UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Weak fault feature extraction and remaining useful life prediction for diagnostics and prognostics of rolling element bearings

by Qing Ni

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of A/Prof Jinchen Ji
A/Prof Benjamin Halkon

University of Technology Sydney
Faculty of Engineering and Information Technology

January 2023

Certificate of original authorship

I, Qing Ni, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Mechanical and Mechatronic Engineering, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Qing Ni

Production Note:
Signature removed prior to publication.

Date: 01/01/2023

Abstract

Rolling element bearings are extensively utilized in modern industry owing to their ability to support the shafts, constrain relative motion, and alleviate frictions between rotating components. In practice, the degradation and failure of rolling element bearings are inevitable as a result of continuously long operating duration, improper maintenance, suffering dynamic stress induced by the dynamic loads of rotating parts, etc. The deterioration or failure of rolling element bearings can result in unscheduled shutdowns of the whole mechanical system, leading to enormous costs due to productivity losses, maintenance, or more seriously, fatal accidents. Therefore, it is of great importance to monitor the condition of rolling element bearings for modern industry. Consequently, this thesis aims to develop a comprehensive condition monitoring framework for rolling element bearing fault diagnosis and prognosis, through which predictive maintenance can be organized in advance, the reliability of the rolling element bearings can be guaranteed, and unexpected accidents can be avoided, contributing to the safely and economically operating of the machinery.

The most frequent rolling element bearing faults are localized faults induced by local material removal. These faults can generate repetitive collisions between the healthy and faulty components during operation. Extracting the repetitive transients caused by rolling element bearing faults is one of the most critical procedures for rolling element bearing diagnosis. The key challenge in extracting these repetitive transients is that the collected rolling element bearing signals are usually multi-components and are surrounded by intensive noise. Therefore, the collected rolling element bearing fault-related repetitive transients are generally relatively weak, especially at the early stage,

and can be overwhelmingly contaminated by background noise, abnormal impulses, gear meshing vibrations, etc. To address the aforementioned challenge, two improved rolling element bearing diagnostic methodologies are proposed to enhance the weak fault information.

Besides the diagnostics of rolling element bearings, the prediction of remaining useful life (RUL) also plays a crucial role in condition monitoring. The construction of health indicator (HI) that can effectively depict the degradation characteristics is the prerequisite for accurate prediction. Most of the existing HIs can somewhat signify the deterioration trend of bearings. However, they are more or less affected by interferences from adjacent components or external environment, and the spurious fluctuations that are unrelated to degradation progression. To this end, this thesis promotes the capability of accurate rolling element bearing RUL prediction by proposing two novel HIs. Together with hyperparameters-adaptively determined networks, two prognostic methodologies that can accurately predict RUL are established.

The effectiveness of the developed methodologies for rolling element bearing diagnostics and prognostics is validated using different datasets. Successful diagnostic and prognostic results demonstrate that the proposed methodologies can generate compelling and innovative tools for rolling element bearing condition monitoring under complicated operating conditions, providing reliable maintenance decisions in modern industry to avoid economic and safe losses resulting from rolling element bearing failures.

Acknowledgments

It was my great honor to meet so many wonderful people during my PhD journey. Here, I would like to express my sincere gratitude to them.

Firstly, I would like to sincerely thank my principal supervisor, A/Prof. Jinchen Ji. He provided me with professional guidance on research projects, insightful suggestions in revising papers, valuable discussions in solving problems, and warm help in daily life. I learned not only research skills from A/Prof. Ji, but also a rigorous attitude in research and an optimistic attitude in life. In addition, I would like to thank my co-supervisor, A/Prof. Benjamin Halkon, for providing me with valuable suggestions, encouragement, and support during my PhD study.

Secondly, I would like to thank my colleagues and friends. It is great to spend such a long time with them. We worked and shared ideas together, which inspired me significantly.

Thirdly, I would like to thank the University of Technology Sydney (UTS) for providing me with the scholarship. Without the financial support from UTS, my PhD career would not be so smooth.

Last but not least, I would like to thank my family and my boyfriend, Ke Feng. Their selfless love and support have encouraged me all the time.

Qing Ni

Sydney, Australia, 2023

List of publications

J-1. **Qing Ni**, J. C. Ji, Ke Feng, and Benjamin Halkon. "A novel correntropy-based band selection method for the fault diagnosis of bearings under fault-irrelevant impulsive and cyclostationary interferences." *Mechanical Systems and Signal Processing*, 153 (2021): 107498. DOI: 10.1016/j.ymssp.2020.107498

J-2. **Qing Ni**, J. C. Ji, Ke Feng, and Benjamin Halkon. "A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis." *Mechanical Systems and Signal Processing*, 164 (2022): 108216. DOI: 10.1016/j.ymssp.2021.108216

J-3. **Qing Ni**, J.C. Ji, and Ke Feng, "Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network." *IEEE Transactions on Industrial Informatics*, 19.2 (2023): 1301-1311, DOI: 10.1109/TII.2022.3169465

J-4. **Qing Ni**, J.C. Ji, Ke Feng, Yongchao Zhang, Dongdong Lin, and Jinde Zheng, "Data-driven bearing health management using a novel multi-scale fused feature and gated recurrent unit." Under review

J-5. **Qing Ni**, J.C. Ji, Benjamin Halkon, Ke Feng, and Asoke K. Nandi, "Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics." Under review

J-6. Ke Feng, **Qing Ni**, Michael Beer, Haiping Du, and Chuan Li. "A novel similarity-based status characterization methodology for gear surface wear propagation

monitoring." *Tribology International*, 174 (2022): 107765. DOI: 10.1016/j.triboint.2022.107765

J-7. Ke Feng, J. C. Ji, and **Qing Ni**. "A novel adaptive bandwidth selection method for Vold–Kalman filtering and its application in wind turbine planetary gearbox diagnostics." *Structural Health Monitoring*, 22.2 (2023): 1027-1048. DOI: 10.1177/14759217221099966

J-8. Ke Feng, J.C. Ji, Yongchao Zhang, **Qing Ni**, Zheng Liu, and Michael Beer, "Digital twin-driven intelligent assessment of gear surface degradation." *Mechanical Systems and Signal Processing*, 186 (2023): 109896. DOI: 10.1016/j.ymssp.2022.109896

J-9. Ke Feng, J.C. Ji, Kesheng Wang, Dongdong Wei, Chengning Zhou, and **Qing Ni**, "A novel order spectrum-based Vold-Kalman filter bandwidth selection scheme for fault diagnosis of gearbox in offshore wind turbines." *Ocean Engineering*, 266 (2022):112920. DOI: 10.1016/j.oceaneng.2022.112920

J-10. Xinglong Wang, Jinde Zheng, **Qing Ni**, Haiyang Pan, and Jun Zhang. "Traversal index enhanced-gram (TIEgram): A novel optimal demodulation frequency band selection method for rolling bearing fault diagnosis under non-stationary operating conditions." *Mechanical Systems and Signal Processing*, 172 (2022): 109017. DOI: 10.1016/j.ymssp.2022.109017

J-11. Jinde Zheng, Shijun Cao, Haiyang Pan, and **Qing Ni**. "Spectral envelope-based adaptive empirical Fourier decomposition method and its application to rolling bearing fault diagnosis." *ISA transactions*, 129 (2022): 476-492. DOI: 10.1016/j.isatra.2022.02.049

Table of Contents

List of Figures

List of Tables

Abbreviation

ADMM	Alternate direction method of multipliers
AI	Artificial intelligence
AMD	Adaptively mode decomposition
AM-FM	Amplitude-modulated-frequency-modulated
BSF	Ball spin frequency
BPFI	Ball pass frequency of the inner race
BPFO	Ball pass frequency of the outer race
BLIMFs	Bandwidth-limited intrinsic mode functions
BLSTM	Bidirectional long short-term memory
BO	Bayesian optimization
BO-GRU	Bayesian optimization and gated recurrent unit
CBM	Condition-based maintenance
CMSWD	Composite multi-scale wasserstein distance
CNN	Convolutional neural network
CWRU	Case Western Reserve University

DBN	Deep belief networks
DDS	Drivetrain Diagnostics Simulator
DOF	Degree of freedom
DRS	Discrete/random separation
EEMD	Ensemble empirical mode decomposition
EMD	Empirical mode decomposition
ES	Envelope spectrum
ESC	Energy spectrum based on correntropy
EWT	Empirical wavelet transform
FCF	Fault characteristic frequency
FEC	Fault energy based correntropy
FIVMD	Fault information-guided variational mode decomposition
FIR	Finite impulse response
FK	Fast kurtogram
FTF	Fundamental train frequency
GG	Generalized Gaussian
GGCS	Generalized Gaussian cyclostationary

GGS	Generalized Gaussian stationary
GOA	Grasshopper optimization algorithm
GOA-GRU	Grasshopper optimization algorithm and gated recurrent unit
GP	Gaussian process
GRU	Gated recurrent unit
HES	Hilbert envelope spectrum
HI	Health indicator
HS	Health stage
ICS2	Indicator of second-order cyclostationarity
IFB	Informative frequency band
IMF	Intrinsic mode function
ISO	International Organization for Standardization
JS	Jensen-Shannon
KL	Kullback-Leibler
LMD	Local mean decomposition
LSTM	Long short-term memory
LP	Linear prediction

LR	Linear rectification
MAE	Mean absolute error
MFS	Machinery Fault Simulator
MOGOA	Multiobjective grasshopper optimization algorithm
MSCNN	Multiscale convolutional neural network
MSWD	Multi-scale wasserstein distance
NREL	National Renewable Energy Laboratory
PCA	Principal component analysis
PF	Product function
PHI	Physical health indicator
RCC	Ratio of cyclic content
RFCA	Ratio of fault characteristic amplitude
RMS	Root mean square
RMSE	Root mean square error
RNN	Recurrent neural network
RUL	Remaining useful life
SANC	Self-adaptive noise cancellation

SC	Spectral correlation
SDOF	Single degree of freedom
SE	Squared envelope
SES	Squared envelope spectrum
SK	Spectral kurtosis
STFT	Short-time Fourier Transform
SWT	Synchrosqueezing wavelet transform
TSA	Time-synchronous averaging
UESTC	University of Electronic Science and Technology of China
UNSW	University of New South Wales
VHI	Virtual health indicator
VMD	Variational mode decomposition
WD	Wasserstein distance
WPT	Wavelet packet transform
XJTU-SY	Xi'an Jiaotong University and Changxing Sumyoung Technology Co., Ltd.

Chapter 1. Introduction

1.1. Background

An ever-increasing demand for the reliability and safety of machinery has arisen in the modern industry due to the requirements for automation, integration, and precision of industrial innovation. As one of the most vital components in rotating machinery, bearings are commonly utilized in various modern industrial applications, such as helicopters, high-speed trains, wind turbines, processing machinery [1-3], etc. (some examples are shown in Figure 1.1). In practice, the bearings are one of the most frequent failure components in rotating machinery as a result of often operating under harsh conditions, such as continuously long work duration and heavy dynamic loads from the rotating components. According to a motor reliability survey [4], approximately 40% failure of the induction motor is derived from the bearings faults. And in one statistical report conducted by the National Renewable Energy Laboratory (NREL) of America [5], the bearing faults account for 76.2% of wind turbine gearbox failures. The deterioration of bearings can result in safety critical failures and performance degradation of the whole machinery, leading to unplanned shutdowns, increases in maintenance costs, or even fatal accidents [6-10]. Therefore, monitoring the condition of bearings is of great significance to modern industry.

