. Nat. Sci. Edn.* **35**(6), 627–634 (2014)
16. Gao, Y.: Automobile application engineering, 3rd edn. People's Communications Press, Beijing, China (2004)
17. Wu, X., Zhang, B., Jia, Z., et al.: Optimization analysis of wheel positioning parameters on vehicle handling stability. *Tractor Farm Transporter* **275**(3), 27–30 (2020)
18. Mao, L., Dai, R., Jiang, G.: K&C analysis and optimization of multi-link rear suspension of a vehicle. In: 2020 China Society of Automotive Engineering Annual Meeting Proceedings (3) (2020)
19. Huang, J.: Research on simulation and optimization of handling stability of pickup truck based on ADAMS/ CAR. Nanchang University (2010)
20. Zhang, L., Xia, Y., Zhang, H., et al.: Review on optimal design of vehicle handling stability. *Shandong Transp. Sci. Technol.* **6**, 21–24 (2017)

21. Chen, X., Lan, F., Chen, J., et al.: Design and ride comfort analysis of micro electric vehicle suspension system. *J. Chongqing Univ. Technol. (Natural Science)* **386**(8), 30–37 (2018)
22. Yu, G., Zhao, Z., Wu, C., et al.: Optimization design and handling stability analysis of a bus front independent suspension. *Mach. Design Manuf.* (6), 74–77+82 (2020)
23. Zhang, Z., Zhu, J.: FSAE racing vehicle modeling and handling stability simulation. *Mach. Design Manuf.* **1**, 75–79 (2020)
24. Wang, R., Su, X., Wang, W.: Analysis and optimization of suspension characteristics in vehicle handling and stability. *Mach. Design Manuf.* **5**, 96–99 (2015)
25. Tian, D., Ma, X., Xi, W.: Optimization design of suspension wheel positioning parameters based on ADAMS. *Automobile Pract. Technol.* **16**, 79–82 (2019)
26. Xu, S., Peng, J., Ding, H., et al.: McPherson front suspension system structure optimization design. *Automobile Pract. Technol.* **324**(21), 92–96 (2020)
27. Zhang, Z., Shi, X., Li, J.: Motion simulation and optimization design of vehicle suspension system. *J Mach. Design* **32**(9), 30–33 (2015)
28. Prabhakar, D.: Optimization and effects of suspension parameter on front suspension of SAE Baja vehicle using ADAMS. *Int. J. Eng. Tech. Res.* **5**(9) (2016)
29. Zhang, P., Wang, H., Zhang, X., et al.: Structural parameter optimization design of torsion beam semi-independent suspension based on multi-island genetic algorithm. *Appl. Mech. Mater.* **893**, 109–115 (2019)
30. Shi, W., Wang, C., Li, Z.: Improving light bus handling and stability by anti-roll bar and bushing adjustment. *SAE Technical Papers* (2015)
31. Mahmoodi-Kaleibar, M., Javanshir, I., Asadi, K., et al.: Optimization of suspension system of off-road vehicle for vehicle performance improvement. *J Central South Univ.* **20**(4), 902–910 (2013)
32. Zhang, L., Zhang, S., Zhang, W.: Multi-objective optimization design of in-wheel motors drive electric vehicle suspensions for improving handling stability. *Proc. Inst. Mech. Eng. Part D J. Automobile Eng.* **233**(8), 2232–2245 (2019)
33. Wang, D., Chen, N., Qin, H.: Design and optimization of electric vehicle wheel motor and multi-link suspension system. *Mach. Design Manuf.* **351**(5), 231–234 (2020)
34. Wang, J., Liu, P., Yang, F., et al.: Kinematics optimization of double-arm front suspension for in-wheel motor drive electric vehicles. *Automot. Eng.* **43**(3), 305–312 (2021)
35. Li, A., Liao, L., Zhang, J., et al.: Kinematics analysis and multi-objective optimization of a vehicle suspension. *J. Chongqing Univ. Technol. (Natural Science)* **406**(6), 40–45+70 (2019)

Analysis and Decompose of Nine Degrees of Freedom Motion Simulator Relative Positional Precision



Bo Li, Huadong He, Yinjun Lian, Xia Wu, Tongling Fu, Weiling Zhao, and Huibo Zhang

Abstract Nine Degrees of Freedom Motion Simulator (9-DOF-MS) is the key equipment for calibration of Camera-type Rendezvous & Docking Sensor (CRDS) in spacecraft space rendezvous & docking Guidance Navigation and Control (GNC) sub-system, and it must be high with relative position precision. For meeting this demand, the components of errors impacting this system's integral indexes are analyzed systemically in this paper. At first, the relationship and interactions among the components of system errors are analyzed. Then the error model is built. By decomposing and redistributing the systematic precision index, 9-DOF-MS designed fulfils the precision requirements.

Keywords Nine degrees of freedom · Motion simulator · Relative position precision · Error analysis · Index decomposition

1 Introduction

Camera-type Rendezvous & Docking Sensor (CRDS) is composed of the camera installed on the tracking aircraft and the target marker on the target aircraft. Its performance is the key to the success of the rendezvous and docking [1, 2]. Normally, the ground calibration and verification of CRDS is realized by Nine Degrees of

B. Li

School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China
e-mail: 18222646720@163.com

B. Li · H. He · Y. Lian · X. Wu · T. Fu

Tianjin Key Laboratory of Microgravity Environment Simulation Technology, Tianjin 300458, China

B. Li · H. He · Y. Lian · X. Wu · T. Fu · W. Zhao

Tianjin Institute of Aerospace Mechanical and Electrical Equipment, Tianjin 300458, China

H. Zhang (✉)

School of Mechatronics Engineering, Harbin Engineering University, Harbin 150001, China
e-mail: zhb_hit@163.com

Freedom Motion Simulator (9-DOF-MS). Therefore, the 9-DOF-MS is one of the key ground support equipment for spacecraft rendezvous and docking. Our original “6 + 3” 9-DOF-MS is mainly used for the test verification of the GNC subsystem level of rendezvous and docking [3], which does not require high relative position accuracy of the simulator. The 9-DOF-MS described in this article is a product-level calibration device for CRDS. Furthermore, the 9-DOF-MS puts forward higher requirements on system accuracy. For example, the overall position error of the full stroke is smaller than 0.5 mm. At the same time, due to cost and development cycle factors, the traditional “6 + 3” type of motion simulator is difficult to meet the needs of the project. Therefore, a new “5 + 4” form of free distribution is adopted.

With reference to the previous research and development experience, the influencing factors of the full-stroke comprehensive position error mainly include straightness, disjointness, non-perpendicularity, and positioning accuracy of straight line and rotation. Therefore, establishing a mathematical model of the system’s comprehensive position error and each sub-error, analyzing the influence of the latter on the former, taking into account the feasibility of the project, and decomposing the indicators of each sub-error is an important basis for project implementation, and it has significant meaning.

In this paper, the error components that affect the comprehensive position accuracy of the system are identified firstly. Secondly, the correlation model between the component ring error and the closed ring error is established by referring to the calculation method of dimensional chain. Furthermore, the error accuracy of each component ring is calculated according to the error of the closed ring. Finally, the index decomposition of relative position accuracy is realized by using the iterative method.

2 The Simulator

Nine degrees of freedom motion simulator is shown in Fig. 1, consisting of two large mechanical subsystems: the high precision turntable system (tracking turntable, target turntable), the linear motion subsystem (x direction motion sub-system, y direction motion sub-system, and z direction motion sub-system). Among them, the two turntables are all three-axis turntables, from the outside to the inside in turn are the yaw axis, the pitch axis and the rolling axis. In the initial state, the yaw axis is the vertical axis, besides the pitch axis and the rolling axis are the horizontal axis.

3 Analysis of Influencing Factors

In view of the fact that the relative position error of the 9-DOF-MS has many influencing factors, if the overall analysis method is adopted, the calculation process

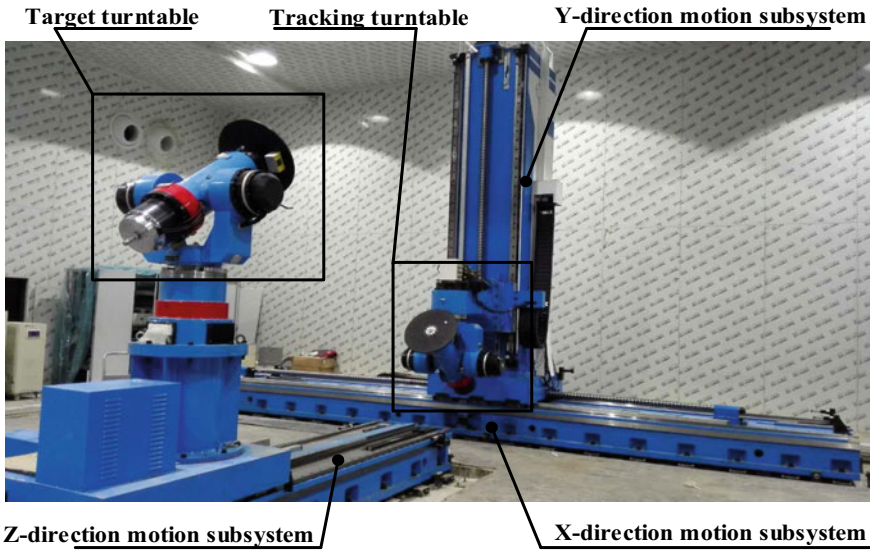


Fig. 1 Constitute of 9 degrees of freedom motion simulation system

will be cumbersome, which is not conducive to highlighting the main influencing factors, and even makes the analysis impossible [4, 5]. Therefore, it is necessary to separate the influencing factors into the rotary position error source of the turntable and the linear system position error source. After the separation, the calculation is performed separately, and the obtained results are superimposed to obtain the overall relative position error [6, 7]. Identify of error constituents affecting relative positional precision is shown in Table 1.

Table 1 Identify of error constituents affecting relative positional precision

Comprehensive index analysis item	Components	Error source
relative position error analysis	Position error introduced by turntable rotation	Disjoint degree
		non-vertality degree
		control precision
	Geometric error of three-way linear system	Position control error
Vertality error		
		The straightness of guide rail and the corresponding derived error source

4 Comprehensive Position Index Influence Analysis

4.1 Turntable Disjoint Degree

In order to analyze the influence of the disjoint degree of the turntable on the relative position error of the two simulators, the process analysis diagram is shown in Fig. 2. Considering the error source of the three-axis disjoint degree of the turntable, the mathematical model is established, as shown in Fig. 3 [8], where r is the distance between the load plate and the three-axis center.

In the general, the vector coordinate of r_{OM} with respect to the $o_0 - x_0y_0z_0$ coordinate system is

$$r_{OM} = \begin{bmatrix} r \cdot \cos\varphi_2 \cdot \sin\varphi_1 \\ r \cdot \sin\varphi_2 \\ r \cdot \cos\varphi_2 \cdot \cos\varphi_1 \end{bmatrix} \tag{1}$$

Considering the influence of disjoint degree, the vector coordinate of $r_{OM'}$ relative to $o_0 - x_0y_0z_0$ coordinate system is

$$r_{OM'} = \begin{bmatrix} (r \cdot \cos\varphi_2 \pm \Delta \cdot \sin\varphi_2)\sin\varphi_1 \pm \Delta \cdot \cos\varphi_1 \\ r \cdot \sin\varphi_2 \pm \Delta\cos\varphi_2 \\ (r \cdot \cos\varphi_2 \pm \Delta \cdot \sin\varphi_2)\cos\varphi_1 \pm \Delta \cdot \sin\varphi_1 \end{bmatrix} \tag{2}$$

The vector coordinate of the position error $r_{MM'}$ relative to the $o_0 - x_0y_0z_0$ coordinate system is

$$r_{MM'} = \begin{bmatrix} \pm\Delta \cdot \sin\varphi_1 \cdot \sin\varphi_2 \pm \Delta \cdot \cos\varphi_1 \\ \pm\Delta \cdot \cos\varphi_2 \\ \pm\Delta \cdot \cos\varphi_1 \cdot \sin\varphi_2 \pm \Delta \cdot \sin\varphi_1 \end{bmatrix} \tag{3}$$

The yaw angle and the pitch angle are two relatively independent quantities, such as $\varphi_1 \in (-50^\circ, 50^\circ)$, $\varphi_2 \in (-32.5^\circ, 32.5^\circ)$, Disjoint degree $\Delta \in (0, 0.35)$. Therefore, by bringing in the maximum value of each component of $r_{MM'}$ (not the maximum value at the same time), the three-dimensional components are as follows:

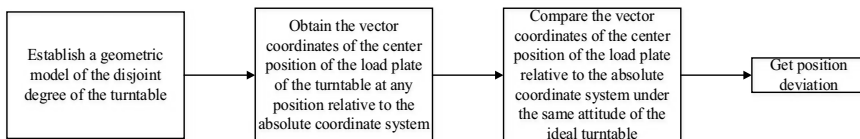


Fig. 2 Analysis procedure of turntable disjointness error affect

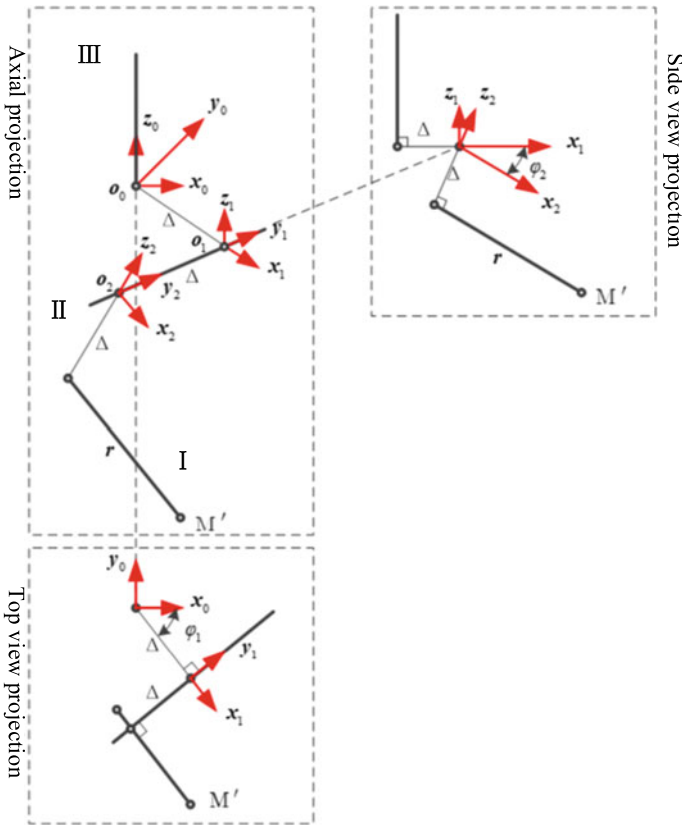


Fig. 3 Disjointness error model for three axis turntable

$$i = (\pm\Delta \cdot \sin\varphi_1 \cdot \sin\varphi_2 \pm \Delta \cdot \cos\varphi_1)_{max} = 0.369, \tag{4}$$

$$j = (\pm\Delta \cdot \cos\varphi_2)_{max} = 0.35, \tag{5}$$

$$k = (\pm\Delta \cdot \cos\varphi_1 \cdot \sin\varphi_2 \pm \Delta \cdot \sin\varphi_1)_{max} = 0.389. \tag{6}$$

Through the above calculation, it can be seen that the three-dimensional component value is more significant, but the disjointness of the turntable is a steady-state error. After processing and assembly, the disjointness values have been determined and can be detected by corresponding means. For the most effective to avoid the position error introduced by the disjoint degree, the measurement result can be used to set it as a part of the oretical model. Finally the corresponding position compensation is performed through motion control.

4.2 Non-Verticality of the Turntable

In order to analyze the influence of the non-verticality of the turntable on the relative position error of the two simulators, the analysis is carried out according to the process shown in Fig. 4, considering the error source of the three-axis non-verticality of the turntable.

The basic assumptions are as follows [9]:

- (1) The three axes of the turntable are not perpendicular, but the three axes intersect at a point.
- (2) The non-verticality of the turntable only occurs between two adjacent axes, such as the yaw axis (axis III) and the pitch axis (axis II), the pitch axis (axis II) and the rolling axis (axis I). The rotation axis and the yaw axis are disjoint.
- (3) The non-perpendicular angle ϵ between the axis I and the axis II should be inside the plane formed by the two intersecting lines of the axis II and the Z axis, as shown in Fig. 5a. For the convenience of calculation, the angle between axis II-Z axis plane and the XOZ plane is extremely small, which can be approximately regarded as in the XOY plane, and the angle between axis III and X axis is ϵ , as shown in Fig. 5b.

A sphere with radius r is established in the Cartesian coordinate system, where r is the distance between the center of the load plate and the intersection of the three axes. The 3 axes of the turntable are in this sphere. The three axes of the turntable intersect at a point, which is at the center of the sphere. The axis I is in the XOZ plane, the axis II is in the YOZ plane, and the axis III coincides with the Y axis. Due

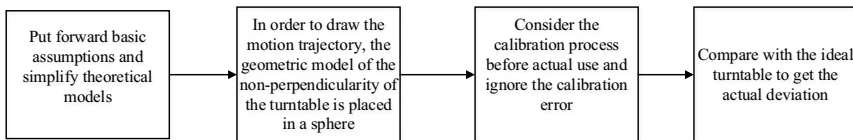


Fig. 4 Analysis procedure of turntable vertical error affect

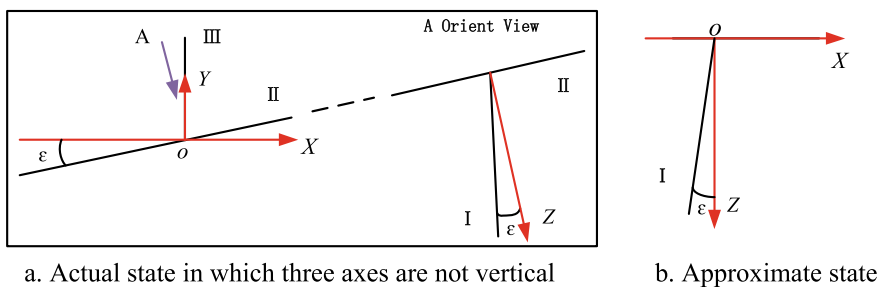


Fig. 5 The approximate principle for three axis vertical error

to the non-perpendicularity factor, there is an angle ε between the axis II and the Y axis. Due to the non-perpendicularity factor, there is an angle ε between the axis I and the Z axis. These situations are shown in Fig. 6a.

Since the turntable system has to be calibrated in the normal direction of the load plate after the assembly is completed, the axis I and the Z axis coincide, which is achieved by the rotation axis III at the ε angle. After the yaw axis rotates, the position of the axis II changes, and the new state is shown in the Fig. 6b. In this state, only the movement of axis II may introduce the position error of the load plate. The center position of the load plate forms a trajectory EF on the spherical surface during the 90° rotation of axis II, as shown in the Fig. 6b.

After the axis II rotates at an angle ψ , the coordinate of the position N' of the center point in the load plate on the trajectory \widehat{EF} is (x_0, y_0, z_0) , and the relationship is as follows:

$$N' \text{ on the sphere} : x_0^2 + y_0^2 + z_0^2 = r^2 \tag{7}$$

$$EN' \text{ is perpendicular to axis II} : x_0 + y_0\varepsilon - (z_0 - r)\varepsilon = 0 \tag{8}$$

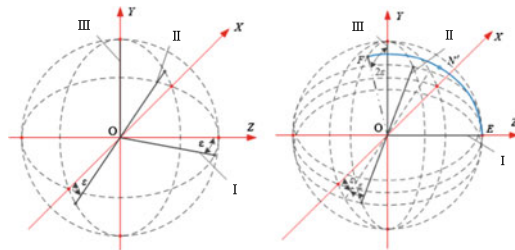
$$\text{Angle between } OE \text{ and } ON' : \cos\theta = z_0/r \tag{9}$$

Another point $F(0, -2\varepsilon r, r)$ is on the trajectory. Through the combination of the above conditions, the calculated coordinate relationship of the N' point is as follows:

$$x_0 = (1 - \cos\theta + \sin\theta)\varepsilon \cdot r, \tag{10}$$

$$y_0 = r \cdot \sin\theta, \tag{11}$$

$$z_0 = r \cdot \cos\theta. \tag{12}$$



a. Error model obtained by hypothesis b. The axis I coincides with the Z axis after calibration

Fig. 6 Vertical error model for three axis turntable

When considering the non-verticality model of the turntable, after the turntable is adjusted and calibrated to the ideal position, only the rotation of axis II produces an X-axis error component. When the maximum angular stroke of rotation is 32.5° , the X-axis error component is the largest with a magnitude of 0.01 mm. This deviation is mainly caused by the eccentricity of the rolling bearing. Although there is a certain periodicity, it is not a steady-state error and cannot be compensated.

4.3 Three-Dimensional Linear Motion

In order to analyze the influence of the linear system error on the relative position error of the two simulators, the analysis is carried out according to the process shown in Fig. 7.

The 9-DOF-MS space model (excluding the three-axis turntable part) is established, as shown in Fig. 8. The coordinate system is established at the intersection of the Z-direction and X-direction rails. The reference coordinate system Z-axis and Z-direction rail fitting straight line coincides. The X axis of the reference coordinate system is on the horizontal plane and orthogonal to the Z axis. The direction of Y axis is determined by the right-hand rule.

The main factors affecting the position accuracy of the target are as follows [10]:

θ_1 is the pitch angle of the target caused by the level change of the Z-direction guide rail in the direction of movement, as shown in Fig. 9.

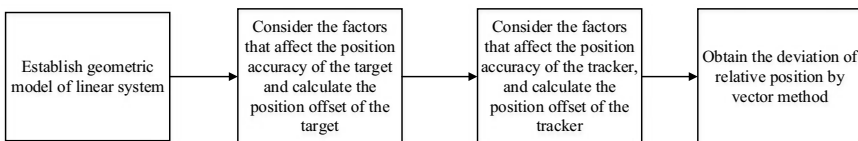


Fig. 7 Analysis procedure of linear system error

Fig. 8 The error model for 9 degrees of freedom motion simulator's linear system

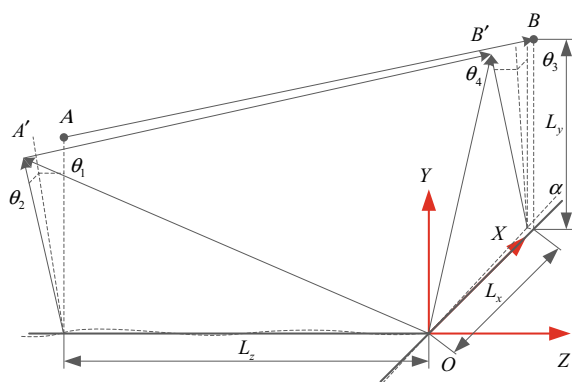
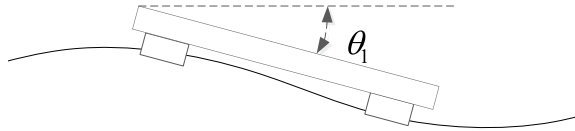


Fig. 9 Pitching angle error due to nonlinearity of the rail



θ_2 is the roll angle to the target simulator due to the height deviation of the Z-direction double guide rails, as shown in Fig. 10.

Δz is the position control error of the target device in the Z direction, as shown in Fig. 11.

W_z^1 is the height runout of the Z-direction guide rail, as shown in Fig. 12.

W_z^2 is the lateral runout of the Z-direction guide rail, as shown in Fig. 13.

The coordinates of the nominal position coordinate point of the target in the reference coordinate system can be expressed as $A(0, h, -L_z)$, while the coordinates of the actual position coordinate point of the target are expressed as $A'(\pm h \cdot \theta_2 \pm W_z^2, h \pm W_z^1, -L_z \pm \Delta z \pm h \cdot \theta_1)$. The vector radius between A and B is expressed as follows:

$$r_{A'A} = \begin{bmatrix} \pm h \cdot \theta_2 \pm W_z^2 \\ \pm W_z^1 \\ \pm \Delta z \pm h \cdot \theta_1 \end{bmatrix} \tag{13}$$

The main factors affecting the position accuracy of tracker are as follows:

α is the degree of non-perpendicularity between the X-direction guide rail and the Z-direction guide rail (rad).

θ_3 is the pitch angle of the target caused by the level change of the X-direction guide rail in the direction of movement (rad).

θ_4 is the roll angle to the target simulator due to the height deviation of the X-direction double rail (rad).

Δ_x is the X-direction position control error of the target (mm).

Δ_y is the Y-direction position control error of the target (mm).

Fig. 10 Rolling angle error due to variance between two rails

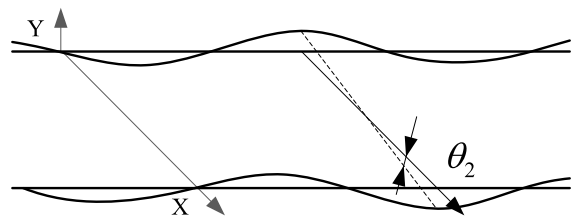


Fig. 11 Absolute position control error region

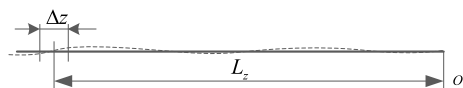


Fig. 12 Sketch diagram of rail vertical fluctuate



Fig. 13 Sketch diagram of rail lateral fluctuate



W_x^1 is the height runout of the X-direction guide rail (mm).

W_x^2 is the lateral runout of the X-direction guide rail (mm).

W_y^1 is the height runout of the Y-direction guide rail (mm).

W_y^2 is the lateral runout of the Y-direction guide rail (mm).

The nominal position coordinate point B of the tracker in the reference coordinate system is $(L_x, L_y, 0)$, while the actual position coordinate point B' of the target device is $(x_{B'}, y_{B'}, z_{B'})$, where

$$x_{B'} = L_x \pm \Delta_x \pm (L_y \pm \Delta_y) \cdot \theta_3 \pm W_y^2, \tag{14}$$

$$y_{B'} = L_y \pm \Delta_y \pm W_x^1, \tag{15}$$

$$z_{B'} \pm (L_y \pm \Delta_y) \cdot \theta_4 \pm (L_x \pm \Delta_x) \cdot \alpha \pm W_x^2 \pm W_y^1. \tag{16}$$

The $r_{BB'}$ vector expression is:

$$r_{BB'} = \begin{bmatrix} \pm \Delta_x \pm (L_y \pm \Delta_y) \cdot \theta_3 \pm W_y^2 \\ \pm \Delta_y \pm W_x^1 \\ \pm (L_y \pm \Delta_y) \cdot \theta_4 \pm (L_x \pm \Delta_x) \cdot \alpha \pm W_x^2 \pm W_y^1 \end{bmatrix}. \tag{17}$$

The relative position error should be the difference between the actual position vector $r_{A'B'}$ and the theoretical position vector r_{AB} of the tracker and the target. Because of $r_{A'B'} + r_{B'B} + r_{BA} + r_{AA'} = \mathbf{0}$, thus

$$r_{A'B'} - r_{AB} = r_{A'A} - r_{B'B}. \tag{18}$$

By calculating the above expression, it can be expressed as:

$$r_{A'B'} - r_{AB} = \begin{bmatrix} \pm h \cdot \theta_2 \pm W_z^2 \pm \Delta_x \pm (L_y \pm \Delta_y) \cdot \theta_3 \pm W_y^2 \\ \pm W_z^1 \pm \Delta_y \pm W_x^1 \\ \pm \Delta_z \pm h \cdot \theta_1 \pm (L_y \pm \Delta_y) \cdot \theta_4 \pm (L_x \pm \Delta_x) \cdot \alpha \pm W_x^2 \pm W_y^1 \end{bmatrix}. \tag{19}$$

After expanding various equations, accumulating all errors, and removing the second-order small quantities that have little effect on the system error, the X-direction component of the maximum relative position error is obtained, as follows:

$$i = h \cdot \theta_2 + W_z^2 + \Delta_x + L_y \cdot \theta_3 + W_y^2. \tag{20}$$

The Y component of the maximum relative position error is expressed as follows:

$$j = W_z^1 + \Delta_y + W_x^1. \tag{21}$$

The Z component of the maximum relative position error is expressed as follows:

$$k = \Delta_z + h\theta_1 + L_y\theta_4 + L_x\alpha + W_x^2 + W_y^1. \tag{22}$$

By using the probability method to calculate the superposition of each component ring, Eqs. (20), (21), and (22) are converted into Eqs. (23), (24), and (25), as follows:

$$i = \sqrt{(h \cdot \theta_2)^2 + (W_z^2)^2 + (\Delta_x)^2 + (L_y \cdot \theta_3)^2 + (W_y^2)^2}, \tag{23}$$

$$j = \sqrt{(W_z^1)^2 + (\Delta_y)^2 + (W_x^1)^2}, \tag{24}$$

$$k = \sqrt{(\Delta_z)^2 + (h \cdot \theta_1)^2 + (L_y \cdot \theta_4)^2 + (L_x \cdot \alpha)^2 + (W_x^2)^2 + (W_y^1)^2}. \tag{25}$$

After the initial values of the error components of each component of the linear system are substituted into the above-mentioned dimensional chain, the requirements cannot be met. Considering the feasibility, the error component is iterated until the relative position error reaches about 60% of the required value. And the final result is the superimposed components of each component index that affect the comprehensive index are shown in Table 2. It can be seen from the table that the comprehensive position deviation is estimated to reach 0.271 mm, leaving a margin for unestimable measurement and compensation errors.

Table 2 Integral relative positional errors of whole range

	straight line	Turntable (disjoint)	Turntable (not vertical)	Turntable (control)	Sum
<i>i</i> /mm	0.171	Steady-state error, Compensation elimination	$0.01 \times \sqrt{2}$	$0.017 \times \sqrt{2}$	0.209
<i>j</i> /mm	0.066		0	$0.017 \times \sqrt{2}$	0.09
<i>k</i> /mm	0.147		0	0	0.147
$\sqrt{i^2 + j^2 + k^2}$ /mm	0.257		0.014	0.034	0.271

4.4 Index Decomposition

The initial comprehensive position error is calculated by substituting the initial value into the dimensional chain. There is a certain deviation from the expected index requirement. Therefore, each sub-error index needs to be corrected again. After several rounds of repeated calculations, the appropriate sub-error is obtained. The sub-error results of the last iteration are analyzed and compared with the initial value to form Table 3. In the table, the data of the project realization value is combed and compared. It can be seen that the implementation and control of the project are strictly based on the decomposition index.

5 Conclusions

In this paper, the disjoint degree, non-verticality degree and linear system error models of the turntable system are established, and the action forms and processing methods of various errors are analyzed. The more economical error values of each sub-error can be obtained through the probability calculation of the dimensional chain. In the entire distribution process, the realizability of each system is strictly considered, and problems that cannot be realized by engineering are avoided. The implementation process of this project is strictly controlled according to the results of this article. Currently, 45 positions of the target and 34 positions of the tracker have been measured. There are total of 1530 arrangement results, and the comprehensive position accuracy meets the 0.5 mm requirement. This index analysis and decomposition method has relatively general adaptability, and has reference value for the index realization analysis of other complex and high-precision mechanical systems.

Table 3 Initial values and validated values of linear system error components

Number	Sub-error	Initial value	Definite value	Project realization value
1	X-axis position control error $\Delta_x(\text{mm})$	0.06	0.03	0.018
2	Y-axis position control error $\Delta_y(\text{mm})$	0.05	0.03	0.022
3	Z-axis position control error $\Delta_z(\text{mm})$	0.1	0.03	0.030
4	Level fluctuation error of Z-axis movement direction $\theta_1(\text{rad})$	$0.01 \times (\pi/180)$	$0.002 \times (\pi/180)$	$0.002 \times (\pi/180)$
5	Z-axis lateral levelness fluctuation error $\theta_2(\text{rad})$	$0.005 \times (\pi/180)$	$0.002 \times (\pi/180)$	$0.0018 \times (\pi/180)$
6	X-axis movement direction levelness fluctuation error $\theta_3(\text{rad})$	$0.01 \times (\pi/180)$	$0.002 \times (\pi/180)$	$0.0015 \times (\pi/180)$
7	X-axis lateral levelness fluctuation error $\theta_4(\text{rad})$	$0.005 \times (\pi/180)$	$0.002 \times (\pi/180)$	$0.0013 \times (\pi/180)$
8	Non-verticality error of Z-direction and X-direction guide rail $\alpha(\text{rad})$	$0.01 \times (\pi/180)$	$0.002 \times (\pi/180)$	$0.0019 \times (\pi/180)$
9	X-direction rail height run-out error $W_x^1(\text{mm})$	0.05	0.03	0.024
10	X-direction guide rail lateral runout error $W_x^2(\text{mm})$	0.05	0.03	0.012
11	Y-direction rail height runout error $W_y^1(\text{mm})$	0.05	0.03	0.018
12	Y-direction guide rail lateral runout error $W_y^2(\text{mm})$	0.05	0.03	0.010
13	Z-direction rail height run-out error $W_z^1(\text{mm})$	0.1	0.03	0.021
14	Z-direction guide rail lateral runout error $W_z^2(\text{mm})$	0.1	0.03	0.019

Acknowledgements This research was funded by the Natural Science Foundation of Hebei Province of China, grant number E2019202132; the Science and Technology Research Project of Colleges and Universities of Hebei Province, grant number BJ2019049; the Science and Technology on Space Intelligent Control Laboratory, grant number 6142208200101; the Innovation Foundation Project of CAST, grant number CAST-2021-01-11; the Tianjin Key Laboratory of Micro-Low Gravity Environment Simulation Technology Open Subjects: WDZL-2020-01; and the 2021 Postgraduate Innovation Funding Project of Hebei Province of China, grant number CXZZSS2021028.

References

1. Dezhu, G., Chunhui, Z., Lin, Z., et al.: LD lighting system for camera-type rendezvous and docking sensor. *Aerosp. Control Appl.* **40**(5), 57–62 (2014)
2. Yong, L., Yu-ping, S., Bowen, Q.: Research on the technologies of dynamic test of movement simulator with high-precision industrial measurement. *Geomatics Spat. Inform. Technol.* **37**(7), 47–49 (2014)
3. Xinbang, Z., Liangdong, L., Shenzhao, L.: Motion simulators for rendezvous simulation test. *Aerosp. Control Appl.* **35**(2), 51–55 (2009)
4. Xinglin, Z., Long, M.: Design and application of three-dimensional movement simulator under space environment. *Spacecraft Environment Eng.* **26**(5), 442–446 (2009)
5. Jiangwei K.: The error model research of spacecraft docking motion simulator. Harbin Institute of Technology, Harbin (2005)
6. Yizha, L., Ke, Z., Xiaosong, Z., Zhifei, Z.: Research on total error model and error compensation of CNC machine tool. *Manuf. Technol. Mach. Tool* **7**, 46–51 (2003)
7. Chang, Z., Shenhui, L.: Mechanism accuracy. Higher Education in the Higher Education Press, vol. 7, pp. 1–3 (1995)
8. Kai, H., Jiangang, S., Mingming, Q.: Analysis on performance indexes of a six-degree-of-freedom motion simulator. *Electron. Opt. Control* **15**(6), 83–87 (2008)
9. Hao, T., Yang, Z., Dawei, Z.: Movement simulator modeling and simulation in integrate test platform for docking mechanism. *J. Astronaut.* **28**(4), 997–1001 (2007)
10. Zhigang, S., Yongxiang, C.: Technical research of the error compensation of three coordinates CNC machine tool. *Mach. Tool Electr. Apparatus* **1**, 22–24 (2004)

A Study on Vibration Response in the Baseplate of a Delta 3D Printer for Condition Monitoring



Xinfeng Zou, Zhen Li, Fengshou Gu, and Andrew D. Ball

Abstract In recent years, the necessity of implementing sensor-based process condition monitoring (CM) in additive manufacturing (AM) has attracted the attention of many foreign governments and academic institutions. To verify the feasibility of the condition monitoring on additive manufacturing, an experiment is carried out. The experiment focuses on the abnormal status of the 3D printer to explore the relationship between the printing signals and the printing quality. There are two methods are used to process the experimental data. One method is the Short-Time Fourier Transform (STFT), which is used frequently used in this report. Its processing result indicates that the signal changes in both the time domain and the frequency domain. The other method is the Mean function. By the comparison of the two methods, the mean function turns out to be better than STFT at presenting the differences in detail and proving the consistency of the signals and the 3D printer features. This experiment lays a firm foundation and points out the directions for future research, such as mathematical simulation, etc.

Keywords Condition monitoring · Additive manufacturing · Vibration signals · STFT

1 Introduction

Currently, 3D printing technology has a wide range of applications in personal and industrial consumption fields such as medicine, mold making, automobile manufacturing, cultural creativity, aerospace, and construction [1]. Fused deposition modeling (FDM) technology, one of the additive manufacturing methods, has the advantages of simple structure, easy to use and low cost, it has become one of the most popular

X. Zou · Z. Li (✉) · F. Gu
School of Automation Industrial, Beijing Institute of Technology, Zhuhai 519000, China
e-mail: Zhen.Li@hud.ac.uk

X. Zou · Z. Li · F. Gu · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_65

811

and widely used technologies in kinds of additive manufacturing technologies. It has been used in many fields such as academic research, education, and industrial design in many applications.

However, most of the 3D printing equipment currently on the market are not equipped with a relevant process monitoring system, and it is impossible to know the various defects or failure status information that may occur during the printing process, and take corresponding remedial measures in time, and the quality of the printed products is consistent, dimensional accuracy and comprehensive performance can't meet the requirements of application in the professional engineering field. Therefore, the application of many 3D printing products is still limited to model making or design verification, and to a certain extent, 3D printing technology and its products are restricted with further promotion [2].

2 Literature Review

At present, the stability of the FDM processing process, the overall quality and consistency of the product, etc. still can't meet the needs of industrial applications. It limits the further expansion of the application scenarios of FDM additive manufacturing technology. It is necessary to study the corresponding monitoring technology to achieve the real-time identification and prevention of typical defects or abnormal printing conditions in FDM. There are several faults or defects of the quality problems occur in FDM process during printing a 3D model, such as too large surface roughness, low precision, and low strength. When these faults or defects occur, due to the changes in the internal structure of the material, especially the shrinkage deformation process caused by the temperature field and the uneven stress caused by the melting phase transition of the material, it will have characteristics that can be distinguished from the normal signals. In order to ensure and improve the overall quality of additive manufacturing printed products, many researchers have investigated into modeling the process, feedback control, printing route scheduling, printing process parameter optimization, material performance analysis [3–6]. Melvin and others have used a video microscopy system in powder bed fusion (PBF) to monitor its processing process to study the material powder forming characteristics in the laser sintering process [7]. Rao et al. first studied the multi-sensor fusion monitoring technology for FDM printing conditions [8]. Zou et. al investigated the probability of the condition monitoring on FDM printing process based on multi-signals [9–11].

3 Experiment Methodologies

To investigate if condition monitoring (CM) could be applied in additive manufacturing, there are several experiments have been done. In this chapter, two sections are illustrated as below.

3.1 Experimental Cases Analysis

From above, it is well known that the printing model is formed by the accumulation layer by layer based on a printing baseplate, as a result, leveling the baseplate of the delta 3D printer for the accumulation operation is one of the important setting before printing. This paper focuses on a delta 3D printer which has a complicated and precisely movement system. The leveling baseplate method of the delta 3D printer is according to three points positioning principles. Hence, how to evaluate the leveling status of each positioning point is key operation to leveling the baseplate. As shown in Table 1, there are several experiment cases should be followed based on mixed combination.

According to Table 1, there are three kinds of parameter groups are described. SC, the signal channel, includes four channels named Baseplate channel, Nozzle channel, Column channel and Sound channel, independently. LS represents the status of the delta 3D printer, including $LS - N$, $LS - A1$, $LS - A2$ and $LS - A3$. For example, $LS - N$ means the normal status of the delta 3D printer, i.e., the three levelling points of the baseplate are normal status, as shown in Table 2.

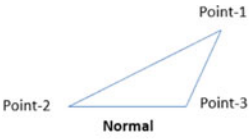
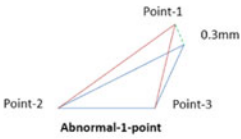
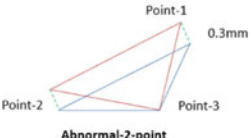
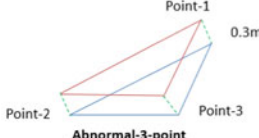
Phase is the third parameter group, including $P - Su$, $P - So$ and $P - Ho$ respectively. The different parameter groups combination can reflect different experiment results. As a consequence, the data analysis will be based on the three kinds of parameter groups as follow.

$$f(\text{experimentcases}) = \left(\left[\begin{matrix} SC - B \\ SC - N \\ SC - C \\ SC - S \end{matrix} \right], \left[\begin{matrix} LS - N \\ LS - A1 \\ LS - A2 \\ LS - A3 \end{matrix} \right], \left[\begin{matrix} P - Su \\ P - So \\ P - Ho \end{matrix} \right] \right)$$

Table 1 The experiment cases combination table

Parameter group 1 Signal channel (SC)	Parameter group 2 Leveling Status of the delta 3D printer (LS)	Parameter group 3 Phase (P)
Channel-baseplate (SC-B)	Normal-3 points are good (LS-N)	Support phase (P-Su)
Channel-nozzle (SC-N)	Abnormal-one-point faults (LS-A1)	Solid phase (P-So)
Channel-column (SC-C)	Abnormal-two-points faults (LS-A2)	Hollow phase (P-Ho)
Channel-sound (SC-S)	Abnormal-Three-points faults (LS-A3)	

Table 2 The four kinds of baseplate leveling cases

LS-N (Normal status)	LS-A1 (Only one point inclination 0.3mm)
	
LS-A2 (Two points inclinations 0.3mm)	LS-A3 (All of the points (three points) inclinations 0.3mm)
	

3.2 Experimental Preparation

As shown in Fig. 1, there is a delta 3D printer which is regarded as the experiment platform. All the CM experiment will focus on the printer. The model of the Delta printer is MOOZ-3, produced by DOBOT. There are three vibration sensors are equipped with the delta 3D printer, including on the baseplate, neighbor to nozzle and on the column, respectively. Moreover, there is a sound sensor is set close to the printer. All of the sensors are produced by YMC Piezotronics, Inc.

On the other hand, there is a data collector which has four channels, to obtain the dynamic data in-real time. The purpose, equipped with three vibration sensors and

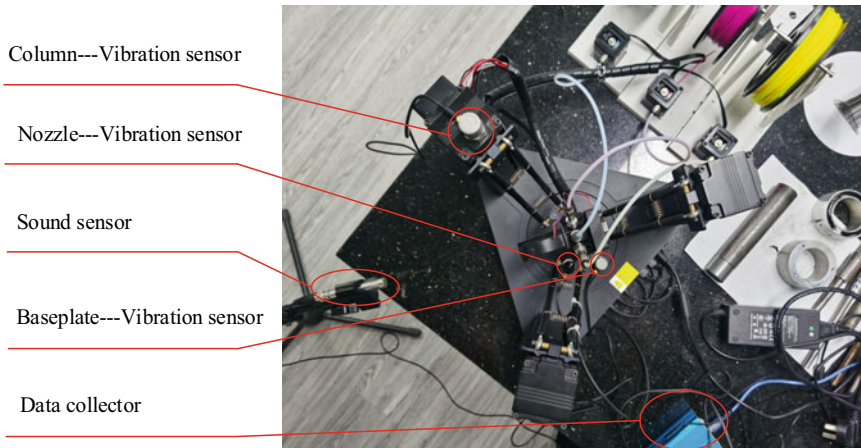


Fig. 1 Delta 3D printer equipped with several sensors

one sound sensor, is to investigate the relationships between the tiny vibrations and the status of the printer when the printer is working. The reasons to locations of the vibration sensors on the baseplate and nozzle, because they are the two important parameters of the delta printer, and these two parameters would cause some poor printing qualities with the printing model.

4 Data Analysis

According to $f(experimentcases)$, there are kinds of experiment results have been proposed. As we known that $f(experimentcases)$ has three group parameters, if all the parameters are proposed to compare each other simultaneously, it would cause a fuzzy experiment result. As a result, in this chapter, two parameters should be fixed and only one parameter can be changed, hence, the experiments results can be analyzed precisely as shown in Fig. 2. Consequently, there are two experiment cases have been investigated.

4.1 Different Phases with the Same Case

To reveal the frequency comparison on the same case with different phase, the $f(experimentcases)$ are described as below.

$$f(experimentcases) = \left(\left[\begin{matrix} SC - B \\ SC - N \\ SC - C \\ SC - S \end{matrix} \right], [LS - N], \left[\begin{matrix} P - So \\ P - Ho \end{matrix} \right] \right)$$

$SC - B$ to $SC - S$ means the data signal from different signal channels should have a comparison, at the same time, $LS - N$ remains as the normal levelling points

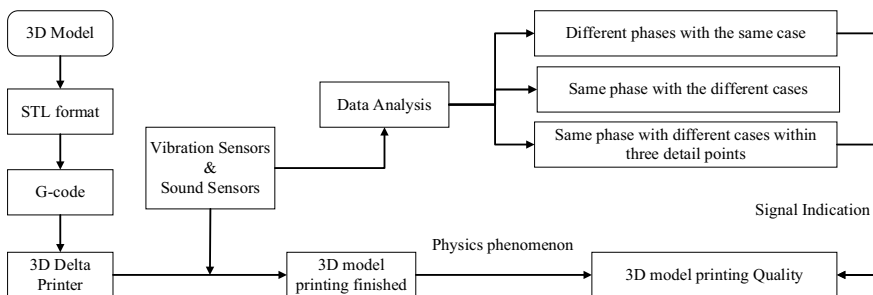


Fig. 2 The $f(experimentcases)$ analysis framework

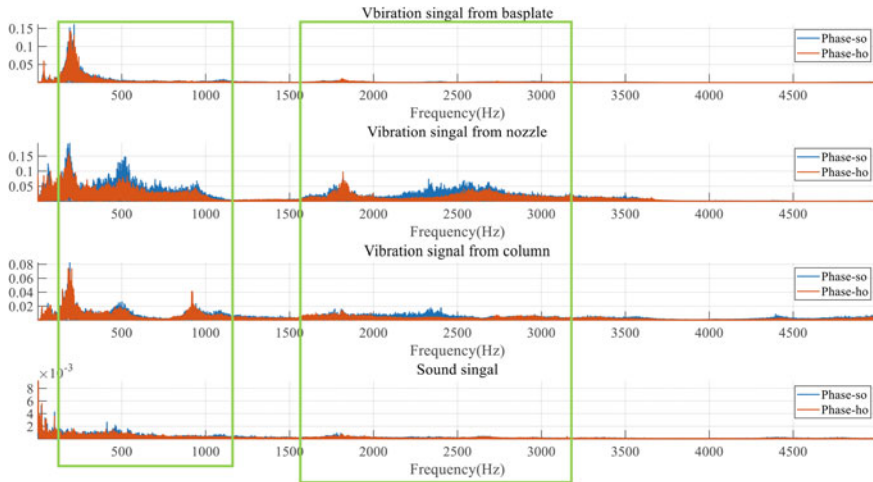


Fig. 3 The frequency comparison on the same case with different phases

status of the delta printer baseplate, moreover, two phases including printing the solid phase and printing the hollow phase with a same printing leveling status.

As shown in Fig. 3, the frequency comparison between $P - So$ and $P - Ho$ based on the same $LS - N$ (the normal status of the baseplate of the 3D delta printer) with four channels respectively. The blue line is the frequency of $P - So$ and the red line is the frequency of $P - Ho$. It is obviously to obtained that the amplitude of $P - So$ is stronger than $P - So$, especially around 180–1200 Hz, and 1500–3700 Hz.

On the other hand, the frequency around 180 Hz is obviously obtained in $SC - N$, $SC - B$ and $SC - C$, except $SC - S$. The result demonstrates that there is a vibration around 180 Hz occurred during the 3D delta printer is working. Furthermore, this vibration energy exists in the 3D printer including the baseplate, nozzle and column, and has no transmission to the sound sensor. Meanwhile, the frequencies around 1800 Hz are strong enough that the phenomenon occurred from $SC - N$ to $SC - S$ reveals the vibration around 1800 Hz doesn't disappear during the vibration transmission, i.e., the frequency around 1800 Hz may be caused by printing process.

4.2 Same Phase with Two Different Cases

According to the conclusion above, there is another cases study has been proposed. Since 3D printing process is a non-linear system, the signals with two different phases ($P - So$ and $P - Ho$) can be analyzed by STFT method and Mean functions based on the signal from LS-N. Moreover, the investigation frequency zone can be shorted within 2000–3700 Hz to get more detail information.

As shown in Fig. 4, the STFT processing results on $P - So$ with $LS - N$ and $LS - A3$ have been demonstrated, independently. The upper photos of Fig. 4a, b are the entire STFT processing including $P - So$ and $P - Ho$, and the lower photos of Fig. 4a, b are the STFT processing only including P-So characteristics with alignment, meanwhile, all of the frequency windows are set from 2000 to 3700 Hz. Nevertheless, the different between the two signals are hard to get by the STFT method consequently.

To investigate the different of the signals between $LS - N$ and $LS - A3$, mean function method has been proposed and the result are described as Fig. 5. There

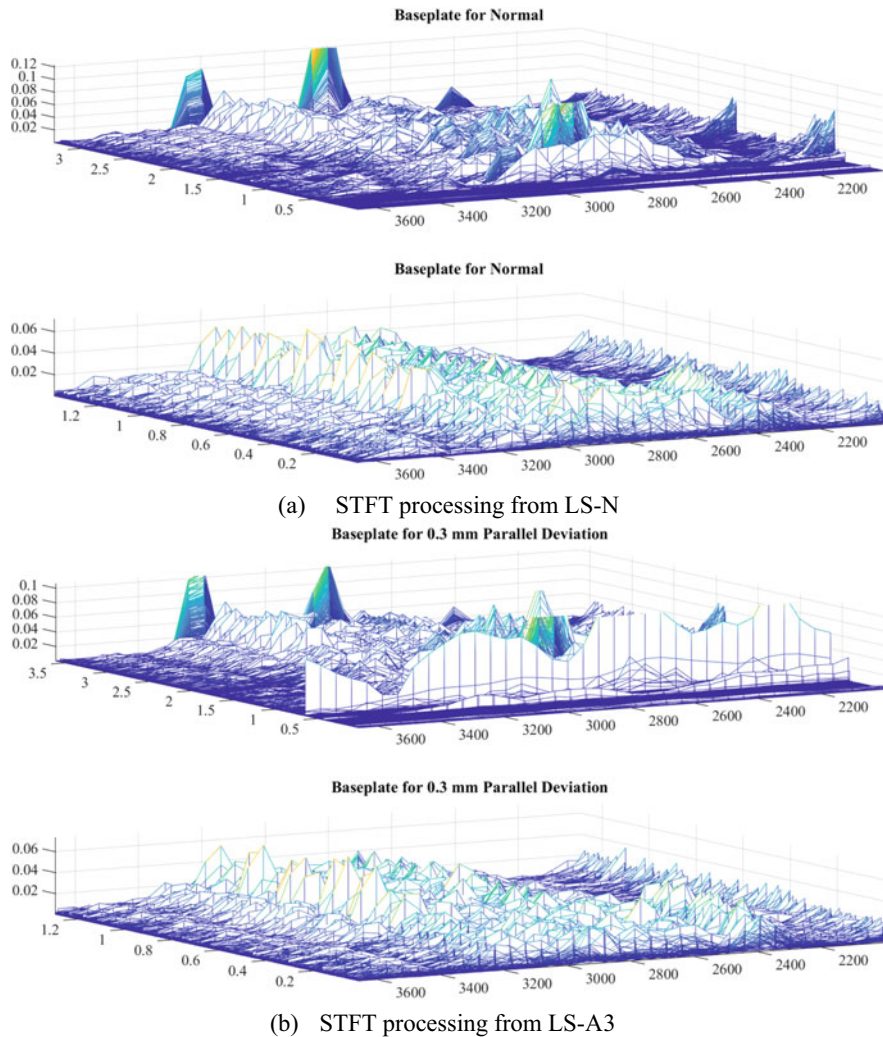


Fig. 4 STFT processing from $LS - N$ and $LS - A3$, independently

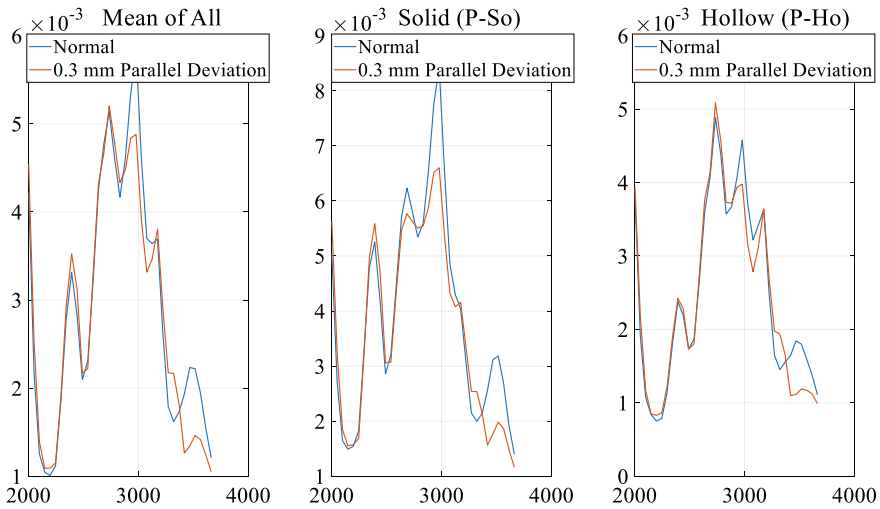


Fig. 5 Mean function processing result comparison between the signal $LS - N$ and $LS - A3$

are three comparison results are described. The left figure demonstrates the average result with $P - So$ and $P - Ho$, and it is clearly to observe the amplitude of $LS - N$ is stronger than $LS - A3$, in addition, the comparison result in $P - So$ is more clearly than $P - Ho$. As a consequence, the printing process with $P - So$ is more effectively transmit the vibration power from the nozzle to baseplate around 3000 Hz. On the other hand, compare with the STFT method, the mean function method seems more effectively since the processing result is more obviously in this experiment case study.

After demonstrating the different signals from $P - So$ and $P - Ho$ with the same LS , the signal of $P - So$ is more obviously. Hence, there are comparisons with signals of $P - So$ from different LS . Figure 6a reveals that there is an obviously different comparison when the printer is printing the solid phase between 2000 to 3700 Hz. However, the comparison curve from Fig. 6b, c and d are very familiar, Consequently, baseplate can reflect more effectively with the printing process than others. Furthermore, the amplitude value from nozzle is stronger than others' location, but with a high coincidence result between the two curves.

4.3 Same Phase with Different Cases Within Three Detail Points

Based on the data analysis results above, the high frequency is more likely produced during the printing process, and the frequency zone from 2000 to 3700 Hz maybe

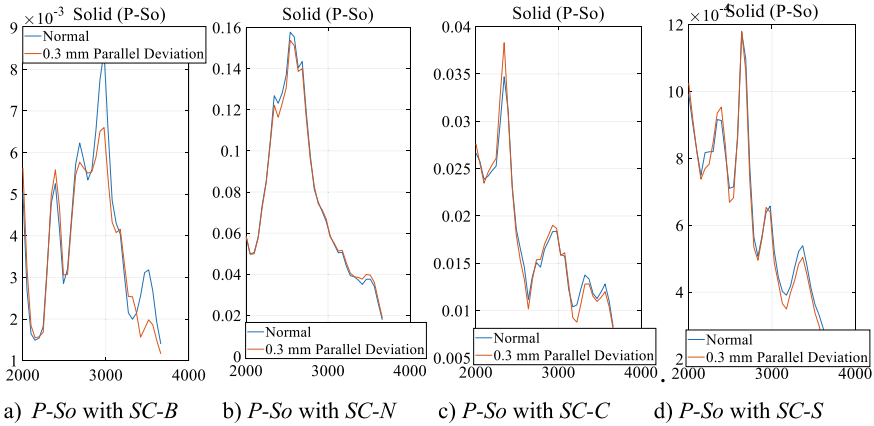


Fig. 6 The signals of printing *P-So* in *LS-N* with different *SC*

Table 3 The three kinds of baseplate leveling cases

LS-A1-point NO.1	LS-A1-point NO.2	LS-A1-point NO.3

describe the leveling status more clearly, furthermore, the signal of *P-So* from baseplate is more obviously than others. According to $f(experimentcases)$ illustrates, the signals of the same phase with different cases within three detail points would be investigated in this section.

$$f(experimentcases) = \left([SC - B], \begin{bmatrix} LS - A1 - pointNo.1 \\ LS - A1 - pointNo.2 \\ LS - A1 - pointNo.3 \end{bmatrix}, \begin{bmatrix} P - So \\ P - Ho \end{bmatrix} \right)$$

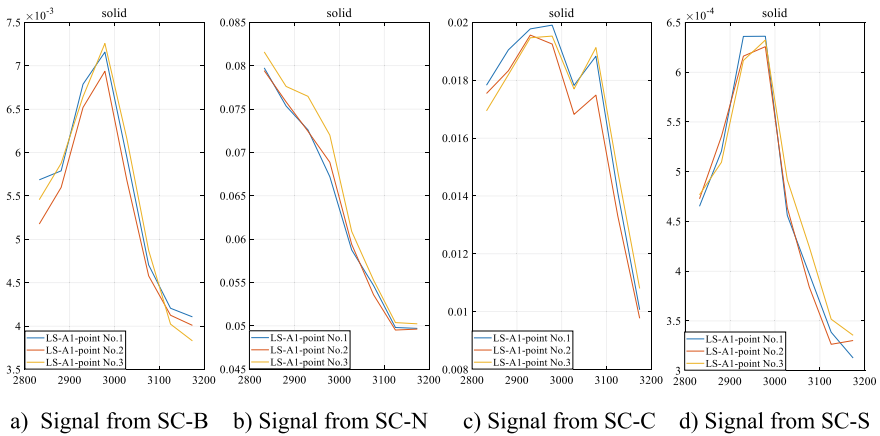


Fig. 7 The same phase with different cases within three detail points

LS – A1-point NO.1, from $f(experimentcases)$, means leveling status with only one abnormal point, and the point is named NO.1, as shown in Table 3. NO.1, NO.2 and NO.3 are three independent point.

Based on Fig. 7, there are two indications can be proposed. Firstly, the different of the three signals from (a)–(d) shows that (a) and (b) are more obviously than (c) and (d), i.e., the different of the three signals are easy to be detected on the baseplate and nozzle during the printer is printing the solid characteristic. On the other hand, no matter which signal cases and channels, the different of the three signals when printer is printing the hollow characteristic is no clearly.

In addition, the vibration signals of SD2-01 and SD2-02 are very closely, but very different with vibration signals of SD2-03. At the same time, vibration signals of SD2-03 is the lowest amplitude than the other two point’s. This kinds of the phenomenon also reveals that the distance between nozzle and baseplate, the long distance is, the weak vibration signal transmission is. Furthermore, according to Fig. 6, the experiment result interpreted that SD2-03 point already belongs to the abnormal status at the beginning of the experiment.

5 Conclusion

The experiment conducted above is mainly to investigate the vibration responses in a 3D printing process and printing quality for online condition monitoring based on the data-driven method. The author collected the vibration signal of the baseplate, the nozzle and the column and analyzed the time and frequency information through STFT data processing. Furthermore, to investigate the printing quality, the author focuses on the factor of leveling baseplate, then, mean function method is utilized into the data process. There are three conclusions can be deduced from this investigation:

- (1) Compared with STFT method, mean function or RMS method are more effectively in this case. Because mean function or RMS method can conduct a tiny difference from the original signals during the printing features are very similar.
- (2) Since there are vibration signals occurred during the printing process, some of the vibration signals could be caused by the physics 3D printer framework, and the vibration could be transmitted from nozzle to baseplate, such as the frequency around 180 Hz. In addition, some of the vibration may be caused by printing process, this kind of vibration can be detected by acoustic, such as the frequency around 1800 Hz.
- (3) The more distance between nozzle and baseplate, the lower amplitude of the signal with high frequency during the printing process, moreover, the probability of the warping phenomenon with the printing 3D model are increasing.

References

1. Wang, C.Y., et al.: The evolution and development frontiers of 3D printing technology based on literature analysis. *Journal* **34**(9), 72–77 (2015)
2. Everton, S.K., Hirsch, M., et al.: Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing. *J. Mater. Design* **95**, 431–445 (2016)
3. Heralić, A., Christiansson, A.-K., et al.: Height control of laser metal-wire deposition based on iterative learning control and 3d scanning. *Opt. Lasers Eng.* **50**(9), 1230–1241 (2012)
4. Jin, Y.-a., He, Y., et al.: A parallel-based path generation method for fused deposition modeling. *Int. J. Adv. Manuf. Technol.* **77**(5–8), 927–937 (2014)
5. Sood, A.K., Ohdar, R.K., et al.: Experimental investigation and empirical modelling of FDM process for compressive strength improvement'. *J. Adv. Res.* **3**(1), 81–90 (2012)
6. Lee, J., Huang, A.: Fatigue analysis of FDM materials. *Rapid Prototyping J.* **19**(4), 291–299 (2013)
7. Melvin III, L.S., Das, S. et al.: Video microscopy of selective laser sintering. In: *Proceedings of the Solid Freeform Fabrication Symposium, DTIC Document*, pp. 34–41 (1994)
8. Rao, P.K., Liu, J., et al.: Online real-time quality monitoring in additive manufacturing processes using heterogeneous sensors. *J. Manuf. Sci. Eng.* **137**(6), 61007 (2015)
9. Zou, X., Zeng, L., et al.: An investigation of vibration responses in a 3D printing process for online condition monitoring. In: *Conference 2020, InCoME*, vol. 105, pp. 335–344, EI (2020)
10. Zeng, L., Zou, X.: Research on fault analysis and prediction algorithm based on delta 3D printer. In: *Conference 2020, InCoME*, vol. 105, pp. 498–504, EI (2020)
11. Zeng, L., Zou, X.: Theoretical and experimental study on the levelling and printing quality control of delta 3D printer. In: *Conference 2020, ICEDME*, pp. 555–558 EI (2020)

Analysis of Metamaterials-Based Acoustic Sensing Enhancement



Shiqing Huang, Yubin Lin, Lichang Gu, Rongfeng Deng, Fengshou Gu, and Andrew D. Ball

Abstract Acoustic sensing is a non-destructive technology that plays an essential role in condition monitoring. For high-quality data collection, condition monitoring relies on various sensing methods that further complicate the wiring of the system. Moreover, weak signals such as evanescent waves carrying valuable information are usually hard to capture. With the emerging field of metamaterials, such issues could be optimized and solved. This paper presents a metamaterial that is designed by two kinds of typical unicells, purely geometrically, with the aim to enhance the acoustic signal without any external power source. As a result of transmission through the designed metamaterials, the acoustic pressure level at a particular range of frequency is efficiently enhanced. Furthermore, the frequency shift of the enhancement is achieved by altering specific structural parameters, which demonstrates its tunable characteristics. This study intends to provide ideas for the design of acoustic metamaterials for applications such as remote sound measurement, energy harvesting, fault diagnosis, etc.

Keywords Metamaterials · Sensing · Sound enhancement · Acoustics

1 Introduction

Acoustic sensing has a significant position in engineering and has demonstrated the technical feasibility and effectiveness of using sound signals for condition monitoring. The advantages of acoustic sensing for condition monitoring are obvious,

S. Huang · Y. Lin (✉) · L. Gu · R. Deng · F. Gu
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, Guangdong, China
e-mail: Yubin.Lin@hud.ac.uk

S. Huang
e-mail: Shiqing.Huang@hud.ac.uk

S. Huang · Y. Lin · L. Gu · R. Deng · F. Gu · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1, UK

non-invasive detection, economical, and online testability of sound signals, so it has been widely used [1–3]. Nevertheless, Acoustic sensing also has its limitations, such as sound signals tend to mix with background noise and reflect waves, affecting the quality of the acquitted data. When it comes to acoustic sensors arrays, fabrication and installation of which could be arduous. Meanwhile, the placement of acoustic sensors requires to set in the same way during the monitoring process.

In the last thirty years, the emergence of acoustic metamaterials has provided us with novel ideas to solve the existing challenges in acoustic sensing, and enable various extraordinary effects, such as subwavelength imaging, negative refraction, invisible cloaking, etc. Liu et al. proposed wrapping the lead-ball in soft material and using epoxy resin as the base structure to achieve ultra-low frequency acoustic bandgap [4]. Subsequently, Z. Yang et al. applied this design idea to two-dimensional structures by embedding a mass block in the center of the elastic film, the effective density of the composite material was found to be negative at the resonance frequency of the system [5]. Zlong Liang et al. proposed a acoustic metamaterials-surface and observed in both mass density and bulk modulus being negative simultaneously [6]. The studies mentioned above indicate that such unique properties of metamaterials have tremendous potential engineering applications. More importantly in this study, as a propagational medium for sound; acoustic metamaterials are commonly designed without electrical power components, which reduces the intricacy of the wiring sensors systems.

2 Methodology

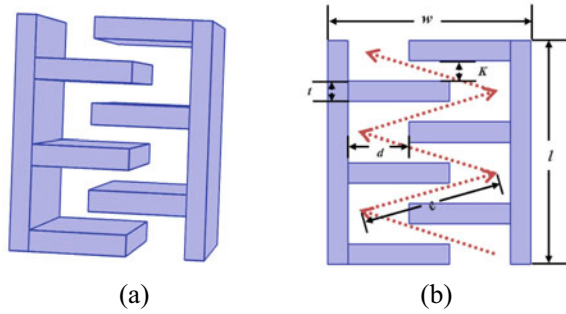
Coiling space structure [7–10] was firstly proposed by Liang et al. in 2012, it has obtained various exotic characteristics such as zero, negative refractive index at different frequencies, showing that the design can be flexible and the properties can be further extended for certain scenarios. Coiling space refers to a material with a particular refractive index through a multi-folded acoustic air duct that increases the path of sound waves in specific directions in which reduces the equivalent sound velocity This study uses two type of coiling space unicells to assemble a metamaterial, which is designed to achieve acoustic magnification.

2.1 Metamaterials Unicell Design

A three-dimensional unicell structure is shown in the Fig. 1a. The geometrical parameters of this unit are l in length and w in width, and it has interlaced internal structures, where the blue part represents the frame of the design, and it is placed in the air as the propagation medium. This unit has a Z-shape channel for sound transmission.

The path has the same width d , the number of coiling paves is N . The number of curls N refers to the number of times the sound wave turns back and forth as it

Fig. 1 **a** Space coiling unicell in 3D. **b** Geometric parameters and wave propagate direction



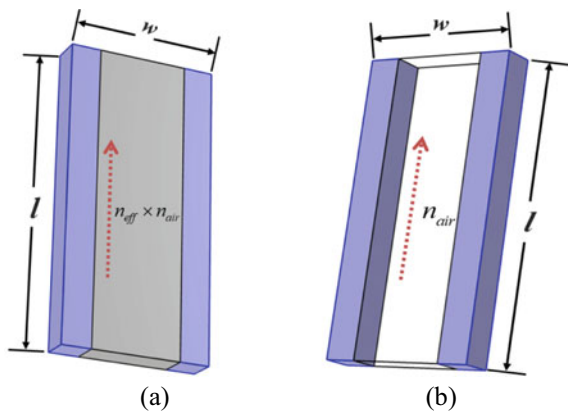
transmits through the Z-shape channel. As a result, the effective refractive index of this unit can be calculation by the following equation.

$$n_{eff} = \frac{N \times \sqrt{(w - d)^2 + (k + t)^2}}{l}$$

where n_{eff} is the effective refractive index, N is the wave path number (in Fig. 1 the this number is 2.5). w, d, k, t, l , are the structural geometrical parameters in Fig. 1b. When the frequency is below the cutoff frequency, the sound waves travel from the entrance to the exit of the unit structure in Z-shape path rather than in a straight line. Shown in Fig. 2.

In this process, the transmission length of the acoustic wave is increased by several times, leading to a high relative equivalent refractive index of the unit structure, and the value of the refractive index depends on the ratio of the total transmission length of the acoustic wave to the length of the unit.

Fig. 2 **a** Equivalent medium of coiling space. **b** Without coiling space equivalent model



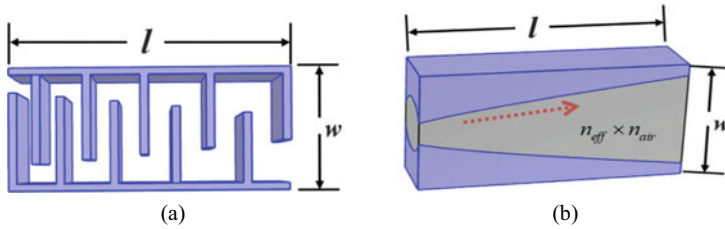


Fig. 3 a Unicell with gradient. b Equivalent model unicell with gradient

2.2 Gradient Unicell Design

Until now, coiling space metamaterials have been configured so that the Z-shape path width (k in Fig. 1b) is the same throughout its entire length. By applying this design methodology, it is demonstrated that an internal limit is placed on the amplitude of an acoustic signal, irrespective of cell size or operating frequency, another unicell is proposed, the gradient coiling space structure, in which the metamaterials’ acoustic impedance changes relatively gradually as the channel width changes. It is possible to realize the change in path width in many different ways. However, for the purposes of this study we will examine a single case in which the channel width progresses in a specific way, which is the width (k in Fig. 3a) varies with gradient. A constant common ratio is used to control the width of the channel over its entire length rather than having a constant width throughout.

2.3 Periodic Structure Design

To magnify the sound signals, the unicells mentioned above are being assembled periodically to a metamaterial, as Fig. 4 shows, each unicells are isolated from the path of

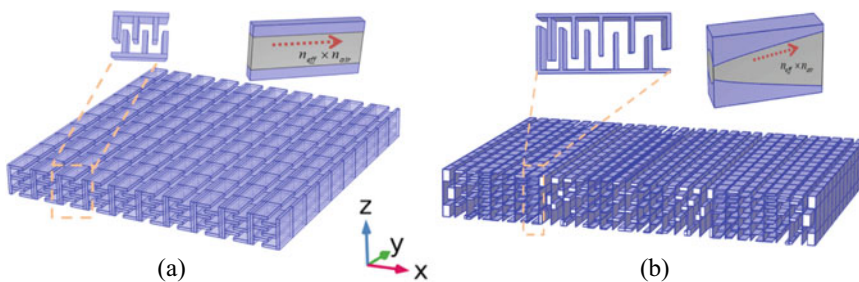


Fig. 4 Two types Metamaterial in 3D view. a Homogenous path metamaterials. b Gradient path metamaterials

other adjacent units, forming an array in the direction of x and y . Two distinct characteristics are realized by such designs [5]: the first one is the resonance modes, which ensure complete acoustic transmission with high selectivity, and another being the increased path length, which effectively increases the refractive index. As independent properties, they are not fascinating, but when these characteristics are combined, they exhibit exotic characteristics. In previous research [5], structure analogous with Fig. 4a have been prove that they are capable of enhancing sound emission in form of double-plate, and the sound source has to be place between these to plates so that the sound signal can be magnified.

Since most of the sound sources come from the external field, in order to better deal with these applications, the above structure is further adjusted in this research. The second unicell design focuses on a particular structure of geometric gradual variation of channel width, which is similar a shape of the speaker as Fig. 3b shows. Hence, A gradient coiling space structure is designed in which the gradual variation of channel width results in the acoustic impedance variation of the metamaterials. When sound signal propagates from external field, it transmits through the metamaterials then sound wave are being focused due to the gradient effective refractive index both in x and z direction, thus acoustic signal has been enhanced.

3 Numerical Simulation Analysis

In this study, Finite elements analysis is carried out to analyze the characteristics of the designed metastructrues, to theoretically predict the sound magnification of the metamaterial, numerical calculations are performed using COMSOL, a finite element analysis and solver software package. In this analysis, the metamaterial is set to be sound-hard-walls using pressure-acoustics-frequency-domain, air as the medium, and incident plane waves in the normal direction as the sound source with sound pressure amplitude 1 Pa.

3.1 Gradient Metamaterial Acoustic Signals Magnification

Firstly, the metamaterial with gradient has been analyzed in 2D as the Fig. 5a shows. It demonstrates that acoustic signal can be enhanced at the center above after propagating the designed structure, due to the gradient design of a unicell and the overall arrangement of the unicells, the high refractive index is achieved so that the acoustics signals has been focused above the metamaterial. A tested point is set to the sound focus region (shown in Fig. 5b) to plot out the acoustic pressure response to the frequency range under 1500 Hz. It shows that the sound sources generated at the exit of each cell (upper side) will interfere with each other in the far-field and then produce a focusing region, the amplitude distribution of the sound pressure field obtained by numerical simulation is shown in Fig. 5c. It demonstrates that the designed gradient

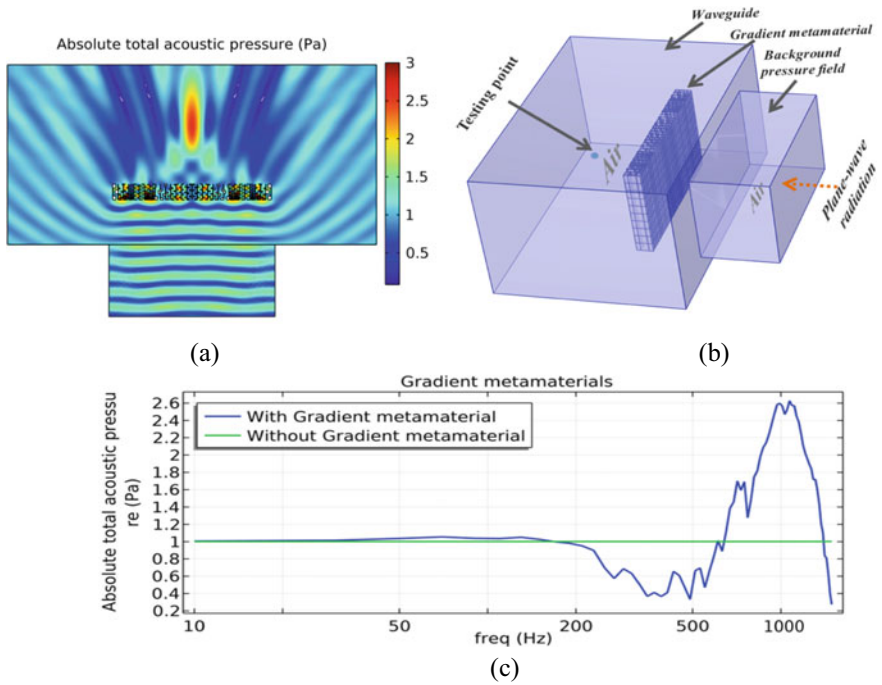


Fig. 5 **a** Gradient metamaterials acoustic enhancement. **b** Equivalent 3D simulation sketch. **c** Schematic of comparison

metamaterial enhances the pressure in the range around 1000 Hz. As the plane wave propagates through the metamaterials, it is clearly visible in the upper side of the metamaterial, and the sound signal is confined to a narrow region, indicating the metamaterials can manipulate the plane wave to achieve strong directivity efficiently hence the sound can be amplified.

3.2 Combined Metamaterials Acoustic Signals Magnification

With the structure designed above in 2.3 (Fig. 4a), in this section it is combined with the gradient structure, and carried out in the same scenario as Fig. 5b. Sound pressure level can be further enhanced in this particular arrangement. As the Fig. 6 shows, the characteristic of the second metamaterial is fully utilized to focus the acoustics signal, after that, the first metamaterial acts as a Fabry–Perot (FP) resonance, sound signal is enhanced between the two plates due to the high reflective. As Fig. 6b shows, comparing to the previous test (Fig. 5c), acoustic pressure at the frequency around 1900 Hz can be further magnified approximately by 25%, and with higher selectivity as results.

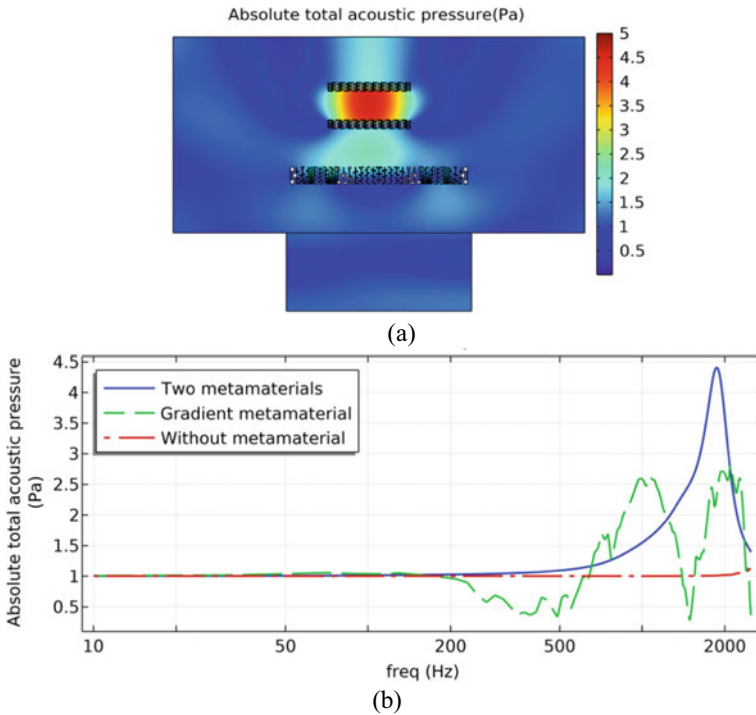


Fig. 6 a Two kinds of metamaterials. b Schematic of comparison

4 Conclusion

This paper presents a metamaterial that is designed by two kinds of typical unicells-coiling space with high refractive index. The simulation results indicate that it is capable of acoustics wave magnification: the metamaterials with gradient manifests characteristic of sound focusing thus sound signal can be magnified; Based on this gradient metamaterial, on which two plates of structures without gradient are combinedly applied, the result shows that sound signals can further be enhanced within this structure. In contrast to the conventional sound signal enhancement, the proposed metamaterials can operate without any electrical components but a significant improved magnification, and thus may facilitate potential applications. It is the intention of this paper to provide ideas for the design of acoustic metamaterials for applications such as remote sound measurement, energy harvesting, and fault diagnosis, etc.

References

1. Ma, J., Zhang, H., Shi, Z., et al.: Modelling acoustic emissions induced by dynamic fluid-asperity shearing in hydrodynamic lubrication regime. *Tribol Int* (2020)
2. Li, M., Appadoo, R., Hu, W., et al.: Condition monitoring of reciprocating compressor based on acoustic imaging (2020)
3. Mondal, D., Haba, U., Gu, F., et al.: Airborne acoustic signature analysis for fault diagnosis of reciprocating compressors using modulation signal bi-spectrum. In: 2019 25th International Conference on Automation and Computing (ICAC) (2019)
4. Liu, Z., Zhang, X., et al. Locally resonant sonic materials. *Science* **289**(5485), 1734–1734 (2000)
5. Liang, Z., Li, J.: Extreme acoustic metamaterial by coiling up space. *Phys. Rev. Lett.* **108**(11), 114301 (2012)
6. Liang, Z., Willatzen, M., Li, J., et al.: Tunable acoustic double negativity metamaterial. *Sci Rep*
7. Xie, Y., Popa, B.I., Zigoneanu, L., et al.: Measurement of a broadband negative index with space-coiling acoustic metamaterials. *Phys. Rev. Lett.* **110**(17), 175501.1–175501.4 (2013)
8. Ghaffari, R., Nikolajczyk, J., Holt, G., et al.: Horn-like space-coiling metamaterials toward simultaneous phase and amplitude modulation. *Nat. Commun.* **9**, 1349 (2018)
9. Song, K., Lee, S.H., Kim, K., et al.: Emission enhancement of sound emitters using an acoustic metamaterial cavity. *Rep* **4** (2014)
10. Yuan, B.G., Liu, J.Y., Liu, C., et al.: Resonant tunneling compression and evanescent wave amplification by an acoustic metalens. *Appl. Acoust.* **178**, 107993 (2021)

A Novel Cylindrical Mechanical Metastructure for Drone Vibration Isolation



Yubin Lin, Shiqing Huang, Lichang Gu, Rongfeng Deng,
Solomon Okhionkpwonyi, Qingbo He, Fengshou Gu, and Andrew D. Ball

Abstract Drone technologies are widely used for various purposes in many fields. However, the onboard imaging platform is severely compromised by low-frequency vibration during flight, which cannot be suppressed by a general vibration isolation method, leading to poor image quality and failure to fly. In this paper, a novel cylindrical mechanical metastructures (CMMS) vibration isolator was proposed to overcome the drawback of general vibration isolators based on the vibration of the drone imaging platform. By using additive manufacturing, a prototype of the CMMS was fabricated, and experiments were carried out to verify the mechanical properties. In research results, the CMMS isolator has been observed to suppress the full-band frequency vibration of the drone, enabling the imaging platform to operate in a stable environment.

Keywords Vibration isolation · Mechanical metastructure · Drone system · Imaging platform

1 Introduction

Drones are used frequently in the military, industry, and academics due to their excellent efficiency and flexibility. There are several types of aerial surveillance

Y. Lin · S. Huang · R. Deng (✉)
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, China
e-mail: Rongfeng.Deng2@hud.ac.uk

Y. Lin
e-mail: Yubin.Lin@hud.ac.uk

Y. Lin · S. Huang · L. Gu · R. Deng · S. Okhionkpwonyi · F. Gu (✉) · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK
e-mail: F.Gu@hud.ac.uk

Q. He
School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

systems used in the following applications: search and rescue [1, 2], engineering mapping [3], patrol inspection to health monitoring of civil facilities [4–6]. The majority of these uses are typically performed by drones equipped with the proper imaging platform, for example, cameras or thermal imaging devices for collecting relevant image data. Nevertheless, the imaging platform is seriously affected by different frequencies and amplitudes of airborne vibrations, leading to reduced image quality and even information loss [7]. Since drone technology is developing at an exponential rate, increased demands are placed on drone vibration suppression and isolation.

In general, by using passive vibration isolation, rubber isolators are installed between the imaging platform and the drone (see Fig. 1) [8–10]. Despite the fact that passive vibration isolation systems with elastic materials can effectively dampen high-frequency vibration on imaging platforms, the natural frequency of the system will increase, resulting in insufficient low-frequency vibration reduction. Additionally, low-frequency vibrations are typically associated with large amplitudes. As a result, many scholars suggest using active control to isolate vibrations [11–14]. Vibration isolation with active means has better isolation performance, and it is more robust. It is also more stable compared with other passive vibration isolation methods. Vibration isolation systems can be complex due to their structure. Controllers and precise feedback systems are additional devices that are required for the drone, which undoubtedly increase its workload, increase the operating cost, and reduce its endurance. Therefore, improving the performance of imaging platforms through passive vibration isolation is of great significance for drone applications.