Condition monitoring is a vital part of predictive maintenance (also known as condition-based maintenance (CBM)) [11]. Generally, bearing condition monitoring

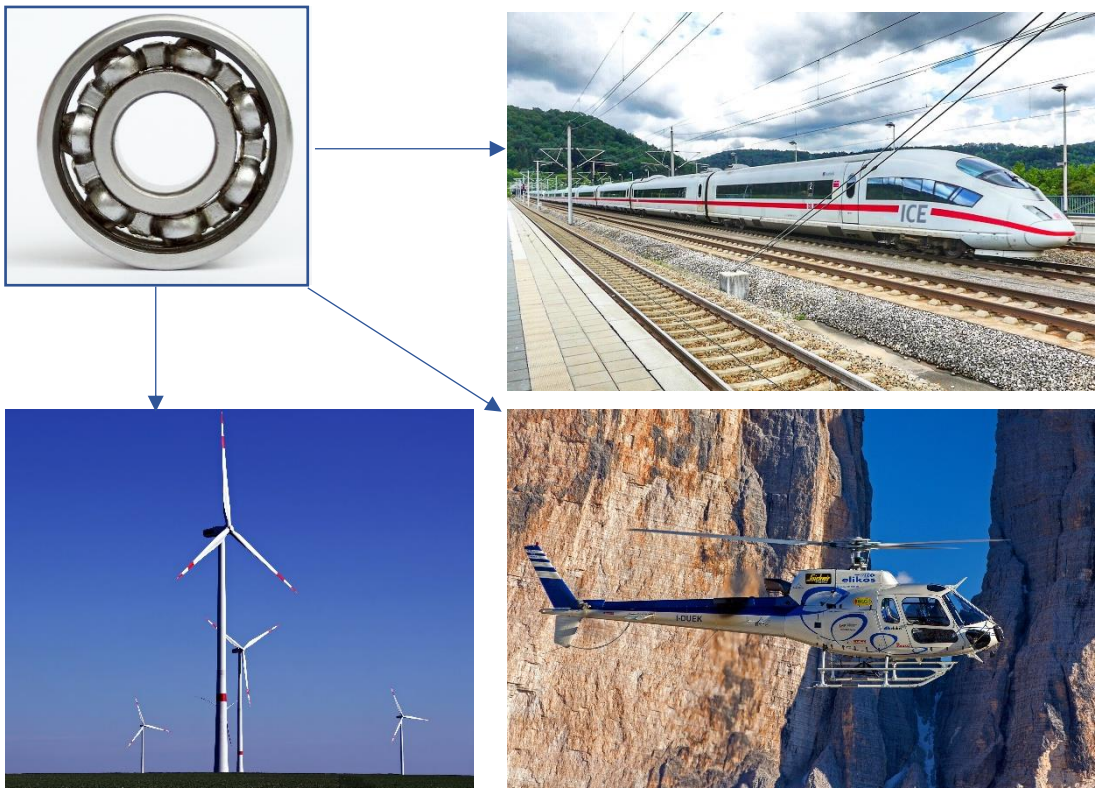


Figure 1.1. Applications of the rolling element bearings

includes two essential aspects: diagnostics and prognostics [12]. Bearing diagnostics aims at the detection, isolation and identification of bearing faults [13]. It should be noted that even though the faults of bearings are diagnosed, it's not necessary to replace them immediately unless the bearings no longer meet their operational requirements. Alternatively, the severity of bearing faults can be tracked so that a reliable prediction about how long the machine can operate safely can be made, thus allowing the maintenance to be scheduled to maximize the operational life of bearings [14]. This process is known as the prognostics of bearings, through which the remaining useful life (RUL) of the bearings is predicted. With the bearings' diagnostic and prognostic techniques, bearings' current status can be monitored and future RUL can be predicted, thus potential breakdowns of bearings can be estimated using operating data, enabling the machines to operate safely and economically. Therefore,

the condition monitoring techniques of bearings will be studied in this thesis, with a particular interest in diagnostics and prognostics.

Rolling element bearings are the most frequently used bearings in modern industrial applications. The initial failures of rolling element bearings usually result from local material removal, including pitting, spalling, etc., between contacting surfaces (ball, inner race, outer race) [15]. The local material removal, known as local faults, can generate a series of short-duration impulses. These impulses, which emerged from collisions between the faulty and healthy parts of rolling element bearings, repeat themselves periodically based on the fault location, geometry, and kinematics of the rolling element bearings. Thus the rolling element bearing faults can theoretically be identified by this frequency of repetition (known as fault characteristic frequency (FCF)). However, it is not always valid to identify the rolling element bearing faults by simply using the FCF. This is because the vibrations collected by sensors are usually close interactions between the rolling element bearings and other rotating components through distinct paths, which makes the measured signal sensitive to the position of sensors and contains entangled interferences from other vibration sources [16]. This phenomenon can be commonly found in modern industrial systems, such as wind turbine gearbox, helicopter gearbox, high-speed train gearbox, etc., where massive vibrations are generated from adjacent rotating components and mask the rolling element bearing fault signature. Hence, methodologies that can enhance the rolling element bearing fault features are usually required to be applied in advance. Techniques based on informative frequency band (IFB) selection [17], which design a filter to extract the frequency band with the richest fault information, and adaptively mode decomposition (AMD) [18] that adaptively decomposes the signal into several mode functions, are two widely utilized approaches for enhancing the weak rolling element bearing fault signature. Therefore, in this thesis, two novel methodologies focusing on the IFB selection and AMD will be proposed parallelly at first, with which the fault features can be extracted and enhanced, thus enabling the accurate diagnosis of rolling element bearings.

Diagnosing the initial faults accurately from entangled interferences is a critical step towards rolling element bearing CBM. However, immediately shutting down the

machine and replacement of the faulty rolling element bearing is not necessary on many occasions. The rolling element bearings can operate after initial failure until they no longer meet operational requirements [14]. Therefore, prognostic methodologies for rolling element bearing RUL prediction are subsequently focused on in this thesis. In the context of CBM, the rolling element bearings' prognosis is usually realized through four technical steps, namely data acquisition, health indicator (HI) construction, health stage (HS) division, and RUL prediction [19], among which the HI construction plays a vital role. Constructing a suitable HI that can reveal the deterioration characteristics is the prerequisite for accurate RUL estimation, as the predicted RUL is ultimately obtained through establishing the relationship between the HI and RUL. Constructing proper HIs is challenging as the HIs are more or less affected by the interferences from adjacent components or external environment and the spurious fluctuations that are not related to degradation progression. These phenomena can overwhelm the rolling element bearing degradation trend and thus decrease the prediction accuracy. Therefore, novel HIs that can better infer the degradation progression will be proposed in this thesis to promote the capability of accurate rolling element bearing RUL prediction.

1.2. Literature review

1.2.1.Rolling element bearing failure mechanism

This section presents the underlying science that can affect the response signal for rolling element bearing condition monitoring, with a particular interest in rolling element bearing fundamental knowledge, failure modes, and kinematics.

There are extensive categories of bearings available in modern industry, and they can generally be divided into two classes: sliding bearings and rolling element bearings [20]. Sliding bearings refer to the bearings whose contacting surfaces move to each other without the help of rolling contact, and the relative motion is aided by the lubricant. While rolling element bearings use the ball or roller between two surfaces, and the relative motion is facilitated by the cooperation of rolling motion and

lubrication [20]. Rolling element bearings are the most extensively utilized bearings in industrial systems, and thus this thesis focuses on the condition monitoring of rolling element bearings. There are various types of rolling element bearings, including ball bearings and roller bearings (cylindrical roller bearings, taper roller bearings, barrel roller bearings, etc.) commercially available, they are designed with their unique pros and can be optimally chosen according to the application [21]. For brevity, in the rest of this thesis, the bearing is utilized to refer to rolling element bearing unless otherwise specified.

Bearings may fail at any stage due to various reasons during service life. Common factors that can result in bearing failure are summarized as: improper design (inappropriate selection of bearing type and size), improper installation (misalignment, looseness, insufficient preload, excessive preload), overloading, inadequate lubricant, contaminant, repeated stress of contacting surfaces during operating, etc. [22-26]. In practice, several aforementioned failure causes usually interact simultaneously, resulting in one or more particular failure modes. According to International Organization for Standardization (ISO) 15243: 2017, bearing failure modes can be divided into six categories, including rolling contact fatigue, wear, corrosion, electrical erosion, plastic deformation, cracking and fracture [27]. If normal and ideal operating conditions of bearings are guaranteed, i.e., bearings are appropriately selected, properly installed, not overloading, adequately lubricated, and clean, fatigue will be the universal cause for failure as the repeated stress of contacting surfaces is inevitable when operating. Generally, in the early stage of bearing failure, rolling contact fatigue can result in changes both in microstructure and macrostructure [27]. Microstructure changes usually initiate from subsurface and can cause damages such as microcracks. Macrostructure changes can be manifested as material removal from surfaces, usually in the form of spalling. The macrostructure changes usually result from the microstructure changes, as a consequence of the microcracks propagating, reaching the surface, and forming spalling [27].

The local material removal between contacting surfaces can make the local fault part strike the healthy part and then generate a series of short-duration impulses. These impulses repeat themselves almost periodically and can thus be detected by this periodical frequency (known as FCF). The FCF is determined by the fault location,

geometry, and kinematics of bearings and will be briefly introduced below [24]. A typical rolling element bearing consists of four fundamental components, i.e., inner race, outer race, rolling elements, and cage. A schematic that geometrically reflects the bearing structure is drawn in Figure 1.2. The most critical geometrical quantities of the bearings are the rolling element diameter D_r, the pitch diameter D_p, the contact angle θ, and the number of rolling elements *n*.

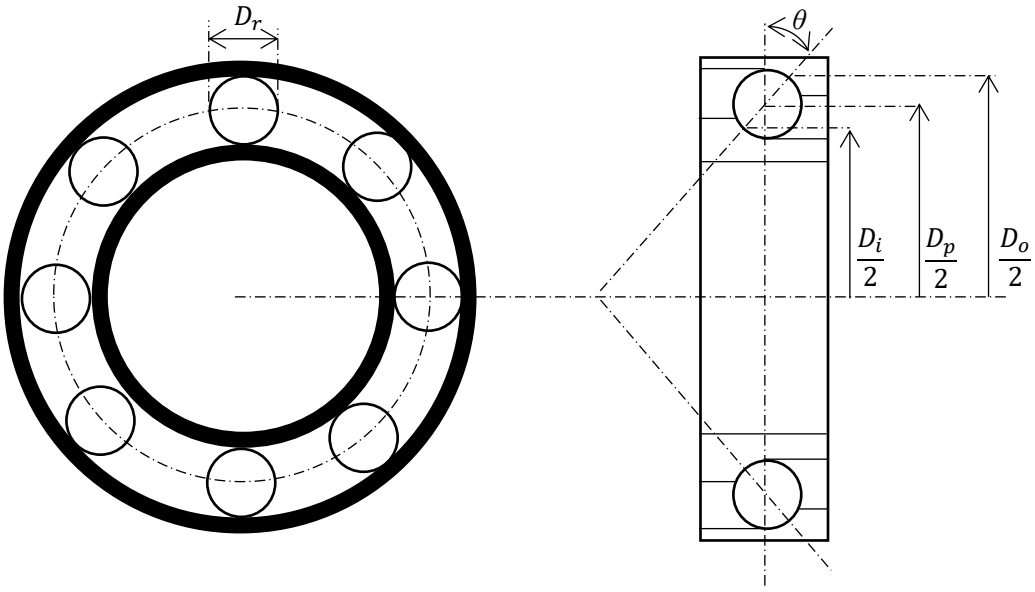


Figure 1.2. Schematic of the rolling element bearing

Assuming the rolling element bearing, as shown in Figure 1.2, operates without slip, the inner race diameter D_i and outer race diameter D_o can be expressed as [24]

$$D_i = D_p - D_r \cos(\theta) \tag{1-1}$$

$$D_o = D_p + D_r \cos(\theta) \tag{1-2}$$

The inner race and outer race circumferential velocity (m/s) can thus be calculated through

$$V_i = \omega_i D_i / 2 \tag{1-3}$$

$$V_o = \omega_o D_o/2 \tag{1-4}$$

where the ω_i and ω_o represent the angular velocity (rad/s) of inner and outer races. The cage circumferential velocity V_c (m/s) is the mean of the inner and outer races' circumferential velocities and can be expressed by

$$V_c = \frac{\frac{\omega_i D_i}{2} + \frac{\omega_o D_o}{2}}{2} = \frac{\omega_i (D_p - D_r \cos(\theta)}{4} + \frac{\omega_o (D_p + D_r \cos(\theta))}{4} \tag{1-5}$$

Transform the cage circumferential velocity (m/s) to angular velocity (rad/s) and then the frequency (Hz, cycle per second) by

$$\begin{aligned} f_c = \frac{\frac{V_c}{D_p}}{2\pi} &= \frac{\frac{\omega_i \left(1 - \frac{D_r}{D_p}\cos(\theta)\right)}{2} + \frac{\omega_o \left(1 + \frac{D_r}{D_p}\cos(\theta)\right)}{2}}{2\pi} \\ &= \frac{f_i \left(1 - \frac{D_r}{D_p}\cos(\theta)\right)}{2} + \frac{f_o \left(1 + \frac{D_r}{D_p}\cos(\theta)\right)}{2} \end{aligned} \tag{1-6}$$

where f_c is the cage frequency (also known as the fundamental train frequency (FTF)), f_i and f_o are the rotating frequencies of inner race and outer race, respectively. For bearing with *n* rolling elements, the rotating frequency for *n* rolling elements relative to the inner race, known as the ball pass frequency of the inner race (BPFI), can be calculated by

$$f_{bpfi} = n|f_c - f_i| = \frac{n|f_o - f_i| \left(1 + \frac{D_r}{D_p}\cos(\theta)\right)}{2} \tag{1-7}$$

Similarly, the ball pass frequency of the outer race (BPFO) can be calculated by

$$f_{bpfo} = n|f_c - f_o| = \frac{n|f_o - f_i|\left(1 - \frac{D_r}{D_p}\cos(\theta)\right)}{2} \quad (1\text{-}8)$$

The frequency for *n* rolling elements rotating around their axes, known as the ball spin frequency (BSF), can be calculated by $f_{bpfi} \times \frac{D_i}{D_r}$ or $f_{bpfo} \times \frac{D_o}{D_r}$, which both yield

$$f_{bsf} = \frac{n|f_o - f_i|\left(1 + \frac{D_r}{D_p}\cos(\theta)\right)}{2} \times \frac{D_p - D_r\cos(\theta)}{D_r}$$

$$= \frac{n|f_o - f_i|\left(1 - \frac{D_r}{D_p}\cos(\theta)\right)}{2} \times \frac{D_p + D_r\cos(\theta)}{D_r} \quad (1\text{-}9)$$

$$= \frac{n|f_o - f_i|D_p\left(1 - (\frac{D_r}{D_p}\cos(\theta))^2\right)}{2D_r}$$

Note that the BSF calculated by Equation (1-9) is the theoretical frequency that the rolling elements pass the same race. Generally, the rolling elements will strike both the inner race and outer race per cycle, leading to the dominance of 2 × BSF on most occasions [28].

1.2.2.Rolling element bearing diagnostic techniques

The bearing diagnostic techniques are generally developed with two purposes [29]. The first and fundamental one is to identify the status of bearings and the locations of the faults. The second purpose is to separate and/or enhance the bearing fault signature from vibrations of other components and backgrounds, which are usually dominant in the collected signal and lead to incorrect diagnostic results.

The diagnostic techniques that focus on the first purpose can be concluded into three categories: model analysis, signal processing, and artificial intelligence (AI) analysis. The signal processing technique is the most classical and universal one among them. Time domain analysis [30-32], frequency domain analysis [33-35], and time-frequency analysis [36-38] are three primarily utilized bearing signal processing methods for this purpose. In modern industry, the bearings are usually installed in complex systems, making the collected signal entangled with vibrations from other rotating components and ambient noise. These fault-unrelated vibrations can dominate the collected signal and reduce the effectiveness of diagnostic techniques that merely focus on the first purpose. Therefore, methodologies that can separate and/or enhance the bearing fault signature from vibrations of other components and backgrounds are the second purpose of bearing diagnostics and usually need to be performed in advance.

Equations (1-6), (1-7), (1-8), and (1-9) provide theoretical basis for bearing diagnostics. They are, however, kinematic frequencies based on the assumption of no slip. In practice, there is a slightly random variation of each contacting angle θ with the location of each rolling element, making the rolling elements tend to run at different speeds. While the cage guarantees all the rolling elements run at the same speed, which leads to some slight random slippage [28]. This random slippage is inevitable, and though very slight (usually 1%-2% of the FCF [28]), it is able to change the structure of the bearing fault signal, making the signal not truly periodic but cyclostationary. Figure 1.3 presents numerically generated bearings' fault signals without and with 2% random slippage. The numerically generated bearings' fault

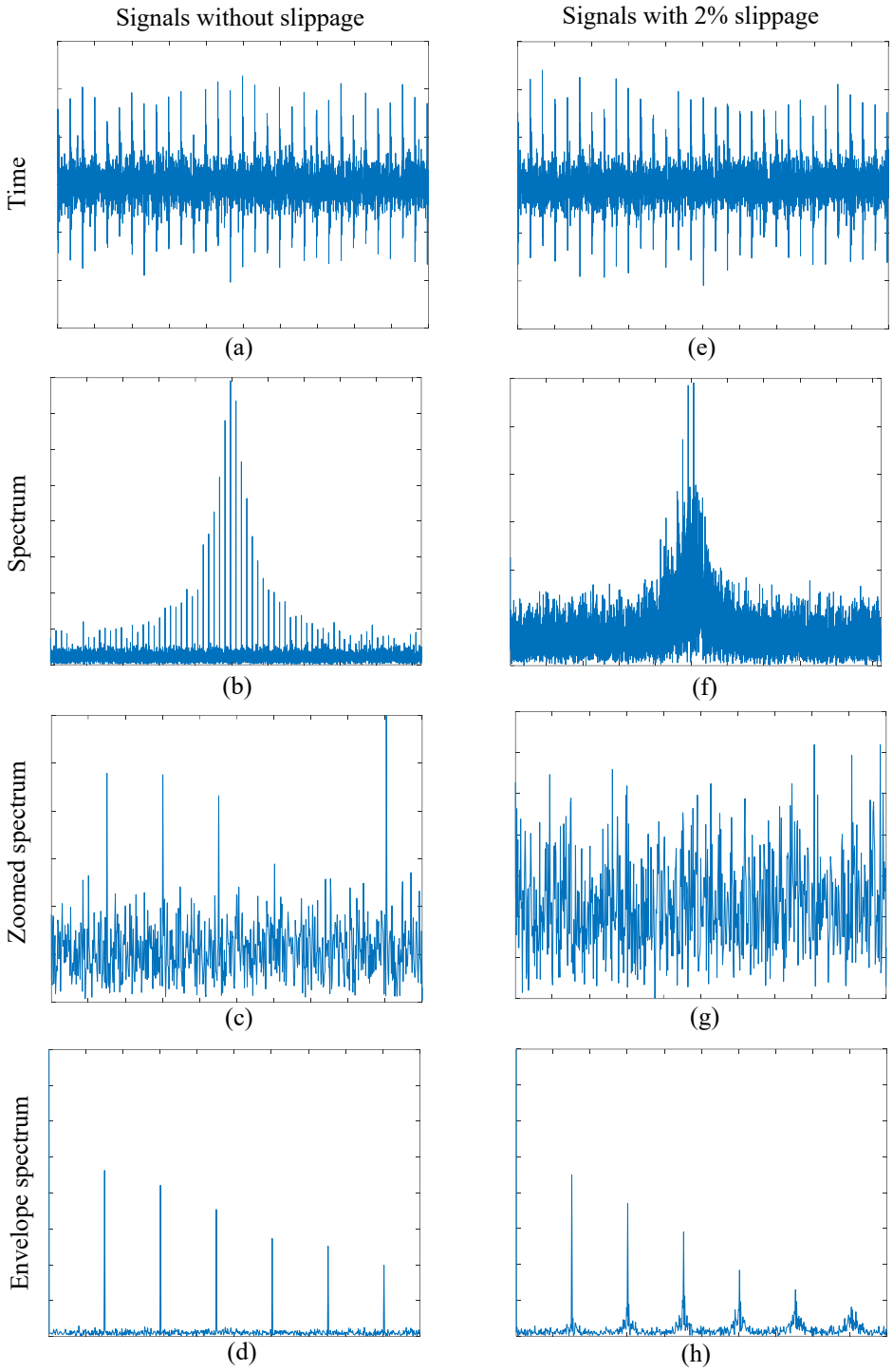


Figure 1.3. Demonstration on how slippage affect the bearing signal: (a) bearing fault signal without slippage; (b) raw spectrum of (a); (c) zoom (b) from 0 to 1024 Hz; (d) envelope spectrum of (a); (e) bearing fault signal with slight random slippage; (f) raw spectrum of (e); (g) zoom (f) from 0 to 1024 Hz; (h) envelope spectrum of (e)

signals are established based on Equation (2-11), where a tiny fluctuation τ_k between the theoretical and realistic T of k^{th} collision is used to simulate slippage. The signal

without random slippage is generated by setting τ_k to zero, while the signal with 2% random slippage is generated by setting τ_k as a random number that follows $N(0, 0.02T)$ (N represents Gaussian Distribution). The cyclostationarity of bearings renders the possibility to separate the bearing fault signatures from the truly periodic signals, known as the deterministic signals, such as gears' and shafts' signals [39]. Some typical methodologies used for this purpose are time-synchronous averaging (TSA) [40], self-adaptive noise cancellation (SANC) [41], linear prediction (LP) [42], discrete/random separation (DRS) [43, 44], and cepstrum editing [45].

Another brunch of techniques to achieve the second purpose is enhancing the bearing fault information by selecting and filtering frequency bands that bearings' fault information is the richest. The fast kurtogram (FK) [17] is the most typical as well as widely used method for locating the most informative frequency band and is usually applied together with the envelope spectrum, which is known as envelope analysis and has become a benchmark for bearing diagnostics. Besides the traditional band selection methods, AMD methodologies that can adaptively decompose the signal into several mode functions are also commonly employed for enhancing bearing fault signature. Some AMD methods have been proven to be able to act essentially as dyadic filter banks [46-48] and can be considered as adaptive filter techniques to enhance the bearing fault signature. The diagnostic part of this thesis focuses on these two categories of methodologies and will be detailly reviewed in the following sections.

1.2.3.Informative frequency band selection methodologies

There exist extensive studies on how to accurately obtain the IFB before the implementation of demodulation analysis. The kurtogram [49], as the first systematic method for IFB selection, is a pioneering solution for the band selection issue. Spectral kurtosis (SK), with the definition of the normalized fourth-order cumulant of the real or imaginary part of each spectral line after the short-time Fourier transform (STFT) of the signal [50], is the theoretical basis of the kurtogram. By taking advantage of SK, the impulsiveness can be well-described in different frequency ranges. Then in Antoni's work [51], the SK was systematically interpreted with the theory of Wold-Cramer decomposition, making it appropriate for nonstationary signal analysis. Subsequently, the practical applications of SK were further explored in the study [49], where the kurtogram based on STFT was proposed. The SK at each frequency range under different window lengths can be expressed by kurtogram. However, theoretically, there are infinite combinations of window lengths and frequencies, thus resulting in excessive computation burden and limiting the industrial uptake of kurtogram. Accordingly, the FK was proposed by Antoni [17] to reduce the computation effort. By using finite impulse response (FIR) filters with 1/3-binary tree structure, SK values can be calculated only in a limited number of bandwidths and center frequencies, which significantly improves the computational efficiency of the FK and facilitates it to be a benchmark for IFB selection in bearing fault diagnosis.