In the last decade, a great deal has been learned about metastructures, which are artificial materials designed for unique mechanical properties [15–17]. Metastructures with nonlinear mechanical properties are used extensively in energy capture and vibration isolation, such as negative stiffness mechanical metastructures [18–20], multi-stable mechanical metastructures [21], quasi-zero stiffness mechanical

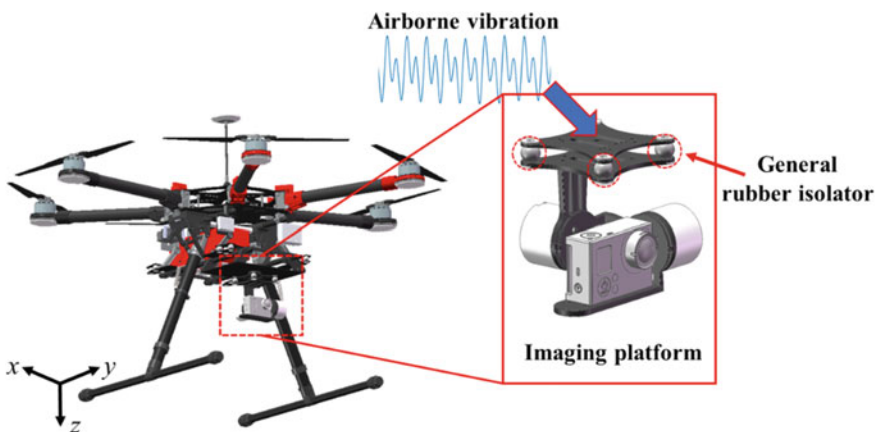


Fig. 1 Industrial drone with imaging platform

metastructures [22, 23], etc. The exotic mechanical properties of the mechanical metastructure can be enhanced with the growth and improvement of additive manufacturing technology. Research shows that lightweight mechanical metastructures have a significant vibration isolation effect and are able to isolate low-frequency vibrations in micro-mechanisms, which opens up a new solution for the stabilization of drone imaging platforms.

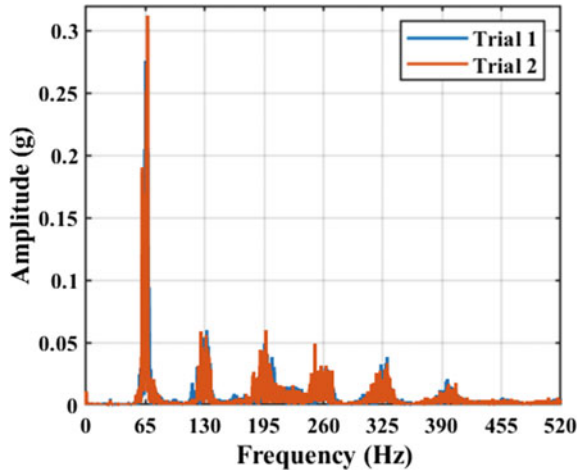
A novel cylindrical mechanical metastructure (CMMS) vibration isolator is proposed in this study using the vibration characteristics of the drone, and the mechanical properties of the CMMS are experimentally investigated. Secondly, the mechanical properties of an isolator for general rubber and an isolator for recombination structures were compared. Finally, the vibration isolation performance of the CMMS is then confirmed on the drone.

2 Vibration Characteristics of Drone System

Typically, drone systems are composed of multiple rotating motor-propeller systems, their rotation will be the primary excitation source [24]. Similarly, the rotating mechanical system of the drone will be similar to its frame. As a result, along with the fundamental frequency of the motor's rotation, the frequency multiplier of that rotation frequency should also be considered. This is because it may produce a considerable amount of vibration at high frequency.

To obtain the vibration of the drone frame accurately, the hover mode, generally used during the drone's working operation, is selected for vibration signal analysis. Considering the limitations of traditional acceleration sensors and the need to record high-frequency vibration signals from rotating machinery, a wireless acceleration sensor with high sampling frequency provides a viable solution for the condition monitoring of rotating machinery [25]. As the intensity of vibration is highest in the vertical direction, the sensor is installed at the connection position of the imaging platform to continuously collect the vibration data in the vertical direction two times in different environments. In the spectrum of the signal (see Fig. 2), it can be seen that the impact frequency of vibration is mainly concentrated at 65 Hz, as well as a small vibration impact between 2 and 6 harmonics of the 1st order, which makes it important to isolate the vibration at these frequencies to improve the drone platform stability.

Fig. 2 Spectrum in hover flight mode



3 Metastructure and Fabrication

3.1 Structural and Design Parameters

The design process of the novel CMMS vibration isolator proposed in this study shows in Fig. 3. The unit of the CMMS (see Fig. 3a) is based on a “frame-spring-frame” system in which a pre-deformed ring generates the “spring” effect. When the ring is compressed radially, it shows high nonlinear stiffness, which significantly improves vibration isolation efficiency [26]. As shown in Fig. 3b, the one-dimensional (1D) structure will be connected by a pre-deformed ring unit with the same parameters and convoluted to form a CMMS (see Fig. 3c). The geometrical

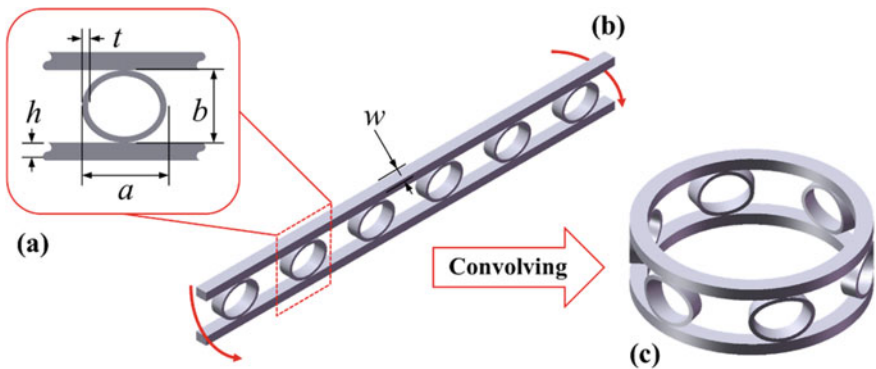


Fig. 3 Structural and design parameter (a) units cell; (b) 1D metastructure isolator; (c) cylindrical metastructure isolator

parameters are depicted as follow: the pre-deformed ring long shaft is a , the short shaft is b , its thickness is t , the height of the upper and lower frame connecting the ring is h , the overall width of the 1D metastructure is w , and the number of pre-deformed rings in the CMMS is n .

3.2 Fabrication via Additive Manufacturing Technology

Due to the complexity of three-dimensional (3D) CMMS, it cannot be fabricated by traditional manufacturing methods. However, the possibility of manufacturing complex metastructures is rapidly increasing as additive manufacturing technology matures. This study utilized a 3D printer based on Fused Deposition Modeling (FDM) to fabricate a CMMS vibration isolator prototype using TPU-95A, a thermo-plastic polyurethane material with high toughness. Table 1 provides details about the prototype parameters, only 9.5 g of CMMS are measured as weight.

4 Experiments

To obtain dynamic mechanical properties, CMMS vibration isolators are subjected to frequency response function (FRF) testing via exciter sweep excitation. As seen in Fig. 4 on the exciter output, the CMMS isolator is connected. To avoid other factors influencing the results of the imaging platform, it was replaced by a mass equal to the quality of the imaging platform, such as the bending mode of the imaging platform. For further analysis, two accelerometers measure both the input excitation signal and the vibration output signal of the mass, and then the FRF is obtained.

Furthermore, to test the vibration isolation effect of the CMMS isolation system when it is driven by the main vibration frequency of the drone, the acceleration signals of the accelerometer are recorded, and the corresponding transmissibility is calculated as $T = 20\log|Acc_{out}/Acc_{in}|$.

Table 1 Metastructure parameters

Parameters	Value (mm)	Parameters	Value
a	22	h	4.5 mm
b	18	w	5.0 mm
t	1.5	n	6

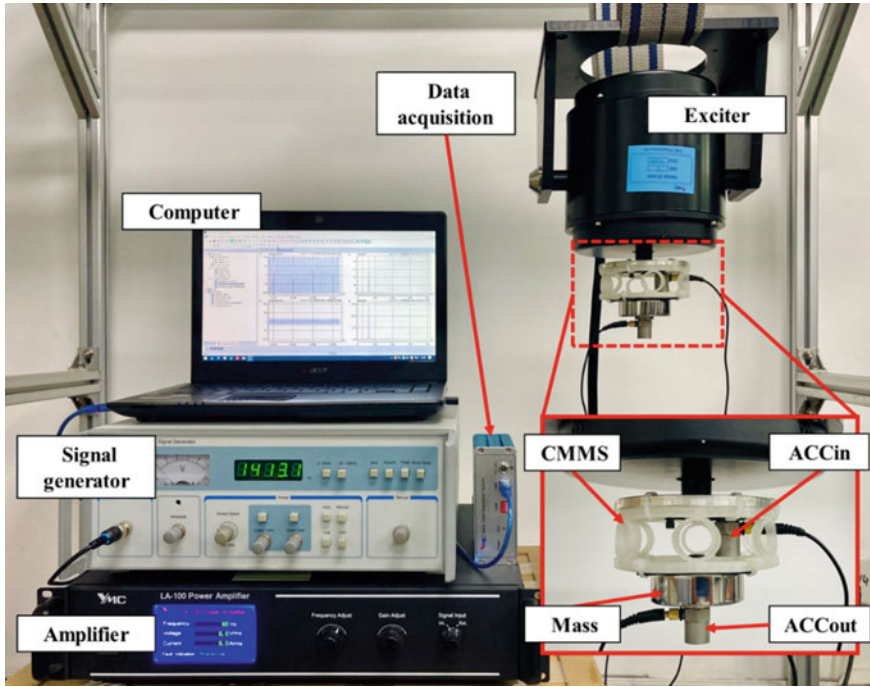


Fig. 4 Experimental setting

5 Data Analysis and Discussion

5.1 Mechanical Properties of the CMMS

According to Fig. 5(a), CMMS isolation is measured by experiment. CMMS's isolator provides excellent low frequency vibration isolation compared to the general rubber isolator. After 28.72 Hz, there is a vibration isolation effect, which is 30 percent lower than rubber isolators in general. A transmissibility measurement in Fig. 5(b) confirms a satisfactory low-frequency vibration isolation effect. CMMS isolators are inferior to general rubber isolators at high frequencies, caused by their high damping coefficient and second-order natural frequencies.

To enhance vibration isolation effects further, the dual-layer CMMS isolator will be studied further, and the performance differences between the double layer CMMS isolator and the composite isolator will be compared. According to the FRF, a composite isolator with rubber and CMMS is able to further reduce the effective vibration isolation frequency. However, this may also reduce the second natural frequency and worsen vibration isolation around 195 Hz. Composite isolators with dual-layer CMMS are capable of improving low-frequency vibration isolation, as

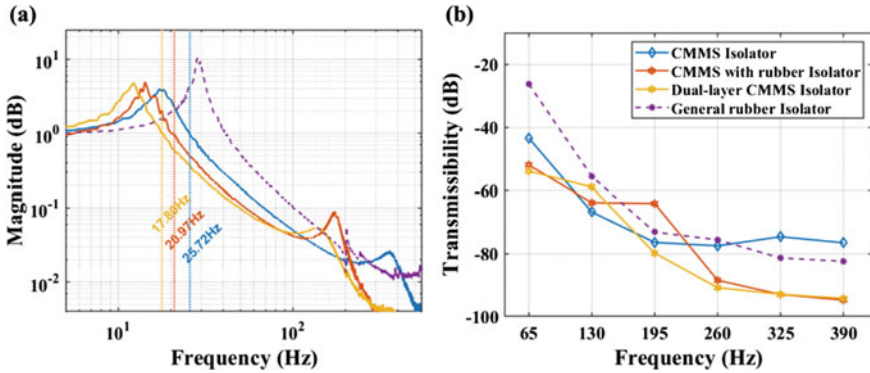


Fig. 5 Cylindrical isolator with different combination. **a** Frequency response function. **b** Experimental transmission

well as improvements in high-frequency vibration isolation. In comparison with a general rubber isolator, this vibration isolation is exceptional.

5.2 *Vibration Isolation Performance of the CMMS with Different Combinations*

For testing, the vibration isolator was installed on a drone imaging platform to verify the performance of the CMMS isolator in practical work. Figure 6 displays the time-domain signals of a drone imaging platform using different combinations of vibration isolators. Clearly, the single-layer CMMS isolator is able to eliminate vibration impacts better than general rubber isolation, especially when the composite isolator employs dual-layer CMMS.

Figure 7 shows the vibration spectrum after various vibration isolators are tested, showing the vibration isolation effect that corresponds to the test environment. In comparison to the general rubber isolator, CMMSs is 97.5% more vibration-isolating in the vertical direction, while dual-layer CMMS is 119.4%.

6 Conclusions

This paper analyzes the vibration characteristics of drones and determines the effect of the drone on the vertical translational vibration of the imaging platform to solve the problem of the unstable drone imaging platform.

An improved vibration isolation device, called CMMS vibration isolation device, is proposed to replace general ones. In experimental research, experimental properties of CMMS were examined and fabricated using additive manufacturing. By

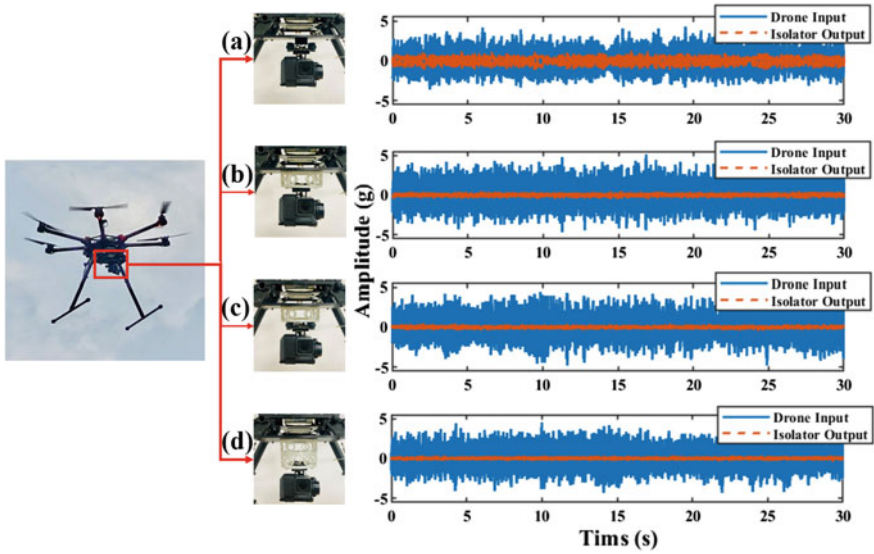


Fig. 6 Time-domain vibration signals of different isolators on drone. (a) General rubber isolator; (b) CMMS isolator; (c) CMMS with rubber isolator; (d) Dual-layer CMMS isolator

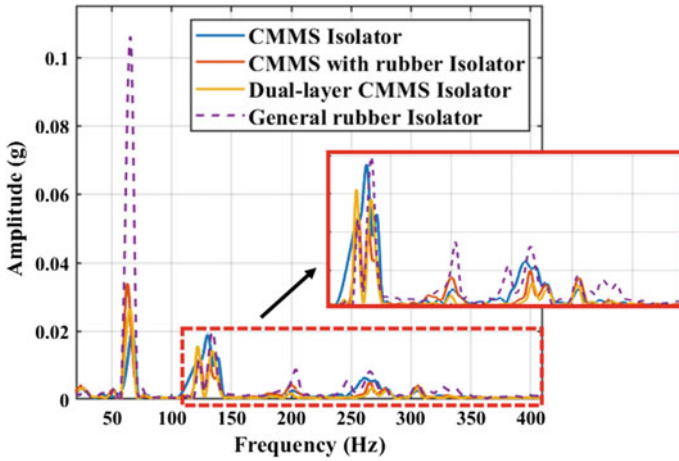


Fig. 7 The spectrum of different isolators on drone

using the single-layer CMMS isolator, low-frequency vibrations of the drone can be effectively suppressed. Dual-layer CMMS composite vibration isolators will help further improve the vibration isolation performance and ensure that drones capture images of high quality.

In the future, the influence factors of CMMS isolator structure parameters on mechanical properties will be discussed in detail, and other nonlinear mechanical

factors will be introduced for modeling and analysis so that the proposed CMMS isolator is not only limited to specific drone use but also potential in other application scenarios.

Acknowledgements Supports from the Guangdong Province Scientific Research Platform (2020KTSCX188) and Innovating major training projects of Beijing Institute of Technology, Zhuhai (XKQC-2019-06) are gratefully acknowledged.

References

1. Ramasamy, M., Ghose, D.: A heuristic learning algorithm for preferential area surveillance by unmanned aerial vehicles. *J. Intell. Rob. Syst.* (2017)
2. Kotaro, H., Kai, W., Mizuho, W., Takahiro, I., Makoto, K., Yoshiaki, B., et al.: Design of UAV-embedded microphone array system for sound source localization in outdoor environments. *Sensors* **17**(11), 2535 (2017)
3. Turner, I.L., Harley, M.D., Drummond, C.D.: UAVs for coastal surveying. *Coastal Eng.* **114**(Aug.), 19–24 (2016)
4. Khadka, A., Fick, B., Afshar, A., Tavakoli, M., Baqersad, J.: Non-contact vibration monitoring of rotating wind turbines using a semi-autonomous UAV. *Mech. Syst. Signal Process.* **138**(Apr.), 106446.1–106446.14 (2020)
5. Yoon, H., Shin, J., Spencer, B.F.: Structural displacement measurement using an unmanned aerial system. *Comput.-aided Civil Infrastruct. Eng.* (2018)
6. Chen, G., Liang, Q., Zhong, W., Gao, X., Cui, F.: Homography-based measurement of bridge vibration using UAV and DIC method. *Measurement* **170**(2) (2020)
7. Liang, C.K., Chang, L.W., Chen, H.H.: Analysis and compensation of rolling shutter effect. *IEEE Trans. Image Process* **17**(8), 1323–1330 (2008)
8. Changshuai, Y., Haitao, L., Siwei, G.: Vibration test and vibration reduction design of UAV load radar. In: Proceedings of the 2019 4th International Conference on Automation, Control and Robotics Engineering—CACRE2019 (2019)
9. Semke, W.H.: Vibration reduction for camera systems onboard small unmanned aircraft. In: Conference Proceedings of the Society for Experimental Mechanics Series, pp. 313–321 (2019)
10. Maës, J.L., Binczak, S., Lhenry, V.: A passive stabilization solution for camera embedded onboard small planes. In: Integrated Communications, Navigation & Surveillance Conference. IEEE (2014)
11. Stuckel, K.J., Semke, W.H., Baer, N., Schultz, R.R.: A high frequency stabilization system for UAS imaging payloads (2011)
12. Bartel, T., Atzrodt, H., Wilczynski, D.: Active vibration reduction on a quadcopter. In: International Conference on Noise and Vibration Engineering, Leuven, Belgium, pp. 131–141. ISMA 2018 KU Leuven, Leuven, Belgium, 17–19 Sept 2018
13. Verma, M., Lafarga, V., Baron, M., Collette, C.: Active stabilization of unmanned aerial vehicle imaging platform. *J. Vib. Control* **26**(19–20), 1791–1803 (2020). <https://doi.org/10.1177/1077546320905494>
14. Verma, M., Lafarga, V., Dehaeze, T., Collette, C.: Multi-degree of freedom isolation system with high frequency roll-off for drone camera stabilization. *IEEE Access* (2020)
15. Wu, L., Wang, Y., Chuang, K., Wu, F., Wang, Q., Lin, W., et al.: A brief review of dynamic mechanical metamaterials for mechanical energy manipulation. *Mater. Today* (2020)
16. Jiang, T., He, Q.: Dual-directionally tunable metamaterial for low-frequency vibration isolation. *Appl. Phys. Lett.* **110**(2), 021907 (2017). <https://doi.org/10.1063/1.4974034>
17. Li, C., Jiang, T., He, Q., Peng, Z.: Stiffness-mass-coding metamaterial with broadband tunability for low-frequency vibration isolation. *J. Sound Vib.* **489**, 115685 (2020)

18. Debeau, D.A., Seepersad, C.C., Haberman, M.R.: Impact behavior of negative stiffness honeycomb materials. *J. Mater. Res.* **33**(3), 290–299 (2018)
19. Chen, S., Wang, B., Zhu, S., Tan, X., Wu, L.: A novel composite negative stiffness structure for recoverable trapping energy. *Compos. Part A Appl. Sci. Manuf.* 105697 (2019)
20. Tan, X., Wang, B., Chen, S., Zhu, S., Sun, Y.: A novel cylindrical negative stiffness structure for shock isolation. *Compos. Struct.* **214**(APR.), 397–405 (2019)
21. Bcab, C., Lcab, C., Bing, D.D., Hla, B., Wlab, C., Df, E.: Novel multifunctional negative stiffness mechanical metamaterial structure: tailored functions of multi-stable and compressive mono-stable. *Compos. Part B: Eng.* **204** (2020)
22. Wu, L., Wang, Y., Zhai, Z., Yang, Y., Jiang, H.: Mechanical metamaterials for full-band mechanical wave shielding. *Appl. Mater. Today* **20** (2020)
23. Fan, H., Yang, L., Tian, Y., Wang, Z.: Design of metastructures with quasi-zero dynamic stiffness for vibration isolation. *Compos. Struct.* **243**
24. Verbeke, J., Debruyne, S.: Vibration analysis of a UAV multirotor frame. In: *International Conference on Noise and Vibration Engineering* (2016)
25. Feng, G., Hu, N., Mones, Z., Gu, F., Ball, A.: An investigation of the orthogonal outputs from an on-rotor MEMS accelerometer for reciprocating compressor condition monitoring (2016)
26. Lu, Z.Q., Gu, D.H., Ding, H., Lacarbonara, W., Chen, L.Q.: Nonlinear vibration isolation via a circular ring. *Mech. Syst. Signal Process.* **136**, 106490 (2019)

Design and Simulation of Broadband Piezoelectric Energy Harvester with Multi-Cantilever



Weiqliang Mo, Shiqing Huang, and Na Liu

Abstract Piezoelectric bimorph cantilever is a typical collecting structure for vibration energy, however it can not adapt to the low frequency and random of vibration excitation in natural environment. In this paper, the physical model of linear vibration system of piezoelectric bimorph cantilever is analyzed, where the piezoelectric energy harvester with multi-cantilever is designed for its disadvantages. By increasing the cantilever beam with different natural frequencies, the energy harvester has the characteristics of broadband. By measuring the dynamic strain under sweep excitation in the simulation of Comsol, Compare the output voltage and working bandwidth between the piezoelectric bimorph cantilever and the piezoelectric energy harvester with multi-cantilever, verify the broadband characteristics of the latter. This paper also designs the rectifier circuit, to convert alternating current from the energy harvester into direct current.

Keywords Piezoelectric · Cantilever · Vibration · Natural frequency · Broadband

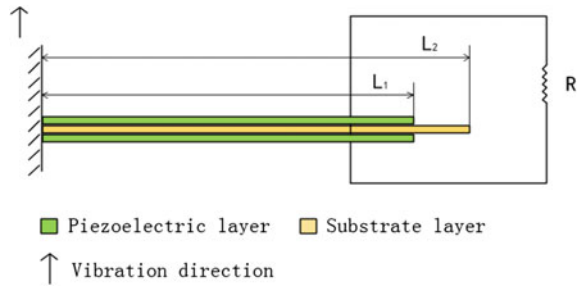
1 Introduction

With the continuous development of the Internet of Things (IoT) and the gradual construction of smart cities, many small electric devices around us, such as smart wearable devices and monitoring devices, can realize information exchange and sharing through the Internet. But these small devices or distributed sensors need to be powered by a separate power source, If we can harness energy from the environment to self-energized the low-power electronics, This will be a very big boost for the development of IoT technology and smart cities [1].

Vibration is a common natural phenomenon and contains a lot of energy. Using vibration energy harvesting technology, which collects vibration energy and uses it in low-power electronic devices. Piezoelectric bimorph cantilever beam is one of the

W. Mo (✉) · S. Huang · N. Liu
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, Guangdong,
China
e-mail: Vacant_mo@163.com

Fig. 1 Piezoelectric bimorph cantilever



typical vibration energy harvesting structures, which is shown in Fig. 1. The structure is simple and low production cost. When working, the free end of the cantilever beam generates inertia force under the external vibrational excitation, which makes the cantilever beam reciprocating deformation. The piezoelectric element pasted on the cantilever beam outputs voltage under the action of alternating stress, thus realizing the conversion of mechanical energy to electric energy. The piezoelectric bimorph cantilever is designed based on the resonance principle [2]. Only when the resonant frequency of the structure is close to or match with the vibration frequency of the environment, it can have higher output voltage and output power. However, vibration in the natural environment is usually characterized by low frequency and randomness, especially in the key working areas of piezoelectric energy harvester, such as highways, bridges, machine tools, etc., with a wide range of vibration frequency. Therefore, it is required that the design of vibration energy harvester must have a wide working frequency band in the frequency range, so that it can effectively respond to the frequency of the vibration source in the environment [3].

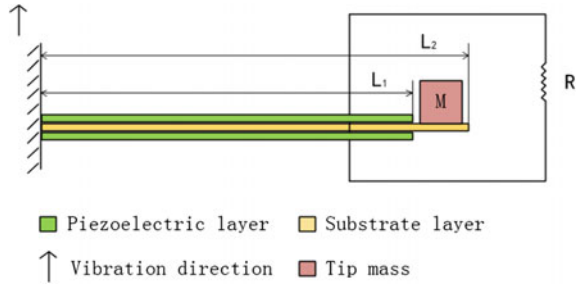
According to the principle of piezoelectric bimorph cantilever and the physical model of its linear system, a broadband piezoelectric energy harvester structure with multi-cantilever is proposed in this paper, the finite element analysis and relevant verification is carried out according to the obtained data.

2 Structure and Analysis of Piezoelectric Cantilever

2.1 Theoretical Analysis of Cantilever Structure

Piezoelectric bimorph cantilever is the most representative and simple energy harvester, which is shown in Fig. 1. The upper and lower layers of the piezoelectric material are the main part of the power generation. For higher output voltage, the two piezoelectric layers are connected in series in the circuit. The piezoelectric layer is made of PZT-5H, and the intermediate layer can be made of copper. In order to improve the energy conversion efficiency and adjust the resonant frequency of the

Fig. 2 Piezoelectric bimorph cantilever with mass



energy harvester, a mass is usually added at the end of the cantilever as shown in Fig. 2 [4].

Piezoelectric bimorph cantilever is a linear vibration system, the equivalent mass of the energy harvester is assumed to be m_i , the equivalent damping is c_i , the equivalent stiffness is k_i , the vibrational displacement of base is z_f , the vibrational displacement of the mass on the free end of the cantilever i to the base is assumed to be z_i , then the dynamic equation of the cantilever i can be expressed as,

$$m_i \ddot{z}_i + c_i \dot{z}_i + k_i z_i = -m_i \ddot{z}_f \tag{1}$$

Both sides of the equal sign divide by m_i , the equation can be written as,

$$\ddot{z}_i + 2\zeta_i \omega_{ni} \dot{z}_i + \omega_{ni}^2 z_i = -\ddot{z}_f \tag{2}$$

$\omega_{ni} = \sqrt{k_i/m_i}$ represents the undamped natural frequency of the cantilever, $\zeta_i = c_i/2\sqrt{k_i m_i}$ represents the damping ratio of the cantilever. According to the dynamics equation, when the vibration element is a cantilever, the response vibration is related to the natural frequency and damping ratio of the system.

2.2 Finite Element Simulation Analysis

Comsol software is a large general finite element analysis software integrating structure, fluid, electric field, magnetic field and sound field analysis, it has the function of realizing multi-physical field and multi-field coupling analysis. In this paper, the piezoelectric analysis module in the software is used to carry out the multi-physical field coupling analysis of solid mechanics and electrostatic field on the piezoelectric bimorph cantilever, research on the change rule between frequency and structure on the piezoelectric bimorph cantilever [5]. According to the schematic diagram of energy harvester in Figs. 1 and 2, the physical model of piezoelectric bimorph cantilever is constructed in the simulation software, as shown in Fig. 3. The structure of the two cantilevers is consistent, except the model shown in Fig. 3b adds a mass.

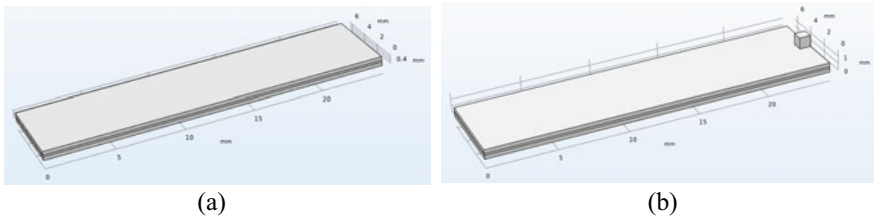


Fig. 3 **a** simulation model without mass, **b** simulation model with mass

The natural frequency of the model shown in Fig. 3 was analyzed. The dimensions of the model are: 24.53 mm long, 6.4 mm wide and 0.67 mm high. The thickness of piezoelectric layer is 0.265 mm. The mass block is a square with a size of 0.6 mm. It is found that the first natural frequency of the simulation model without mass is 545 Hz, when the mass is added, the first natural frequency is 540 Hz. Therefore, the natural frequency of the cantilever can be changed effectively by adding mass. By scanning the above two models in frequency domain, when the frequency of vibrational excitation approaches or matches the natural frequency of the model, the output voltage has a significant increase, as shown in Fig. 4. Moreover, the output

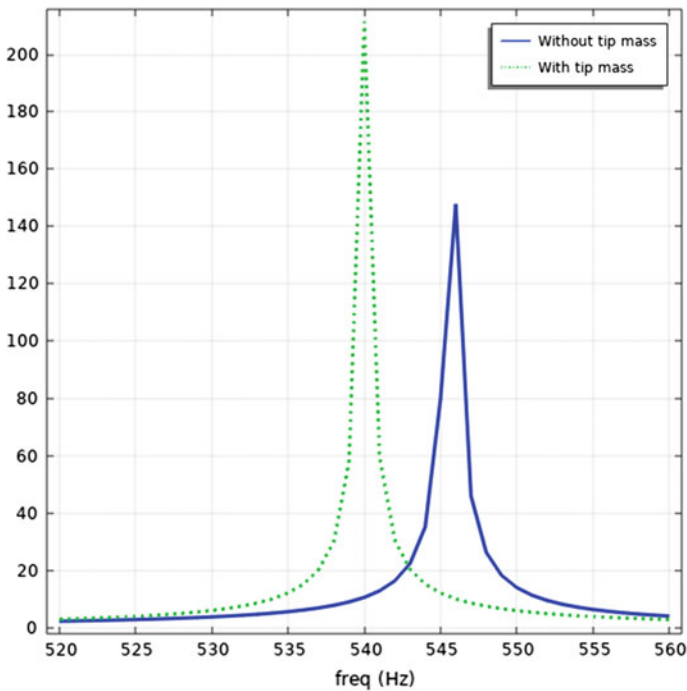


Fig. 4 Influence of mass on frequency and output voltage

voltage can be as high as 210 V when the mass is added, which is higher than without mass. Therefore, adding mass can improve the efficiency of energy conversion.

3 Energy Harvester with Multi-Cantilever

3.1 Structure of Energy Harvester

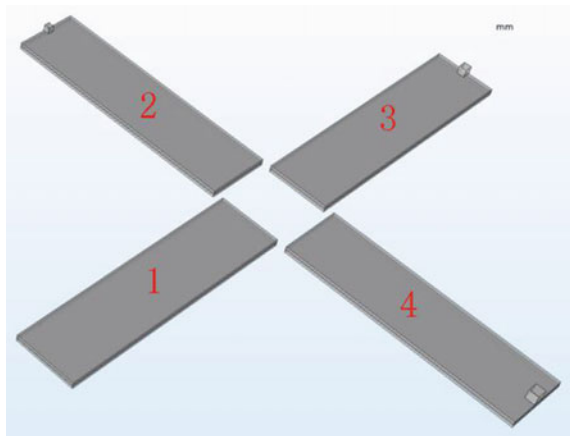
According to the above analysis, although a single piezoelectric bimorph cantilever can achieve collecting vibrational energy, it has a high requirement on the frequency of the excitation vibration source. The natural frequency and energy conversion efficiency of the cantilever can be changed by adding mass. In this paper, the energy harvester with multi-cantilever is designed base on linear multi-frequency resonance method, that is, the working frequency band is widened by increasing the number of resonant peaks in the frequency range [6].

Figure 5 shows the piezoelectric energy harvesting structure of four cantilevers. The structure of the no.1 cantilever is consistent with the ordinary bimorph cantilever. Small(size of 0.6 mm), medium(size of 0.8 mm) and large(size of 1 mm) mass blocks are fixed at the free ends of no. 2, 3 and 4 cantilevers respectively, so as to realize the different natural frequencies of each cantilevers.

The analysis shows that each cantilever is an independent vibration element in the energy harvester. According to expression (2), the motion equation of the whole multi-frequency system can be expressed as,

$$\ddot{z} + 2\zeta\omega_n\dot{z} + \omega_n^2z = -\ddot{z}_f \tag{3}$$

Fig. 5 Energy harvester with multi-cantilever



$\mathbf{z} = [z_1, z_2, z_3, z_4]^T$ is the displacement vector for the entire multifrequency system, so $\mathbf{z}_f = z_f [1, 1, 1, 1]^T [1, 1, 1, 1]^T$.

Given by Thomson and Dahleh [7], the damping ratio matrix is as follows,

$$\boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 & & & \\ & \zeta_2 & & \\ & & \zeta_3 & \\ & & & \zeta_4 \end{bmatrix} \quad (4)$$

The natural frequency matrix is as follows,

$$\boldsymbol{\omega}_n = \begin{bmatrix} \omega_{n1} & & & \\ & \omega_{n2} & & \\ & & \omega_{n3} & \\ & & & \omega_{n4} \end{bmatrix} \quad (5)$$

For the whole energy harvester, under the same external vibration excitation, the vibrational response of each cantilever is only related to their own natural frequency and damping ratio, and the displacements of different cantilever have no influence on each other, they move independently [8].

3.2 Broadband Characteristics of Energy Harvester with Multi-Cantilever

In order to research the broadband characteristics of the multi-cantilever energy harvester structure, the amplitude-frequency characteristics of the structure are analyzed. According to the dynamic equation in expression (2), the amplification factor of the cantilever beam as follow,

$$\beta_i(\omega) = \frac{A_i(\omega)}{B} = \frac{(\omega/\omega_{ni})^2}{\sqrt{[1 - (\omega/\omega_{ni})^2]^2 + (2\zeta\omega/\omega_{ni})^2}} \quad (6)$$

$A_i(\omega)$ represents the amplitude of the vibrational displacement z of the free end relative to the fixed end of the cantilever. B is the amplitude of external excitation, that is, the amplitude of the vibration displacement z_f of the base. The vibration angular frequency of the base is $\omega = 2\pi f$, f is the vibrational frequency of the base. The damping ratio ζ_i is 0.1. Natural frequency analysis by Comsol software shows that the natural frequency of the whole multi-cantilever structure is 546, 544, 540 and 535 Hz. Under the excitation of the same vibration source, the output voltage of no. 1–4 cantilever is shown in Fig. 6.

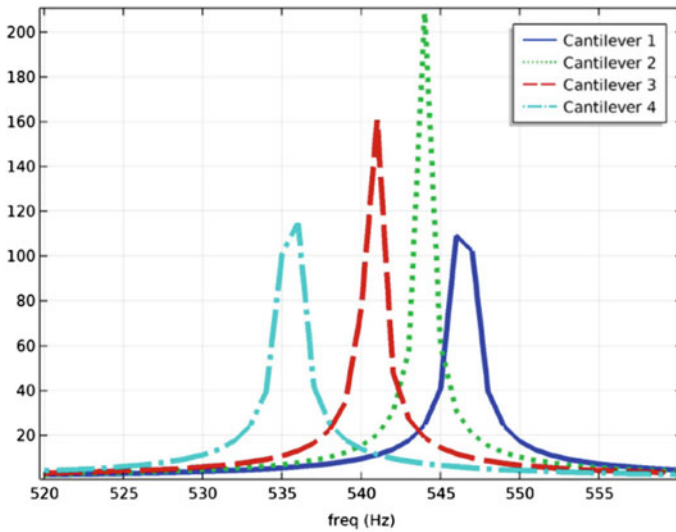


Fig. 6 Output voltage of no. 1–4 cantilever beam

The four curves in the Fig. 6 are the frequency—voltage response curves of four cantilevers with different natural frequencies. Because different vibration elements in a multi-frequency system have different natural frequencies, the resonant peaks of multiple frequency response curves are in different frequency bands, and the peaks do not overlap [9]. Therefore, it can be seen that the total working bandwidth of the multi-frequency system is equal to the sum of the bandwidth of each working frequency band, and the total working bandwidth of the multi-frequency system is greatly expanded, thus presenting broadband characteristics. By connecting the output voltage of each vibration unit in series, the total output voltage of the energy harvester with multi-cantilever can be obtained. Figure 7 shows the total working bandwidth is 534–547 Hz.

Through the natural frequency analysis by Comsol software, it can be seen that the high-order natural frequencies of the multi-cantilever structure are 3288, 3274, 3256 and 3230 Hz, and the excitation frequency range of the vibration source is set to 3200–3350 Hz. The output characteristics of high-order natural frequencies of the multi-cantilever energy harvester are shown in Fig. 8. In the figure, the output voltage is reduced and the maximum value is only close to 9 V, but it still shows the characteristics of broadband. The total working bandwidth is 3220–3300 Hz.

Although the total operating frequency band of multi-frequency system can be effectively widened by the accumulation of resonant frequency bands in different vibration elements, the response amplitudes of different vibration elements are not accumulative. For example, when the vibrational excitation frequency reaches around 540 Hz, the output voltage of no. 3 cantilever in Fig. 6 will increase significantly, but the response amplitude of other cantilever is still small due to the mismatch between the resonant frequency and the vibrational excitation frequency. It means that under

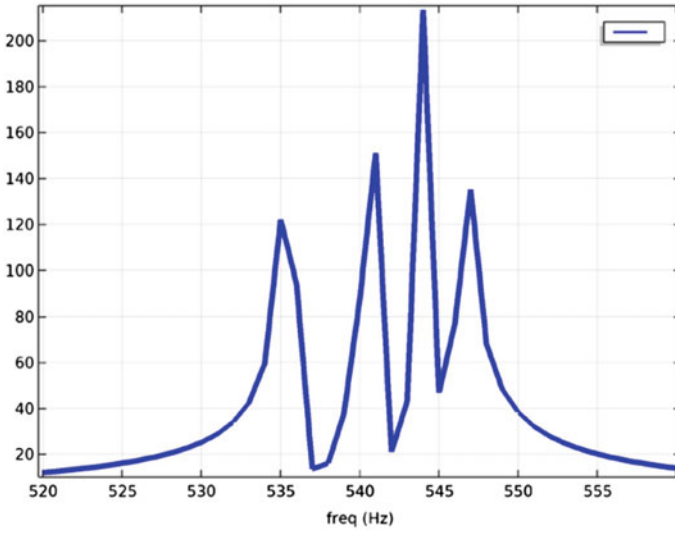


Fig. 7 Output voltage of energy harvester with multi-cantilever

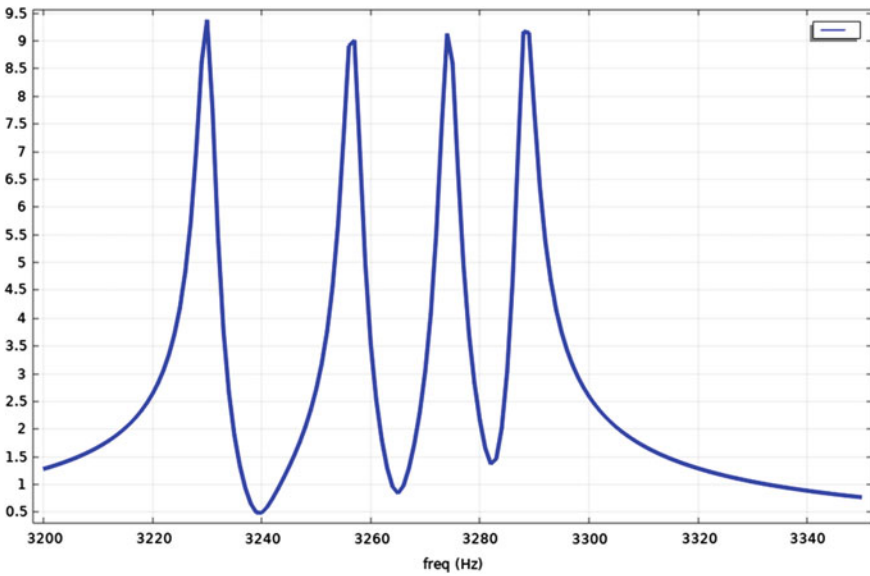
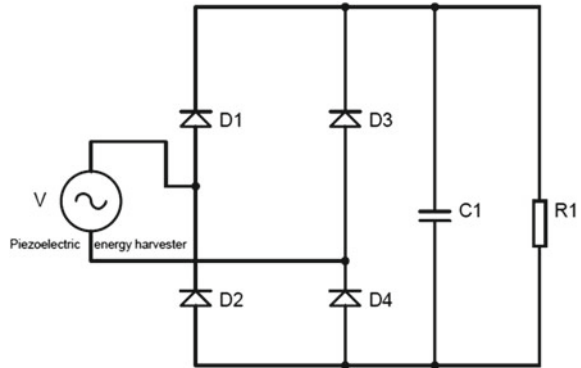


Fig. 8 High-order output characteristics of energy harvester

Fig. 9 Bridge rectifier circuit



the vibration excitation of any particular frequency, at most one vibration unit in the multi-frequency system will be in effective working state, and the vibration unit in the non-resonant state will not work effectively. This is also the reason for the low utilization rate of space structure in existing multi-frequency vibration energy harvester.

4 Energy Harvesting Circuit

4.1 Circuit Design

Usually, the microelectronic equipment uses dc power supply, and the energy harvester generates alternating current, so the rectifier circuit needs to convert the ac electrical signal into the direct current signal suitable for the use of electronic equipment. Given by Elie Lefevre et al. [10], a simple and common energy harvesting circuit is a bridge rectifier circuit, as shown in Fig. 9.

On the basis of the bridge rectifier circuit, the energy harvester circuit as shown in Fig. 10 is improved. The circuit consists of an energy harvester, a capacitor (energy storage device), two inductors, six diodes and a load.

4.2 Circuit Analysis

When the piezoelectric device is in the first half of the vibrational cycle, the terminal A of energy harvester output positive voltage, the current through diode D2, inductor L1, energy storage device capacitor C1, diode D4 back to the terminal B of energy harvester, inductor L1 and capacitor C1 are on charge. When the positive output voltage of terminal A is equal to or less than the voltage of capacitor C1, inductor

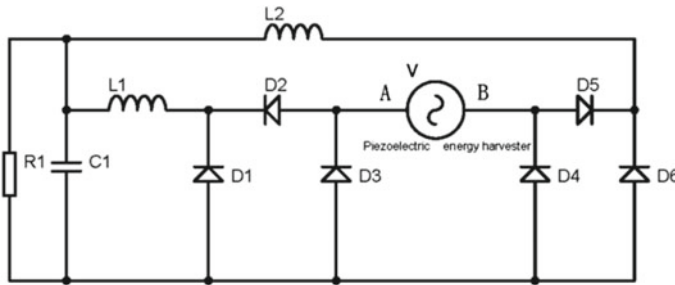


Fig. 10 Energy harvesting circuit

L1 is in the state of negative voltage. At this time, diode D1 is on state, capacitor C1, inductor L1 and diode D1 form a loop, and inductor L1 charges energy storage device C1.

When the piezoelectric device is in the second half of the vibrational cycle, the terminal B of energy harvester output positive voltage, the current through diode D5, inductor L2, energy storage device capacitor C1, diode D3 back to the terminal A of energy harvester, inductor L2 and capacitor C1 are on charge. When the positive output voltage of terminal B is equal to or less than the voltage of capacitor C1, inductor L2 is in the state of negative voltage. At this time, diode D6 is on state, capacitor C1, inductor L2 and diode D6 form a loop, and inductor L2 charges energy storage device C1.

5 Conclusion

In this paper, the theoretical model of piezoelectric bimorph cantilever is analyzed, compared to the structure with mass. And the simulation is carried out by Comsol software, which verifies that the piezoelectric cantilever must work in a specific frequency band. Adding mass can change the working frequency band of cantilever and improve the conversion efficiency of vibration energy.

In view of the limitation of single cantilever on energy collecting, this paper designs energy harvester with multi-cantilever. Through simulation analysis, it is verified that this structure can superimpose the working frequency bands of multiple vibration units, and achieve the purpose of expanding the total working frequency band of multi-frequency system. Compared with a single piezoelectric cantilever structure, energy harvester with multi-cantilever can collect energy under multiple frequency vibration excitation sources, and the energy collection efficiency is higher.

At the same time, the shortcomings of the device: under the vibration excitation of a specific frequency, only one vibration unit is in the working state, resulting in the low utilization rate of the space, which affects the further improvement of energy

conversion efficiency. In view of the existing problems, the research will continue in the subsequent design.

Acknowledgements This paper is one of the phased achievements of "Design and Research of Low-Power Energy Collection Circuit" (2019KQNCX196), a provincial key platform and major scientific research project of Guangdong universities.

References

1. Zhong Lin Wang, F., Tao Jiang, S., Liang Xu, T.: Toward the blue energy dream by triboelectric nanogenerator networks. *Nano Energy* **39**, 9–23 (2017)
2. Lu Wang, F., Libo Zhao, S., Zhuangde Jiang, T.: High accuracy comsol simulation method of bimorph cantilever for piezoelectric vibration energy harvesting. *AIP Adv.* **9**, 095067 (2019)
3. Jaeyun Lee, F., Bumkyoo Choi, S.: Development of a piezoelectric energy harvesting system for implementing wireless sensors on the tires. *Energy Convers. Manage.* **78**, 32–38 (2014)
4. Wang Hai, F., Qiu Wanqun, S., Zhou Xuan, T.: A multiple mass broadband piezoelectric energy harvesting device. *Piezoelectrics Acoustooptics* **37**(6), 1003–1011 (2015)
5. Qian, F., Hajj, M.R. and Zuo, L.: Bio-inspired bi-stable piezoelectric harvester for broadband vibration energy harvesting. *Energy Convers. Manag* **222**, 113174 (2020)
6. Song Juan, F., Liu Qicai, S., Lu Haozan, T.: Design and research of circular bifurcation piezoelectric energy harvester. *Instrum. Tech. Sensor* **7**, 38–42 (2020)
7. Thomson, W.T., Dahleh, M.D.: *Theory of vibration with application*. Prentice Hall, Upper Saddle River (1997)
8. Majid Khazaei, F., Alireza Rezaniakolaie, S., Lasse Rosendahl, T.: A broadband macro-fiber-composite piezoelectric energy harvester for higher energy conversion from practical wideband vibrations. *Nano Energy* **76**, 104978 (2020)
9. Xiaobo Rui, F., Yu Zhang, S., Zhoumo Zeng, T.: Design and analysis of a broadband three-beam impact piezoelectric energy harvester for low-frequency rotational motion. *Mech. Syst. Signal Process.* **149**, 107307 (2021)
10. Elie Lefeuvre, F., David Audigier, S., Claude Richard, T.: Buck-boost converter for sensorless power optimization of piezoelectric energy harvester. *IEEE Trans. Power Electron.* **22**(5), 2018–2025 (2007)

A Mobile Pipeline Leak Monitoring Robot Based on Power Spectrum Correlation Analysis and Sound Pressure Location



Weijie Tang, Rongfeng Deng, Baoshan Huang, Fengshou Gu, and Andrew D. Ball

Abstract Pipes, like blood vessels, play an important role in industry and people's life. Once a pipeline leak occurs, it will bring huge economic losses and even cause serious accidents. Usually, pipeline leak monitoring is carried out by manual inspection or the installation of many sensors, which have great limitations. In this paper, an intelligent mobile robot is proposed for more effectively monitoring the large-scale pipeline systems. Equipped with microphone the robot can be set with the inspection paths, realize the monitoring of leakage anomaly by power spectrum and correlation analysis of sound signal according to the collected real-time data, and pinpoint the location of leakage by calculating the sound pressure. The experimental results show that this method is convenient and effective in a typical industrial environment.

Keywords Pipeline leak · Condition monitoring robot · Correlation analysis · Sound pressure positioning

1 Introduction

Pipeline transportation has the advantages of less economic investment, low energy consumption, high transportation efficiency and easy to realize automatic management. Therefore, pipeline has been widely used in many scenes in production and life. Pipelines are used in many ways and may leak due to corrosion, wear, or third-party damage during their use. Pipeline leakage accidents can cause economic losses and even casualties. Pipeline leakage can be realized through manual inspection. The staff can judge whether there is leakage by monitoring the leakage noise on the

W. Tang · R. Deng (✉) · B. Huang
School of Industrial Automation, Beijing Institute of Technology, Zhuhai, China
e-mail: Rongfeng.Deng2@hud.ac.uk

W. Tang
e-mail: Weijie.tang@hud.ac.uk

W. Tang · R. Deng · F. Gu · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, UK

pipeline. This will cost a lot of manpower and material resources, and it depends on the experience of workers. This can also be done by installing many sensors, but this will greatly increase production costs and make monitoring difficult due to the uncertainty of the location of the leak.

With the development of robot technology, all kinds of robots appear in people's vision. Intelligent inspection robot has incomparable advantages over manual inspection in timely detection of safety hazards. Mobile inspection robot equipped with a variety of sensors, along the set inspection path. Generally, acoustic analysis technology is used to detect pipe leakage, which is very convenient and effective [1–3]. The robot collects environmental data and processes it in real time. When abnormalities are found, it can be further analyzed and prompt management personnel for processing and scheduling in the first time. There is no doubt that intelligent inspection robot will play a very important role in the future intelligent factory.

Due to the difference of fluid properties, operating conditions, pipeline materials and shapes, pipeline leakage rules are different. So far, there is no model that can be fully applicable to all leakage situations. This paper discusses the problem of leakage monitoring and positioning of compressed air in plastic pipelines.

Correlation analysis refers to the analysis of two or more variables with correlation to measure the degree of correlation between them. Correlation analysis has strong anti-jamming ability and is also the most direct method to judge signal similarity. Correlation analysis is a very effective data analysis method, which can be applied not only in pipeline leakage fault diagnosis, but also in other fault diagnosis analysis [4–10].

In this paper, the formation mechanism of pipeline leakage noise is studied, and a method based on correlation analysis of power spectrum is proposed to identify whether there is leakage, because power spectrum can more reflect the internal characteristics of signals compared with time domain. Once the leakage situation is found, the patrol inspection robot is guided to find the location of the leakage sound source according to the relationship between sound pressure and distance. Compared with the method of sensor array, this method cannot accurately locate the sound source, but it makes full use of the mobile robot's flexible motion characteristics and can find the approximate location of the sound source by using a sensor and a simple algorithm. The method of leakage detection and location is very convenient and effective.

2 Characteristics of Leakage Noise

Compressed air pipeline leakage can be regarded as jet flow process. The rapid mixing of the high-speed airflows from the leakage outlet with the surrounding air will cause the local fluid to produce strong pulsating turbulence, thereby produce loud sounds.

In the early 1950s, the mathematical model of jet noise was established by Lighthill based on his similarity, which has played a great role in the study of jet noise. According to Lighthill's theory, the aerodynamic flow generated by turbulence is

regarded as the radiation of the quadrupole sound source, and the sound power of the jet noise is proportional to the eighth power of the jet velocity [11, 12], which is the famous eighth power law.

$$P_T = K \frac{\rho_a^2 \cdot D^2}{\rho_0 \cdot C_0^5} v^8 \quad (1)$$

where K as Lighthill constant, usually value is $3 \times 10^{-5} \sim 1.8 \times 10^{-4}$, D is the diameter of the leakage hole, ρ_0 is the density of the surrounding air, ρ_a is the jet density, C_0 is the speed of the sound, v is the leakage velocity.

The leakage hole is a uniform circular hole, and the wave front can be regarded as a spherical wave, and its sound intensity is:

$$I = \frac{P_m^2}{2\rho_0 C_0} \quad (2)$$

The sound power of leakage is:

$$W = IS = \frac{P_m^2}{2\rho_0 C_0} \cdot 4\pi r^2 \quad (3)$$

where P_m is the maximum sound pressure, r is the radius of the sphere.

According to Eqs. (1) and (2), the maximum sound pressure of the leakage sound field is:

$$P_m = \sqrt{\frac{K}{2\pi}} \cdot \frac{\rho_a D v^4}{r C_0^2} \quad (4)$$

It can be seen from the above equation that the sound pressure is proportional to the diameter of the leakage hole and the fourth power of the leakage velocity. Inversely proportional to the distance from the source.

3 Experimental Scene

The experimental scene as Fig. 1. A condition monitoring robot is patrolling in the laboratory according to the set path as Fig. 2. A sampling point is set next to the equipment to be monitored. When the robot moves to the sampling point, it stops moving and collects audio data for 5 s.

Diameters of the leakage holes are set as 0.4 mm, 0.6 mm and 0.8 mm respectively, as shown in Fig. 3. The pressure of the pipeline is 0.6 MPa, as shown in Fig. 4.

The Turtlebot2 mobile robot is equipped with a PC and can move to different sampling points. For each aperture size, 5 s of audio data are collected at four points A, B, C, and D.

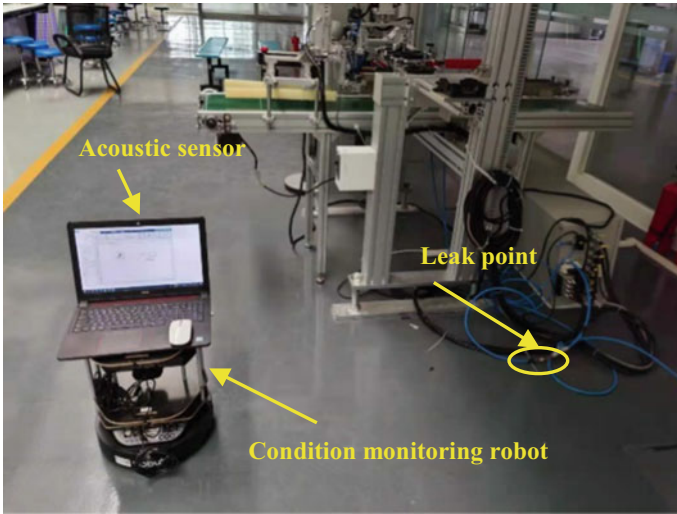


Fig. 1 The experimental scene

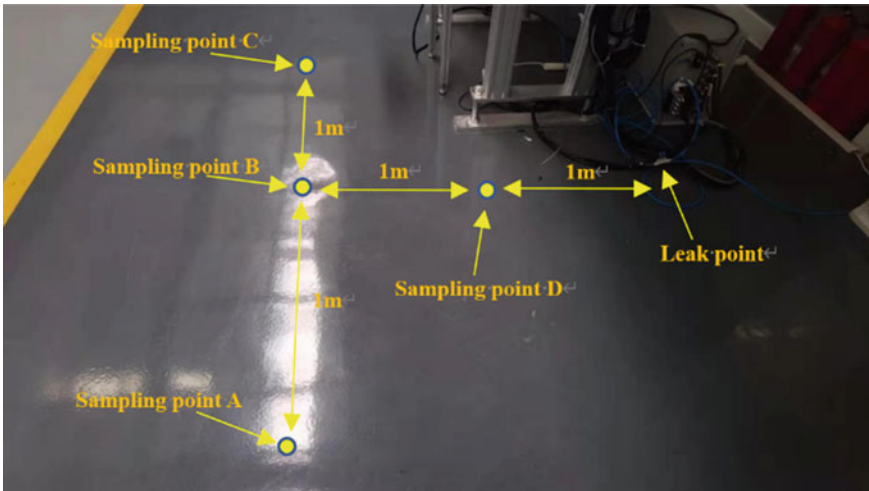


Fig. 2 The sampling points

4 Power Spectrum Correlation Analysis to Identify Leakage

The cross-correlation function between two continuous functions $x(t)$ and $y(t)$ are defined as:

Fig. 3 Size of leak hole



Fig. 4 The pressure of the pipeline



$$R_{xy}(\tau) = \sum_{n=-\infty}^{\infty} x(t)y(t + \tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)y(t + \tau)dt \tag{5}$$

Suppose signal 1 is described as:

$$A(t) = a(t) + s(t) \tag{6}$$

Here, $a(t)$ is the selected fault signal, and $s(t)$ is the noise.

Suppose signal 2 is described as:

$$B(t) = b(t) + v(t) \tag{7}$$

Here, $b(t)$ is the selected fault signal, and $v(t)$ is the noise.

$$\begin{aligned}
 R_{AB}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T A(t)B(t + \tau)dt \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [a(t) + s(t)][b(t + \tau) + v(t + \tau)]dt \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T a(t)b(t + \tau)dt + \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T a(t)v(t + \tau)dt \\
 &\quad + \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s(t)b(t + \tau)dt + \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s(t)v(t + \tau)dt \\
 &= R_{ab}(\tau) + R_{av}(\tau) + R_{sb}(\tau) + R_{sv}(\tau)
 \end{aligned}$$

In general, there is no correlation between noise and signal.

$$R_{av}(\tau) \approx 0, R_{sb}(\tau) \approx 0$$

There is little correlation between the noises.

$$R_{sv}(\tau) \approx 0$$

So

$$R_{AB}(\tau) \approx R_{ab}(\tau)$$

This indicates that the cross-correlation analysis has a strong noise suppression function.

Different from the time domain signal, the power spectrum represents the change of the signal power with the frequency, that is, the distribution of the signal power in the frequency domain, and the power spectrum reflects the internal characteristics of the signal. As can be seen from Eq. (1), the leakage sound power is closely related to the size of the leakage aperture and the leakage velocity, so it is very meaningful to use the power spectrum to conduct correlation analysis.

When there is no leakage, the power spectrum of audio signals collected at four points A, B, C and D is shown in Fig. 5.

The power spectrum of the audio signal collected at four points A, B, C and D with the leakage aperture of 0.4, 0.6 and 0.8 mm is shown in Fig. 6.

Correlation analysis (at zero lag) is carried out directly without considering signal delay. Taking the leak-free audio collected at point D as the benchmark, the correlation analysis of the power spectrum of the audio signals collected at points A, B and C is made, as shown in Table 1.

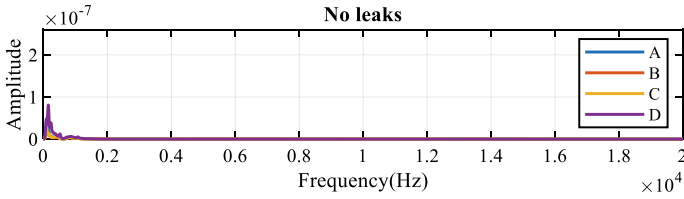


Fig. 5 The power spectrum of the signal collected at four points A, B, C and D when there are no leaks

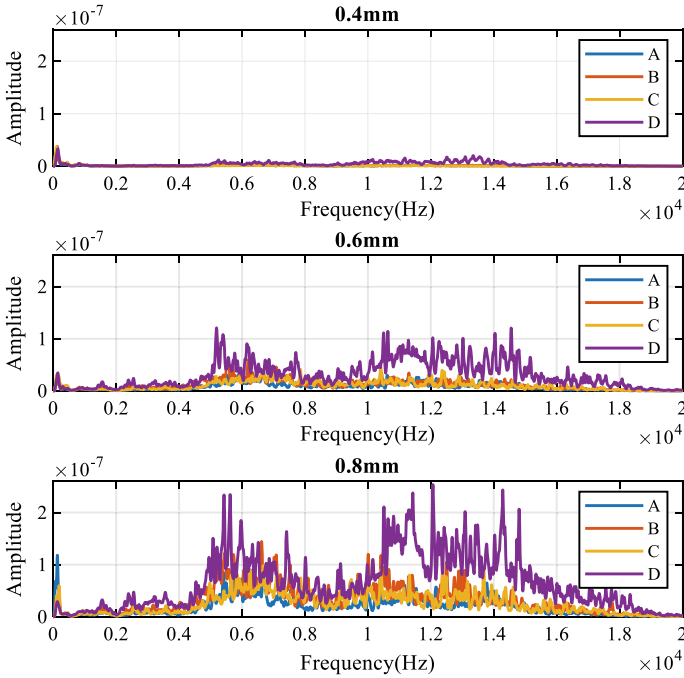


Fig. 6 The power spectrum of leakage signals of different sizes collected at different points

Table 1 The power spectrum correlation calculation results of the audio collected by A, B, C and D without leakage

No leak no leak	A	B	C	D
D	0.8883	0.8695	0.8831	1

In the case of leakage, with aperture of 0.4, 0.6 and 0.8 mm, the correlation analysis of power spectrum between the audio signals collected at point A, point B, point C and point D and the audio signals collected at point D without leakage is carried out in the frequency band above 500 Hz, as shown in Tables 2, 3 and 4.

Table 2 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.8 mm and no leak

Leak (0.4 mm) no leak	A	B	C	D
D	0.3676	0.4130	0.5605	0.1686

Table 3 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.6 mm and no leak

Leak (0.6 mm) no leak	A	B	C	D
D	0.1537	0.1530	0.1684	0.1227

Table 4 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.4 mm and no leak

Leak (0.8 mm) no leak	A	B	C	D
D	0.1426	0.1361	0.1385	0.1222

By comparing the data in Tables 1, 2, 3 and 4, through the correlation analysis of power spectrum, obvious differences can be found, and abnormal situations occur.

The next step is to explore whether the detected anomalies are the result of leaks. Human voice and knocking sound are introduced, and their power spectra are shown in Fig. 7. Correlation analysis of power spectrum between leakage conditions of different sizes and human voice and knocking sound is conducted, as shown in Tables 5, 6 and 7.

It can be seen from the above results that the correlation between the leakage signal and the human voice and the tapping sound is very low. In the case of leakage, the

Fig. 7 The power spectrum of the human voice and knock sound

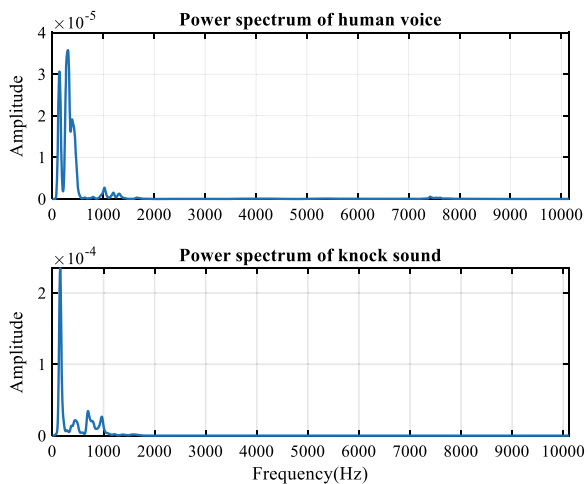


Table 5 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.4 mm and human voice and knock sound

Leak (0.4 mm) sound type	A	B	C	D
Human voice	0.2676	0.2972	0.3572	0.1083
Knock sound	0.2324	0.2751	0.4385	0.0629

Table 6 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.6 mm and human voice and knock sound

Leak (0.6 mm) sound type	A	B	C	D
Human voice	0.1169	0.1359	0.1375	0.0878
Knock sound	0.0400	0.0330	0.0504	0.0180

Table 7 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.8 mm and human voice and knock sound

Leak (0.8 mm) sound type	A	B	C	D
Human voice	0.1163	0.1196	0.1083	0.0891
Knock sound	0.0301	0.0201	0.0264	0.0162

correlation analysis of the power spectrum of the data collected at different aperture and sampling points is carried out, as shown in Tables 8, 9, 10, 11, 12 and 13.

From the above analysis, the correlation coefficient between the leakage signal collected at different aperture and different position and the non-leakage, human

Table 8 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.4 mm

Leak (0.4 mm) leak (0.4 mm)	A	B	C	D
A	1	0.9131	0.8963	0.8274
B	0.9131	1	0.9145	0.7822
C	0.8963	0.9145	1	0.7186
D	0.8274	0.7822	0.7186	1

Table 9 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.6 mm

Leak (0.6 mm) leak (0.6 mm)	A	B	C	D
A	1	0.8930	0.8665	0.8680
B	0.8930	1	0.9038	0.8478
C	0.8665	0.9038	1	0.8342
D	0.8680	0.8478	0.8342	1

Table 10 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.8 mm

Leak (0.8 mm) leak (0.8 mm)	A	B	C	D
A	1	0.8909	0.8959	0.8763
B	0.8909	1	0.9010	0.8506
C	0.8959	0.9010	1	0.8544
D	0.8763	0.8506	0.8544	1

Table 11 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.4 and 0.6 mm

Leak (0.6 mm) leak (0.4 mm)	A	B	C	D
A	0.9158	0.8803	0.8812	0.8523
B	0.8457	0.8964	0.8659	0.7924
C	0.7785	0.8038	0.8307	0.7266
D	0.8228	0.7948	0.8240	0.8754

Table 12 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.4 and 0.8 mm

Leak (0.8 mm) leak (0.4 mm)	A	B	C	D
A	0.9162	0.8613	0.8780	0.8576
B	0.8542	0.8807	0.8448	0.8027
C	0.7811	0.7897	0.8006	0.7197
D	0.8360	0.8315	0.8364	0.8865

Table 13 The correlation calculation results of the power spectrum of the leakage audio under the aperture of 0.6 and 0.8 mm

Leak (0.8 mm) leak (0.6 mm)	A	B	C	D
A	0.9415	0.8665	0.8814	0.8815
B	0.9012	0.9479	0.9037	0.8474
C	0.8764	0.9073	0.9492	0.8389
D	0.8799	0.8553	0.8396	0.9469

voice and percussion sound is very small, while the correlation coefficient of the leakage signal is very large. The cross-correlation analysis of power spectrum can be used to determine whether there is leakage, but because the correlation coefficient between the leakage signals of different aperture is very close, it is difficult to distinguish which aperture leakage belongs to.

5 Sound Pressure Location of Leakage Source

In the fourth part, the inspection robot detects whether there is a leak using the correlation analysis of power spectrum. According to Eq. (4), sound pressure is proportional to the fourth power of the diameter of the leakage hole and the leakage velocity. Inversely proportional to the distance from the source. The sound pressure is then used to guide the robot to locate the source of the leak.

It can be clearly found from Fig. 8 that the measured sound pressure increases when the leak aperture becomes larger. The robot’s patrol path moves from point A to point C. After analyzing the data collected at points A, B and C, it is found that point B is closer to the sound source, so the robot returns to point B. After reaching point B, the robot was still unable to locate the sound source. Therefore, the robot rotates counterclockwise along the forward direction at point B and collects audio data from four angles of 0°, 90°, 180° and 270°, as shown in Fig. 9.

During the experiment, the data collected with an aperture of 0.4 mm were used for comparative analysis again, as shown in Fig. 10.

According to the data comparison and analysis in Fig. 10, when the robot rotates to 90°, the measured sound pressure value is the largest, indicating that the sound source is close to this direction. In fact, the microphone is facing the direction of the leakage sound source currently to guide the robot to move in this direction. In the experiment, the robot moved to point D, collected data again, and compared the collected results with points A, B and C for verification, as shown in Fig. 11.

It can be clearly seen from Fig. 11 that point D is closest to the sound source and has the largest value. Point B is farther away from the sound source, and its value is smaller than that of Point D. Point A and point C are further away, so they are smaller. As can be seen from the above experiments, this method takes advantage of the qualitative relationship among distance, direction, and sound pressure as well as the flexibility of robot movement. It does not need to increase the number of

Fig. 8 The relation between the power value of the sound aperture and different position source at different

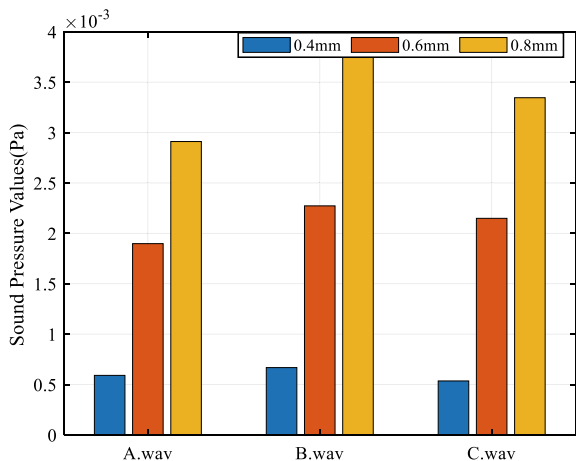


Fig. 9 The robot collects audio data in different directions

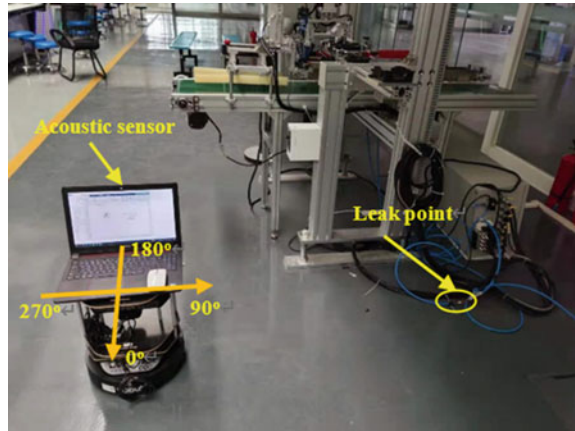
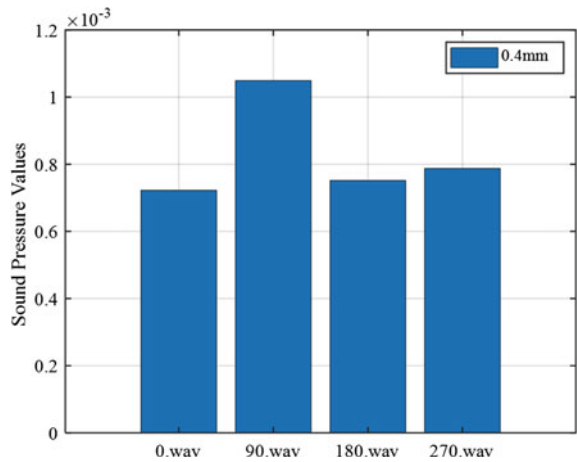


Fig. 10 Comparison of data collected from four directions of 0°, 90°, 180° and 270°

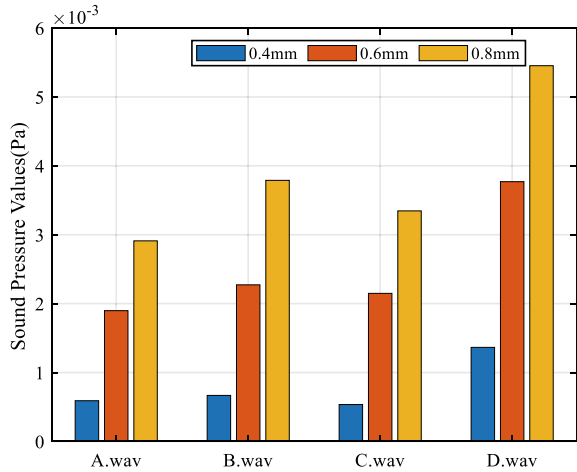


additional sensors, but only needs one sensor to guide the inspection robot to find the leak point.

6 Conclusion

The leakage noise of pipeline compressed air is discussed in this paper. The relationship between the sound pressure and the leakage aperture and the distance from the sound source is obtained by using Lighthill equation. A power spectrum containing more characteristic information is proposed to carry out correlation analysis. The patrol robot can judge whether there is abnormal leakage. After determining the

Fig. 11 Comparison of data collected at four positions A, B, C and D



existence of leakage, the sound pressure is used to guide the robot to find the location of the leakage source. This method is convenient, fast, economical, and efficient, and easy to be deployed on the inspection robot. However, it also has some defects, which should be improved in the future research. For example, it is impossible to distinguish which size of leakage hole is generated, and the location of sound source is not very accurate.

Acknowledgements This work was supported by the School Scientific Research Fund Project of Beijing Institute of Technology, Zhuhai, under Grant No. XK-2018-29 and Innovating Major Training Projects of Beijing Institute of Technology, Zhuhai (XKQC-2019-06).

References

1. Fantozzi, M., Di Chirico, G., Fontana, E., Tonolini, F. (1993). Leak inspection on water pipelines by acoustic emission with cross-correlation method. In: Annual Conference Proceeding, American Water Works Association, Engineering and Operations, pp. 609–621, June 1993
2. Fuchs, H.V., Riehle, R.: Ten years of experience with leak detection by acoustic signal analysis. *Appl. Acoust.* **33**(1), 1–19 (1991)
3. Liston, D.A., Liston, J.D. Leak detection techniques. *J. N. Engl. Water Works Assoc. JNEWA* **6**, **106**(2) (1992)
4. Pereira, E.L., Deschamps, C.J.: Numerical analysis and correlations for radial and tangential leakage of gas in scroll compressors. *Int. J. Refrig* **110**, 239–247 (2020)
5. Ebrahimi-Moghadam, A., Farzaneh-Gord, M., Deymi-Dashtebayaz, M.: Correlations for estimating natural gas leakage from above-ground and buried urban distribution pipelines. *J Nat. Gas Sci. Eng.* **34**, 185–196 (2016)
6. Ma, D., Deng, J., Zhang, Z.: Correlation analysis for online CO2 leakage monitoring in geological sequestration. *Energy Procedia* **37**, 4374–4382 (2013)
7. Zhang, B., An, L.S., Sheng, G.Q., Li, Y.: Error analysis of acoustic pyrometer based on cross-correlation method. *Electr. Power Sci. Eng* **1** (2006)

8. Jinshan, L., Chunhong, D., Ni, Z.: Fault diagnosis of gearbox based on multifractal detrended cross-correlation analysis. *J. Mech. Transm.* **1**(40), 91–94 (2016)
9. Bourne, M.M., Whaley, J., Dolan, J.L., Polack, J.K., Flaska, M., Clarke, S.D., Tomanin, A., Peerani, P. and Pozzi, S.A.: Cross-correlation measurements with the EJ-299-33 plastic scintillator. *Nucl. Instrum. Methods Phys. Res. Sect. A: Accelerators, Spectrometers, Detectors and Associated Equipment* **784**, 460–464 (2015)
10. Clarke, S.D., Flaska, M., Pozzi, S.A., Peerani, P.: Neutron and gamma-ray cross-correlation measurements of plutonium oxide powder. *Nucl. Instrum. Methods Phys. Res., Sect. A* **604**(3), 618–623 (2009)
11. Lighthill, M.J. On sound generated aerodynamically I. General theory. *Proc. Royal Soc. London. Series A. Math. Phys. Sci.* **211**(1107), 564–587 (1952)
12. Bose, T.: Lighthill's theory of aerodynamic noise. In: *Aerodynamic Noise*, pp. 53–60. Springer, New York, NY (2013)

A Review of Acoustic Emission Monitoring on Additive Manufacturing



Zhen Li, Xinfeng Zou, Fanbiao Bao, Fengshou Gu, and Andrew D. Ball

Abstract Additive manufacturing has the characteristics of gradual accumulation of materials in the manufacturing process. It is often superimposed layer by layer in the process of material physical shape change, which may be accompanied by hot melting, liquid material solidification, particle sintering and other processes. Due to the influence of physical environment, machine state, manufacturing principle and other factors in the whole process, performance defects of parts may occur. The traditional monitoring methods such as vision, optics and CT tomography have limitations, or can only observe the defects on the outer surface, or it is difficult to find the defects in time in the processing process, or the micro defect identification accuracy is not enough. A series of research on acoustic emission detection technology, due to the high sensitivity to high-frequency signals, can observe various phenomena of the machine itself in the processing procedure, and monitor the spatial micro faults of the whole part in the process of parts made of additive materials.

Keywords Additive Manufacturing (AM) · Acoustic Emission (AE) · Condition monitoring · Machine learning

Z. Li · X. Zou (✉) · F. Bao

School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, Guangdong, China

e-mail: xfzou_mo@outlook.com

Z. Li

e-mail: Zhen.Li@hud.ac.uk

Z. Li · X. Zou · F. Gu · A. D. Ball

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1, UK

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_70

1 Introduction and Application Trend of Additive Manufacturing

Different from the material removal method in traditional NC machining, AM uses the method of accumulation to process parts. Generally, the materials usually include liquid, powder and wire [1], has obvious digital characteristics, integrating new materials, computer software, control technology and other technologies. It includes two parts: first, in the data processing stage, the computer-aided design model is hierarchically and analyzed into the operation procedures of the corresponding machine tool; The second is layered manufacturing. According to the analyzed machine operation procedures, materials are superimposed from bottom to top [2, 3, 4].

AM is suitable for rapid manufacturing of complex structures. It is generally used for personalized customization, especially for manufacturing special parts [5, 6]. AM is often used for prototype verification and small batch trial production in the process of product R&D. With the maturity of technology, the scope of application is little by little expanded from manufacturing prototype and model to direct manufacturing of working parts [7]. AM is progressively applied to manufacturing parts with special and difficult structures [8]. The development of AM technologies has solved the limitation that it is difficult to manufacture parts with complex graphics and structure through computer model. The key point and main difficulty of AM technologies is that the mechanical characteristics of the manufactured parts are not high compared with the material reduction method [9, 10]. In customization requirements and some special application scenarios, such as medical printing dentures and various human implants, there is an increasing demand for performance. There have been studies on using composites as construction materials to improve hardness [11], promoting the performance of parts is the direction of future efforts [12]. There are many reasons for the low mechanical properties. Complex geometry, unreasonable process parameters or changes in the properties of raw materials may cause product defects [13]. There are many methods to evaluate the quality of completed parts, such as optical backscatter reflection (OBR) [14]. X-ray computed tomography (CT) internal voids [15, 16]. The visual evaluation of surface roughness is transformed into the problem of visual image texture roughness recognition [17]. AE monitoring technology can also be used for quality monitoring of AM [18]. The goal of reducing the processing cost can be achieved by finding the defects and problems of parts in-situ of manufacturing rather than after the parts are processed [19].

2 Principle, Development and Application of Acoustic Emission

At present, AE technology has developed into a mature nondestructive testing method, which is applied in many industries and fields for all kinds of testing and monitoring [20, 21]. It mainly includes the following aspects:

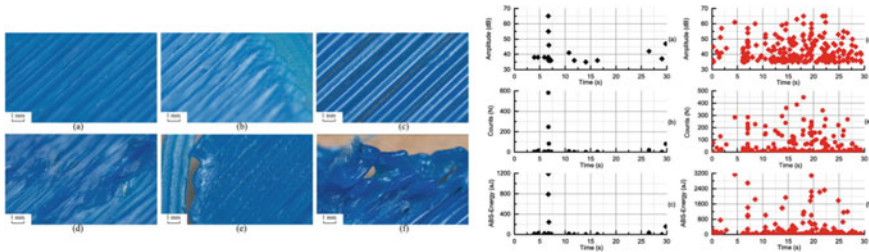


Fig. 1 The distribution of AE hit rate corresponding to printing failure [18]

Detection of damage point of closed container [22]. Welding point quality monitoring [23]. Various load and failure tests [24]. Deterioration detection of masonry infrastructure [25]. State detection of NC cutting tools [26]. Bearing residual life prediction [27]. Grinding wheel condition monitoring in grinding process [28]. In the process of AM, AE will occur during the processing of workpieces [18]. Typical faults, such as warpage and deformation, can be effectively monitored by the time-domain distribution of AE hits (see Fig. 1).

3 Acoustic Emission Phenomenon in Additive Manufacturing

3.1 Melting Phase Transformation Process

AE phenomena in the process of material phase change, such as the melting process of water from solid ice to liquid. AE during the melting of homogeneous liquid can be detected. Therefore, AE is a suitable technology to monitor the phase transition process [29].

Vorontsov researched the phase transformation process of high-purity aluminum. The change of melt structure is reflected by AE. Experiments (see Fig. 2) show that the AE can reflect the information of a melting process of substances with crystal structure [30].

In the process of laser welding, there is also AE in the process of material phase transformation. L. Schmidt studied AE monitoring in this process, it shows that using AE technology for process monitoring has a good effect [31]. Plastic products have structural complexity and defects in production. In Florian Muller’s research, a fresh AE monitoring production method is adopted, which is compared with the traditional monitoring methods of injection mold cavity pressure and workpiece temperature [32]. The monitoring method of acoustic emission provides a unused idea.

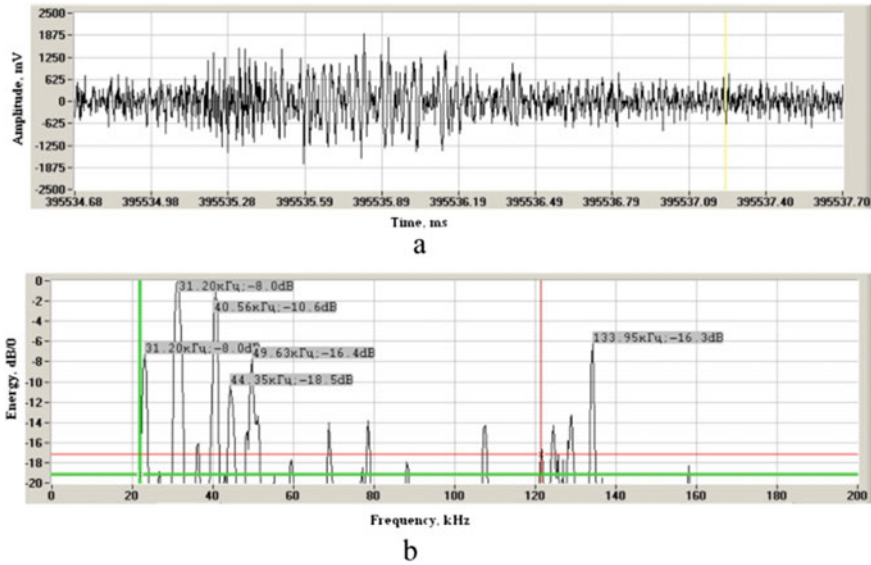


Fig. 2 AE signal of aluminum at 840 °C and frequency domain distribution [30]

Some processes and construction processes of AM, along with the melting process, have been studied by using AE to monitor parts in the process of material phase transformation. F. Li proposed a real-time monitoring method based on AE [33].

J. Nam established a framework for health monitoring and diagnosis during melt deposition modeling (see Fig. 3). Acceleration, acoustic emission and thermocouple are used to collect data. The health monitoring and diagnosis model of FDM process is established by using multiple groups of data [34].

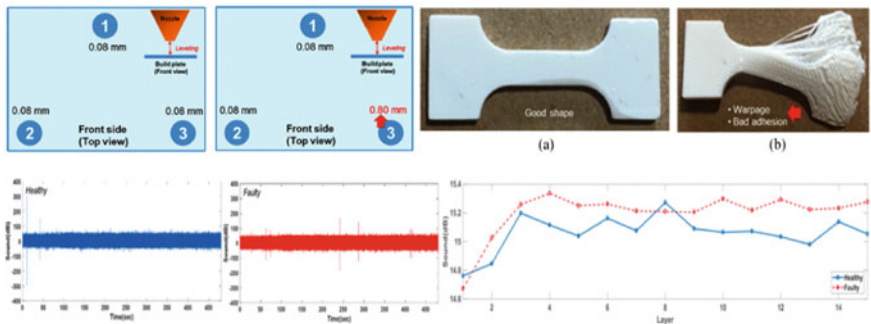


Fig. 3 AE signal in two states of health and fault during FDM modeling [34]

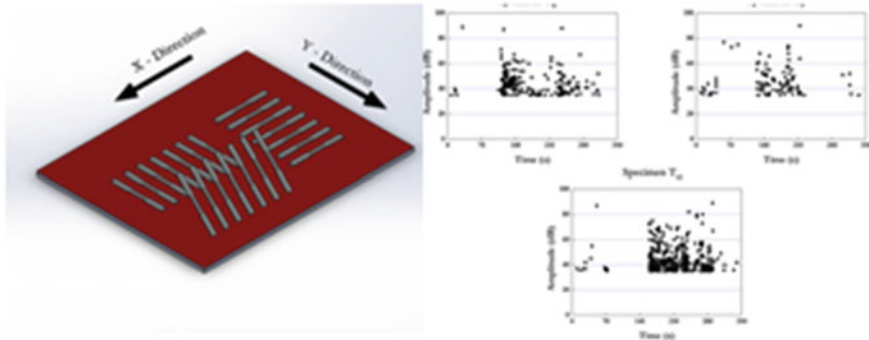


Fig. 4 AE distribution of parts in different construction directions in tensile test [39]

3.2 Pressure and Tensile Test

In order to judge the internal defects of product made by AM, such as filament falling off [35], the AE phenomenon in the experiment can be collected through pressure and tension experiments, and the defects of parts can be judged through the signals of materials in the process of pressure or tension loading [36]. This method can be used as an effective mechanical property detection method [37]. The fracture process of FDM wire feeding mechanism is studied by Z. Yang. A framework based on instantaneous line break detection is proposed, and the original AE waveform is analyzed [38].

Barile first used a selective laser melting technique from different structural directions (see Fig. 4). Evaluate the AE signals of parts under tensile load. The results show that AE technology can effectively monitor the damage in the tensile process [39].

3.3 Stereo Lithography Appearance Process

Resin 3D printing is a relatively early 3D printing technology. In dental research, the use of resin light curing in tooth filling technology is a very mature technology, and its mechanism and construction process are similar to resin light curing 3D printing. N. Choi proved the AE phenomenon in the light curing process through dental research. The curing was monitored by AE technology (Fig. 5). Practice has proven that AE technology can be used as an evaluation index of edge disintegration fracture [40].

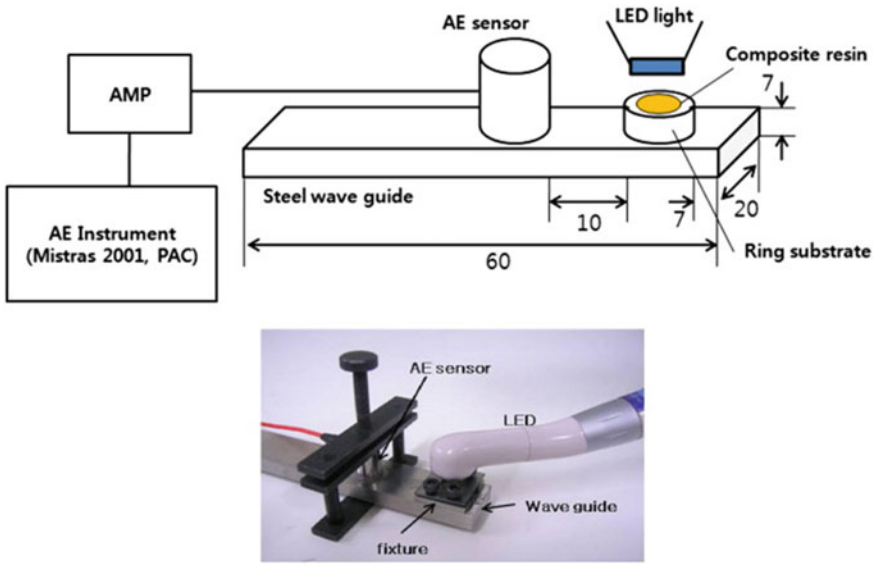


Fig. 5 Schematic diagram of UV curing experiment monitored by AE [40]

3.4 Stacking Process: Warpage, Deformation, Fracture, Shrinkage

AE technology can detect small signals. As a kind of AM, FDM has curled and relaxation defects. F. Li proposed a real-time monitoring method of part deformation based on acoustic emission. Defects are identified by sensing and digital signal processing techniques [33]. Collect and identify acoustic emission signals in normal, loose and curled states [41]. Other materials can also use AE technology. The fracture strain and service life of magnesium alloy were evaluated by tensile test [42]. The application of glass/carbon fiber reinforced plastics (GFRP/CFRP) in AM has attracted more and more attention. However, the residual stress generated in the manufacturing process of parts and assemblies will lead to warpage and affect the mechanical properties of the composite structure [43].

4 Application of Acoustic Emission in Additive Manufacturing

AE technology is widely used in am process monitoring and part performance testing, due to the high adoption rate. The application of acoustic emission signal analysis will face the challenge of massive data.

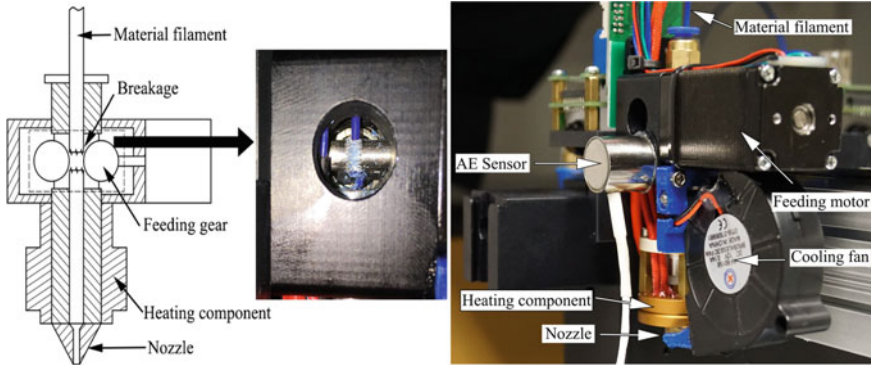


Fig. 6 Identify FDM machine fault according to AE hit

4.1 Machine Condition Monitoring

The reliability of desktop FDM printer may affect the product quality during operation. In order to improve the final printing effect, the monitoring of machine status and the analysis of multi factor influence in the printing process can provide effective guidance for the last printing quality. H. Wu research proposed a method of monitoring FDM machine by using AE technology [44].

H. In Wu's experiment, the AE sensor was installed at the nozzle of 3D printer (see Fig. 6). According to the collected AE hits, machine learning was used to distinguish between normal and abnormal states of the machine [45]. J. Liu used acoustic emission sensors to diagnose faults during material extrusion of FDM [46].

4.2 Fault Classification and Degree Evaluation

There are different types and degrees of part faults and quality defects in the process of AM. Judging the type or degree of fault by signal is conducive to accurately monitoring the whole process [47, 48]. AE hit is the data intercepted from waveform signals through threshold setting. It is a data suitable for fault classification using deep learning method.

S. A. Shevchik's research proves that AE signals can be used for fault classification in selective laser melting technology (SLM). By adjusting the process parameters of the machine to simulate the faults in manufacturing processes, holes of different sizes and densities are formed in the parts. Define poor, medium and high part quality standards, can be classified by AE signal [49].

In the laser powder layer fusion (LPLF) process, the part density is used as the standard to measure the part performance (see Fig. 7). The density measurement method

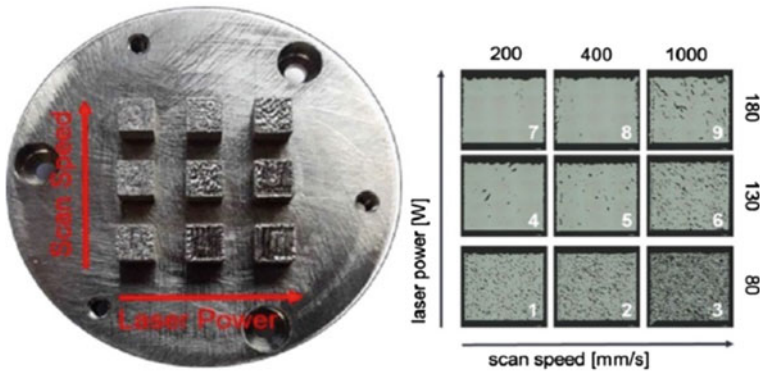


Fig.7 Density is used to mark AE signal [50]

is used to calibrate the AE data. Through supervised learning, data are classified according to density measurements [50].

4.3 Fault Location

The application in AM process is affected by specific experimental conditions and physical media, so it is difficult to locate. K. Ito monitors micro defects in SLM process in real time. The time and location of AE events were compared with the specimen cross section observed by X-ray computed tomography (CT) (see Fig. 8). It proves the feasibility of AE monitoring technology for monitoring the time and location of AE events in the process of SLM [51].

5 Conclusion

It is considered that AE can effectively detect the quality of parts and monitor the manufacturing process. It is worth pointing out that AE technology has more research in metal printing. AE technology is more and more studied in AM and applied to machines with different mechanisms. It is also of reference value to refer to processing processes similar to AM process mechanism, such as welding process, injection molding, 3D printing, building laminated manufacturing, etc. Summarizing the research contents in recent years, we can know the following application modes of AE technology in AM: Monitoring the machine state; Location of defects; Evaluate the degree of manufacturing defects; Classify manufacturing defects.

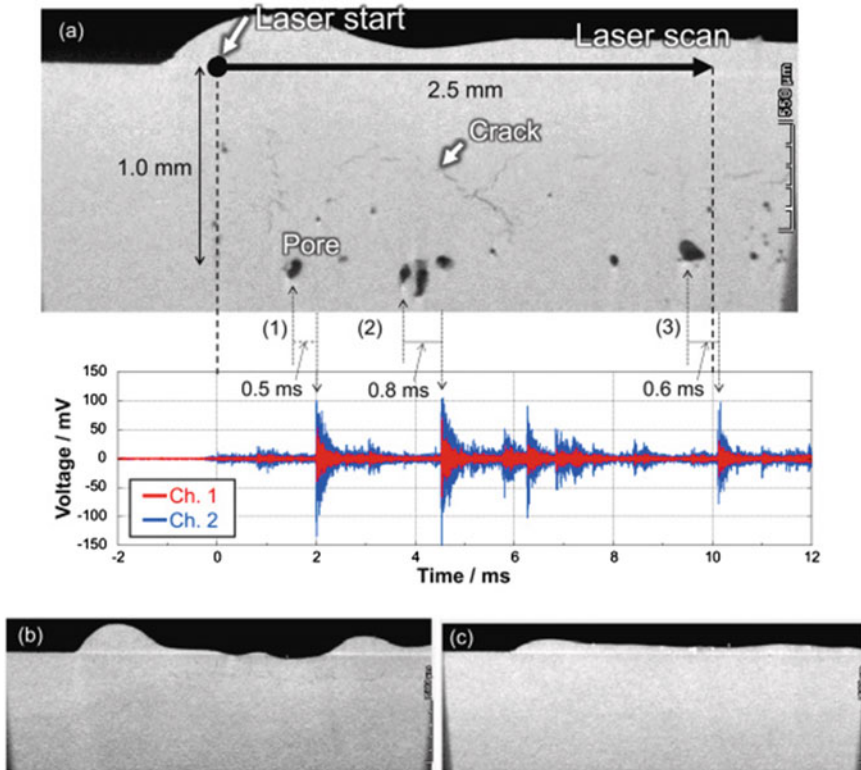


Fig. 8 CT scanning tomography corresponds to the AE signal [51]

References

1. Behera, D., Cullinan, M.: Current challenges and potential directions towards precision microscale additive manufacturing—Part I : direct ink writing/jetting processes. *Precis. Eng.* **68**(Apr 2020), pp. 326–337 (2021). <https://doi.org/10.1016/j.precisioneng.2020.12.009>
2. Behera, D., et al.: Current challenges and potential directions towards precision microscale additive manufacturing—Part II : laser-based curing, heating, and trapping processes. *Precis. Eng.* **68**(Apr 2020), pp. 301–318 (2021). <https://doi.org/10.1016/j.precisioneng.2020.12.012>
3. Chizari, S. et al.: Current challenges and potential directions towards precision microscale additive manufacturing—Part III : energy induced deposition and hybrid electrochemical processes. *Precis. Eng.* **68**(Apr 2020), pp. 174–186 (2021). <https://doi.org/10.1016/j.precisioneng.2020.12.013>
4. Behera, D., et al.: Current challenges and potential directions towards precision microscale additive manufacturing—part IV : future perspectives. *Precis. Eng.* **68**(April 2020), pp. 197–205 (2021). <https://doi.org/10.1016/j.precisioneng.2020.12.014>
5. Jared, B.H., et al.: Scripta Materialia additive manufacturing: toward holistic design. *Scr. Mater.* **135**, 141–147 (2017). <https://doi.org/10.1016/j.scriptamat.2017.02.029>
6. Olsson, A., Helsing, M.S., Rennie, A.R.: New possibilities using additive manufacturing with materials that are difficult to process and with complex structures (2017)

7. Kim, H., Lin, Y., Tseng, T.B.: A review on quality control in additive manufacturing. *3*(June 2017), pp. 645–669 (2018). <https://doi.org/10.1108/RPJ-03-2017-0048>
8. Zhang, K., Cheng, G., Xu, L.: Topology optimization considering overhang constraint in additive manufacturing. *Comput. Struct.* **212**, 86–100 (2019). <https://doi.org/10.1016/j.compstruc.2018.10.011>
9. Okarma, K., Fastowicz, J., Lech, P., Lukin, V.: Quality assessment of 3D printed surfaces using combined metrics based on mutual structural similarity approach correlated with subjective aesthetic evaluation. *Appl. Sci.* **10**(18), 1–19 (2020). <https://doi.org/10.3390/APP10186248>
10. Field, V., Jiang, S., Chen, P., Zhan, Y., Zhao, C.: Applied sciences theoretical and computational analysis on the melt flow behavior of polylactic acid in material extrusion additive manufacturing under (2020)
11. Zhou, W., Kamata, K., Dong, M., Nomura, N.: Laser powder bed fusion additive manufacturing, microstructure evolution, and mechanical performance of carbon nanotube-decorated titanium alloy powders. *Powder Technol.* **382**, 274–283 (2021). <https://doi.org/10.1016/j.powtec.2020.12.066>
12. *Physics Procedia* 5 (2010) 617–622 www.elsevier.com/locate/procedia. **5**, 617–622 (2010). <https://doi.org/10.1016/j.phpro.2010.08.089>
13. Dana, H.R., Barbe, F., Delbreilh, L., Ben Azzouna, M., Guillet, A., Breteau, T.: Polymer additive manufacturing of ABS structure : Influence of printing direction on mechanical properties. **44**(May), 288–298 (2019). <https://doi.org/10.1016/j.jmapro.2019.06.015>
14. Wang, S., Lasn, K., Elverum, C.W., Wan, D., Echtermeyer, A.: Novel in-situ residual strain measurements in additive manufacturing specimens by using the optical backscatter reflectometry. *Addit. Manuf.* **32**(Jan), p. 101040 (2020). <https://doi.org/10.1016/j.addma.2020.101040>
15. Petru, J., Kritikos, M.: Porosity analysis of additive manufactured parts using CAQ technology (2021)
16. Hernandez-contreras, A., Ruiz-huerta, L., Caballero-ruiz, A.: Extended CT void analysis in FDM additive manufacturing components
17. Yu, H., Peng, C., Zhao, Z., Bai, L., Han, J.: Visual texture-based 3-D roughness measurement for additive manufacturing surfaces. *IEEE Access* **7**, 186646–186656 (2019). <https://doi.org/10.1109/ACCESS.2019.2919682>
18. Wu, H., Yu, Z., Wang, Y.: Experimental study of the process failure diagnosis in additive manufacturing based on acoustic emission. *Measurement* **136**, 445–453 (2019). <https://doi.org/10.1016/j.measurement.2018.12.067>
19. Hossain, S., Taheri, H.: In Situ process monitoring for additive manufacturing through acoustic techniques. *J. Mater. Eng. Perform.* **29**(10), 6249–6262 (2020). <https://doi.org/10.1007/s11665-020-05125-w>
20. Technique, A.E.: Fracture behavior of permeable asphalt mixtures with steel slag under low temperature based on (2020)
21. Stepanova, K.A., Kinzhagulov, I.Y., Yakovlev, Y.O., Kovalevich, A.S.: Applying laser-ultrasonic and acoustic-emission methods to nondestructive testing at different stages of deformation formation in friction stir welding. **56**(3), pp. 191–200 (2020). <https://doi.org/10.1134/S1061830920030122>
22. Lain, E.: ScienceDirect instrumented tests on composite pressure vessels (type IV) under internal water pressure, **6** (2020). <https://doi.org/10.1016/j.ijhydene.2020.09.160>
23. A. Dejangs, O. Kurtov, and P. Van Rymenant, “Acoustic emission as a tool for prediction of nugget diameter in resistance spot welding. *J. Manuf. Process.* **62**(Dec 2020), pp. 7–17 (2021). <https://doi.org/10.1016/j.jmapro.2020.12.002>
24. Hu, Y., Li, S., Li, D., Vadim, S.: Yield criteria and strength conditions considering comprehensive mechanical and acoustic emission characteristics of tension—torsion composite deformation of nylon materials. *Compos. Struct.* **243**(March), p. 112278 (2020). <https://doi.org/10.1016/j.compstruct.2020.112278>
25. Alexakis, H., Liu, H., Dejong, M.J.: Damage identification of brick masonry under cyclic loading based on acoustic emissions. *Eng. Struct.* **221**(June), p. 110945 (2020). <https://doi.org/10.1016/j.engstruct.2020.110945>

26. Sun, S., Hu, X., Zhang, W.: Detection of tool breakage during milling process through acoustic emission, pp. 1409–1418 (2020)
27. Motahari-nezhad, M., Jafari, S.M.: Bearing remaining useful life prediction under starved lubricating condition using time domain acoustic emission signal processing. *Expert Syst. Appl.* **168**(March 2020), p. 114391 (2021). <https://doi.org/10.1016/j.eswa.2020.114391>
28. Lopes et al.: An efficient short-time Fourier transform algorithm for grinding wheel condition monitoring through acoustic emission, pp. 585–603 (2021)
29. Kuznetsov, D.M., Smirnov, A.N., Syroeshkin, A.V.: New ideas and hypotheses : acoustic emission on phase transformations. **78**(11), 2273–2281 (2008). <https://doi.org/10.1134/S1070363208110492>
30. Vorontsov, V.B., Pershin, V.K.: Experimental research of phase transitions in a melt of high-purity aluminum. *J. Cryst. Growth* **480**, 170–174 (2017). <https://doi.org/10.1016/j.jcrysgro.2016.10.067>
31. Schmidt, L., et al.: ScienceDirect ScienceDirect new methodology to analyze the functional and physical architecture of acoustic process monitoring in laser beam welding acoustic process monitoring laser beam welding design existing products for an assembly oriented product family identification. *Procedia CIRP* **94**, 763–768 (2020). <https://doi.org/10.1016/j.procir.2020.09.139>
32. Mu, F., Rath, G., Lucyshyn, T., Kukla, C., Burgsteiner, M., Holzer, C.: Presentation of a novel sensor based on acoustic emission in injection molding, pp. 4744–4749 (2013). <https://doi.org/10.1002/app.38083>
33. Li, F., Yu, Z., Yang, Z., Shen, X.: Real-time distortion monitoring during fused deposition modeling via acoustic emission. (2020). <https://doi.org/10.1177/1475921719849700>
34. Nam, J., Jo, N., Kim, J.S., Lee, S.W.: Development of a health monitoring and diagnosis framework for fused deposition modeling process based on a machine learning algorithm. **234**, 324–332 (2020). <https://doi.org/10.1177/0954405419855224>
35. Li, F., Yu, Z.: Failure characterization of PLA parts fabricated by fused deposition modeling using acoustic emission. 7(Sept 2019), pp. 1177–1182 (2020). <https://doi.org/10.1108/RPJ-09-2019-0247>
36. Casiez, N., Deschanel, S., Monnier, T., Lame, O.: Acoustic emission from the initiation of plastic deformation of polyethylenes during tensile tests. *Polymer (Guildf)* **55**(25), 6561–6568 (2014). <https://doi.org/10.1016/j.polymer.2014.09.044>
37. Dikshit, V., Prasanth, A., Ling, Y., Leong, S., Yee, W., Wei, J.: Crack monitoring and failure investigation on inkjet printed sandwich structures under quasi-static indentation test. *Mater. Des.* **137**, 140–151 (2018). <https://doi.org/10.1016/j.matdes.2017.10.014>
38. Yang, Z.: Filament breakage monitoring in fused deposition modeling using acoustic emission technique, pp. 1–16 (2018). <https://doi.org/10.3390/s18030749>
39. Barile, C., Casavola, C., Pappalettera, G., Vimalathithan, P.K.: Mechanics of materials acoustic emission descriptors for the mechanical behavior of selective laser melted samples : an innovative approach. *Mech. Mater.* **148**(Dec 2019), pp. 103448 (2020). <https://doi.org/10.1016/j.mechmat.2020.103448>
40. Choi, N., Gu, J., Arakawa, K. Composites : part A acoustic emission characterization of the marginal disintegration of dental composite restoration. **42**, 604–611 (2011). <https://doi.org/10.1016/j.compositesa.2011.01.019>
41. Li, F., Yu, Z., Shen, X., Zhang, H.: Status recognition for fused deposition modeling manufactured parts based on acoustic emission. **01005** (2019)
42. Oosumi, K. et al.: 日本機械学会論文集 **81**(823), 1–12 (2015). <https://doi.org/10.1299/transjsme>
43. Wu, T., Tinkloh, S., Tröster, T., Zinn, W., Niendorf, T.: Measurement and analysis of residual stresses and warpage in fiber reinforced plastic and hybrid components (2021)
44. Wu, H., Wang, Y., Yu, Z.: In situ monitoring of FDM machine condition via acoustic emission. *Int. J. Adv. Manuf. Technol.* 1483–1495 (2016). <https://doi.org/10.1007/s00170-015-7809-4>
45. Wu, H., Yu, Z., Wang, Y.: Real-time FDM machine condition monitoring and diagnosis based on acoustic emission and hidden semi-Markov model, pp. 2027–2036 (2017). <https://doi.org/10.1007/s00170-016-9548-6>

46. Liu, J., Hu, Y., Wu, B., Wang, Y.: An improved fault diagnosis approach for FDM process with acoustic emission **35**(Sept), 570–579 (2018). <https://doi.org/10.1016/j.jmapro.2018.08.038>
47. Tr, W.: Acoustic emission for determining early age concrete damage as an important indicator of concrete quality/condition before loading (2020)
48. Niknam, S.A., Li, D., Das, G.: An acoustic emission study of anisotropy in additively manufactured Ti-6Al-4V, pp. 1731–1740 (2019)
49. Shevchik, S.A., Kenel, C., Leinenbach, C., Wasmer, K.: Acoustic emission for in situ quality monitoring in additive manufacturing using spectral convolutional neural networks. *Addit. Manuf.* **21**, 598–604 (2018). <https://doi.org/10.1016/j.addma.2017.11.012>
50. Eschner, N., Weiser, L., Häfner, B., Lanza, G.: Classification of specimen density in laser powder bed fusion (L-PBF) using in-process structure-borne acoustic process emissions. *Addit. Manuf.* **34**(Oct 2019), 101324 (2020). <https://doi.org/10.1016/j.addma.2020.101324>
51. Ito, K., Kusano, M., Demura, M., Watanabe, M.: Detection and location of microdefects during selective laser melting by wireless acoustic emission measurement. *Addit. Manuf.* **40**(Sept 2020), 101915 (2021). <https://doi.org/10.1016/j.addma.2021.101915>

Modelling and Vibration Signal Analysis for Condition Monitoring of Industrial Robots



Huanqing Han, Dawei Shi, Lichang Gu, Nasha Wei, and Fengshou Gu

Abstract Industrial robots are widely used in modern factories. Robot faults and abnormal working state will lead to the shutdown of the production line inevitably. Robot condition monitoring can improve production capacity. However, due to the changes of robot in dynamic working state, this is a challenge. This paper presents a methodology of condition monitoring for industrial robots using vibration signals. The main purpose of this paper is to identify the occurrence of the fault and its different degrees. Experiments was performed on a 6-dof industrial robot (IR). Firstly, the Frequency Response Function of the IR was obtained by using the Experimental Modal Analysis method. And the characteristic frequency in each axis was found. Then, based on the Short-time Fourier Transform analysis method, the vibration data under normal conditions and different degrees of abnormal working conditions were analysed. In some characteristic frequency bands, the amplitude will increase with the increase of the binding force at the joint. Finally, this trend was further verified by the calculation of RMS value. The results show that the proposed frequency domain and model analysis method can monitor the operating condition of industrial robots.

Keywords Industrial robots · Condition monitoring · Frequency Response · Model analysis · Vibration analysis

H. Han (✉) · F. Gu

School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, China
e-mail: Huanqing.Han@hud.ac.uk

H. Han · D. Shi · L. Gu · F. Gu

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

N. Wei

Department of Mechanical Engineering, Taiyuan University of Technology, Shanxi 030024, China

1 Introduction

Industrial robots have become one of the most critical manufacturing equipment in modern factories, and have been widely used in many fields, such as automobile, electronic equipment and aerospace. In 2021, the number of online industrial robot has reached 630,000, with an average annual growth rate of 14% [1]. In order to improve work efficiency and product quality, more and more industrial robots are used in automatic and intelligent production lines. Therefore, once the industrial robot fails or some abnormal operating condition occurred, it may increase the downtime of the production process and lead to incalculable economic losses [2]. Therefore, aim to reduce the shutdown loss, the condition monitoring and fault diagnosis technology of industrial robots has been widely studied recently.

For the sake of monitoring the operating condition of IR, it is essential to understand its vibration mode. However, the Frequency Response Function (FRF) of a 6-DOF IR varies with arm parameters (or posture) [3]. Focus on determine the FRF of a 6-DOF IR, off-line methods (such as Experimental Modal Analysis, EMA) usually been selected. The impact hammer experiment is carried out on the robot structure, and the FRF is determined by the corresponding excitation force and the vibration response of the structure. Nguyen et al. [4] have used the data-driven method to model the 6-DOF IR, and carried out modal analysis combined with the application of industrial robot in milling. Wu et al. [5] also used EMA in their research that the low-order dynamic behavior prediction method of 5-DOF hybrid robot based on the minimum generalized coordinate set. In addition, mathematical modeling is also a common method. Wang et al. [6] proposed a novel method to simplify the dynamic model of industrial robot based on interval method. Ding et al. [7] proposed a complete system for mechanical parameter identification of mechanical manual based on artificial bee colony algorithm.

At present, there are many methods to monitor the operating condition of IR. Ferrar et al. [8] used visual monitoring method, but the accuracy was insufficient. Hecke et al. [9] analyzed Acoustic Emission (AE) signals, but the cost of sensor is too high. Saixuan Chen et al. [10] used current signals for monitoring, but current signals need to use IR internal signal. Algburi et al. [11] used encoder signal, which also needs to use internal signal or external structure. The most commonly method is vibration analysis, which has low cost and convenient signal acquisition. Jaber et al. [12] used vibration signals for condition monitoring of industrial robots. While vibration signal is analyzed, the selection of analysis method is also very important. Because analysis method determines whether useful information can be obtained from the signal that full of noise. Sun et al. [13] used the traditional frequency domain analysis method to study the health status of IR gears. Jaber et al. [14] combined the wavelet and neural network algorithms to realize the black corrosion fault monitoring of internal mechanical structures. Yang et al. [15] studied the fault diagnosis method in dynamic process based on modeling and data-driven methods.

In this research, vibration signals were used for the condition monitoring of IR. And the feasibility verification has been carried out in the early study [16]. This

paper took a further discusses on experimental verification combined with the model analysis. The vibration characteristics of industrial robot were analyzed based on EMA method. And Short-time Fourier Transform (STFT) method was used on the vibration signal process and analysis. Finally, the signal processing results were comprehensively analyzed through the characteristics at the characteristic frequency and model. The main purpose of this research is to identify the occurrence of the fault and its different degrees.

The rest of this paper is organized as follows. Section 2 describes the theoretical backgrounds of IR and accelerometer. Section 3 presents the modelling of industrial robot using EMA and the behavior of FRF. Section 4 describes the experiment facilities of condition monitoring on IR. Section 5 presents the result and discussion about this research. And finally, Sect. 6 addresses the main conclusion and future work.

2 Background: The Industrial Robot and Accelerometer

2.1 The Industrial Robot

Industrial robots are composed of links and joints that, allow the robot to generate complex motions. [17] Every joint is mainly composed of motor, reducer and other mechanical links. The number of degree-of-freedom (DOF) is the same as the number of joints, which named J1, J2, J3, J4, J5, J6, separately. Typical industrial robots have 6 DOFs, which allows them to realize translational and rotational motion. The typical configuration of an industrial robot is shown in Fig. 1.

In order to enable the industrial robot to complete the specified task, IR needs the rotation of each joint and transfer the motion through the links. And finally synthesize the working path at the working point of the hand. The action of each joint requires the motor as the actuator and the reducer as the transmission system. So that each joint can produce the corresponding torque. The simplified model of single joint can be seen in Fig. 2.

Where, u is armature voltage, θ_1 is the motor shaft angle, θ_2 is the rotation angle of load shaft and n is the reduction ratio of reducer.

As seen in Fig. 2, the rotation angle of load shaft θ_2 is mainly determined by the parameters of motor and reducer. If the mathematical model of each joint is known and the conversion conditions between joints are known, the overall mathematical model of industrial robot can be obtained. However, the mechanical parts of the robot are complex. And the mechanical parts may bend due to load. So, the joints may have small effects that are difficult to calculate, such as elasticity and mechanical friction. These small effects are generally ignored in the actual modeling. Meanwhile, if each joint is regarded as a separate part and the whole IR is regarded as a multiple-degrees-of-freedom (MDOF) system composed of multiple joints connected in series. Then the simple model of industrial robot can be described in the way of lumped-mass model.

Fig. 1 6-DOF industrial robot

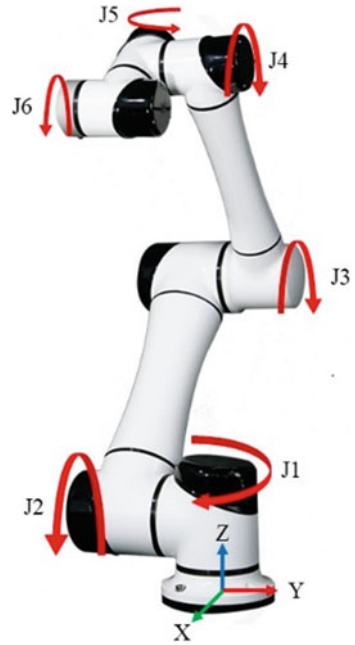
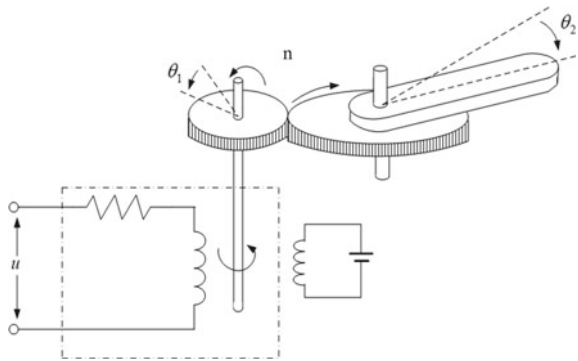


Fig. 2 Single joint motor-gear-load model



The free body diagram of six masses as shown in Fig. 3. Based on D'Alembert's principle, the equations of motion are written as

$$\begin{cases}
 m_1 \ddot{U}_1 + (c_1 + c_2) \dot{U}_1 - c_2 \dot{U}_2 + (k_1 + k_2)U_1 - k_2 U_2 = 0 \\
 m_2 \ddot{U}_2 - c_2 \dot{U}_1 + (c_2 + c_3) \dot{U}_2 - c_3 \dot{U}_3 - k_2 U_1 + (k_2 + k_3)U_2 - k_3 U_3 = 0 \\
 m_3 \ddot{U}_3 - c_3 \dot{U}_2 + (c_3 + c_4) \dot{U}_3 - c_4 \dot{U}_4 - k_3 U_2 + (k_3 + k_4)U_3 - k_4 U_4 = 0 \\
 m_4 \ddot{U}_4 - c_4 \dot{U}_3 + (c_4 + c_5) \dot{U}_4 - c_5 \dot{U}_5 - k_4 U_3 + (k_4 + k_5)U_4 - k_5 U_5 = 0 \\
 m_5 \ddot{U}_5 - c_5 \dot{U}_4 + (c_5 + c_6) \dot{U}_5 - c_6 \dot{U}_6 - k_5 U_4 + (k_5 + k_6)U_5 - k_6 U_6 = 0 \\
 m_6 \ddot{U}_6 - c_6 \dot{U}_5 + c_6 \dot{U}_6 - k_6 U_5 + k_6 U_6 = 0
 \end{cases} \quad (1)$$

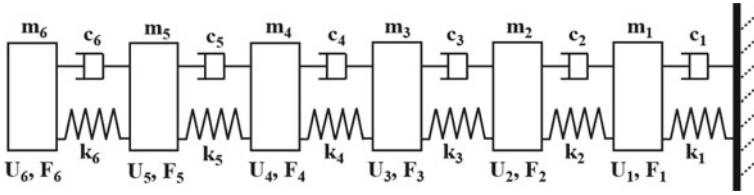


Fig. 3 The lumped-mass model of a 6-DOF industrial robot

Or writing in simply matrix form we get

$$[m]\{\ddot{U}\} + [c]\{\dot{U}\} + [k]\{U\} = \{0\} \tag{2}$$

In Eq. (2), $[m]$ is mass matrix which includes lumped mass and consistent mass. $[c]$ is damping matrix which often been ignored or obtained experimentally. $[k]$ is stiffness matrix which will be formed by assembling all the elements together to form a structure. The size of $[m]$, $[c]$ and $[k]$ are same as the DOF of IR. Therefore, the equation of motion for an undamped vibration of the IR system can be written as

$$[m]\{\ddot{U}\} + [k]\{U\} = \{0\} \tag{3}$$

And then FRF can be calculated based on differential equation.

2.2 Three-Axis Accelerometer

In order to get the vibration data during the operation of industrial robot, a wireless acquisition system named On Rotor Sensing (ORS) was produced. The role of ORS system is to measure the acceleration in three axes and transmit the data to the mobile phone through Bluetooth, including power module, acceleration sensor module, Bluetooth module and processor module. Its structure is shown in Fig. 4,

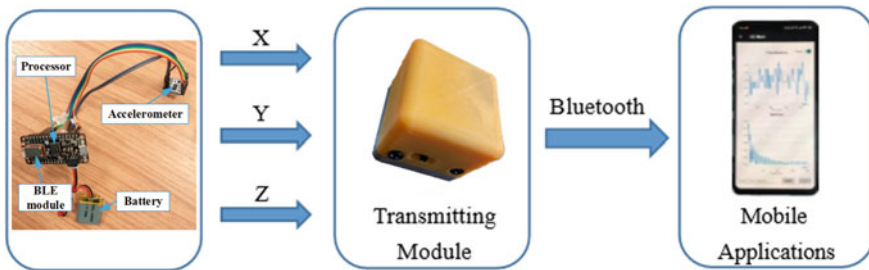


Fig. 4 Three-axis accelerometer based on ADXL345

Table 1 Parameter details of the three-axis accelerometer

Module name	Parameter details
Power module	Rechargeable lithium battery
Accelerometer	ADXL345 (Frequency Range: 1600 Hz, dynamic range16g)
BLE module	BLE 5.0 wireless transmission
Processor	nRF52840 (Cortex-M4F based microcontroller)

and the parameters of each module are shown in Table 1. The function of the acceleration sensor module is to collect the vibration data generated by the machine in the working process. The signals collected by the sensor can be used as the data for evaluating the health status of the machine, such as IR.

3 EMA Experiments and FRF Behaviour

The vibration response of the robot was measured by a three-axis accelerometer mounted at the hand of IR. Experiment was performed on the 6-dof industrial robot as Fig. 1 shown. To excite the robot structure, an impact hammer was used to apply an impulse excitation [18]. Then the FRF of robot in X, Y and Z directions can be calculated, with the assume that the cross coupling of robot vibration response can be ignored [19]. During the experiment, the three-axis acceleration sensor was installed on the end flange (hand) of the IR. The motion path of the hand is shown in Fig. 5, that reciprocating at point 1 and point 2. During the reciprocating, three times impacts were applied to J3 and J4 joints respectively to calculate the FRF when knocking the two joints. The magnitudes of the FRFs calculated from modal impact hammer tests performed in the X, Y, and Z directions are shown in Fig. 6.

The results show that the FRF in three directions shows nonlinearity. In addition, there are several vibration characteristic frequencies in each direction, that 25, 105 and 170 Hz in the X direction, 37, 67, 170 and 225 Hz in the Y direction and 37, 67, 170 and 225 Hz in the Z direction. And the amplitude in the Y direction is the largest under this motion, which is more suitable for analyzing. Therefore, in the subsequent vibration analysis, this paper will take the data in the Y direction as the analysis object.

4 Experiment Facilities of Condition Monitoring on IR

The experiment has been performed on the 6-dof industrial robot as Fig. 1 shown. Accelerometer was installed on the hand of industrial robot, see Fig. 7. In addition, the acceleration sensor is installed on the base as a group of comparative tests, which

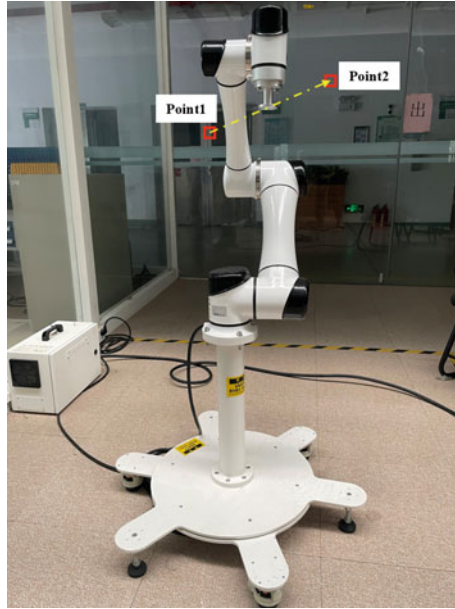


Fig. 5 Robot work path during experiments

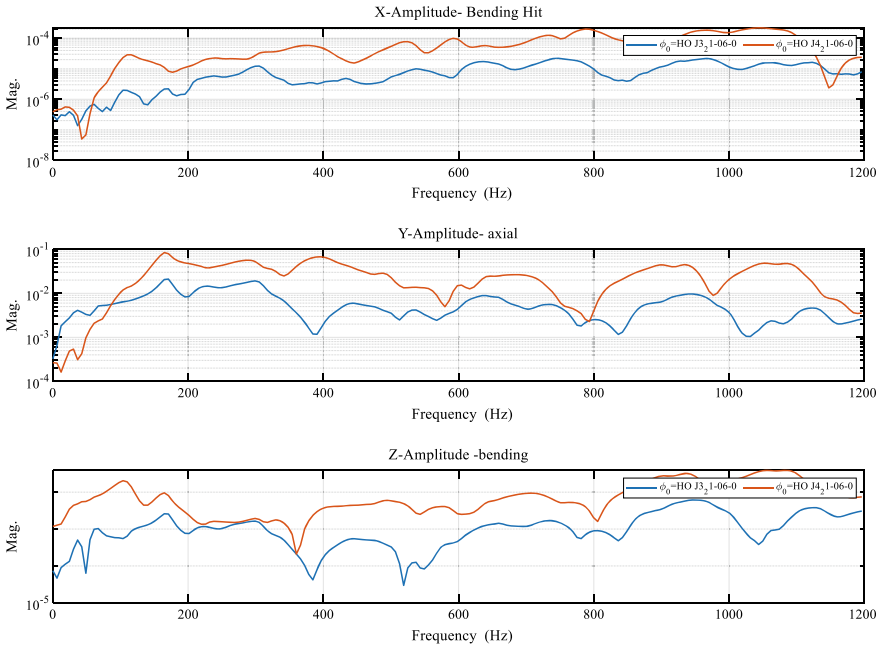


Fig. 6 Robot FRF magnitudes determined from EMA experiment along the X, Y and Z direction

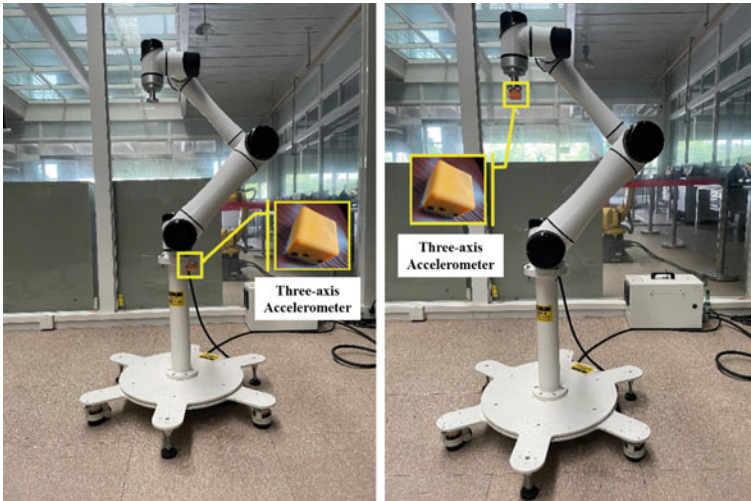


Fig. 7 Sensor installation position

can observe whether the obtained laws are the same. The abnormal working condition simulated in this paper is different degrees of joint binding force. The different tightness of stainless-steel clasp (see Fig. 8) is used to simulate the different degrees

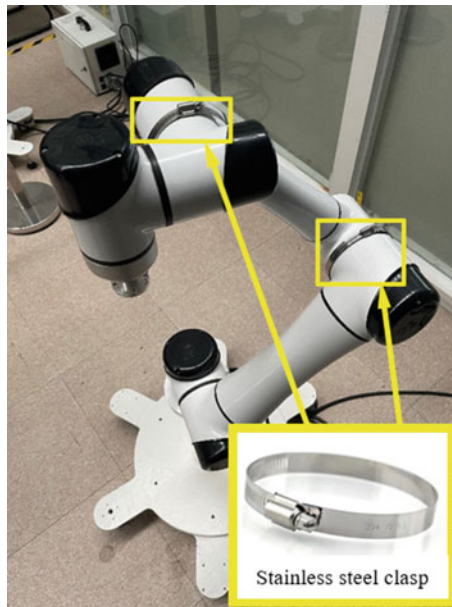


Fig. 8 Fault position

of binding force. Three different degrees of binding force were set, namely: “Normal”—without stainless steel clasp, “Force”—moderate binding force and “Strength Force”—large binding force. The purpose is to monitor the occurrence and degree of abnormal working conditions or faults. In addition, in order to eliminate the contingency caused by different fault locations. Two groups of comparative tests were carried out in J3 and J4 respectively. During the experiment, the motion path of the hand is the same as the EMA experiments that reciprocating at point1 and point2 (see Fig. 5).

In summary, 4 groups of experimental data were obtained in the experiments. When the sensor is installed on the hand, 2 groups data were obtained, that one on J3 and one on J4 under abnormal working conditions. And each group of data contains vibration data of three different degrees of binding force, which is the same when installed on the base.

5 Results and Discussion

After obtaining time-domain signals, the vibration signals can be processed based on the signal processing method. The signal processing method used in this paper is STFT. The time-domain vibration signals of different binding forces are shown in Fig. 9 which as an example when the sensor installed on the hand and the fault simulation in J3 joint. When the IR moves under the simple path planned in this

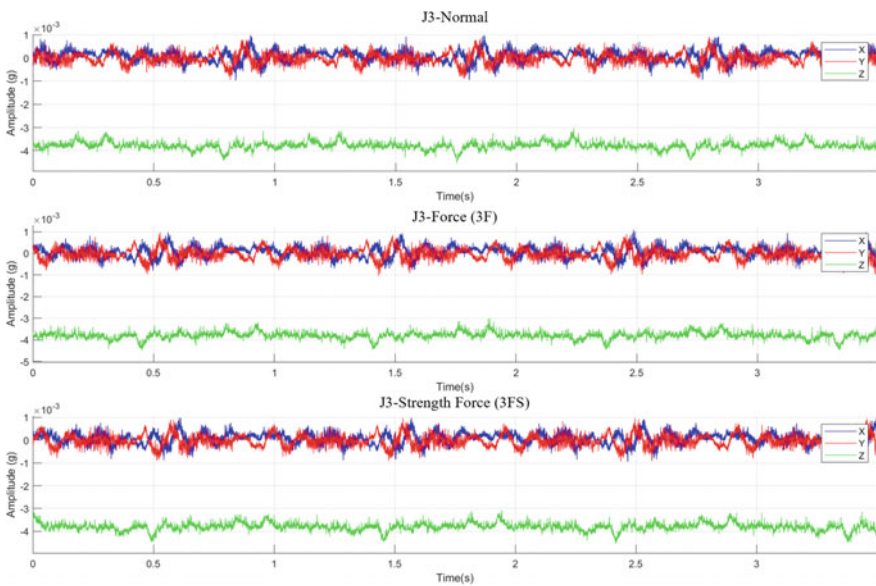


Fig. 9 Time domain vibration signal (Sensor on hand and Faults on J3)

paper, the amplitude will change in all three axes. As mentioned in Sect. 3, onected for processing in the analysis to explore the effectiveness of vibration signal in industrial robot condition monitoring.

After STFT processing of four groups time domain signals, the results shown in Fig. 10 were obtained. It can be seen that the frequency domain data shows different characteristics under different degrees of binding force. As shown in Fig. 10a, b, when the sensor was installed on the hand, the data under different binding forces will show an obvious peak near 37, 175 and 228 Hz, which is consistent with the characteristic frequency obtained by the model analysis in Sect. 3. And different amplitudes will be generated at these characteristic frequencies when IR working under different degrees of binding forces. As can be seen from Fig. 10c, d, when the sensor was installed on the base, it also has a similar law. It can be seen from the contour figure that the amplitude variation trend at these characteristic frequencies increases with the increase of the binding force at the joint.

In addition, in order to compare the calculations and set up trends that could be analyzed with the goal to find some pattern. The total root mean square (RMS) value of amplitude data was calculated and plotted in Fig. 11. It also presents an upward trend with the increase of binding force. Therefore, it is feasible to monitor the occurrence and degree of abnormal working conditions of IR using vibration analysis.

In the kinematic model of the industrial robot, damping, like the binding force in this research, is also an item that cannot be ignored. Although damping matrix often been ignored, the accuracy of model will increase if damping matrix can be determined. The reason for the amplitude rise needs to be analyzed in combination with the single joint model (see Fig. 2). On the premise that the load shaft outputs the same angle, the greater the mechanical damping at the joint, in order to overcome this resistance, the motor must output greater torque, which will further lead to greater joint vibration.

This study only demonstrates the feasibility of industrial robot condition monitoring through simple modeling and experimental methods. But the modeling method is still simple and the factors considered are not comprehensive, resulting in the lack of model accuracy. In addition, because the industrial robot is a complex nonlinear system, it is difficult to analyze the whole IR. So, the next step of the fault diagnosis method will start from key parts of IR, such as the reducer in the joint.

6 Conclusion

In this study, a monitoring method was proposed to identify the occurrence of the fault and its different degrees of IR based on vibration signal and analyzed combined with the model. The proposed method only uses a three-axis acceleration sensor to monitor the condition of the IR. In the proposed method, firstly, the mathematical modeling and EMA methods were used to establish the model of the 6-DOF IR and its FRF was obtained. The characteristic frequencies were determined base on

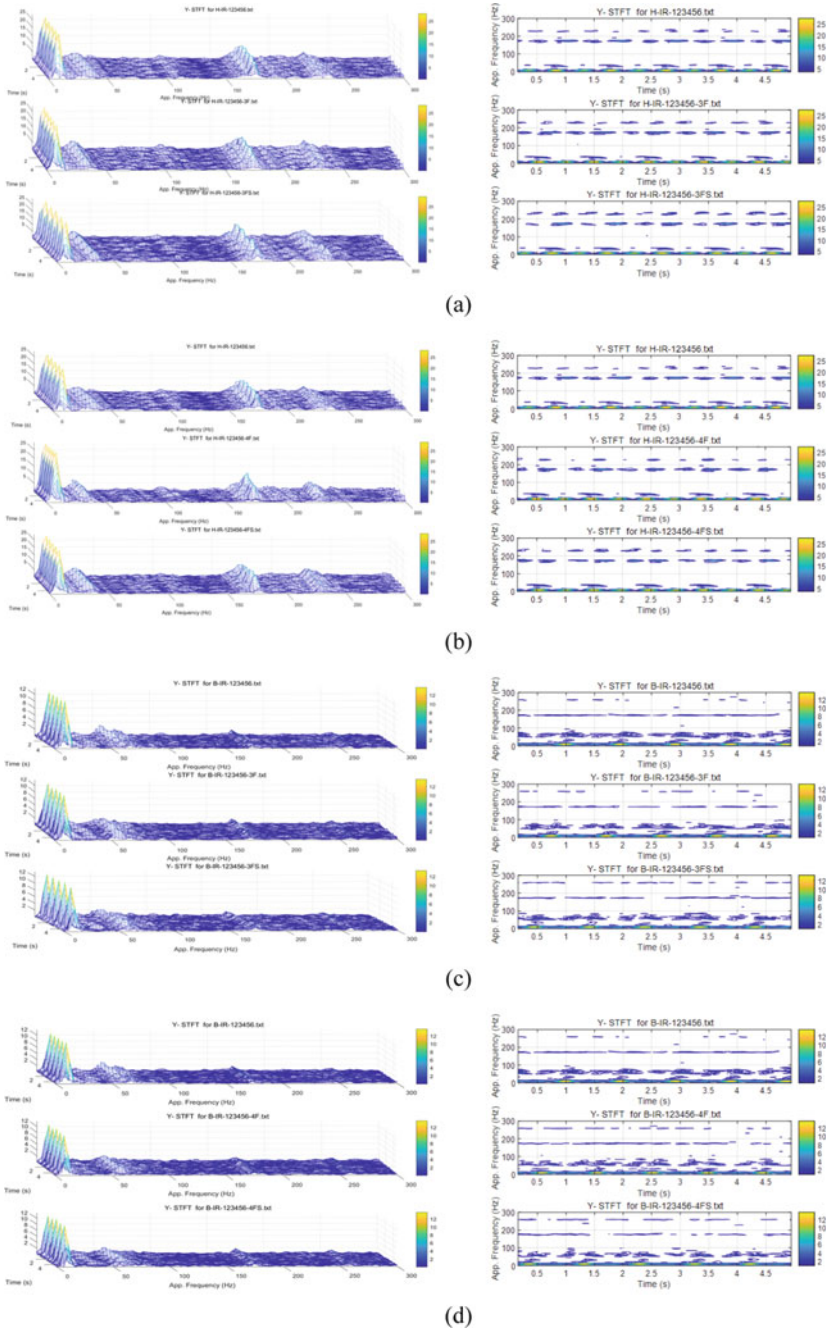


Fig. 10 The STFT result of different conditions (0 ~ 300 Hz)

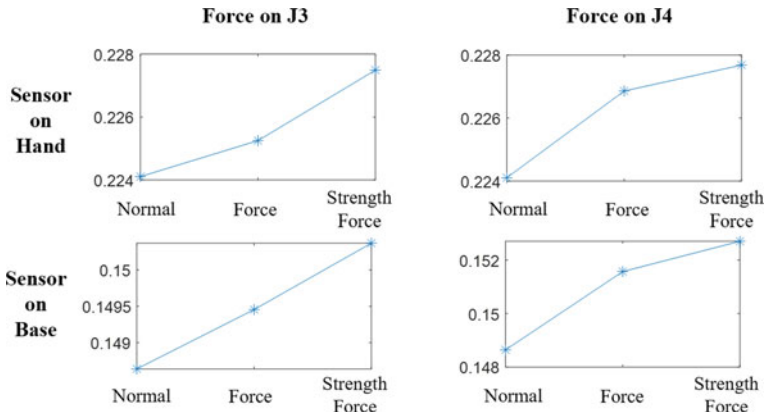


Fig. 11 RMS value under different conditions

FRF. Then, the three-axis acceleration sensor, named ORS, was used to collect the vibration data under different working conditions at the same joint. In order to eliminate the influence of different sensor installation position and fault position, three groups of comparative experiments were set up. Then, the time-domain vibration data were processed by STFT method, and the time–frequency diagrams under different conditions were obtained. Through the analysis of the vibration characteristics at three characteristic frequencies (37, 175 and 228 Hz) corresponding to the FRF in the time–frequency diagram, it is concluded that the amplitude will increase with the increase of the binding force at the joint. The internal reason is that the damping matrix in the model changes, which further enhances the oscillation of the system. In order to further present the law at the data level, the RMS value was calculated. The results show that it also presents an upward trend under different situation with the increase of binding force. Therefore, the proposed method is feasible. And still needs to be further studied in modeling and analysis methods in the future.

References

1. Wu, Q., Liu, Y., Wu, C., et al.: An overview of current situations of robot industry development. In: ITM Web of Conferences, vol. 17, pp. 03019 (2018)
2. Wang, S., Zhang, X., Xu, W., et al.: Energy-efficient concurrent assessment of industrial robot operation based on association rules in manufacturing. In: Proceedings of the 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), pp. 1–6 (2018)
3. Mejri, S., Gagnol, V., Le, T., et al.: Dynamic characterization of machining robot and stability analysis. *Int. J. Adv. Manuf. Technol.* **82**(1–4), 351–359 (2016)
4. Nguyen, V., Cvitanic, T., Melkote, S.: Data-driven modeling of the modal properties of a 6-dof industrial robot and its application to robotic milling. *J. Manuf. Sci. Eng.* **141**(12), 1–24 (2019)
5. Wu, L., Dong, C., Wang, G., et al.: An approach to predict lower-order dynamic behaviors of a 5-DOF hybrid robot using a minimum set of generalized coordinates. *Rob. Comput.-Integr. Manuf.* **67**, 102024 (2021)

6. Wang, K., Leonard, F., Abba, G.: A novel approach for simplification of industrial robot dynamic model using interval method. In: 2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp. 1697–1703. IEEE, Besançon, France (2014)
7. Ding, L., Wu, H., Yao, Y., et al.: Dynamic model identification for 6-DOF industrial robots. *J. Rob.* **2015**, 471478 (2015)
8. Ferrara, A., Incremona, G., Sangiovanni, B.: Sliding mode fault diagnosis with vision in the loop for robot manipulators. *New Trends Robot Control* **270**, 81–105 (2020)
9. Hecke, B.H., Yoon, J., He, D.: Low speed bearing fault diagnosis using acoustic emission sensors. *Appl. Acoustics* **105**(Apr), 35–44 (2016)
10. Chen, S., Luo, M., He, F.: A universal algorithm for sensorless collision detection of robot actuator faults. *Adv. Mech. Eng.* **10**(1), 168781401774071 (2018)
11. Algburi, R.N.A., Gao, H.: Health assessment and fault detection system for an industrial robot using the rotary encoder signa. *Energies, MDPI, Open Access J.* **12**(14), 1–25 (2019)
12. Jaber, A., Bicker, R.: Fault diagnosis of industrial robot bearings based on discrete wavelet transform and artificial neural network. *Int. J. Prognostics Health Manage.* **7**(2), 13 (2016)
13. Sun, H., Zhang, J.: Health monitoring of strain wave gear on industrial robots. In: IEEE 8th Data Driven Control and Learning Systems Conference, pp. 1166–1170. IEEE, Dali, China (2019)
14. Jaber, A., Bicker, R.: Industrial robot backlash fault diagnosis based on discrete wavelet transform and artificial neural network. *J. Qual. Maintenance Eng.* **4**(1), 21–31 (2016)
15. Yang, Q., Li, X., Cai, H., et al.: Fault prognosis of industrial robots in dynamic working regimes: find degradation in variations. *Measurement* **173**, 108545 (2021)
16. Han, H., Lin, Y., Gu, L., et al.: Vibration analysis based condition monitoring for industrial robots. In: Zhen, D., et al. (eds.) *Proceedings of IncoME-V & CEPE Net-2020*. IncoME-V 2020. Mechanisms and Machine Science, vol. 105, pp. 186–195. Springer, Cham (2021)
17. Kim, Y., Park, J., Na, K., et al.: Phase-based time domain averaging (PTDA) for fault detection of a gearbox in an industrial robot using vibration signals. *Mech. Syst. Signal Process.* **138**(Apr), 106544 (2020)
18. Nguyen, V., Melkote, S.: Hybrid statistical modelling of the frequency response function of industrial robots. *Rob. Comput.-Integr. Manuf.* **70**, 102134 (2021)
19. Cen, L., Melkote, S.N.: Effect of robot dynamics on the machining forces in robotic milling. *Procedia Manuf.* **10**, 486–496 (2017)

Study on Fault Mode of Hybrid Electric Vehicle



Guibo Liao, Fanbiao Bao, and Baoshan Huang

Abstract Vehicle reliability, battery life, and increased costs due to increased system complexity will hinder the marketization of hybrid electric vehicles. Improving vehicle reliability is the basis for improving product safety and performance. The dissertation studies the failure modes and failure laws of hybrid electric vehicles, and uses hybrid electric buses of electric vehicle demonstration operation companies as test objects to conduct road assessment tests to verify the matching and optimization of the entire vehicle and improve its performance and reliability. Develop a fault monitoring and acquisition analysis system for electric vehicle battery systems and power switching systems, and establish a mathematical model of the failure mode to summarize the rules for the maintenance and use of electric vehicles and the operation of electric vehicles.

Keywords Hybrid electric vehicle · Failure mode · Parameter estimation · Law of distribution

1 Significance of the Study on Fault Mode and Fault Law of Electric Vehicles

Automobile reliability is an important index to measure vehicle safety quality and fatigue life. For automobile products, reliability is closely related to personal safety and economic benefits. A car is made up of many assemblies, parts and components. In particular, compared with the traditional fuel vehicles, electric vehicles add batteries, motors and other systems with high failure rate, and many electronic systems and components constitute the control system. If a part is damaged, a component fails, it may cause an accident, causing serious consequences [1]. Improving

G. Liao
Zhuhai City Polytechnic, Zhuhai, Guangdong Province, China

F. Bao (✉) · B. Huang
Beijing Institute of Technology, Zhuhai, Guangdong Province, China
e-mail: 83667425@qq.com

the reliability of automobiles is the basis of improving the safety and performance of products. The improvement of reliability requires the continuous accumulation of experience and technical achievements, which requires long-term reliability research and test, summarizing experience and finding out the deficiencies, so as to improve and optimize the products.

2 Research Methods and Technical Routes Adopted

2.1 Assessment and Evaluation of Automobile Reliability

The measurement of reliability is mainly based on mathematical statistics. The measurement used in this study to evaluate the reliability of complete electric vehicles is as follows (note: since the operating mileage L is the basic data unit of the research object, the definition here is defined as mileage L : In reliability studies, the fault distribution function $F(L)$ is usually taken as the main research object. Since the size of $F(L)$ directly reflects the probability of fault occurrence, the cumulative fault in mileage L , and the functional relationship between fault and mileage L , it is also called $F(L)$ as cumulative fault probability. From the definitions of reliability and unreliability, we can know that $R(L)$ and $F(L)$ are the probability of whether the car will fail, and the value range is:

$$0 < R(L) \leq 1$$

$$0 < F(L) \leq 1$$

$$R(L) + F(L) = 1$$

The relationship between probability density function and fault distribution function $F(L)$ is as follows:

If $F(L)$ is continuously differentiable, then, Fault probability density $f(L)$:

$$f(L) = \frac{dF(L)}{dL} \quad (1)$$

Mean fault interval mileage $L(m)$ (MTBF): Life characteristics of MTBF products: rated life $L(0.9)$, median life $L(0.5)$, characteristic life $L(0.368)$.

2.2 Fault Statistics and Pattern Identification

From the perspective of automotive reliability engineering, the primary goal is to eliminate or effectively reduce the hazard of the fault, followed by vigorously reduce the failure frequency [2]. Generally speaking, vehicle state, fault cause and result are three elements of fault analysis. The main steps of fault analysis are as follows:

1. Investigation.
2. Mode identification.
3. Of fault mechanism.
4. Measures.

2.3 Fault Mode Analysis Method

1. Mode influence and hazard analysis (FMEACA).
2. Tree analysis (FTA).
3. Analysis graph.
4. System analysis.

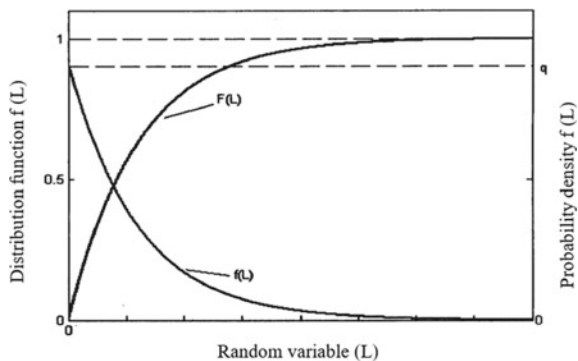
2.4 Commonly Used Vehicle Fault Theory Distribution

As can be seen from the bathtub curve shown in Fig. 1, when the product is in the accidental failure period, the exponential distribution is observed [3], and the failure rate is a constant, that is:

$$\lambda = \lambda(L) \tag{2}$$

1. *distribution*

Fig. 1 Distribution function and probability density function curve of exponential distribution



If the normal distribution of random variable L is $N(\mu, \sigma^2)$, then its probability density function $f(L)$ is:

$$f(L) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(L-\mu)^2}{2\sigma^2}} \tag{3}$$

where μ is the Average Value, σ is the Standard Deviation. Distribution function $F(L)$ is:

$$F(L) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^L e^{-\frac{(l-\mu)^2}{2\sigma^2}} dl \tag{4}$$

If the fault mileage of the product is a random variable with normal distribution, its characteristic values are as follows:

$$L_m = \mu, L_{0.5} = \mu$$

In order to facilitate the calculation of distribution function $F(L)$, it is generally necessary to convert normal distribution $N(\mu, \sigma^2)$ to standard normal distribution $N(0,1)$ in practical engineering application.

Set $L \sim z = \frac{L-\mu}{\sigma}$, set:

$$z = \frac{L - \mu}{\sigma} \tag{5}$$

Then, the probability density function $f(Z)$ of $Z \sim N(0,1)$ is:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{6}$$

The distribution function $\Phi(z)$ of z is:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} \tag{7}$$

If we know the z value, we can get $\phi(z)$ by looking it up in the standard normal distribution function table.

$$F(L) = \Phi\left(\frac{L - \mu}{\sigma}\right) \tag{8}$$

2. *Lognormal distribution*

The probability density function $f(L)$ and distribution function $F(L)$ of lognormal distribution are as follows:

$$f(L) = \frac{1}{L\sigma\sqrt{2\pi}} e^{-\frac{(lnL-\mu)^2}{2\sigma^2}} \tag{9}$$

$$F(L) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^L \frac{1}{l} e^{-\frac{(lnL-\mu)^2}{2\sigma^2}} \tag{10}$$

3. Weibull distribution

$$f(L) = \frac{m}{t_0} \left(\frac{L-r}{t_0}\right)^{m-1} e^{-\left(\frac{L-r}{t_0}\right)^m} \tag{11}$$

$$f(L) = 1 - e^{-\left(\frac{L-r}{t_0}\right)^m} \tag{12}$$

$$\lambda(L) = \frac{f(L)}{R(L)} = \frac{m}{t_0} \left(\frac{L-r}{t_0}\right)^{m-1} \tag{13}$$

Three-parameter Weibull distribution (m, t₀, r).

(m—Shape parameter t₀—Scale parameter r—Positional arguments).

When m < 1, it reflects the process characteristics of the early failure period of the product—as time goes by, the failure decreases [4].

When m = 1, the failure rate is constant, which reflects the process characteristic of accidental failure period of the product—failure rate. Under this condition, Weibull distribution is approximately an exponential distribution.

When m >1, it reflects the product’s wear out fault phase process characteristic—failure over time Increase sharply;

Where, when m = 3~4, Weibull distribution is approximately normal distribution, and in actual engineering, the shape parameters encountered are generally within the range of 0.5–5.0.

3 Fault Collection and Analysis of Hybrid Electric Vehicle in Demonstration Operation

3.1 Operating Status and Configuration of Operating Vehicles

A total of 67 hybrid buses have been operated by the demonstration operation company. According to the mode of city bus, the hybrid electric bus demonstration runs in the city [5]. The 67 hybrid electric buses in the demonstration run all use the Dongfeng EQ 6110HEV, it has experienced four generations of functional sample vehicles, performance sample vehicles, industrial commodity sample vehicles and commercial vehicles.

3.2 The Acquisition Scheme of Fault Data

The fault management subsystem of hybrid electric vehicle collects and manages fault information from three categories: vehicle fault, motor fault and battery fault [6].

3.3 Failure Data Analysis of Operating Vehicles

3.3.1 Relevant Statistics of Fault Occurrence Mileage

The data of fault occurrence mileage listed in Table 1 were collected and counted, and the histogram of fault occurrence mileage and frequency it was shown in Fig. 2.

Table 1 Fault mileage record table

i	Li	i	Li	i	Li	i	Li	i	Li
1	67	22	1216	43	1776	64	2499	85	5798
2	380	23	1256	44	1789	65	2567	86	5902
3	397	24	1279	45	1867	66	2589	87	6600
4	430	25	1293	46	1897	67	2597	88	6821
5	450	26	1298	47	1912	68	2634	89	6934
6	740	27	1341	48	1925	69	2867	90	7035
7	751	28	1387	49	1945	70	2903	91	8132
8	933	29	1394	50	1948	71	3031	92	8356
9	938	30	1398	51	2098	72	3097	93	8860
10	1023	31	1423	52	2112	73	3687	94	8923
11	1055	32	1438	53	2301	74	3698	95	9102
12	1079	33	1458	54	2310	75	3778	96	9204
13	1089	34	1483	55	2329	76	3987	97	9531
14	1110	35	1490	56	2378	77	4012	98	9567
15	1121	36	1523	57	2397	78	4110	99	9725
16	1129	37	1596	58	2420	79	4206	100	9767
17	1133	38	1630	59	2436	80	4309	101	9857
18	1140	39	1678	60	2457	81	4850	102	9886
19	1155	40	1697	61	2433	82	4899	103	9914
20	1187	41	1705	62	2451	83	4970	104	9943
21	1208	42	1721	63	2486	84	5035	105	9985

Note i represent the serial number of the fault; Li represents the mileage km at the time of failure

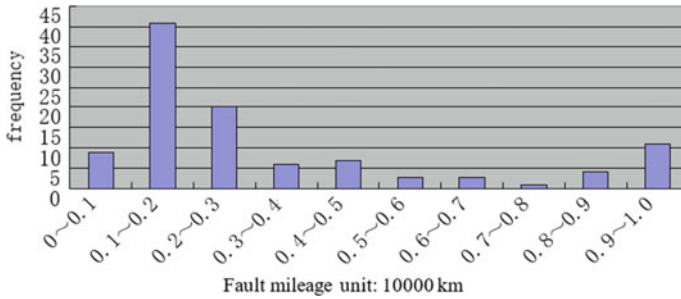


Fig. 2 Diagram of frequency of fault occurrence

Vehicle characteristic life (i.e. characteristic trouble-free working mileage) is in the range of 3000–4000 km, and the median life (i.e. median trouble-free working mileage) is in the range of 1000–4000 km.

3.3.2 Parameter Estimation of Theoretical Distribution

The probability density function of Weibull distribution $\omega(m, t_0) \sim f(L)$ 、Distribution function $F(L)$ and the failure rate function $\lambda(L)$:

$$f(L) = \frac{m}{t_0} \left(\frac{L}{t_0}\right)^{m-1} e^{-\left(\frac{L}{t_0}\right)^m} \tag{14}$$

$$FL = 1 - e^{-\left(\frac{L}{t_0}\right)^m} \tag{15}$$

$$\lambda(L) = \frac{m}{t_0} \left(\frac{L}{t_0}\right)^{m-1} \tag{16}$$

1. Linear regression

The relation between reliability $R(L)$ and unreliability $F(L)$:

$$R(L) + F(L) = 1 \tag{17}$$

It is concluded that:

$$R(L) = e^{-\left(\frac{L}{t_0}\right)^m} \tag{18}$$

Formula 3–5 is linearized, and the logarithms of both sides are taken twice in succession to obtain:

$$\ln \ln \frac{1}{R(L)} = m \ln L - m \ln t_0 \tag{19}$$

$$\text{set } y = \ln \ln \frac{1}{R(L)}, x = \ln L, A = -m \ln t_0$$

Linear function:

$$y = mx + A \tag{20}$$

By linear regression of the original function 3–5, the parameter estimation of the original function is transformed into the estimation of the parameters m and A of the linear function 3–8.

2. Parameter estimation

Through the sequence x_i, y_i , the least square method is adopted to carry out linear fitting of the function relation 3–7, that is, the parameter values m and a can be obtained.

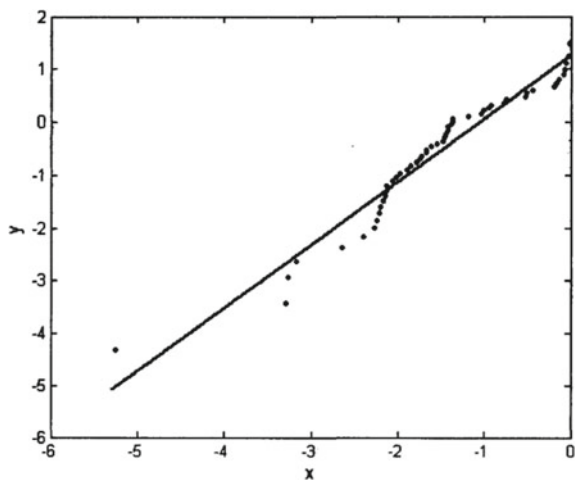
In reliability function engineering, the median rank is the approximate value of the distribution function $F(L)$ of unreliability, and the approximate formula of the median rank is:

$$H_i = \frac{i - 0.3}{n + 0.4} \quad (i = 1, 2, 3 \dots n = 105) \tag{21}$$

From formula 3–6 and L values of each fault mileage point in Table 1 x_i and y_i can be obtained, where $n = 1, 2, 3 \dots 105$. That is, 105 sequences x_i, y_i are obtained, and linear fitting is performed by MATLAB software. The fitting curve is shown in Fig. 3.

The fitting parameters: $m = 1.257, A = 1.362$, plug in formula 6–7 and you get $t_0 = 0.392$.

Fig. 3 Fitting curve



Where, when $m = 3\sim 4$, Weibull distribution is approximately normal distribution, and in actual engineering, the shape parameters encountered are generally within the range of 0.5–5.0.

3.3.3 Hypothesis Testing of Theoretical Distributions

In order to evaluate the degree to which the data points conform to the fitting line, goodness of fit γ^2 is defined. The expression is as follows:

$$\gamma^2 = \frac{m^2(\sum_{i=1}^n x_i^2 - nx^{-2})}{(\sum_{i=1}^n y_i^2 - ny^{-2})} \leq 1 \tag{22}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \tag{23}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \tag{24}$$

where $i = 1, 2, 3 \dots n$, n is equal to 105, x and y have the same meaning, $t_{n-2, 1-\frac{\alpha}{2}}$ is used to determine the critical goodness of fit γ_c^2 , and the inequality is verified:

$$\gamma^2 > \gamma_c^2 \tag{25}$$

To determine whether there is A linear relationship between y and x , $y = mx + A$. In practical applications, critical goodness of fit is generally taken.

It is calculated that:

$$\gamma^2 = 0.9 \quad \gamma_c^2 = 0.95 > 0.9$$

If the test requirements are met, the linear relationship between y and x can be determined to be reliable, that is, the fault range data studied can be determined to be from the whole subject to Weibull distribution.

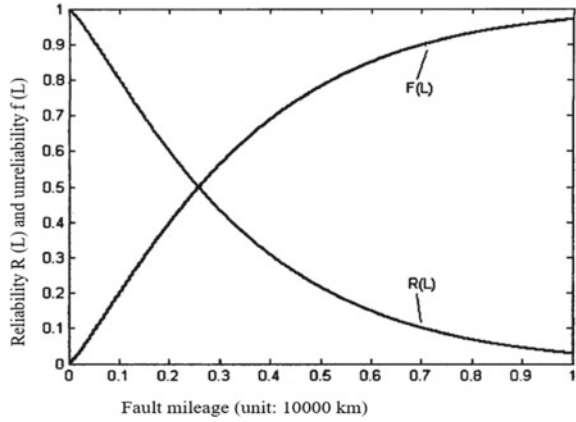
3.3.4 The General Trend of Electric Vehicle Malfunctions

Fault Distribution Trend Analysis

The distribution parameters are known, and the corresponding reliability function $R(L)$ and fault distribution function $F(L)$ can be obtained.

$$RL = e^{-\left(\frac{L}{0.392}\right)^{1.275}} \tag{26}$$

Fig. 4 Fault distribution function and reliability function diagram



$$F L = 1 - e^{-\left(\frac{L}{0.392}\right)^{1.275}} \tag{27}$$

Failure Probability Density Trend Analysis

Fault probability density function $f(L)$ is:

$$f(L) = \frac{1.257}{0.392} \left(\frac{L}{0.392}\right)^{0.257} e^{-\left(\frac{L}{0.392}\right)^{1.275}} \tag{28}$$

Failure Rate Trend Analysis

Failure rate function $\lambda(L)$ is:

$$\lambda(L) = \frac{1.257}{0.392} \left(\frac{L}{0.392}\right)^{0.257} \tag{29}$$

Comprehensive Study on the Overall Trend of Fault Occurrence

According to Figs. 4, 5 and 6 and the trend analysis of each function curve, it is consistent with the observation conclusion of automobile operation fault data [7].

Fault Reliability Characteristic Quantity Analysis and Evaluation

See Table 2.

Fig. 5 Failure probability density function diagram

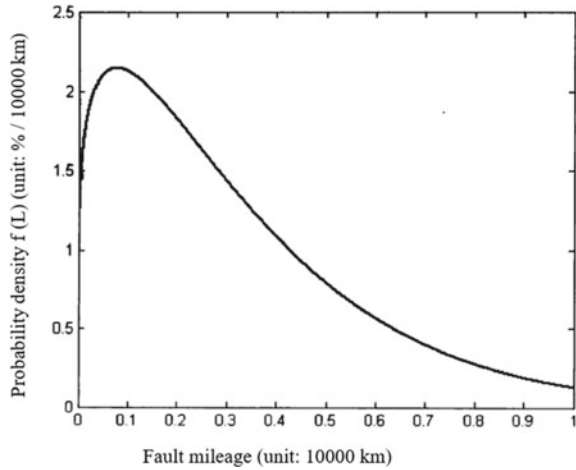


Fig. 6 Failure rate function diagram

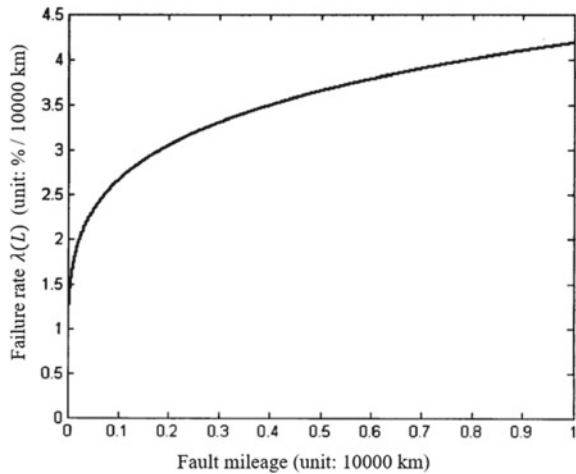


Table 2 Estimate of reliability characteristics of hybrid electric vehicles

Feature amount	Estimated value (km)
Mean time between failures (MTBF)	3740
Characteristic life	3660
Median life	2760
Rated life	697

According to GB/T 19750-2005, the MBTF limit of hybrid electric city bus is 3000 km. Based on this limit, the MBTF (3740 km) of electric vehicle in demonstration operation basically meets the limit requirements and meets the reliability index requirements.

4 Conclusion

Hybrid electric vehicle demonstration operation management in the development stage, in order to make the industrialization and marketization of electric vehicles can be further, must improve the electric car reliability ordinary consumers in order to achieve the requirements of normal use, through the demonstration operation of hybrid electric vehicle research and analysis of fault rule, suggested that the next step of hybrid perfect improvement can focus on the following aspects:

1. The active anti-electromagnetic interference capability of each control system (master controller, battery management system, motor controller, etc.) and CAN bus. In the overall arrangement of vehicles, the installation position of electromagnetic interference sources such as motors and frequency converters should be carefully arranged, so that they do not cause electromagnetic interference to the control system and CAN bus.
2. The relevant protocols of each control system of the vehicle should have unified standards, and the algorithm procedures of the control strategy should be meticulous, precise and perfect. The instructions and operations of the control system should be more user-friendly through multiple debugging tests, so that the driver can have more buffer time in thinking and operation.
3. The dynamic matching research can not only meet the basic requirement of the dynamic performance of the vehicle, but also need to further carry out the vibration modal analysis of the vehicle. Vibration is an important reason that affects the life of battery, motor and other key components, and also causes some safety problems in operation [8].
4. The processing quality and assembly quality of conventional mechanical parts need to be further improved. New energy vehicles are improved on the basis of traditional vehicles. While carrying out technological innovation on energy storage devices, power devices and transmission devices, more attention should be paid to the process quality and assembly quality of traditional mechanical components.

References

1. Wu, G.: Research and development of intelligent hybrid fault diagnosis system for electronically controlled engine. Guangdong University of Technology (2016)
2. Zhang, X.: Research on fault diagnosis method of automatic transmission based on computer. Shanghai University of Engineering and Technology (2016)
3. Wang, L.: Design of vehicle instrument data update system based on CAN network. Chengdu University of Technology (2016)
4. Zhang, H.: Development of fault diagnosis system for pure electric vehicle based on vehicle controller. Hunan University (2016)
5. Suresh, J.S., Jongkun, L.A.: TPM-based architecture to secure VANET. *Indian J. Sci. Technol.* **8**(15), (2015)

6. Zhao, T., Chen, C., Wei, L., et al.: An anonymous payment system to protect the privacy of electric vehicles. In: 2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6. IEEE (2014)
7. Sumra, I.A., Hasbullah, H.B., Manan, J.A.: Using TPM to ensure security, trust and privacy (STP) in VANET. In: 2015 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW), pp. 1–6. IEEE (2015)
8. Jhou, J.S., Chen, S.H., Tsay, W.D., et al.: The implementation of OBD-II Vehicle diagnosis system integrated with cloud computation technology. In: 2013 Second International Conference on Robot, Vision and Signal Processing (RVSP), pp. 9–12. IEEE (2013)

A New Two-Dimensional Condition-Based Maintenance Model by Using Copulas



Hanyang Wang, Ming Luo, and Fengshou Gu

Abstract This paper introduces a new two-dimensional condition-based maintenance model for complex and repairable machining systems like computer numerical control machining tools. A joint distribution of condition and reliability indicators is constructed by using copula. The maintenance threshold is set on the cumulative hazard rate conditioning on intensity of work. A numeric example with assumed settings is provide to demonstrate the relationship between the maintenance threshold and expected cost rate. This is the first model jointly considering condition and reliability indicators in maintenance area; and being benefited by the features of copula, this model can be easily extended to model dependences among multiple indicators in practice.

Keywords Condition-based maintenance · Two-dimensional maintenance · Copula · Conditional hazard rate · Expected cost rate · Data-driven maintenance · CNC

1 Introduction

Maintenance of machines plays an important role in manufacturing operation, and maintenance cost, the cost incurred by the user or owner to keep the machines in good working condition, contributes to a considerable proportion of the total expense of manufacturing enterprise. In practice, there are two main categories of maintenance actions, preventive maintenance (PM) and corrective maintenance. Conventionally, PM is implemented in the form of system overhaul or component replacement based on a predetermined time schedule to prevent malfunction; while corrective maintenance is conducted on the failures of system to restore the system. PM is called

H. Wang · F. Gu

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

M. Luo (✉)

Huddersfield Business School, University of Huddersfield, Huddersfield HD1 3DH, UK
e-mail: m.luo@hud.ac.uk

time-based maintenance (TBM) in some literatures due to the nature of its schedule [1]. In recent years, Condition-based Maintenance (CBM) has received more attention from the researchers. CBM is a maintenance approach that emphasizes on combining data-driven reliability models with sensor data collected from monitored operating systems to develop strategies for condition monitoring and maintenance [2]. In recent years, CBM practice is facilitated by advanced condition monitoring technology, which can dynamically reflect the health status and deterioration process of machines.

Nowadays, computer numerical control (CNC) machine tools are the cornerstone of modern manufacturing industry and are widely used in various industries. CNC machine tools are generally composed of numerical control system, spindle, feed shaft, knife library and other components. Due to the complexity of CNC tools mechanism, processing conditions load change, spindles, tools and other key components wear leading to its performance deterioration, seriously affecting the quality of parts processing. A well-designed CBM strategy of CNC tools can ensure product quality and improve the competitiveness of enterprises.

The advanced monitoring technologies, such as computer-based or IoT-based (internet of things) sensors, facilitate continuous monitoring, which can constantly monitor the system condition and trigger an alarm when a maintenance is required. In terms of condition monitoring of CNC tools, the mainstream methods are indirect monitoring, which is conducted through collecting the vibration, current, power, cutting force, torque, temperature and other data in the processing process, and establishing the mapping relationship between signal and tool state [3–11]. Huang et al. [12] proposed a tool wear prediction method based on the fusion of deep convolutional neural network multi-domain characteristics. Xie [13] analyzed the intelligent monitoring algorithm of tool wear and presented the architecture of digital twin system for monitoring of tool conditions based on cloud computing. According to a large number of cutting force signals obtained through cutting experiments, the researcher found characteristic vectors in cutting force signals could be reflected in cutting force signals by multi-resolution wavelet analysis, and the state recognition of tool wear was realized by neural network technology [14].

Currently, CBM research heavily relies on the stochastic degradation models. The degradation, which can be viewed as damage to a system, accumulates over time and eventually leads to a product failure when the accumulated damage reaches a failure threshold [15]. The data collected through continuous monitoring can reflect the health conditions of CNC tools, because the failures of most machine system are found attributed to some underlying degradation mechanism, such as wear of a mechanical component, resistance of an electronic component, capacity of a battery. To prevent the failure occurring or to extend the residual useful life of the machine, maintenance actions should be implemented when the degradation metric reaches a maintenance threshold before failure.

The degradation models are well developed in literatures. A well-known example is the Wiener degradation process. By this model, the first-passage-time or first-hitting-time, i.e. the time of the degradation first reaching a fixed threshold, follows an inverse Gaussian distribution. This property can assist us in scheduling maintenance effectively. Stochastic processes like Wiener process are dominating in this

research area, because they perform well in modelling the randomness in degradation processes caused by inherent randomness and environmental factors, even though complexity makes them not really handy to use for engineers [16]. Considering the richness of operating data collected through monitoring the machines like CNC tools, a data-driven CBM strategy model should be construct with incorporating multiple condition indicators of CNC.

Regarding the reliability modeling of CNC, one of the key reliability concepts is the time between failures (TBF) which is normally assumed independently and identically distributed (i.i.d.). The common types of random distribution used to model TBF include Weibull [17], Log-normal [18], Exponential [19], Gamma [20], and Inverse Gaussian [21] distributions. However, in practice, the i.i.d. assumption of TBF is not always held. To model the reliability of CNC machine tools, Wang et al. (2011) [22] proposed a three-parameter Weibull mixture model with a case study.

As a complex and repairable system, multiple failures could occur during the life time of CNC. The failure arrival process can be modelled by Poisson process [23], non-homogeneous Poisson process [24, 25], renewal process [26], etc. Guofa Li et al. 2019 [27] proposed a reliability evaluation method based on mixture variable parameter power law model (MVPPLM) with considered the influence of working condition difference in the traditional reliability modeling of numerical control (NC) machine tools. In metal cutting (i.e., machining) processes, tool wear was shown to be a main source of quality variation [28]. Li Hao et al. (2017) [29] propose an interaction model that utilizes a linear model to represent the impact of tool wear on quality degradation and a stochastic differential equation model to capture the impact of quality degradation on the instantaneous rate of tool wear.