Since then, a series of FK-based selection algorithms have been proposed for the improvement of IFB selection. For example, Lei and Lin [52] developed a wavelet packet transform (WPT)-based IFB selection method, where the STFT or FIR filter was replaced by the WPT. The decomposition effect was improved significantly

owing to the improved performance of WPT over the STFT or FIR filter under non-stationary conditions. In Ref. [53], the spectrum was decomposed into bands with different initial resolutions using multi-scale clustering, through which bandwidth optimization was achieved. The results of the above-mentioned methods and other related research which focus on improving the performance of the filters or frequency bands division procedure are, unfortunately, affected by the fault-irrelevant impulsive interferences due to their impulsive-based nature [54]. In addition, these impulsive-based methods are vulnerable under fault-irrelevant cyclostationary interferences, because of the dual sensitivity to non-Gaussianity and non-stationarity of kurtosis. In study [55], Borghesani et al. demonstrated the relationship between the kurtosis and squared envelope spectrum (SES), and found that the fault-irrelevant cyclostationary interferences could make the kurtosis-based tools prone to fail. The fault-irrelevant impulsive and/or cyclostationary interferences, therefore, have significant impacts on the IFB selection procedure, and may ultimately lead to the misdiagnosis of bearing faults.

To suppress the impulsive interferences, additional methods such as the Autogram [56], alpha stable distribution-based gram [57], and protrugram [58] were recently developed. Among these methods, the protrugram gained the most attention, as it first shifted the focus from impulsiveness to cyclostationarity — one of the other essential symptoms in faulty bearings. Compared to FK, kurtosis of the time-domain envelope is replaced by kurtosis of the envelope spectrum (ES) in protrugram analysis. Moreover, the protrugram changes the variable bandwidths in FK to a fixed band which is usually applied as approximately 3-5 times the bearing FCF. Another remarkable work, coined as infogram, was proposed by Antoni [59]. This

thermodynamics-based method is constituted by the SE-infogram and SES-infogram. In the infogram, the SE-infogram is calculated through the negative entropy of the squared envelope (SE) in time-domain, with the capability of measuring the impulsiveness of the signal. Similarly, the SES-infogram is defined as the negative entropy of the SES, indicating the cyclostationaity of the analyzed signal. Upon integrating the two grams, a high impulsiveness band with repetitive behavior can be selected.

The protrugram and infogram, on the one hand, alleviate the influence of fault-irrelevant impulsive noises, on the other hand, still perform poorly when fault-irrelevant cyclostationary noise exists as a result of ignoring the realistic cyclic periods induced by bearing faults. To address this issue, the ratio of cyclic content (RCC) [55] and indicator of second-order cyclostationarity (ICS2) [60] can both be utilized to estimate the cyclostationarity of specific frequencies. The RCC and ICS2 are both calculated by summing the targeted components, but normalized by the sum of all components or the zero-frequency component in SES respectively. By considering the realistic cyclic periods, the influence from the fault-irrelevant cyclostationary noise can be alleviated to a great extent and the bearing's faults can subsequently be diagnosed correctly. However, it was demonstrated that external impulsive components with high-level energy can compromise SES [61], invalidating the RCC or ICS2-based IFB selection methods.

Based on the above literature review, the FK has been utilized as a benchmark for the IFB selection. Although designed to identify the most impulsive part of the signal, the FK is inevitably affected by the fault-irrelevant impulsive and cyclostationary interferences due to the dual sensitiveness to the impulsiveness and cyclostationarity

of the kurtosis. Some improved methods select the IFB based on cyclostationarity, which are able to suppress the influence of impulsiveness to some extent, but still surrender under high-level external impulsive interferences. The fault-irrelevant impulsive and cyclostationary interferences, therefore, may result in an inappropriate selection of band for demodulation, which remains a challenging topic to be studied.

1.2.4.Adaptive mode decomposition methodologies

The AMD method, which can decompose a signal based on the signal's characteristics, is one of the feasible options to enhance fault features in bearing fault diagnostics. Owing to the adaptive decomposition property, extensive researches have been conducted towards the adaptive signal decomposition methods, including the empirical mode decomposition (EMD) [62], local mean decomposition (LMD) [63], variational mode decomposition (VMD) [64], etc. Among various adaptive signal decomposition methods, the EMD is the pioneered one which was proposed by Huang et al. [62] in 1998. It is a decomposition algorithm that can iteratively decompose the signal into several so-called intrinsic mode functions (IMFs). Two restrictions are used to filter the IMFs: the difference between the local extremes and zero crossings in the entire time range should be zero or at most one; the average of the local maxima and minima must be zero at any time. Since then, a series of researches have been devoted to the EMD [65], studying the behaviors of EMD based on mathematical foundation and applications in practical engineering scenarios. For example, the behaviors of EMD in the stochastic situation were explored in Ref. [46] via numerically generated signals, demonstrating that the EMD can generate IMFs through a kind of dyadic filter bank resembling procedure. Yu et al. [66] decomposed

the complicated signal into several IMFs using the EMD method, together with Hilbert marginal spectrum the bearing fault was successfully diagnosed.

In addition, considering the iteratively sifting and mode mixing issues of EMD, attempts to improve the performance of EMD such as the ensemble empirical mode decomposition (EEMD) [67], LMD [63], synchrosqueezing wavelet transform (SWT) [68] and empirical wavelet transform (EWT) [69] were made in recent years. The EEMD [67] is a noise-assisted method that improves the performance of EMD by adding Gaussian white noise and multiple averages. To extract the bearing fault feature with physical meaning in the collected nonlinear and nonstationary signal, Lei et al. [70] integrated the EEMD with wavelet neural network, through which the locomotive roller bearings' faults were identified. To reduce computational burden and solve the contamination problems induced by the added noise, Xue et al. [71] proposed an adaptively fast EEMD method to determine the upper-limit frequency of the added noise, significantly saving the computational cost of EEMD for bearing diagnosis. Different from the EMD method, the LMD [63] method treats a signal as the sum of several product functions (PFs). Taking the LMD as a process for demodulation, Cheng et al. [72] extracted the modulation features of gear and bearing faults separately, achieving a successful diagnosis of the gear and bearing faults. Liu et al. [73] combined the second-generation wavelet de-noising and LMD methods to propose a hybrid bearing fault diagnosis method for the locomotive. However, the decomposition procedure of EEMD and LMD is still iterative and does not take the structure of the fault feature into account. The SWT [68] is a kind of frequency-adaptive decomposition method which shares a similar philosophy of EMD, but a different procedure that fuses the wavelet analysis and reallocation method is utilized

to construct the sub-signals. By using this novel decomposition procedure, a signal can be decomposed into several components that are the sum of several approximate harmonics. Similarly, the EWT [69] decomposes a signal into a series of IMFs using a novel wavelet filter bank, which enables the sub-signals to be more suitable to characterize the fault feature. The SWT and EWT have been successfully applied on various occasions, including quantifying the damage level [74], diagnosing the ball bearing fault [75], identifying compound faults of wheel-bearings [76] and extracting weak bearing faults of wind turbine [77]. However, the fault-related components sometimes may be slipt into different modes due to the binary band allocation [69].

To avoid the above-mentioned problems and effectively extract the fault feature in multi-component signals, the VMD [64] was proposed as an alternative to the EMD method. The VMD method can adaptively decompose a signal into several quasi-orthogonal bandwidth-limited intrinsic mode functions (BLIMFs) which are defined as a kind of function with sparsity property and most energy is concentrated around a center frequency f_c. Owing to the non-iterative calculation and adaptive narrowband determination procedures, the VMD method has demonstrated its strong capability in solving the mode mixing and misclassification problems, and thus lots of studies have been conducted towards VMD since it was put forward. Some researches were focused on verifying the properties and superiorities of the VMD. For instance, the performance of VMD under noisy and nonuniformly sampled data was studied in Ref. [47], in which the stability and robustness of VMD were demonstrated under a wider range of stochastic processes. The VMD was utilized to extract the fault feature in rotor system in Ref. [48], where the equivalent filtering characteristics were proved and better performance of VMD than EMD, EEMD and EWT in extracting multiple

signatures was demonstrated. The combination of VMD and other methods under complicated operational conditions is another research focus. Li et al. [78] integrated the VMD and the convolutive blind-source separation methods, proposing a multi-dimensional VMD technique that can track the bearing crack under large driving-speed variational condition. Multi-Teager energy operator was incorporated into VMD in Ref. [79], in which the multi-resonance bands were identified and employed to enhance the weak fault-related feature of bearings.

Although better performance of VMD than traditional decomposition methods has been demonstrated [80], the success of VMD depends highly on the pre-set parameters, including the mode number and bandwidth control parameter. Therefore, the selection of proper parameters that can optimize the performance of VMD has become a key research issue. Some researchers devoted to improving VMD performance based on the mode number. For example, an independence-oriented VMD technique [81] was proposed by utilizing correlation analysis. The modes with high correlation coefficients were used to optimize the mode number, through which the information loss and over decomposition problem was solved. Ref. [82] proposed an adaptive mode number selection method, by comparing the statistical characteristics such as the permutation entropy, extreme value and kurtosis of the decomposed modes, the proposed method was found to be less affected by modulation phenomenon when determining the mode number. These above-mentioned methods, although improve the performance of VMD to some extent, they ignore the optimization of the bandwidth control parameter, which may lead to an improper decomposition when the frequency components are close to each other and the noise is high in the collected signal [83]. Accordingly, some studies were focused on

optimizing both the mode number and bandwidth control parameter. Pu and Yang [84] proposed a precise feature extraction method that determined the mode number and bandwidth control parameter by calculating the energy ratio of the envelope spectrum between the central frequency components and BLIMFs. Jiang et al. [85] established a coarse-to-fine decomposition method in which the original signal was divided into one target coarse mode at first, then the bandwidth control parameter was optimized by selecting the BLIMF with maximum kurtosis. Furthermore, intelligence optimization algorithms are also prevalent options for the optimizing of VMD decomposition parameters. Zhu et al. [86] obtained the mode number and bandwidth control parameter synchronously through the artificial fish swarm algorithm. Considering the ability of kurtosis in revealing the impulsiveness of signals, kurtosis was employed as the objective function to optimize the VMD decomposition parameters. Zhang et al. [87] utilized the grasshopper optimization algorithm (GOA) to improve the VMD method, in which the products of the kurtosis and the correlation coefficients between BLIMFs and the original signal were used as the objective function. Miao et al. [88] also selected the optimal mode number and bandwidth control parameter through the GOA, but a different objective function — the ensemble kurtosis, defined as the product of kurtosis and envelope spectrum kurtosis, was employed. Yan and Jia [89] used the cuckoo search algorithm to determine the decomposition parameters with the mean envelope entropy as the objective function. These intelligence optimization algorithm-based methods, although taking the advantage of different optimization algorithm bases, the pivotal points are still the appropriate utilization of objective functions which are tied to fault symptom. Therefore these intelligence optimization algorithms are in essence identical to the

traditional VMD-improved methods, optimizing the performance of VMD through appropriate indicators.

Based on the above literature review, the indicators used in the improved-VMD methods can be roughly divided into three categories: the impulsiveness indicators (the most representative is kurtosis), the cyclostationarity indicators (the most representative is kurtosis of envelope spectrum) and the disorder indicators (the most representative is entropy). On the one hand, these indicators can reflect the fault symptom to some extent as the faulty bearings can generate a series of repetitive transients, thus resulting in the increase of impulsiveness, cyclostationarity and disorder. On the other hand, these indicators are blind, which means that these indicators are not solely sensitive to the bearing fault symptom. In other words, the vibration interferences from other adjacent components and abnormal impulses can also lead to the increase of these indicators, thus render the fault feature extraction of bearings under interferences, especially at the early stage, to be a challenging research task.