By reviewing the literatures in CBM, CNC monitoring, degradation research, and reliability modeling areas, we can see that developing optimal CBM policies for complex CNCs is desirable and valuable in both academia and industry; however, CNCs are monitored on multiple dimensions in practice, and one or a few variables may not represent the condition of machine appropriately. Meanwhile, the majority of CBM studies are focusing on single random degradation process, which models inherent and environment-driven randomness well, but complexity makes them not really handy to use, especially when various factors are considered. In this paper, a copula-based method of constructing two-dimensional CBM models is introduced, which allows further expansion to consider more factors, and the parameters could be estimated stepwisely in practice.

2 The Joint Distribution of Condition Indicator and Reliability Index

In CBM research and practice, the condition of a machine is monitored and measured based on various metrics, such as vibration, temperature, contaminants, wear, noise, etc. [16]. Some of them are non-cumulative which can reflect the condition at a

specific moment, e.g. noise; meanwhile, some other metrics can indicate the degradation increment and present the degradation process of component or system, e.g. wear.

From the engineering point of view, the degradation increment can be treated as a superposition of many small effects in operation. As a combination of these effects, the behaviour of degradation increment can be approximated by a probability distribution. If the increment is considered normally distributed, and the increments in non-overlapping time intervals are believed independent, this process can be modelled by a basic Wiener process [15]. A basic Wiener degradation process $\{X(t), t \geq 0\}$ is expressed as:

$$X(t) = \mu\Lambda(t) + \sigma B(\Lambda(t)) \tag{1}$$

where μ is the drift parameter indicating the rate of degradation, σ is the volatility parameter, $B(\cdot)$ is the standard Brownian Motion, and $\Lambda(\cdot)$ is a monotone increasing function. $X(t)$ is used to represent the transformed system degradation in some scenarios. For example, $X(t)$ could be the logarithm of the degradation when the degradation value is required non-negative, then the degradation follows a Geometric Brownian motion (GBM), see Fig. 1.

Various metrics are monitored while various stochastic processes are used to model them in research. However, considering the simplicity of model construction and usability of models in practice, it is necessary to develop a data-driven model which can easily incorporate or be compatible to multiple condition indicators. In

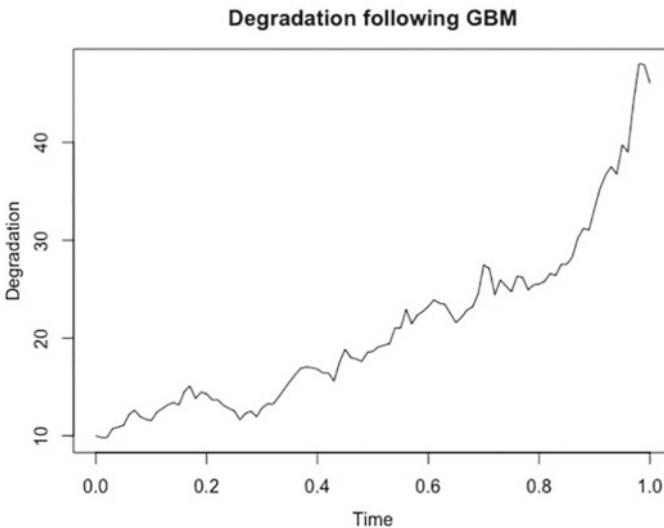


Fig. 1 Simulation of a GBM degradation

this research, the tool wear is selected as the only condition indicator to demonstrate the modelling strategy.

Tool wear in the machining process is a common phenomenon, and the deterioration of tool performance caused by wear is the main factor affecting product quality [1]. In operation, the tool wear is influenced by various factors, such as surface finish quality, ambient temperature, abrasive particles in the environment, lubrication characteristics, load, running distance or running time, material characteristics, etc. The wear should be treated as a degradation process with non-negative increments; in the related area, Gamma process is widely used to model this type of degradation [30]. However, in practice, when a machine fails or its system degradation reaches a given threshold, we can observe the values of multiple condition indicators and time; for instance, we can have a pair of wear level and time at the failure or threshold. After observing the machine for multiple operation circles, we can have a set of wear level values. The wear level could be treated as a continuous random variable and assumed following a Gamma distribution, $W \sim \Gamma(\alpha, \beta)$, where α is shape parameter, and β is rate parameter. Actually, $\Gamma(\alpha, \beta)$ is the distribution of the sum of α i.i.d. exponentially distributed variables. The probability density function (PDF) is

$$f_W(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} e^{-\beta w}, \tag{2}$$

and the cumulative distribution function (CDF) is

$$P(W \leq w) = F_W(w) = \int_0^w f_W(x) dx. \tag{3}$$

where, $\Gamma(\alpha)$ is the Gamma function. W can be named as wear to failure (WTF) in this research.

The reliability of CNC tool can be modelled by the reliability function, $F_T(t)$, which is the CDF of a non-negative random variable T and gives the probability that a machine fails or its degradation indicator hits a given threshold before or at time t . Then, the argument of $F_T(t)$ could be time to failure (TTF) or first passage time (FPT). By the similar means, we have the survival function

$$S(t) = P\{T > t\} = 1 - F_T(t) \tag{4}$$

which gives the probability that a machine will survive beyond time t , i.e. does not fail before or at time t . The survival density function is $s(t) = -f_T(t)$. Based on the reliability function and survival function, the hazard rate is defined as

$$\lambda_t(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t \cdot S(t)} = \frac{f_T(t)}{S(t)} \tag{5}$$

which gives the conditional rate that a machine has survived for t but will not survive for an additional time Δt . If T follows an exponential distribution, $\lambda_t(t) = \lambda$ is a constant and λ is the rate parameter of this exponential distribution.

The distribution of wear to failure, $F_W(w)$, and the distribution of TTF, $F_T(t)$ could be dependent, for example, using a machine heavily can lead to high wear level and early failure. Therefore, the joint distribution of W and T is constructed by incorporating a bivariate copula. The joint distribution of W and T is

$$F(t, w) = C(F_T(t), F_W(w)) = C(u, v) \tag{6}$$

where $C(\cdot)$ is a bivariate copula, $F_T(t) = u \in [0, 1]$, and $F_W(w) = v \in [0, 1]$. The density of this joint distribution is

$$f(t, w) = c(F_T(t), F_W(w)) \cdot f_T(t) \cdot F_W(w) = c(u, v) \cdot f_T(t) \cdot f_w(w) \tag{7}$$

where, $c(\cdot)$ is the copula density.

This joint distribution function can reflect reliability on $T - W$ plane. $F(t, w)$ gives the probability that the machine fails before wear reaching w and age reaching t . Then the two-dimensional survival function is

$$\begin{aligned} S(t, w) &= 1 - F(t, +\infty) - F(+\infty, w) + F(t, w) \\ &= 1 - C(u, 1) - C(1, v) + C(u, v) \\ &= 1 - u - v + C(u, v). \end{aligned} \tag{8}$$

The value given by $S(t, w)$ is the probability that the machine does not fail before age reaching t and wear reaching w . This two-dimensional survival function can be represented through survival copula as well:

$$S(t, w) = \tilde{C}(1 - u \cdot 1 - v), \tag{9}$$

where \tilde{C} is the survival copula of C .

In practice, the joint distribution can be easily estimated by estimating marginal distribution and copula separately. Several families of copulas are well explored by research communities, for example, the widely used Gaussian copula and Archimedean copulas. The various dependence structures between random variables can be approximately modelled by selecting proper copula according. Nested bivariate copulas and multivariate copulas can be used to construct multivariate joint distributions when needed.

3 Conditional Based Maintenance

According to the joint distribution introduced above, the conditional PDF of TTF given the occurrence of $W = w$ is

$$\begin{aligned}
 f_{T|W}(t|w) &= \frac{f(t, w)}{f_W(w)} \\
 &= \frac{c(u, v) \cdot f_T(t) \cdot f_W(w)}{f_W(w)} \\
 &= c(u, v) \cdot f_T(t).
 \end{aligned}
 \tag{10}$$

The conditional CDF of T given the occurrence of $W = w$ is

$$\begin{aligned}
 F_{T|W}(t|w) &= \int_0^t c(u, v) \cdot f_T(x) dx \\
 &= \frac{\partial C(u, v)}{\partial v}.
 \end{aligned}
 \tag{11}$$

The conditional hazard rate given the occurrence of $W = w$ is

$$\lambda_{T|W}(t) = \frac{f_{T|W}(t)}{S_{T|W}(t)} = \frac{c(u, v) \cdot f_T(t)}{1 - \frac{\partial C(u, v)}{\partial v}}
 \tag{12}$$

The conditional hazard function, i.e. conditional cumulative hazard rate is

$$\Lambda_{T|W}(t) = \int \lambda_{T|W}(t)
 \tag{13}$$

As we know $\lambda_{T|W}(t) = \frac{f_{T|W}(t)}{S_{T|W}(t)} = -\frac{s_{T|W}(t)}{S_{T|W}(t)}$, then

$$\Lambda_{T|W}(t) = \int \lambda_{T|W}(t) = -\int \frac{s_{T|W}(t)}{S_{T|W}(t)} = -\ln S_{T|W}(t).
 \tag{14}$$

If sufficient historical data were collected from the same machine or the machine in the same type under the same working condition, the conditional hazard rate could be a condition indicator of the machine. It gives the rate that the machine survived for t but will not survive for an additional infinitesimal time conditioning on the wear w . It may show, for example, working intensity can influence the remaining useful life of a machine. Therefore, the maintenance threshold can be set on $\lambda_{T|W}(t)$ as λ_M ; however, because $\lambda_{T|W}(t)$ is not essentially monotonic, the maintenance threshold can be set on $\Lambda_{T|W}(t)$ as Λ_M when needed. This means a maintenance should be conducted at time T_M , when $\lambda_{T|W}(T_M) = \lambda_M$ or $\Lambda_{T|W}(T_M) = \Lambda_M$.

According to Eqs. (6), (7), (12), and (14), when λ_M (or Λ_M) is given, the value of T_M is determined by a function of wear w , $g(w|\Lambda_M)$. Then the expected value of

T_M , i.e. the expected time to maintenance, is

$$E(T_M) = \int g(w|\lambda_M) f_W(w) dw. \tag{15}$$

Assume the maintenances are perfect, which means the machine will be restored as good as new after each maintenance. Then, the expected number of maintenances for time t is

$$N(t) = \frac{t}{E(T_M)} \tag{16}$$

The cost of each maintenance can be varying due to some random factors associated, but the expected cost of a maintenance, C_M , could be estimated by experts or according to empirical data. The expected total cost of maintenance for time t is

$$TC = C_M N(t) = \frac{C_M t}{E(T_M)}. \tag{17}$$

However, in practice, the total useful life of a machine is influenced by many factors, and some of them are beyond the machine itself, for example, a new generation of machine is introduced to the market. Then, the expected cost rate is more important than total maintenance cost from the perspective of management. The expected cost rate, C_r , is

$$C_r = \frac{TC}{t} = \frac{C_M}{E(T_M)}. \tag{18}$$

The optimal CBM policy can be determined by finding the optimal maintenance threshold which can minimize the expected cost rate. It means the following optimization problem needs to be solved:

$$\begin{aligned} \text{Min } C_r &= \frac{C_M}{E(T_M)} \\ \text{Subject to } E(T_M) &\geq T_O \\ 0 < \Lambda_M < \Lambda_O \end{aligned} \tag{19}$$

where T_O and Λ_O are set according to operating needs.

4 Numeric Example

To demonstrate the model, the Clayton copula is used in this research, but the actually form of copula in practice should be determined according to the data collected or experts' experience. The Clayton copula is:

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} \tag{20}$$

where $\theta \in [-1, 0) \cup (0, \infty)$ is the parameter. The probability density of the Clayton copula is:

$$c(u, v) = (1 + \theta)(uv)^{-(1+\theta)}(u^{-\theta} + v^{-\theta} - 1)^{-\frac{2\theta+1}{\theta}} \tag{21}$$

The survival copula associated to the Clayton copula is:

$$\tilde{C}(\tilde{u}, \tilde{v}) = \left(\tilde{u}^{1-\theta} + \tilde{v}^{1-\theta} - 1 \right)^{-\frac{1}{\theta-1}} \tag{22}$$

where, $\tilde{u} = 1 - u$ and $\tilde{v} = 1 - v$.

According to Eqs. (12), (20), and (21), we have

$$\begin{aligned} \lambda_{T|W}(t) &= \frac{c(u, v) \cdot f_T(t)}{1 - \frac{\partial C(u, v)}{\partial v}} \\ &= \frac{(1 + \theta)(uv)^{-(1+\theta)}(u^{-\theta} + v^{-\theta} - 1)^{-\frac{2\theta+1}{\theta}} \cdot u'}{1 - (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-1} v^{-\theta-1}}. \end{aligned} \tag{23}$$

where $u' = f_T(t)$. According to Eqs. (4), (11), (14) and (20), we have

$$\begin{aligned} \Lambda_{T|W}(t) &= -\ln S_{T|W}(t) \\ &= -\ln \left(1 - \frac{\partial C(u, v)}{\partial v} \right) \\ &= -\ln \left(1 - (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-1} v^{-\theta-1} \right). \end{aligned} \tag{24}$$

Obviously, Eq. (24) is much simpler than Eq. (23), to demonstrate the model, we set the maintenance threshold as $\Lambda_{T|W}(t) = \Lambda_M$. Then, we have

$$\begin{aligned} -\ln \left(1 - (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-1} v^{-\theta-1} \right) &= \Lambda_M \\ \rightarrow (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-1} &= (1 - e^{-\Lambda_M}) \cdot v^{\theta+1} \\ \rightarrow u^{-\theta} &= \left[(1 - e^{-\Lambda_M})^{-\frac{\theta}{1+\theta}} - 1 \right] \cdot v^{-\theta} + 1. \end{aligned} \tag{25}$$

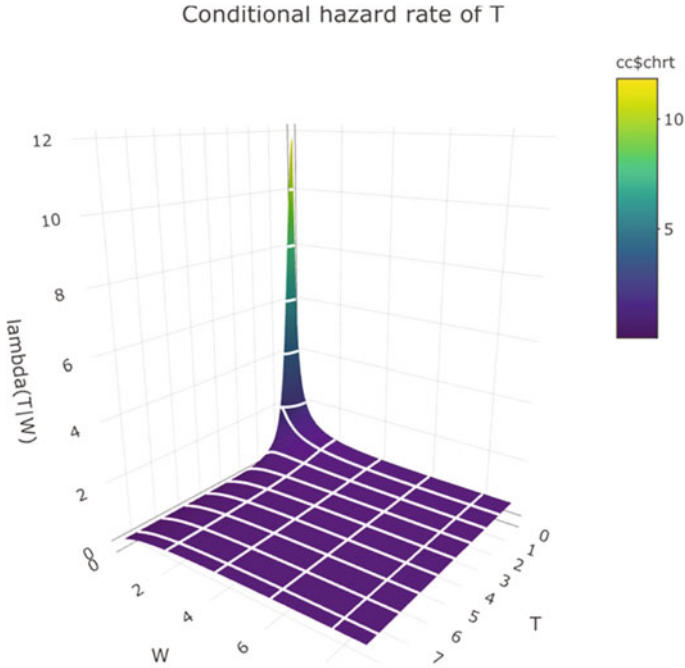


Fig. 2 Surface of conditional hazard rate $\lambda_{T|W}(t)$

Assume $W \sim \Gamma(\alpha = 3, \beta = 1)$, $T \sim \text{Exp}(\lambda = 0.5)$, and $\theta = 1$; the maintenance threshold is set as $\Lambda_M = 6$. According to Eq. (23), the surface of conditional hazard rate $\lambda_{T|W}(t)$ is plotted in Fig. 2.

On Fig. 2 we can see the conditional hazard rate decreases dramatically when wear and time grow. It means the proposed assumption of marginal distributions and copula might only reflect the early stage reliability of machine, as the Clayton copula has more probability concentrated in the lower tail ($u \rightarrow 0, v \rightarrow 0$), i.e. it has strong lower tail dependence. The early stage refers to the initial wear phase of tool, in which the tool wears quickly due to the roughness of the sharpened cutting-edge surface. It has a small actual contact area with the machining surface, allowing the stress on the cutting edge and the machining surface to concentrate, so a narrow surface is quickly ground on the rear tool face, allowing the contact pressure between the cutting edge and the machining surface to decrease and the tool wear rate to gradually decrease and stabilize until the end of the initial wear phase.

In practice, the marginal distributions and copula should be estimated based on real data which may require more complicate models to reflect the actual dependence structures. For instance, the Gumbel copula could be applied on the final stage of machine life, and a combined or stepwise copula could be used to model the whole lifecycle.

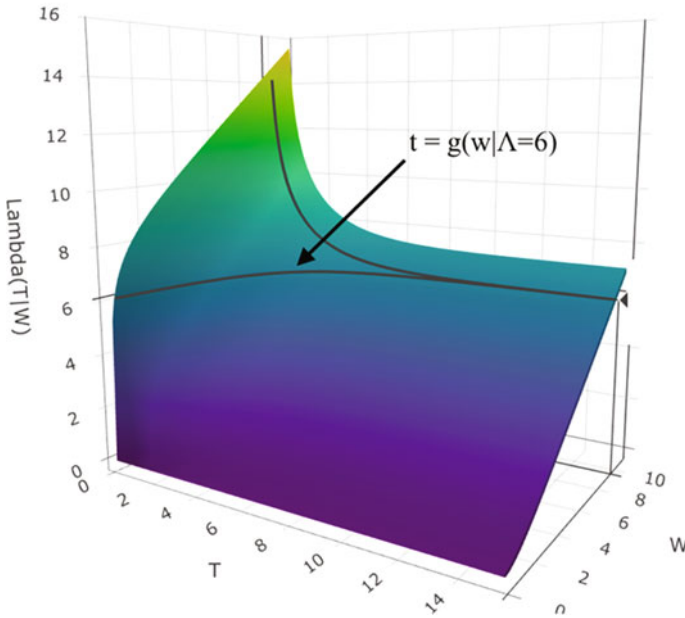


Fig. 3 The surface of conditional cumulative hazard rate and the relationship between T_M and W when $\Lambda_M = 6$

Then, according to Eqs. (24) and (25), the surface of conditional cumulative hazard rate $\Lambda_{T|W}(t)$ is shown by Fig. 3, and the relationship between T_M and W on the plane $\Lambda_M = 6$ is marked.

The function $T_M = g(w|\Lambda_M)$ is determined by Eq. (25), the curves of this function with different values of Λ_M are plotted in Fig. 4.

The expected value of T_M is calculated associated with the distribution of which is shown by Eq. (15), $E(T_M|\Lambda_M = 3) = 60.57826$, $E(T_M|\Lambda_M = 6) = 124.5343$, and $E(T_M|\Lambda_M = 9) = 189.9809$. We can see the expected value of T_M is increasing with the maintenance threshold Λ_M .

By the same means, the relationship between the expected cost rate $C_r = \frac{C_M}{E(T_M)}$ and the maintenance threshold Λ_M can be plotted after more simulations. Set the expected maintenance cost $C_M = 100$, the relationship between C_r and Λ_M is shown by Fig. 5.

According to Fig. 5, the function $C_r = h(\Lambda_M)$ is monotonically decreasing. Therefore, the optimization problem stated by Eq. (19) can be solved at $\Lambda_M = \Lambda_O$ if. This result means, if the assumptions could reflect the early stage of a machine, the user should conduct maintenance as late as they could or should not do maintenance in the early stage.

Fig. 4 The relationships between T_M and W when $\Lambda_M = 3, 6$ and 9

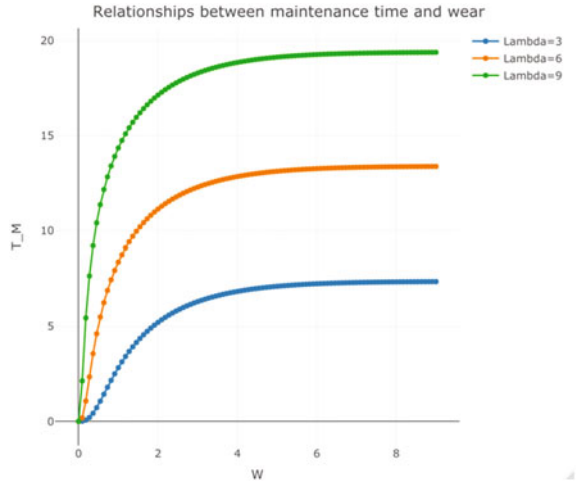
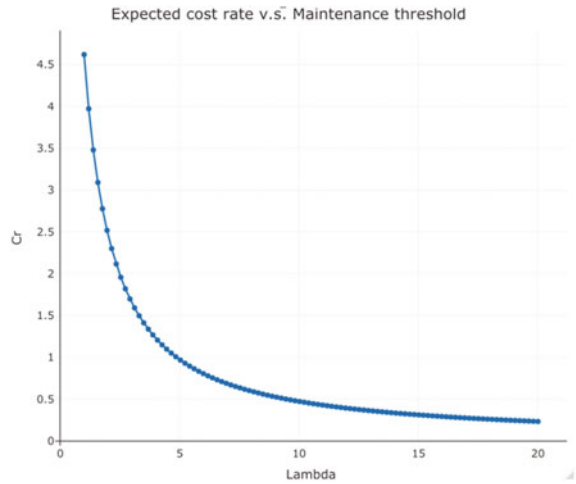


Fig. 5 Expected cost rate versus maintenance threshold



5 Conclusion

This paper has introduced an easy-to-use and extensible framework which allows to incorporate multiple monitored factors in condition-based maintenance scheduling; even though, the assumptions of this study are too simple to reflect the details in practice. However, a data-driven model, it is easy-to-use by fitting to the data collected in operation, and benefited by the properties of copula, the model could be extended to higher dimension by constructing multivariate copulas. In the future research, the model could be improved by further studies with considering the following, but

not limited to, issues: models of the different stages of machine operation, multidimensional copula selection and construction, time-varying dependence, the effects of uptime and downtime, the cost of monitoring and preventive maintenance, etc.

References

1. Halila, F., Czarnota, C., Nouari, M.: New stochastic wear law to predict the abrasive flank wear and tool life in machining process. *J. Eng. Tribol.* **228**(11), 1243–1251 (2014)
2. Alaswad, S., Xiang, Y.: A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliab. Eng. Syst. Saf.* **157**, 54–63 (2017)
3. Byrne, G., Dornfeld, D., Inasaki, I., Ketteler, G., König, W., Teti, R.: Tool condition monitoring (TCM)—the status of research and industrial application. *Ann CIRP* **44**(2), 541–567 (1995)
4. Chao, P., Hwang, Y.D.: An improved neural network model for the prediction of cutting tool life. *J. Intell. Manuf.* **8**, 107–115 (1997)
5. Jemielniak, K., Kwiatkowski, L., Wrzosek, P.: Diagnosis of tool wear based on cutting forces and acoustic emission measures as inputs to a neural network. *J. Intell. Manuf.* **9**, 447–455 (1998)
6. Rehorn, A.G., Jiang, J., Orban, P.E.: State-of-the-art methods and results in tool condition monitoring: a review. *Int. J. Adv. Manuf. Technol.* **26**, 693–710 (2005)
7. Scheffer, C., Heyns, P.S.: An industrial tool wear monitoring system for interrupted turning. *Mech. Syst. Signal Process.* **18**, 1219–1242 (2004)
8. Purushothaman, S.: Tool wear monitoring using artificial neural network based on extended Kalman filter weight updation with transformed input patterns. *J. Intell. Manuf.* (2009). <https://doi.org/10.1007/s10845-009-0249-y>(2009)
9. Sick, B.: On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. *Mechanical systems and signal processing*, **16**(4), 487–546 (2002)
10. Audy, J.: An appraisal of techniques and equipment for cutting force measurement. *J. Zhejiang Univ. - Sci. A7*, 1781–1789 (2006)
11. Wang, X., Wang, W., Huang, Y. et al.: Design of neural network-based estimator for tool wear modeling in hard turning. *J Intell Manuf* **19**, 383–396 (2008)
12. Huang, Z.W., Zhu, J.M., Lei, J.T., et al.: Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations. *J. Intell. Manuf.* **31**, 953–966 (2020)
13. Xie, N., Kou, R., Liu, X.: Research on digital twin system for monitoring of tool conditions based on cloud computing. *Mech. Manuf.* **59**(03), 78–82+92 (2021)
14. Xiang, S., Altintas, Y.: Modeling and compensation of volumetric errors for five-axis machine tools. *Int. J. Mach. Tools Manuf.* **11**, 65–78 (2016)
15. Ye, Z.-S., Xie, M.: Stochastic modelling and analysis of degradation for highly reliable products. *Appl. Stoch. Model. Bus. Ind.* **31**, 16–32 (2015)
16. Ahmad, R., Kamaruddin, S.: An overview of time-based and condition-based maintenance in industrial application. *Comput. Ind. Eng.* **63**, 135–149 (2012)
17. Yang, Z., He, J., Wang, J., et al.: Bayesian method to solve the early failure period of numerical control machine tool. *Proc. Inst. Mech. Eng. B J. Eng. Manuf.* (2016)
18. Jia, Y., Wang, M., Jia, Z.: Probability distribution of machining center failure. *Reliab. Eng. Syst. Saf.* **50**(1), 121–125 (1995)
19. Sadeghi, M., Roghanian, E.: Reliability optimization for non-repairable series-parallel systems with a choice of redundancy strategies: Erlang time-to-failure distribution. *Proc. Inst. Mech. Eng. Part O: J. Risk Reliab.* **231**(3), 1748006X1771761 (2017)
20. Wang, Y., Jia, Y., Qiu, J., et al.: Load spectra of CNC machine tools. *Qual. Reliab. Eng. Int.* **16**(3), 229–234 (2015)

21. Peng, W., Li, Y.F., Yang, Y.J., et al.: Bayesian degradation analysis with inverse Gaussian process models under time-varying degradation rates. *IEEE Trans. Reliab.* **99**, 1–13 (2017)
22. Wang, Z., Yang, J., Wang, G., Zhang, G.: Application of three-parameter Weibull mixture model for reliability assessment of NC machine tools: a case study. *Mech. Eng. Sci.* 2718–2726 (2011)
23. Noorossana, R., Sabri-Laghaie, K.: System reliability with multiple failure modes and time scales. *Qual. Reliab. Eng. Int.* **32**(3), 1109–1126 (2016)
24. Wang, S., Wu, Y., Lu, M., Li, H.: Discrete nonhomogeneous Poisson process software reliability growth models based on test coverage. *Qual. Reliab. Eng. Int.* **29**(1), 103–112 (2013)
25. Xu, J.Y., Yu, D., Xie, M., Hu, Q.P.: An approach for reliability demonstration test based on power-law growth model. *Qual. Reliab. Eng. Int.* **33**(8), 1719–1730 (2017)
26. Yang, Q., Hong, Y., Zhang, N., Li, J.: A copula-based trend-renewal process model for analysis of repairable systems with multitype failures. *IEEE Trans. Reliab.* **66**(3), 590–602 (2017)
27. Li, G., Zhu, H., He, J., Wu, K., Jia, Y.: Application of power law model in reliability evaluation of machine tools by considering working condition difference. *Qual. Reliab. Eng. Int.* **35**, 136–145 (2019)
28. Huang, Q., Shi, J., Yuan, J.: Part dimensional error and its propagation modeling in multi-operational machining processes. *Manuf. Sci. Eng.* **125**, 255–262 (2003)
29. Hao, L., Bian, L., Gebraeel, N., Shi, J.: Residual life prediction of multistage manufacturing processes with interaction between tool wear and product quality degradation. *IEEE Trans. Autom. Sci. Eng.* **14**(2), (2017)
30. van Noortwijk, J.M.: A survey of the application of gamma processes in maintenance. *Reliab. Eng. Syst. Saf.* **94**, 2–21 (2009)

Rotary Valve to Improve the Problem of Big End and Needle Glue Overflow in Contact Dispensing Process



Gaolian Huang, Shifei Zhang, Zhiguo Liu, Gaobo Xiao, Chucheng Chen, and Fanbiao Bao

Abstract Since the discovery of the piezoelectric effect, the application of piezoelectric technology has been increasing. The rotary valve has precise and controllable characteristics for the contact dispensing process, and has become an indispensable part of the dispensing industry. For a long time, the working mode of piezoelectric ceramic valve controller is relatively single. The traditional point mode, line mode and single channel can no longer adapt to the development of industrial automation. With the increasing complexity of dispensers, customers have put forward more and more customized requirements. In order to better adapt to the development of dispensers, reduce the difficulty of control, and shorten the development cycle, a more flexible and convenient controller system is needed.

Keywords Piezoelectric ceramic valve · Dispensing process · Air pressure control

1 Introduction

With the continuous update and enrichment of electronic products, people's requirements for products are developing in the direction of being smaller, lighter, thinner and more stable. At the same time, the application of dispensing technology in electronic products has become more and more extensive, such as mobile phones and cameras, and the process requirements are becoming more and more stringent.

The traditional double-sided adhesive bonding method or the simple dispensing bonding method is gradually replaced. The performance requirements of the product require the dispensing process to ensure good bonding performance during the assembly process, while ensuring other more process performances, such as waterproof, air tightness, drop test, electrical conductivity, thermal conductivity and

G. Huang · S. Zhang · Z. Liu · G. Xiao · C. Chen
Zhuhai Bojay Electronics Co., Ltd., Zhuhai, Guangdong Province, China

F. Bao (✉)
Beijing Institute of Technology, Zhuhai, Guangdong Province, China
e-mail: 83667425@qq.com

other requirements. The application of dispensing technology in electronic products is becoming more and more extensive, such as mobile phones, cameras, LEDs, smart wearable devices, new energy, SMT industries, etc. Process requirements are becoming more and more stringent, more refined and specific, such as glue height, glue width, glue amount, appearance requirements (big end problem of the start and end points, wire drawing, splash, scattered spots, glue overflow, bubbles) [1].

In order to solve the process problem of large end point and glue overflow after closing glue in contact dispensing, we designed this rotary dispensing valve. By using the rotary dispensing valve body in the contact dispensing process, the problem of large end-point glue volume during the dispensing process and glue overflow of the needle after closing the glue can be well solved. It is verified by testing that the improvement effect is good, the operation is simple and convenient during the use process, and it has better dispensing effect and stability for the application of medium and low viscosity fluid contact dispensing as well.

2 Problems with Traditional Dispensing Methods

For the contact dispensing process, the selected valve types are syringe air pressure, striker valve, and screw valve. In the daily test and use process, the usual problems encountered are the amount of glue at the end point and the problem of glue overflow after closing the glue.

2.1 Syringe Air Pressure

1. *Operating principle*

The syringe air pressure is a relatively simple way to dispense glue. The dispensing needle is directly connected to the glue outlet of the syringe. The dispensing controller controls the air pressure at the upper end of the syringe to dispense and adjust the amount of glue [2].

2. *Process problem*

During the dispensing process, due to the influence of the residual pressure of the air pressure, it is easy to cause the needle to overflow after the glue is turned off, as shown in Fig. 1.

2.2 Striker Valve

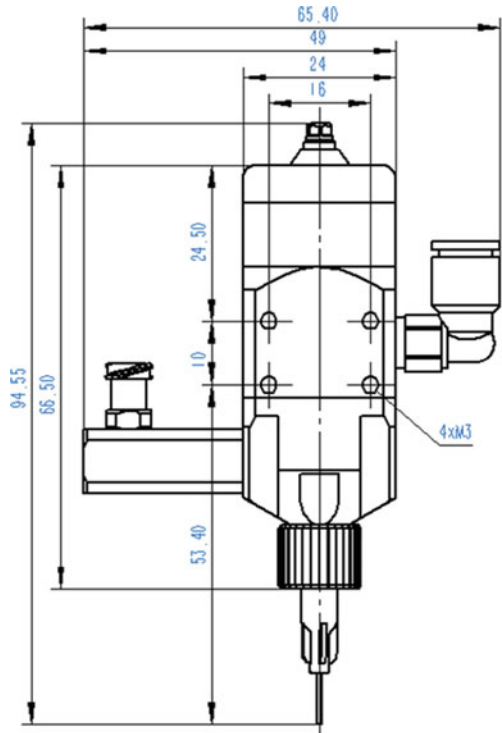
1. *Operating principle*

The pressure controller controls the lift of the valve body striker to switch on and off the glue. As shown in Fig. 2 [3].

Fig. 1 Syringe air pressure



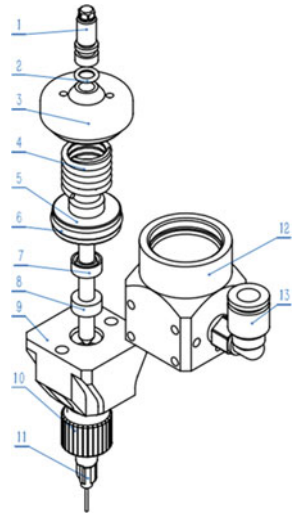
Fig. 2 Striker valve



2. *Process problem*

The strength of the striker is not easy to control during the lifting process, and it is easy to have too much glue at the end, as shown in Fig. 3.

Fig. 3 Exploded view of Striker valve



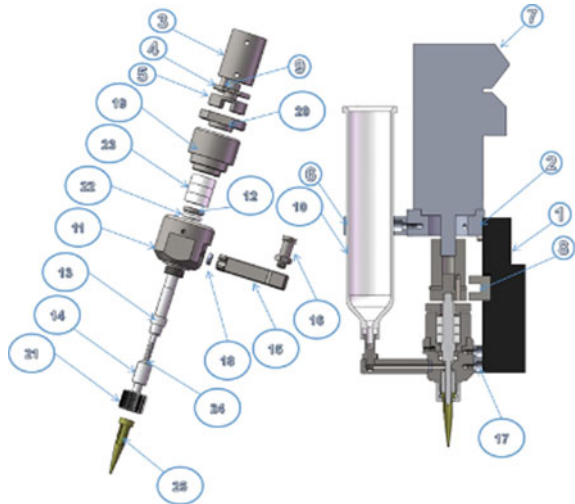
2.3 Screw Valve

1. *Operating principle*

The feed pressure is adjusted by the air pressure controller, the glue is supplied into the screw cavity, and the glue is squeezed out from the needle by the rotation of the rotor in the cavity, as shown in Fig. 4 [4].

2. *Process problem*

Fig. 4 Screw valve



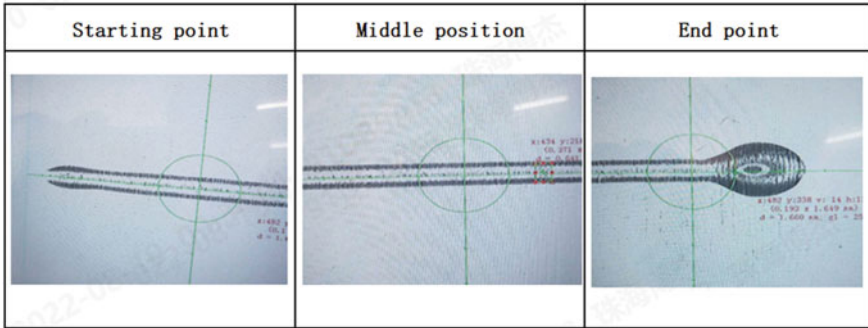
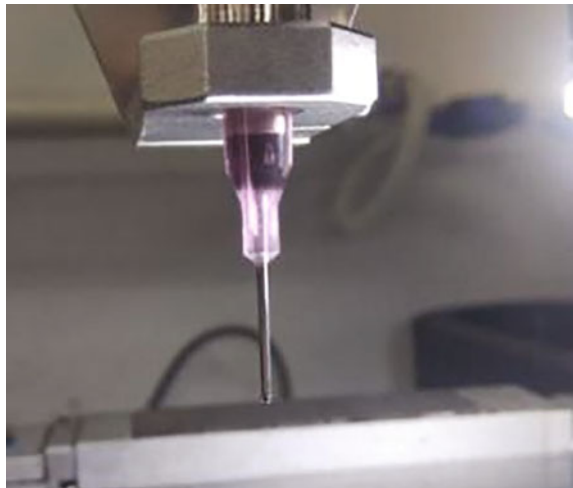


Fig. 5 Big head problem of the end point

Fig. 6 The problem of needle overflow after closing the glue



The rotor and the stator are not completely airtight. After the glue is turned off, under the influence of the residual pressure of the air pressure, there will still be glue hanging at the needle, and the response speed of the on-off valve is slow, as shown in Figs. 5 and 6 [5].

3 Introduction of Rotary Dispensing Valve

3.1 Rotary Valve

Figures 7, 8, 9 and Table 1 show the structure and parts required for the design of the rotary dispensing valve [6].

Fig. 7 Dimensions of the rotary valve

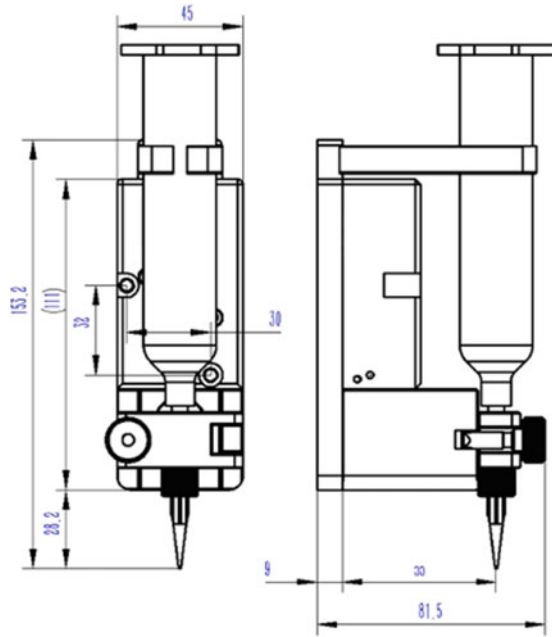
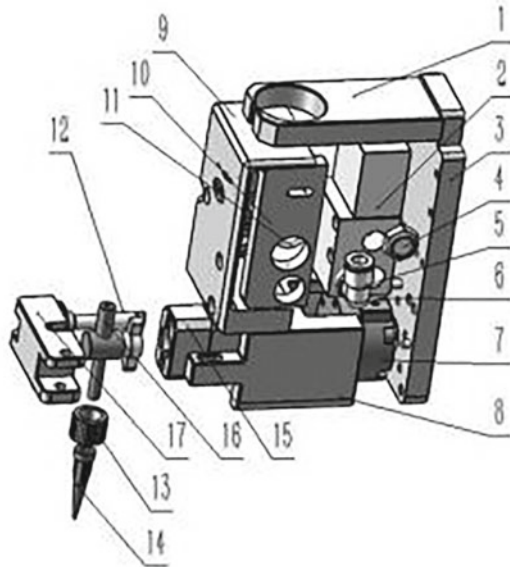


Fig. 8 Exploded view of rotary valve



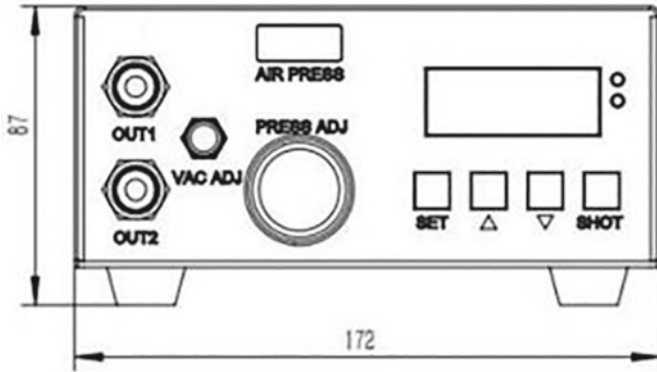


Fig. 9 Dimensions of the air pressure controller

Table 1 Parts required for rotary valve

1	Syringe holder	8	Rotary spool fixing fixture
2	Electromagnetic valve	9	Enclosure mounting
3	Backplane fixture	10	Side panel
4	Air outlet	11	Sealing ring
5	Air inlet	12	Locking components
6	O-ring	13	Needle fixing nut
7	Rotary valve	14	Needle

3.2 Operating Principle

The air pressure controller OUT1 is the normal air pressure and is connected to the air inlet of the solenoid valve. The solenoid valve signal line and the air pressure controller signal line are connected to the external machine (dispenser). The air pressure controller OUT2 is the on-off air pressure, which is connected to the adapter tube at the upper end of the glue cylinder to provide the glue supply air pressure when dispensing glue, and the air pressure is disconnected when the glue is not being dispensed to reduce the influence of residual pressure. When opening the glue, the dispensing controller OUT2 provides the supply air pressure, the external machine provides the solenoid valve 24 V signal, and the solenoid valve A work position is turned on to make the rotary valve rotate 90°. The cavity hole in the middle of the rotor in the rotary valve communicates with the glue cylinder and the needle, and the glue passes through the cavity hole of the rotor under the action of the feeding pressure, and the glue is dispensed through the needle. When closed, the air pressure controller OUT2 closes the supply air pressure, the external machine closes the solenoid valve 24 V signal, and the solenoid valve B work position is turned on to make the rotary valve rotor rotate back to 0°. At this time, the glue in the cavity hole is cut off, and

the air pressure controller OUT2 turns off the air pressure to complete the glue off action.

3.3 The Characteristics of Rotary Valve

1. With good airtightness, water is used as the medium, and the feed pressure is 0.5Mpa, no dripping.
2. Adopting direct-acting solenoid valve, the response speed of switch glue is fast, and there is no glue overflow after the glue is closed.
3. Adopt the air pressure controller to supply the material normally, no need to suck back, and the glue volume is stable.
4. Applicable glue has a wide range of viscosity, especially suitable for glues without filler and reacting with metal particles.
5. It can be used with tapered needle to achieve better process effect.

3.4 Process Parameters

Design the rotary valve through various process parameters, as shown in Table 2.

Table 2 Various process parameters

Products	Parameters	Unit
Minimum dispensing volume	0.0005	ml
Minimum dispensing diameter	200	μm
Minimum operating frequency	400	Times/minute
Maximum feeding pressure	0.6	Mpa
Max flux	1.8	L/min
Suitable for glue viscosity	5000~150,000	Cps
Drive material	Aluminum alloy	/
Accuracy deviation	±5%(W >0.005 ml)	/
Dimensions	45*81.5*153.2	Mm
Weight	1.8	KG

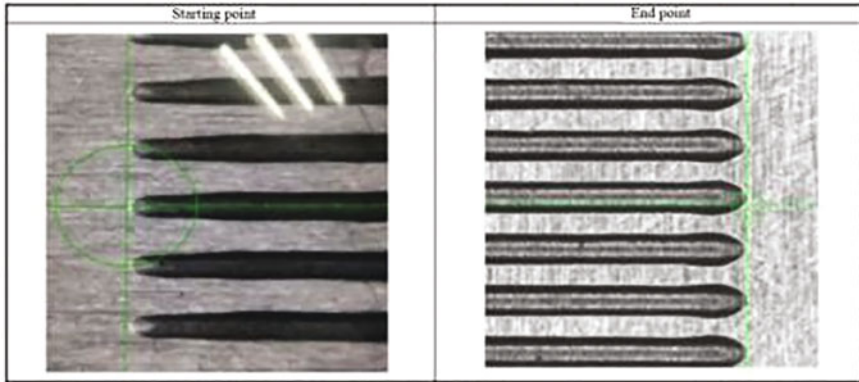


Fig. 10 Dispensing process

3.5 *Dispensing Process Effect*

The effect of the dispensing process is shown in Fig. 10.

4 Conclusion

In the application of dispensing process, although new non-contact dispensing methods are used more frequently, due to the characteristics of certain fluids and the needs of customers' specific applications, contact dispensing methods are still indispensable. It has been verified through testing that the rotary dispensing valve can better improve the problem of needle overflow caused by residual pressure and excessive instantaneous amount of glue when the valve is closed. In addition, it has greater advantages and better dispensing effects compared to other types of valve bodies in low-viscosity fluid applications, with good air-tight characteristics.

Acknowledgements Funding from the Zhuhai City Innovation Team Project in 2018 Foundation is gratefully acknowledged. The project name is R&D and industrialization of automatic precision dispensing equipment based on machine vision. The project approval number is ZH01110405180012PWC.

References

1. Fan, Z., Rong, W., Wang, L., Sun, L.: Control and experiment of piezoelectric actuated micro-dispenser. *Opt. Precis. Eng.* (2016)
2. Gu, S., Wang, Z., Jiang, H., Liu, J., Lu, S., Liu, Q.: Piezoelectric drive hydraulic amplified jetting system. *Opt. Precis. Eng.* (2015)

3. Lu, S., Jiang, H., Gu, S., Liu, Q., Liu, J., Liu, X.: Design and experiment of hydraulic piezoelectric driven jet dispensing valve. *J. Sichuan Univ. (Engineering Science Edition)*. (2015)
4. Kunchala, K.: 3D printing high density ceramics using binder jetting with nanoparticle densifiers. *Mater. Des.* (2018)
5. Deng, G., Cui, W., Zhou, C., Li, J.: A piezoelectric jetting dispenser with a pin joint. *Optik*. (2018)
6. Khan, S., Nguyen, T.P., Lubej, M., Thiery, L., Vairac, P., Briand, D.: Low-power printed micro-hotplates through aerosol jetting of gold on thin polyimide membranes. *Microelectron. Eng.* (2018)

The Positioning Accuracy Measurement of the Dispenser and Compensation Method



Xiong Huang, Gaobo Xiao, Zhiguo Liu, Shifei Zhang, Qiwen Wu, and Fanbiao Bao

Abstract This article describes a method for measuring the positioning accuracy of dispenser and the compensation method. This method uses customized high-precision calibration boards, CCD imaging technology and vision algorithms to measure the XY two-dimensional positioning accuracy of the machine, and compensate the measured error results to the motion system to improve the positioning accuracy of the machine. Since the deviation is calculated by visually grabbing the Mark points on the calibration board, it is lower in cost, easier to operate, and more efficient than the traditional laser interferometer measurement method. Furthermore, the traditional laser interferometer can only perform single-axis compensation, while the calibration plate method can compensate XY at the same time, which truly improves the positioning accuracy.

Keywords Dispenser · Positioning accuracy · Two-dimensional compensation

1 Introduction

1.1 Purpose and Significance of the Study

Nowadays, with the widespread use of electronic products, the market demand for dispensers is also expanding. However, domestic dispensers generally have low accuracy and insufficient glue dispensing, while the electronic products are becoming lighter and smaller, with higher requirements for accuracy. Consequently, many manufacturers in the purchase of dispensers can only reluctantly look for world

X. Huang · G. Xiao · Z. Liu · S. Zhang · Q. Wu
Zhuhai Bojay Electronics Co., Ltd, Zhuhai, Guangdong Province, China

F. Bao (✉)
Beijing Institute of Technology, Zhuhai, Guangdong Province, China
e-mail: 83667425@qq.com

brands. Therefore, in the high-precision area, continuous efforts of equipment manufacturers are expected. There is no doubt that only quality and service can make the manufacturers stand out when the market competition becomes fierce.

Due to various reasons such as processing and assembly, the positioning accuracy of the dispenser equipment is difficult to be guaranteed. Taking a dispenser equipment of our company as an example, within the range of $400\text{ mm} \times 400\text{ mm}$, after the equipment is assembled, the positioning deviation of the machine with good performance can be up to 0.04 mm , while the positioning deviation of the machine with poor performance can exceed 0.1 mm .

As a result, a simple and low-cost method is needed to improve the positioning accuracy.

1.2 Main Content

1. Design and manufacture the calibration board: design a suitable calibration board according to the XY stroke of the equipment, the camera field of view of the equipment, and the accuracy requirements. The Mark of the calibration board should form a checkerboard, and ensure that the range of the Mark can cover the range of the main dispensing stroke. Meanwhile, the algorithm should capture the Mark to calculate the deviation.
2. Write the positioning photography process: program the positioning photography process, to make XY move according to the specified distance, and to ensure that the camera can accurately shoot each Mark point.
3. Write the Mark grasping algorithm: after each Mark is shot, the vision needs to calculate the deviation and save it to form a compensation file.
4. Write the motion compensation algorithm: the program loads the compensation file. When motion positioning, interpolation compensation is performed according to the grid interval of the calibration plate where the positioning point falls.

This method is designed for an on-line automatic dispenser of our company, but it is also suitable for other types of dispensers or similar equipment. Customizing the calibration board of different sizes to realize the compensation of positioning accuracy. For the convenience of description in the paper, the subsequent names of the equipment will be referred to as the dispenser [1].

1.3 Reasons for Poor Positioning Accuracy

1.3.1 Possible Reasons

1. The perpendicularity of x-axis and y-axis assembly of the equipment cannot reach absolute 90°.
2. The unevenness of optical grating ruler leads to the scaling of moving distance.
3. The deformation of beam or guide rail.
4. The parallelism between Y-axis spindle and support rail is not enough, resulting in extrusion during movement.

1.3.2 Impacts

1. Poor dispensing accuracy leads to deviation of dispensing position or deformation of dispensing shape.
2. There will be deviation when reusing different machine programs, and when the dispensing track is moved to other machines, there will be deviation.

2 Design of Calibration Board

The maximum effective stroke of XY of the dispensers is 400 mm × 400 mm. The origin of the device is defined as the lower left corner, the X-axis coordinate is positive to the right, and the Y coordinate is positive to the front. The positioning camera is a 130 W pixel CCD with a minimum field of view of 4 mm, as shown in Fig. 1.

According to these information, the calibration plate is designed, which is 420 mm long and 420 mm wide, and the range of internal dots is 400 mm × 400 mm. The diameter of the circle is 2 mm, and the center distance between two adjacent circles is 4 mm. There are 101 circles in the transverse direction and 101 circles in the longitudinal direction, as shown in Fig. 2.

In order to ensure accuracy, a glass calibration plate is selected with an accuracy of ± 2 μm. It can be customized by a qualified calibration board supplier, as shown in Fig. 3 [2].

Fig. 1 Equipment picture



Fig. 2 Schematic diagram of calibration board design

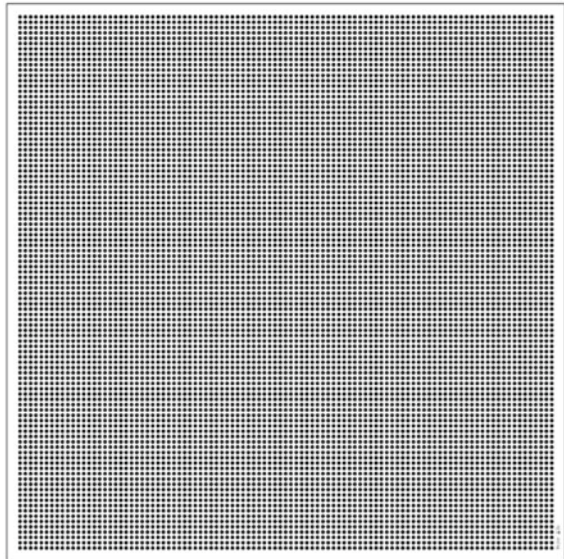
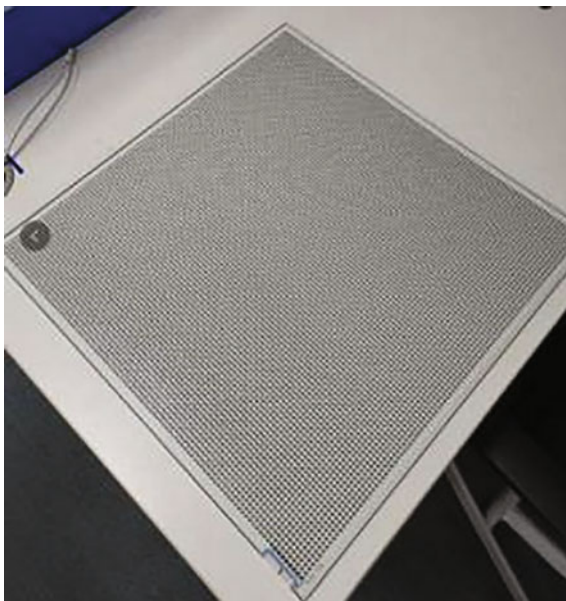


Fig. 3 Physical image of calibration board



3 Process of Positioning Accuracy Error Measurement

3.1 Fixing of Calibration Board

Lay the calibration board flat on the dispensing track, try to make the edge of the calibration board parallel to the X-axis, reduce the error caused by the simultaneous movement of X and Y, and also reduce the error introduced by the angle calculation.

Use the altimeter on the machine to measure the height of the four corners of the calibration board to ensure the levelness of the calibration board.

Use clamps to fix the calibration board to prevent shaking as shown in Fig. 4.

3.2 Camera Focus and Calibration

3.2.1 Camera Focus

Move the camera to the top of a dot on the calibration board and adjust the Z axis to ensure that there is only one complete dot in the camera field of view. If the field of vision is not suitable, it can be adjusted by increasing or decreasing the circle.

Fine adjust the focal length of the camera to ensure that the edge of the dot in the image is clear, as shown in Fig. 5 [3].

Fig. 4 Calibration board placement diagram

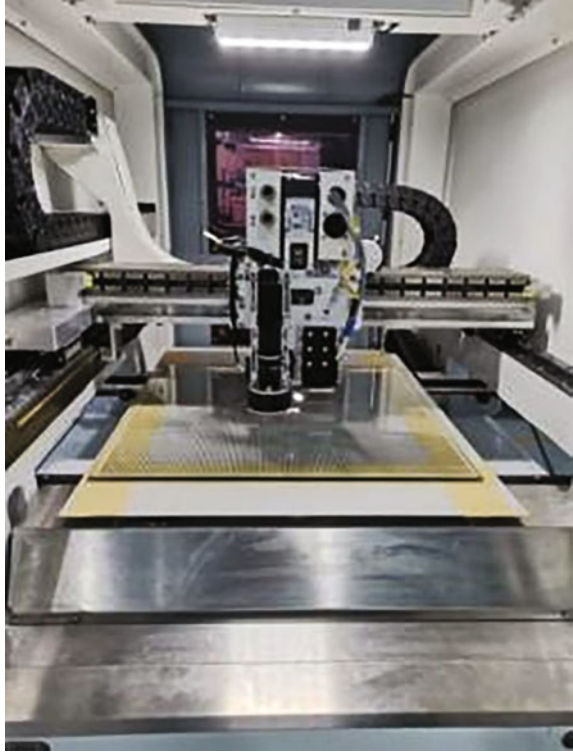
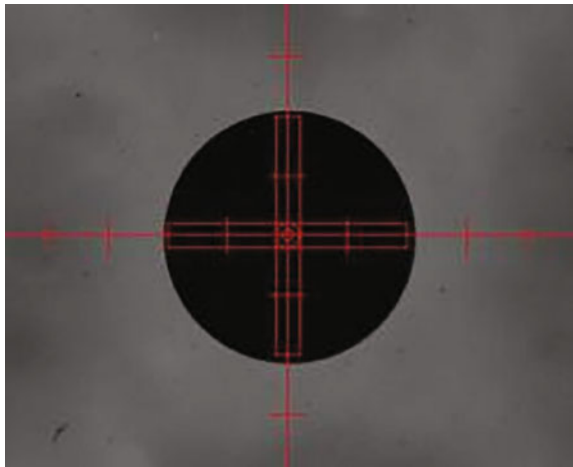


Fig. 5 Dot imaging diagram of calibration plate



3.2.2 Camera Calibration

Use the visual nine-point calibration method that comes with the dispenser system for calibration. Get the ratio of pixels to motion. For example, 0.006 mm/pix.

3.3 Mark Point Photograph

3.3.1 Photographing Path Design

In order to conveniently set the photographing order and range of the Mark points of the calibration board, and also to correct the parallelism between the calibration board and the X-axis, we have written an interface to input the coordinates of the first dot on the calibration board and the last dot on the first line, as well as the number of rows and columns, and the dot interval, automatically generating photo spots.

The coordinate of the first dot, that is, the starting point of the angle line in the figure below, is the X, Y motion coordinates when the center of the first dot on the indicator board is at the center of the field of view [4].

The coordinates of the last dot in the first row, that is, the end point of the angle line in the figure below, are the X, Y motion coordinates when the center of the last dot in the first row of the indicator board is at the center of the field of view, as shown in Fig. 6.

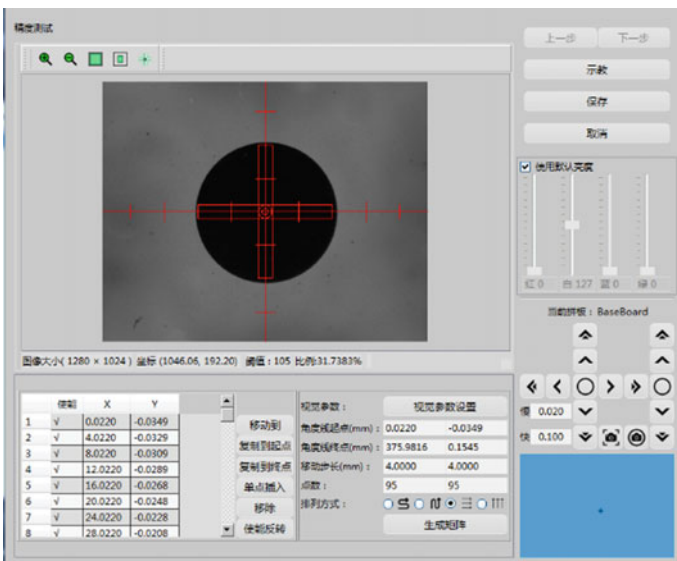
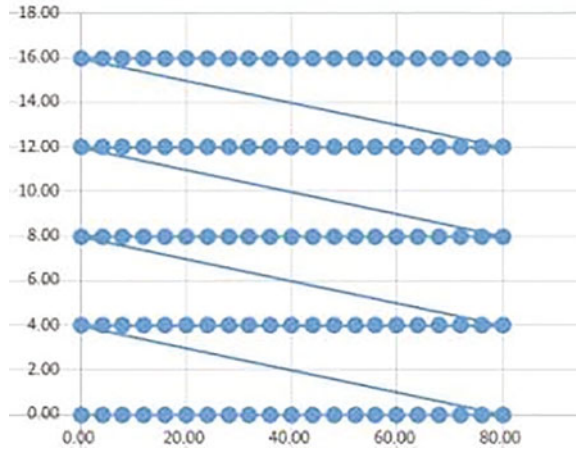


Fig. 6 Setting diagram of photo spot

Fig. 7 Schematic diagram of the photo path



The generated coordinate points are separated by 4 mm, and automatically cut to the beginning of another line when the end of the line is reached.

During operation, start from the first point at the lower left corner of the calibration board, take photos point by point according to the line, move to the first point of the next line after taking one line, take photos point by point according to the line, and the combined distance of X and Y of the motion axis is 4 mm, as shown in Fig. 7.

3.3.2 Precautions for Photographing

To ensure accuracy, the moving speed should not be too fast and should not exceed 50 mm/s. After the moving, take a picture with a delay of 50 ms to ensure that the motion stops steadily.

3.4 Deviation Calculation and Storage

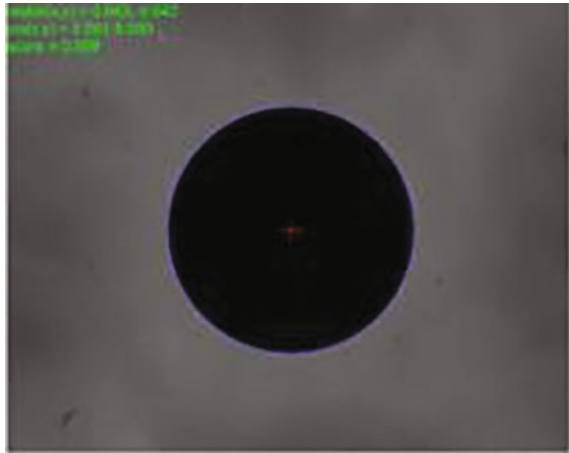
Image algorithm is used to capture the center of the circle and get the deviation value (unit pixel) of the center of the circle from the center of the field of view. Then, according to the number of MM corresponding to each pixel calibrated before, convert it into mm unit.

Write the photographing coordinates, X offset and Y offset into the configuration file.

After all the mark points are taken, the positioning deviation table, that is, the two-dimensional compensation table, can be obtained, as shown in Fig. 8.

The abscissa indicates the number of points, and the ordinate indicates the deviation. X offset refers to the deviation in the X direction from the first dot to the last

Fig. 8 Deviation value before compensation data



	A	B	C	D
1	photoX	photoY	offsetX	offsetY
2	3.157	127.5006	0.00021	0.00013
3	7.1573	127.501	-0.00103	-5.7E-05
4	11.1573	127.5014	0.000382	-0.0003
5	15.1573	127.5017	-0.0012	0.000126
6	19.1573	127.502	0.001015	0.001386
7	23.1573	127.5023	0.000896	0.000868
8	27.1573	127.5027	0.000121	0.001131
9	31.1573	127.503	0.000595	0.001738
10	35.1573	127.5033	0.001332	0.001846
11	39.1573	127.5037	0.001685	0.001454
12	43.1573	127.504	0.001219	0.000575
13	47.1573	127.5043	0.001943	0.001028
14	51.1573	127.5047	0.000758	0.000696
15	55.1573	127.505	-0.00099	0.0013
16	59.1573	127.5053	-0.00096	0.002036
17	63.1573	127.5056	-0.00273	0.001986
18	67.1573	127.506	-0.00191	0.00229
19	71.1573	127.5063	-0.00235	0.002915
20	75.1573	127.5066	-0.00177	0.003573
21	79.1573	127.507	-0.00194	0.003839
22	83.1573	127.5073	-0.00146	0.004231
23	87.1573	127.5076	-0.00241	0.003541
24	91.1573	127.508	-0.00185	0.003703
25	95.1573	127.5083	-0.00192	0.002718

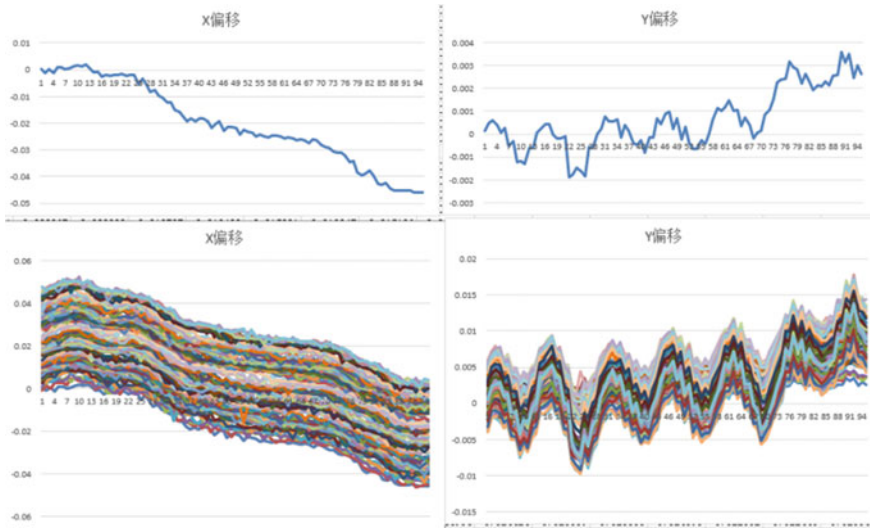


Fig. 9 Deviation value before compensation

dot in each row, and Y offset refers to the deviation in the Y direction from the first dot to the last dot in no column [4].

The upper part of Fig. 9 is the deviation data of a single row or single column; the lower part is the deviation data of all rows or columns.

4 Compensation and Verification

4.1 Compensation Principle

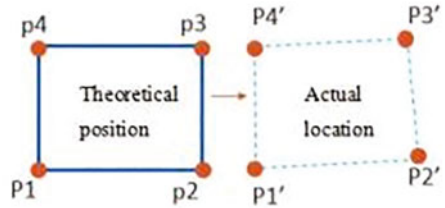
The first two data in the table are the photographing points (x, y), the latter two data are the deviation values (dx, dy), and the corrected points (x', y') can be obtained according to the deviation values [5].

The calibration board can be seen as a bunch of dense small rectangles. Take the photographing coordinates of the dots at the 4 corners of each small rectangle and perform affine transformation with the corrected 4 points to obtain 100*100 transformation matrices, as shown in Fig. 10.

4.2 Transformation Function

As a formula 1:

Fig. 10 Transformation matrix



$$p_{1'} = p_{10} * p_{11} + p_{12} * p_{13} + p_{14} \tag{1}$$

$$p_{2'} = p_{20} * p_{11} + p_{22} * p_{13} + p_{24} \tag{2}$$

4.3 Transformation Matrix

As a formula 2:

$$\begin{bmatrix} p_{1'} \\ p_{2'} \\ 1 \end{bmatrix} = \begin{bmatrix} p_{10} & p_{11} & p_{12} \\ p_{20} & p_{22} & p_{24} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{10} \\ p_{11} \\ 1 \end{bmatrix} \tag{3}$$

When issuing a motion instruction, determine which cell the motion point belongs to, perform the corresponding transformation, obtain the compensated coordinate, and then issue the compensated motion coordinate to the motion system.

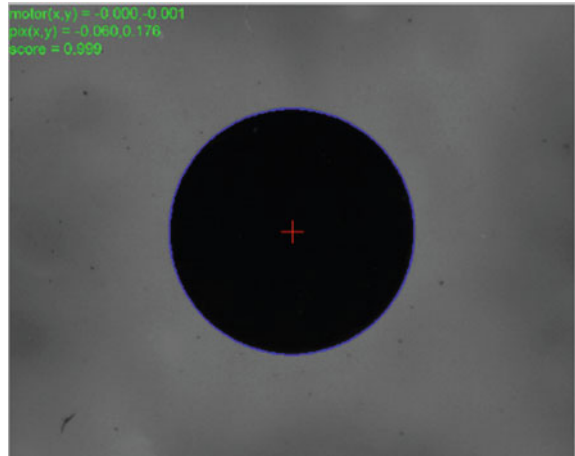
4.4 Verification Method

Based on the center of the first dot of the calibration plate, a series of points are generated along the direction of the calibration plate, 4 mm horizontally and 4 mm vertically.

Go to each point to take a picture and calculate the deviation value. As shown in Fig. 11.

From the data, the positioning accuracy is greatly improved after compensation. The positioning accuracy before compensation is about 0.05, while after compensation is about 0.01 mm.

Fig. 11 Deviation value after compensation



5 Factors Affecting Compensation Accuracy

5.1 Camera Accuracy

At present, the pixel accuracy of the camera equipped with the dispenser is 0.006 mm. If a higher-precision compensation effect is expected, you need to replace a higher-resolution camera.

5.2 Calibration Board Accuracy

The calibration board is customized by a professional manufacturer with an accuracy of ± 0.002 mm. If a higher-precision compensation effect is expected, you need to make a higher-precision calibration board.

5.3 Fixing of Calibration Board

When the calibration board is fixed, the levelness must be ensured. Insufficient levelness will result in deviations in the photographing calculation. Moreover, it must be firm and will not deviate. During the shooting of motion, any shaking will introduce errors.

5.4 Machine Stability

The machine must be placed on the ground steadily, without shaking during movement. Therefore, during the measurement process of the calibration board, the movement speed should not be too fast, and it needs to be stopped steadily before photographing.

6 Conclusion

1. Due to various reasons, the positioning accuracy of the dispenser is difficult to meet the needs of high-end customers.
2. Using laser interferometer and other methods to compensate, the cost is too high, the operation is complicated, and the time-consuming is long.
3. This compensation method is simple, efficient and low-cost. After compensation, the positioning accuracy can basically meet customer needs.

Acknowledgements Funding from the Zhuhai City Innovation Team Project in 2018 Foundation is gratefully acknowledged. The project name is R&D and industrialization of automatic precision dispensing equipment based on machine vision. The project approval number is ZH01110405180012PWC.

References

1. Peng, X., Shi, J., Wu, Q., Tang, L.: Automatic dispenser system design based on image positioning. *Sci. Technol. Innov. Herald* (2017)
2. Lin, X.: Research on control system of automatic dispenser. *Sci. Technol. Outlook* (2016)
3. Peng, X., Shi, J., Wu, Q., Tang, L.: Design and implementation of automatic dispenser system based on visual control technology. *China New Commun.* (2017)
4. Li, X., Liu, B., Ouyang, X., Bai, Y., Dang, J., Lei, J.: Design of intelligent dispenser control system based on FPGA and SRAM. *Chinese J. Sci. Instrum.* (2009)
5. Wu, F., Qu, D., Ma, Z., Li, C.: Research on positioning system of dispenser based on fusion of multi-sensor information. *Comput. Measur. Control* (2017)

X/Y/Z High-Speed and High-Precision Operation Platform Design



Jun Guo, Gaobo Xiao, Zhen Xing, Zhiguo Liu, Shifei Zhang, and Yongqi Wu

Abstract This paper mainly describes the structure design of a high-speed and high-precision operation platform in the automatic dispenser industry. Firstly, through the analysis of the development of domestic automatic dispenser and the mechanism of automatic dispenser, the importance of X/Y/Z high-speed and high-precision operation platform is summarized. Referring to the relevant information, this paper analyzes and compares the existing automatic dispenser mechanisms in the market, describes their advantages and disadvantages, and makes its own improvement on the basis of these mechanisms. This design mainly includes the design of x-axis module, Y-axis module and z-axis module, which mainly involves the material comparison and parameter design of casting base, selection of linear motor, linear guide rail, servo motor, precision grinding screw rod, grating ruler and photoelectric sensor, design parameter check, material comparison of x-axis beam and tolerance analysis of important parts. Furthermore, this design chooses to use SolidWorks three-dimensional modeling software to complete the X\Y\Z high-speed and high-precision operating platform structure design, 2D assembly drawing and drawing of its main parts.

Keywords X/Y/Z high-speed and high-precision operating platform design · Linear motor · Linear guide · Precision screw rod · Bearing pedestal

J. Guo · G. Xiao · Z. Xing · Z. Liu · S. Zhang
Zhuhai Bojay Electronics Co., Ltd., Zhuhai, Guangdong Province, China

Y. Wu (✉)
Beijing Institute of Technology, Zhuhai, Guangdong Province, China
e-mail: 1048618721@qq.com

1 Introduction

1.1 *Research Background and Significance of Automatic Dispenser*

The self-developed glue dispenser has a low degree of automation and requires manual control of the glue head to the designated position. After reaching the designated position, manual squeezing of glue results in a high defect rate of traditional manual dispensing products. Gradually, companies began to use dispensing equipment for industrial production dispensing. Its advantage is that workers can control the amount of glue that is operated by the dispensing equipment, and the dispensing is relatively uniform. The disadvantage is that the manual operation of dispensing is low in automation, accuracy and production efficiency, which makes it less competitive compared with foreign products. Therefore, domestic small and medium-sized enterprises must purchase high-precision and highly automated dispensing equipment in order to improve their market competitiveness. Facing the current situation, the XYZ high-speed and high-precision operation platform structure of the automatic dispensing machine is studied in this project. Design a XY axis composed of linear motor and linear guide rail, with a Z-axis composed of servo motor, screw and guide rail, to develop a three-axis automatic dispensing machine that can operate in a three-dimensional space at high speed and high precision. In this way, the competitiveness of the domestic electronics market as well as industrial automation can be improved [1].

1.2 *The Development of Automatic Dispenser*

The development and production applications of various emerging industries have made the dispenser industry an indispensable role. Dispensing machines can be seen in the fields of food, medical equipment, PCB single-chip microcomputers, production and textiles, and laser welding. It is expected that in the upcoming new industry chain, there will still be many potential markets for dispensers, and the development of dispensers will also move toward higher automation and intelligence.

For simple dispensing equipment, although the control of motion feed has been improved through domestic research and innovation, enabling it to produce some complicated and difficult graphics, there are still some defects such as poor dispensing accuracy, insufficient motion sensitivity, high noise, short life of moving parts such as lead screws, and low automation. Therefore, the development and upgrade direction of domestic dispensers is still high-precision feed motion control, high-automation, and high-quality.

1.3 The Main Content of the Design

The main content is to design an X/Y/Z high-speed and high-precision operation platform based on the traditional manual dispenser. It is required to use linear motors and linear guides in the X and Y directions to achieve high-speed, high-precision, and high-stability operation. Using servo motors, lead screws, and linear guides, the Z-axis can achieve a speed of 250 mm/s and a load of ≤ 3 kg to stably run to the specified position [2].

2 Main Design Parameters of X/Y/Z High-Speed Operation Platform

The X/Y/Z high-speed operating platform is shown in Fig. 1.

2.1 Main Functions and Scope of Application

The X/Y/Z high-speed operation platform is suitable for the precise movement of the dispenser in the X, Y, and Z directions [3].

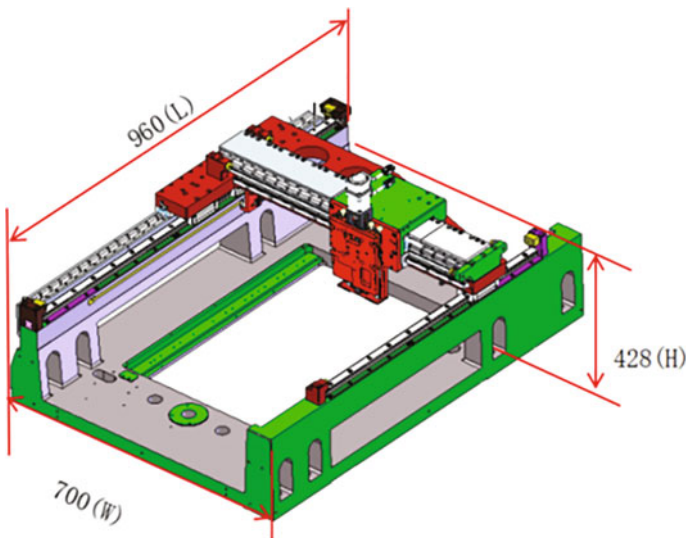


Fig. 1 X/Y/Z high-speed operating platform

2.2 Description of Main Technical Parameters

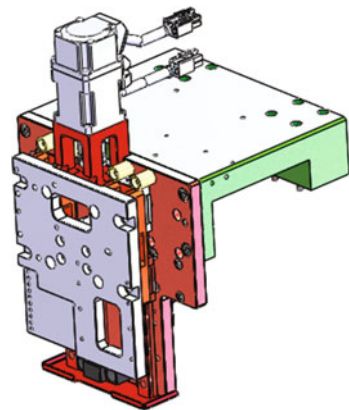
The X/Y/Z high-speed operating platform is shown in Fig. 2.

1. Material of pedestal: HT250 (grey cast iron)
2. Processing technology of pedestal: casting
3. Dimensions: 960(L) * 700(W) * 477(H) mm
4. Working size: 520(L) * 394(W) * 70(H) mm
5. Main functional components:
6. X direction – linear motor + linear guide + grating ruler
7. Y direction – linear motor + linear guide + grating ruler
8. Z direction – servo motor + screw + guide rail
9. High speed: speed can reach 1.0 m/s, and acceleration can reach 1.0G
10. High precision (repeat positioning accuracy): X/Y can reach $\pm 10 \mu\text{m}$, and Z can reach $\pm 20 \mu\text{m}$
11. High resolution: using $0.5 \mu\text{m}$ grating ruler, high resolution can be achieved

3 X/Y/Z High-Speed Operation Platform Module Composition

The X/Y/Z high-speed operating platform module is composed of the following three components. The installation sequence and accuracy guarantee are all based on the Y-axis pedestal. In addition to the strict control of the processing accuracy of the key parts of each component, the assembly accuracy of the components also has strict requirements, so as to achieve the requirements of high precision, high speed, and high stability [4].

Fig. 2 Positioning of X/Y/Z high-speed operation platform



4 X/Y/Z High-Speed Operation Platform Module Design Principle

4.1 Introduction of Y-axis Module

The Y-axis module is mainly composed of a pedestal, a motor, a linear guide, a photoelectric sensor, a grating ruler, and a limit block module, as shown in Fig. 3.

4.1.1 Casting Base Design

The casting pedestal is an open-mold casting made of HT250. This material and processing method have characteristics of: high strength, wear resistance, heat resistance, and good shock absorption, and the tensile strength can reach 250 MPa.

Pedestal size: 960 * 700 * 198. The design principle is determined based on the industry product size and the dispensing range of the equipment itself, as shown in Fig. 4.

Pedestal weight: 166.92 kg. This automated dispenser requires high precision and high speed operation, so the platform needs to be strong and stable, especially using casting technology. In addition, the mold casting process has good processing consistency, and also has great advantages in price and quality, which is convenient for cost and quality control.

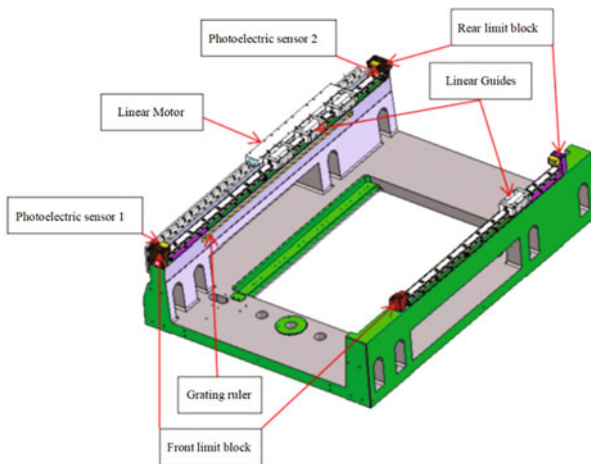
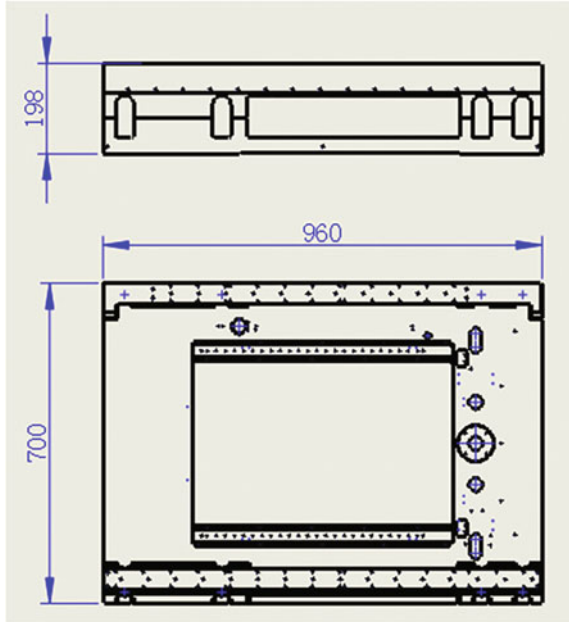


Fig. 3 Y-axis module

Fig. 4 Casting pedestal



4.1.2 Linear Motor Selection

The linear motor has the characteristics of simple structure, high positioning accuracy, fast acceleration and high efficiency.

According to the Y-axis load and stroke, the appropriate linear motor model we need can be calculated.

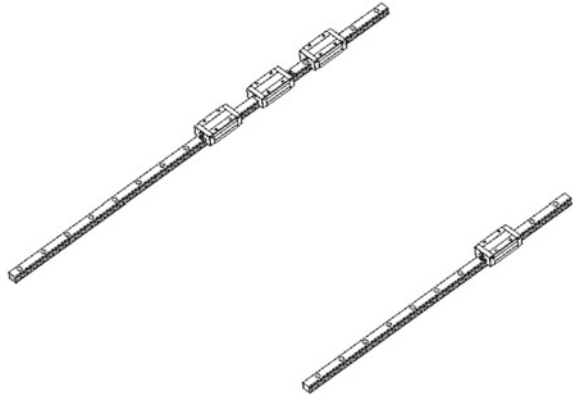
4.1.3 Linear Guide Selection

Due to the load and high precision requirements, we use THK's preloaded ball linear guide, as shown in Fig. 5.

4.1.4 Photoelectric Sensor Selection

The photoelectric sensor has the advantages of high response frequency, high detection accuracy, non-contact detection, and long service life. It is used for limit position detection on X/Y/Z-axis [5].

The output signal type is NPN, the maximum working voltage is 24 Vdc, the working current is 50 mA, the wiring length is 1 m, the detection distance is 5 mm, and the outline size is L37 * W18 * H18 mm.

Fig. 5 Ball linear guide

4.1.5 Grating Ruler Selection

Grating ruler (grating ruler displacement sensor) is a measurement feedback device that uses the optical principle of grating.

The size of the grating ruler is $W6 * T0.2$ mm, the grating pitch is 20 μm , the linear accuracy can reach $\pm 3 \mu\text{m/m}$, the installation method is surface mount, and the sensor is adapted.

4.2 Introduction of X-axis Module

4.2.1 X-axis Beam Design Principle

It spans 2 Y-axis guide rails and carries the weight of X-axis motor and Z-axis. At the same time, it needs to ensure high-speed and high-precision operation, so stable hardware support is necessary.

The size of the beam we designed is $713.5 * 361 * 28$ mm, the material is AL6061, the natural color is matt anodized, and the ultrasonic cooling method removes the stress. At the same time, there is a notch designed to reduce weight and reduce Y-axis load.

4.2.2 Other Standard Parts Selection

The linear motor, linear guide, photoelectric sensor, grating ruler, and limit block module are the same as the Y-axis design principle, because the stroke, load, accuracy and other conditions are not used, the selection is slightly different [6].

4.3 Z-axis Module Composition

It is mainly composed of Z-axis sliding table, servo motor, ball screw, linear guide, photoelectric sensor, coupling, bearing pedestal, and limit post, as shown in Fig. 6.

4.3.1 Z-axis Module Power Mechanism

It is composed of servo motor, coupling, ground ball screw, linear guide and bearing pedestal.

The upper and lower limit position adopts EE-SX951-R photoelectric sensor: the output signal type is NPN, the maximum working voltage is 24Vdc, the working current is 50 mA, the wiring length is 1 m, the detection distance is 5 mm, and the external dimension is L37 * W18 * H18 mm.

4.3.2 Servo Motor Selection Calculation Principle

The module is controlled by a servo motor, as shown in Fig. 7 [7].

It can be obtained that:

Motor speed, such as formula 1.

$$N_M = V/P \tag{1}$$

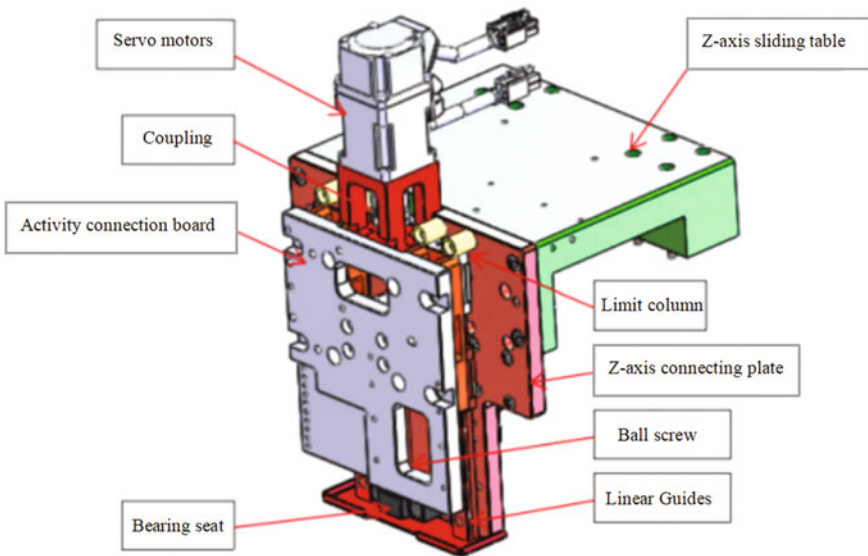


Fig. 6 Axis module

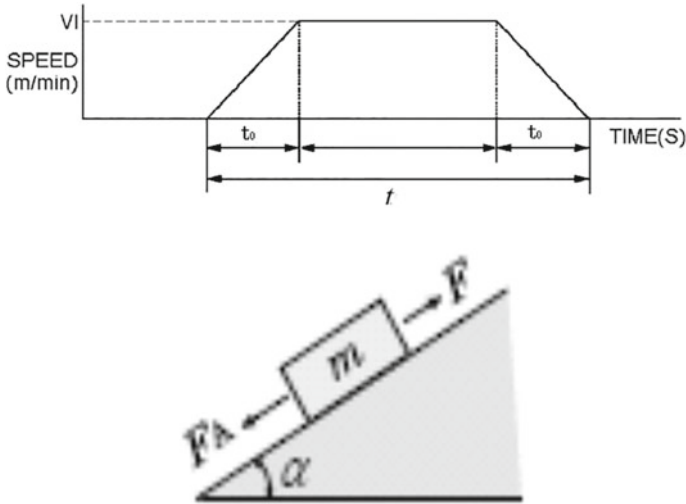


Fig. 7 Servo motor

Load torque calculation:

Axial load, such as formula 2.

$$F = F_A + mg \sin a + \mu \cos a \tag{2}$$

Load torque, such as formula 3.

$$T_L = \frac{F P_B}{2\pi \eta} \tag{3}$$

Starting torque calculation:

Linear motion platform and load inertia, such as formula 4.

$$J_L = m \left(\frac{P_B}{2\pi} \right)^2 \tag{4}$$

Ball screw inertia, such as formula 5.

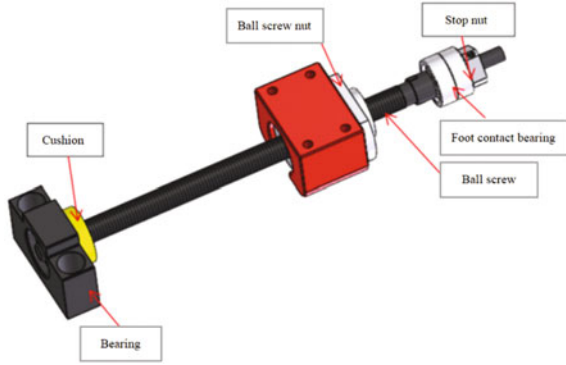
$$J_B = \frac{\pi}{32} \rho L_B D_B^4 \tag{5}$$

Coupling inertia, such as formula 6.

$$J_C = \frac{1}{8} m D_C^2 \tag{6}$$

Total load inertia, such as formula 7.

Fig. 8 Precision ground screw



$$J_L = J_L + J_B + J_C \tag{7}$$

Starting torque, such as formula 8.

$$T_s = \frac{2\pi N_M J_L}{60t_1} \tag{8}$$

According to the above selection calculation, it can be determined that the following motors can meet the functional requirements of the Z-axis.

4.3.3 Precision Grinding Screw Selection

According to the comprehensive calculation of load, stroke, speed, accuracy, space, etc., we choose 0805 specially customized precision grinding screw.

Due to the selection of standard parts, the supporting bearing pedestal is too large, we use the form of bearing + machined parts, and design the bearing pedestal according to the size of the space, so that the Z-axis as a whole is more compact and lighter, and it can also increase the dispensing range in the Y-axis direction.

As shown in Fig. 8, angular contact bearings mainly bear larger unidirectional axial loads and can work at higher speeds. The greater the contact angle, the greater the load capacity. The Z-axis only need to bear axial load, therefore, angular contact bearings are used here.

4.3.4 Coupling Selection Principle

The coupling has the ability to compensate for the offset (including axial offset, radial offset, angular offset or comprehensive offset) between the two shafts due to incorrect manufacturing and installation, deformation or thermal expansion caused by work, reducing the shock and vibration.

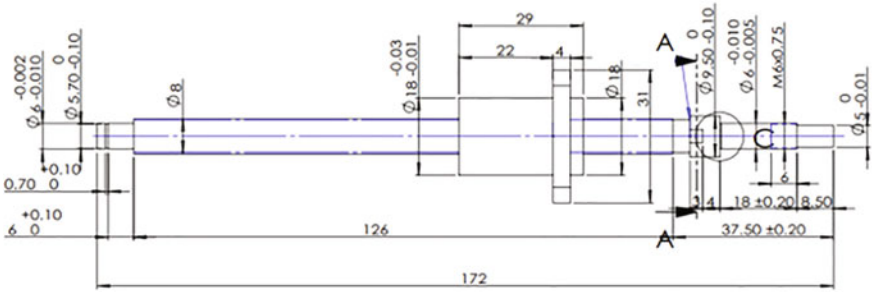


Fig. 9 Coupling selection

As shown in Fig. 9, because the servo motor shaft is $\phi 8$ and the screw rod is $\phi 5$, the coupling is DBM01-D19-d5-e8.

4.3.5 Buffer Column Material Selection

The buffer column is used at the upper limit of the Z-axis. When the stroke exceeds the sensor, the slider connector will hit the buffer column to protect it. Therefore, the material of the buffer column should not be too soft, as too soft would hardly provide protection; nor too hard or metal, which will damage the parts and cause a loud impact.

4.3.6 Design Principle of Main Processing Parts

Z-axis sliding table: It is mainly connected to the X-axis linear guide, which is the benchmark for guaranteeing the accuracy of all Z-axis components. To be able to carry all the load weight of the Z-axis, and to ensure the basis of high-speed and high-precision operation, but not to increase the load and affect the X-axis load, we choose AL6061 material with anodized surface. Therefore, it is necessary to have a high machining accuracy on the assembly surface.

Z-axis connecting plate: It is mainly connected with X-axis linear motor and Z-axis sliding table. It is the basic part of Z-axis components and the basis of assembly. In order to ensure the accuracy of the assembled Z-axis relative to the X/Y axis, the surface needs to have good flatness and parallelism requirements. Especially the height difference and flatness of the motor fixing pedestal and the screw fixing pedestal and the linear guide surface will directly affect the smoothness of the screw operation.

Other machined parts: The entire Z-axis needs to ensure the accuracy requirements after assembly. In addition to the above two basic parts, the thickness, flatness, dimension reference and tolerance requirements of other parts related to the screw

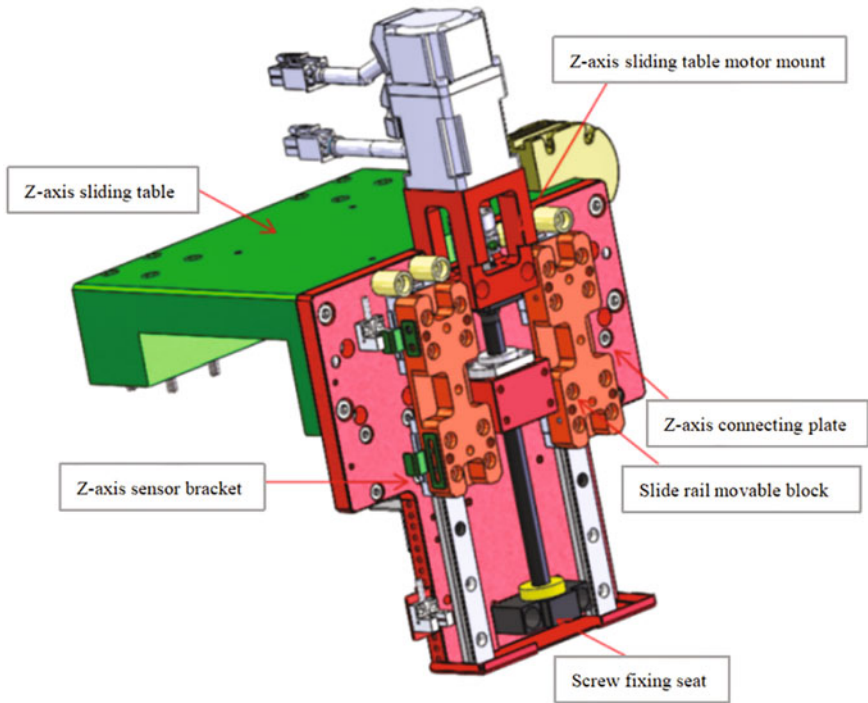


Fig. 10 Z-axis sliding pedestal

rod also require special attention. These may cause the screw rod to jam and affect the smoothness, thereby affecting the accuracy and service life [8].

Sensor bracket: The lower limit adopts the method of long holes and multi-holes reserved, and the limit sensing position can be adjusted according to the thickness/height of different products, as shown in Fig. 10.

5 Conclusion

The paper completed this time is the design of the XYZ high-speed operating platform. The following are some conclusions of this completed design paper:

5.1 The Overall Structural Design Background

The design principle of XYZ high-speed operation platform is mainly that according to different industries, the product size, weight, accuracy and speed requirements

are different. There will also be different sizes of mechanism frames and power parts selection and matching. High-precision online dispensers have a wide range of application prospects: FPC, consumer electronics, automotive, medical, semiconductor packaging, precision assembly LED, LCD industry application requirements. Based on the verification data of different customers and different products, we have comprehensively determined that the design of this X\Y\Z high-speed operating platform is designed based on the dispensing area of 360 * 380 mm.

5.2 *Single Module Frame Design*

1. Z-axis module production: According to the requirements of speed, stroke, load, accuracy, etc., standard parts such as servo motor, screw rod, guide rail, coupling, etc. are selected and assembled with Solid works software.
According to the weight of these standard parts and parts, plus other accessories, estimate the weight of the entire Z-axis for the selection of the X-axis linear motor.
2. X-axis module production: According to the weight of the Z-axis and the dispensing area, select the power of the linear motor and the model of the linear guide (preload, accuracy, length...).
According to the load and stroke, use the Solid works software to draw the X-axis measurement bracket, assemble the selection standard room, and design the limit sensor as well as the hard limit.
3. Y-axis design: Production of casting base—First, determine the maximum product size and height according to the dispensing area to determine the size of the base's inlet and outlet gaps.
According to the X and Z-axis load, and the effective dispensing area, select the Y axis linear motor and linear guide (high precision, medium preload, stroke...).
Selection of grating ruler and selection of installation reference.

5.3 *XYZ Matching Design*

First of all, according to the size of the dispensing area, we can draw the simulated DUT in solids work, the size is 400 * 400 * 5 mm;

Then, match the X\Y\Z axis modules together with DUT-3D for inspection.

5.4 *Detailed Design*

According to the assembled modules, adjust and refine some details such as hole positions, sizes, positions, tolerances, etc., and sort out a complete 2/3D drawing.

Acknowledgements Funding from the Zhuhai City Innovation Team Project in 2018 Foundation is gratefully acknowledged. The project name is R&D and industrialization of automatic precision dispensing equipment based on machine vision. The project approval number is ZH01110405180012PWC.

References

1. Hu, F.: Interchangeability in Technical Measurement Basis. Higher Education Press (2010)
2. Jiao, Y., Zhang, J., Xu, C.: Engineering Drawing. Higher Education Press (2008)
3. Xu, X.: Interchangeability and Measurement Technology Basis. Hunan University Press (2005)
4. Yang, K., Cheng, G., Li, Z.: Fundamentals of Mechanical Design. Higher Education Press (2012)
5. Luo, S., Zhu, S.: Concise Handbook of Mechanical Curriculum Design. Chemical Industry Press (2006)
6. Liu, H.: Mechanics of Materials, 5th edn. Higher Education Press (2012)
7. Tao, Y., Wang, H.: Fundamentals of Engineering Materials and Mechanical Manufacturing, 2nd edn. Chemical Industry Press (2016)
8. Zhang, C., Yang, W., Yang, P., Lu, S.: Mechanical Design. Xi'an Technological University Press (2016)

Differentiable Architecture Searched Network with Tree-Structured Parzen Estimators for Rotating Machinery Fault Diagnosis



Jingkang Liang, Yixiao Liao, and Weihua Li

Abstract Deep learning is widely used in the field of rotating machinery fault diagnosis. However, manually designing the neural network structure and adjusting the hyperparameters for specific fault diagnosis task are complex and requires a lot of expert knowledge. Aiming at these problems, Differentiable Architecture Searched Network with Tree-Structured Parzen Estimators (DASNT) is proposed for fault diagnosis. Differentiable Architecture Search (DARTS) is utilized to automatically search network structure for specific fault diagnosis task. Tree-Structured Parzen Estimators (TPE) is utilized to optimize the hyperparameters of the network searched by DARTS, which can further improve the fault diagnosis accuracy. The results of comparison experiments indicate that the network architecture and hyperparameters optimized by DASNT can achieve superior fault diagnosis performance.

Keywords Fault diagnosis · Neural architecture search · Hyperparameter optimization

1 Introduction

Modern electromechanical equipment tends to be complicated, sophisticated, high-speed, and intelligent. As an important part of electromechanical equipment, the failure of rotating machinery will cause serious security risk and economic loss. Fault diagnosis is utilized to diagnose the fault type, location, and cause of the equipment. By analyzing the signal of rotating machinery through the fault diagnosis algorithm,

J. Liang · Y. Liao

School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China

W. Li (✉)

Shine-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou 511442, China
e-mail: whlee@scut.edu.cn

Guangdong Artificial Intelligent and Digital Economy Laboratory (Guangzhou), Guangzhou 510335, China

the potential fault can be recognized in time, so that the management personnel can timely maintain the equipment that may fail, so as to ensure the reliable operation of the equipment and achieve greater economic benefits. Therefore, the fault diagnosis of rotating machinery has great significance.

Nowadays, many researches about deep learning for fault diagnosis are carried out. Zhang et al. [1] proposed Deep Convolutional Neural Network with Wide First-layer Kernel (WDCNN), which extracted short-time features through the first layer large convolutional kernel and achieved a high accuracy on the CWRU bearing data set. Guo et al. [2] proposed an adaptive deep convolutional neural network for fault diagnosis of rotating machinery. Fault types were classified through the first convolutional neural network, and then the fault degree was classified according to different fault types. Jing et al. [3] proposed to use one-dimensional convolutional neural network to classify the frequency spectrum of vibration signals, so as to realize fault diagnosis of gearbox. Wang et al. [4] proposed a wavelet-based convolution neural network for machinery fault diagnosis, the vibration signal is processed by wavelet transform into a multi-scale spectrogram image as the input of convolutional neural network, and ReLU activation function and dropout technique had been used to enhance the network performance.

Deep learning can automatically extract the features of the signal, avoiding the artificial feature engineering. A good hand-designed neural network can be used for fault diagnosis and achieve high accuracy. For specific fault diagnosis task such as fault diagnosis for bearing or fault diagnosis for bevel gear, however, the design process of the neural network structure is complex, and heavily relies on expert knowledge. In addition, there are many hyperparameters in the model that need to be set. To obtain an appropriate hyperparameters set needs constant trial and error, which takes a lot of time. What's more, a network that is efficient for one fault diagnosis task might perform less efficient or even bad for another fault diagnosis task due to the difference between the two data sets. Therefore, a method that can automatically optimize network structure and hyperparameters based on different fault diagnosis task is necessary.

Neural architecture search (NAS) can automatically search appropriate neural network structures by defining search spaces, search strategies and evaluation criteria, which can achieve comparable performance with those of manually designed network by experts. Common NAS methods such as reinforcement learning (RL) [5] and evolutionary algorithms (EA) [6] search in discrete and non-differentiable search space, which cost a large amount of computing resources. Unlike those non-differentiable NAS methods based on RL and EA, DARTS is a gradient based NAS method, which is more efficient. In the field of fault diagnosis, NAS has also been applied. Wang et al. [7] proposed a NAS method based on reinforcement learning for fault diagnosis of rolling bearings. Zhou et al. [8] proposed a gradient based method called differentiable architecture search (DARTS) for aeroengine fault diagnosis. Zhang et al. [9] proposed a DARTS-based method considering searching cost, performance and complexity at the same time, which makes NAS task for fault diagnosis more efficient. Although these automatically searched neural networks can get

good results, their hyperparameters are still manually determined, which may not be the optimal one.

Aiming at the problem of hyperparameter adjustment in neural network, hyperparameter optimization (HPO) can automatically search for better hyperparameter and save time for the designer. The common HPO methods include grid search, random search [10] and Bayesian optimization. Grid search traverses all possibilities, which is easy to use, but consumes a lot of computing resources. Random search is simple and efficient, but not as good as Bayesian optimization methods. Tree-Structured Parzen Estimators (TPE) [11] is a commonly used HPO method based on Bayesian optimization, which is widely used in the tuning of neural networks. TPE chooses hyperparameters based on all the previous results, and skips the choices that are likely to perform bad, which is more efficient than grid search and random search. Yoo et al. [12] proposed to use TPE to optimize the depth of VGG network and obtain better results than Grid Search. Bergestra et al. [13] used TPE to optimize the neural network for image classification tasks, which outperformed random search in both optimization efficiency and accuracy. In the field of fault diagnosis, Nguyen et al. [14] proposed a Long Short-Term Memory (LSTM) neural network for nuclear power plant fault prognostics and used TPE for hyperparameters optimization. However, generally, HPO can only optimize the hyperparameters when the neural network architecture is fixed, while the hyperparameters that determine the structure of the neural network can hardly be optimized due to too complex computation.

In summary, the current fault diagnosis methods present the following problems:

1. In order to get good results for specific fault diagnosis tasks, specific neural network architectures need to be designed manually, which requires expert knowledge and takes a lot of time.
2. A large number of hyperparameters need to be adjusted in the neural network, and different hyperparameters sets will affect the testing results. Manually adjusting hyperparameters also heavily relies on expert knowledge and is time consuming.

This paper proposes a method combining NAS and HPO to realize automated neural architecture search and automated hyperparameter optimization for fault diagnosis. In this method, DARTS is used to search specific neural network structures for fault diagnosis tasks, while TPE is used to optimize the hyperparameters of the network searched by DARTS, so as to obtain optimal performance for specific fault diagnosis task. To be specific, the contributions of this paper are summarized as follows.

1. This paper proposes a Differentiable Architecture Searched Network with Tree-Structured Parzen Estimator (DASNT) to automatically generate optimal network structure and hyperparameters for specific fault diagnosis task, which is more automatic and less dependent on expert knowledge while compared to other manual methods.
2. This paper verifies that the combination of DARTS and TPE can achieve state of the art (SOTA) fault diagnosis accuracy. The proposed method performs well on

CWRU bearing fault diagnosis and transmission fault diagnosis tasks, especially under a noisy environment.

The rest of this paper is organized as follows. Section 2 introduces the proposed methods. In Sect. 3, comparison experiments are conducted to verify the performance of the proposed method with CWRU data set and transmission fault data set. Section 4 is the summary of the whole paper.

2 Methodology

Differentiable Architecture Searched Network with Tree-Structured Parzen Estimator (DASNT) will be introduced in detail in this section. The proposed DASNT for rotating machinery is described in Fig. 1. It can be divided into four steps: (1) Data acquisition and preprocessing (2) using DARTS to search an optimal neural network for target fault diagnosis task; (3) using TPE to optimize the hyperparameters of the DARTS network, so as to better train the network. (4) Finally, a trained network is obtained for fault diagnosis task. In this section, we will focus on DARTS and TPE algorithms.

2.1 DARTS

DARTS is a gradient based NAS method. It maps the search space to a continuous space with Softmax’s approach, and then the network architecture can be optimized by gradient descent, which is more efficient than the non-differentiable approach.

Figure 2 is the schematic illustration of DARTS algorithm. The objective of DARTS is to search operations in the search space to form a cell, and to form a larger

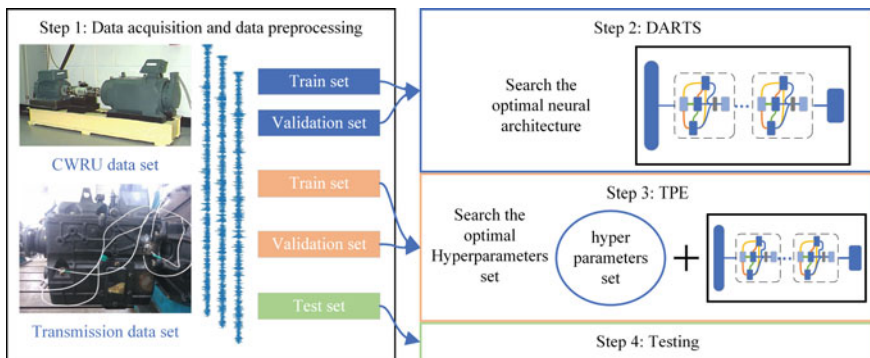


Fig. 1 An overview of DASNT

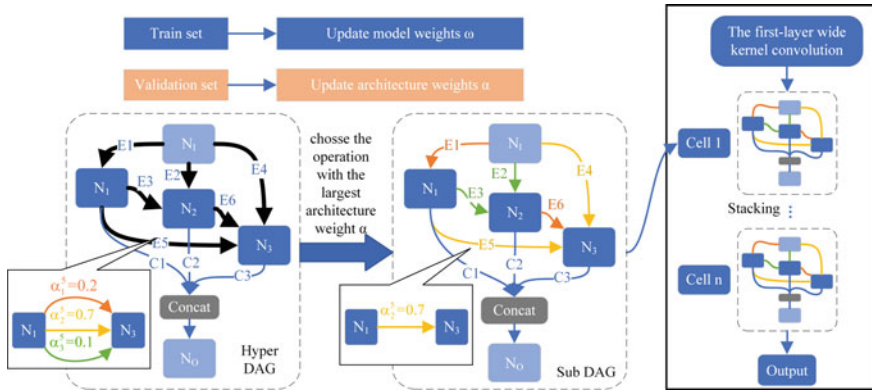


Fig. 2 Illustration of DARTS algorithm

neural network by stacking cells. A cell consists of nodes, and the node represents the feature map. Each node is connected to all its predecessors, forming a directed acyclic graph (DAG). So the problem can be thought of finding a subgraph from a hypergraph. The connecting edge between each node in the hypergraph contains all considered operations, and the operations with the largest architecture weights α are chosen to constitute the cell of the subgraph. For instance, as shown in Fig. 2, the cell consists of 5 nodes and 9 edges, where N_1 and N_0 are input and output nodes, N_1, N_2, N_3 are the intermediate nodes, $E_1, E_2, E_3, E_4, E_5, E_6$ are the connecting edges which contains all the candidate operations, C_1, C_2, C_3 are connecting edges, output node N_0 is depth-wise concatenation of all the intermediate nodes. There are 3 candidate operations in the search space, represented in red, yellow and green. After the update, the architecture weights of the three operations between N_1 and N_3 are 0.2, 0.7 and 0.1, and the operation with the largest architecture weight is remained, which is $\alpha_2^5=0.7$. Each edge selects the operations with the largest architecture weights to form the cell. There are two types of cells, the normal cell and the reduction cell. Normal cell doesn't change the dimension of the input, while the reduction cell reduces the feature size by half and doubles the number of channels. By adding a convolutional layer in front of the first cell, stacking the cells and then follow with a classifier, the searched neural network is formed.

The forward calculation of the hyper DAG can be regarded as the inner product between the architecture weights matrix and the search space:

$$f_{cell}(x) = \underbrace{\begin{bmatrix} \alpha_1^1 & \cdots & \alpha_M^1 \\ \vdots & \ddots & \vdots \\ \alpha_1^{\frac{N(N+1)}{2}} & \cdots & \alpha_M^{\frac{N(N+1)}{2}} \end{bmatrix}}_{\text{architecture weights matrix}} \otimes \underbrace{\begin{bmatrix} OP_1(x) \\ \vdots \\ OP_M(x) \end{bmatrix}}_{\text{Search space}} \quad (1)$$

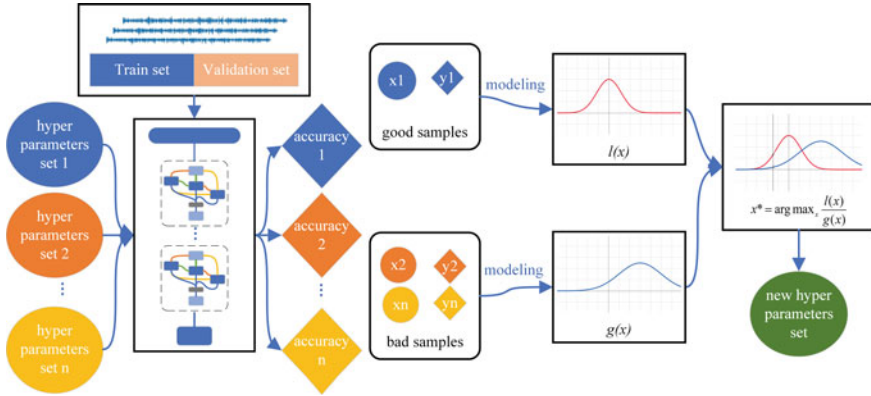


Fig. 3 Illustration of TPE algorithm

where OP_i represents the i -th operation of the search space. α_i^j represents the architecture weight of the i -th operation on the j -th edge. N represents the number of nodes. The i -th node has i edges connecting with its i predecessors, so a cell with N nodes has $\frac{N(N+1)}{2}$ edges in total.

The optimizing process can be regarded as a bilevel optimization problem, and the equation is shown as follows:

$$\begin{aligned} \alpha^* &= \arg \min_{\alpha} L_{val}(\omega^*(\alpha), \alpha) \\ s.t. \omega^*(\alpha) &= \arg \min_{\omega} L_{train}(\omega, \alpha) \end{aligned} \tag{2}$$

where ω^* is the model weights under fixed α , which is optimized in the train set, and α^* is the optimized architecture weights, which is optimized in the validation set. Both ω and α are optimized with gradient decent alternately, and the best neural architecture is searched after certain times of iterations.

In the experiment of this paper, there are 8 candidate operations in the search space of DARTS, which is shown in Table 1. Each cell has 4 intermediate nodes.

2.2 TPE

In DASNT, TPE is applied to search the optimal hyperparameters set, so as to further improve the network accuracy. Figure 3 is the schematic illustration of TPE algorithm. A series of hyperparameters sets are used to train the DARTS network, and the corresponding accuracy on validation set is obtained as the evaluation value. Hyperparameters sets are divided into good and bad samples based on the evaluation values. After that, the TPE defines the models of likelihood probability for each of the two groups as follows:

Table 1 Candidate operations in the search space of DARTS

Operation	Description
None	Zero operation
Skip_connect	Identity mapping
Max_pool	Max pooling
Avg_pool	Average pooling
Sep_conv_3	Separable convolutions, kernel size = 3
Sep_conv_5	Separable convolutions, kernel size = 5
Dil_conv_3	Dilated separable convolutions, kernel size = 3
Dil_conv_5	Dilated separable convolutions, kernel size = 5

$$p(x|y) = \begin{cases} l(x), & y > y^* \\ g(x), & y \leq y^* \end{cases} \tag{3}$$

where y^* is the threshold value of accuracy, which is usually set to 75% of the historical highest accuracy of validation set. $l(x)$ means the probabilities that the hyperparameter x will achieve a accuracy higher than y^* , while $g(x)$ means the probabilities that the hyperparameter x will achieve a accuracy lower than y^* . Then the new hyperparameters set is chosen based on the expected improvement (EI) formula, which can be defined as follows:

$$EI(x) = \frac{l(x)}{g(x)} \tag{4}$$

At each iteration, the objective is to find the hyperparameter set x^* that maximizes the EI value:

$$x^* = \arg \max_x \frac{l(x)}{g(x)} \tag{5}$$

The formula indicates to find x^* which has a higher probability in the good group and less likely in the bad group. After the new hyperparameters set is found, it is regarded as historical experience for the later iterations and to update $l(x)$ and $g(x)$. The process will end in certain times of iterations, and the hyperparameters set with the best historical accuracy will be chosen as the TPE search results.

3 Experimental Study

CWRU bearing fault diagnosis and transmission fault diagnosis cases are utilized for the verification of the proposed method. The hand-designed network WDCNN, which is a state of the art (SOTA) method for fault diagnosis, is carried out as the baseline. WDCNN with TPE (WDCNN-TPE) and DARTS [8] are also carried out

Table 2 Hyperparameters search space

Hyper-parameter	Optional value	Description
Batch_size	16/32/64/128/256/512	Batch size
First_kernel size	16/24/32/40/48/56/64/72/80/88/96/106/112/120/128	The kernel size of the first convolutional layer
Resize	4/8/16/32	The ratio of the input to output dimensions of the first convolutional layer
Normalization	True/False	Whether to normalize the data or not
Enc_step	0/14/28/56/92	The step size of data enhancement
Lr	0.1/0.01/0.001/0.0001	Learning rate
Lr_scheduler	True/False	Whether to use cosine annealing learning rate scheduler or not
Optimizer	SGD/Adam	Stochastic gradient descent optimizer and Adam optimizer
Epochs	10/15/20/25	Training epochs

as comparison experiments. All the experiments are conducted on the GTX 960 graphics card.

In the DARTS phase of the proposed method, the parameters of the first convolutional layer and the number of cells in DARTS network are pre-determined hyperparameters. The number of cells is 2, the first-layer convolution kernel size is 64, and the stride is 8 as the parameters of DARTS network. DARTS will be run for 50 epochs.

In the TPE phase of the proposed method, there are a number of hyperparameters to be optimized, and each hyperparameter has a number of possible options. The details of the hyperparameters are shown in Table 2. There are 230,400 combinations of hyperparameters and we will run TPE for 100 iterations.

3.1 Data Description

The first case is CWRU bearing fault diagnosis. The signal type is acceleration signal and the sample length is 2048. According to different damage degree of inner ring,

Table 3 The CWRU data set

Damage position	Inner ring			Ball			Outer ring			None
Damage diameter ($\times 10^{-3}$ inch)	7	14	21	7	14	21	7	14	21	0
Load	0/1/2/3HP									
Label	0	1	2	3	4	5	6	7	8	9

Table 4 The transmission data set

State of bearing	Inner ring 0.2 mm				Inner ring 2 mm		Inner ring normal			
State of gear	BT	M	MD	N	MD	BT	BT	M	MD	N
Rotating speed	500/750/1000/1250 rpm									
Label	0	1	2	3	4	5	6	7	8	9

ball and outer ring, there are ten types of faults. In order to enable the neural network to learn the characteristics of the same damage state under different loads at the same time, each state was operated at four working conditions. There are 1000 pieces of data for each label, and the train set, validation set and test set are divided according to the ratio of 4:4:2 in the DARTS phase and 6:2:2 in the TPE phase. Details are shown in Table 3.

The second case is transmission fault diagnosis. The signal type is acceleration signal and the sample length is 2048. Broken teeth (BT), mild wear (M), moderate wear (MD), and normal (N) are the four states of gears. The inner ring has three health states. Therefore, there are totally 10 different fault types, which is composed of different bearing states and gear states. Each state was operated at four working conditions. There are 1000 pieces of data for each label, and the train set, validation set and test set are divided according to the ratio of 4:4:2 in the DARTS phase and 6:2:2 in the TPE phase. Details are shown in Table 4.

3.2 Results and Discussion

The experimental results of DASNT on the CWRU data set and transmission data set are shown in Table 5 and Figs. 4, 5. Experiments are conducted under different signal to noise ratio (SNR). All the experiments results are average by five trials to reduce randomness. The following observations and conclusions can be drawn:

1. In all fault diagnosis tasks, the accuracy of WDCNN-TPE is higher than that of WDCNN, DASNT is higher than that of DARTS, which indicates that the TPE is useful for finding the optimal hyperparameters set so as to achieve a higher accuracy.
2. In all fault diagnosis tasks, the accuracy of DARTS is higher than that of WDCNN, which indicates that DARTS performs better than WDCNN.

Table 5 The experimental results

Tasks	WDCNN	WDCNN-TPE	DARTS	DASNT
CWRU -10 SNR	49.55 ± 0.63	50.74 ± 0.70	53.57 ± 0.61	56.34 ± 0.70
CWRU -5 SNR	72.23 ± 0.80	74.43 ± 1.29	75.56 ± 0.97	79.49 ± 0.41
CWRU 0 SNR	91.44 ± 0.66	92.84 ± 0.70	94.47 ± 0.30	95.33 ± 0.57
CWRU 5 SNR	97.88 ± 0.30	98.88 ± 0.32	99.07 ± 0.35	99.62 ± 0.10
CWRU 10 SNR	99.28 ± 0.10	99.57 ± 0.09	99.52 ± 0.13	99.88 ± 0.07
CWRU noise free	99.43 ± 0.08	99.92 ± 0.07	99.66 ± 0.13	99.98 ± 0.04
Transmission -4 SNR	52.09 ± 1.86	52.33 ± 2.77	65.65 ± 0.64	67.91 ± 1.12
Transmission -2 SNR	63.54 ± 1.89	63.71 ± 0.96	72.30 ± 1.24	79.67 ± 0.18
Transmission 0 SNR	70.59 ± 3.21	73.72 ± 3.05	81.03 ± 0.55	84.76 ± 0.60
Transmission 2 SNR	80.13 ± 0.44	82.51 ± 0.88	87.11 ± 0.78	88.67 ± 0.59
Transmission 4 SNR	85.13 ± 1.64	90.07 ± 0.77	89.68 ± 0.83	94.31 ± 0.43
Transmission noise free	95.94 ± 1.20	99.65 ± 0.05	97.97 ± 0.55	99.87 ± 0.12

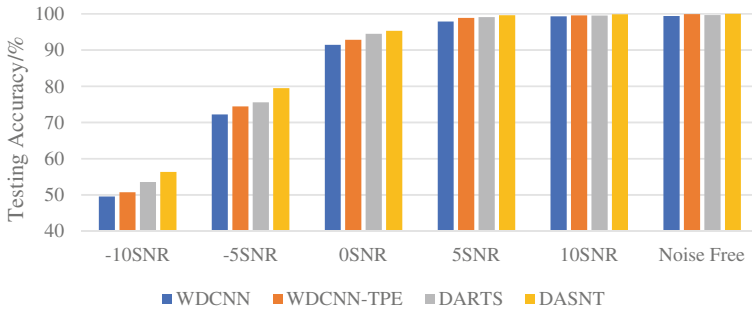


Fig. 4 CWRU bearing fault diagnosis results

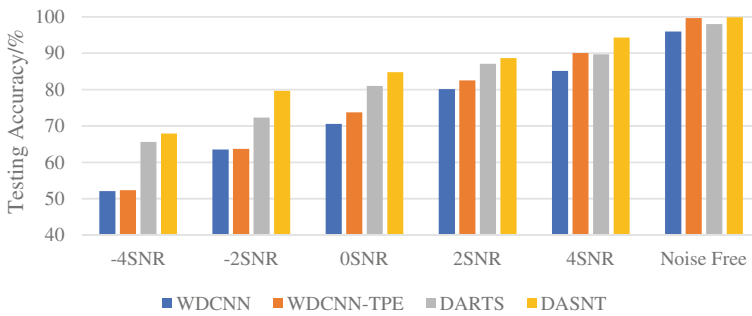


Fig. 5 Transmission fault diagnosis results

3. On both CWRU and transmission fault diagnosis cases, DASNT outperforms all other methods under different SNR level. Especially in noisy environment, e.g., $-10/-5$ SNR on CWRU data set, $-4/-2$ SNR on transmission data set, the accuracy of DASNT is significantly higher than that of WDCNN-TPE. This result shows that DASNT has better noise resistance than WDCNN-TPE.
4. In CWRU noise free task, the accuracy of the four methods is above 99%, which proves that the fault diagnosis task of CWRU data set is relatively easy. In transmission noise free task, the accuracy of WDCNN is significantly lower than other methods, because WDCNN is customized for CWRU bearing fault diagnosis case and cannot get good results if it's directly used in other fault diagnosis task without any modification. DASNT doesn't have this drawback because it can automatically search the optimal network structure and hyperparameters for every specific fault diagnosis task.
5. By comparing CWRU 0 SNR task with transmission 0 SNR task, it can be found that in CWRU 0 SNR task, the accuracy of WDCNN-TPE is slightly lower than that of DASNT (92.84%–95.33%), while in transmission 0 SNR task, the accuracy of WDCNN-TPE is significantly lower than that of DASNT (73.72%–84.76%). This result indicates that WDCNN is customized for CWRU bearing case, even the hyperparameters have been optimized, it cannot reach a good result for other fault diagnosis task because it's network architecture is not the optimal one.

In conclusion, on both CWRU and transmission fault diagnosis cases, DASNT outperforms the hand-designed baseline WDCNN and comparison methods such as WDCNN-TPE and DARTS, which proves that DASNT can effectively search the optimal neural network architecture and the corresponding hyperparameters set for specific fault diagnosis task. Moreover, the experiments shows that DASNT has strong generalization ability and noise resistance for different fault diagnosis tasks.

4 Conclusion

This paper proposed a differentiable architecture searched network with TPE (DASNT) for rotating machinery fault diagnosis, which combines the advantages of DARTS and TPE to search network structure and hyperparameters automatically. On both CWRU and transmission fault diagnosis cases, DASNT outperforms other SOTA methods including hand-designed WDCNN and WDCNN-TPE, and a gradient based NAS method called DARTS, which proves that DASNT is more effective than manual network designing and manual hyperparameters tuning. Furthermore, for different fault diagnosis tasks, DASNT performs strong generalization ability and noise resistance.

Acknowledgements This work was supported in part by Key-Area Research and Development Program of Guangdong Province under Grant 2021B0101200004, in part by the National Natural Science Foundation of China under Grant 51875208, in part by the National Key Research and

Development Program of China under Grant 2018YFB1702400, and in part by the Key-Area and Development Program of Guangdong Province under Grant 2019B010154002.

References

1. Zhang, W., Peng, G., Li, C., et al.: A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **17**(2), 425 (2017)
2. Guo, X., Chen, L., Shen, C.: Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* **93**, 490–502 (2016)
3. Jing, L., Zhao, M., Li, P., et al.: A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* **111**, 1–10 (2017)
4. Wang, J., Zhuang, J., Duan, L., et al.: A multi-scale convolution neural network for featureless fault diagnosis. In: 2016 International Symposium on Flexible Automation (ISFA). IEEE, pp. 65–70 (2016)
5. Zoph, B., Le, Q.V.: Neural architecture search with reinforcement learning. arXiv preprint (2016). [arXiv:1611.01578](https://arxiv.org/abs/1611.01578)
6. Real, E., Moore, S., Selle, A., et al.: Large-scale evolution of image classifiers. In: International Conference on Machine Learning. PMLR, pp. 2902–2911 (2017)
7. Wang, R., Jiang, H., Li, X., et al.: A reinforcement neural architecture search method for rolling bearing fault diagnosis. *Measurement* **154**, 107417 (2020)
8. Zhou, Z., Li, T., Zhao, Z., et al.: Differentiable architecture search for aeroengine bevel gear fault diagnosis. In: 2020 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD). IEEE, pp. 270–274 (2020)
9. Zhang, K., Chen, J., He, S., et al.: Differentiable neural architecture search augmented with pruning and multi-objective optimization for time-efficient intelligent fault diagnosis of machinery. *Mech. Syst. Signal Process.* **158**, 107773 (2021)
10. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **13**(2), (2012)
11. Bergstra, J., Bardenet, R., Bengio, Y., et al.: Algorithms for hyper-parameter optimization. *Adv. Neural Inf. Process. Syst.* **24** (2011)
12. Yoo, S., Haider, M.A., Khalvati, F.: Estimating optimal depth of vgg net with tree-structured parzen estimators. *J. Comput. Vis. Imaging Syst.* **3**(1), (2017)
13. Bergstra, J., Yamins, D., Cox, D.: Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: International Conference on Machine Learning. PMLR, pp. 115–123 (2013)
14. Nguyen, H.P., Liu, J., Zio, E.: A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by Tree-structured Parzen Estimator and applied to time-series data of NPP steam generators. *Appl. Soft Comput.* **89**, 106116 (2020)

A Review of Fault Diagnosis Methods for Marine Electric Propulsion System



Dongqin Li, Rongfeng Deng, Zhexiang Zou, Baoshan Huang,
and Fengshou Gu

Abstract With the rapid development of power electronics technology and the proposal of intelligent ships, electric propulsion systems on ships are becoming more and more widespread. As the power source for ship navigation, timely and accurate diagnosis and prediction of faults of electric propulsion system play a vital role in the operation safety of ships. This paper summarises the common faults of electric propulsion systems, reviews the latest developments and applications of fault diagnosis techniques based on fault signal analysis in electric propulsion system fault diagnosis, and discusses the advantages and disadvantages of typical methods in the light of the latest literature and current research problems. The paper concludes by proposing future trends in fault diagnosis and prediction for ship electric propulsion systems.

Keywords Marine electric propulsion system · Fault diagnosis · Signal analysis

1 Introduction

1.1 Marine Electric Propulsion System

Marine electric propulsion systems have the advantages of energy-saving, environmental protection, mobility and flexibility, less space required and less propulsion noise and vibration than traditional ones using internal combustion engines. The development of intelligent ships has also led to an expansion of their applications. An

D. Li · R. Deng · Z. Zou (✉) · B. Huang
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, China
e-mail: zhexiang.zou@hud.ac.uk

D. Li · R. Deng · Z. Zou · F. Gu
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK

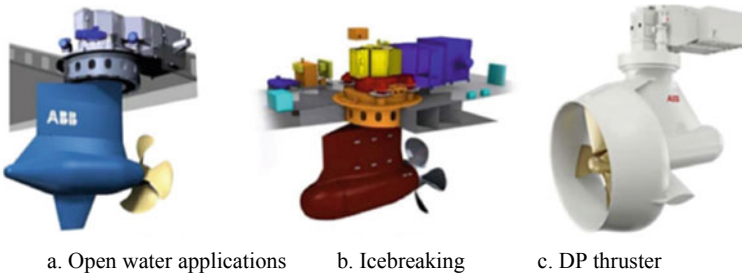


Fig. 1 The common pod propulsion systems [2]

increasing number of manufacturers are also opting for electric propulsion systems driven by electric motors when building ship propulsion systems [1]. Figure 1 shows several common types of pod propulsion systems [2].

The electric propulsion system consists mainly of motor, drive system and propeller. The two ends of the driveline are connected to the motor and the propeller, respectively. When the motor is running, the motor's power is transmitted to the propeller through the drive system, which drives the propeller to rotate and thus propel the ship forward [3]. Figure 2 shows the structure of a ship's electric propulsion system.

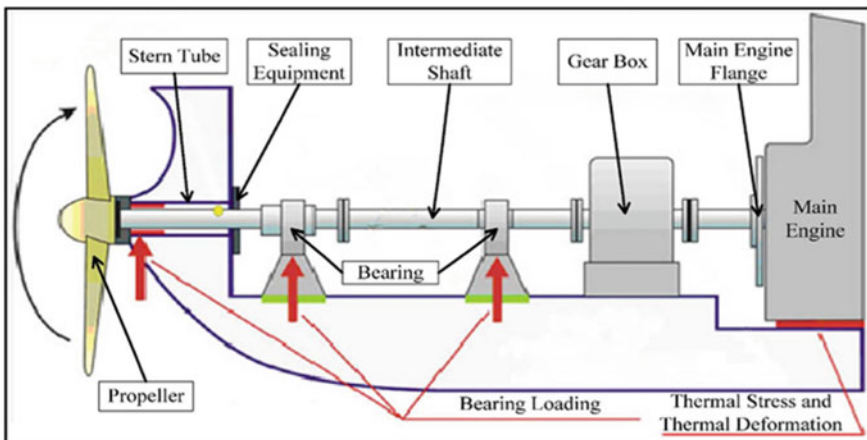


Fig. 2 Marine propulsion system structure diagram [4]

1.2 Common Types of Failures in Electric Propulsion System

During a ship's voyage, any breakdown can cause a major safety incident and seriously threaten the lives of all personnel on board [5, 6]. According to statistics, the electric propulsion system is one part of the ship that is most likely to fail [7]. Therefore, the stability of the electric propulsion system is very crucial to the ship. It is a vital task to find out the faults of the electric propulsion system in time and accurately to avoid accidents.

1.2.1 Motor Faults

The motors commonly used in the propulsion system include induction motors, synchronous motors, permanent magnet motors, etc. [8]. In recent years, AC motors are far more widely used than DC motors in electric propulsion systems [9].

1. Induction Motor (IM) fault

Compared to other motors, IMs have low cost, simple construction, reliability and high efficiency. The faults of AC induction motors consist of two main categories: electrical faults and mechanical faults. Electrical faults mainly include broken rotor bars, stator winding faults, etc. Mechanical faults mostly include bearing faults, rotor shaft eccentricity, etc. [10–13].

2. Permanent magnet synchronous motor (PM) fault

PMs have the advantages of low weight, low noise, high efficiency and low heat generation. It outperforms other types of motors in applications where high-speed operation and precise torque control are required. Faults in PM synchronous motors consist of three main components: electrical faults, mechanical faults and demagnetisation faults [14]. Electrical faults are mainly stator winding faults and stator winding failures. Mechanical faults are mainly bearing faults and eccentric faults [15].

1.2.2 Driveline Failure

The drivelines is mainly used to transmit power and consists of a gearbox or drive shaft system.

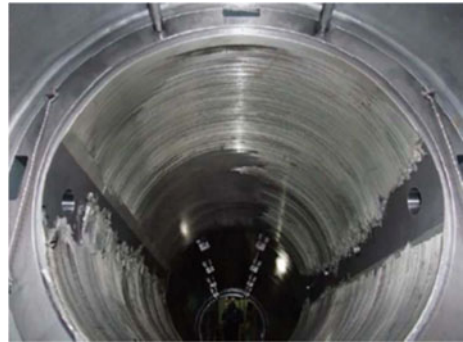
1. Gearbox failure

The gearbox is mainly used to change the speed and transmit power and is mainly composed of gears, shafts, bearings and other parts. Common faults include gear wear, gear broken teeth, shaft and gear misalignment, bearing failure (Fig. 3) and so on [16, 17].



Fig. 3 Fatigue fracture of reducer shaft [18]

Fig. 4 Stern tube bearing wear [21]



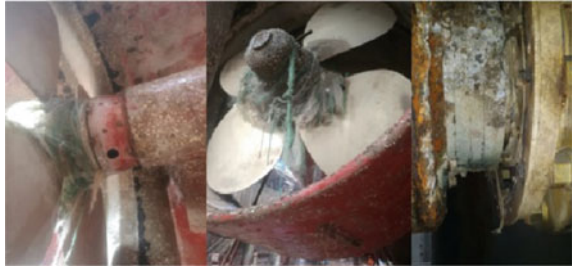
2. *Driveshaft system failure*

Common faults of the transmission shaft system include stern bearing failure (Fig. 4), stern shaft oil leakage, stern shaft high temperature, propulsion shaft fatigue, etc. [19, 20].

1.2.3 Propeller Failure

Ships may run aground or strike hard objects while underway, and propellers may become entangled in underwater ropes, fishing nets and other debris. These may lead to propeller blade curling, blade deformation, and even blade breakage [22]. Also, as the propeller is always underwater, the propeller is prone to scaling, causing the rotor to run unbalanced. Propellers can also generate new cyclically varying centrifugal and hydrodynamic forces during operation due to failure, which can further lead to shaft system and main engine failure. Figure 5 shows the propeller and tail shaft seal section entangled in the fishing net.

Fig. 5 The propeller and stern seal parts are entangled by the fishing nets [23]



2 Motor Fault Diagnosis Methods

2.1 Analysis Based on Vibration Signal

All operating equipment, whether or not there is a fault, will produce vibration signals. We can determine whether the equipment is faulty by analyzing the difference between the vibration signal with and without fault. Among the diagnosis methods of mechanical faults, the diagnosis method based on vibration signal is generally considered to be the most effective. The vibration signal can be obtained through the speed and accelerometer, and then the vibration amount is transformed into electrical signals through the supporting circuit, which is also the most commonly used vibration signal measurement method. When the motor is running normally, a small vibration signal will be generated, and when the motor fails, a larger vibration pulse signal will be generated. Therefore, the vibration signal can be used to identify the motor failure. Zhen et al. [24] used weighted average ensemble empirical mode decomposition (WAEEMD) to construct the eigenmode functions derived from the vibration signal decomposition into new signals. Then modulation signal bispectrum (MSB) was used to extract the fault features of the new signals, thus realising the fault diagnosis of rolling bearings. Medoued et al. [25] used the diagnosis method based on axial vibration analytical signal (AVAS) to complete the fault diagnosis and analysis of the asynchronous motor. Considering the difficulty of fault data acquisition during the operation of rotating machinery, Zhao et al. [26] proposed an intelligent fault diagnosis method for rotating machinery based on an unsupervised deep learning network. This method automatically extracts fault information from the original vibration signals and realised fault diagnosis for rolling motor bearings. Wang et al. [27] used the end-to-end fault diagnosis model based on a convolutional neural network (CNN) to complete the fault detection of bearings and gearboxes. This method does not need to extract fault features manually and has a high diagnosis rate.

2.2 Analysis Based on Current Signal

Motor Current Signature Analysis (MCSA) method is the lowest cost, non-invasive electrical fault detection technology. The main principle is that when the motor gear is worn or the tooth is broken, the rotor is broken, or the shaft is eccentric, the system will generate additional fluctuating torque. The motor will generate a corresponding electromagnetic torque to balance this torque, resulting in a non-linear current signal in the stator current [28]. The feature extraction methods of motor current signal mainly include Fourier transform and its special transform, wavelet transform and its special transform, Hilbert spectrum analysis, modulation signal bispectrum (MSB) and so on [29].

Dehina et al. [30] carried out Fast Fourier transform (FFT) on the motor's stator current and vibration signals. They analysed the spectrum information of the two signals under the static eccentricity fault of the motor, respectively. And the two kinds of signal analysis methods are compared. The conclusion is that compared with the motor vibration signal, the analysis of the motor stator current is easier to reflect the fault. Glowacz et al. [31] performed continuous wavelet transform (CWT) on the current signal of the squirrel cage asynchronous motor. By analysing the frequency spectrum of the current signal, the fault characteristics of the broken bar of the motor are extracted, and the broken bar fault of the motor under no-load operation is detected. Rangel-Magdaleno et al. [32] proposed a rotor initial broken bar detection method based on Hilbert spectrum analysis. Through the current signal acquisition, empirical mode decomposition, Hilbert spectrum analysis and other steps, the detection of motor broken bar in the early stage were realised. Systems et al. [33] compared the performance of the conventional bispectrum (CB) method and the modulation signal bispectrum (MSB) method in the analysis of induction motor rotor faults (including rotor bar breaking, gearbox wear, etc.). Experiments show that MSB has a significantly better diagnostic effect than the CB method due to its unique non-linear modulation detection and random noise suppression capabilities.

2.3 Analysis Based on Sound Signal

Suppose the vibration noise of the permanent magnet synchronous motor is too large. In that case, it may affect other modules and even make the hull unable to operate normally by analysing the frequency spectrum characteristics of the vibration noise by detecting whether the permanent magnet synchronous motor is operating normally [34]. Sangeetha et al. [35] used the Rational-dilation Wavelet Transform (RADWT) method to extract fault features from the acoustic signals generated by three-phase asynchronous motors. They then used the multiple regression method to estimate the feature signals' torque and feature signals' torque and finally accurately judge the fault location of the asynchronous motors. Delgado-Arredondo et al. [36] proposed A fault diagnosis method based on acoustic and vibration signals. In the experiment,

the sound and vibration signals were extracted and analysed respectively to diagnose bearing fault, rotor broken bar, and mechanical imbalance.

3 Driveline Fault Diagnosis Methods

3.1 Analysis Based on Vibration Signal

Currently, feature extraction methods used for vibration analysis mainly include time-domain analysis (such as reference [37]), frequency-domain analysis (such as reference [38]), time–frequency analysis (such as reference [39]) and so on.

The gearbox vibration of the Marine propulsion system is a non-linear mixed vibration composed of different vibration sources. It is challenging to identify all the fault types effectively using traditional vibration analysis methods. Li et al. [40] proposed a non-linear blind source separation (BSS) analysis technique for gearbox hybrid fault detection under variable speed conditions. Using this method, the decoupled fault components are obtained, and then the indexes of the fault components are compared with the benchmark to realise the identification of mixed faults. Experimental results show that this method can detect the mixed faults of the Marine propulsion gearbox with large speed variation. Similarly, for the fault diagnosis of gears and bearings in planetary gearboxes, Xiao et al. [41] proposed a fault detection method based on vibrational resonance (VR) to realise the diagnosis of weak faults by resonating weak fault signals.

In actual navigation, the vibration of the ship propulsion shaft is coupled by transverse vibration, longitudinal vibration and torsional vibration. When there is an unbalanced force on the propulsion shaft, strong coupling vibration is prone to occur, which will cause the fatigue fracture of the shaft system [42, 43]. The dual-frequency laser Doppler torsional vibration measuring instrument can be used to obtain the torsional vibration information of the ship's propulsion system shaft system. The acquisition of vibration signals enables the acquisition of the shaft system's rotational and longitudinal vibration information characteristics [44]. Han and Lee [38] designed a test device that can simulate the high torsional vibration of the ship's propulsion shaft. And through the FFT transformation of the vibration torque and related signals, it is verified that the fracture of the propulsion shaft of a ship is caused by the high torsional vibration caused by the transverse-torsional coupled vibration. Li et al. [39] used the wavelet transform algorithm to carry out time–frequency conversion for the collected fault vibration signals and then input the converted data into the deep convolution neural network (DCNN) model for fault diagnosis of bearing running state. Experiments verify the effectiveness of the method. Grządziela et al. [37] analysed the vibration signal of the propeller system of a minesweeper in the time domain, and through the analysis of a series of signal characteristics, developed and verified the method of the non-coaxial detection of the main engine shafting.

3.2 Analysis Based on Oil Conditions

The mechanism of the oil analysis method [45–48] is that when two surfaces in relative motion are in abnormal wear condition, the wear rate will be relatively high, and corresponding abrasive particles will be generated according to different wear failure forms. By judging the concentration and composition of the abrasive particles in the oil, the wear of the equipment can be analysed, or whether the stern shaft oil leakage.

Li et al. [49] used the spectral element analysis method to analyse the content of related elements in oil; used infrared spectrum analysis method to compare new oil, in-use oil and biological oil; and used ferrospectrum wear particle analysis method to analyse the size of wear particles in the oil. Finally, the cause of abnormal heat in the stern tube of a ship is found out: the oil purity is not enough to lead to the deterioration of lubricating oil performance, making the bearing bush serious wear, and then cause abnormal heat in the stern tube. Zhao [50] realised the monitoring of oil lubrication state, the supervision of the Marine gear box's running state and wear condition, and the detection of the gear box's waterproof sealing condition by analysing oil. It also provides a way to change oil scientifically.

3.3 Analysis Based on Acoustic Emission Signal

The propeller shaft of a ship may crack due to intense vibration during its rotation. In response to this situation, acoustic emission technology can diagnose faults on the drive shaft. Arifianto et al. [51] used acoustic emission technology [52–55] to collect the mixed sound signal when rotating the drive shaft. By performing time–frequency blind source separation of the mixed sound signal, the crack position of the ship's propulsion system transmission shaft is judged. Elasha et al. [56] used acoustic emission technology to detect bearing defects in a planetary gearbox. Also, they found through experiments that acoustic emission can identify bearing defects faster than vibration analysis.

4 Propeller Fault Diagnosis Methods

4.1 Analysis Based on Hydrodynamic Performance

When a propeller blade fails, the corresponding position of the blade changes the hydrodynamic performance [57–60]. Hu et al. [61] used hydrodynamic analysis to obtain the pulse pressure maps of the monitoring points under different propeller failures. They used FFT to extract the characteristic parameters for propeller failure

detection. Wang et al. used CFD to calculate the pressure values of each monitoring point on the propeller surface to achieve fault diagnosis and prediction of the propeller.

4.2 Analysis Based on Vibration Signal

As propellers work underwater in a harsh working environment, which makes detection very inconvenient, it is often necessary to borrow indirect detection methods to achieve this. When a propeller blade is damaged, the fault signal is displayed in a vibration signal through the drive shaft. By analysing the time domain and spectrograms of the analogue signals generated by the rotational vibrations of the shaft system, the fault parameters in the case of a broken propeller blade can be obtained [62]. Ngui et al. [63] used continuous wavelet transform (CWT) to extract features of turbine vibration signals and then input the extracted feature signals into an artificial neural network (ANN) model for blade fault diagnosis. Experimental results show that the method has a high accuracy rate for fault diagnosis of turbomachinery blades.

4.3 Analysis Based on Current Signal

When a propeller fails, it causes a change in torque, which causes a change in motor flux. The current signal can reflect these changes. Huang et al. [64] carried out a spectrum analysis of the propulsion motor stator current signal through simulation experiments to diagnose propeller fault. The dynamic blades of a tidal stream turbine (TST) are prone to biofouling and, consequently, unbalanced the turbine rotor due to a long time on the seabed. Lofisedi et al. [65] proposed a conventional bispectrum (CB) diagnostic method to analyse and diagnose the situation. Bispectral analysis of a permanent magnet synchronous generator (PMSG) can identify the frequencies associated with biofouling imbalance.

5 Conclusion

This article mainly focuses on introducing the failures of each component of the ship's electric propulsion system and each component's failure signal analysis methods. The advantages and disadvantages of common signal analysis methods are shown in Table 1. Compared with electric motors and drivelines, there are relatively few literatures on the analysis of ship propulsion propeller failures. However, it is very important and necessary to conduct fault analysis and research on the ship's electric propulsion system, and it is also the trend of ship development.

Table 1 Comparison of the advantages and disadvantages of common signal analysis methods

Methods	Advantages	Disadvantages
Vibration signal	The measurement method is relatively simple, and the theory is relatively mature The detection effect of common gear faults is better	For precision mechanical structures, the sensor is not easy to install Sensitivity is affected by the position of the sensor Easy to be interfered by mechanical resonance
Sound signal	Non-invasive detection method, flexible and convenient installation location It is more sensitive to early gear and bearing failure than the vibration analysis method	The diagnostic effect is affected by the sensor installation position and other background noise
Oil analysis	Can be applied to places with the harsh working environment High detection efficiency, high information integration, accurate and reliable diagnosis results, and early fault prediction	Difficult to extract wear particles online for analysis, and offline analysis is relatively lagging No unified monitoring standards
Hydrodynamic performance	Important means of fluid analysis	Susceptible to interference from turbulence and biological fouling There is an error between the theoretical analysis method and the real situation
Current signal	Non-invasive detection method, easy to obtain the current signal and will not interfere with the original system	It strongly depends on the current loop bandwidth, low signal-to-noise ratio, and is not sensitive enough for applications under certain conditions

With the development of artificial intelligence, the fault diagnosis technology of marine electric propulsion system will develop to more valuable fault prediction technology. Considering the complexity of the ship’s electric propulsion system and the need for real-time monitoring of the overall status during navigation, more and more researchers have also proposed many intelligent fault diagnosis methods based on artificial neural networks, support vector machines, expert systems, etc. It is used for fault diagnosis of ship’s electric propulsion system. Due to the complexity of the actual system, it is difficult to collect enough data from the corresponding fault location to train an accurate detection model, and there are relatively few relevant reference materials. Therefore, most intelligent fault diagnosis methods have corresponding deficiencies. The intelligent fault diagnosis methods in many references also use simulation methods to generate data and verify the fault diagnosis effect.

Therefore, it is still necessary to carry out research in combination with practice to achieve more accurate and comprehensive fault diagnosis.

References

1. Bjørum, L.O.: Development of a digital twin for condition monitoring, focusing on electrical propulsion systems for marine application. no. June, p. 107 (2019)
2. Jan, B., Hansen, F., Wendt, F.: History and state of the art in commercial electric ship propulsion, integrated power systems, and future trend. pp. 1–14 (2015)
3. Li, H.H.C., Liu, N., Su, J.: Vibro-acoustic responses of a coupled propeller-shaft-hull system due to propeller forces. *Ocean Eng.* **173**, 460–468 (2019)
4. Vizentin, G., Vukelić, G., Srok, M.: Common failures of ship propulsion shafts. *Pomorstvo* **31**(2), 85–90 (2017). <https://doi.org/10.31217/p.31.2.1>
5. He, J., Li, Y., Cao, J., Li, Y., Jiang, Y., An, L.: An improved particle filter propeller fault prediction method based on grey prediction for underwater vehicles. *Trans. Inst. Meas. Control* **42**(11), 1946–1959 (2020). <https://doi.org/10.1177/0142331219901202>
6. Zhang, S.K., Fan, L., Gao, J., Pu, J., Xu, K.: Fault diagnosis of underwater vehicle and design of intelligent self-rescue system. *J. Coast. Res.* **83**, 872–875 (2019)
7. Li, Z., Peng, Z.: A new non-linear blind source separation method with chaos indicators for decoupling diagnosis of hybrid failures: a marine propulsion gearbox case with a large speed variation. *Chaos Solitons Fractals* **89**, 27–39 (2016). <https://doi.org/10.1016/j.chaos.2015.09.023>
8. Dimitrov, L., Kanturska, S.: Features in the selection and operation of AC motors for electric propulsion system in ship. In: 2017 15th International Conference on Electrical Machines, Drives and Power Systems ELMA 2017—Proceeding. pp. 228–232 (2017). <https://doi.org/10.1109/ELMA.2017.7955438>
9. Cardoso, A.J.M.: Evolution and Development Prospects of Electric Propulsion Systems of Large Sea Ships. pp. 296–303 (2020)
10. Jain, J.K., Ghosh, S., Maity, S.: Concurrent PI controller design for indirect vector controlled induction motor. *Asian J. Control* **22**(1), 130–142 (2020). <https://doi.org/10.1002/asjc.1911>
11. Glowacz, A., et al.: Detection of deterioration of three-phase induction motor using vibration signals. *Meas. Sci. Rev.* **19**(6), 241–249 (2019). <https://doi.org/10.2478/msr-2019-0031>
12. Ciabattoni, L., Ferracuti, F., Freddi, A., Monteriu, A.: Statistical spectral analysis for fault diagnosis of rotating machines. *IEEE Trans. Ind. Electron.* **65**(5), 4301–4310 (2018). <https://doi.org/10.1109/TIE.2017.2762623>
13. Faiz, J., Takbash, A.M., Mazaheri-Tehrani, E.: A review of application of signal processing techniques for fault diagnosis of induction motors—part I. *AUT J. Electr. Eng.* **49**(2), 109–122 (2017). <https://doi.org/10.22060/eej.2017.13219.5142>
14. Wang, Z., Yang, J., Ye, H., Zhou, W.: A review of permanent magnet synchronous motor fault diagnosis. In: 2014—Conference Proceeding Transportation Electrification Asia-Pacific (ITEC Asia-Pacific). IEEE, (2014). <https://doi.org/10.1109/ITEC-AP.2014.6940870>
15. Abdelli, R., Bouzida, A., Touhami, O., Ouadah, M.: Static eccentricity fault modeling in permanent—magnet synchronous motors. In: 2016 8th International Conference on Modelling, Identification and Control, vol. 1, pp. 364–368 (2016)
16. Wu, J., Wu, C., Cao, S., Or, S.W., Deng, C., Shao, X.: Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines. *IEEE Trans. Ind. Electron.* **66**(1), 529–539 (2019). <https://doi.org/10.1109/TIE.2018.2811366>
17. Wu, J., Su, Y., Cheng, Y., Shao, X., Deng, C., Liu, C.: Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system. *Appl. Soft Comput. J.* **68**, 13–23 (2018). <https://doi.org/10.1016/j.asoc.2018.03.043>

18. Han, H.S., Lee, K.H.: Experimental verification for lateral-torsional coupled vibration of the propulsion shaft system in a ship. *Eng. Fail. Anal.* **104**(January), 758–771 (2019). <https://doi.org/10.1016/j.engfailanal.2019.06.059>
19. Shin, S.H.: Effects of propeller forces on the propeller shaft bearing during going straight and turning of ship. *J. Soc. Naval Architects Korea* **52**(1), 61–69 (2015)
20. Shubao, W.: Analysis of propulsion shafting high temperature failure on a large container ship and countermeasures. *J. Shanghai Sh. Shipp. Res. Inst.* **42**(3), 46–50 (2019)
21. Lee, J.U.: Application of strain gauge method for investigating influence of ship shaft movement by hydrodynamic propeller forces on shaft alignment. *Measur. J. Int. Measur. Confed.* **121**(July 2017), 261–275 (2018). <https://doi.org/10.1016/j.measurement.2018.02.067>
22. Ali, M., Shaikh, S.: Data acquisition system & real time monitoring of the parameters of induction motor via wireless communication. no. IEEEC, pp. 5–9 (2020)
23. Qiongjun, L.X.X., Peiliang, M.A.: Structure and maintenance of propeller shaft sealing device for large and medium-sized ships. *Sh. Eng.* **41**, 196–198 (2019)
24. Zhen, D., Guo, J., Xu, Y., Zhang, H., Gu, F.: A novel fault detection method for rolling bearings based on non-stationary vibration signature analysis. *Sensors (Basel)* **19** (2019)
25. Medoued, A., Mordjaoui, M., Soufi, Y., Sayad, D.: Induction machine bearing fault diagnosis based on the axial vibration analytic signal. *Int. J. Hydrogen Energy* **41**(29), 12688–12695 (2016). <https://doi.org/10.1016/j.ijhydene.2016.02.116>
26. Zhao, X., Jia, M.: A novel unsupervised deep learning network for intelligent fault diagnosis of rotating machinery. *Struct. Heal. Monit.* **19**(6), 1745–1763 (2020). <https://doi.org/10.1177/1475921719897317>
27. Wang, Y., Zhou, J., Zheng, L., Gogu, C.: An end-to-end fault diagnostics method based on convolutional neural network for rotating machinery with multiple case studies. *J. Intell. Manuf.* (2020). <https://doi.org/10.1007/s10845-020-01671-1>
28. Chen, Z., Wang, T., Gu, F., Haram, M., Ball, A.: Gear transmission fault diagnosis based on the bispectrum analysis of induction motor current signatures. *Jixie Gongcheng Xuebao/J. Mech. Eng.* **48**(21), 84–90 (2012). <https://doi.org/10.3901/JME.2012.21.084>
29. Ali, M.Z., Liang, X.: Induction motor fault diagnosis using discrete wavelet transform. In: 2019 IEEE Canadian Conference of Electrical and Computer Engineering CCECE 2019, no. c, pp. 1–4 (2019). <https://doi.org/10.1109/CCECE.2019.8861923>
30. Dehina, W., Boumehraz, M., Kratz, F., Fantini, J.: Diagnosis and comparison between stator current analysis and vibration analysis of static eccentricity faults in the induction motor. In: Proceedings—2019 4th International Conference on Power Electronics and Their Applications ICPEA 2019, vol. 1, no. September, pp. 1–4 (2019). <https://doi.org/10.1109/ICPEA1.2019.8911193>
31. Granda, D., Aguilar, W.G., Arcos-Aviles, D., Sotomayor, D.: Broken bar diagnosis for squirrel cage induction motors using frequency analysis based on mcsa and continuous wavelet transform. *Math. Comput. Appl.* **22**(2), 30 (2018). <https://doi.org/10.3390/mca22020030>
32. Rangel-magdaleno, J., Peregrina-barreto, H., Ramirez-cortes, J., Cruz-vega, I.: Hilbert spectrum analysis of induction motors for the detection of incipient broken rotor bars. *Measurement* **109**, 247–255 (2017). <https://doi.org/10.1016/j.measurement.2017.05.070>
33. Huang, B., Feng, G., Tang, X., Gu, J.X., Xu, G.: A performance evaluation of two bispectrum analysis methods applied to electrical current. *Energies* **12**, 1438 (2019)
34. Qiu, L., et al.: Analysis of electromagnetic force and deformation behavior in electromagnetic tube expansion with concave coil based on finite element method. *IEEE Trans. Appl. Supercond.* **28**(3), 1–5 (2018). <https://doi.org/10.1109/TASC.2017.2789287>
35. Sangeetha, P.B., Hemamalini, S.: Rational-dilation wavelet transform based torque estimation from acoustic signals for fault diagnosis in a three-phase induction motor. *IEEE Trans. Ind. Inf.* **15**(6), 3492–3501 (2019). <https://doi.org/10.1109/TII.2018.2874463>
36. Delgado-Arredondo, P.A., Morinigo-Sotelo, D., Osornio-Rios, R.A., Avina-Cervantes, J.G., Rostro-Gonzalez, H., de Romero-Troncoso, R.J.: Methodology for fault detection in induction motors via sound and vibration signals. *Mech. Syst. Signal Process.* **83**, 568–589 (2017). <https://doi.org/10.1016/j.ymsp.2016.06.032>

37. Grządziela, A., Musiał, J., Pająk, M.: A method for identification of non-coaxiality in engine shaft lines of a selected type of naval ships. *Pol. Marit. Res.* **22**(1), 65–71 (2015)
38. Han, H.S., Lee, K.H.: Experimental verification for lateral-torsional coupled vibration of the propulsion shaft system in a ship. *Eng. Fail. Anal.* **104**(June), 758–771 (2019). <https://doi.org/10.1016/j.engfailanal.2019.06.059>
39. Li, G., Deng, C., Wu, J., Chen, Z. and Wang, Y.: Real-time intelligent fault diagnosis using deep convolutional neural networks and wavelet transform. In: 2018 IEEE 8th International Conference on Underwater System Technology: Theory and Applications, pp. 1–5, (2018)
40. Li, Z., Peng, Z.: A new non-linear blind source separation method with chaos indicators for decoupling diagnosis of hybrid failures : A marine propulsion gearbox case with a large speed variation. *Chaos, Solitons Fractals* **000**, 1–13 (2015). <https://doi.org/10.1016/j.chaos.2015.09.023>
41. Xiao, L., Tang, J., Zhang, X., Xia, T.: Weak fault detection in rotating machineries by using vibrational resonance and coupled varying-stable non-linear systems. *J. Sound Vib.* **478**, 115355 (2020). <https://doi.org/10.1016/j.jsv.2020.115355>
42. Lugni, C., Greco, M., Faltinsen, O.M.: Influence of yaw-roll coupling on the behavior of a FPSO: an experimental and numerical investigation. *Appl. Ocean Res.* **51**, 25–37 (2015). <https://doi.org/10.1016/j.apor.2015.02.005>
43. Han, H.S., Lee, K.H., Jeon, S.H., Park, S.H.: Lateral-torsional coupled vibration of a propulsion shaft with a diesel engine supported by a resilient mount. *J. Mech. Sci. Technol.* **31**(8), 3727–3735 (2017). <https://doi.org/10.1007/s12206-017-0715-y>
44. Zhang, C.: Vibration measurement method of the ship propulsion shafting. **5**(12), 34–38 (2018)
45. Xin, P., Zong, Y.A.N.: Research status of intelligent oil analysis technology for mechanical equipment. (2020)
46. Bo, L.M.X.: Development and prospect of oil monitoring technology. *Guangzhou Chem. Ind.* **49**(5), 1–3 (2021)
47. Li, L.W.L.: Study on fault diagnosis method of planetary gearbox. Application 112–113 (2017)
48. Zongqing, Y., Xinyu, P., Mingxin, C.: Prediction of gear wear trend based on oil analysis. *Mod. Manuf. Eng.* 3–8 (2020)
49. Li, Y.Z.Q., He, S., Li, Q., He, W., Qin, C.: Abnormal wear failure analysis of large ship’ s stern tube based on oil monitoring. *Lubr. Eng.* **46**(May), 137–141 (2021)
50. Changchang, Z.: Research on monitoring and diagnosis of lubrication and wear status of marine machinery. South China University of Technology (2018)
51. Arifianto, D.: Crack Detection of Propeller Shaft on board Marine Ship using Microphone Array. *Journal of Physics Conference* (2018)
52. Bai, R.: Rotating machinery fault diagnosis based on acoustic emission technology. (2017)
53. Guo, Y.: Research on acoustic emission monitoring system of rotating machinery. Shenyang University of Technology (2020)
54. Tian, C.: Research on acoustic emission monitoring system based on LabVIEW. Shenyang University of Technology (2020)
55. Shao, Y.: Study on acoustic emission signal propagation characteristics of rolling bearing. Shenyang University of Technology (2020)
56. Elasha, F., Greaves, M., Mba, D., Fang, D.: A comparative study of the effectiveness of vibration and acoustic emission in diagnosing a defective bearing in a planetary gearbox. *Appl. Acoust.* **115**, 181–195 (2017). <https://doi.org/10.1016/j.apacoust.2016.07.026>
57. Chong, Z.: Research on hydrodynamic performance of new auxiliary propulsion device for ships. Jiangsu University of Science and Technology (2020)
58. Hao-dong, P.A.N., Zhi-guang, W., Chun-hu, L.I.U.: Analysis of hydrodynamic performance of propeller in streamlined high-speed ROV. *Sh. Sci. Technol.* **42**(12) 2020
59. Mengfang, G., Baowei, S.: Analysis of hydrodynamic performance of a low- speed current turbine blades. *Mech. Sci. Technol. Aerosp. Eng.* (2020)
60. Liu, S.: Research on the influence of propeller hydrodynamic force on the dynamic alignment of shafting. Dalian University of Technology (2020)

61. Hu Z.C.G.-Z, Hu Y.H., Chen Y.-Z: Hydrodynamic characteristics analysis and diagnostic application of broken blade propeller. *Sh. Sci. Technol.* **42**(7), (2020)
62. Lijian, O.: Fault diagnosis of propeller blade breakdown based on vibration method. *Guangdong Shipbuild*, pp. 35–40 (2019)
63. Ngui, W.K., Leong, M.S., Shapiai, M.I., Lim, M.H.: Blade fault diagnosis using artificial neural network. *Int. J. Appl. Eng. Res.* **12**(4), 519–526 (2017)
64. Huang, H., Jian-xin, C.H.U., Ai-di, S.H.E.N.: Diagnosis of propeller fault of electric propulsion ship based on stator current. *Navig. China* **37**, 28–31 (2014)
65. Saidi, L., Benbouzid, M., Diallo, D., Amirat, Y., Elbouchikhi, E., Wang, T.: Higher-order spectra analysis-based diagnosis method of blades biofouling in a PMSG driven tidal stream turbine. *Energies* **13** (2020)

Research on the Influence of Crack Parameters on the Vibration Characteristics of Gas Turbine Compressor Blades



Weiwen Yu, Shixi Yang, Hongwei Chi, Zhisheng Peng, and Jun He

Abstract As one of the core components of the gas turbine compressor, the blades are prone to cracks and even breakages when working in harsh environments such as high pressure, high speed, and alternating heavy loads. The generation and propagation of cracks will change the natural vibration characteristics of blades. Therefore, it is necessary to research the influence of different types of crack faults on the natural vibration characteristics of blades, which can provide a theoretical reference for the blade crack fault diagnosis. Firstly, the finite element model of the healthy gas turbine compressor blade was established, and the modal parameters were analyzed; secondly, in order to verify the accuracy of the finite element model, the blade modal experiment platform was built to carry out the modal experiment of the healthy blade based on the moving hammer method, and the influence of different sensor installation methods on the modal test results was analyzed; finally, the modal parameters of the blade with different types of crack faults were analyzed based on the finite element model, and the mapping relationship between the crack faults and the natural vibration characteristics of the blade was established. The results show that crack type variation would affect the natural vibration characteristics of the blade, which will lead to modal coupling and modal shape switching characteristics. As a result, the same order vibration mode of blade has different mode shapes at different crack positions and shapes. The results of this paper may serve as a theoretical basis for the diagnosis of compressor blade crack faults based on vibration characteristics.

W. Yu · S. Yang (✉) · J. He (✉)

School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China
e-mail: yangsx@zju.edu.cn

J. He

e-mail: hjshenhua@zju.edu.cn

W. Yu

e-mail: yuweiwen121@163.com

H. Chi

Zhejiang Rancon Turbine Machinery Co., Ltd, Hangzhou 311100, China

Z. Peng

Hangzhou Steam Turbine & Power Group Co, Ltd, Hangzhou 310016, China

Keywords Compressor blade · Blade crack · Modal analysis · Finite element method

1 Introduction

As one of the core components of the gas turbine compressor, the blades are prone to cracks and even breakages when working in harsh environments such as high pressure, high speed, and alternating heavy loads for a long time. Not only cracks during work affect the performance of the engine, but also the broken blades that detach at high speed pose a serious threat to the safety of other components [1, 2]. The generation and propagation of cracks will change the natural vibration characteristics of blades, accompanied by coupled vibration with frequency veering and mode shape switching as the main characteristics. Therefore, in order to diagnose blade fracture failure in time, it is necessary to study the effect of cracks on the vibration frequency of the blade.

Currently, many researchers have conducted a lot of research on the influence of cracks on vibration characteristics. Based on fracture mechanics and plane stress theory, Cai [3] proposed a calculation method of cross-sectional equivalent bending stiffness of the cracked blade using a beam model. In order to obtain the crack parameters, a method based on measurement of natural frequencies presented for detection of the location and size of a crack in a cantilever beam was proposed by Maiti [4] and Nandwana [5]. Kuang [6] used Galerkin's method to study the effect of crack position and depth on blade vibration mode. After that, the finite element method was introduced. Tsai [7] calculated the rotational vibration characteristics of the blade by the finite element method. Orhan [8] studied the free and forced vibration characteristics of a cracked cantilever beam via the finite element method. Witek [9] verified that vibration is the main cause of blade failure through a combination of finite element and test methods, but did not give a specific relationship between the two. Ding [10] and Ge [11] studied the effects of crack shape, depth, and position on the natural frequency of blades based on the finite element method analysis. Zhang [12] studied the effects of crack parameters on the natural vibration and forced vibration characteristics of the blade, and discussed the frequency and vibration mode changes in the vicinity of the frequency veering zone. Free and fixed plate blade root states were simulated with finite element analysis to reveal the crack effect on the natural vibration frequency of the blade by Li [13]. Vibration based damage detection technique has been used by various researchers [14–18] in various applications from wind turbine to composite structures. However, some finite element studies have been conducted on the variation regulation of blade natural frequencies with different crack parameters, most of which have not verified the accuracy of the finite element model.

In this paper, the influence of crack parameters on the natural frequency of blades is revealed by studying the natural frequency of 12Cr12Mo stainless steel blades. The influence of how the sensor is installed is considered. The finite element model is

modified with the experimental data of the moving hammer method, and the modified model is used to analyze the influence of different crack lengths and positions on the natural vibration of the blade, which provides theoretical guidance for predicting crack location or crack length by vibration frequency change to diagnose the blade crack fault.

This paper is organized as follows. In Sect. 2, the blade vibration problem is introduced. In Sect. 3, the influence of different sensor installation methods on the modal experiment results is analyzed, and the finite element model is modified with the experimental data. In Sect. 4, the modified model is used to analyze the influence of different crack lengths and positions on the natural vibration of the blade. Finally, conclusions are summarized in Sect. 5.

2 Modal Analysis Theory

Modes are the natural vibration characteristics of mechanical structures. Each mode has a specific natural frequency, damping ratio, and mode shape. The process of obtaining each mode by theoretical calculation or experimental analysis is called modal analysis, which is an important method for structural dynamic design and fault diagnosis for equipment and the basis of various dynamic analyses.

For a multi-degree-of-freedom vibration system, the dynamics of the elastic structure can be solved according to the D'Alembert principle to derive the dynamic balance equation, which can be written as

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{F} \tag{1}$$

where $\ddot{\mathbf{u}}$, $\dot{\mathbf{u}}$, and \mathbf{u} are the nodal acceleration vector, velocity vector, and displacement vector, respectively, \mathbf{M} , \mathbf{C} , and \mathbf{K} denote the mass, damping matrix, and stiffness matrices, while \mathbf{F} is the external force of the system.

When the system vibrates freely, the external force is zero and the damping of the system can be neglected. Then, the above linear differential equation can be simplified to the following homogeneous equation:

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{0} \tag{2}$$

For a linear system, the form of the above solution is

$$\mathbf{u} = \boldsymbol{\varphi} \cdot \sin(\omega t + \theta) \tag{3}$$

where $\boldsymbol{\varphi}$ is the mode shape eigenvector, ω is the natural frequency (unit: rad/s), t is time (unit: s).

The available characteristic equation is as follows:

$$(\mathbf{K} - \omega^2\mathbf{M})\boldsymbol{\varphi} = \mathbf{0} \tag{4}$$

By solving the eigenvalue problem, the natural frequency and mode shape of the blade can be obtained.

3 Modal Analysis for Normal Blades

3.1 Experimental Modal Analysis

Taking the first stage moving blade of a certain compressor as the research object, the modal test is carried out by using the moving hammer method. Namely, the acceleration sensor is fixed at a certain point on the blade, and the vibration state of each point on the blade is obtained by changing the striking position of the hammer, as shown in Fig. 1. According to the structural characteristics of the tested blade, the blade is divided into eight rectangular grids evenly and equally, the grid size is about 30 mm × 30 mm, and the positions of 15 tapping points are determined. The grid division and tapping sequence are shown in Fig. 2. The model of the sensor used is DYTRAN 3035B, and the weight is 3 g. Furthermore, in order to compare the effects of different sensor installation methods on the experimental results, this experiment carried out a single-sensor modal experiment and a dual-sensor modal experiment. In the single-sensor modal experiment, a single sensor was pasted on the 15# test point, while two sensors were pasted on the front and back of the 15# test point in the dual-sensor modal experiment. The vibration modes are shown in Figs. 3 and 4.

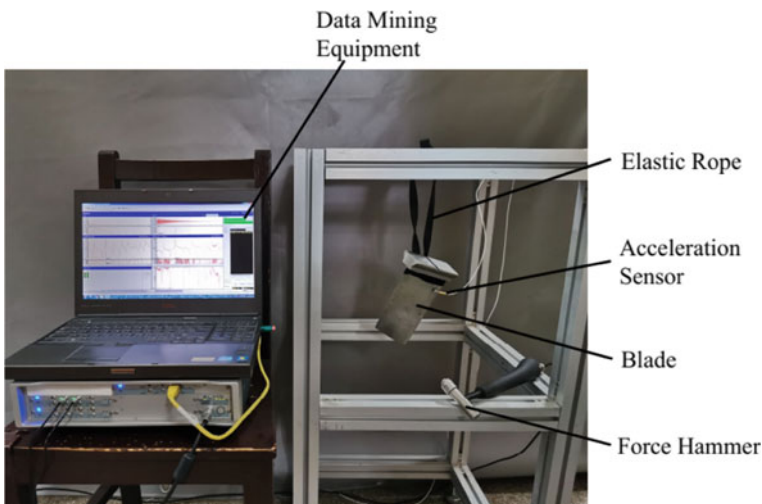


Fig. 1 Modal analysis test of compressor blade

Fig. 2 Testing point distribution figure

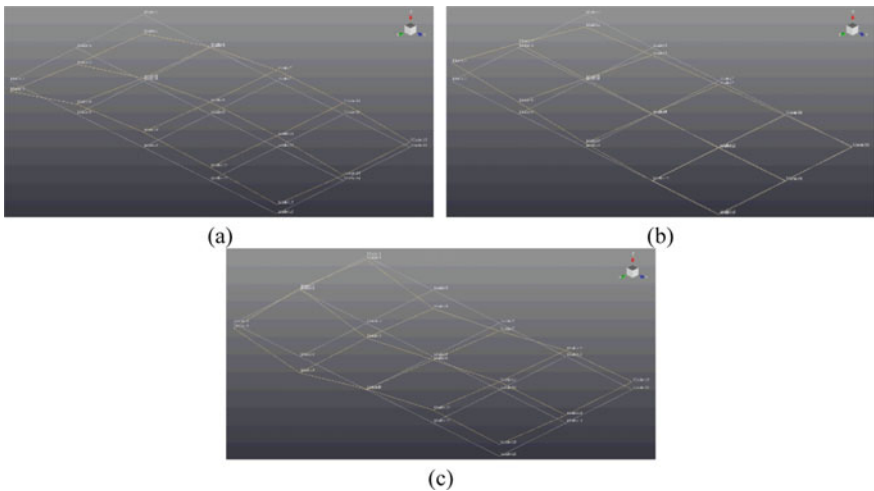


Fig. 3 The first three modes in single sensor experiment: **a** 1st mode, **b** 2nd mode, and **c** 3rd mode

3.2 Finite Element Analysis

Modal Analysis is carried out for the compressor blade through FEM-based technique using ANSYS Workbench2019 R3 software. The blade height is 164 mm, the root chord length is 78 mm, and the weight is 0.87 kg. The material model is 12Cr12Mo stainless steel with elastic modulus $E = 219 \text{ GPa}$, density $\rho = 7750 \text{ kg/m}^3$, and Poisson's ratio $\mu = 0.31$. The SOLID187 element is selected to mesh the blades, as shown in Fig. 5.