1.2.5.Rolling element bearing prognostic techniques

In the context of prognostic health management, the methodologies for bearing prognostics can be classified into model-based and data-driven approaches. The model-based methods represent the system's degradation progression by establishing models using mathematical knowledge. Models, such as the quadratic regression model [90], exponential model [91], Wiener-based model [92], Paris' Law model [93], Archard wear model [94], and Lemaitre damage model [94], are effective for bearing RUL estimation. The model-based approaches have demonstrated their effectiveness when appropriate mathematical models are established [95]. However, the results of the model-based approaches rely heavily on the accuracy of the models, which requires precise model parameters and high-level professional knowledge. In

industrial applications, the mechanical system is usually intricate and some model parameters are unavailable, making it incredibly challenging to establish such a precise model to reveal the system responses considering all coupling effects. On the contrary, data-driven methods can depict the degradation progression by directly utilizing the operating data [96]. The data-driven methods, therefore, are more generic used in modern industry.

In general, data-driven-based bearing prognostics involve two main approaches. The first category can be achieved through four steps, including data acquisition, HI construction, HS division, and RUL prediction [19], among which HI construction plays a vital role. Constructing a suitable HI to reveal the deterioration characteristics is the prerequisite to accurately estimate the RUL, as the predicted RUL is ultimately obtained through establishing an underlying relationship between the HI and RUL. The second category is the pure deep learning-based approach. Some recently developed pure deep-learning methods realize RUL prediction without the HI construction procedure. For example, research works [97-101] input the raw signal to predict the bearing RUL directly. That doesn't indicate that HI construction is no longer needed. Instead of manually extracting features and explicitly expressing them with HIs, these deep learning methods automatically extract the features through the network structure without expressing them explicitly, and then utilize these implicit features to predict the bearing RUL directly. Thus, these deep learning methods essentially integrate HI construction and RUL prediction together. In addition, these methods are tremendously dependent on the quality of the input data. If data are insufficient (for example, the training data are not enough to represent the overall distribution and result in distribution discrepancy between training data and testing data) or imperfect (for example, noise is included in the training data), the

performance of these methods will deteriorate. Therefore, HI construction still holds its unique advantage and will be studied in this thesis.

The HIs can be categorized into physical HIs (PHIs) and virtual HIs (VHIs) according to their construction strategies. The PHIs refer to the HIs developed based on the physical mechanism of deterioration. The root mean square (RMS) is the most commonly used HI for data-driven bearing prognostics since the growth of failure is usually accompanied by the increase of energy, and RMS is an intuitive HI to reveal the energy of the signal. The RMS was utilized as one time-domain feature to capture the degradation characteristics of bearings in Ref. [102], after which the enhanced Kalman filter combined with the expectation-maximization algorithm was utilized for RUL prediction. The RMS was improved, and its derived version — relative RMS was employed for accurate RUL estimation in Ref. [103]. In addition, the bearing fault signal can be characterized as a series of periodically repeated impulses, making impulsiveness and cyclostationarity two promising properties for bearing degradation characterization [104]. The kurtosis and the indicator of second-order cyclostationary (ICS2) are typical indicators of impulsiveness and cyclostationarity. In Ref. [105], kurtosis and the other eleven statistic features were used as the input of a support vector regression network to compare the RUL prediction results. The correlation between failure mechanism and second-order cyclostationarity was explored, then the ICS2 of the informative band was utilized for degradation progression tracking in Ref. [106]. Besides the energy-based, impulsiveness-based, and cyclostationarity-based HIs, the disorder of the system can also be adopted to capture the characteristics during deterioration, with entropy being the most representative one. In research work [107], the Rényi entropy of the bearing signal's wavelet packet coefficients was

calculated and utilized to reflect the bearing degradation trend. The sample entropy was employed to depict the degree of disorder in the wind turbine transmission system in Ref. [7], which successfully achieved severity tracking of the wind turbine bearings. Compared with PHIs, the VHIs unveil the bearing degradation progression using a virtual characterization [108]. Principal component analysis (PCA) is an extensively utilized tool to establish VHIs. For example, the PCA was employed to fuse three-channel images from multiple sensors in Ref. [109], addressing the difficulties caused by massive data. The use of machine learning methods is another branch for VHIs construction. In Ref. [110], the measured signal was analyzed to extract various statistical features, after which the recurrent neural network (RNN) was utilized to fuse these statistical features into one composite HI with a predefined failure threshold.

Another critical step in data-driven bearing prognostics is RUL prediction, which is usually achieved through utilizing machine learning methods to establish the mapping between the HI and RUL. Conventional approaches used in this step involve Gaussian process regression [111], support vector machines [112], relevance vector machines [113], etc. Deep learning methodologies, such as the convolutional neural network (CNN), deep belief network (DBN), and RNN, have also been widely applied in recent decades [114]. The Gated Recurrent Unit (GRU) is a variant of RNN with simpler architecture and can maintain historical information without eliminating former information over time, making it a superior method for time-series prediction [115]. There are extensive studies utilizing GRU for bearing RUL prediction. For example, in Ref. [116], the GRU was utilized to achieve accurate bearing RUL estimation based on the reconstructed and selected features. In Ref. [117], dual attention gates were

integrated into the GRU, through which the bearing RUL was predicted using the RMS at different moments as input. In Ref. [118], by reutilizing previous timestep's state information, the GRU was improved and employed to predict the RUL with an unsupervised Gaussian mixture model and Kullback-Leibler (KL) divergence-based HI. The GRU network will be used for RUL prediction in this thesis. However, the GRU network, as well as other machine learning methods, on the one hand, can learn the potential relationship between HI and RUL; on the other hand, the hyperparameters set in the network affect the final prediction results significantly. These hyperparameters are usually manually determined through numerous trials and errors, which are human-labor-consuming and not applicable to different operating conditions.

Based on the above literature review, to realize accurate bearing RUL prediction, constructing a proper HI is critical. To construct such a proper HI, there are two challenges that most of the existing HIs encounter. First, the bearing vibrations are usually entangled with noise, emerging from the external background or adjacent internal components. Second, there are usually spurious fluctuations in the extracted HI, which are irrelevant characteristics during the degradation progression of bearing. These two phenomena can overwhelm the underlying bearing failure characteristics and result in poor predictive performance. In addition, how to define an appropriate prediction model is also essential and needs to be further explored.

1.3. Research objectives

This thesis focuses on developing compelling and innovative techniques for bearing condition monitoring, complementing existing capabilities for bearing diagnostic and

prognostic techniques for complex systems. To achieve this aim, the following research objectives are included.

1) To develop a novel IFB selection methodology for bearings diagnostics under complex interferences.

Traditional IFB selection methods such as the FK are sensitive to impulsiveness and cyclostationarity [119], which makes them less effective for complex bearing system diagnostics. Thus a novel method to select the IFB without the influence of fault-irrelevant impulsive and cyclostationary interferences will be proposed in this thesis.

2) To develop an improved AMD method for adaptatively decomposing signals under complex interferences for bearings diagnostics.

The VMD is an advanced AMD technique. However, its success depends highly on the pre-set mode number and bandwidth control parameter. In practical engineering, heavy background noise, abnormal impulses, and vibration interferences from other internal components often bring significant challenges in selecting mode number and bandwidth control parameter [81]. Therefore, an improved VMD method will be proposed to determine the optimal mode number and bandwidth control parameter for extracting the weak bearing repetitive transient in this thesis.

3) To construct proper HIs so that the bearing degradation process can be better expressed.

Constructing a proper HI is the fundamental step to achieve high predictive accuracy. Most of the existing HIs can signify the deterioration trend of bearing to some extent. Generally, they are more or less affected by: interferences from adjacent components or external environment; and spurious fluctuations. These phenomena can overwhelm the bearing degradation information and thus decrease the prediction accuracy. Therefore, HIs that are less affected by the interferences and fluctuations will be proposed to better infer the bearing degradation process in this thesis.

4) To develop systematic prognostic schemes utilizing the proposed HIs and appropriate prediction models.

Various data-driven prediction models have been used for RUL prediction. However, the selection of hyperparameters can affect the prediction models' performance. Therefore, optimization algorithms integrated prediction models will be utilized to adaptively determine some important hyperparameters in this thesis. Together with the proposed HIs, systematic prognostic schemes with high accuracy and generalizability will be established, which may apply to different industrial situations.

1.4. Outline of the thesis

This thesis comprises six chapters. The relationship between these chapters is demonstrated in Figure 1.4. Summaries of these chapters are listed as follows:

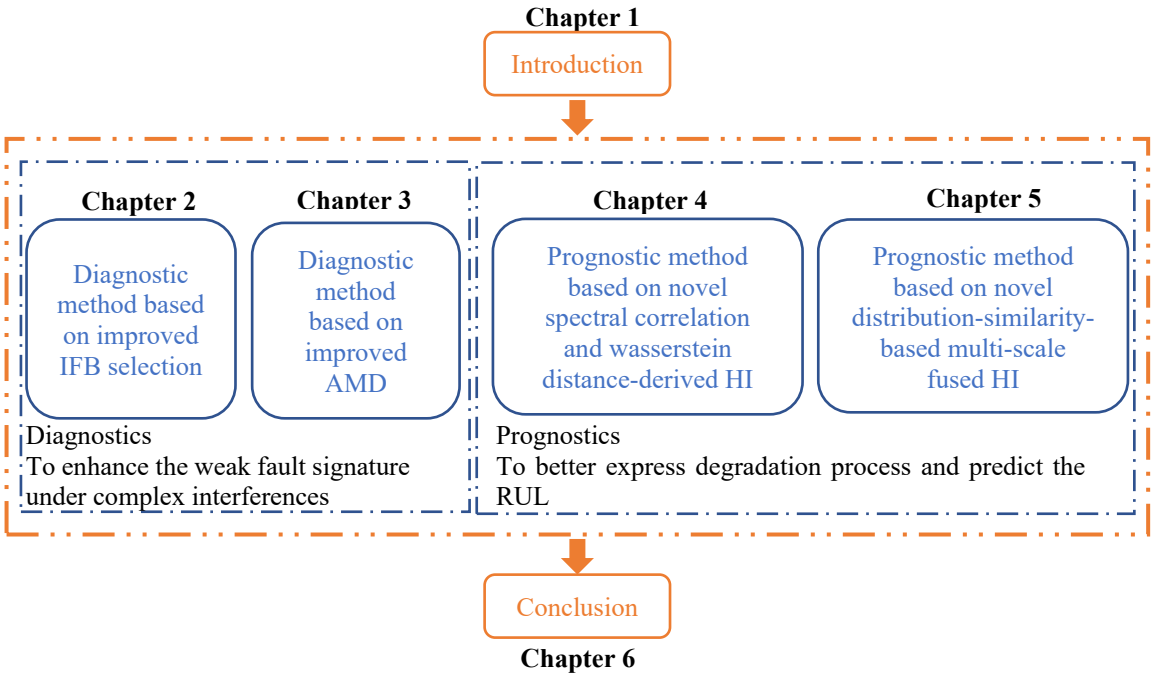


Figure 1.4. Outline of this thesis

Chapter 1: The background, literature review of bearing fault mechanism, IFB selection techniques for bearing diagnostics, AMD methods for bearing diagnostics,

methodologies for bearing prognostics, research objectives, and outline are introduced in this chapter.

Chapter 2: This chapter focuses on developing an improved IFB selection method for bearings diagnostics under complex interferences. A novel and robust IFB selection method based on the fault energy of correntropy is proposed in this chapter, through which the IFB can capture the fault symptom without being influenced by the fault-irrelevant impulsive and cyclostationary interferences.