The blade is in the free state, and the vibration shapes obtained after the solution are shown in Fig. 6.

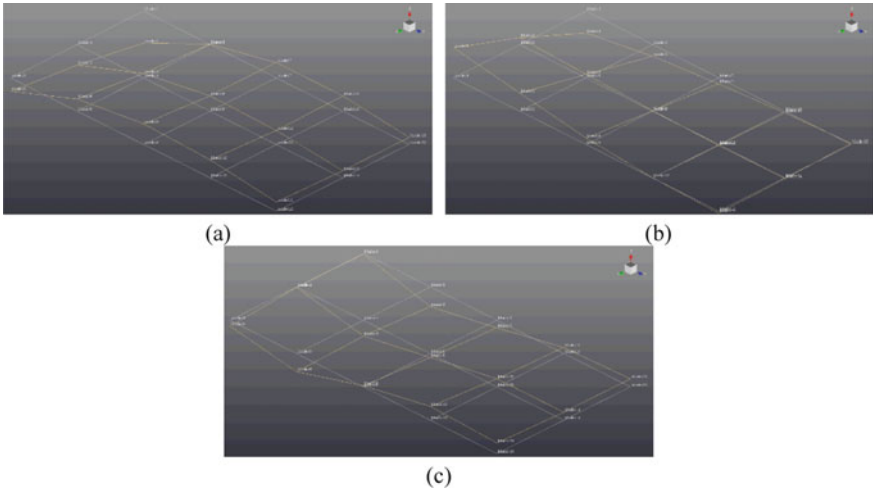


Fig. 4 The first three modes in dual sensor experiment: **a** 1st mode, **b** 2nd mode, and **c** 3rd mode

Fig. 5 The finite element model of the blade



3.3 Comparison of Experimental Modal Analysis and Finite Element Analysis

It shows that the first three modes obtained by finite element analysis are consistent with those obtained by experimental analysis, which verifies the validity of the ANSYS finite element model. As shown in Table 1, the finite element analysis results are consistent with the experimental results, and the natural frequency values of each order are roughly the same. Although there is a certain error, the error is within the acceptable range, which can explain the effectiveness of the finite element model. In addition, the results obtained by the single sensor and dual sensor tests are almost

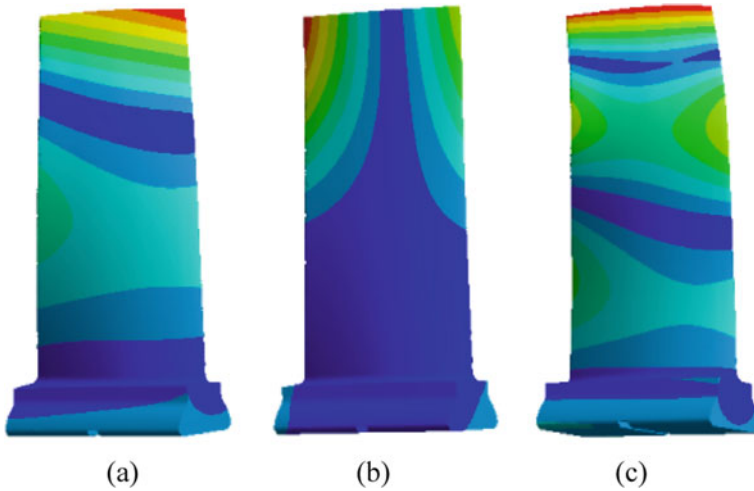


Fig. 6 The first three modes obtained by finite element analysis: **a** 1st mode, **b** 2nd mode, and **c** 3rd mode

Table 1 Comparison of finite element analysis and experimental modal analysis results in free state

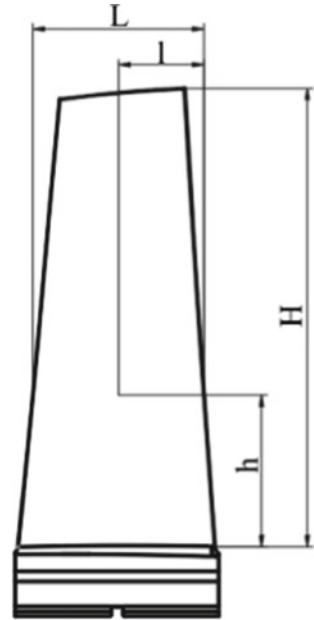
Mode	Natural frequency		Relative error (%)	Natural frequency		Relative error (%)
	Single sensor	ANSYS		Dual sensor	ANSYS	
1	884.73	877.31	0.84	883.77	877.31	0.73
2	1667.26	1649.50	1.07	1665.45	1649.50	0.95
3	2364.17	2345.30	0.80	2352.03	2345.30	0.29

the same. The reason is that the mass of the sensor is much smaller than the blade, so the sensor has little effect on the result.

4 Influence of Crack Parameters on Blade Mode

Crack parameters mainly include crack length and crack position. This paper studies the influence of crack length and position on the natural vibration of the blade. The finite element analysis model of the cracked blade is constructed on the basis of the blade model, as shown in Fig. 7. l is the length of the crack, L is the chord length of the blade root, h is the distance between the crack and the blade root, and H is the height of the blade. Define the ratio of crack length to blade width (l/L) as the crack length ratio (L_c), which represents the relative length of the crack; Define the ratio of crack height to blade height (h/H) as the crack height ratio (H_c), which represents the relative crack height position. The blades in the actual compressor can

Fig. 7 Blade crack analysis model



be approximated as fixed blade root. The finite element model of the cracked blade can be solved for the natural frequency of different crack lengths and different crack positions under the constraint of the fixed blade root. The curves of first three-order natural vibration frequency varying with crack length and crack height are drawn in Figs. 8 and 9, whose variables refers to [19]. As can be seen in Fig. 8, when the crack propagates at the same location, the natural vibration frequency decreases gradually. The results show that the longer the crack, the greater the impact on the stiffness of the blade, and the greater the impact on the natural frequency of the blade. As can be seen in Fig. 9, the closer the crack is to the leaf root, the greater the effect on the natural vibration frequency.

Figure 10 shows the first twelve natural frequencies where the crack height ratio was fixed at $H_c = 0.4$, and the crack length was varied as $0 \leq L_c \leq 0.9$. It is worth noting that the blade exhibits characteristics such as frequency veering with the crack length (circled in Fig. 10). Frequency veering refers to the phenomenon that the trajectory of the system characteristic value converges with some system characteristic parameters but does not cross and then separates [20]. The essence of frequency veering is vibration coupling and mode shape switching between modes. Besides, keep the crack length ratio fixed at $L_c = 0.5$ while the crack height varied. The results are shown in Fig. 11. When the crack is longer, the change of the crack position will also cause frequency veering, modal coupling, and mode shape switching characteristics. When the crack length is small, the variation of the crack position has little effect on the frequency.

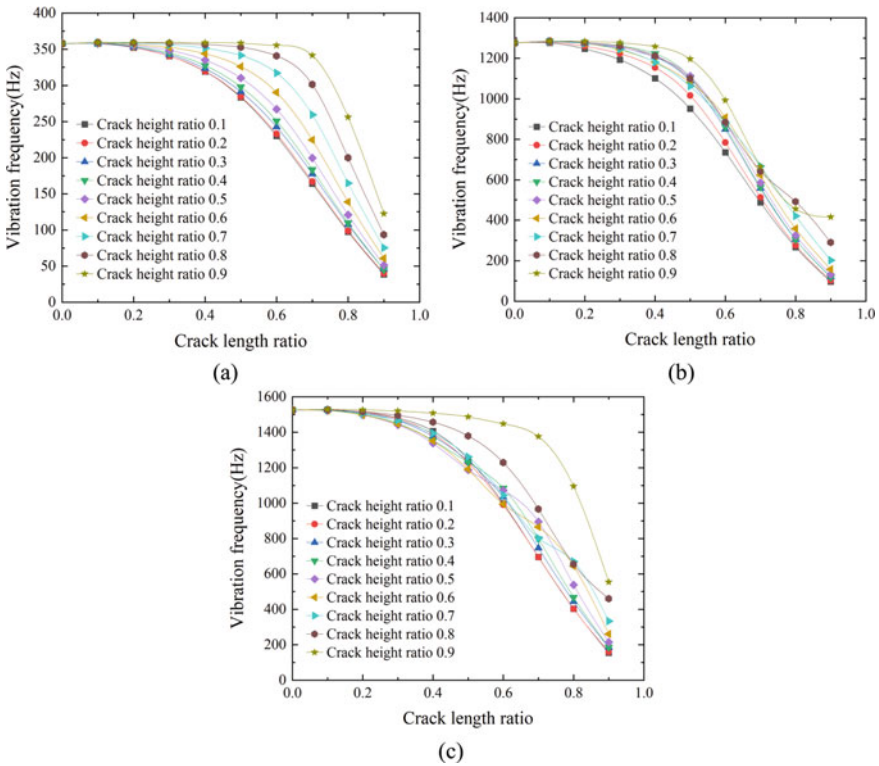


Fig. 8 Curve of natural vibration frequency with crack length in root fixed state: **a** 1st mode, **b** 2nd mode, and **c** 3rd mode

In order to observe the veering regions and the variations in the mode shapes more clearly, representative cases are analyzed. As shown in Fig. 12, when the crack length ratio increases in the veering region, the modes 9 and 10 start to mix and then switch. Figure 13 shows the veering region caused by the variation of crack position. In the 5th and 6th modes, the frequency veering and mode shape switching is completed between the two modes as the crack position increases.

5 Conclusion

This paper studies the influence of the length and position of the crack on the natural vibration characteristics of the first stage moving blade of the Gas Turbine compressor, and the conclusions are as follows:

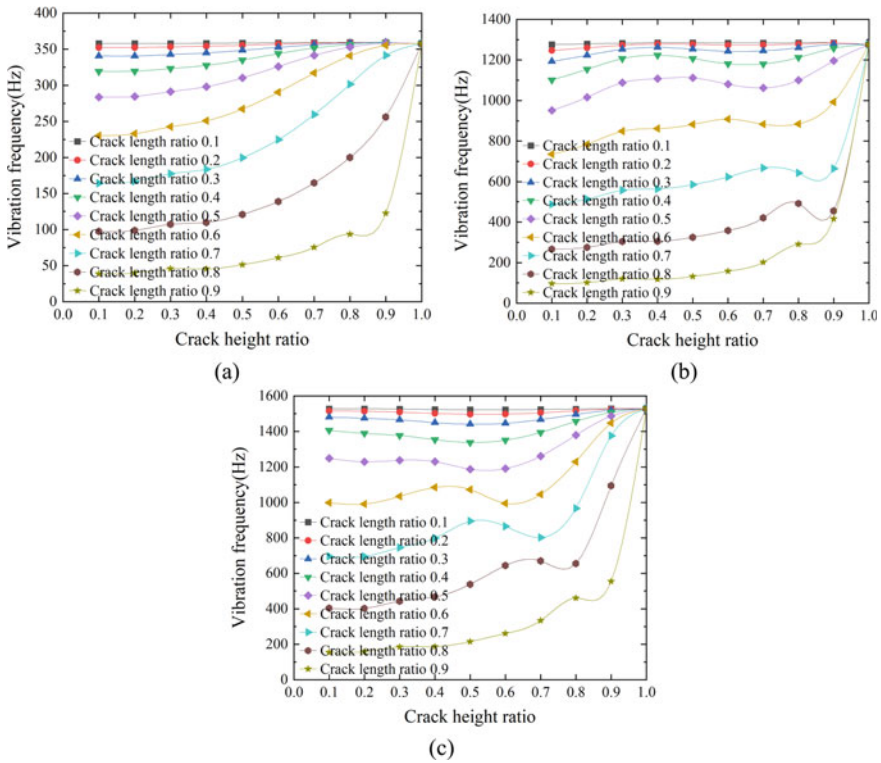


Fig. 9 Curve of natural vibration frequency with crack height in root fixed state: a 1st mode, b 2nd mode, and c 3rd mode

Fig. 10 Crack length compared with natural frequency

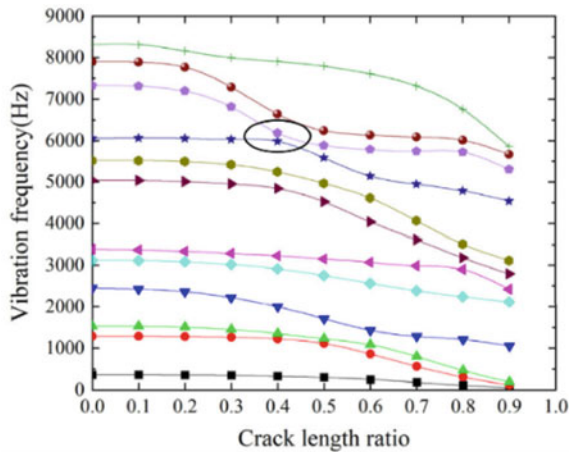


Fig. 11 Natural frequency varying with crack location

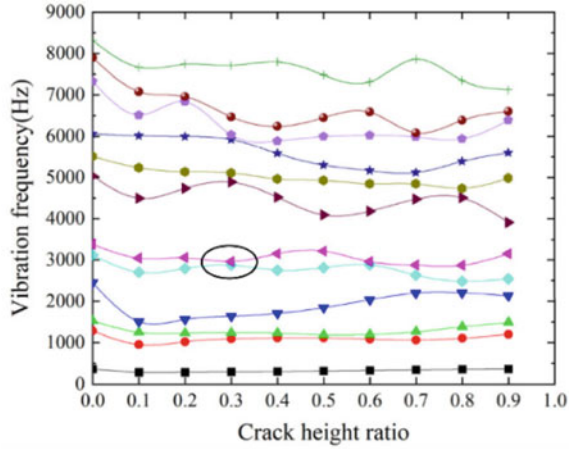
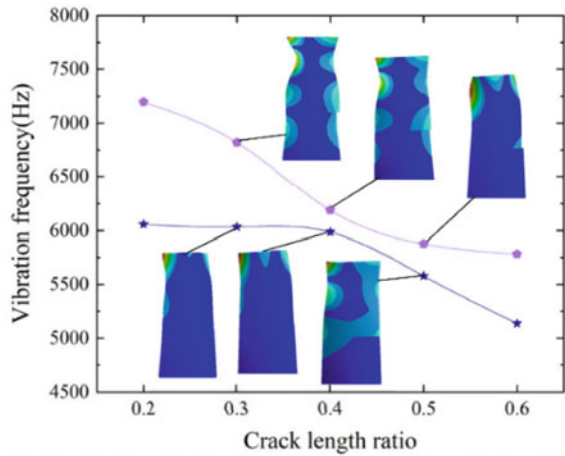
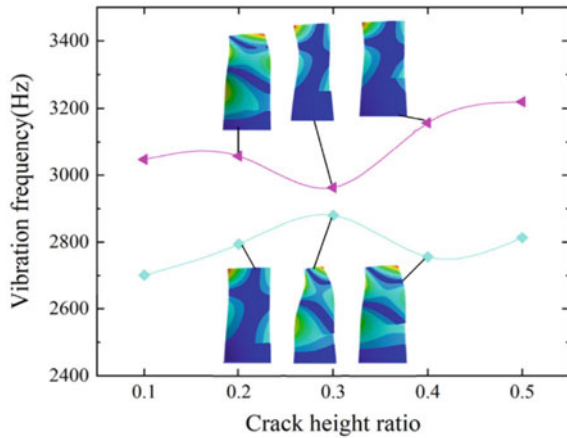


Fig. 12 Frequency veering and mode shape of 9 and 10 mode varying with crack length



- (1) The maximum relative error between the natural frequency of the blade solved by ANSYS finite element analysis and the one based on the moving hammer method test is 1.07%, which confirms the accuracy of the finite element model.
- (2) Cracks can reduce the blade stiffness and affect its natural frequency. The longer the crack is, the greater the change in stiffness matrix, and the greater the impact on the natural frequency of the blade; when the crack position tends to the free end, the impact of the crack on the natural frequency of the system gradually decreases.
- (3) The length and location of the crack both affect the natural frequency of the blade. The increase of the crack length and the change of the position can lead to the frequency veering and mode shape switching characteristics, which

Fig. 13 Frequency veering and mode shape of 5 and 6 mode varying with crack height



explains why mode shapes alter with the variation of crack length and location in the same order of vibration mode.

Acknowledgements The authors acknowledge the support from the National Natural Science Foundation of China (Grant No. U1809219) and the Key Research and Development Project of Zhejiang Province (Grant No. 2020C01088).

References

- Li, S.H., Cai, L.M.: Fan blade crack fault diagnosis based on the analysis of pneumatic signals. *J. Vib. Shock* **36**(19), 227–231 (2017)
- Zhang, Y.G.: Key technology research on non-contact online monitoring for fume turbine blade vibration. Tianjin University (2008)
- Cai, W.H.: Equivalent bending stiffness calculation for rotating blade crack section. *J. Jilin Univ.* **32**(3), 41–44 (2002)
- Maiti, S.K., Lele, S.P.: Modeling of transverse vibration of short beams for crack detection and measurement of crack extension. *J. Sound Vib.* **257**(3), 559–583 (2002)
- Nandwana, B.P., Maiti, S.K.: Detection of the location and size of a crack in stepped cantilever beam based on measurement of natural frequencies. *J. Sound Vib.* **203**(3), 435–446 (1997)
- Kuang, J.H., Huang, B.W.: The effect of blade crack on mode localization in rotating bladed disks. *J. Sound Vib.* **227**(1), 85–103 (1999)
- Tsai, G.C.: Rotating vibration behavior of the turbine blades with different groups of blades. *J. Sound Vib.* **271**(3–5), 547–575 (2004)
- Orhan, S.: Analysis of free and forced vibration of a cracked cantilever beam. *NDT & E Int.* **40**(6), 443–450 (2007)
- Witek, L.: Experimental crack propagation and failure analysis of the first stage compressor blade subjected to vibration. *Eng. Fail. Anal.* **16**(7), 2163–2170 (2009)
- Ding, Z.J., Ji, G.Y.: Blade crack diagnosis based on finite element method. *J. Power Equip.* **29**(10), 1385–1388, 1392 (2010)
- Ge, Y.Q., An, L.S.: Study of the influence of crack parameters on the blade natural frequencies. *J. Power Eng.* **28**(4), 519–522 (2008)

12. Zhang, J.H., Yang, S., et al.: Influence of crack parameters on frequency veering characteristic of aero engine blade. *J. Vib. Shock* **33**(20), 7–11 (2014)
13. Li, C.W., Li, J., Fang, Y.W.: Simulation of the crack geometry effect on the natural vibration frequency of a plate blade. *Strength Mater.* **54**(4), 97–102 (2020)
14. Krawczuk, M., Ostachowicz, W., Zak, A.: Dynamics of cracked composite material structures. *Comput. Mech.* **20**(1), 79–83 (1997)
15. Ghoshal, A., Sundaresan, M.J., et al.: Structural health monitoring techniques for wind turbine blades. *J. Wind Eng. Ind. Aerodyn.* **85**, 309–324 (2000)
16. Ge, M., Lui, E.M.: Structural damage identification using system dynamic properties. *Comput. Struct.* (2005)
17. Rucka, M., Wilde, K.: Application of continuous wavelet transform in vibration based damage detection method for beams and plates. *J. Sound Vib.* **297**(3–5), 536–550 (2006)
18. Rolfes, R., Gerasch, W., et al.: Early damage detection system for towers and rotor blades of offshore wind turbines. In: *Proceedings of the 3rd European Workshop on Structural Health Monitoring*, Granada, Spain, pp. 5–7, June 2006
19. Saito, A., Castanier, M.P., Pierre, C.: Estimation and veering analysis of nonlinear resonant frequencies of cracked plates. *J. Sound Vib.* **326**(3–5), 725–739 (2009)
20. Ren, X.M., Nan, G.F., et al.: Studying frequency veering characteristics of aircraft engine blade with beam function combination method. *J. Northwest. Polytechnical Univ.* **27**(2), 269–273 (2009)

Research on Vibration Characteristics of Last Stage Blade Based on Blade Tip-Timing Technology



Xinyu Hu, Daming Zhuang, Jun He, Haizhou Huang, and Shixi Yang

Abstract As the key component of the steam turbine, the steam turbine blade needs to work in a complex and rigorous operating environment, which easily leads to blade cracks or even fractures and other faults. Excessive vibration is one of the main causes of blade failure, which may affect the safe and stable operation of the equipment. Therefore, it is significant to detect and analyze blade vibration characteristics. Blade tip-timing (BTT) technology has the advantages of non-contact and simple installation, which is widely used in online blade vibration monitoring of turbomachinery. In this paper, the research of using BTT technology to measure the vibration characteristic parameters of the last stage moving blade of a steam turbine with integral shroud and snubber is carried out. Firstly, a finite element model of the last stage blade is built, the stress distribution and mode shape of the blade are obtained through simulation analysis. Secondly, the blade vibration measuring experiment is accomplished on a dynamic balancing test-bed, and the synchronous vibration parameters such as resonance speed are calculated correctly under lowing speed working condition based on BTT technology. Furthermore, the strain gauge method is used simultaneously to verify the accuracy of the measurement results. The resonance frequency and engine order of the blade are measured successfully. The analysis results show that parameters such as resonance speed identified by BTT method are consistent

X. Hu · J. He (✉) · S. Yang (✉)

The State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, China

e-mail: hjshenhua@zju.edu.cn

S. Yang

e-mail: hjshenhua@zju.edu.cn

X. Hu

e-mail: xinyuhu@zju.edu.cn

D. Zhuang

Hangzhou Steam Turbine Co., Ltd, Hangzhou, China

H. Huang

Huadian Electric Power Research Institute Co., Ltd, Hangzhou, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_80

with that measured by strain gauge method. The research results prove the effectiveness of BTT technology in measuring the vibration parameters of the last stage blade with integral shroud and snubber, which can provide reference for design rationality verification and vibration characteristics detection of the steam turbine blade.

Keywords Blade tip-timing · Finite element method · Steam turbine blade · Vibration characteristic · Strain gauge

1 Introduction

With the development of thermal power technology and the adjustment of energy structure policies, the steam turbine is developing towards high parameters and large capacity, which puts forward higher requirements for the efficiency, reliability and operation range of steam turbine. As the key component of the steam turbine, the steam turbine blade needs to work in the complex and rigorous operating environment of high temperature, high pressure, and high speed, which easily leads to fatigue damage and other faults. In particular, due to its longer length and greater centrifugal force, the last stage blade is susceptible to unstable airflow impacts and is easier lead to blade cracks or even fractures. According to statistics, 70–80% [1] of forced shutdown accidents of steam turbines are caused by blade damage. Excessive blade vibration is one of the main causes of blade failure. Related investigations show that 60–70% [2] of blade damage is caused by vibration. Therefore, vibration measurement and analysis of steam turbine blade is of great significance to ensure safe and reliable operation of the steam turbine.

Rotating blade vibration detection technology has been widely researched by scholars in China and abroad. According to the measurement method, it is mainly divided into contacting methods and non-contacting methods. Traditional contacting measurement methods need to mount strain gauges [3, 4] on the blade surface and measure the vibration frequency and dynamic stress of the blade by monitoring the strain change. Wang et al. [5] built the turbine blade dynamic stress measurement system and completed dynamic stress measurement of turbine blades under high temperature and high-speed conditions. Lu et al. [6] investigated the vibration characteristics of compressor blades through finite element analysis, natural mode experiments and dynamic stress tests, and evaluated the dynamic performance of the blades. The strain gauge will interfere with the dynamic properties of blades. In addition, due to its complicated installation process and the limited number of monitoring blades, the strain gauge measurement method has gradually been used for experimental comparison and verification. Alternatively, non-contacting measurement methods overcome the above defects. In particular, BTT [7] is well-known because of its simple structure and wide monitoring range. Wang et al. [8] introduced a key phase interpolation method to solve the larger error caused by acceleration, deceleration or speed fluctuation with once-per revolution probe. Due to the high under sampling characteristics of blade tip timing data, Wu et al. [9] proposed a

blade parameter identification technology based on enhancing sparse decomposition, and the effectiveness of the algorithm has been verified. Guo et al. [10] put forward a new method to identify the parameters of synchronous resonance vibration without once-per revolution sensor, which makes it more generally applicable special condition. Guru et al. [11] used BTT method for pre-emptive prediction of rotor blade damage through carefully monitoring blade natural frequencies in conjunction with the blade tip position during an engine test. Du Toit et al. [12] put forward a hybrid approach, comprising a stochastic Finite Element Model (FEM) based modal analysis and Bayesian Linear Regression based BTT technique, to identify and classify the blade damage. BTT technology is mainly used for vibration measurement, condition monitoring, and crack identification of aeroengine and gas turbine blades without shroud and snubber. Due to the limitation of the shroud in the last stage blade of the steam turbine, it is difficult to obtain the tip timing signal by traditional tip timing method. Therefore, BTT technology is rarely applied to the study of the vibration characteristics of the last stage blade of the steam turbine.

In this paper, BTT vibration measurement technology is used to measure the vibration characteristic parameters of the last stage moving blade of a steam turbine with integral shroud and snubber. This study mainly focused on the following contents: (1) A finite element model of the last stage blade is built, and stress distribution and mode shape of the blade are obtained through simulation analysis. (2) According to the structural characteristics of the last stage blade of the steam turbine, the layout of the BTT sensor is optimized. BTT sensor is installed opposite to the blade trailing edge near blade tip, and the center line of the sensor is parallel to the rotor axis. Then, the blade vibration measuring experiment is accomplished on a dynamic balancing test-bed using BTT technology. (3) The strain gauge method is used simultaneously for comparative analysis. Comparing the BTT and the strain gauge measurement results, the reliability of the BTT measurement result is verified. The comparison results show that BTT can accurately identify the resonance speed, which can provide reference for design rationality verification and vibration characteristics detection of the steam turbine blade.

2 FEM Model Analysis

2.1 Modelling

The last stage moving blade of the steam turbine with integral shroud and snubber is studied in this paper.

There are more than 80 moving blades with torsion curved surface and fir-tree root, the material is 17-4PH, the density is 7810 kg/m^3 , the elastic modulus is 206.1 GPa, and the Poisson's ratio is 0.3.

The last stage blade is analyzed by ANSYS Workbench2019 R3 software. Hexahedron and tetrahedron meshes were set up, and the mesh size is controlled appropriately. Considering that the entire last stage blade has a cyclic symmetric structure, only a single blade is analyzed using a cyclic symmetry method to simplify the calculation. The blade root is fully constrained, and contact pairs are set between shrouds and snubbers. At the same time, the centrifugal load is applied to the blade. Stress distribution and mode shape of the blade were obtained through simulation analysis.

2.2 Simulation Results Analysis

The stress distribution of the blade in the working state is analyzed through static simulation. Since the air force is much smaller than the centrifugal force, only the centrifugal force is considered in the simulation. Figures 1 and 2 show the von Mises stress distribution of the blade and the third intensity theoretical stress distribution near the measuring point of blade root at the design rotating speed respectively. It shows that the stress near the root of the blade is relatively large. Therefore, the strain gauge should be pasted near this position.

According to the theory of modal analysis, we should pay more attention to the low-order modes of the blade. With the increase of rotor speed, the vibration of adjacent blades in the last stage will be coupled with each other, and the nodal diameter vibration will appear. Based on the study of the vibration characteristics of a single blade, this paper conducts modal analysis of the last stage blade. The first-order fourth-nodal diameter mode shape of the last stage blade at design speed

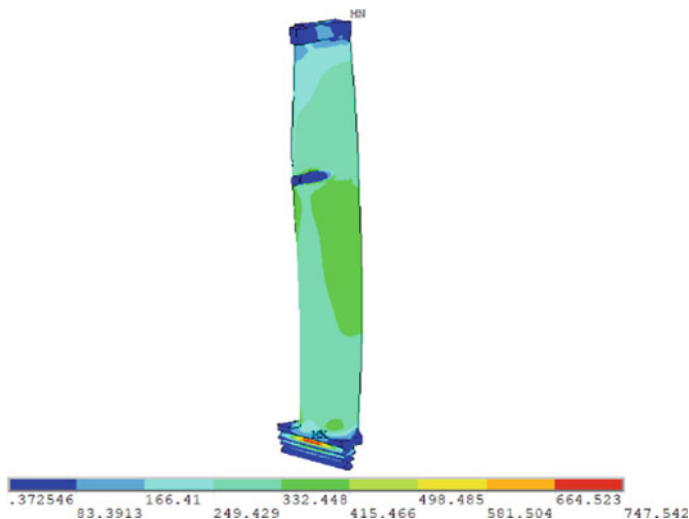


Fig. 1 Von Mises stress distribution

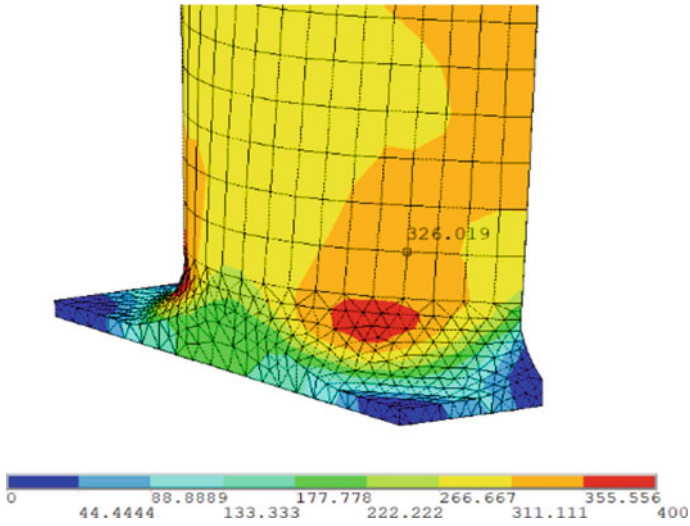


Fig. 2 The third intensity theoretical stress distribution

is in Fig. 3. As shown in Fig. 3, the blade vibration amplitude changes periodically, and the largest vibration amplitude appears at the blade tip. This is the main reason for installing the BTT sensor close to the blade tip.

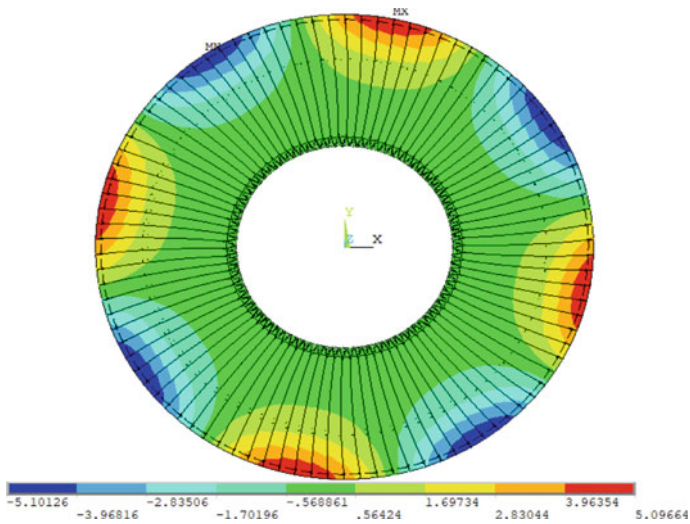


Fig. 3 First-order fourth-nodal diameter mode shape

3 Experimental Investigation

3.1 BTT Experiment

Traditional BTT technology uses tip timing sensors mounted radially on the turbine casing to measure the Times of arrivals (TOAs) of the passing blade tip. However, limited by the shroud of the last stage interlocked blade of the steam turbine, the traditional radially installed tip timing sensors cannot measure TOAs accurately.

This paper improves the traditional tip timing method to prove the effectiveness of BTT technology in measuring the vibration parameters of the last stage blade despite the limitation of the shroud ring. As shown in Fig. 4, to obtain the TOAs effectively, BTT sensor is installed opposite to the blade trailing edge near blade tip, and the center line of the sensor is parallel to the rotor axis. Based on the single degree of freedom hypothesis, if the blade vibrates in the direction of rotation, the blade trailing edge will arrive in “advance” or “delay” time compared to the non-vibrating state. According to this time difference sequence, the blade vibration amplitude signal sequence can be obtained. In this paper, the speed-vector-end-tracking method [13, 14] is used to scan the entire working area through the change of the rotating speed. When passing through the resonance region, the vibration amplitude increases significantly. The maximum value minus the minimum value of the vibration amplitude-frequency curve is the maximum blade resonance amplitude. At this time, the scanning frequency is also the blade resonance frequency. Combining the natural frequency and the Campbell diagram, the blade vibration order and the accurate frequency of the resonance can be further obtained.

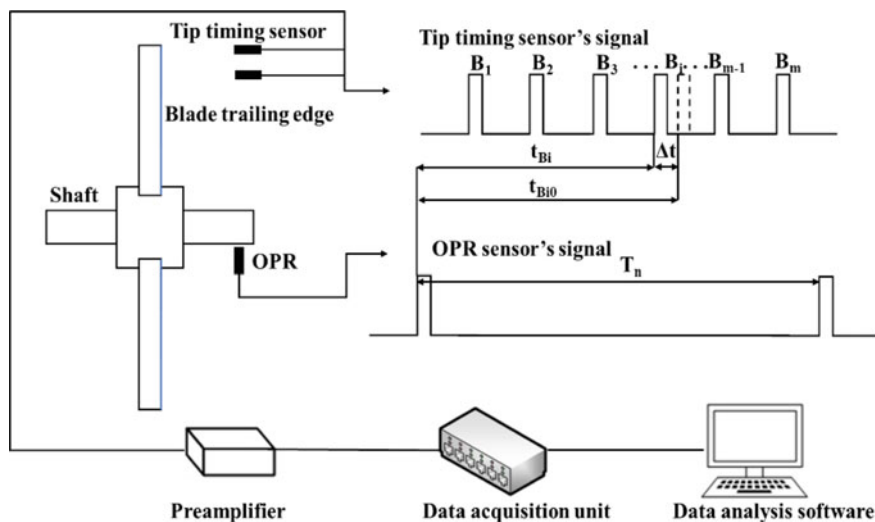


Fig. 4 Improved blade tip timing system

The blade vibration measuring experiment is accomplished on a dynamic balancing test-bed using BTT technology. The eddy current sensors are selected to acquire blade tip timing signals. As shown in Fig. 5, four eddy current sensors are installed on the bracket at equal intervals of 10° . During the experiment, the compressed gas with stable pressure and flow rate is continuously ejected from the nozzle shown in Fig. 6 to excite the blade vibration.

The blade will hardly vibrate at a low rotating speed. Therefore, the installation angle of the tip timing sensor is calibrated in advance at 1000 rpm to obtain the theoretical arrival time of the blade. During the experiment, the synchronous vibration of the last stage blade is measured by the variable speed sweep method. The rotor accelerates from 1000 rpm to the maximum speed steadily. After a period of stable operation, the rotor decelerates from the maximum test speed to 1000 rpm. The four blade tip timing sensors measure the arrival information of each blade synchronously.

Fig. 5 Tip timing sensors

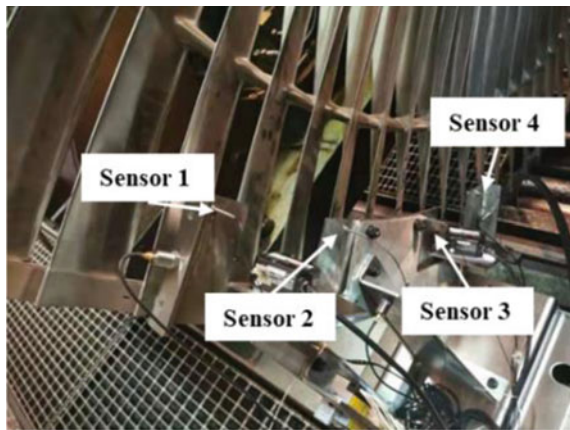
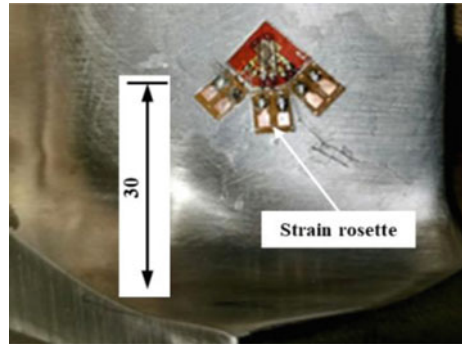
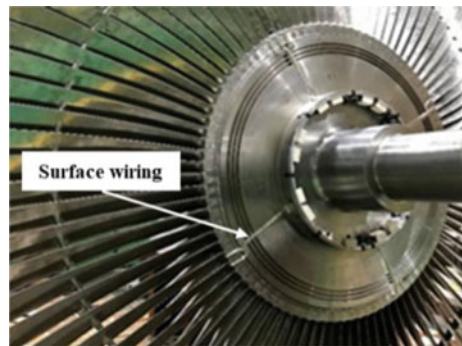


Fig. 6 Compressed air supply nozzles



Fig. 7 Strain rosette**Fig. 8** Surface wiring

3.2 Strain Gauge Experiment

In order to verify the accuracy of the measurement results, the strain gauge method is used simultaneously. According to the stress distribution of the blade at 3000 rpm in Sect. 2.2, and considering the reliability of the strain gauge position, the strain gauges are pasted at a distance of 30 mm from the blade root. Figure 7 shows the strain rosette attached to the blade. As shown in Fig. 8, given the influence of dynamic balance and monitoring the number of blades, eight strain gauges are symmetrically pasted on the blades, including uniaxial strain gauges and rosettes.

4 Results and Discussion

4.1 BTT Analysis

The raw key phase signal is shown in Fig. 9, and each pulse represents one revolution of the rotor. The rotating speed calculated from key phase signal is drawn in Fig. 10.

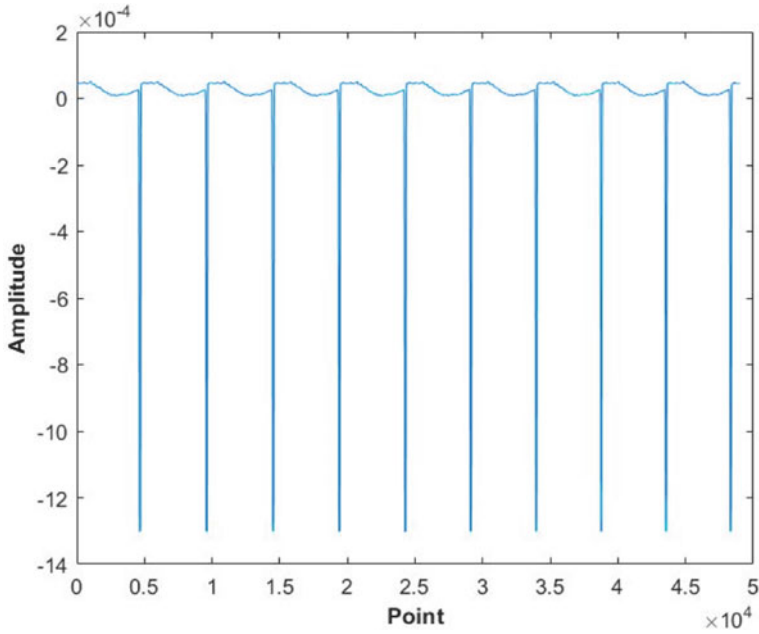


Fig. 9 Key phase signal

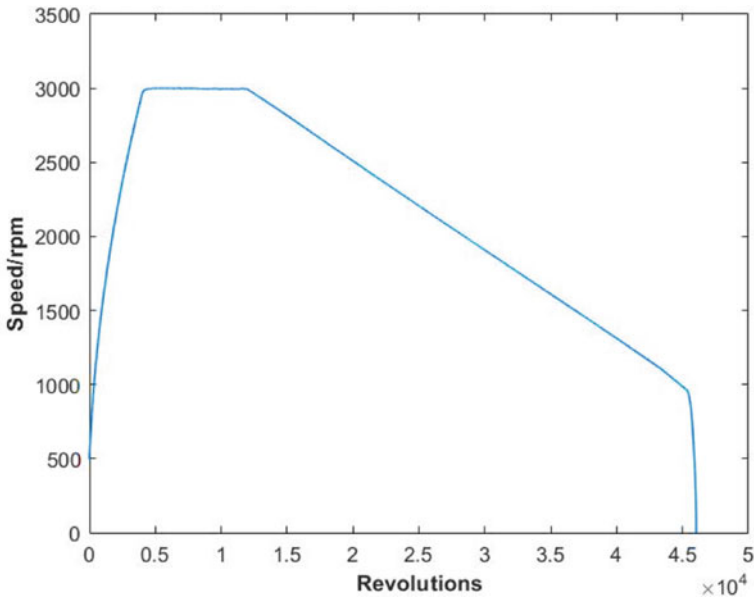


Fig. 10 Rotating speed

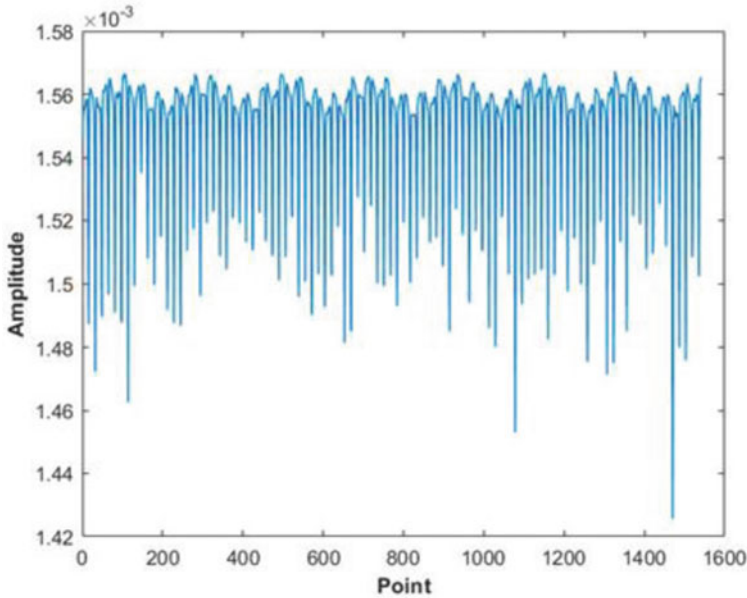


Fig. 11 Blade tip timing signal

In this paper, the data collected during the deceleration process from the maximum experimental speed to 1000 rpm is selected for analysis.

Figure 11 shows the pulse signal at the arrival time of blades in one circle. Due to the errors in blade manufacturing and experimental installation, the amplitude of each pulse which represents the TOA of each blade is not quite similar. Based on the speed-vector-end-tracking method in Sect. 3.1, the blade vibration displacement is calculated correctly under lowing speed working condition. From Fig. 12, there are two synchronous resonance regions in the process of deceleration. Through parameter identification, the resonance speeds are 1850 rpm and 2500 rpm respectively. Furthermore, combined with the Campbell diagram, the amplitude and frequency information could be obtained using the classical least square fitting method.

4.2 Stress Gauge Analysis

In order to clarify the relationship between the amplitude-frequency characteristics of the blade strain and the rotating speed during the lowing speed process, the strain data is transformed by the fast Fourier transform (FFT) for every 1 rpm reduction of the rotation speed. According to the blade vibration waterfall chart drawn in Fig. 13, there are two obvious resonance regions in the speed range of 1800–2000 rpm and 2300–2500 rpm. *The above two resonance regions correspond to the 5th and 4th*

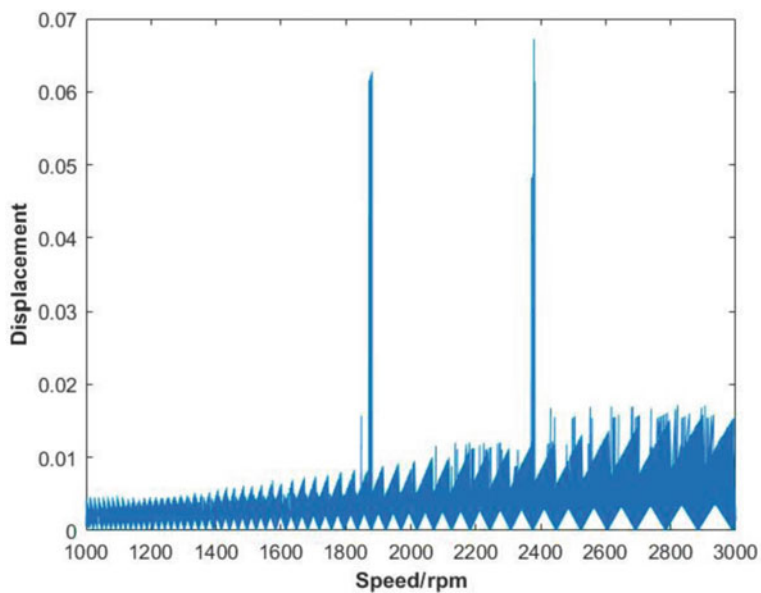


Fig. 12 Vibration displacement change

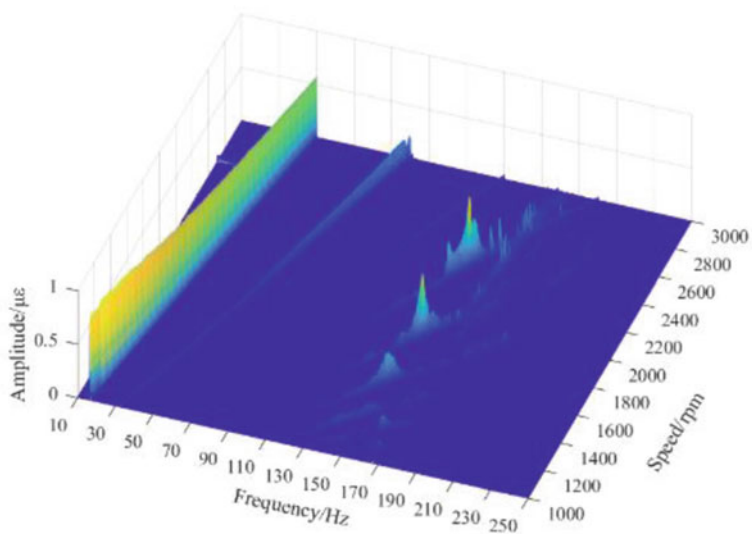


Fig. 13 Waterfall chart

vibration orders respectively, and the 4th and the 5th order analysis diagrams are obtained by waterfall chart. As shown in Fig. 14a, b, when the blade vibration amplitude reaches to the maximum, the resonance speed of the 4th and 5th engine order are 2379 rpm and 1869 rpm respectively. There is no triple point resonance vibration in the design rotation speed range of the last stage blade, which meets the dynamic frequency control requirements.

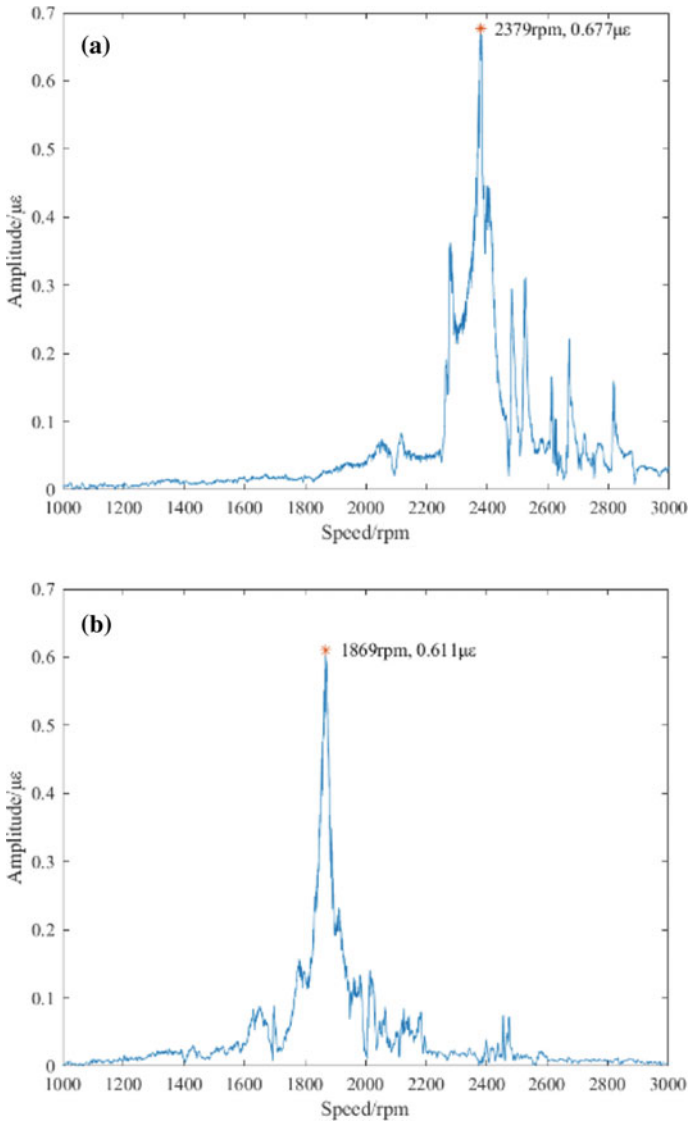


Fig. 14 a Order analysis—4th order, b order analysis—5th order

Table 1 Comparison of resonance speed measured by BTT and strain gauge methods

Engine order	BTT (resonance speed/rpm)	Strain gauge (resonance speed/rpm)	Relative error (%)
4	2832	2379	0.13
5	1880	1869	0.59

4.3 Comparison of BTT and Strain Gauge Methods

As shown in Table 1, the first resonance speed measured by BTT and strain method are 1880 rpm and 1869 rpm respectively, and the second resonance speed are 2832 rpm and 2379 rpm respectively. Taking the measurement result of the strain gauge method as a reference, the errors of the first resonance speed and the second resonance speed measured by the BTT method are 0.59% and 0.13% respectively. The resonance speeds measured by BTT and strain gauge are roughly the same. The above analysis results show that the BTT measurement results are accurate and reliable, and the BTT technology can be applied to measure the vibration parameters of the last stage blades of the steam turbine.

5 Conclusion

This paper proposes an improved BTT method to measure the vibration characteristic parameters of the last stage moving blade of a steam turbine with integral shroud and snubber. Different from the traditional tip timing method using tip timing sensors mounted radially on the turbine casing to measure the TOAs of the passing blade tip, in this paper, TOAs are measured accurately by sensing the blade trailing edge, which can get rid of the limitation of shroud ring. In addition, the strain gauge method is used simultaneously to verify the accuracy of the measurement results. The analysis results show that parameters of resonance speed identified by the BTT method are consistent with strain gauge method. The results prove the effectiveness of BTT technology in measuring the vibration parameters of the last stage moving blade with integral shroud and snubber, which can provide reference for the design rationality verification and vibration characteristics detection of the steam turbine blade.

Acknowledgements The authors acknowledge the support from the National Natural Science Foundation of China (Grant No. U1809219) and the Key Research and Development Project of Zhejiang Province (Grant No. 2020C01088).

References

1. Wang, J., Feng, T.X.: Study on the reliability of steam turbine blades in transformation. *J. Shandong Ind. Technol.* **35**(14) (2013)
2. Sheng, W., Li, Q.: Vibration modal analysis of the last stage blade of 350MW steam turbine. *Turbine Technol.* **61**(01), 23–25 (2019)
3. Knappett, D., Garcia, J.: Blade tip timing and strain gauge correlation on compressor blades. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* **222**(G4), 497–506 (2008)
4. Arpin-Pont, J., Gagnon, M., Tahan, S.A.: Strain gauge measurement uncertainties on hydraulic turbine runner blade. In: 26th Iahp Symposium on Hydraulic Machinery and Systems, vol. 15, pp. 1–7 (2013)
5. Wang, F.M., Zhu, J.Q.: Test method for the combined high and low cycle fatigues of turbine blade based on the whole engine operation. *J. Aerosp. Power* **33**(10), 2343–2350 (2018)
6. Lu, Y.H., Mei, Q.: Vibration stress test and dynamical evaluation for high rotational speed compressor blades through numerical and experimental method. *J. Propul. Technol.* **35**(10), 1398–1403 (2014)
7. Neumann, M.: A laser-optical sensor system for blade vibration detection of high-speed compressors. *Mech. Syst. Sig. Process.* **64–65**, 337–346 (2015)
8. Wang, W.M., Ren, S.Q.: The blade vibration measurement research based on the key phase interpolation method. *J. Vib. Meas. Diagn.* **37**(02), 361–365+409 (2017)
9. Wu, S.M., Hu, H.F.: Enhancing sparse decomposition based blade vibration parameter identification. *J. Mech. Eng.* **55**(19), 19–27 (2019)
10. Guo, H.T., Duan, F.J.: Blade resonance parameter identification based on tip-timing method without the once-per revolution sensor. *Mech. Syst. Sig. Process.* **66–67**, 625–639 (2016)
11. Guru, S.S., Shylaja, S.: Pre-emptive rotor blade damage identification by blade tip timing method. *J. Eng. Gas Turbines Power-Trans. ASME* **136**(7) (2014)
12. Du Toit, R.G., Diamond, D.H.: A stochastic hybrid blade tip timing approach for the identification and classification of turbomachine blade damage. *Mech. Syst. Sig. Process.* **121**, 389–411 (2019)
13. Heath, S., Imregun, M.: An improved single-parameter tip-timing method for turbomachinery blade vibration measurements using optical laser probes. *Int. J. Mech. Sci.* **38**(10), 1047–1058 (1996)
14. Zhang, Y.G., Duan, J.: Key Technology Research on Non-contact Online Monitoring for Fume Turbine Blade Vibration. Tianjin University (2008)

A Simulation Study of an Energy Harvester Operating on a Vertical Rotor System



LiChang Gu, Yubin Lin, Rongfeng Deng, Dawei Shi, Wang Wei, Zhixia Wang, Qishan Chen, Fengshou Gu, and Andrew D. Ball

Abstract The paper presents a novel magnetic coupled piezoelectric energy harvester for supplying an on-rotor sensing (ORS) IoT device. It operates based on a rotating piezoelectric beam and a fixed permanent magnet placed remotely. When the free end of the beam rotates passing through the magnet fixed on stators, an impulsive magnetic force will excite the beam to vibrate and produce electricity. A lumped electromechanical model is calibrated by fixed beam tests and subsequently used to competently evaluate the basic configuration and performance of the harvester. Simulations has verified that the harvester can performs outstandingly not only in the resonance frequency band of the beam but also the frequency range lower than half of the resonance frequency, thanks to the impulsive excitations produced by the when the beam tip passing the fixed magnet. Simulation studies also shows that this harvester can operates for both horizontal and vertical rotor systems.

Keywords Energy harvesting · Vibration · Magnet-piezoelectric beam · On-rotor sensing

L. Gu · Y. Lin · R. Deng (✉) · D. Shi · F. Gu · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield H1 3DH, UK
e-mail: rongfeng.deng2@hud.ac.uk

L. Gu
e-mail: Lichang.Gu@hud.ac.uk

Y. Lin · R. Deng · F. Gu
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, China

W. Wei · Z. Wang
School of Mechanical Engineering, Tianjin University, Tianjin 300350, China

Q. Chen
YMC Piezotronics Inc., Yangzhou 225009, China

1 Introduction

In recent years an on-rotor sensing (ORS) approach has been actively studied as it allows the dynamic behaviour of a rotor systems to be acquired directly and provides more accurate information for condition monitoring [1]. An ORS device consists typically of a vibration sensor, a microprocessor, wireless module and a power supply unit. Such an ORS node can be mounted on either shaft surface of shaft end so that lateral and axial vibrations of the complicated rotors such as that of gears [2], machining spindles [3], robots [4], wheelsets and so on can be measured with high signal-to-noise ratio (SNR) and transmitted wirelessly to host devices such as mobiles and fog commuting devices, thereby for more accurate remote monitoring.

Currently, ORS nodes are powered by battery or power capacitance units. These power supply units need frequent recharging or replacement of the batteries, leading to additional system service costs. This takes out the outstanding merits of services-free and high reliability with using wireless measurement systems and thus limited the wide applications of such systems. To overcome this bottleneck a large volumes of research works have been carried out in recent years, which are comprehensively overviewed in a number of successive survey works which include [5] overviewing different types of possible harvesting sources for wireless IoT devices, [6] examining different micro/small-scale energy storage systems to assess the integrated design's overall efficiency. Alex [7] reviewing the current state of energy harvesting progress in harvesting methods, energy storage technologies, and harvesting system architectures for self-sustaining wireless sensor networks, and [8] summarising different configuration between magnet and piezoelectric beams. Amongst many numerous harvesting approaches, piezoelectric energy harvesting (PEH) methods are the most popular one due to its merits of high energy density, easy miniaturization, simple structure, and easy miniaturisation, which are also particular requirements for ORS applications.

According to the sources of driving harvesters based piezoelectric beams, the research progresses can be reviewed based three main categories: gravitational, centripetal, and magnetic forces. Gravitational and centripetal force can be easily formalised in a rotating object, it is natural to design harvesting mechanisms based such forces. Tzern et al. [9] in 2008 proposed a gravity counterweight-based harvester in which a downward gravity force by an offset mass produces the gravity torque that intends to stop the rotations of the generator rotor associated with the mass and thus, the relevant rotations between generator and host rotor to generate electricity by magnetic induction. In 2010, Wang et al. [10] investigated using three well-designed weights to help pendulum to adjust its natural frequency to meet the wheel rotation frequency. In 2012, Wang et al. [11] proposed an electromagnetic generator system with a weighted pendulum-type pivots at a centre off the rotor centre. It oscillates due to periodic change of the tangential component of gravitational force, can match up with the rotational frequency for a large angle and angular velocity so as to generate more power. In 2021, Z. Wang et al. [12] made a comprehensive model and optimised the gravity counterweigh and all-in-one DC generator for train wheelset monitoring.

Guan et al. in 2016 [13] proposed the piezoelectric beam-based harvester, in which a mass on the tip of a piezoelectric beam is set close to the rotating centre to reduce the influences of centripetal force on system balances. Deng et al. in 2020 [14] designed a variable cross-section piezoelectric beam mechanisms in that the length of the beam is adjusted with rotational speed to tune the resonance frequency so that the operating speed can match with the frequency as close as possible for more effective energy harvesting.

External magnetic forces are investigated by many scholars to overcome the strength limitation of gravity effects. Zhou et al. in 2013 [15] proposed a piezoelectric energy harvester excited by the magnetic force by altering the angular orientation of its external magnets for enhanced broadband frequency response. Zou et al. in 2017 [16] suggested two piezoelectric cantilever beams whose free ends point to the rotating shaft, both the gravity force and centrifugal force are used to excite the vibrations of the piezoelectric beams. Wu et al. in 2018 [17] studied a piezoelectric energy harvester excited by the magnetic force. The non-contact magnetic force was employed for exciting the piezoelectric cantilever vibration to produce electric output. Multiple magnets on rotational plate with piezoelectric fixed were exponentially investigated for more effective harvesting. To enhance the power generation capability over a wider bandwidth, Fu et al. in 2019 [18] used two fixed magnets one below and the other above the tip magnet on the piezoelectric beam. When the low side magnet rotates through the tip magnet on the fixed piezoelectric beam and the magnet above, the tip magnet exhibits bistable vibration behaviours, thus widening the effective bands. More recently, Wang et al. [19] in 2021 studied to use piezoelectric beams arranged perpendicularly for achieving wideband harvesting. The tip magnet on the primary piezoelectric interact with magnets on a rotor in radial direction so that the beam system exhibit multi modal characteristics.

Specifically, a number of prototype harvesters proposed for monitoring vehicle wheels and tyres are particularly interested as such scenarios are very closer to ORS applications, in which all the harvesters rotate with rotors. In 2016, Zhang et al. [20] studied a harvester with a piezoelectric beam with tip magnetic mass rotates around an off-centre of with the vehicle wheel. The rotation centre of the beam is offsite the centre of the wheel. Bothe gravitational and magnetic forces are as the excitation to dive the beam for electricity generation. used for monitoring rotating wheels. In 2019, Rui et al. [21] used the same beam configuration as [20] but excited by gravitational force only. It avoids the drawback of magnet attracting metallic debris on road that may cause influences on the monitoring system. The beam stiffening by the centrifugal force is investigated increasing with rotational speed.

Previous studies have made significant progress in using piezoelectric beams for energy harvesting. However, they all focused on horizontal rotors systems i.e., the axis of rotation is more or less parallel to the Earth's surface, which exclusively need the assistances of gravitational effects. As such, it is not workable for vertical rotors where ORS systems frequency operates. To fill up this gap, this study proposes a new magnet- piezoelectric beam harvester workable in both horizontal and vertical rotors.

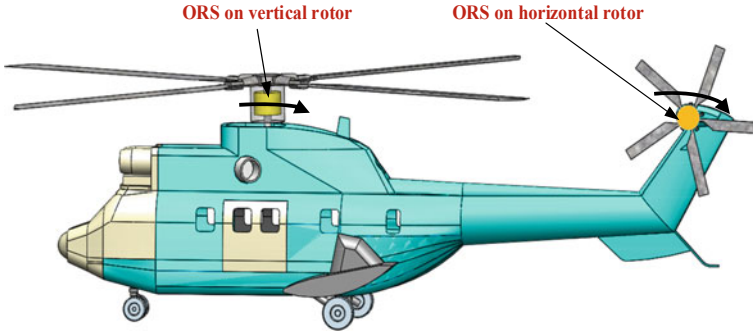


Fig. 1 A typical scenario of ORS applications

Following sections of the paper presents the prototype configuration, lumped parameters models and simulation analysis of the harvester, in which construct features and output performances are investigated.

2 Harvester Configuration and Operation Principle

2.1 ORS Operation Setups

A typical scenario of ORS applications is monitoring helicopters, as illustrated in Fig. 1. ORS units can be installed at the end of the rotors to monitor blades and their associated transmission systems. In addition to desired performance of vibration acquisition, transmission, and compactness, the supply units will operate with different inclinations as helicopters operates with different manoeuvres. This means that the harvesting devices should be workable in different rotor pitch angles. i.e., gravity effect based harvesters [16, 17] probably are not fully effective as the gravity effect will disappear or reduce significantly. This means that a more generic harvester must be developed to meet different rotor configurations.

2.2 Prototype Configuration Based on Magnet-Piezoelectric Beams

Based on the great efforts and advances carried out for powering wheel monitoring sensor systems [20] and different configuration between magnet and piezoelectric beams [8], a prototype harvester for ORS applications is proposed as shown in Fig. 2 to accommodate to the ORS device. It consists of mainly multiple piezoelectric cantilever beams with tip magnets and one externally fixed magnet with

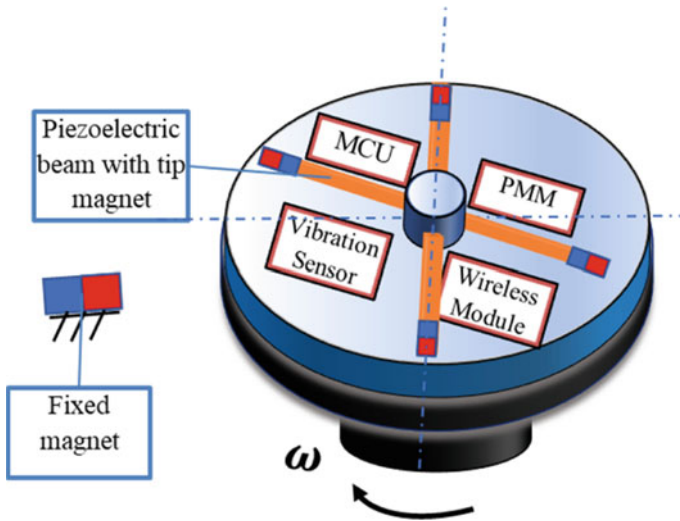


Fig. 2 A prototype of magnetic piezoelectric beam harvester

magnetisation in radial direction i.e. The beam rotates along with the ORS device comprised of a vibration sensor, a microcontroller (MCU), a wireless transmission (WT) and power management module (PMM). As the magnet at the beam end pass through the fixed magnet, a non-contact impulsive excitation is produced by the magnets and drives the beam to vibration and generate electric output to power all ORS devices straightforward. This beam configuration operates based on monostable vibrations, thereby thus avoiding complicated designs and potential instability of nonlinear based ones. Therefore, it can be easily integrated with other ORS modules to construct a compacted system, being reliable for operation, simpler in construction and more convenient for size minimisation and application. Moreover, the harvester can operate in different rotor inclinations, therefore meeting the requirements for different operating scenarios of ORS system.

3 Modelling for a Magnet-Piezoelectric Beam Harvester (MPBH)

To gain an understanding of harvester output performances, a dynamic model for a single magnet-piezoelectric beam harvester is developed, which will be the basis for multiple beam analysis. Many different models have been developed for such a piezoelectric beam-based harvesters. The distributed models based finite element analysis (FEA) provide more accurate analysis. However, it takes considerable computing time and may be inconvenient for parametric analysis as it needs large volumes of calculations. In addition, uncertainties in parameter values such as flux density

and piezoelectric constants can impact the calculation accuracy. Therefore, this study adopted a lumped electromechanical model to efficiently analyse the vibration behaviour of the beam and the electric output performances.

3.1 Lumped Electromechanical Models for a Magnet-Piezoelectric Beam

As shown in Fig. 3a, there is an impulsive force generated when the two magnets are close to each other. It will then cause the beam to vibrate. Considering the small vibration displacement of tip mass m_t and low stiffness beam k , the vibration of the tip mass can be simplified to be a single degree freedom (SOF) system, as shown in Fig. 3b including the coupling effect of electricity generation.

According to simplified mode, vibration motion y and electric output $v(t)$ can be related by refereeing to the models used in [23] as:

$$m\ddot{y} + c\dot{y} + ky + \varepsilon_p v(t) = F_w \cos(\theta) + F_t \tag{1}$$

in which the gravitational forces F_w by tip mass m_t :

$$F_w = m_t g \cos(t\omega_r) \tag{2}$$

and the electric voltage $v(t)$ output can be predicted by

$$C_p \dot{v}(t) + \frac{v(t)}{R} = \varepsilon_p \dot{y}(t) \tag{3}$$

where the equivalent mass is $m = m_p + m_t$, which is the combination of the beam and the tip mass. ε_p is a linear electromechanical coupling coefficient; C_p is the

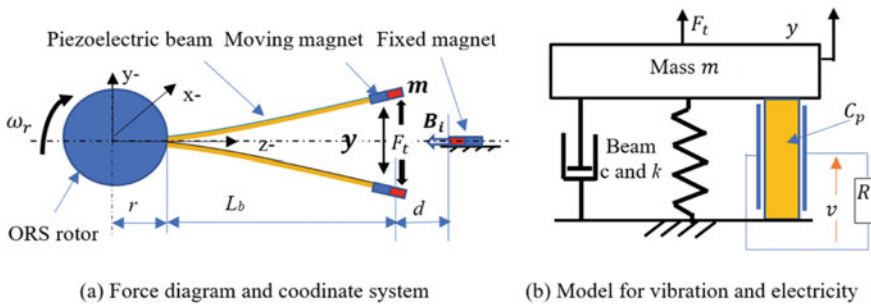


Fig. 3 Vibration model for the magnet-piezoelectric beam harvester

capacitance of the piezoelectric element; R is load resistance and θ is pitch angle of rotor systems, which is angle between the rotor axis and the earth's surface, being 0 for horizontal rotors and 90° for vertical rotors.

F_x i.e., F_y are the noncontact magnetic forces in the lateral direction perpendicular to the magnetization direction of two cuboidal magnets. This force is the primary one to induce the lateral vibration of the beam in y-direction. For two magnets of dimensions: $2A \times 2B \times 2C$ and $2a \times 2b \times 2c$ respectively, magnetic forces F_x , F_y and F_z can be calculated according to [22].

$$F_{x,y,z} = \frac{B_1 B_2}{4\pi\mu_0} \sum_{i=0}^1 \sum_{j=0}^1 \sum_{k=0}^1 \sum_{l=0}^1 \sum_{p=0}^1 \sum_{q=0}^1 (-1)^{i+j+k+l+p+q} \vartheta_{x,y,z}(U_{ij}, V_{kl}, W_{pq}, R_m) \quad (4)$$

$\vartheta_{x,y,z}$ represents the force functions corresponding ϑ_x , ϑ_y and ϑ_z to forces F_x , F_y and F_z respectively:

$$\begin{aligned} \vartheta_x = & \frac{1}{2} (V_{kl}^2 - W_{pq}^2) \ln(R_m - U_{ij}) + U_{ij} V_{kl} \ln(R_m - V_{kl}) \\ & - V_{kl} W_{pq} \tan^{-1} \frac{U_{ij} V_{kl}}{R_m W} + \frac{1}{2} R_m U_{ij} \end{aligned} \quad (5)$$

$$\begin{aligned} \vartheta_y = & \frac{1}{2} (U_{ij}^2 - W_{pq}^2) \ln(R_m - V_{kl}) + U_{ij} V_{kl} \ln(R_m - U_{ij}) \\ & - U_{ij} W_{pq} \tan^{-1} \frac{U_{ij} V_{kl}}{R W_{pq}} + \frac{1}{2} R_m U_{ij} \end{aligned} \quad (6)$$

$$\begin{aligned} \vartheta_z = & -U_{ij} W_{pq} \ln(R_m - U_{ij}) - V_{kl} W_{pq} \ln(R_m - V_{kl}) \\ & - U_{ij} V_{kl} \tan^{-1} \frac{U_{ij} V_{kl}}{R_m W_{pq}} - R_m W_{pq} \end{aligned} \quad (7)$$

where $U_{ij} = d_x + (-1)^j A - (-1)^i a$, $V_{kl} = d_y + (-1)^l B - (-1)^k b$, $W_{pq} = d_z + (-1)^q C - (-1)^p c$; $R_m = \sqrt{U_{ij}^2 + V_{kl}^2 + W_{pq}^2}$.

In which d_x , d_y and d_z are the distances from the co-centre of two magnets in x-, y- and z-axis respectively. For the coordinate system set in Fig. 3, they vary with the angular displacement, in particular, d_y and d_z are functions of rotation speed ω_r and vibrations $y(t)$ and $z(t)$ at y-axis and z-axis respectively.

$$d_x = d_{x0} + x(t) \quad (8)$$

$$d_y = L_b \sin(\omega_r t + \alpha_0) + d_{y0} + y(t) \quad (9)$$

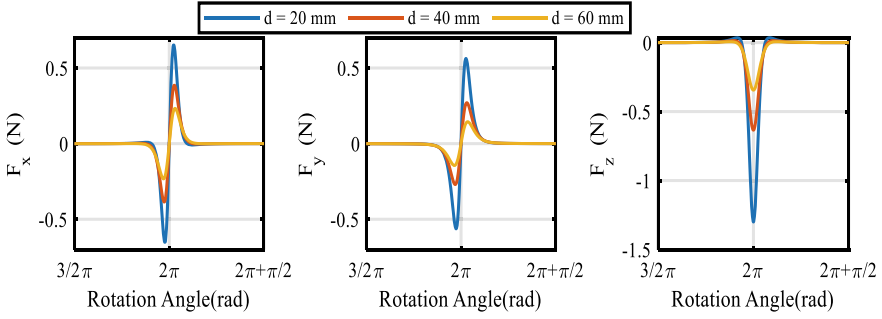


Fig. 4 Magnet force variation with distances and rotor positions when magnetization in z-axis

$$d_z = d_{z0} + [L_b - L_b \cos(\omega_r t + \alpha_0)] + z(t) \tag{10}$$

Figure 4 shows those three forces varying with distances and rotor positions when the magnetization direction of two cuboidal magnets aligns in z-axis. Owing to the superposition of magnetic fields, the magnetic forces exhibit two typical impulses in the direction of both x-axis and y-axis: a positive impulse and a negative impulse. On the other hand, only one negative impulse is generated in z-axis. Moreover, the closer the two magnets between two magnets in z-axis, the stronger of the pulses. In addition, the repulsive force in z-direction F_t has higher amplitude. These impulsive behaviours will result in different vibration behaviours and thus obtain different electric outputs.

3.2 Simulation Implementation

To gain quantitative understandings of the characteristics of vibration and electric outputs of the beam, a numerical analysis was conducted by solving the differential equations of Eqs. (1) and (3) using the Runge–Kutta method in MATLAB. It allows the time-varying impulse $F_t(x, d)$ and linear coupling ε_p to be analysed effectively. To be in accordance with the test setup for the piezoelectric beams available, main parameters are setup based on Table 1 for simulation studies, aiming at achieving in-depth understanding the effect of main operating parameters such as magnet sizes and the distance between tow magnets. The load resistance is in the range of effective power harvesting and the outputs can be recorded by common measurement and data acquisition systems. Coupling coefficient ε_p was determined by measuring the attenuation rate of voltage output records through an offline, which is based on the damping effect of coupling as shown in Eqs. (1) and (3).

In addition, both the stiffening and the softening effect due to beam axil forces: centrifugal force $F_c = (r + L_b m \omega^2)$ and magnet F_z for axially loaded beams is taken

Table 1 Key system parameters

Symbol and units	Values	Model
Magnet ($w \times h \times l$) mm ³	10 × 10 × 10	N35 neodymium
Piezoelectric beam (L × W × b) mm ³	60 × 30 × 0.3	PIEZOELECTRIC5
L_b rotating radius of beam (mm)	115	
ε_p Coupling coefficient (N/V)	0.0005	Calibrated by fixed PIEZOELECTRIC beam tests
B_i magnetic flux density (Tesla)	1.17	Calibrated by fixed PIEZOELECTRIC beam tests
Cp capacitance of piezoelectric	-8.5×10^5	Calibrated by fixed PIEZOELECTRIC beam tests
R Electric load (kΩ)	44.7	Calibrated by fixed PIEZOELECTRIC beam tests

into account based on the change in the natural frequency of a rotating cantilever [1, 24], which is estimated by referring to

$$\omega_{0r} = \omega_0 \sqrt{\left(1 + \frac{5(F_c - F_z)L_b^2}{14EI}\right)} \quad (11)$$

where the natural frequency of a cantilever without axial force is $\omega_0 = \beta_0^2 \sqrt{\frac{EI}{mL_b^3}}$.

3.3 Test Evaluation

As a number of parameters such as magnetic flux density and piezoelectric constants have high uncertainties and cannot be accurately specified by suppliers, experiments were carried out to tune these parameters. The experiments were based on a test setup where the piezoelectric beam with tip magnet is vertically fixed and a rotor a cubic magnet rotates to excite the piezoelectric beam to vibrate and the electric voltage generated by the vibrating beam was recorded by a YMC9004 ADC system at 6.250 kHz with 24 bit resolution. The system is detailed in Fig. 5. In this setup, the beam is only excited by the magnetic force induced by the rotating magnet on the rotational plate, avoiding the influence from the other sources such as gravitational forces of tip mass, unsteady speed, twists of the beam, wind turbulences and so on when piezoelectric beam is spinning. In addition, the output voltage can be captured by the stationary ADC devices for more accurate analysis. Currently ORS nodes are able to operate with data resolutions from 12 to 16 bits, much lower than that of 24 bits of the stationary ADC in use, being unable to capture the details of impulsive responses with high dynamic ranges.

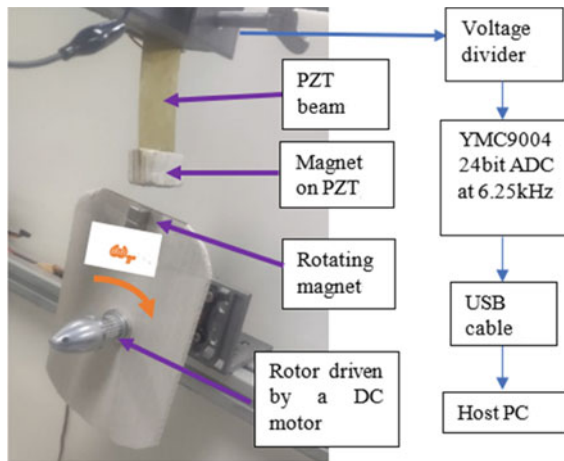


Fig. 5 Layout of a fixed PIEZOELECTRIC beam test system

Figure 6 presents typical voltage waveforms captured at two different magnet distances when the rotor spins at 9 Hz. Two successive oscillating profiles can be observed in the figure. The spikier one with shorter duration corresponds to the effect of magnetic repulsion as the magnet on rotor moves through the beam, and the smoother one with wider duration corresponds to free attenuation responses of the

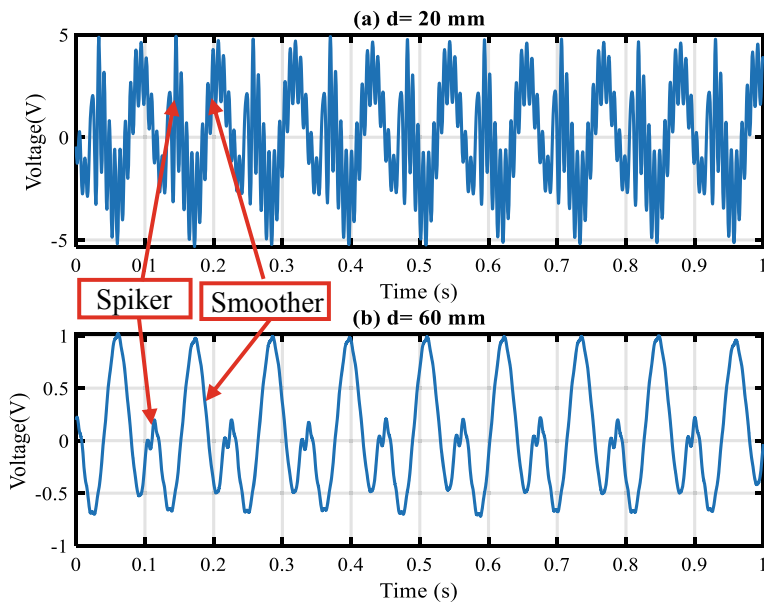


Fig. 6 Measured voltage waveforms at different magnet distances

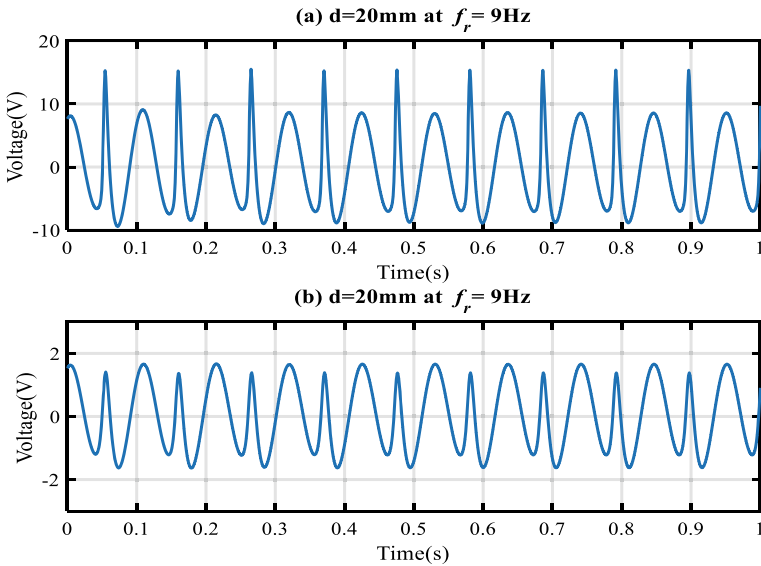


Fig. 7 Predicted voltage waveforms at different magnet distances

beam when the magnet moves away. The magnetic repositions are stronger when the distance between two magnet is smaller, as shown in Fig. 6a, Consequently, it can excite up more high modes of piezoelectric beam and thus present free attenuation oscillations in voltage waveforms. In contrast, when the distance is larger, the repulsive impulse is smaller, and thus less modes are induced, showing smaller attenuation responses, as shown in Fig. 6b.

Comparing with predicted one shown in Fig. 7 has found that predicted and measured waveforms are in good agreement. The spiker one with short duration in the predicted waveforms are more observable and so does the smoother one. In addition, the amplitudes also agreeable in that the closer the magnets the higher the amplitudes. However, the predictions do not show responses at higher-order beam modes as the model in Eq. (1) only takes into account the 1st mode of the beam. Nevertheless, this simplified model is sufficiently good for promptly analysing the dynamic responses and designing the prototype harvester.

More experiments and analysis were carried out based on the fixed piezoelectric beam test setup the beam damping coefficient c was firstly determined from the attenuating vibration displacement measured by a non-contact high speed camera. Then electromechanical coupling coefficient ϵ_p is identified by finding the difference of the displacement attenuations between electricity generation and free vibration case. Moreover, experiments for successive rotor speed increases show that the impulsive magnetic forces enable to excite the 1st modes at the lower rotation frequencies. As shown in Fig. 8, the spectrum peaks around resonant 12.5 Hz show higher magnitudes for low speed operations (below 6 Hz), showing this impulse based harvester can have a good low frequency performance.

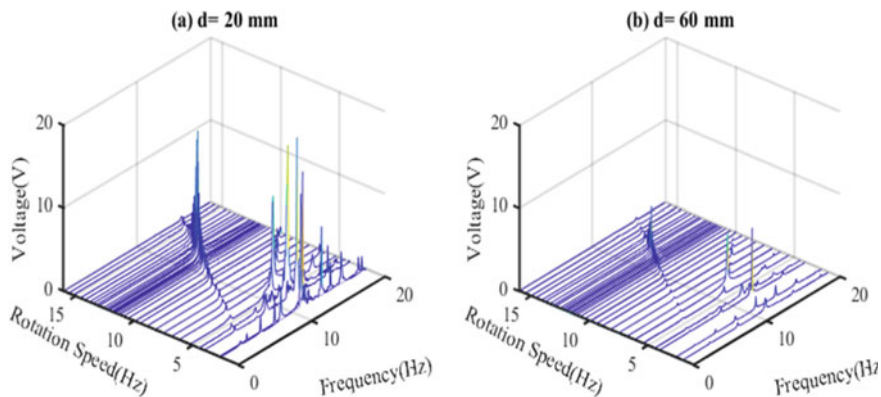


Fig. 8 Voltage spectra of measured voltages

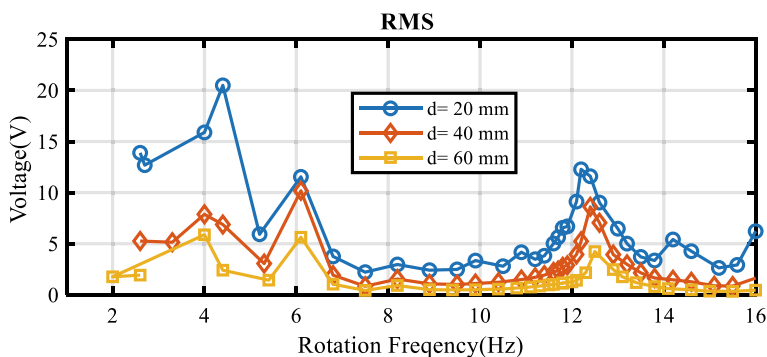


Fig. 9 RMS voltage values at different magnetic distances and rotation speeds

For more accurate analysis, voltage RMS values were calculated from waveforms and presented in Fig. 9. It can be seen that the harvester can also produce high voltages in the speed range below 6 Hz, or half of the 1st resonance at 12.5 Hz. These low frequency outputs are comparable to that at the resonances where most harvester mechanisms operating in. With these harvesting characteristics it is easier to configure other beams on the rotor with different resonances to obtain an even wider band xharvester (Fig. 10).

4 Simulation Studies

Having confirmed the correctness of models and associated programs solving the governing equation in MATLAB, subsequence simulation studies were carried out to evaluate the performance of the impulse-based harvester. Considering conveniences

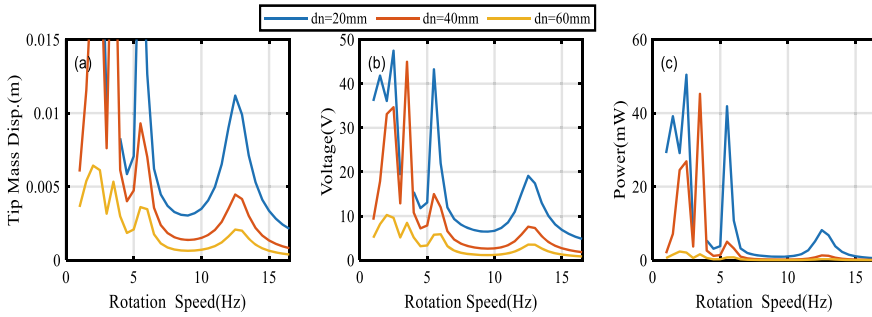


Fig. 10 Predicted outputs for different rotation speeds with different magnet distances

for setting up the fixed magnets in ORS applications, the simulation studies in this study focus on the influence of magnet distances and sizes. These two factors are critical as rotating machines often has limited spaces to accommodate the external magnet. Moreover, the output performances of the harvester element are also examined for not only the scenarios of vertical spinning rotators but also the horizontal ones, which will demonstrate the suitability of generic applications.

4.1 Outputs on Vertical Rotor Systems

For different magnet distances, both the electric voltage outputs and vibration displacements exhibit wideband performances, as shown in Fig. 10. The electric power, which is calculated based on a 44.7 kΩ load resistor, also presents similar wideband characteristics. However, when magnet is too close to the beam tip such as for the distances at 20 and 40 mm, vibration amplitudes in the low frequency range have very high peaks. Such extremes may cause certain damages to the beams and should be avoided. It means that the magnet should be placed with a good distance such as 60 mm away for more steady and prolonged operation. In addition, this distance can be safer and less intrusive as it is sufficiently far from the rotor.

For the magnetic distance at 60 mm, magnetic size is increased to improve output performance. Figure 11 are the output comparison between the sizes of fixed magnets increased by 20% and 40% respectively. By doing so, it not only leverages total magnetic flux density but also widen angular range of repulsive excitations. It can be seen that the outputs because relatively smooth across different frequency bands. This means that increase the size is a better way to improve output performance. In particular, it suggests that a large magnet is preferred for a large angular range of excitation, rather than use high increase the flux intensity by using high quality magnet, which can smoothen low frequency magnitudes for better lifetime of piezoelectric beams.

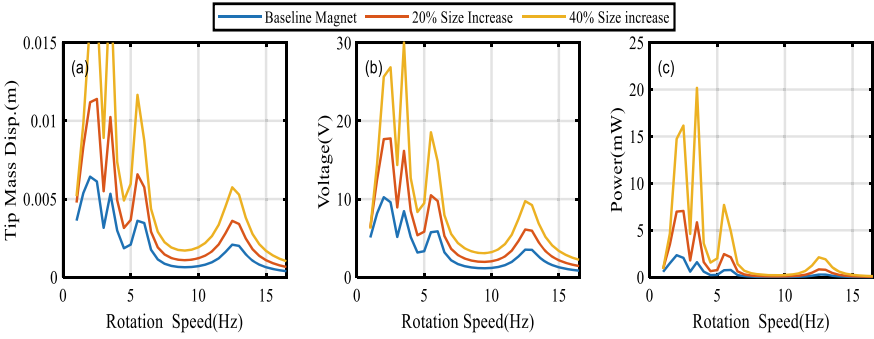


Fig. 11 Predicted outputs for different rotation speeds under different magnet sizes

4.2 Outputs on Horizontal Rotors

When the harvester attached to horizontal rotors, the gravitational force of tip mass will also excite the beam to vibrate periodically and produce electric outputs. It shows in Fig. 12 that gravitational forces can produce good outputs only around the resonant frequency at 12.5 Hz because of its stationarity of sinusoidal excitations, shown in Eq. (2).

When both gravitation and magnetic forces coexist, the harvester produces enhanced results. As shown in Fig. 13c, the power output becomes flatter across the full frequency band, which is more acceptable in terms of widening frequency bands for more effective energy harvesting.

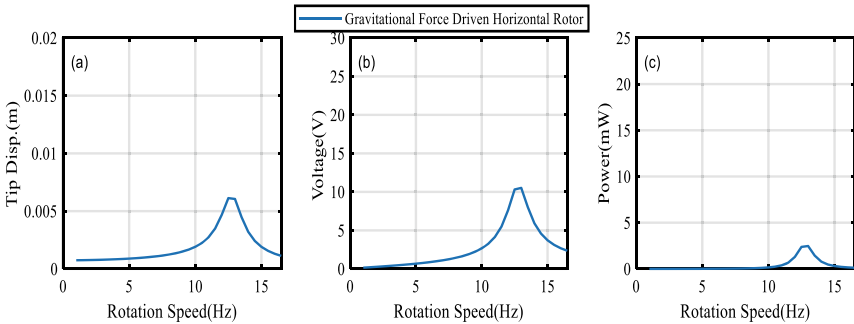


Fig. 12 Predicted outputs for different rotation speeds under gravitational force alone

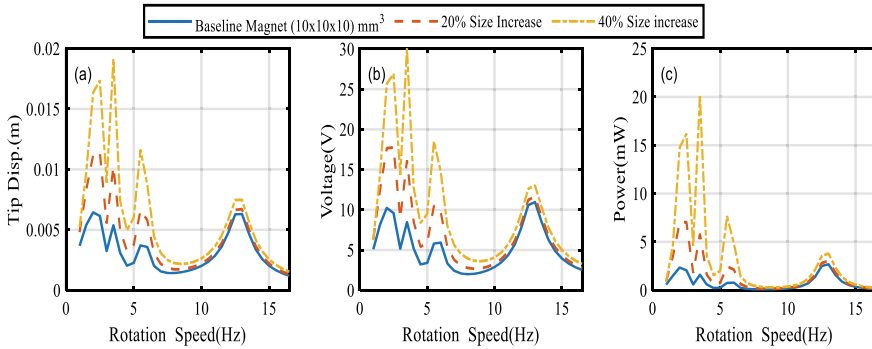


Fig. 13 Predicted outputs for different rotation speeds under both gravitation and magnetic forces

5 Conclusion

The novel prototype energy harvester has been proposed for accommodating ORS applications. Its dynamic responses and electric outputs are studied numerically using a validated lumped parameter model. Analysing the simulation results has led to main findings as follows:

- In addition to the expected high outputs at resonances, the proposed structures can also have higher output performance in low frequency range—about 50% of the 1st resonant frequency owing to the impulsive inputs.
- Large size of magnets with weak magnetic strength are a more effective approach to reducing low frequency vibrations while widening the frequency bands for effective harvesting. In contrast, placing stronger magnet close to the beam may lead too high dynamic responses of the beam, which can cause damages to the piezoelectric beam.
- In addition, the harvester can also operate in horizontal rotators with enhanced output at resonances due to gravitational effect.

References

1. Feng, G., Hu, N., Mones, Z., Gu, F., Ball, A.D.: An investigation of the orthogonal outputs from an on-rotor MEMS accelerometer for reciprocating compressor condition monitoring. *Mech. Syst. Signal Process.* **76–77**, 228–241 (2016)
2. Mones, Z., Zeng, Q., Hu, L., Tang, X., Gu, F., Ball, A.D.: Planetary gearbox fault diagnosis using an on-rotor MEMS accelerometer. In: 2017 23rd International Conference on Automation and Computing (ICAC), pp. 1–6 (2017)
3. Li C., Li, B., Gu, L., Feng, G., Gu, F., Ball, A.D.: Online monitoring of a shaft turning process based on vibration signals from on-rotor sensor. In: 2020 3rd World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM), pp. 402–407. IEEE (2020)

4. Han, H., Lin, Y., Gu, L., Xu, Y., Gu, F.: Vibration analysis based condition monitoring for industrial robots. In: *International Conference on Maintenance Engineering*, pp. 186–195. Springer, Cham, 2020.
5. Tang, X., Wang, X., Cattley, R., Gu, F., Ball, A.D.: Energy harvesting technologies for achieving self-powered wireless sensor networks in machine condition monitoring: a review. *Sensors* **18**(12), 4113 (2018)
6. Calautit, K., Nasir, D.S., Hughes, B.R.: Low power energy harvesting systems: state of the art and future challenges. *Renew. Sustain. Energy Rev.* **147**, 111230 (2021)
7. Alex, W.: Survey of energy harvesting technologies for wireless sensor networks. *IEEE Access* **9**, 77493 (2021)
8. Jiang, J., Liu, S., Feng, L., Zhao, D.: A review of piezoelectric vibration energy harvesting with magnetic coupling based on different structural characteristics. *Micromachines* **12**(4), 436 (2021)
9. Toh, T.T., Mitcheson, P.D., Holmes, A.S., Yeatman, E.M.: A continuously rotating energy harvester with maximum power point tracking. *J. Micromech. Microeng.* **18**(10), 104008 (2008)
10. Wang, Y.-J., Chen, C.-D., Sung, C.-K.: Design of a frequency-adjusting device for harvesting energy from a rotating wheel. *Sens. Actuators, A* **159**(2), 196–203 (2010)
11. Wang, Y.-J., Chen, C.-D., Sung, C.-K.: System design of a weighted-pendulum-type electromagnetic generator for harvesting energy from a rotating wheel. *IEEE/ASME Trans. Mechatron.* **18**(2), 754–763 (2012)
12. Wang, Z., Wang, W., Fengshou, Gu., Wang, C., Zhang, Q., Feng, G., Ball, A.D.: On-rotor electromagnetic energy harvester for powering a wireless condition monitoring system on bogie frames. *Energy Convers. Manage.* **243**, 114413 (2021)
13. Guan, M., Liao, W.-H.: Design and analysis of a piezoelectric energy harvester for rotational motion system. *Energy Convers. Manage.* **111**, 239–244 (2016)
14. Deng, L., Jiang, J., Zhou, L., Zhang, D., Fang, Y.: Design and simulation of a frequency self-tuning vibration energy harvester for rotational applications. *Microsyst. Technol.* **27**(7), 2857–2862 (2021)
15. Zhou, S., Cao, J., Erturk, A., Lin, J.: Enhanced broadband piezoelectric energy harvesting using rotatable magnets. *Appl. Phys. Lett.* **102**(17), 173901 (2013)
16. Zou, H.-X., Zhang, W.-M., Li, W.-B., Wei, K.-X., Gao, Q.-H., Peng, Z.-K., Meng, G.: Design and experimental investigation of a magnetically coupled vibration energy harvester using two inverted piezoelectric cantilever beams for rotational motion. *Energy Convers. Manage.* **148**, 1391–1398 (2017)
17. Wu, W.-H., Kuo, K.-C., Lin, Y.-H., Tsai, Y.-C.: Non-contact magnetic cantilever-type piezoelectric energy harvester for rotational mechanism. *Microelectron. Eng.* **191**, 16–19 (2018)
18. Fu, H., Yeatman, E.M.: Rotational energy harvesting using bi-stability and frequency up-conversion for low-power sensing applications: theoretical modelling and experimental validation. *Mech. Syst. Signal Process.* **125**, 229–244 (2019)
19. Wang, Z., He, L., Zhang, Z., Zhou, Z., Zhou, J., Cheng, G.: Research on a piezoelectric energy harvester with rotating magnetic excitation. *J. Electron. Mater.* **50**(6), 3228–3240 (2021)
20. Zhang, Y., Zheng, R., Shimono, K., Kaizuka, T., Nakano, K.: Effectiveness testing of a piezoelectric energy harvester for an automobile wheel using stochastic resonance. *Sensors* **16**(10), 1727 (2016)
21. Rui, X., Zhoumo Zeng, Yu., Zhang, Y.L., Feng, H., Huang, X., Sha, Z.: Design and experimental investigation of a self-tuning piezoelectric energy harvesting system for intelligent vehicle wheels. *IEEE Trans. Veh. Technol.* **69**(2), 1440–1451 (2019)
22. G. Akoun and J. - Yonnet, “3D analytical calculation of the forces exerted between two cuboidal magnets. *IEEE Trans. Magnet.* **20**(5), 1962–1964. doi: <https://doi.org/10.1109/TMAG.1984.1063554> (1984)
23. Suresh, K., Shankar, K., Sujatha, C.: A lumped electromechanical model for output power from partially covered piezoelectric energy harvester. In *2019 IEEE 2nd International Conference on Power and Energy Applications (ICPEA)*, pp. 94–98. IEEE, April (2019)

24. Valle, J., Fernández, D., Madrenas, J.: Closed-form equation for natural frequencies of beams under full range of axial loads modelled with a spring-mass system. *Int. J. Mech. Sci.* **153**, 380–390 (2019)
25. Al-Ashtari, W., Hunstig, M., Hemsel, T., Sextro, W.: Frequency tuning of piezoelectric energy harvesters by magnetic force. *Smart Mater. Struct.* **21**(3), 035019 (2012)

Image-Based 3D Shape Estimation of Wind Turbine from Multiple Views



Minghao Huang, Mingrui Zhao, Yan Bai, Renjie Gao, Rongfeng Deng, and Hui Zhang

Abstract This paper addresses the problem of reconstructing depth and silhouette images of wind turbine from its photos of multiple views using deep learning approaches, which aims for wind turbine blade fault diagnosis. Some previous multi-view based methods have extracted each photo's silhouette and combined them into separate channels as the input of convolution; others use LSTM to combine a series of views for reconstruction. These approaches inevitably need a fixed number of views and the output result is divergent if the order of the input views is changed. So, we refer to a network, SiDeNet (Wiles and Zisserman, Learning to predict 3d surfaces of sculptures from single and multiple views. *Int J Comp Vision*, 2018), which has a flexible number of input views and will not be affected by the input order. It integrates both viewpoint and image information from each view to learn a latent 3D shape representation and use it to predict the depth of wind turbine at input views. Also, this representation could generalize to the silhouette of unseen views. We make the following contributions to SiDeNet: improving the resolution of predicted images by deepening network structure, adopting 6D camera pose to increase the degrees of freedom of viewpoint to capture a wider range of views, optimizing the loss function of silhouette by applying weights on edge points, and implementing silhouette refinement with point-wise optimizing. Additionally, we conduct a set of prediction experiments and prove the network's generalization ability to unseen views. Evaluating predicted results on a realistic wind turbine dataset confirms the high performance of the network on both given views and unseen views.

M. Huang · M. Zhao · Y. Bai · R. Gao · H. Zhang (✉)
Programme of Computer Science and Technology, BNU-HKBU United International College,
Zhuhai 519087, China
e-mail: amyzhang@uic.edu.cn

M. Huang
e-mail: n830026049@mail.uic.edu.cn

R. Deng
School of Industrial Automation, Beijing Institute of Technology, Zhuhai 519088, China

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield
HD1 3DH, UK

Keywords 3D representation · Multi-view reconstruction · Silhouette prediction · Depth prediction · Wind turbine dataset

1 Introduction

Wind energy is one of the most technologically mature and widely used renewable energy sources in the world. The number and usage of wind turbines has increased rapidly, and the frequency of accidents has also increased. At the same time, blade failures is the most expensive core component in wind turbines, accounting for more than 20% of the total number of failures. In order to reduce the economic losses caused by this, new and effective wind turbine blade monitoring methods have become extremely important. Generally, sensors are installed on the wind turbine blade for fault diagnosis [1]. Due to the limitations and deficiencies of the sensor, the quality of the detection results mainly depends on the sensor itself and the use of the sensor, and the results often conflict. Therefore, the computer vision technology that does not need to use the sensor has received more attention. In order to obtain the 3D data of the wind turbine blade, researchers stick many markers on the wind turbine blade [2]. However, this method always takes a lot of time and requires higher image resolution and acquisition speed. Meanwhile, in order to diagnose the shape and status of wind turbine blades, we need to reconstruct it from a single or multiple photos.

This work proposes wind turbine blade (WTB) fault diagnosis based on a real-time 3D reconstruction system. In order to implement this system, a series of continuous views of the wind turbine must be generated to form a global 3D reconstruction of the wind turbine. Therefore, our task is to predict depth images of wind turbines from distinct perspectives. Due to the lack of prior knowledge, it is difficult to infer the concavities of the unseen views on the basis of single or multi view reconstruction. An alternative way to predict new unseen views is to use silhouette. Inspired by the deep learning network SiDeNet, we can use a variable number of views as input to predict the depth of a given view and the contour of a new view. The network can learn 3D information from single or multiple views in order to infer 3D shapes by combining the information of each view into global information. On this basis, SiDeNet can also be extended to new invisible shapes.

We also notice that a higher resolution of predicted images can increase the clarity of details. And by incorporating camera pose as 6 degree of freedom parameters, the method can be more universality applied. As our real dataset consists of several frames captured in a fixed camera viewpoint, rather than multiple-view data, single-view data, or data that focuses on only a small range of viewpoints is collected, we conducted experiments related to the number of views, and revealed the importance of the rich viewpoints in training data to compensate for the small number of views that we actually collect. Finally, in order to improve the edge accuracy of the predicted silhouette, we apply an improved silhouette weighted loss, and implement

an optimized silhouette edge prediction method according to PointRend [3]. Then, a system that can generate a series of wind turbine views is well trained and evaluated.

Our work has the following contributions: (1) The encoder-decoder architecture and its loss function in SiDeNet are refined for learning 3D shapes and predicting depth and silhouette. (2) A dataset of complex wind turbine is synthesized, which shows that the learned 3D representation is sufficient for new view synthesis of a set of unseen objects with complex shapes and even textures. (3) By incorporating camera pose as 6° of freedom parameters, the method can be more universality applied.

This paper is organized as follows. Section 2 introduces basic knowledge of existing methods of several issues related to image-based reconstruction. Section 3 presents the main framework of SiDeNet and our improvements. Section 4 gives the description and analysis of proposed dataset. The results of conducted experiments are shown and evaluated in Sect. 5 which is followed by conclusions in Sect. 6.

2 Related Works

2.1 Sensors and Markers

The fault detection of fan blades mainly relied on the use of sensors [4] and markers [5]. For the sensor, the detection technology was mainly based on vibration signals [30]. However, under complex operating conditions, the signal is not sensitive to surface fault detection and is susceptible to environmental factors. Another detection technology based on acoustic emission technology requires densely arranged sensors on the surface of the blade [6]. This method is often limited by the acoustic emission sensor, and its orientation often affects the quality of the detection results.

Today's machine vision-based wind turbine blade detection can avoid the use of sensors, for example, Corten and Sabel [7] tried to use photogrammetry technology [8] to measure blades by presetting markers on wind turbine blades and towers. Ozbek [9] uses a system that includes 4 CCD cameras and a high-power flashlight to measure the working status of the fan through marked points. Poozesh [10] adds markers on the surface of the blade and uses a bunch of stereo cameras to capture its geometry. Moreno [11] uses a vision-based deep learning method to automatically monitor each part of the blade surface using a camera installed on a robotic system to detect damage. Combining the above methods and the existing literature, in the blade structure failure detection method, markers are added on the blade surface. However, these methods often have long detection cycles, high costs, and high camera resolution requirements.

2.2 3D Reconstruction with Multi-images

The single image method needs to apply a priori as a constraint to reconstruct 3D information, because the 3D shape of the model cannot be inferred from the feature correspondence between multiple images. For example, the prior may be class-based modeling for deviation from the average shape. This method was first proposed by Blanz and Vetter [12]. Another application is to use texture or lighting priors to recover various complex 3D shapes [13]. But this method that requires additional information as a priori is not suitable for our project.

The self-occlusion problem can be overcome by providing multiple viewpoints. Several methods to integrate information from different angles are proposed. A classic method is that given multiple views of an object, a 3D shape can be generated by combining the characteristic points of these views using structure-from-motion [14]. Our framework based on Wiles and Zisserman's SiDeNet [15] is optimized, using max-pooling to combine feature vectors extracted from multiple views into latent feature encoding, which can be used for the depth and silhouette prediction of each view.

2.3 Silhouette

The initial method [16] used a series of silhouette images with known camera extrinsics to predict the visual hull, which was achieved by using voxels in the 3D representation. This is an improvement over other traditional methods that take into account reconstruction shapes with certain geometric and photometric constraints [17]. When testing constraints on shapes, they require multiple views and cannot infer an invisible view of the object. In our work, we construct a latent 3D representation by combining the encoding of multiple views. Given the input of the viewpoint information, the representation can be generalized to the silhouettes of the invisible part of the object.

2.4 Depth

The traditional depth estimation method has the problem of matching errors in scenes with drastic changes in lighting conditions [18]. Yao [19] proposed an end-to-end depth estimation framework based on deep learning, which improves the accuracy of depth estimation by restoring the dense structure of scenes from multiple perspectives with a certain degree of overlap. Chen [20] used the predicted depth information combined with ground truth to form a three-dimensional point cloud, and then used the 3D point cloud algorithm to optimize the depth regression. However, since the generation of 3D point clouds requires a lot of resources, we adopt a more efficient

method, which is based on U-Net [21] and Pix2Pix [22] using 2D images for depth recovery.

3 Proposed Method

Inspired by the architecture of SideNet, we use a framework includes two branches, i.e., silhouette prediction and depth prediction at given viewpoints. The 6D pose [23] of the wind turbine refers to the translation and rotation of the camera coordinate system relative to the world coordinate system, including the three rotation angles of the three camera directions and the camera position vector ($t = [x, y, z]$).

3.1 Main Architecture

The overall network structure is Encoder-Decoder, which is classical in U-Net, as shown in Fig. 1. The encoder is implemented using convolutional layers. The layer parameters and design are based on the encoder of the pix2pix architecture [22] and the U-Net architecture [21]. Given a set of images from each viewpoint, the encoder takes these images I_i and the corresponding viewpoint parameters, including translation and spatial rotation (I_i, t_i, R_i) as input, and then encodes them and broadcast over the feature channels. After convolution, the feature vector of each viewpoint is combined into a single latent vector x . x incorporates features from each viewpoint,

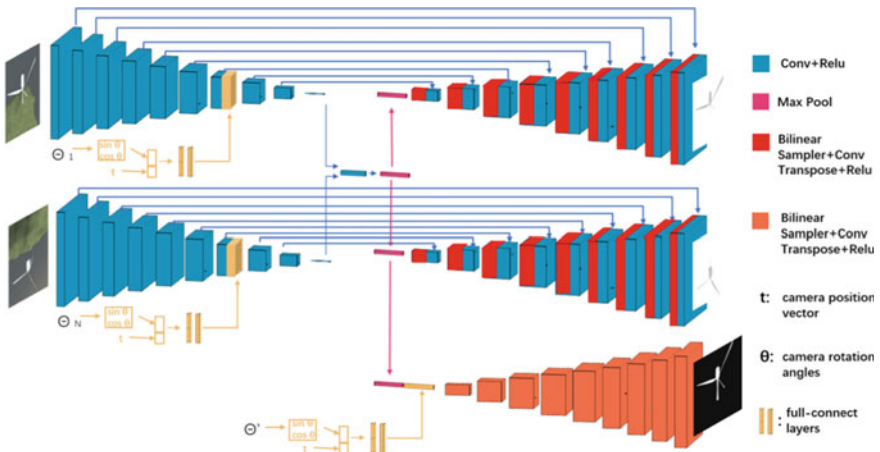


Fig. 1 The architecture of the network. Weights are shared across encoders. The blue arrows denote concatenation (over the feature channels). The feature vectors are combined to form a feature encoding (indicated by pink blocks)

including image and viewpoint parameters. Then through max pooling, x represents the most “confident” features of each viewpoint, and can be used for depth and silhouette prediction using the decoder. The decoder of each viewpoint depth prediction takes x as an input to recover the surface concavities of the object with depth $d_1 \dots d_N$ in each given viewpoint, by using a transposed convolution and up-sampling layer with skip connections (taken from the corresponding input branch for feature reuse). The layer parameters also come from the pix2pix and U-Net. Similarly, the silhouette decoder uses x as input to predict the silhouette S at a new viewpoint (t_i, R_i) , with a feature channel broadcasting of new viewpoint information. The layers in the silhouette decoder are the same as those in the depth decoder without skip connections.

In order to have a more flexible depth estimation, we set the convolution encoder with camera pose parameters with 6 degrees of freedom into two full-connect layers [24]. The broadcasting operation is the same as before. Specifically, we first calculate the sin and cos of the 3 camera rotation angles, combine them with the camera position vector (t), and pass them through two full-connect layers. The output is concatenated to each image encoder. These two layers can help reduce noise in camera poses and encourage the network to work well on each view. By incorporating the camera pose into the feature channel of the image encoder, we enable the network to learn information about the features of the wind turbine from specific views.

3.2 Silhouette Refinement

Silhouette estimation is always one of the most difficult steps in image segmentation. In order to obtain a good silhouette prediction, we adopt the method of PointRend [3], which effectively improves the accuracy of the silhouette. Basically, PointRend performs point-based segmentation prediction at adaptively selected locations based on an iterative subdivision algorithm.

In Fig. 2, we add a coarse-to-fine rendering module between the last up-sampling layer and the output result on selected feature points, while original up-sampling remains unchanged. The module selects a set of points (red dots) from the output feature map according to the adaptive subdivision algorithm and makes prediction for each point independently with a small MLP [25] and then inserts these predicted results to the corresponding points of the output silhouette image. The result in Fig. 3 shows the improvement in silhouette edge clarity.

3.3 Loss Functions

This section introduces the use of multi-task loss functions, using binary cross-entropy loss to predict silhouettes and MSE loss to predict depth.

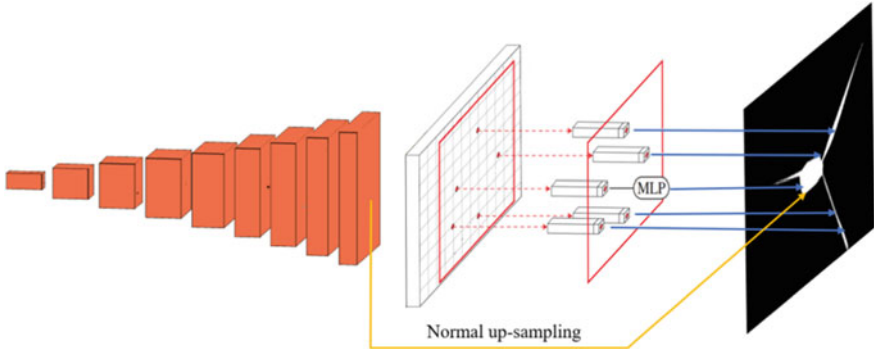


Fig. 2 The coarse-to-fine rendering module of PointRender is a sub-branch between the last layer and output image where the corresponding points are refined and inserted. This achieves a point-wise refinement of uncertain feature points and improves the silhouette

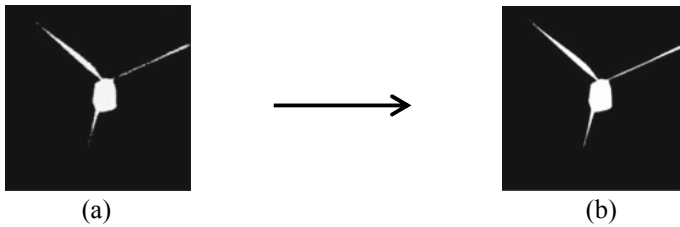


Fig. 3 Improvement of predicted silhouette. **a** Silhouette without point-wise refinement. **b** Silhouette with point-wise refinement. The result shows that the accuracy has improved

Depth Loss. The mean value of the difference between the target depth and the predicted depth is simply the absolute difference of all pixels of the target and predicted depth divided by the number of pixels. This operation allows the model no longer guess the absolute position of the object. It reduces ambiguity [15].

$$L_{depth} = \sum_{i=1}^N |d_i - d_{i_{gt}}| \tag{1}$$

where d is the predicted depth, d_{gt} is the ground truth of depth, N is the pixel number.

Silhouette Loss. Silhouette is a binary mask, i.e., 0 or 1. The binary cross-entropy is used to represent the average error of all pixels, which is also called the error prediction rate of all pixels. We use an improved loss: the use of parameter can weight more on the edge of silhouette, so that the silhouette edge can be predicted more accurately [15].

$$L_{sil} = \sum_{i,j} w_{i,j} \left(S_{i,j}^{gt} \log(S_{i,j}) + (1 - S_{i,j}^{gt}) \log(1 - S_{i,j}) \right) \quad (2)$$

$$w_{i,j} = \begin{cases} dist_{i,j}, & \text{if } dist_{i,j} \leq T \\ c, & \text{otherwise} \end{cases}$$

where S is predicted silhouette, S^{gt} is ground truth of silhouette, i, j are position of pixel, c is default constant weight, T is default threshold of distance.

3.4 Interpretability

First, in the Encoder structure, the image is input to the convolutional layer of each encoder. In a certain layer, the corresponding viewpoint information composed of the 6D camera extrinsic is encoded and broadcast over the feature channels of each input convoluted image. Then through the convolutional layer, the broadcast feature channel of the viewpoint information is integrated into the original feature channels of the image. At the end of each encoder, a $512 \times 1 \times 1$ feature vector is generated. All these feature vectors are concatenated into $512 \times N \times 1$ feature maps (N is the number of input viewpoints). Then through max pooling, the largest one is selected among the 512 mappings of size $N \times 1$, and the output is a $512 \times 1 \times 1$ feature vector x . x incorporates the largest of the N elements of each feature map, which means that it contains some viewpoint and image information. Among them, each viewpoint of the 512 features is the most important, and contains the most confident features of each viewpoint. Therefore, x can encode the properties of 3D shapes useful for both depth prediction and silhouette prediction in new views.

Then for the decoder of depth branch, skip connections and up-sampling layers with transposed convolution are used for depth reconstruction. Up-sampling helps to restore the resolution and pixels of the image, and skip connections help to reuse the features of the image in the encoder and recover its information. For the silhouette branch, x is input together with a feature channel broadcasting of new viewpoint information, and starting from noise, the cost function is minimized by using binary cross-entropy loss. When input to the network, the reconstructed silhouette should give a feature vector x' that is the same as x . This is because among the viewpoint-dependent features in x , those who are confident in the viewpoints close to the new viewpoint will be given a larger weight, and then the parameters of these related units will be updated through back propagation, and the most prominent features in the new poses will be introduced slowly, and finally a reconstructed silhouette will be obtained.

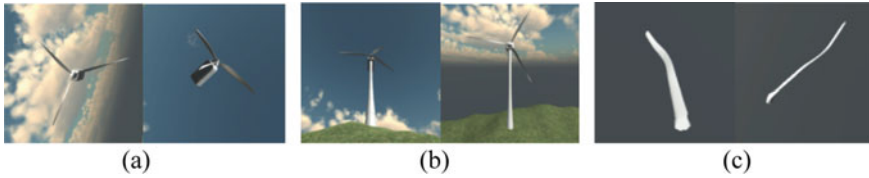


Fig. 4 Example images rendered from the dataset. **(a, b)** The wind turbine dataset. **(c)** The wind turbine blade dataset. These examples show various wind turbine models with rich shapes and textures, which helps the network generalize to real datasets

4 Dataset

Our work uses a series of wind turbine models to manually generate a data set, shown in Fig. 4. These models accurately simulate real wind turbines with realistic textures, including images and their corresponding pose information. Specifically, we align the center of the gravity and orientation of wind turbine itself with the world coordinate system ($(0,0,0)$ represent the center coordinates), and then randomly select target points on the surface of a sphere with a radius of 2 centered at $(0,0,0)$. Then images are generated with corresponding pose parameters.

As for rendering, the wind turbine model is normalized to a uniform size and 5~10 images of each wind turbine model are rendered from viewpoints uniformly and randomly selected within a specific range determined by the camera rotation angles and position vector, as the wind turbine is rotated about horizontal axis or vertical axis and has a variable distance from camera. In particular, the image sets of each model as data samples can cover the full-view of the wind turbine. These characteristics of our dataset help the network fully extract the information of various real turbine models from almost all possible views, and combine them to generate new views. In other words, the network parameters have been fitted to a wide range of views during training, so new views that may be close to the views already seen in training can be successfully generated in the test.

The dataset of wind turbine is divided into training set, validation set, and test set at the turbine level, and the proportions are 70%, 10%, and 20%, respectively. For each iteration, wind turbines are randomly selected, from which a subset of the 5–10 rendered views are selected.

5 Experimental Results

5.1 Improvement of Resolution

The result of depth/silhouette prediction is shown in Table 2. The prediction error of the results of different architectures of SiDeNet is shown in Table 1. The prediction

Table 1 The performance of different architectures

Model	Input/output size	Improved loss	Degree of freedom	Depth error	Silhouette error
SiDeNet 256 _{1dof}	256 × 256	×	1	0.078	0.101
SiDeNet 256 _{1dof}	256 × 256	✓	1	0.077	0.096
SiDeNet 256 _{6dof}	256 × 256	✓	6	0.072	0.093
SiDeNet 1024 _{1dof}	1024 × 1024	✓	1	0.061	0.065
SiDeNet 1024 _{6dof}	1024 × 1024	✓	6	0.063	0.064




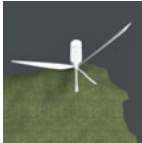


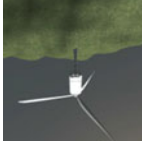

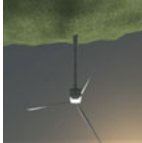


accuracy is high enough because compared with the ground truth, the loss of both depth and silhouette accurately lies in the range from 0.05 to 0.1, which is low enough for L1 loss and binary cross entropy loss, as explained with Eqs. (1) and (2). Therefore, the predicted images of depth/silhouette in Table. 2 show that with the loss smaller than 0.1, the depth will have very little error on each pixel's depth value compared with ground truth, and also silhouette will have a clear visual hull that is nearly the same as the one in ground truth because even though a few edge points are wrong predicted, they were refined with PointRend introduced in Sect. 3.

Table 1 compares the performance of SiDeNet 256 and SiDeNet 1024 with 1 or 6 degrees of freedom, as well as using an improved loss on silhouette prediction. This shows that although the difference in results driven by the increased degree of freedom is minimal, the converging speed during training is much faster because we have more feature variables related to the viewpoint, which helps the network encode the feature maps of multiple views more efficiently and perfectly. This helps to improve the robustness and compatibility of our multiple-view reconstruction task. In addition, along with the PointRend refinement, our weighted loss function improves performance. In SiDeNet 1024, we increase the resolution of the input image from 256 to 1024. This can improve the resolution of predicted result. As a result, the accuracy of both depth and silhouette is increased.

5.2 The Effect of the Number of Views

Training with more views can predict better than training with fewer views, and since the network sees more viewpoints during training, this can improve the prediction of those unseen views which are far from the range of predictable new views from input viewpoints. On this basis, when testing with a smaller number of views, the network can make use of information learned from the additional views. It shows that using fewer views than training will have certain negative impacts on the silhouette of

Table 2 Predicted images of all available views (0°~180°) of a wind turbine model

Viewpoint	Input views	Depth/Silhouette
0°		
45°	Nil	
60°		
80°	Nil	
120°		
150°		
180°	Nil	

For “Input views”, “Nil” means it is a new unseen viewpoint for silhouette prediction, otherwise there is an RGB image of the input view and this viewpoint is for depth prediction



Fig. 5 **a** The effect of view number in testing on performance. **b** The performance of different viewpoint choices

unseen new views, and have no effect on the depth of input views. Given the limitation of view number in real data set, we have to find the reduced number of views that will not greatly affect the prediction performance of silhouette. We completed this experiment by reducing the number of views in testing (6 views in initial training). The result is shown in Fig. 5a. Therefore, we know that an acceptable silhouette accuracy is 0.084 which is with 4 views (2 views reduced).

5.3 The Effect of Viewpoint Choice of Generating New Views

In order to test the viewpoint-encoding and generalization ability of SiDeNet, we control the viewpoint range of dataset by adjusting 6 degrees of freedom: only the horizontal rotation angle is changed in a range from 0° to 120° , and the other 5° of freedom are fixed. The experimental results of silhouette prediction of different viewpoint choices are shown in Fig. 5b. It shows that the network is able to generate new views with high accuracy when the new viewpoint is selected near the viewpoints used in training (Viewpoints from 100° to 140°). This infers that we have to broaden the range of views and increase the number of views during training, so that we can train a network with stronger generalization ability.

5.4 Predict Results

Based on the above experiment, we test the trained network using a data sample that contains 4 input views (0° , 60° , 120° , 150°) and predict the depth images at these 4 given views as well as the silhouette images at new unseen views (45° , 80° , 180°) as shown in Table 2, which reveals that adding more views to the training will more accurately generate new views close to the given view.

6 Conclusions

This paper introduces a multi-view 3D shape reconstruction system for wind turbine blade fault diagnosis. We make improvements on SiDeNet, in which the view-dependent encoders and feature combiner stimulates the network to integrate image information from all the input views. Therefore, the network can learn global feature that include the 3D shape features of all views, thereby being able to predict the depth of input views and generalize to the silhouette of unseen views. After verifying the prediction results, we obtain the following conclusions. Firstly, the increase in the degree of freedom of the viewpoint allows the network to take images of a wider viewpoint range as input, so that the shape information from more viewpoints can be learned, and the generalization ability of unseen views is improved. Secondly, the resolution increase makes the depth and silhouette images with more details. Third, the effectiveness of PointRend's adaptive points selection and point-wise refinement is remarkable. In addition, experiments have been conducted to prove that using more views in training, the network can use fewer views during testing and guarantee higher prediction performance. The experiments have also proved that the network more accurately predicts the silhouette of a new view that is close to the viewpoint of the views in training. These reflect the correlation and continuity between the image features of continuous viewpoints that the network can learn, which prove the generalization ability of the network to predict continuous views of the turbine. Finally, a series of consecutive views can be generated in high definition and combined into a reconstructed model.

Acknowledgements The work was supported by the National Natural Science Foundation of China (Project no. 62076029) and an internal funding from United International College.

References

1. Poozesh, P., Baqersad, J., Niezrecki, C., Avitabile, P., Harvey, E., Yarala, R.: Large-area photogrammetry based testing of wind turbine blades. *Mech. Syst. Signal Process.* **86**, 98–115 (2017)
2. Sarrafi, A., Mao, Z., Niezrecki, C., Poozesh, P.: Vibration-based damage detection in wind turbine blades using Phase-based Motion Estimation and motion magnification. *J. Sound Vib.* **421**, 300–318 (2018)
3. Kirillov, A., Wu, Y., He, K., Girshick, R.: PointRend: Image Segmentation As Rendering. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2020)
4. Xiaowei, L.: Research on fan Blade Detection Based on Fiber Grating Sensing Technology [D]. Nanjing University of Aeronautics and Astronautics (2013)
5. Wang, W., Chen, A.: Target-less approach of vibration measurement with virtual points constructed with cross ratios. *Measurement* **151**, 107238 (2020)
6. Zifeng, Q.: Research on Fan Blade Surface Damage Detection Based on Computer Vision [D]. Beijing Jiaotong University (2019)
7. Corten, G.P., Sabel, J.C.: Optical motion analysis of wind turbines. In: Proceedings of European Union Wind Energy Conference, Goteborg, Sweden, 1996 May 20–24 (1996)

8. Johnson, J.T., Hughes, S., van Dam, J.: A stereo-videogrammetry system for monitoring wind turbine blade surfaces during structural testing. *ASME Early Career Tech. J.* **8**(1), 1–1 (2009)
9. Ozbek, M., Rixen, D., Erne, O., Sanow, G.: Feasibility of monitoring large wind turbines using photogrammetry. *Energy* **35**(12), 4802–4811 (2010)
10. Poozesh, P., Baqersad, J., Niezrecki, C., Harvey, E., Yarala, R.: Full Field Inspection of a Utility Scale Wind Turbine Blade Using Digital Image Correlation. *CAMX, Orlando, FL*, vol. 10, issue no. 2.1, pp. 2891–2960 (2014)
11. Vučina, D., Ćurković, M., Novković, T.: Classification of 3D shape deviation using feature recognition operating on parameterization control points. *Comput. Ind.* **65**(6), 1018–1031 (2014)
12. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, pp. 187–194 (1999)
13. Barron, J., Malik, J.: Shape, illumination, and reflectance from shading. *IEEE Trans. Patt. Anal. Mach. Intell.* **37**, 1670–1687 (2015)
14. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press, USA (2003)
15. Wiles, O., Zisserman, A.: Learning to predict 3d surfaces of sculptures from single and multiple views. *Int. J. Comp. Vision* (2018)
16. Laurentini, A.: The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(2), 150–162 (1994)
17. Kolev, K., Klodt, M., Brox, T., Cremers, D.: *Continuous Global Optimization in Multiview 3D Reconstruction*, vol. 84, pp. 80–96. Springer (2009)
18. Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N.: Deeper depth prediction with fully convolutional residual networks. In: *Fourth International Conference on 3d Vision*. IEEE (2016)
19. Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L.: *MVSNet: Depth Inference for Unstructured Multi-view Stereo* (2018)
20. Chen, R., Han, S., Xu, J., Su, H.: Point-Based Multi-View Stereo Network, pp. 1538–1547. <https://doi.org/10.1109/ICCV.2019.00162> (2019)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention* (2015)
22. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017)
23. Xiang, Y., Schmidt, T., Narayanan, V., Fox, D.: Posecnn: A Convolutional Neural Network for 6d Object Pose Estimation in Cluttered Scenes (2017)
24. Kok, F., Charles, J., Cipolla, R.: Footnet: an efficient convolutional network for multiview 3d foot reconstruction. In: *Asian Conference on Computer Vision* (2020)
25. Yan, X., Yang, J., Yumer, E., Guo, Y., Lee, H.: Perspective transformer nets: LEARNING single-view 3D object reconstruction without 3D supervision. In: *Advances in Neural Information Processing Systems* (2016)

Health Status Assessment of Marine Diesel Engine Based on Testability Model



Ru Xiao, Guojun Qin, Zeyun Zhou, and Min Wang

Abstract To solve the problem of maintenance lag caused by long-term ocean voyage, a health assessment method is proposed based on the testability model for diesel engine. Firstly, the testability model is applied to generate the “fault-test” correlation matrix and accurately describe the interaction of each module and the fault signal propagation in the system structure; Then, the current health state corresponding to the bottom fault mode can be quickly deduced by using the model to infer the test information entropy; Finally, the health status of the whole diesel engine is evaluated by mapping to the health status of the upper structure through the support vector machine algorithm. The method can be used to determine the maintenance requirements in advance and improve the accuracy of fault prediction.

Keyword Diesel engine · Health assessment · Testability model · Correlation matrix · Support vector machine

Due to long-time sea voyage and operation, the maintenance lag of ocean-going ships is inevitable, which brings great security risks. As the core of marine power plant, diesel engine’s health assessment is of great value for fault early warning and reducing the probability of failure.

In equipment health assessment, the “fault-test” correlation matrix, as the basis of fault reasoning, is mainly used to describe the interaction between various modules and the propagation relationship of fault signals. To establish correlation matrix, a testability model must be given firstly. At present, multi-signal flow graph method is usually adopted in testability modeling mostly [1, 2]. In this method, “test” data are

R. Xiao · G. Qin (✉) · Z. Zhou · M. Wang
Hunan Zhixin Technology Development Co., Ltd, Changsha, China
e-mail: qgjnudt@163.com

G. Qin
College of Electrical and Information Engineering, Hunan University, Changsha, China

G. Qin · M. Wang
College of Information and Mechatronics Engineering, Hunan International Economics University, Changsha, China

usually sampled from actual measurement points or sensors, and faults are detected and isolated by describing the logical relationship between fault and test. Most of the faults isolated by this method just belonged to the bottom layer fault units [3, 4]. Although this method has been widely used in electrical equipment [5, 6], the bottom layer fault modes based on fault-test information contained in existing testability models are not enough to estimate the health status of whole system due to the small quantity of test points for complex mechatronical systems. Therefore, a health state assessment method based on testability model is proposed in this paper. The health state of the system is obtained by reasoning based on hierarchical state model and input test data. The method is also verified and applied in marine diesel engine.

1 Testability Model Construction

To construct testability model, multi-signal flow graph method is used firstly, which mainly includes the establishment of system structure, fault mode, test, signal and directed edge. Then a data model of state relationship between different levels is established. As shown in Fig. 1, the system structure tree is constructed in a top-down way. Each node in the tree represents a module, and the hierarchy of each module is marked by hierarchy labels such as system, subsystem, LRU and failure mode. In the realization by software, the same module node can be used as a page, and its sub module nodes are linked with its parent page in a directed edge way. Each test has many properties, such as method, type, output (result), cost, steps, etc. The test output has two states of “pass” or “fail”.

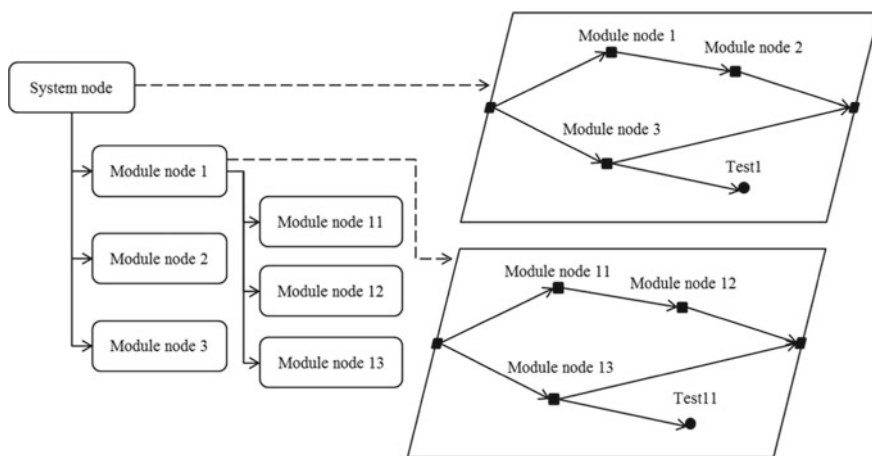


Fig. 1 Testability model construction

On this basis, the signals representing the characteristic, state or attribute set parameters [7] of the system or its constituent units can be attached to each module and test to realize the quantitative description of the object parameters or the qualitative description of the functional state. In the process of modeling, the same signal which is transmitted between different unit modules must be the same name, so as to avoid the error of signal transmission relationship and dependency relationship; Signals must be independent on each other, so as to ensure the simple and correct modeling process.

2 Entropy-Based Fault Reasoning

Based on the testability model, the correlation matrix can be calculated first, and then the information entropy method can be used to infer the state of the bottom fault mode.

2.1 Calculation of Correlation Matrix

Correlation matrix can be calculated by reachability algorithm, in which the search is begun with a failure mode. Firstly, along the flow direction of directed edge, it searched the past paths, and recorded the measuring points it passed, until it cannot flow, the traversal ends or returns to the node it has passed. Then, a correlation matrix FT [8] (as shown in Eq. (1)) is generated by comparing whether the failure mode and the tested signal have dependency relationship.

$$FT_{MN} = \begin{bmatrix} ft_{11} & \cdots & ft_{1N} \\ \vdots & \ddots & \vdots \\ ft_{M1} & \cdots & ft_{MN} \end{bmatrix} \tag{1}$$

Where, the row represents the failure mode f_1, \dots, f_M , column is the test $t_1 \dots t_N$. In the FT matrix, if test t_j is caused by fault f_i , this means $f_{ij} = 1$, otherwise $f_{ij} = 0$. The row vector $F_i = [ft_{i1}, ft_{i2} \dots ft_{iN}]$ represents the performance of each test when f_i occurs. The column vector $T_j = [ft_{1j}, ft_{2j} \dots ft_{Mj}]^T$ represents the possible fault mode when t_j occurs. In application, one fault has multiple tests, and the one test may be caused by the multiple fault modes. For failure mode f_i , all tests can be written as $T(f_i) = \{t_j | t_j \in T, ft_{ij} = 1\}$.

2.2 Bottom Fault State Reasoning

Taking correlation matrix as discrete information source, each object is regarded as a group of random objects, and the information entropy of information source is analyzed by adding failure rate. After the test is added into the information source matrix, the change of its information entropy can indicate the change of the system state, so as to judge whether there is a fault. Assume that for the discrete information source $D = d$, The information quantity defined by Shannon's information entropy is

$$I(d_i) = -P(d_i)\log_2 P(d_i) \quad (2)$$

where d_i represents the occurrence of failure mode f_i , while the probability of d_i appearing in D

$$P(d_i) = \frac{1}{N} fr_i \sum_{j=1}^N ft_{ij} \quad (3)$$

To distinguish the health status, the correlation matrix element ft_{ij} is changed: when the correlation is 1, $f_{ij} = 1$, otherwise $f_{ij} = -1$. N represents the number of tests, and fr_i represents the failure rate of the i failure mode.

The Shannon information entropy of the discrete information source can be obtained by

$$H(D) = - \sum_{i=1}^N P(d_i)\log_2 P(d_i) \quad (4)$$

The information entropy training of the discrete information source is as follows: Firstly, the information table of the information source is reorganized and the information entropy A is calculated under the condition of failure training of fault mode. Secondly, under the normal condition, the information table of the information source is reorganized and the information entropy B is calculated.

The information entropy is calculated after the test set is added into the information table. If the information entropy is A and basically unchanged, the fault mode is failure state; If the information entropy is B , it is in normal state; If the information entropy is between $(-A/2, B/2)$ and is not zero, the fault mode state is suspect; In other cases, the state of the fault mode is unknown. So, It is Fault state reasoning from tests to bottom fault modes.

3 System Health Assessment Based on Support Vector Machine

After the initialization of the health state of the bottom fault mode, the state mapping from the bottom fault mode is carried out layer by layer through the support vector machine algorithm to the top system node, so as to obtain the health state of the whole system.

Support vector machine (SVM) algorithm [9, 10] can maximize the classification interval between two samples of different categories by constructing an optimal classification surface. Multi-classification support vector machines [11, 13] can first construct several binary classification support vector machines and then combine them to solve the existence of multi-classification problems.

Suppose there are n classes of data training samples:

$$\{m_1^1, \dots, m_{t_1}^1, \dots, m_1^k, \dots, m_{t_k}^k, \dots, m_1^n, \dots, m_{t_n}^n\} \tag{5}$$

where $k = 1 \dots n$ is label for category, t_k represents the number of training samples of class k data. The total number of data training samples is $t_1 + t_2 + \dots + t_n$; In addition, Assume $f(\cdot)$ is a discriminant function on the sample space R^d , then the steps of constructing the n class (assuming $n = 4$) SVM classifier are as follows:

(1) Training of classifier

Firstly, the samples belonging to category 1 are regarded as positive samples, and the rest are negative samples, and the two classes SVM classifier $f_1(\cdot)$ is trained; Then, according to the same method, the samples belonging to category 2 are taken as positive samples, and the rest of the samples are taken as negative samples to train the two class classifier $f_2(\cdot)$. And so on, using the same method to train the second class classifier $f_3(\cdot)$; The combination of the three two class classifiers can form a four class classifier.

(2) Application of classifier

In the application, first of all, the actual data sample m_i is substituted into the second class SVM classifier $f_k(\cdot)$, and get its output $f_k(m_i)$. Then, find the category y corresponding to the maximum value in the function $f_k(m)$, $k = 1, 2, \dots, n$. The category y is corresponding to m_i :

$$y = \arg \max \{f_1(m_i), f_2(m_i), \dots, f_n(m_i)\} \tag{6}$$

Considering that there are four kinds of system health states in this paper: “fault”, “normal”, “suspected” and “unknown”, a multi-state classifier is constructed by using one to many method. The health state of the lower structure is imported into the multi fault classifier for calculation. If $y = 1, 2, 3, \text{ or } 4$, the health status of the upper structure corresponds to “fault”, “normal”, “suspected” and “unknown” respectively.

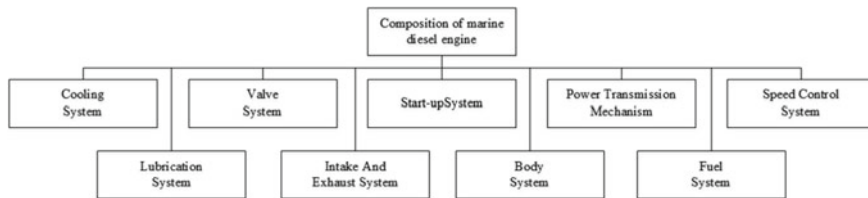


Fig. 2 Diesel engine system structure

4 Application of Marine Diesel Engine Health Assessment

4.1 Testability Modelling of Marine Diesel Engine

According to the structure of marine diesel engine (as shown in Fig. 2), the model of the system is constructed from top to bottom, and all modules are divided hierarchically according to system, subsystem, LRU and fault mode. As the only system layer, diesel engine is regarded as the root node of modeling, while nine subsystems are taken as the first level sub nodes, and the subsystems are divided step by step. For example, the cooling system is divided to many secondary sub-nodes, such as seawater pump LRU, fresh water pump LRU, heat exchanger LRU, thermostat LRU, etc. while thermostat LRU includes two failure modes: thermostat stuck and thermostat installed in the wrong direction. At the same time, each function module contains corresponding attributes, such as failure rate, maintenance cost, working time, port, additional signal, etc.

On the basis of hierarchical division, corresponding tests can be added at appropriate positions, such as oil pressure, starting detection, runaway, speed fluctuation, power detection, exhaust color, exhaust temperature, etc. Each test itself includes test type, method, cost, level, additional signal and other attributes. By connecting modules and between modules and tests with directed edges, the signal flow path can be formed, and the testability model can be constructed (as shown in Fig. 3).

4.2 Bottom Fault Mode Reasoning

The correlation matrix generated on the basis of the test item model is shown in Table 1. The row in the table represents the failure mode, the column represents the test, and the second column is the failure rate.

Firstly, the information entropy of each failure mode corresponding to different health states (failure, normal, suspected, unknown) is trained, and then the experimental test set (normal test and abnormal test) are input. On the basis of the above correlation matrix, the information entropy algorithm is applied to fault reasoning,

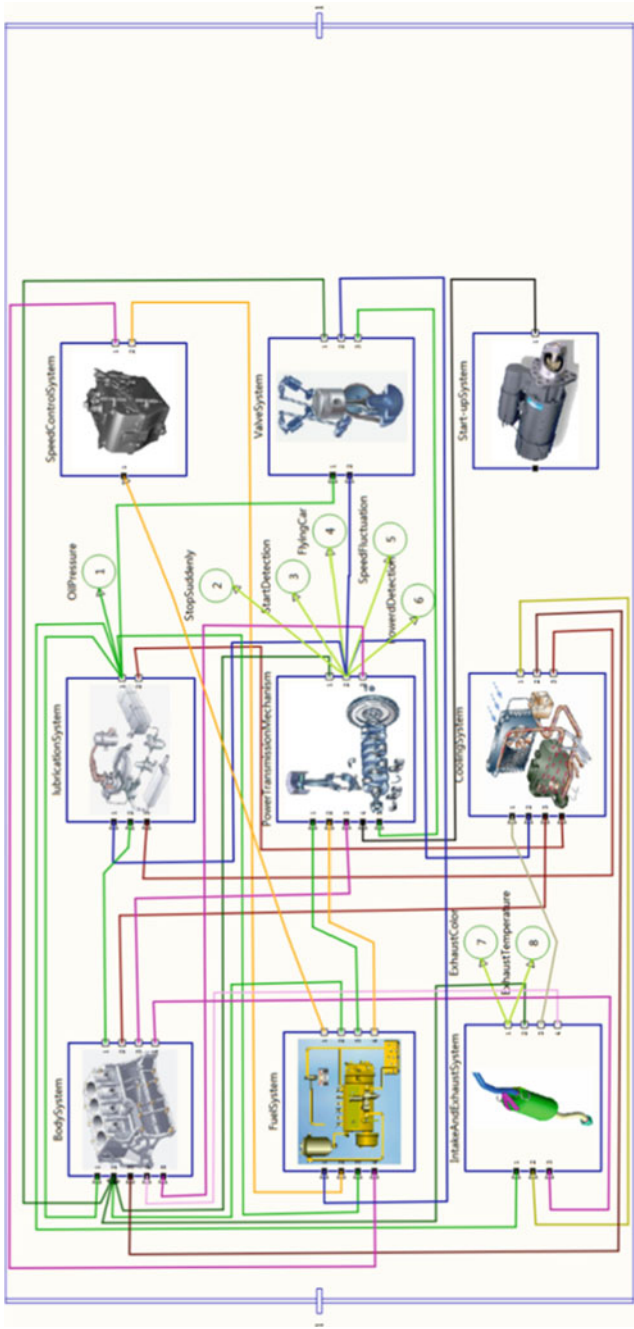


Fig. 3 Testability model of marine diesel engine system

Table 1 Correlation matrix

Fault mode	FR (1/h)	Abrupt stop	Start-up detection	Flying run	Speed fluctuation	Power sensing	Exhaust color
Belt	1e-06	0	1	0	0	0	0
Crankshaft	1e-06	0	0	0	0	1	0
Thermostat	1e-06	1	0	0	0	0	0
Regulating valve	1e-06	0	0	1	0	0	0
Seawater pump	1e-06	0	0	0	1	0	1
Wire way	1e-06	0	0	0	0	0	0
Brownout	1e-06	0	1	0	0	0	0
Valve leakage	1e-06	1	0	0	0	1	0
Low intake pressure	1e-06	0	0	0	0	0	0
Flying hammer	1e-06	0	0	0	1	0	0

and the fault occurrence status (fault, normal, suspected and unknown) of the fault mode can be obtained (as shown in Fig. 4).

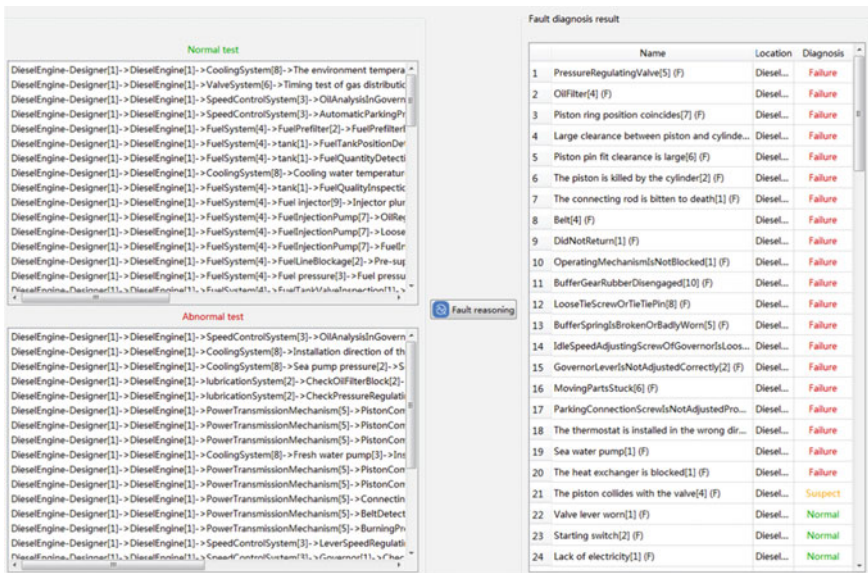


Fig. 4 Fault reasoning

4.3 Health Status Assessment of Marine Diesel Engine System

Taking the diesel engine system and its lower 9 subsystems as an example, the configured training matrix is shown as Fig. 5. In this figure, the first column of the lower table represents the health status of the diesel engine system, and the last nine columns represent the health status of the nine subsystems in the lower layer. Each row corresponds to a set of sample data, and the multi-dimensional sample data sets are trained to the mapping model of this layer by SVM algorithm. And so on, from the diesel engine system to the subsystem, to the LRU, and finally to the failure mode, all matrix training is carried out. After the training, the health state model is generated and saved.

Figure 6 shows the simulation results of health evaluation of actual diesel engine faults. The health status estimates for each level of the entire diesel system are shown on the left of level tree and the right side of health level chart, where red represents failure, green represents normal, yellow represents suspect, and gray represents unknown.



Fig. 5 Health state modeling based on SVM

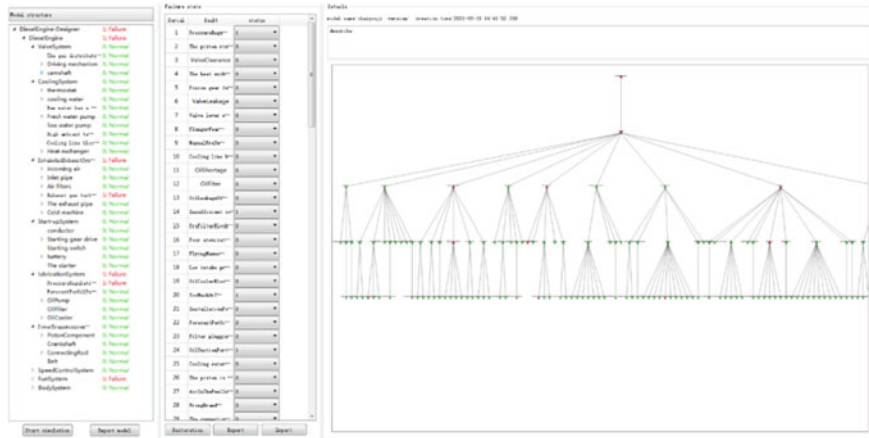


Fig. 6 Health status assessment of marine diesel engine system

5 Conclusion

At present, the commonly used testability models are mostly applied for the diagnosis of the bottom layer fault modes, and cannot obtain the health status of the whole system. Actually, bottom layer fault modes often have a direct impact on the health status of the upper components. In this paper, based on the testability modeling, information entropy of the bottom fault modes are firstly trained, and then the state mapping models of each level are trained by the support vector machine method from top to bottom. Finally, the overall health state model of the system is formed. The application of health assessment on Marine diesel engines shows that this method can better identify and predict the health status of complex mechanical and electrical equipment, which is expected to provide important reference information for maintenance decisions, so as to better control the potential failure risk.

Acknowledgements This paper is supported by the Key Research and Development Program from Hunan Province (No. 2018GK2073) and the Scientific Research Project from Hunan Education Department (No. 19B326).

References

1. Simpson, W.R., Sheppard, J.W.: System complexity and integrated diagnostics. *IEEE Des. Test Comp.* **8**(3), 16–30 (1991)
2. Peng, D., Yang, C., Sun, P., et al.: Testability modelling and optimization technology of ship electrical equipment based on multi-signal flow diagram. *IOP Conf. Ser. Mater. Sci. Eng.* **1043**(2), 022057, 11 (2021)

3. Lee, S.H., Kim, S., Kim, J.M., et al.: Fourier and wavelet transformations application to fault detection of induction motor with stator current. In: International Conference on Intelligent Computing. Springer-Verlag (2010)
4. Ocak, H., Loparo, K.A.: HMM-based fault detection and diagnosis scheme for rolling element bearings. *J. Vib. Acoust.* **127**(4), 299–306 (2005)
5. Liu, Q., Liang, T., Dinavahi, V.: Deep learning for hardware-based real-time fault detection and localization of all-electric ship MVDC power system. *IEEE Open J. Indus. Electron. Soc.* **1**(1), 194–204 (2020)
6. Al-Araimi, A.: Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Eng. Appl. Artif. Intell.* **7**(16), 657–665 (2003)
7. Qiguo, Y.: Modeling and analysis of testability of field replaceable modules based on multi-signal model. *Comp. Measure. Contr.* **24**(5), 35–38 (2016). (in Chinese)
8. Zhang, X., Li, Z.: A testability modeling method based on “functional fault transfer relation model. *Comp. Sci. Comput.* **37**(6), 44–47 (2019)
9. Shannon, C.E.: A mathematical theory of communication. *AT&T Tech. J. Acm Sigmobility Mob. Comput. Commun. Rev.* **5**(1), 3–55 (2001)
10. Liang, J., Shi, Z., Wierman, D.: Information entropy, rough entropy and knowledge granulation in incomplete information systems. *Int. J. Gen Syst* **35**(6), 641–654 (2006)
11. Kun, Z., Ying-Jie, T., Nai-Yang, D.: Unsupervised and semi-supervised two-class support vector machines. In: IEEE International Conference on Data Mining Workshops. IEEE (2006)
12. Chih-Wei, H., Chih-Jen, L.: A comparison of methods of multiclass support vector machines. *IEEE Trans. Neural Networks* **13**(2), 415–425 (2002)
13. Gjorgjevikj, D.: A multi-class SVM classifier utilizing binary decision tree. *Informatica* **33**(2), 225–233 (2009)

Modelling the Dynamics of a CNC Spindle for Tool Condition Identification Based on On-Rotor Sensing



Chun Li, Dawei Shi, Bing Li, Hongjun Wang, Guojin Feng, Fengshou Gu, and Andrew D. Ball

Abstract Cutting tool plays an important role in modern manufacturing industry, however, tool wear is unavoidable during machining which could reduce the efficiency. Aiming at studying an appropriate and efficient tool condition monitoring method to improve the accuracy of finished parts, the roughness of the turned surface, a novel On-Rotor Sensing (ORS) is installed on the rotating workpiece to obtain vibration signals. To get an in-depth understand of the vibration data, a multi-degree-of-freedom (MDOF) system consisted of spindle, chuck and workpiece is established and its multi-mode natural frequency is obtained by finite element model (FEM) method. It is found that the dynamic response of the spindle rotor determines machining accuracy in the turning process and shows that the first several modes in the frequency range within 2000 Hz are the main responses of the system, which can be effectively captured by the ORS. Especially, the spring stiffness is calibrated based on the FEM results and the accuracy of the dynamic modal responses of this model are verified when the mass of the workpiece decreases during the turning process. According to the results, two frequency bands are advocated for ORS based online monitoring of tool wear conditions.

C. Li · B. Li

School of Industrial Automation, Beijing Institute of Technology Zhuhai, Zhuhai 519088, China
e-mail: Chun.li@hud.ac.uk

C. Li · G. Feng · F. Gu (✉) · A. D. Ball

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK
e-mail: F.Gu@hud.ac.uk

G. Feng

e-mail: Guojin.Feng@outlook.com

D. Shi

College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China

H. Wang

School of Mechanical and Electrical Engineering, Beijing Information Science and Technology University, Beijing 100192, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

1057

H. Zhang et al. (eds.), *Proceedings of IncoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_84

Keywords Vibration · Tool wear condition · Finite element model · On-Rotor Sensing · Dynamic characteristics of spindle

1 Introduction

The spindle rotor system is the core component of the CNC machine tool and the related parts including supporting bearings, chuck and workpiece are also the key components and their dynamic behavior play a decisive role in the machining quality and cutting ability. It is pointed that the information such as vibration, cutting force and tool wear, could be identified by properly characterizing the spindle dynamic conditions [1]. An in-depth understanding of the dynamic characteristics of the spindle rotor system is essential to get the most use of the machine, achieve high performance quality and conduct effective condition monitoring and diagnosis [2].

Lots of researches have been conducted on the dynamic characteristics of the machine tool through modal analysis method for complex system [3]. Vibration mode of the spindle of CK6130 CNC lathe was analyzed though the finite element model (FEM) carried out by ANSYS Workbench [4]. Paper [5] gave a understanding of the material selection and preliminary design of the lathe spindle also by FEM. It is pointed out that the main factors that affect the spindle dynamics are the reassemblies of bearings and preload force in axial direction [6]. Considering vibration behavior during the cutting process, T. Schmitz studied a combination system of machine-spindle-holder-tool and showed that the holder-tool modes interact with the machine-spindle modes [7]. In addition, Jie et al. [4] established the coupling system of the inner and outer rotors of the spindle and calculated the critical speed and vibration modes. Through the co-simulation of ANSYS and ADAMS, the influence of cutting torque on spindle vibration was studied [8].

Furthermore, tool wear often gives rise to small but significant vibration amplitudes that affect the form and finish of machine products [9]. Moreover, even a small stochastic excitation can lead to stochastic resonance in the CNC machine system [10], which will lead to instability of machining process and even cause serious damage. Therefore, it is necessary to establish a coupled spindle rotor model and comprehensively study the dynamic response characteristics under different tool wear conditions.

Although significant development works has been completed about the effect of operation conditions [11] on spindle, such as the bearing [12–14], machine structures [15], cutting parameters [16, 17] and tool setting [18], relatively less effort has been expended to study the dynamic characteristic of the coupled spindle rotor system considering the effect of workpiece mass during the turning process.

Based on the description above, firstly, the FEM of a MDOF system consisted of spindle, chuck and workpiece is established and especially the effect of the reduce of workpiece mass on the natural frequency response of the coupled system is analyzed in Sect. 2. Secondly, the bearing stiffness and damping is calculated according to the empirical formula and adjusted through FEM results to meet the impact test results.

In Sect. 3, an experiment about the tool wear condition identification is set up in a CNC lathe system and two band-pass filters are designed to extract the vibration data obtained by a novel On-Rotor Sensing (ORS) installed on the rotating workpiece according to the FEM and test results. Finally, the tool wear states recognition are experimentally verified and some conclusions are stated.

2 FEM Analysis

2.1 Establishment of Finite Element Model

It is known that each component such as holders, tool, and spindle is coupled with each other in the spindle rotor system [7, 19]. To get a comprehensive understanding of the dynamic characteristics of the rotating system, a multi-degree-of-freedom (MDOF) system consisted of spindle, chuck and workpiece is established in this paper shown in Fig. 1.

Figure 1 shows that the spindle rotor system is supported by two main groups of angular contact ball bearings (ACBB) respectively, the front bearing group (workpiece clamping end), which is made up of three DBB-mounted NSK 7220A bearings. The rear bearing group (pulley end) is made up of two DB-mounted NSK 7218A bearing. The main parameters of the bearings are shown in Table 1. The two bearings at the front end are simplified into one since they are close to each other. Therefore, the front and rear bearings are simplified to two pairs of elastic supports respectively.

In general, the spindle, chuck and workpiece are usually modeled as rigid Timoshenko beams, whereas, the bearings assumed to contribute the flexibility and damping to the system [20]. However, the typical researches usually simplify ACBB to radial linear bearings and damping and ignore that in the axial direction [19–22]. Actually, ACBB not only provide the radial support but also the axial support. Taking this into consideration, 16 springs and damping in radial direction and 8 springs and damping in axial direction respectively are established for the FEM of CNC lathe spindle rotor system. One bearing is simplified into 4 radially distributed support

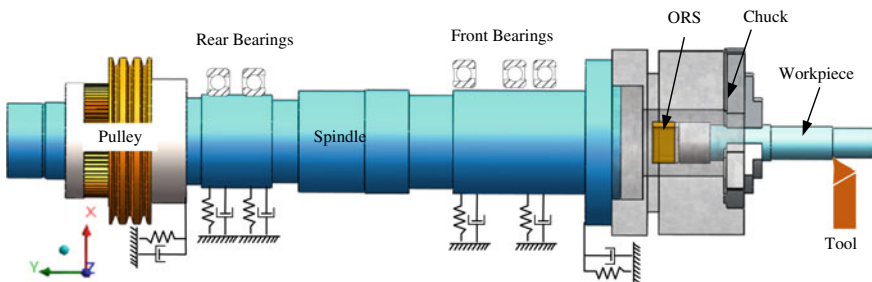


Fig. 1 The FEM of spindle rotor system

Table 1 Main parameters of bearing

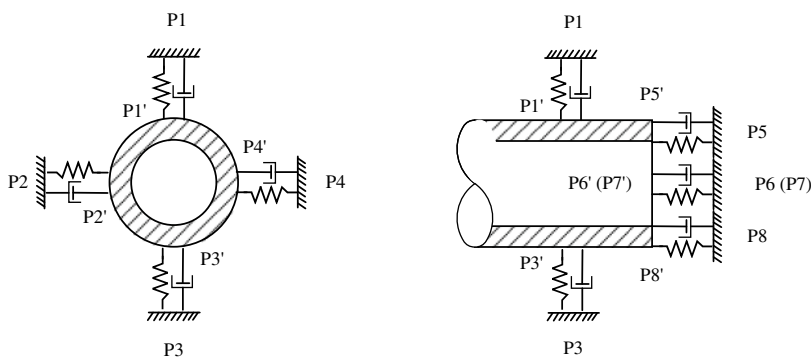
Bearing model	Installation mode	Inner diameter <i>d</i> /mm	Outside diameter <i>D</i> /mm	Width/mm	Circular diameter of roller <i>D_{mp}</i> /mm	Ball diameter <i>D_W</i> /mm	Number of balls (<i>Z</i>)	Cr/ <i>N</i>
NSK 7218A	Back-to-Back (DB)	90	160	30	125	18.26	21	1.18e3
NSK 7220A	Series connection and Back-to-Back (DBB)	100	180	34	140	20.64	21	1.44e3

springs and damping respectively and the axial support is represented by 4 axially supported springs and damping at each end of the support position as shown in Fig. 2.

As shown in Fig. 1, this system is basically axially symmetric, so the structure parameters in *X* and *Z* direction are assumed to be the same. Therefore, the corresponding dynamic equation of MDOF system excited by cutting force can be expressed as follow [23]:

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = K_c F(t) + F_\sigma(t) \tag{1}$$

where, *m* is the mass of the system, *c* is the structural damping coefficient and *k* is the stiffness coefficient. *F*(*t*) is the periodic total cutting force and *K_c* is the coefficient related to the workpiece’s material, heat treatment, cutting angles, depth of cut (DOC). *F_σ*(*t*) is stochastic cutting force and could excite stochastic resonance in the machine system. The corresponding natural frequency is $\omega_n = \sqrt{k/m}$ and the damping ratio is $\zeta = c/2(2m\omega_n)$, Eq. (1) can be normalized by the expression:



(a) Springs and damping in radial direction

(b) Springs and damping in l direction

Fig. 2 Simplified bearing supports

$$\ddot{x}(t) + 2\omega_n \zeta \dot{x}(t) + \omega_n^2 x(t) = \frac{K_c}{m} F(t) + \frac{1}{m} F_\sigma(t) \tag{2}$$

2.2 Bearing Stiffness and Damping Calculation

As for the stiffness of paired bearings cannot be determined simply by superimposing the stiffness of a single bearing, but requires consideration of the mounting combination. The axial stiffness and radial stiffness of a parallel paired bearing could be calculated by Eq. (3) [24] and the recalculation results are shown in Table 2.

$$K_r = 1.7164 \times 10^4 \times \cos^2 a \left(\frac{z^2 D_b F_a}{\sin a} \right)^{1/3} \text{ (N/mm)}$$

$$K_a = 6.866 \times 10^4 \times (F_a D_b z^2 \sin^5 a)^{1/3} \text{ (N/mm)} \tag{3}$$

where, F_{a0} is the bearing preload; Z is the number of balls in a bearing; α is the contact angle and D_w is the bearing ball diameter. For lightly preloaded bearings, the preload forces of the front and rear bearings are taken as 100 N and 370 N respectively according to the instruction manual of NSK [25]. Substituting the bearing parameters in Table 1 and preload forces into Eq. (3), the axial and radial stiffnesses could be calculated respectively as shown in Table 2.

Due to the complex dynamic properties of bearings, which are inevitably affected by vibration, thermal deformation and mating accuracy between the bearing and the spindle during operation [26], the bearing damping cannot yet be derived entirely by calculation. A widely used method is to obtain the bearing damping based on a correction of empirical values which is experimental result for certain bearing sizes and operating conditions. According to experimental measurements, the damping of a bearing with an inner diameter of 60 mm and an outer diameter of 95 mm is 3000 Ns/m for a lightly loaded interference fit. Other bearing needs to be multiplied by the dimension factor α_η defined in Eq. (4):

$$\alpha_\eta = \left(\frac{D_2 + d_2}{D_1 + d_1} \right)^{1.3} \left(\frac{D_1 - d_1}{D_2 - d_2} \right)^{1.9} \left(\frac{d_2}{d_1} \right)^{2.3} \tag{4}$$

Table 2 Stiffness of combination bearing

Bearing	Radial stiffness (N/mm)	Axial stiffness (N/mm)	Damping (N.s/m)
7218A (DB)	8.474×10^5	6.712×10^5	3.812×10^3
7220A (DBB)	13.653×10^5	10.813×10^5	4.356×10^3

where, d_1, D_1 is the inner diameter and outside diameter of the reference bearing and d_2, D_2 is the inner diameter and outside diameter of the actual spindle bearing. And then the bearing damping could be obtained and shown in Table 2.

2.3 Modal Analysis of Spindle Rotor System

The bearing stiffness and damping can be calculated from the above description. However, lots of factors such as installation preload, heat and vibration [27] generated during high-speed rotation in the cutting process will have significant effect on the bearing contact angle, so that the theoretical bearing stiffness obtained from the calculation will deviate from the actual value. Therefore, three different sets of bearing stiffness are set in this paper, as show in Table 3.

Furthermore, the workpiece material is continuously removed and thus its mass is reduced during the cutting process. To study the influence of the workpiece mass on the natural frequency of the spindle system, four different diameters, lengths and masses of workpieces are mounted on a three-jaw chuck and subjected to FEM and FRF experiments shown in Table 4. As can be seen that their diameters, lengths and masses are all different, with the mass and length of the workpiece Mass 4 being much greater than that of the other three workpiece, and as its mass is already large, which will have a greater effect on the spindle rotor's natural frequency induce from Eq. (2).

The natural frequencies obtained through FEM are draw in Figs. 3 and 4. It can be seen from Fig. 3 that for the Stiffness 1, the first 10 orders of the natural frequency with Mass1/Mass2/Mass3/Mas4 are almost the same, especially for that with Mass1, Mass2 and Mass3, where the maximum difference is less than 20 Hz. It can be further

Table 3 Different bearing stiffness sets

	Front bearing radial stiffness (N/mm)	Front bearing axial stiffness (N/mm)	Rear radial bearing stiffness (N/mm)	Rear axial bearing stiffness (N/mm)
Stiffness 1	13.653×10^4	10.813×10^4	8.474×10^4	6.712×10^4
Stiffness 2	13.653×10^5	10.813×10^5	8.474×10^5	6.712×10^5
Stiffness 3	13.653×10^6	10.813×10^6	8.474×10^6	6.712×10^6

Table 4 Workpiece parameters

Workpiece name	Diameter (mm)	Length (mm)	Mass (kg)	Material
Mass 1	22	37	0.2	#45 steel
Mass 2	28	144.5	1.2	#45 steel
Mass 3	65.5	97	5.6	#45 steel
Mass 4	45.5	735	19.2	#45 steel

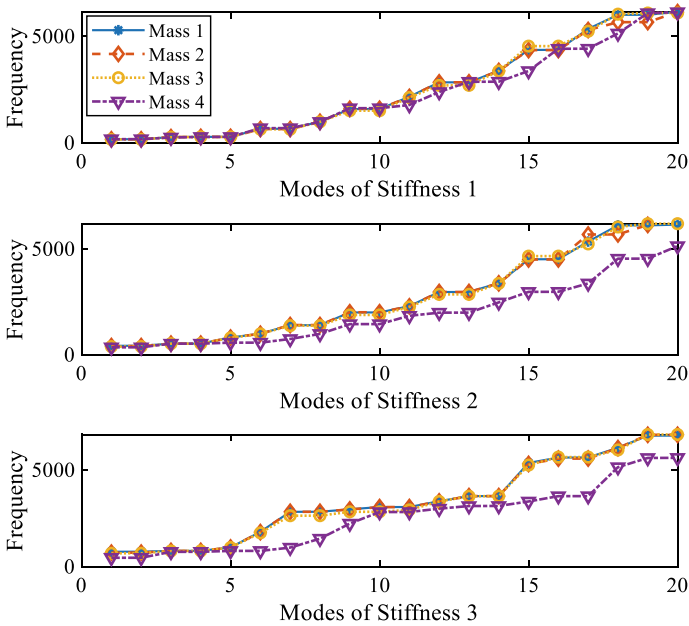


Fig. 3 Natural frequency of spindle system with different workpiece masses

deduced that for the lower bearing stiffness of Stiffness 1, the effect of workpiece mass could be negligible. However, as the stiffness of the bearing support increases, the effect of the workpiece mass on the system frequency becomes more and more significant. The modal behavior of the spindle system differs significantly between clamping the workpiece with the highest Mass 4 and the other three workpieces. Whereas, the modal behavior is basically consistent when clamping workpieces of Mass1, Mass2 and Mass3 at the same stiffness.

Figure 4 shows the effect of bearing stiffness on the natural frequency of a spindle rotor system with the same workpiece. It shows that it varies with different stiffnesses, in particular, the variation of Stiffness 3 differs significantly from that of Stiffness 1 and Stiffness 2, while the latter two have almost identical curves. For the first 10 orders, the natural frequency of Stiffness 2 is higher than that of Stiffness 1. However, the effect of bearing stiffness on the higher order is significantly reduced, especially for Stiffness 1 and Stiffness 2.

Furthermore, the torsional frequency for the same workpiece is exactly the same, only the location where it occurs is different, from which could be inferred that it is not related to the stiffness of the bearing, but rather to the mass of the workpiece. For example, for Mass 1, the 1st torsional frequency is 995.85 Hz at 9th vibration order of the system for Stiffness 1, also 995.85 Hz at 7th for Stiffness 2 and the same 995.85 Hz at 6th for Stiffness 3. As for Mass 4, the 1st torsional frequency is 986.5 Hz for all the three stiffnesses, which shows that the torsional vibration frequency of the spindle system is independent of the bearing stiffness and is only

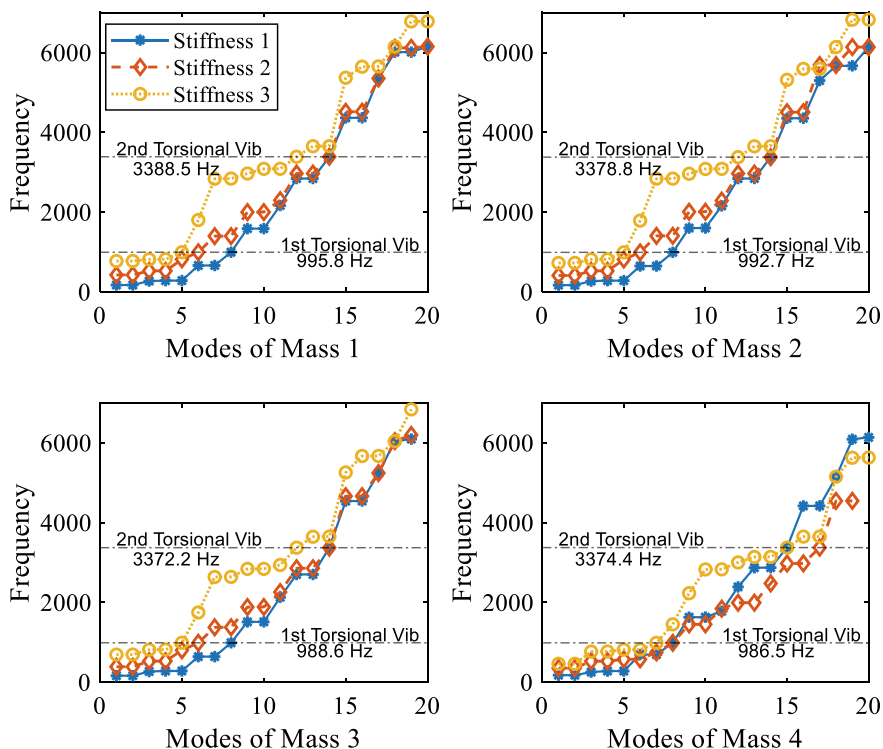


Fig. 4 Natural frequency of spindle system with different bearing stiffnesses

related to the mass of the workpiece, which increases slightly as the mass of the workpiece decreases. On the contrary, bearing stiffness has a greater influence on the other vibration modals.