Chapter 3: This chapter presents a fault information-guided VMD (FIVMD) method for extracting the weak bearing repetitive transient. Two nested statistical models based on the fault cyclic information, incorporated with the statistical threshold at a specific significance level, are used to approximately determine the mode number. Then the ratio of fault characteristic amplitude (RFCA) is defined and utilized to identify the optimal bandwidth control parameter, through which the maximum fault information is extracted.

Chapter 4: This chapter focuses on a systematic prognostic scheme for bearings. The proposed scheme infers the degradation progression by developing a novel HI. This novel HI, derived from the spectral correlation (SC), Wasserstein distance (WD), and linear rectification (LR), can reflect the changes in the probability distribution of all cyclic power-spectra over time. After that, the GRU network is employed to predict the RUL of the bearing system, integrated with the Bayesian optimization (BO) algorithm to tune the optimal hyperparameters adaptively.

Chapter 5: This chapter presents another novel prognostic scheme. A novel HI that fuses distribution-similarity-based multi-scale features is constructed to express the

degradation progression. More specifically, a graph named WDgram is developed to represent the distribution similarity at different frequency scales. Then the high-dimensional multi-scale features are fused to optimize the HI's monotonicity, robustness, and tenability. Subsequently, a GRU network, incorporated with the GOA to tune some hyperparameters adaptively, is established.

Chapter 6: This chapter summarizes the contents and contributions of this thesis. Future research works are also presented in this chapter.

Chapter 2. A novel correntropy-based informative band selection method for bearing diagnostics

2.1. Introduction

This chapter is a modified version of the journal paper titled "A novel correntropy-based band selection method for the fault diagnosis of bearings under fault-irrelevant impulsive and cyclostationary interferences", which was published in Mechanical Systems and Signal Processing (153 (2021): 107498). The work completed in this chapter is dedicated to realizing objective 1 of this research.

Based on the literature review of Chapter 1.2.3, being a classical yet efficient approach, the FK plays an important role in bearing fault diagnosis. However, the IFB selection results are significantly affected by the fault-irrelevant impulsive interferences. Some improved IFB selection methods, such as the protrugram, are able to suppress the influence of impulsiveness to some extent, but still surrender under high-level external impulsive interferences [119]. Furthermore, both FK and protrugram are unable to address the issue of locating IFB properly under fault-irrelevant cyclostationary interferences, which remains a challenging topic to be studied. To this end, a novel method to select the IFB without the influence of fault-irrelevant impulsive and cyclostationary interferences is proposed in this chapter.

By taking advantage of the anti-impulsive property of correntropy, a novel indicator named fault energy based correntropy (FEC) is proposed in this chapter. The FEC, calculating the energy at the targeted FCFs based on correntropy, can solely represent

the bearing fault feature and is not affected by the impulsive interference. Furthermore, an IFB selection method called FECgram is established to overcome the drawbacks of the traditional band selection methods. In the FECgram, the signal is decomposed into sub-frequency-bands using a FK-identical 1/3-binary tree at first, then the FEC of each sub-signal is computed. By using the proposed indicator FEC, the FECgram can detect the energy of the targeted fault frequency without being influenced by fault-irrelevant impulsive and cyclostationary interferences. This renders the FECgram to be an optimal method that can accurately diagnose the bearing's faults in most complex working environments.

The chapter is organized as follows: the FK and protrugram techniques are reviewed in Chapter 2.2, which aims to provide the theoretical basis to interpret two typical (impulsive-based and cyclostationary-based) IFB selection methods. Chapter 2.3 first introduces the fundamentals of correntropy, and then proposes a robust indicator FEC to characterize the energy of targeted FCFs without the influence of fault-irrelevant impulsive and cyclostationary interferences. Furthermore, a novel band selection method named FECgram is established to select the optimal IFB. Chapter 2.4 uses simulation data to validate the effectiveness of FECgram under different levels of impulsive and cyclostationary noise, thereby presenting a comprehensive study of the performance of the FECgram. In Chapter 2.5, three sets of challenging experimental data from three different test rigs, including bearing signals with high-level impulsive noise and outer/inner race faulty planet bearings, are used to demonstrate the superiority of the FECgram. The conclusion and further work are summarized in Chapter 2.6.

2.2. The traditional IFB selection methods

This section briefly reviews the FK and protrugram techniques for impulsive-based and cyclostationary-based IFB selection methods.

2.2.1. The classical impulsive-based band selection method: FK

For the sake of subsequent analysis, this sub-section provides background information on kurtosis, spectral kurtosis, kurtogram, and FK.

2.2.1.1. The kurtosis and spectral kurtosis

Kurtosis, known as the theoretical basis of the FK, is commonly defined as

$$K = \frac{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{N}}{\left[\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}\right]^2} \tag{2-1}$$

where x_i is the raw signal, $\bar{x}$ denotes the mathematical expectation of x_i, N represents the sample number in x_i. When $K = 3$, x_i obeys the Gaussian distribution, which refers to the healthy status of the machinery. The higher value of K gives an indication of the existence of impulsive components, implying the possibility of machine failure [120].

Owing to its impulsiveness-detection ability, kurtosis is commonly used in the condition monitoring of rotating machinery, but it is a global statistical index. The SK was subsequently designed to address the inability of kurtosis to express impulsiveness at each frequency line. The definition of SK can be given by

$$K_x(f) = \frac{\frac{\sum_{t=1}^{N_t}|X(t,f)|^4}{N_t}}{\left[\frac{\sum_{t=1}^{N_t}|X(t,f)|^2}{N_t}\right]^2} \quad (2\text{-}2)$$

where the $X(t,f)$ is the STFT of x_i and N_t represents the number of time segments.

2.2.1.2. The kurtogram and the FK

The detection of impulsiveness in different frequency ranges can be achieved through the SK. It should be noted that a window length needs to be selected to calculate the STFT, and different window lengths can result in different STFT values. Therefore, the SK is actually a function of the frequency and window length, and then the kurtogram was proposed in Ref. [49], using a three-dimensional graph to represent the values of SK according to frequency and window length.

Theoretically the combinations of frequencies and window lengths are infinite, and this brings excessive computational burden to the kurtogram and limits its applications in practical engineering domain. To relieve the kurtogram from the exorbitant computational burden, the FK was established by Antoni [17], utilizing FIR filters with a 1/3-binary tree structure (shown in Figure 2.1). To calculate SK values in the FK, a few combinations of center frequencies and bandwidths which depend on level l are employed. In levels $l = 1, 2, 4, 6, 8, \ldots$, the signal is divided into $N = 2^{\left\lfloor \frac{l}{2} \right\rfloor}$ segments, with bandwidths $b_w = 2^{-(1+\left\lfloor \frac{l}{2} \right\rfloor)} \times F_s$, and center frequencies $f_c = [(n + \frac{1}{2}) \times 2^{-(1+\left\lfloor \frac{l}{2} \right\rfloor)}] \times F_s$ $(n = 0, 1, 2, 3, \ldots, N-1)$, where F_s is the sampling frequency. Similarly, in levels $l = 3, 5, 7, 9, \ldots$, the signal is divided into $N = 3 \times 2^{\frac{l-3}{2}}$ segments, with bandwidths $b_w = \frac{1}{3} \times 2^{\frac{1-l}{2}} \times F_s$, and center frequencies $f_c = [(n +$

$\frac{1}{2}) \times \frac{1}{3} \times 2^{\frac{1-l}{2}}] \times F_s$. The SK value in FK calculated at each combination of center frequency and bandwidth is defined as

$$SK_{f_c,b_w} = \frac{\langle |C_{f_c,b_w}|^4 \rangle}{\langle |C_{f_c,b_w}|^2 \rangle^2} \quad (2\text{-}3)$$

where C_{f_c,b_w} is the filtered signal and $\langle \quad \rangle$ represents the mathematical expectation operator.

The improvement in computational efficiency contributes to the extended application of FK, making it a benchmark approach for IFB selection in bearing condition monitoring.

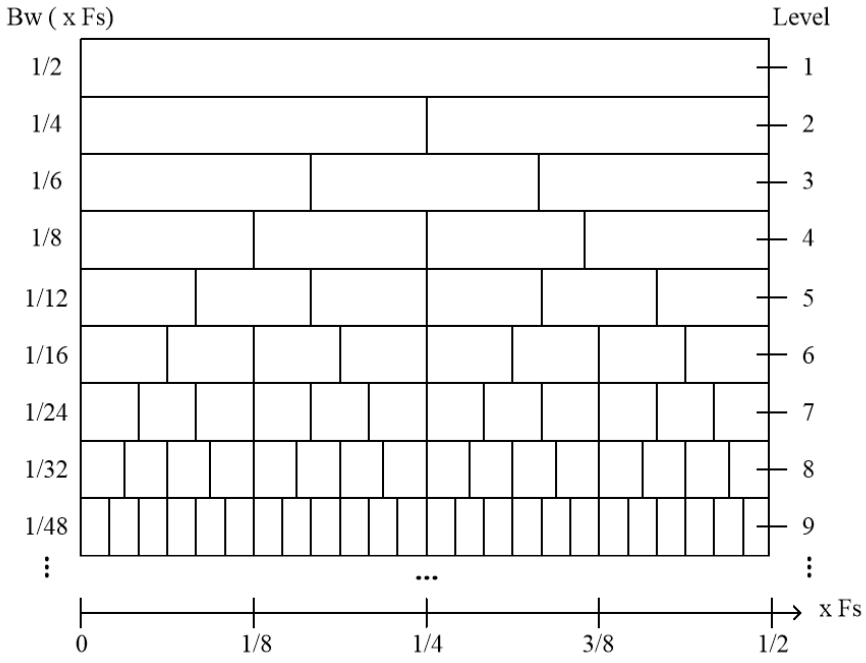


Figure 2.1. The structure of 1/3-binary tree

2.2.2. The classical cyclostationary-based band selection method: Protrugram

Different from the FK, the protrugram [58] focuses on searching the cyclostationarity of the analyzed signal. By replacing the kurtosis of the time-domain envelope to envelope spectrum, the protrugram first shifts attention from impulsiveness to another essential feature (namely the cyclostationarity) of faulty bearings and is often recognized as the improved version of the FK. In addition, the variable bandwidth in the FK is replaced by a fixed narrowband which is suggested as 3-5 times bearing's FCF. In short, in protrugram, the center frequencies are calculated through $f_c = \frac{(2n-1)\times B_w}{2} - (n-1) \times B_{wo}$, where B_{wo} is the overlap bandwidth between two adjacent f_c and $n = 1, 2, \dots, \left\lfloor \frac{\frac{F_S}{2} - B_{wo}}{B_w - B_{wo}} \right\rfloor$. Then the kurtosis for each ES after filtering is computed by

$$K_{f_c,b_w} = \frac{\langle \left| ES_{f_c,b_w} \right|^4 \rangle}{\langle \left| ES_{f_c,b_w} \right|^2 \rangle^2} \quad (2\text{-}4)$$

The filtering procedure of protrugram is shown in Figure 2.2. The shifting to the envelope spectrum domain enables the protrugram to be sensitive to fault-symptomatic cyclostationarity. However, it could be biased by fault-unrelated cyclostationary interferences which generate peaks in the envelope spectrum.

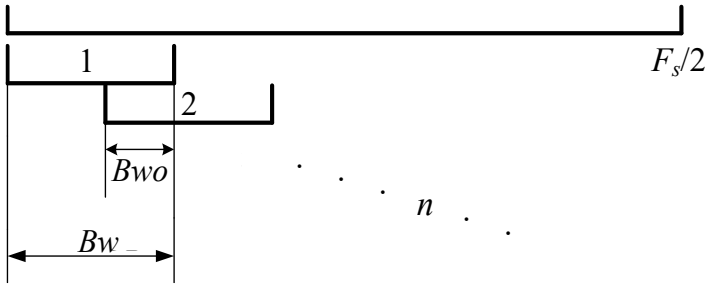


Figure 2.2. The filtering procedure of protrugram

2.3. The novel correntropy-based band selection scheme: FECgram

The FK and protrugram, on the one hand, select the optimal IFB from two essential fault-symptomatic features (impulsiveness and cyclostationarity), which facilitates them to be two typical approaches for demodulation band selection. On the other hand, the FK and protrugram can be affected by fault-irrelevant impulsive and cyclostationary interferences to a greater or lesser extent, as a result of their theoretical basis and their dual sensitivity to impulsiveness and cyclostationarity. To address this issue, a novel correntropy-based method named FECgram is proposed and discussed in detail below.

2.3.1. Correntropy

The correntropy has been recently adopted to suppress impulsive noise, thus providing the theoretical feasibility to develop a robust indicator for the signal in the presence of high impulsive noises. The correntropy proposed by Santamaria et al. [121] in 2006, which can be regarded as the correlation function defined in the reproducing kernel Hilbert space, has shown its strong ability to deal with signals under impulsive noise by adjusting its kernel function. Since then, numerous studies have been conducted

focusing on the development of theory and application of correntropy, including local similarity measure [122], non-linear test [123], time delay estimation [124], and image processing [125]. However, to the best of our knowledge, only a few studies relevant to correntropy have been published in the area of machinery condition monitoring. For example, the cyclic correntropy spectrum was used to diagnose the bearing fault under the impulsive noise environment in Ref. [126]. The correntropy of the intrinsic mode function was used for demodulation in Ref. [127]. A novel indicator based on the frequency components on the informative bands and cyclic correntropy was proposed for gear wear degradation tracking [128]. Unfortunately, the use of correntropy in condition monitoring is still limited and needs further exploration.

Correntropy, a recently developed measure to determine the local similarity of a stochastic process, is defined as

$$C(t_1, t_2) = \langle k(x_{t_1} - x_{t_2}) \rangle \tag{2-5}$$

where $k(x_{t_1} - x_{t_2})$ is a kind of positive-definite function that obeys the Mercer condition and is usually expressed as a Gaussian kernel function on account of its strict positive-definiteness and smoothness. The Gaussian kernel function is written as

$$k(x_{t_1} - x_{t_2}) = \frac{1}{\sqrt{2\pi} \times \sigma} e^{-\frac{\|x_{t_1} - x_{t_2}\|^2}{2\sigma^2}} \tag{2-6}$$

where σ denotes the kernel size and should be chosen carefully as it can affect the Gaussian kernel function directly and significantly. A widely used criterion for kernel size selection is the Silverman criterion. The kernel size calculated by the Silverman

criterion is simultaneously constrained by the dynamics of the stochastic process and the number of samples. The Silverman criterion is expressed as

$$\sigma = 0.9AN^{-\frac{1}{5}} \tag{2-7}$$

where A is the minimum empirical standard deviation and N denotes the sample number of the stochastic process. Expanding Equation (2-5) into a Taylor series yields

$$C(t_1, t_2) = \frac{1}{\sqrt{2\pi} \times \sigma} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} \langle \left\| x_{t_1} - x_{t_2} \right\|^{2n} \rangle \tag{2-8}$$

Equation (2-8) shows that all the even-order moments of $x_{t_1} - x_{t_2}$ are involved in the correntropy, which makes it a superior approach to the conventional autocorrelation function. This is because not only the second-order moments but also all the higher even-order moments are contained in the correntropy. And all the even moments, in fact, can contribute to the final estimation of local similarity owing to the nonlinearity generated by the Gaussian kernel.

2.3.2. The novel indicator FEC and its advantages

Due to its aforementioned property, correntropy can successfully suppress impulsive noise, even large outliers, thus providing theoretical feasibility to develop a robust indicator in high impulsive noise scenarios. However, the cyclostationary interferences induced by fault-irrelevant components can still invalidate the correntropy in fault-symptomatic components extraction. To mitigate the fault-

irrelevant cyclostationary interferences, a robust indicator to calculate the energy at targeted FCFs based on correntropy is derived in the subsequent section.

The indicator to demonstrate the fault energy based on the correntropy is defined as

$$FEC = \frac{\sum_{h=1:H} \max_{n \in TF_h} ESC(n)}{\sum_{n=0}^{n=\left\lfloor \frac{L}{2}-1 \right\rfloor} ESC(n)} \quad (2\text{-}9)$$

where h denotes the h^{th} harmonic and H is the number of harmonics to be analyzed. Due to the frequency deviations caused by slippage [129], TF_1 are considered as a series of targeted frequencies (e.g. a series of frequencies with the center frequencies at BPFO, BPFI, BSF, surrounded by a tiny tolerance band separately), and TF_h are a series of frequencies corresponding to the h^{th} harmonic. In the case of bearings, TF_1 are set as the theoretical calculated bearing FCF with 5 times the frequency resolution as tolerance band. L represents the sample length of the analyzed signal and $ESC(n), n = 1, 2, \dots, N$, denotes the energy spectrum based on correntropy (ESC) and can be written as

$$ESC(n) = |\mathcal{F}[correntropy(|x(t)|^2)]|^2 \quad (2\text{-}10)$$

where $\mathcal{F}$ represents the Fourier transform and $x(t)$ is the signal to be analyzed in the time domain.

From Equation (2-9) and Equation (2-10), it can be concluded that the fault-symptomatic components can be solely extracted by the proposed FEC. Incorporated with the insensitivity to impulsiveness property of correntropy, the superiorities of FEC can be summarized as:

1) It is not affected by impulsive noises, even large outliers.

2) It is insensitive to fault-irrelevant cyclostationary interferences as a result of solely calculating the energy at targeted frequencies using the prior knowledge of the FCFs.

2.3.3. The proposed FECgram and its calculation procedure

In this section, a novel correntropy-based method named FECgram is proposed to overcome the shortcomings of the traditional band selection methods, in which the FEC is employed as the criterion for the IFB selection. The calculation procedure of the FECgram, as shown in Figure 2.3, is discussed below.

Step 1. The raw signal is split into a series of frequency bands with specific center frequencies and bandwidths (i.e. x_{f_c,b_w}). Owing to the high efficiency and extensive acceptance, a similar filter structure to that of the FK, presented in Figure 2.1, is utilized in this chapter. By employing the 1/3-binary tree structure, the same filtered version can be obtained.

Step 2. The correntropy of each squared envelope $\left|x_{f_c,b_w}\right|^2$ after filtering is calculated. Then each energy spectrum ESC_{f_c,b_w} is computed, allowing the establishment of FECgram.

Step 3. Fault energy of targeted FCFs based on the correntropy (i.e. FEC_{f_c,b_w}) is calculated in each filtered signal through $\frac{\sum_{h=1:H} \max_{n\in TF_h} ESC_{f_c,b_w}(n)}{\sum_{n=0}^{n=\left\lfloor \frac{L}{2}-1 \right\rfloor} ESC_{f_c,b_w}(n)}$. It should be noted that there are potential possibilities of damages in the outer race, the inner race and the balls. Therefore, the fault energies of the outer race, inner race and balls should all be calculated separately. In addition, a tiny tolerance band (5 times the frequency

resolution) around each BPFO, BPFI, BSF, and their harmonics is set in case of deviation between actual and theoretical FCFs.

Step 4. A colormap in terms of the center frequency and bandwidth plane is established. The novel method, named FECgram, utilizes the FEC to substitute the traditional kurtosis and can select the optimal IFB using the center frequency and bandwidth with the maximum FEC.

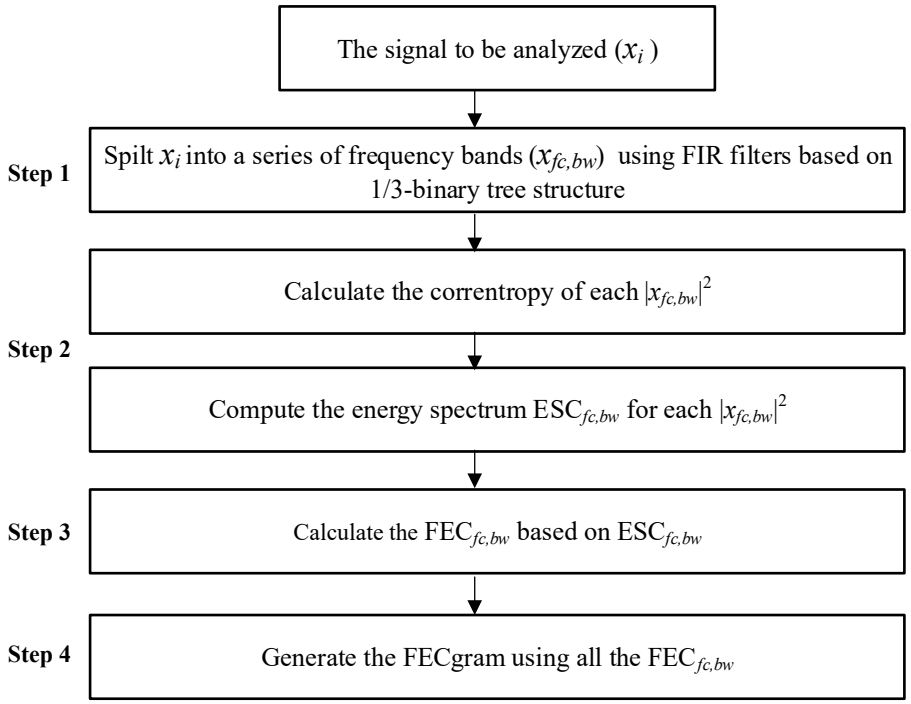


Figure 2.3. The calculation procedure of FECgram

2.4. Validation using numerically generated signals

With the incorporation of FEC, the proposed FECgram can detect the targeted fault energy which eliminates the influence of fault-irrelevant impulsive and cyclostationary interferences, and thus it is an optimal method for IFB selection. In this section, the performance of the FECgram will be validated through numerically generated bearing fault signals under different levels of two typical interferences (i.e.

impulsive and cyclostationary interferences) which often render the FK and protrugram ineffective.

2.4.1. Numerically generated bearing fault signals

The numerical signals of bearing fault used in this section are generated from the following equation [130]

$$x_{bf}(t) = \sum_{k} A_k O(t - kT - \tau_k) + n(t) \tag{2-11}$$

where k is the k^{th} collision, T represents the theoretical value of one fault cyclic period, τ_k indicates the tiny fluctuation between the theoretical and realistic T of k^{th} collision (note: the location of the bearing fault is not changed), A_k denotes the amplitude of k^{th} collision and is given by $A_k = \cos(2\pi \times f_r)$ with f_r being the shaft rotating frequency. $n(t)$ represents the white Gaussian noise induced by the external environment. Furthermore, $O(t)$ represents a kind of oscillating attenuation function determined by the natural frequency f_n and damping ratio ζ. In this chapter, a single degree of freedom (SDOF) system is utilized for illustration purpose, and $O(t)$ can thereby be expressed as

$$O(t) = X_0 e^{-2\pi\zeta f_n t} \sin\left(2\pi\sqrt{1-\zeta^2} f_n t + \varphi_0\right) \tag{2-12}$$

where X_0 denotes the initial displacement and φ_0 represents the initial phase. The parameter values utilized in this simulation study are listed in Table 2-1.