2.4 Frequency Response Analysis

The dynamic behavior of a spindle can be most quickly obtained by measuring its frequency response function (FRF) [3]. To verify the simulation results of FEM, an impact hammer test is carried out and its frequency response is shown in Fig. 5. It can be seen that the variation of natural frequency with the decrease of workpiece mass is consistent with the results of FEM. Furthermore, there are four main resonance frequencies bands within 2000 Hz in this system, which are 244–440 Hz, 709–732 Hz, 1013–1056 Hz and 1465–2045 Hz respectively.

Bearing stiffness is a typical non-linear problem, its groove contact deformation, installation preload, spindle speed, heat, etc. will be the factors affecting the bearing stiffness [28]. Consequently, the bearing stiffness calculated by the empirical formula

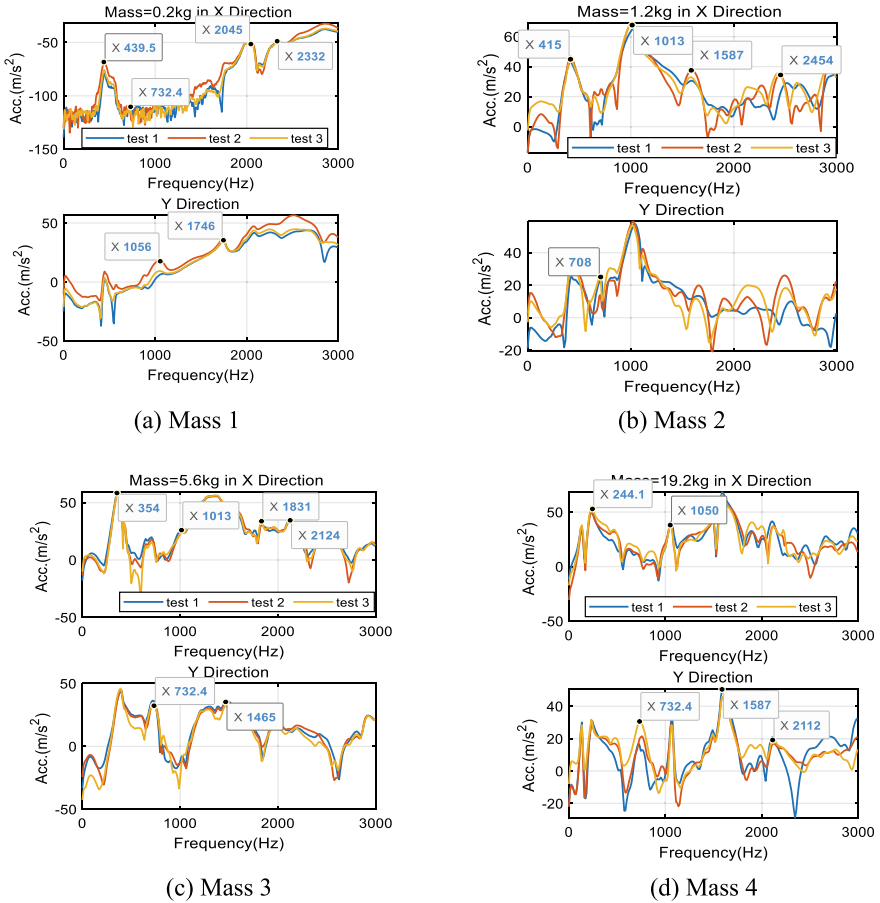


Fig. 5 Impact test results

has a certain discrepancy with the measured value, and it is difficult to measure once the spindle rotor has been installed. Therefore, the bearing stiffness is adjusted according to the impact test results in Fig. 5 and the adjusted stiffness and damping are shown in Table 5.

Table 5 Adjusted bearing stiffness and damping values

Bearing	Radial stiffness (N/mm)	Axial stiffness (N/mm)	Damping (N.s/m)
7218A	10.40×10^5	6.72×10^5	3.812×10^3
7220A	18.40×10^5	7.12×10^5	4.356×10^3

Table 6 The main natural frequencies of the spindle rotor system

Order	Natural frequency modal	Mass 1	Mass 2	Mass 3	Mass 4
1st	1st Rigid torsional	1.77e-02	2.46e-02	1.75e-02	1.68e-02
2nd, 3rd,	1st Front-end transverse vibration	432.59	417.38	391	354.91
4th, 5th,	1st Back-end transverse vibration	547.91	547.99	547.98	547.98
6th	1st Longitudinal vibration	739.83	735.86	723.88	677.66
7th	1st Torsional vibration	995.85	992.71	988.6	986.45
8th, 9th	1st Transverse vibration of whole system	1543	1547	1520.7	1557.4
10th, 11th	2nd Back-end transverse vibration	2085.3	2094.5	1956.3	2067.6

From FEM model, it could be inferred that there is only a rotation freedom (the 1st mode) in the system. And the spindle rotor is basically central symmetric, consequently, the natural frequencies of transverse bending vibration are the same in the *XOY* and *YOZ* plane and only the vibration modal are different in directions. The 6th and 7th results are single roots, and such these modes are longitudinal vibration and torsional vibration respectively [29]. It is pointed out that the fundamental modal frequencies of the machine system are basically within 2000 Hz [9], so only the frequency below 2000 Hz is concerned in this paper.

Based on the adjusted bearing stiffness, the natural frequency of the spindle rotor system with different mass workpieces is recalculated through FEM, and the first 11 orders (excluding the natural frequency of the workpiece) are shown in Table 6. It can be seen that there are mainly five vibration modes, namely transverse vibration at the front end (the clamping workpiece end), transverse vibration at the back end (the pulley end), longitudinal vibration, torsional vibration and transverse vibration of the whole system. The front-end transverse vibration frequency increases considerably as the mass of the workpiece decreases, which is consistent with the theoretical analysis, so as the torsional vibration. In contrast, the back-end transverse vibration frequency does not vary with the mass of the workpiece and the other modal frequencies do not change appreciably too. Consequently, when it needs to study the variation of the dynamic characteristic of the spindle rotor system during the cutting process, it is better to focus on the transverse vibration at the workpiece end.

Furthermore, the frequency below 2000 Hz of interest, with the four frequency bands of 354–433 Hz, 677–739 Hz, 986–995 Hz and 1520–1557 Hz respectively, which are basically consistent with the four main resonance frequencies bands of the FRF experimental results. The smaller differences are due to the errors of the FEM of the spindle rotor. The results show that the MDOF system established in this paper consisted of the workpiece, chuck and spindle is validated as an accurate FEM and could be used for further dynamic characteristic analysis in the machine tool.

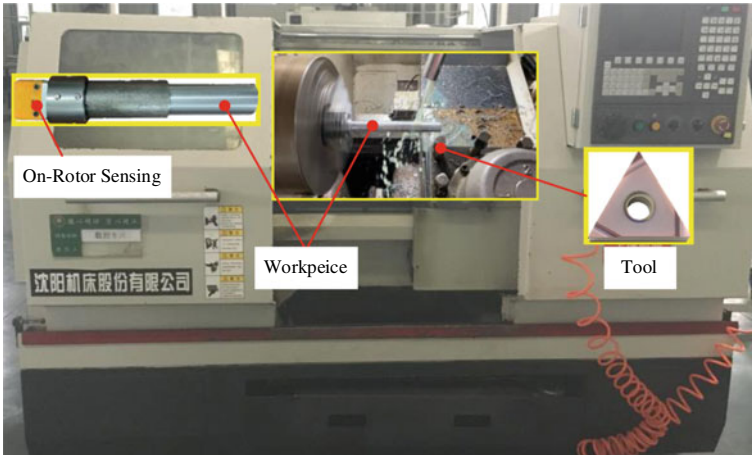


Fig. 6 Tool wear condition identification experiment rig

3 Experimental Verification

3.1 Experimental System Description

The turning experiments conducted on a CAK3665di CNC lathe shown in Fig. 6. The motor power is 4 kW, feed rate is 0.15 mm/r and spindle speed are set to 1000 r/min. The original diameter of the workpiece is 34 mm and cutting it to 28 mm. Two different depths of cut (DOC) are set, respectively 0.5 mm and 1.0 mm.

3.2 Tool Wear Condition Identification

Before the cutting experiment, the tool wear on the flank face is measured by an optical microscope at 29.2 times of magnification. The severity of wear on the flank wear width is denoted by VB amplitudes [30]. Figure 7a–c show the measurement results of the break-in wear insert, steady wear insert, and failure wear insert respectively.

3.3 Vibration Signal Characteristics from ORS

As shown in Fig. 1 that an On-Rotor Sensing (ORS) is mounted directly on the rotating workpiece to obtain the vibration signal during the turning process and such installation way is more sensitive to the dynamic characteristic of the spindle rotor

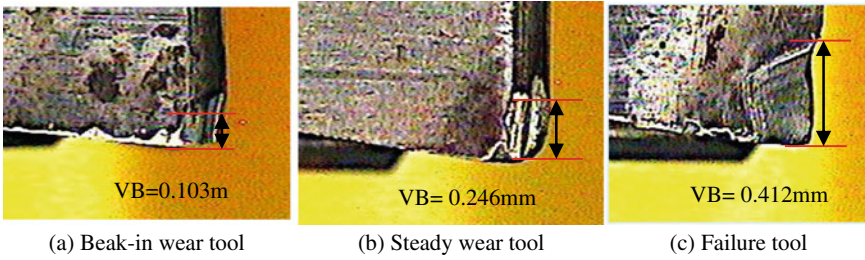


Fig. 7 Tool wear conditions

system [31]. According to the analysis results in Sect. 2, the acceleration signals from the ORS in *X* and *Z* directions represent tangential (torsional) vibration and transverse vibration respectively and both signals are related to the fluctuation of cutting force [32]. Therefore, only the data in *X* direction is extracted and shown in Fig. 8. Figure 8a shows the vibration data when the DOC is 0.5 mm and the cutting workpiece diameter is 34 mm. Figure 8b shows the vibration data when cutting the workpiece to the diameter of 28 mm with DOC of 0.5 mm. It can be seen that the greater the DOC, the greater vibration amplitude. And the vibration amplitude of failure tool is slightly larger than that of the other two tool states.

Figure 8 shows that there are four resonances frequency bands occurred within 2000 Hz for all the tool wear conditions. These resonances frequency bands are around 350–550 Hz, 600–900 Hz, 900–1100 Hz and 1100–1400 Hz, and are consistent with the both results of FEM and FRF. Furthermore, as processing goes on, the mass of the workpiece decreases with the removal of the material, and consequently the four main resonant frequencies increase, which is basically the same with the previous analysis results. Therefore, the resonance frequency bands of 350–550 Hz, 600–900 Hz could be chosen as band-pass filter for further data processing.

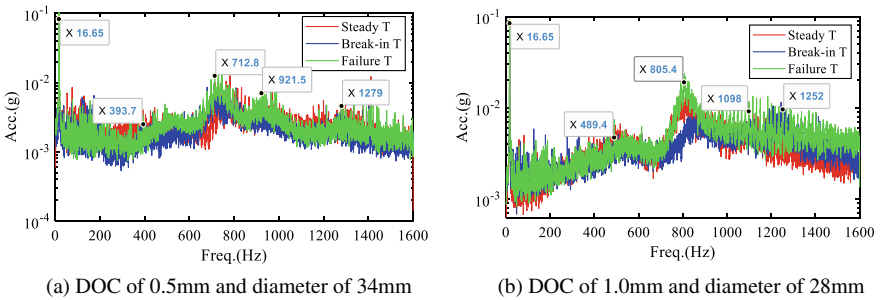


Fig. 8 Vibration signal in *X* direction obtained by ORS

4 Conclusions

Aiming at tool wear conditions identification, this paper investigated the dynamics of CNC spindle rotor coupled with workpieces. Based on FEM method, the influences of workpieces mass and bearing stiffness on vibration modes are quantitatively analyzed and corrected. It is then found that the 2nd, 3rd mode of front-end transverse vibration and the 6th mode of longitudinal vibration are more correlated with the turning process. Both offline FRF tests and online tuning process have confirmed the model analysis results.

Moreover, it is suggested that the two frequency bands of 350–550 Hz and 600–900 Hz with 2000 Hz are more sensitive for monitoring tool conditions, especially the transverse vibration at the front end (the clamping workpiece end) of 350–550 Hz. In the future, the effect of different tool wear on dynamic characteristics of system modes will be focused on for more accurate online monitoring.

Acknowledgements This work is supported by the 2020 Guangdong Province Scientific Research Platform (No. 2020KTSCX188), the Beijing Municipal Science and Technology Project (No. Z201100008320004) and the National Natural Science Foundation of China (No. 2018A030313418).

References

1. Kim, J., Lee, S., Chun, H., Lee, C.B.: Compact curved-edge displacement sensor-embedded spindle system for machining process monitoring. *J. Manuf. Process.* **64**, 1255–1260 (2021). <https://doi.org/10.1016/j.jmapro.2021.02.056>
2. Wu Di, Q.S.: Dynamic analysis for spindle assembly of CK6130CNC Lathe. *Mach. Des. Manuf.* **9**, 158–161 (2016). <https://doi.org/10.19356/j.cnki.1001-3997.2016.09.042>
3. Abele, E., Altintas, Y., Brecher, C.: Machine tool spindle units. *CIRP Ann. - Manuf. Technol.* **59**(2), 781–802 (2010). <https://doi.org/10.1016/j.cirp.2010.05.002>
4. Hong Jie, Z.Z., Han, J.: Analysis of dynamic characteristics of rotor system by integral transfer coefficient method. *J. Beijing Univ. Aeronaut. Astronaut.* **28**, 1–4 (2002)
5. Subbarao, R., Dey, R.: Selection of lathe spindle material based on static and dynamic analyses using finite element method. *Mater. Today Proc.* **22**, 1652–1663 (2019). <https://doi.org/10.1016/j.matpr.2020.02.182>
6. Ritou, M., Rabreau, C., Le Loch, S., Furet, B., Dumur, D.: Influence of spindle condition on the dynamic behavior. *CIRP Ann.* **67**(1), 419–422 (2018). <https://doi.org/10.1016/j.cirp.2018.03.007>
7. Schmitz, T.: Modal interactions for spindle, holders, and tools. *Proc. Manuf.* **48**, 457–465 (2020). <https://doi.org/10.1016/j.promfg.2020.05.069>
8. Qian Shicai, Q.W., Gao, H.: Establishment and performance analysis of rigid—flexible coupling system of machine tool spindle based on ADAMS. *J. Sichuan Mil. Eng.* **6**, 69–70 (2012)
9. Vafaei, S., Rahnejat, H., Aini, R.: Vibration monitoring of high speed spindles using spectral analysis techniques. *Int. J. Mach. Tools Manuf.* **42**(11), 1223–1234 (2002). [https://doi.org/10.1016/S0890-6955\(02\)00049-4](https://doi.org/10.1016/S0890-6955(02)00049-4)
10. Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F.: Stochastic resonance. *Rev. Mod. Phys.* **70**(1), 223–287 (1998). <https://doi.org/10.1103/revmodphys.70.223>

11. Cao, Y., Altintas, Y.: Modeling of spindle-bearing and machine tool systems for virtual simulation of milling operations. *Int. J. Mach. Tools Manuf.* **47**(9), 1342–1350 (2007). <https://doi.org/10.1016/j.ijmactools.2006.08.006>
12. Ozturk, E., Kumar, U., Turner, S., Schmitz, T.: Investigation of spindle bearing preload on dynamics and stability limit in milling. *CIRP Ann. - Manuf. Technol.* **61**(1), 343–346 (2012). <https://doi.org/10.1016/j.cirp.2012.03.134>
13. Fedorynenko, D., Kirigaya, R., Nakao, Y.: Dynamic characteristics of spindle with water-lubricated hydrostatic bearings for ultra-precision machine tools. *Precis. Eng.* **63**, 187–196 (2020). <https://doi.org/10.1016/j.precisioneng.2020.02.003>
14. Xu, K., Wang, B., Zhao, Z., Zhao, F., Kong, X., Wen, B.: The influence of rolling bearing parameters on the nonlinear dynamic response and cutting stability of high-speed spindle systems. *Mech. Syst. Signal Process.* **136**, 106448 (2020). <https://doi.org/10.1016/j.ymsp.2019.106448>
15. Yuan Lin, C., Pin Hung, J., Liang Lo, T.: Effect of preload of linear guides on dynamic characteristics of a vertical columnspindle system. *Int. J. Mach. Tools Manuf.* **50**(8), 741–746 (2010). <https://doi.org/10.1016/j.ijmactools.2010.04.002>
16. Cao, H., Li, B., He, Z.: Chatter stability of milling with speed-varying dynamics of spindles. *Int. J. Mach. Tools Manuf.* **52**(1), 50–58 (2012). <https://doi.org/10.1016/j.ijmactools.2011.09.004>
17. Eynian, M.: Vibration frequencies in stable and unstable milling. *Int. J. Mach. Tools Manuf.* **90**, 44–49 (2015). <https://doi.org/10.1016/j.ijmactools.2014.12.004>
18. Liu, J., Li, F., Yong, J., Lai, T., Zhang, P.: Investigation of spindle-tool assembly dynamics for optical grinding motorized spindles. *Optik (Stuttg)* **216**, 164836 (2020). <https://doi.org/10.1016/j.ijleo.2020.164836>
19. Miao, H., Li, C., Wang, C., Xu, M., Zhang, Y.: The vibration analysis of the CNC vertical milling machine spindle system considering nonlinear and nonsmooth bearing restoring force. *Mech. Syst. Signal Process.* **161**, 107970 (2021). <https://doi.org/10.1016/j.ymsp.2021.107970>
20. Lin, C.W., Lin, Y.K., Chu, C.H.: Dynamic models and design of spindle-bearing systems of machine tools: a review. *Int. J. Precis. Eng. Manuf.* **14**(3), 513–521 (2013). <https://doi.org/10.1007/s12541-013-0070-6>
21. Murgayya, S.B., Suresh, H.N., Madhusudhan, N., SarvanaBhavan, D.: Effective rotordynamics analysis of high speed machine tool spindle—bearing system. *Mater. Today Proc.* (2021). <https://doi.org/10.1016/j.matpr.2021.05.359>
22. Zhang Yimin, Y.Z., Wang, H., Cao, H.: FEM-based frequency reliability analysis of spindle system of CNC lathes. *J. Northeast. Univ.* **36**, 1155–1159 (2015)
23. Turkes, E., Orak, S., Neşeli, S., Sahin, M., Selvi, S.: Modelling of dynamic cutting force coefficients and chatter stability dependent on shear angle oscillation. *Int. J. Adv. Manuf. Technol.* **91**(1–4), 679–686 (2017). <https://doi.org/10.1007/s00170-016-9782-y>
24. Jiang Wei, L.B., Zhou, Y.: Analysis and calculation on rigidity and friction moment of paired angular contact bearings. *Bearing* **8**, 31–35 (2006)
25. NSK bearing replacement guide
26. Wijnant, Y.H., Wensing, J.A.: The influence of lubrication on the dynamic behaviour of ball bearings. *J. Sound Vib.* **4**(3), 579–596 (1999). <https://doi.org/10.1080/10402009208982155>
27. Lynagh, N., Rahnejat, H., Ebrahimi, M., Aini, R.: Bearing induced vibration in precision high speed routing spindles. *Int. J. Mach. Tools Manuf.* **40**(4), 561–577 (2000). [https://doi.org/10.1016/S0890-6955\(99\)00076-0](https://doi.org/10.1016/S0890-6955(99)00076-0)
28. Time-Varying Stiffness Characteristics of Roller Bearing Influenced by Thermal Behaviour Due to Surface Frictions and Different Lubricant Oil Temperatures
29. Rao, S.S.: *Vibration of Continuous Systems* (2007)
30. Ghani, J.A., Rizal, M., Nuawi, M.Z., Ghazali, M.J., Haron, C.H.C.: Monitoring online cutting tool wear using low-cost technique and user-friendly GUI. *Wear* **271**(9–10), 2619–2624 (2011). <https://doi.org/10.1016/j.wear.2011.01.038>
31. Chun Li, A.D.B., Li, B., Gu, L., Feng, G., Gu, F.: Online monitoring of a shaft turning process based on vibration signals from on-rotor sensor. In: 2020 3rd World Conference on Mechanical

Engineering Intelligence Manufacturing, pp. 402–407 (2020). <https://doi.org/10.1109/WCM EIM52463.2020.00091>

32. Li, C., et al.: Tool Condition Monitoring based on Vibration Signal from an On-Rotor Sensor in CNC Turning Process, no. Cm (2021)

Real-Time Condition Monitoring and Health Assessment of Equipment Power Transmission Device Based on Wireless Sensor Network



Cheng Zhe, Jiang Wei, Hu Niaoqing, Zhang Hao, and Zhen Dong

Abstract The high requirements on the integrity and operational reliability of the equipment in the process of carrying out tasks is required by the ships and other major equipment. To ensure the long-term safe and healthy operation of their power transmission devices is one of the key links to achieve this requirement. In response to this urgent need, the real-time monitoring and health assessment method of equipment power transmission based on wireless sensor network is researched, and the related software and hardware prototype systems is constructed, after that, the system testing and use verification in actual tasks of ships are carried out. The verification results show that the prototype system can complete the online collection and analysis of vibration and temperature data of the key parts of the ship's power transmission device, and the real-time monitoring and health assessment of the power transmission device condition is effective. As a result of that, the maintenance workload is reduced effectively while work efficiency is improving.

Keywords Real-time condition monitoring · Health assessment · Power transmission device · Wireless sensor network

C. Zhe (✉) · H. Niaoqing
College of Intelligence Science and Technology, National University of Defense Technology,
Changsha 410073, China
e-mail: chengzhe@nudt.edu.cn

Laboratory of Science and Technology on Integrated Logistics Support, NUDT, Changsha
410073, China

J. Wei
23451 Troops of the Chinese People's Liberation Army, Wenzhou 325000, China

Z. Hao · Z. Dong
School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China

1 Introduction

When the large ships perform long-distance voyage tasks, they are often far away from the rear support and maintenance base for a long time. For this reason, it is necessary to focus on solving an important problem: whether the integrity and reliability of the ship's key subsystems and devices can meet the mission requirements at any time during the voyage. Online testing and data analysis of the key measuring points of the equipment core equipment through the condition monitoring and health assessment system, and the overall health assessment based on the condition information are the effective means to dynamically grasp the condition integrity and mission reliability of the key subsystems in real time. The similar system has been initially applied in major engineering equipments such as aviation, aerospace, electric power, rail transit, etc. [1–5].

One of the difficulties in carrying out condition monitoring and health assessment for ships in active service is to carry out system deployment without changing the basic structure of the power and transmission device and its auxiliary power supply system, so as to avoid affecting the normal operation of the daily maintenance of the ship's power cabin. Wireless multi-function sensors provide an option to achieve this goal. Not only can they sense multiple condition information of the equipment in real time, but also do not need to reconsider the complicated pipeline layout problems in the power cabin. In view of this, this paper constructs a real-time monitoring and health assessment system for the equipment power transmission devices based on wireless sensor networks, and carries out tests and verifications on a certain type of ship equipment.

2 Real-Time Monitoring and Health Assessment System

2.1 Architecture of the System

The equipment monitoring system is mainly composed of some wireless multifunction sensors and a wireless gateway, as shown in Fig. 1. The wireless multifunction sensors can be directly glued to the equipment to measure and wirelessly transmit the vibration and the temperature signal of the ship equipment, including the vertical, horizontal, and axial speed RMS, the true peak value of acceleration, and the effective peak value of acceleration, and the surface temperature data of the equipment. The wireless transmission distance can reach about 100 m in the cabin. The collected data sets are sent to the wireless gateway through the wireless network. The gateway is placed in the cabin with a display screen for the crew to directly observe and read the real-time data. At the same time, the gateway is equipped with a download port to facilitate data export.

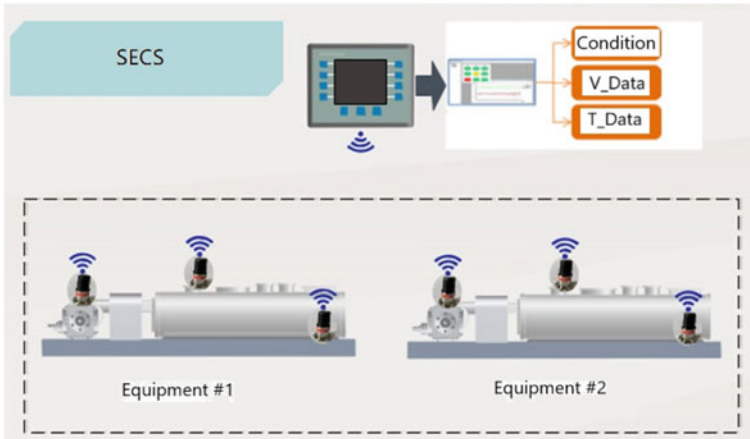


Fig. 1 Schematic diagram of the system's functions

2.2 Wireless Multi-function Sensor

The wireless sensors selected by this system are a new type of multifunctional data acquisition system, as shown in Fig. 2. This type of sensor not only has the ability to synchronously collect the temperature and the three-way vibration data of the monitored equipment in real time, but also can perform data preprocessing and analysis, extract feature vectors that reflect the health of the equipment, and be

Fig. 2 Wireless multi-function sensor



selective to the host computer at specified intervals to send raw data or characteristic parameters wirelessly.

The wireless data transmission is based on the BeeLPW-T protocol, which can self-organize to form a star network topology. The wireless digital signal transmission method eliminates the noise interference caused by long cable transmission, and the entire measurement system has extremely high measurement accuracy and anti-interference ability. Wireless sensor nodes can form a huge wireless sensor network, which supports simultaneous testing of hundreds of measuring points. The node structure is compact, small in size, and easy to install.

2.3 Wireless Gateway System

The wireless gateway selected by this system is a gateway with complete interfaces and comprehensive functions. It can directly use the existing wired network resources through the LAN port to connect the data collected by the sensor to the network and transmit it back to the data center. It can not only be used for remote monitoring, but also be used for short-distance big data transmission (via LAN port). It can form a wireless sensor network through a wireless radio frequency module with indicators for operation, power supply, and failure. The data is displayed on a 10-inch display.

2.4 Software System

The software system of this system has 4 functions, named as follows:

- (1) Real-time display of monitoring data. It mainly includes functions such as historical data curve drawing and display, collection and sending interval setting, data management and preliminary analysis, and automatic setting of alarm values.
- (2) Equipment condition monitoring. The condition monitoring of the equipment is calculated based on all the data of all sensors at each measuring point, and the health condition of the equipment can be judged in real time through multiple dynamic thresholds.
- (3) Analysis of health trends. The system performs health trend analysis and status prediction on key subsystems based on the accumulated historical data.
- (4) Big data processing and management. This system uses a deep learning model to systematically analyze, process and manage a large amount of accumulated historical data to realize the management and prediction of the health status of equipment.

3 System Testing and Verification

3.1 *Brief Introduction*

Aiming at the problem of abnormal detection of power system equipment during the execution of the ship's mission, the real-time monitoring and health assessment system developed by the research institute was installed in the ship's power cabin. In order for the real-time monitoring system to be truly effective, the ship's staff has formulated an equipment inspection system. Each maintenance staff observes the equipment status displayed by the monitoring system every hour, and records the monitoring system data twice a day, with a total of more than 1,000 monitoring equipment, A total of more than 50,000 groups.

According to the feedback from the ship's staff, the system played a significant role in monitoring the status of the ship's main propulsion diesel engine and power station diesel engine during the execution of the task, and provided good hardware support for the use and management of the main and auxiliary engine equipment. Through the health indicator on the display panel, the maintenance staffs can quickly grasp the current health status of the equipment, and can find abnormal situations in time, which reduces the workload and work intensity of the on-duty personnel using the handheld vibration monitoring equipment to measure the status data. At the same time, when abnormal vibration data occurs, it can be measured in a targeted manner, so that manual measurement data has reference data that can be compared horizontally.

3.2 *Data Analysis*

Through a comprehensive analysis of the data of all sensor nodes, not only the operating status of the monitoring system and the operating status of the monitored power and transmission system can be effectively understood, and possible hidden dangers can be eliminated, but also maintenance suggestions and decisions can be provided for similar tasks in the future support.

Figures 3 and 4 show the time history diagram and health history diagram of the vibration of the ship's power engines. From the above data analysis results, the following conclusions can be drawn: In the process of system monitoring, there are many cases of engines vibration exceeding the warning line No. 1, and occasionally exceeding the warning line No. 2. Overall, the health of the equipment is basically above 80 points, and the equipment is operating well.

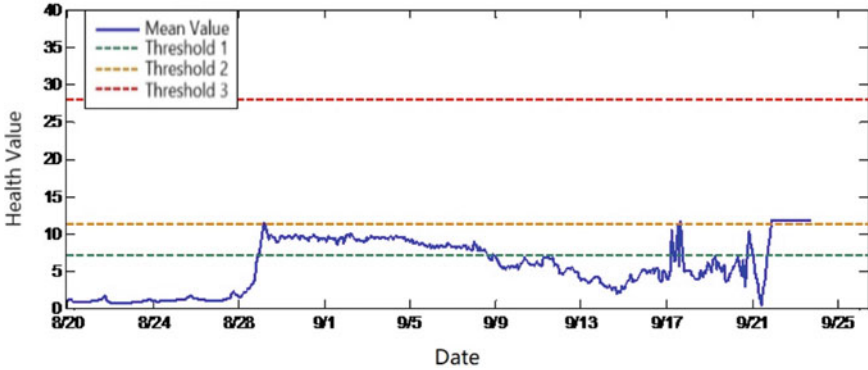


Fig. 3 Vibration history diagram

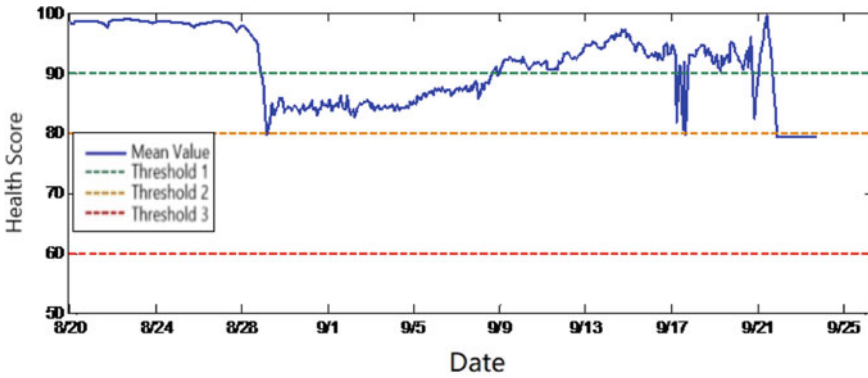


Fig. 4 Health history chart

3.3 Typical Case Verification

According to the information fed back by the maintenance staffs of ship power equipment, the system accurately diagnosed a gearbox failure during the long-term use of the product during the mission, and provided accurate warnings, providing decision-making suggestions for the maintenance personnel to take maintenance measures in a timely manner. The following is an analysis of this failure warning case.

After investigation, the gearbox fault is an oil leakage fault, and the data analysis results are shown in Figs. 5 and 6. The output end of the right shaft of the gearbox vibrates at the corresponding parts due to lack of oil lubrication, and the health of the equipment is significantly reduced; after repairing the faulty part and adding lubricating oil, after a period of observation, the health value immediately returns to the normal range.

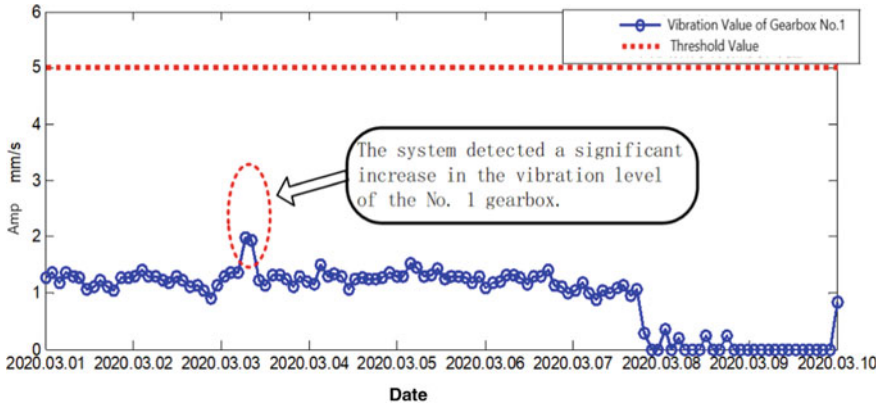


Fig. 5 Vibration analysis in fault diagnosis cases

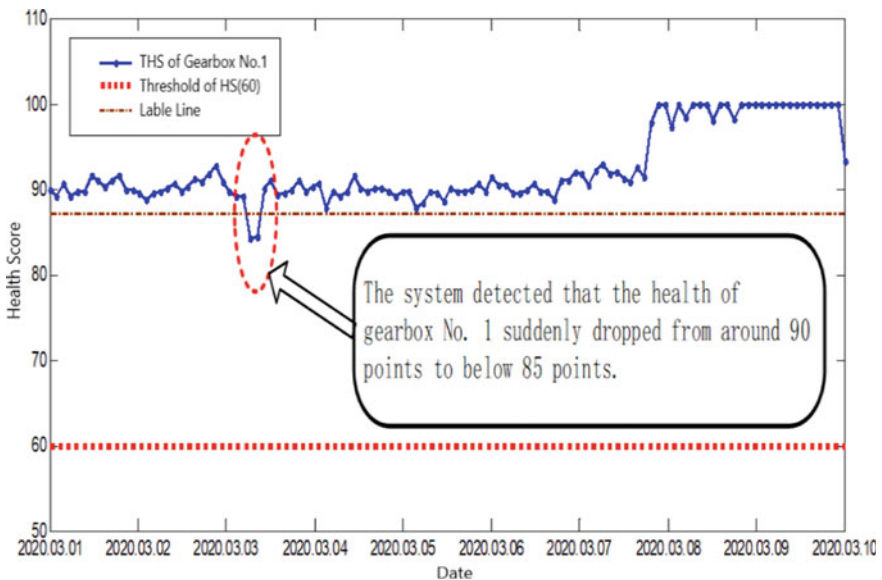


Fig. 6 Health analysis in fault diagnosis cases

4 Conclusion

Through testing and verification of the real-time monitoring and health assessment system developed in this article, the following conclusions can be drawn:

- (1) The system is based on a wireless sensor network to monitor power equipment, which is generally practical and feasible.

- (2) The development of automatic data collection and intelligent state assessment systems has practical application value for reducing the crew's work intensity and increasing the comparability of information;
- (3) Through the state assessment technology based on the analysis and evaluation of the overall health of the equipment, it can warn of hidden troubles and has practical application value for the use and maintenance of ships.
- (4) The existing system needs to be further improved in terms of signal monitoring range, comprehensive status evaluation and fault detection performance, reliability, safety, maintainability, and supportability to fully adapt to the application requirements of the mission site.

References

1. Dou, W., Wang, W.: Design and implementation of ship condition monitoring system based on zigbee. *Electron. Design Eng.* **22**(22), 127–130 (2014)
2. Cai, J., Jiang, Y., Sun, Z.: Implementation and development of health monitoring of A320 aircraft air separator. *J. Civil Aviation* **5**(05), 68–70 (2021)
3. Liu, Y., Liu, X., Sun, L. etc.: Design and implementation of test equipment health management system based on B/S architecture. *Electric Drive* **50**(11), 101–105+116 (2021)
4. Xia, Y., Ding, Q., You, L.: Radar health management system based on big data. *J. Terahertz Sci. Electron. Inform.* **17**(04), 686–690 (2019)
5. Liu, S., Xu, Y., Qiao, Q.: Overview of the health diagnosis and intelligent maintenance management system for large equipment. *J. Ordnance Equipm. Eng.* **38**(12), 297–330+304 (2017)

Research on Fault Detection Method for Special Equipment Under the Condition of Sample Missing



Lei Wei, Zhe Cheng, Niaoqing Hu, Junsheng Cheng, and Guoji Shen

Abstract In the fault diagnosis method of data-driven method, it is difficult to obtain fault data and high cost of experiment due to the particularity of special equipment and health condition for a long time at the beginning of operation. Based on the analysis of slow-changing parameters such as temperature and pressure collected under normal operation, this paper establishes signal prediction models under different conditions and puts forward a historical view. The dynamic threshold method of measuring data eliminates the false alarm and improves the ability of early fault detection at the initial stage of equipment operation, and provides a new idea for fault detection under the condition of only normal samples. It provides scientific and accurate support for fault early warning theory and method of special equipment and realizes the direction of special equipment from regular maintenance and preventive maintenance to condition-based maintenance change.

Keywords Special equipment · Normal sample · Slow variation parameters · Fault detection · Dynamic threshold

1 Introduction

Special equipment is a kind of special equipment widely used in military, aerospace, medical, nuclear power, fire protection and other fields. Due to the extremely harsh environment often encountered in the actual operation process, the parts of special equipment deteriorate until they fail, which leads to serious failure of equipment

L. Wei · Z. Cheng (✉) · N. Hu · G. Shen

College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

e-mail: chengzhe@nudt.edu.cn

Laboratory of Science and Technology on Integrated Logistics Support, NUDT, Changsha 410073, China

L. Wei · J. Cheng

College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410073, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

1081

H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,

Mechanisms and Machine Science 117,

https://doi.org/10.1007/978-3-030-99075-6_86

and major economic losses. Therefore, if effective detection can be carried out in the early stage of failure and maintenance can be carried out in time before serious failure, on the one hand, the safety of equipment and personnel can be protected to the greatest extent, on the other hand, the mode change from regular maintenance and preventive maintenance to condition-based maintenance can be realized in the health management of special equipment, greatly reducing the economic cost of equipment use [1].

At present, the fault diagnosis technologies applied to equipment are mainly divided into four categories: methods based on physical model, methods based on data, methods based on reliability model and methods based on empirical model [2]. Due to the complexity of the actual structure of large-scale equipment, it is often extremely difficult or even impossible to establish its highly accurate physical model. A fault diagnosis technology without involving accurate physical model is more suitable for complex systems. Therefore, data-driven fault diagnosis method has become the mainstream method for fault diagnosis and health management of large-scale special equipment.

Unfortunately, among many fault diagnosis methods based on in-depth learning, there is a common problem that requires a balance between the large number of training data samples and the number of normal samples and the number of fault samples [3]. In [4], the predictive results of different artificial neural network models are multi-level fused, and a wear fault fusion diagnosis model of machine is established, which overcomes the disadvantages of low diagnostic accuracy of a single model. However, this diagnosis method requires a lot of historical data to ensure robust diagnostic accuracy, and the set fault threshold is fixed. In [5], the measured current curve of the ramp under normal operation is used to calculate the similarity with the current curve under fault, and the dynamic fault threshold is set according to different stages. Although the fault diagnosis model of the ramp mechanism can be established only with normal data samples, the dynamic threshold established by the model can not be updated with the change of historical data, and the dynamic threshold of this model can not be updated with the change of historical data. It is discrete in value, only on the basis of static threshold value, different increments of threshold value are increased with the selection of ramp mechanism in different operation stages.

2 Fault Detection Technology with Missing Fault Samples

2.1 Fault Detection Principle

When the system or equipment fails, it often cannot reach the specified performance index, so it shows that at least one monitoring parameter obviously deviates from the normal level externally. In the data-driven fault diagnosis method, the normal operation state of a complex system is described by establishing a mathematical model,

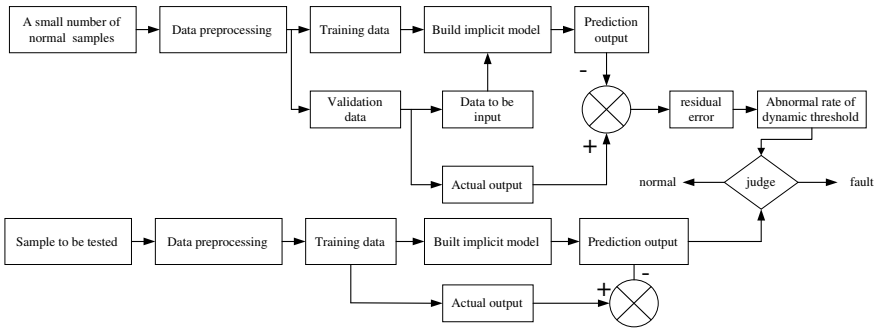


Fig. 1 Fault detection flow chart

and the implicit relationship of relevant measurement variables in the system is established by using multivariate statistical method and machine learning algorithm. When a fault occurs, it will inevitably lead to the change of the implicit relationship of relevant measurement variables in the system. This characteristic is used to evaluate the difference between the system state and the actual state of the current system under normal model parameters. The residual value obtained by difference operation is compared with the set threshold to judge whether a fault has occurred [6]. Under the condition that only a small number of normal samples are obtained, the data-driven fault detection process can be divided into the following processes: training data, establishment of normal system state model, generation of residual between actual value and predicted value of normal system state model, evaluation and decision-making of residual. The flow chart of fault detection is shown in Fig. 1.

2.2 Data Interpolation Method Under the Condition of Missing Samples

In the process of data acquisition of special equipment monitoring system, data loss is common due to random packet loss, different sampling rates of sensors and sensor failures. It will not only cause different multivariate data to be input into training mode models at the same time because of their different dimensions. Moreover, the fault detection model established by using the data with missing values can not effectively detect the actual faults or cause misdiagnosis. Therefore, the missing data interpolation operation in the data preprocessing stage is of great significance to improve the accuracy of the diagnosis model [7]. At present, according to the number of substitute values to replace missing data, the commonly used data interpolation methods can be divided into single value interpolation methods and multiple interpolation methods [8].

- (1) Single Value Interpolation Method

The method of using a single estimated value to interpolate missing data to obtain a complete data set is called single value interpolation method. According to different interpolation methods, single value interpolation methods can be further divided into single value interpolation methods such as mean interpolation method, KNN interpolation method and regression interpolation method.

(2) Multiple Interpolation Method

The core idea of the multiple interpolation method is to use some single value interpolation algorithms to estimate the missing values for many times, obtain multiple estimation results, construct a complete data set, and perform the same operation on all complete data sets to obtain multiple complete estimates, The next step is to comprehensively analyze these estimates to obtain the final value. Multiple interpolation method is used to adopt the idea of multiple estimation, which improves the interpolation accuracy to a certain extent, but inevitably increases the computational complexity.

2.3 *Establishment of Fault Detection Model*

The essence of equipment fault detection model is a pattern recognition problem. It is easy to realize the fault detection function under the condition of obtaining normal samples and fault samples. In practice, it is usually difficult to obtain fault samples. For only normal samples, it is not applicable to the establishment of two classifiers. Therefore, the class I model recognition method trained with normal data is more applicable, and the fault detection is transformed into a class I classification problem. The principle is to treat the health state data obtained from equipment operation as positive samples, train a class of classifier, and input the data to be tested into the trained class of classifier to obtain the classification results. A classical class of classifiers are: one-class support vector machine, LOF (Local Outlier Factor), isolated forest, principal component analysis and so on.

On the vast majority of equipment, the slowly varying parameters represented by temperature often change with different working conditions. Under non-stationary working conditions, the measured values of slowly varying parameters will be different. Therefore, even under the healthy operation state, it may be determined as a fault state by a class of classifiers. However, it is easy to establish corresponding mathematical expressions for the physical model of the sub components constituting large equipment. If the implicit relationship of measurable parameters during normal operation can be analyzed in combination with its operation mechanism in equipment fault detection, the influence of external factors such as temperature and working conditions can be integrated into the fault detection model to further improve the detection accuracy [9].

When the equipment has only one output variable and multiple input variables, the inherent implicit relationship between parameters in the system model is:

$$y = f(x_1, x_2, \dots, x_m) \tag{1}$$

where: y is the output variable, x_m is the input variable, $f(\cdot)$ is the implicit function, $m \in \mathbb{R}$.

In the pattern recognition algorithm, the detection results obtained by fusing multiple fault detection models are often more accurate than those obtained by a single fault detection model. Therefore, an implicit model array fault detection model composed of multiple single implicit relationships in the system is proposed to fuse and determine the detection results of multiple single fault detection models, Under the condition of only a small amount of normal operation data, further improve the fault detection accuracy of the equipment, and realize more sensitive, robust and better fault detection. For an equipment with M input parameters x_1, x_2, \dots, x_M and N output parameters y_1, y_2, \dots, y_N , implicit model arrays composed of parameter implicit models in the system are established. Among the N implicit models, each implicit model uses one of the N output parameters y_1, y_2, \dots, y_N as the output variable, and the other $N-1$ output parameters and input parameters as the input variables of the implicit model. The mathematical expression of this implicit model is shown as follows:

$$\begin{cases} y_1 = f_1(x_1, x_2, \dots, x_M, y_2, y_3, \dots, y_N) \\ y_2 = f_2(x_1, x_2, \dots, x_M, y_1, y_3, \dots, y_N) \\ \dots \\ y_N = f_N(x_1, x_2, \dots, x_M, y_1, y_2, \dots, y_{N-1}) \end{cases} \tag{2}$$

where: y_N is the output variable, x_M is the input variable, $f_N(\cdot)$ is the implicit function, M and $N \in \mathbb{R}$.

The detailed schematic diagram of the system implicit relationship array represented by Eq. (2) is shown in Fig. 2.

2.4 Dynamic Threshold Detection Strategy

In engineering, the fixed threshold fault detection methods represented by Pauta criterion, Shovinat criterion and Grubbs criterion are widely used in equipment fault detection. The fixed threshold detection method does not change the fault detection threshold during the service period of the equipment. This method is not applicable to the equipment under variable working conditions and variable load. It is easy to cause false alarm and missing alarm.

The threshold of fault detection is generally obtained according to the data statistics of historical operation of equipment under normal state. The advantage of dynamic threshold method over fixed threshold method is that it can automatically update the threshold of fault detection according to the real-time operating conditions of equipment, so as to improve the accuracy and robustness of fault detection. In

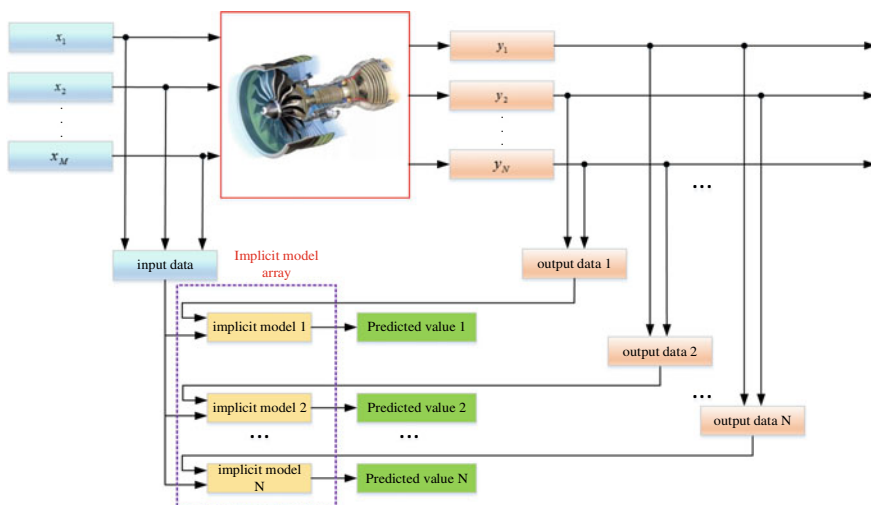


Fig. 2 Schematic diagram of implicit model array

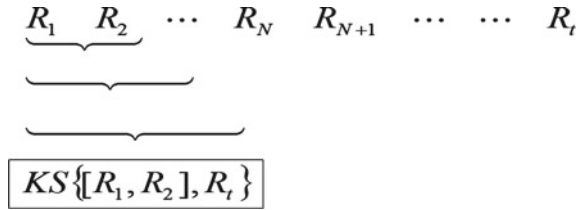
order to improve the updating ability of the newly monitored data to the threshold and improve the sensitivity of fault detection, the sliding window method is usually used to update the dynamic threshold of fault detection to make statistics on the data, that is, the data positions participating in the threshold update are continuously moved with the increase of time, and the samples participating in the threshold update are repeatable. In order to eliminate the influence of false alarm caused by accidental abnormal data, the method of abnormal rate is used to solve this problem. Abnormal rate monitoring refers to the ratio of the number of points regarded as fault data in a sliding window to the number of points of total data in the window [10]. When the number of points in a sliding window is greater than the set ratio, it is determined that there is a fault in this section of sample, otherwise it is determined as normal operation state, and the Eq. (3) describes the abnormal rate:

$$\eta = \frac{N_a}{N} \tag{3}$$

where: η is the abnormal rate, N_a is the number of abnormal data points in the window, N is the total data points in the window.

When the slowly varying parameters such as temperature and pressure are taken as the monitoring objects, due to the inertia in the change of the slowly varying parameters, the slowly varying parameters cannot change in time when the equipment fails. The fault can be detected only after the fault has accumulated for a period of time, and there is no sudden change point in the slowly varying parameters when the sensor fault is eliminated. Therefore, the dynamic threshold setting steps based on the slowly varying parameters are as follows [11]:

Fig. 3 Schematic diagram of subset selection



- (1) Set sliding window size: The minimum feature subset of the original data set is selected according to the K-S test principle, that is, the size of the sliding window. When the result of K-S test is greater than 0.05, it means that the selected subset has the same distribution law as the original data set (parent set). The selection strategy is shown in Fig. 3. Randomly select a small section of data samples as a subset, and then conduct K-S similarity test with the parent set, so as to expand the subset range and know that the similarity test result is valid. At this time, write down the minimum sliding window size. Subset selection is shown in Fig. 3.
- (2) Set failure threshold: In order to make full use of the state index change trend of the previous sliding window, select the sliding window data obtained in step 1, and calculate the fault detection threshold in the window according to Eq. (4):

$$T_i = \frac{\sum_{k=i-N}^i R_k}{N} + R_{th} \tag{4}$$

where: T_i is the exception rate in the window, R_k is the status indicator at a certain time, is the distribution characteristics of equipment operation state indexes are analyzed by nuclear density estimation method and set based on the principle of small probability events, N is the total data points in the window and $N \in \mathbb{R}$, i is the location index in total sample data.

- (3) Move the sliding window step by step and update the threshold according to step 2.
- (4) Repeat step 3 until the thresholds of all sliding windows are updated, and the dynamic threshold curve is obtained by connecting the thresholds.

3 Example

In this paper, 17,420 oil temperature data of a certain type of transformer from July 1, 2016 to June 26, 2018 are selected as the research object. The data acquisition frequency is to collect one data per hour. The data structure is that six input variables determine one output variable. Some data are shown in Table 1:

In order to improve the number of data points, aiming at the problem that the interval between collecting data once an hour and storing data is too large, the

Table 1 Transformer measurable data

date	HUFL	HULL	MUFL	MULL	LUFL	LULL	Oil temperature/°C
2016/7/1 0:00	5.827	2.009	1.598	0.462	4.203	1.340	30.531
2016/7/1 1:00	5.692	2.075	1.491	0.425	4.142	1.371	27.787
...
2018/6/26 19:00:00	10.114	3.549	6.183	1.564	3.716	1.462	9.567

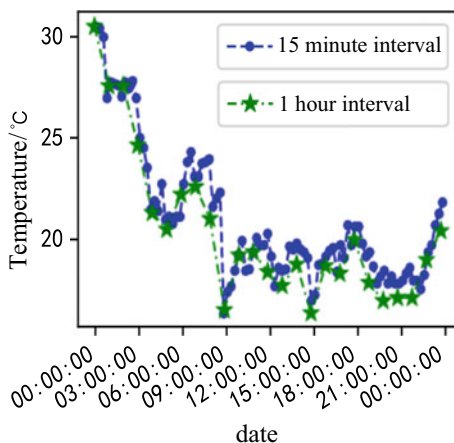
Notes: HUFL represents high useful load, HULL represents high useless load, MUFL represents middle useful load, MULL represents middle useless load, LUFL represents low useful load, LULL represents low useless load

regression interpolation method is used to interpolate the data, and one data point is collected every hour to one data point every 15 min. The data interpolation results of a certain day are shown in Fig. 4.

In this paper, XGBOOST (eXtreme Gradient Boosting) algorithm is selected to extract the implicit model between the data, selects the monitoring data of the transformer from July 1, 2016 to July 10, 2016 as the training data, and selects the health state from July 11, 2016 to August 11, 2016 as the prediction data, calculates the upper and lower limits of the fault detection dynamic threshold from the residual between the real value and the predicted value, and takes the size of the sliding window, The calculation results are shown in Fig. 5. According to the calculation results, if the abnormal rate in each window is higher than 0.35, the transformer fault alarm will be carried out, and if the abnormal rate is lower than 0.15, it will be regarded as normal operation.

Continue to select the data from August 12, 2016 to September 12, 2016 for analysis. The transformer used static threshold to detect the fault. By checking the records, it was found that the residual error between the predicted value and the real value was too large due to the increase of current load, resulting in a false alarm,

Fig. 4 The comparison before and after data interpolation



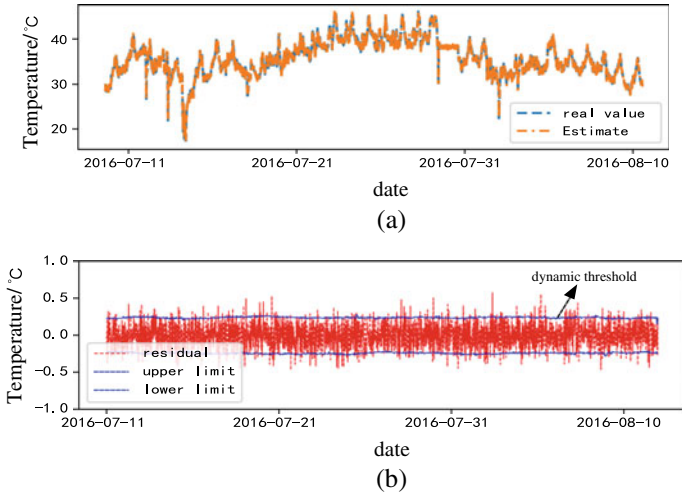


Fig. 5 Prediction error of transformer in healthy state. **a** real and predicted values, **b** residual between real value and predicted value

while the residual error caused by arc discharge fault under normal load was not detected. The use of dynamic threshold fault detection successfully eliminates the false alarm and can not detect the weak temperature change caused by some electrical faults. As shown in Fig. 6, area A is the false alarm of static threshold method and

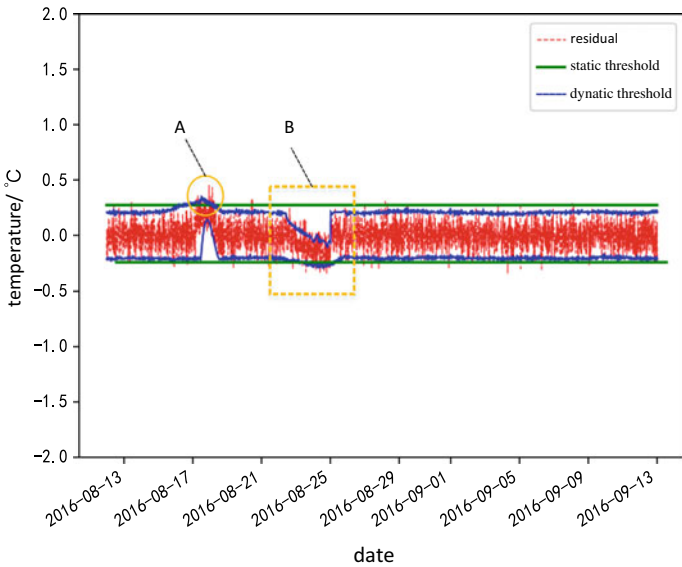


Fig. 6 Comparison of static threshold and dynamic threshold fault detection methods

area B is the missed diagnosis of static threshold method. Because of signal noise and measurement error, some measured temperature values will inevitably exceed the dynamic threshold, but if the abnormal rate does not exceed 0.15 in a sliding window, it is regarded as normal operation.

4 Conclusions

This paper studies the method of equipment fault detection with only a small number of normal samples, excavates the implicit model between variables, establishes the fault detection model based on this, and uses the dynamic threshold method to determine the residual between the predicted value and the actual value. Experiments verify the effectiveness of the scheme and solve the problems of false alarm and missed diagnosis. It has practical significance for equipment health management and fault diagnosis.

On the basis of fault detection for equipment, the fault type can be judged from the curve variation law of measured parameters by using expert knowledge and other techniques, and the maintenance and maintenance efficiency of maintenance managers can be further improved.

References

1. Kaistha, N., Upadhyaya, B.R.: Incipient fault detection and isolation of field devices in nuclear power systems using principal component analysis. *Nuclear Technol.* **136**(2), 221–230 (2001)
2. Yi, K., Enzhe, S., Chong, Y., Quan, D.: Overview of marine diesel engine fault prediction and health management technology. *J. Harbin Eng. Univ.* **41**(01), 125–131 (2020)
3. Yaguo, L., Bin, Y., Xinwei, J., Feng, J., Naipeng, L., Nandi, A.K.: Applications of machine learning to machine fault diagnosis: a review and roadmap. *Mech. Syst. Signal Process.* **138**, 106587 (2020)
4. Xiaojian, X., Zhuangzhuang, Z., Xiaobin, X., Jianbo, Y., Leilei, C., Xinping, Y., Gaodong, W.: Machine learning-based wear fault diagnosis for marine diesel engine by fusing multiple data-driven models. *Knowl.-Based Syst.* **190**, 105324 (2020)
5. Wenjiang, J., Yuan, Z., Xinhong, H., Sheng, G.Q., Infu, N.: Intelligent diagnosis method for turnout fault based on FastDTW. *Pattern Recogn. Artif. Intelligence* **33**(11), 1013–1022 (2020)
6. Zeng, X.J., Yang, M., Bo, Y.F.: Gearbox oil temperature anomaly detection for wind turbine based on sparse Bayesian probability estimation. *Int. J. Electrical Power Energy Syst.* **123**, 106233 (2020)
7. Feng, W., Jiangchuan, L.: Networked wireless sensor data collection: issues, challenges, and approaches. *IEEE Commun. Surv. Tutorials* **3**(4), 673–687 (2011)
8. Xinsheng, P.: Comparative study on missing data interpolation methods. *Stat. Decis.* **24**, 18–22 (2012)
9. Mingliang, B., Jinfu, L., Jinhua, C., Xinyu, Z., Daren, Y.: Anomaly detection of gas turbines based on normal pattern extraction. *Appl. Thermal Eng.* **166**, 114664 (2020)
10. Yanqiong, W., Xiuqin, B.: Research on dynamic threshold generation algorithm for fault diagnosis. *J. Wuhan Univ. Technol. (Traffic Science and Engineering Edition)* **32**(4), 742–745 (2008)

11. Yirong, L., Fulin, W., Jiechao, L., Xiaoli, W.: A novel method for tool identification and wear condition assessment based on multi-sensor data. *Appl. Sci.* **10**(8), 27468 (2020)

Research on Dynamic Impact Force Calculation for Spline Coupling Teeth and Its Suppression



Wenchao Pan, Hai Lan, Zhiyong Han, Lantao Yang, Liming Wang, and Yimin Shao

Abstract Due to the high bearing capacity and good alignment advantages, splines are widely used in various mechanical equipment. However, alternating stress can be easily observed in meshing of spline coupling, which leads to fatigue failure of spline pairs such as cracks. The underlying mechanism and suppression method of impact force of spline pair are seldom studied. In order to settle this problem, this paper proposed a new two-mass nonlinear impact dynamic model of spline pair, and studies the generation mechanism, influence law and suppression method of dynamic impact force of spline pair. Based on Hertzian contact theory, this paper establishes the calculation model of dynamic impact force of spline pair firstly, both the coupling relationship between pressure angle, backlash and eccentricity and impact force are taken into consideration. Then, the calculation and suppression method of impact force of spline pair are investigated and discussed. The results show that the impact force gradually increases with the growth of backlash and eccentricity, and local extremum appears. This paper presents a new model to study and suppress the dynamic impact force of spline teeth, it is found that the root mean square (RMS) value of impact force reaches to the smallest when pressure angle is set as 25.

Keywords Spline coupling · Impact force · Suppression method

1 Introduction

Due to the characteristics of high bearing capacity, good alignment, good guidance and strong interchangeability, involute spline pair is widely used in mechanical

W. Pan · Z. Han · L. Yang · L. Wang (✉) · Y. Shao
State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044,
People's Republic of China
e-mail: lmwang@cqu.edu.cn

H. Lan
Science and Technology on Vehicle Transmission Laboratory, China North Vehicle Research
Institute, Beijing 100821, People's Republic of China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Zhang et al. (eds.), *Proceedings of InCoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_87

1093

transmission systems of aircraft, automobiles, large-tonnage land-based platforms, machine tool manufacturing, agricultural machinery and other fields [1].

However, due to the factors such as heavy load and high impact frequency of splines, the meshing alternating stress of spline pair is easy to cause fatigue fracture and wear failure of spline teeth [2]. Many scholars have studied the anti-fatigue methods of spline pairs. For example, Kahraman et al. [3] deduced the backlash of involute spline pair under pure torsion, and presented its detailed mathematical model. Robins et al. [4] analyzed the influence of backlash on the load and meshing stiffness of involute splines through theory and experiment, and pointed out the meshing sequence of splines under load and that only 25–50% of splines participated in meshing. Cura et al. [5] mainly studied the parallel offset misalignment load distribution of involute spline pairs, and proposed a theoretical method based on non-finite element method. Li et al. [6] investigated the dynamic influence of backlash and load eccentricity on gear meshing force, and found that the backlash can lead to the dynamic change of gear meshing force under the external unsteady disturbance torque. Zhao et al. [7] calculated the reaction force of misaligned spline pair misalignment through the analysis of meshing force, and derived the dynamic equation of misaligned spline pair and the numerical calculation expression of meshing force of key teeth under misalignment condition, and then obtained the influence characteristics of misalignment on its dynamic characteristics by using finite element method. Meanwhile, Zhao et al. [8, 9] also discussed the meshing force and torsional stiffness of involute spline pair rotor system with radial misalignment, and analyzed the dynamic characteristics and stability of spline rotor system, which provided a reliable theoretical basis for the optimization design of involute spline pair [10]. In addition, they established an analytical model of transverse meshing stiffness of spline pair, and discussed the nonlinear relationship between meshing stiffness and torque or meshing force under large deformation [11]. Zhou et al. [12] carried out friction contact analysis on the transmission shaft spline pair of high-speed EMU with bearing both torque and bending moment, and obtained the distribution of contact stress on the bearing tooth surface. The above analyses show that the spline fault and fatigue failure of the spline pair is mainly caused by its impact force, therefore, it is necessary to study the dynamic impact force of spline pair and its suppression methods.

In this paper, based on Hertzian contact theory, a two-mass dynamic model for calculating the impact force of involute spline pair is established considering the backlash, eccentricity and pressure angle of involute spline pair. Then the dynamic equation is solved and the distributions of impact force of spline pair under various working conditions are investigated and discussed. Moreover, the influences of backlash, eccentricity and pressure angle of the involute spline pair on the impact force response are analyzed. Finally, some conclusions are summarized at the end of the paper.

2 Two-Mass Impact Model of Involute Spline

Similar to involute gear, involute spline has basic parameters such as modulus, tooth number and pressure angle. Compared with involute gear, involute spline has smaller crest height, larger backlash and pressure angle. The physical structure model of involute spline pair is shown in Fig. 1.

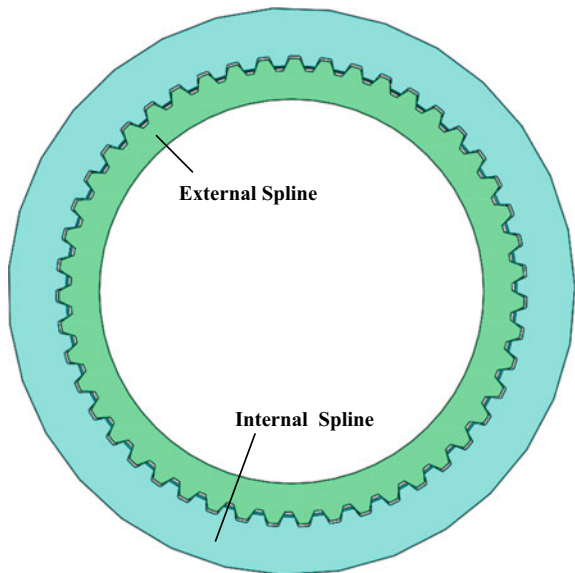
When the spline load is small or even zero, the two tooth surfaces where the external spline and the internal spline are in contact with each other produce instantaneous impact. The driven wheel will accelerate instantaneously and separate from the driving wheel, when the driving wheel gives an impact force to the driven wheel under the small load, which will make it difficult for the driving wheel and the driven wheel to keep in contact, and thus forming a periodic process.

Hertzian contact mechanics model is usually used to describe the elastic interaction between contact surfaces [13]. Based on Hertzian contact theory, the nonlinear contact force is calculated as:

$$F = K\delta^n + \lambda\delta^n\dot{\delta} \tag{1}$$

where $K\delta^n$ denotes the elastic contact force part. $\lambda\delta^n\dot{\delta}$ represents the impact damping force, and $\delta, \dot{\delta}$ are the normal relative deformation and the relative contact speed of the contact surface of the spline pair, respectively. n is a nonlinear index [14]. λ means hysteresis damping coefficient, and its value is related to viscous damping coefficient, shear and volume deformation. K denotes Hertzian stiffness, which depends on material characteristics and curvature radius, and its formula can

Fig. 1 The physical structure model of involute spline pair



be expressed as:

$$K = \frac{4}{3}ER^{1/2} \tag{2}$$

$$E = \frac{E_1E_2}{E_1(1 - \nu_2^2) + E_2(1 - \nu_1^2)} \tag{3}$$

$$R = \frac{r_1r_2}{r_1 + r_2} \tag{4}$$

where E denotes integrated modulus of elasticity. E_1 and E_2 are elastic modulus of internal and external splines, respectively, and ν_1 and ν_2 represent Poisson’s ratio of internal and external splines, respectively. R means the comprehensive radius of curvature. r_1 and r_2 are the pitch circle radii of internal and external splines, respectively. Due to the fact that the involute spline pair is internally engaged, the pitch circle diameters of the driving wheel and the driven wheel are equal. Therefore, the comprehensive radius of curvature R can be deduced as:

$$R = \frac{r_1}{2} = \frac{r_b \tan(\alpha_k)}{2} \tag{5}$$

The relationship between the damping coefficient λ and the velocity before and after impact can be determined by the energy relationship. Based on the nonlinear damping coefficient model proposed by Lankarani and Nikravesh [15], the energy loss equation is:

$$\Delta T = \frac{1}{2}m\dot{\delta}_0^2(1 - e^2) \tag{6}$$

$$m = \frac{m_1m_2}{m_1 + m_2} \tag{7}$$

where, ΔT , m , $\dot{\delta}_0$, e denotes the lost kinetic energy, equivalent mass, relative velocity before impact and recovery coefficient, respectively.

The energy loss caused by collision can also be obtained by integrating along the hysteresis loop:

$$\Delta T = \oint \lambda \delta^n \dot{\delta} d\delta \approx 2 \int_0^{\delta_m} \lambda \delta^n \dot{\delta} d\delta = \frac{2}{3} \frac{\lambda}{K} m \dot{\delta}_0^3 \tag{8}$$

The hysteresis damping coefficient under this model can be denoted as:

$$\lambda = \frac{3K(1 - e^2)}{4\dot{\delta}_0} \tag{9}$$

The equation of impact force can be written as:

$$F = K\delta^n \left[1 + \frac{3(1 - e^2)}{4} \frac{\dot{\delta}}{\dot{\delta}_0} \right] \tag{10}$$

In this paper, two running states of spline pair are discussed as follows:

(1) Normal operating condition

The normal operation condition is shown in Fig. 2, and the internal and external splines are coaxial. Set the rotational angular displacement of the external spline as θ_1 and the rotational angular displacement of the internal spline as θ_2 .

Rotational speed fluctuation of external spline is caused by torsional vibration. According to torsional characteristics, rotational speed fluctuation can be expressed by sinusoidal function:

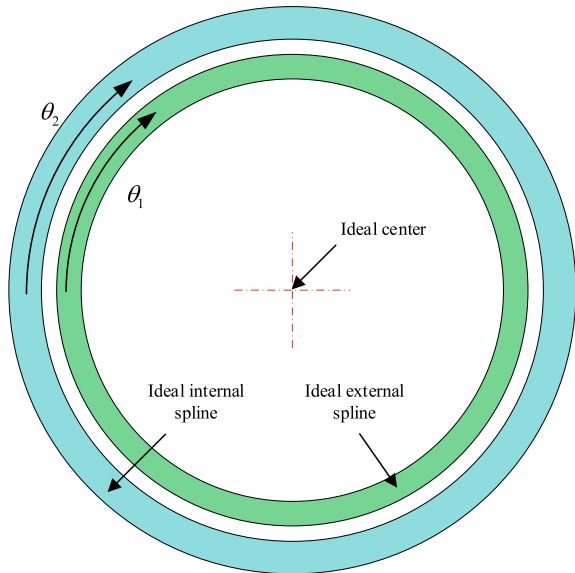
$$w = w_0 + A \sin(2\pi f t) \tag{11}$$

where w, w_0, A, f represent the input speed of external spline, the constant speed, the speed fluctuation amplitude and the speed fluctuation frequency, respectively.

The impact moment produced by the impact between the external spline and the internal spline makes the internal spline rotating, while the external spline is driven by the input rotational speed, which is less affected by the impact moment. The dynamic differential equations of external spline and internal spline are as follows:

$$\ddot{\theta}_1 = 2\pi f A \cos(2\pi f t) \tag{12}$$

Fig. 2 Simplified schematic diagram of normal operating condition of spline pair



$$\ddot{\theta}_2 = (NFR_r - c_2\dot{\theta}_2)/J_2 \tag{13}$$

where, F is the impact force, N means the meshing number of gear teeth, R_r refers to the distance from the impact force to the center, and c_2 denotes the damping coefficient.

Therefore, Eq. (10) can be rewritten as:

$$F = K \left(|\theta_1 - \theta_2|R_r - \frac{d}{2} \right)^{3/2} \left[1 + \frac{3(1 - e^2)}{4} \frac{|\dot{\theta}_1 - \dot{\theta}_2|}{\dot{\theta}_{1,0}} \right] \tag{14}$$

where, d represents tooth clearance and $\dot{\theta}_{1,0}$ means angular velocity of external spline before impact. Hertzian stiffness K can be expressed as:

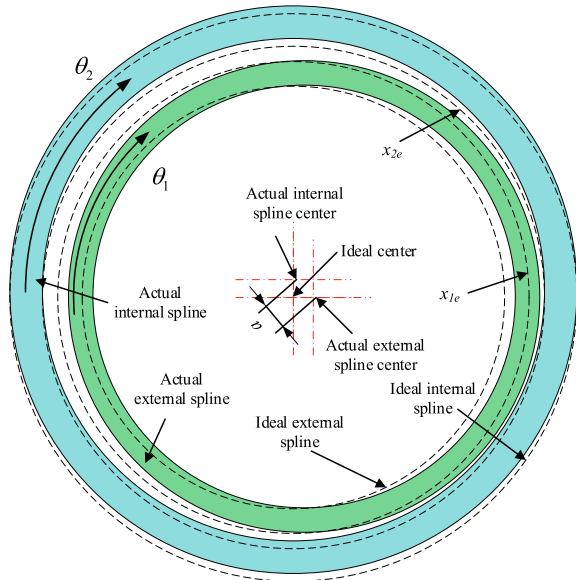
$$K = \frac{4}{3} ER^{1/2} = \frac{2Er_b \tan(\alpha_k)}{3} \tag{15}$$

(2) Axle center offset condition.

The axis offset state is displayed in Fig. 3. The centers of inner and outer splines are parallel, a denotes the eccentricity of inner and outer splines, x_{1e} refers to the translation displacement of external spline, x_{2e} represents the translation displacement of internal splines. The eccentricity a can be expressed as:

$$a = \sqrt{x_{1e}^2 + x_{2e}^2} \tag{16}$$

Fig. 3 Simplified schematic diagram of axle center offset condition of spline pair



Equation (14) can be rewritten as:

$$F = K \left(|\theta_1 - \theta_2| R_r + a - \frac{d}{2} \right)^{3/2} \left[1 + \frac{3(1 - e^2)}{4} \frac{|\dot{\theta}_1 - \dot{\theta}_2|}{\dot{\theta}_{1,0}} \right] \tag{17}$$

Due to the existence of eccentricity a , Hertzian stiffness K can be denoted as:

$$K = \frac{4}{3} E R^{1/2} = \frac{4}{3} E \{ r_b \tan(\arccos[2r_b / (2r_b / \cos \alpha_k - a)] / 2) \}^{1/2} \tag{18}$$

According to the theoretical derivation process of the two-mass dynamic model of spline pair tooth impact shown above, it can be found that the main parameters that affect the spline pair tooth impact include: spline geometry parameters, material parameters, tooth impact recovery coefficient, tooth impact relative speed and relative displacement, tooth side clearance, number of teeth participating in contact, and fluctuation amplitude and frequency of external spline rotational speed.

3 Simulation Calculation and Analysis of Spline Impact Force

The external spline input speed is defined as $n_1 = 1700$ r/min, fluctuation frequency $f = 33$ Hz, load torque $T = 0$, and correlation parameters are presented in Table 1.

In order to study the influences of the backlash, eccentricity and pressure angle on the impact force of spline pair, different parameters of backlash, eccentricity and pressure angle are adopted to jointly simulate this influences, which are displayed in Table 2. The values of the three variables are combined respectively, and a total of 1331 working conditions are obtained.

Figure 4 illustrates the solution result when the eccentricity is 0 mm, the backlash is 0.65 mm, and the pressure angle is 30°. Figure 4a shows the time domain spectrum of impact force. It can be seen that the impact force is distributed periodically, and the meshing of gear teeth has experienced a process of ‘forward collision’ and ‘reverse collision’, and the phenomenon of ‘chasing collision’ appears. The impact force when ‘chasing collision’ occurs is obviously lower than that when ‘chasing collision’ does not occur, which is due to the fact that the multiple collisions disperse the collision energy. The frequency domain spectrum of impact force is shown in Fig. 4b. When

Table 1 Spline pair parameters

Name	Module (mm)	Tooth number	Modulus (GPa)	Poisson ratio	Moment of inertia (kg·m ²)
External spline	4.0	59	210	0.3	23
Internal spline	4.0	59	210	0.3	0.24

Table 2 Joint simulation parameters

Backlash (mm)	Eccentricity (mm)	pressure angle (°)
0.25	0	20
0.30	0.02	22.5
0.35	0.04	25
0.40	0.06	27.5
0.45	0.08	30
0.50	0.10	32.5
0.55	0.12	35
0.60	0.14	37.5
0.65	0.16	40
0.70	0.18	42.5
0.75	0.20	45

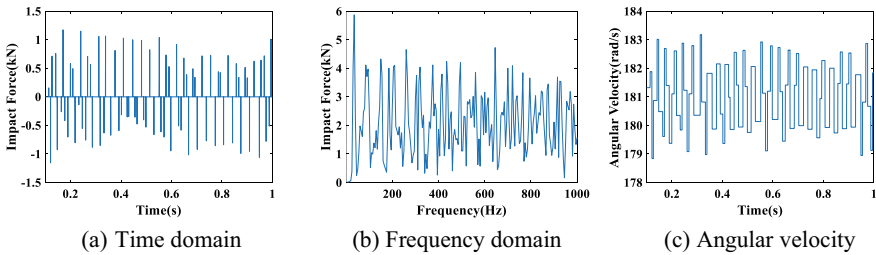


Fig. 4 The calculation results of impact force when eccentricity is 0 mm, backlash is 0.65 mm and pressure angle is 30

the frequency is 35 Hz, the maximum impact force yields 5.87 kN, which is close to the fluctuation frequency of 33 Hz. The angular velocity of the spline is shown in Fig. 4c.

Figure 5 shows the simulation diagram of peak-to-peak impact of spline pair under different backlash, eccentricity and pressure angle. It can be seen from the figure that with the increase of backlash and eccentricity, the peak-to-peak value of impact force gradually rises, and local extreme value appears. When the eccentricity is too large and the backlash is too small, the spline pair will interfere, which will be set zero in the lower right corner in the figure.

As demonstrated in Fig. 6a, the maximum curve of impact force increases with the increase of pressure angle, and the trend is slow at first and then sharp. The forward maximum and reverse maximum of impact force increase in sawtooth shape with the increase of pressure angle. Figure 6b shows the RMS value of impact force under different pressure angles, and the RMS value of impact force is the smallest when the pressure angle is 25°, which indicates that the dispersion degree of impact force is the smallest under this working condition. The two curves displayed in Fig. 6c are the mean value and absolute mean value of impact force.

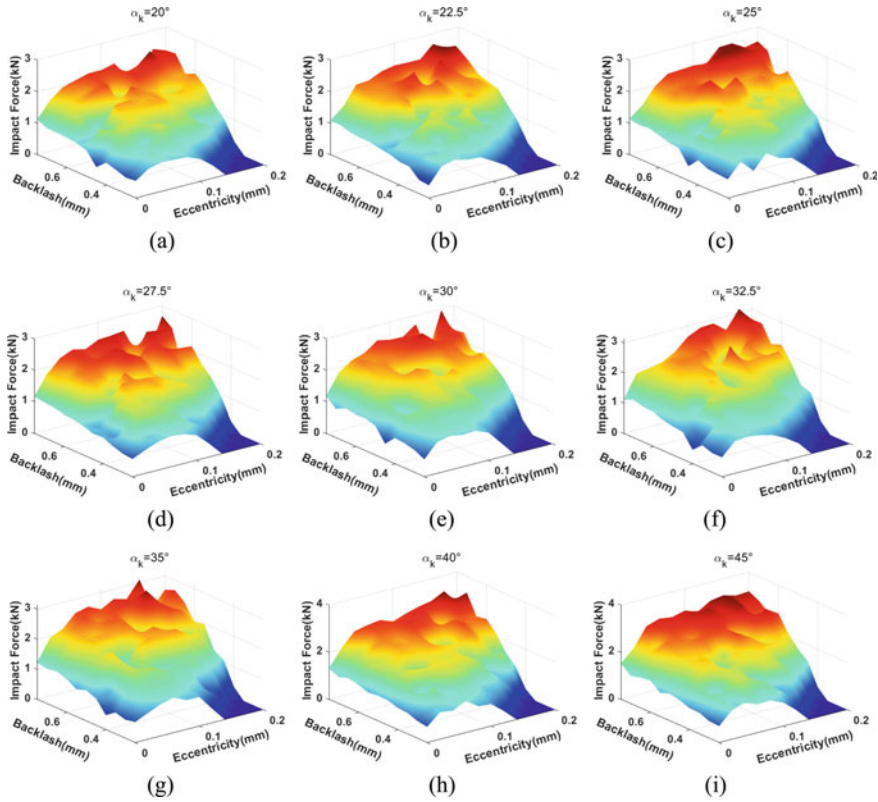


Fig. 5 Impact force peak-to-peak simulation diagram

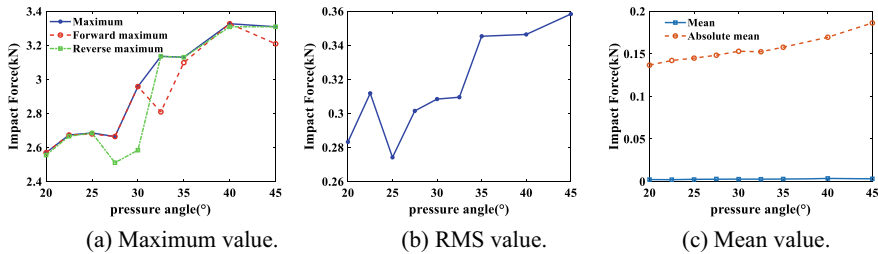


Fig. 6 Relationship between impact force and pressure angle

4 Conclusions

In this paper, based on Hertzian contact theory, a two-mass dynamic model of impact force of involute spline pair is established. Considering different backlash, eccentricity and pressure angle of involute spline pair, the dynamic equation is solved,

and the distribution of impact force of the involute spline pair under various operating conditions is obtained. The results show that with the increase of backlash and eccentricity, the impact force gradually increases, and the local extremum appears. The maximum curve of impact force increases with the raising of pressure angle, and the trend keeps slow at first and then becomes sharp. The forward maximum and reverse maximum of impact force increase in sawtooth shape with the increase of pressure angle. When the pressure angle is 25° , the RMS value of impact force is the smallest.

Acknowledgements The authors are grateful for the National Natural Science Foundation of China (Grant No. 51905053, 52035002), the Graduate Research and Innovation Foundation of Chongqing, China (Grant No. CYB21014).

References

1. Wang, L.M., Shao, Y.M.: Fault feature extraction of rotating machinery using a reweighted complete ensemble empirical mode decomposition with adaptive noise and demodulation analysis. *J. Mech. Syst. Signal Process.* **138**, 106545.1–106545.20 (2020)
2. Bao, W.G., Zhang, B.S., Nie, G.L.: Failure analysis and hardening process research of spline in wind power gearbox. *J. Mach. Manuf.* **56**(08), 84–87 (2018)
3. Kahraman, A.: A spline joint formulation for drive train torsional dynamic models. *J. Sound Vib.* **241**(2), 328–336 (2001)
4. Robins, R.R.: Tooth engagement evaluation of involute spline couplings. *J. Brigham Young Univ.* (2008)
5. Cura, F., Mura, A., Gravina, M.: Load distribution in spline coupling teeth with parallel offset misalignment. *J. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **227**(10), 2195–2205 (2013)
6. Li, C., Xi, X.P., Wu, H.Y.: Dynamic influence of backlash on meshing force. *J. Missile Space Launch Technol.* (3) (2018)
7. Zhao, G.: Study on dynamic characteristics of rotor-misaligned spline coupling system. *J. Vib. Shock* **28**(3), 78–82 (2009)
8. Zhao, G., Liu, Z.S., Ye, J.H.: Meshing force model of misaligned gear coupling and its influence on a rotor system. *J. Harbin Eng. Univ.* (2009)
9. Zhao, G., Liu, Z.S., Ye, J.H.: The misalignment meshing force model of tooth coupling and its influence on the dynamic characteristics of rotor system. *J. Harbin Eng. Univ.* **30**(001), 33–39 (2009)
10. Zhao, G., Liu, Z.S., Chen, F.: Study on the influence of spline on the stability of rotor-bearing system. *J. Vib. Eng.* **03**, 64–70 (2009)
11. Zhao, G., Su, J., Zhai, J.: Study on Nonlinear Meshing Stiffness of Spline. Springer International Publishing, Berlin (2015)
12. Zhou, C.P., Zhang, K.L.: Spline contact analysis of universal joint transmission shaft of high-speed EMU. *J. Mach.* **04**, 8–10 (2008)
13. Dai, D.C., Zhou, J.X., Zhang, J.J., Sun, W.L.: Study on impact vibration characteristics of gear transmission system. *J. Mech. Des. Manuf.* **01**, 8–12 (2019)
14. Li, J.Y., Gou, Z.J., Li, Y.: Dynamic simulation of gear force in gear meshing process based on ADAMS. *J. Mach.* **03**, 15–17 (2005)
15. Lankarani, H.M., Nikravesh, P.E.: Continuous contact force models for impact analysis in multibody systems. *J. Nonlinear Dyn.* **5**(2), 193–207 (1994)