Table 2-1 The parameters utilized in the numerically generated bearing fault signals

ζ	f_r (Hz)	f_n (kHz)	X_0 (m)	φ_0 (°)	τ_k (s)	FCF (Hz)	Signal to noise ratio (dB)
0.02	32	6	1	0	$\sim N(0, 0.01T)$ N denotes Gaussian distribution	150	-6

2.4.2. Validation under different levels of impulsive interferences

The performance of the proposed FECgram under different levels of impulsive interferences will be evaluated in this section. Two classical and fundamental IFB selection methods — FK and protrugram (bandwidth at 3.5 times FCF) will be utilized for comparison. The numerically generated bearing fault signal shown in Chapter 2.4.1 is used and the impulsive interferences are set to be several non-periodic random impulses with different weights. The bearing fault signal with different levels of impulsive interferences can be presented as

$$X(t) = x_{bf}(t) + \sum_{i=1}^{10} x_i(t) \tag{2-13}$$

where $x_i(t)$ denotes several non-periodic impulses. They are modeled to be impulse responses of a SDOF system which has the same expression as Equation (2-12). The damping ratio, natural frequency and initial phase of these impulses are set to be 0.02, 8 kHz and 0° respectively, while the initial amplitudes are assumed as

$$X_{0i} \sim U[0, num] \tag{2-14}$$

where U denotes the uniform distribution and num is the scale factor which enables the simulation of different levels of impulsive interferences by changing the initial

amplitudes. For illustration purpose, the simulated bearing signal and the fault signal with impulsive interferences ($num = 5$) are shown in Figure 2.4.

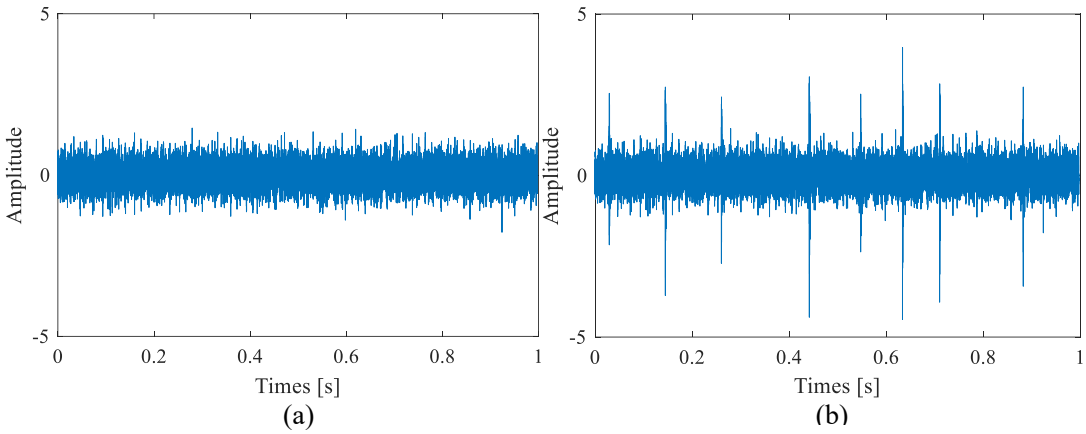


Figure 2.4. (a) Simulated bearing fault signal; (b) the fault signal with impulsive interferences when *num* = 5

To evaluate the performance of the FECgram, the scale factor num is set from 1 to 100. By increasing num, while the other parameters remain unaltered, the circumstances of unchanged fault-symptomatic components with higher and higher fault-irrelevant impulsive interferences can be simulated. The FK, protrugram and FECgram are then applied in each simulated signal for IFB selection, and the suggested bands for demodulation are displayed as the function of num (shown in Figure 2.5). It can be found that the FK selects IFBs at around 8 kHz (which is near the natural frequency of impulsive interferences) at first, then locates IFBs at around 9 kHz and 2 kHz, all of which do not include the natural frequency of the numerically generated bearing fault signal. The protrugram selects IFBs at around 6 kHz when $num < 57$, but above this it locates IFBs fluctuating around 8 kHz. On the contrary, the FECgram consistently provides IFBs at around 6 kHz in this test, which is perfectly in conformity with the natural frequency of the numerically generated bearing fault signal.

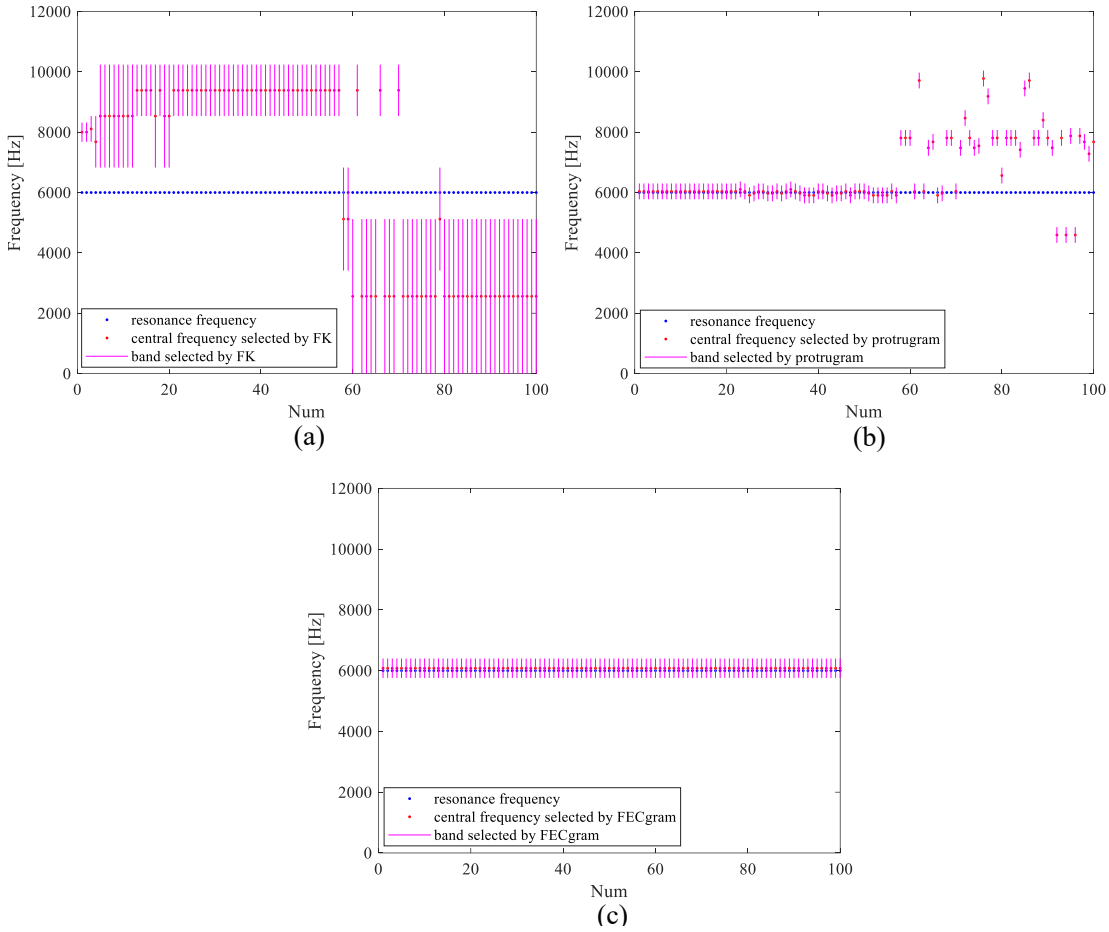


Figure 2.5. The selected IFBs under different levels of impulsive interferences using: (a) FK; (b) protrugram; (c) FECgram

In short, FK becomes ineffective when $num = 1$, which shows that the FK is very sensitive to the impulsive interferences. The protrugram cannot select bands around the bearing natural frequency when $num \geq 57$, showing it is not sensitive to the impulsive interferences to a large extent. In other words, the protrugram can suppress impulsive interferences to a great extent, which makes it effective on most occasions under impulsive interferences, but extremely high levels of external impulsive interferences can overwhelm it. The novel FECgram is not sensitive to the impulsive interferences and performs robustly in this simulated signal test, thus making it a superior method for IFB selection under impulsive interferences.

2.4.3. Validation under different levels of cyclostationary interferences

The performance of FECgram under different levels of fault-irrelevant cyclostationary interferences will be evaluated in this section. The numerically generated bearing fault signals with cyclostationary interferences are set as

$$X(t) = x_{bf}(t) + num \times x_c(t) \tag{2-15}$$

where $x_c(t)$ denotes fault-irrelevant cyclostationary interferences which possess the same expression as Equation (2-11), but with different parameters including $A_k = 1$, $T = 10$ Hz, $f_n = 8$ kHz and $n(t) = 0$. Again, num represents the scale parameter, by increasing it and keeping other parameters unaltered, the unchanged bearing fault components under different levels of cyclostationary interferences can be emulated. For illustration purpose, the simulated bearing fault signal with cyclostationary interferences when $num = 4$ is shown in Figure 2.6.

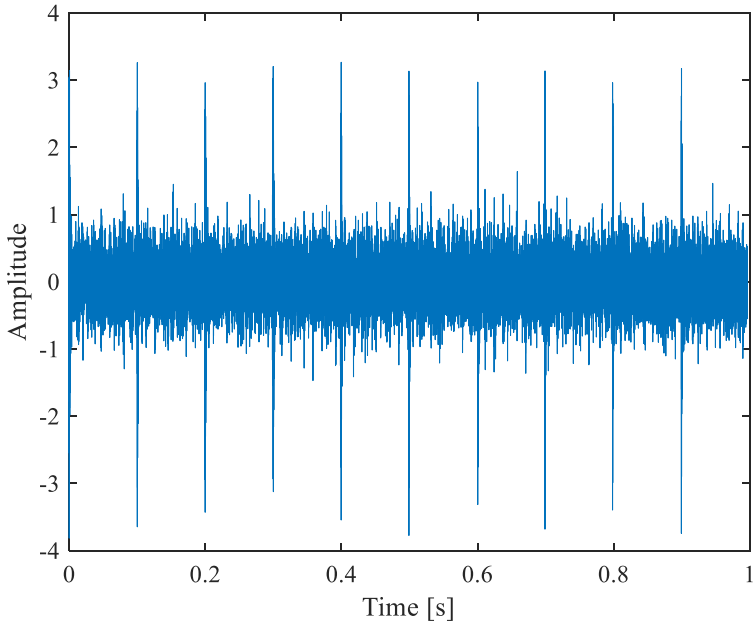


Figure 2.6. Simulated bearing fault signal with cyclostationsry interferences when *num* = 4

By increasing num from 1 to 100, the fault-irrelevant cyclostationary interferences become progressively dominant in the simulated signal, and Figure 2.7 is created to explore the influence of cyclostationary interferences to the FK, protrugram and FECgram. Here, it can be seen that when $num < 22$, the IFBs selected by the FK are around 9 kHz. When $num \geq 22$, the FK locates bands around 2 kHz, both of which do not contain the pre-set bearing natural frequency. The protrugram suggests IFBs at around 6 kHz when $num < 20$, but provides IFBs around 8 kHz which are consistent with the cyclostationary interferences' pre-set nature frequency once $num \geq 20$. As for the FECgram, IFBs around 6 kHz are again selected, locating the demodulation bands which are in line with the simulated bearing fault signal's natural frequency for all scales.

In summary, the FK is prone to be invalid when $num \geq 1$, which reveals that the FK is very sensitive to the cyclostationary interferences. For protrugram, it fails to locate the bands around the bearing natural frequency when $num \geq 20$, demonstrating that the protrugram is not sensitive to the cyclostationary to some extent, but its performance deteriorates with the increase of the irrelevant cyclostationary components. Nevertheless, the FECgram can clearly reflect the targeted fault energy and is not sensitive to cyclostationary interferences at all, thus allowing the FECgram to provide consistently reliable IFBs for bearing diagnosis for these simulated scenarios.

2.5. Experimental validation

The performance of the FECgram will be evaluated using practical bearing fault signals from three different test rigs in this section. Three challenging datasets,

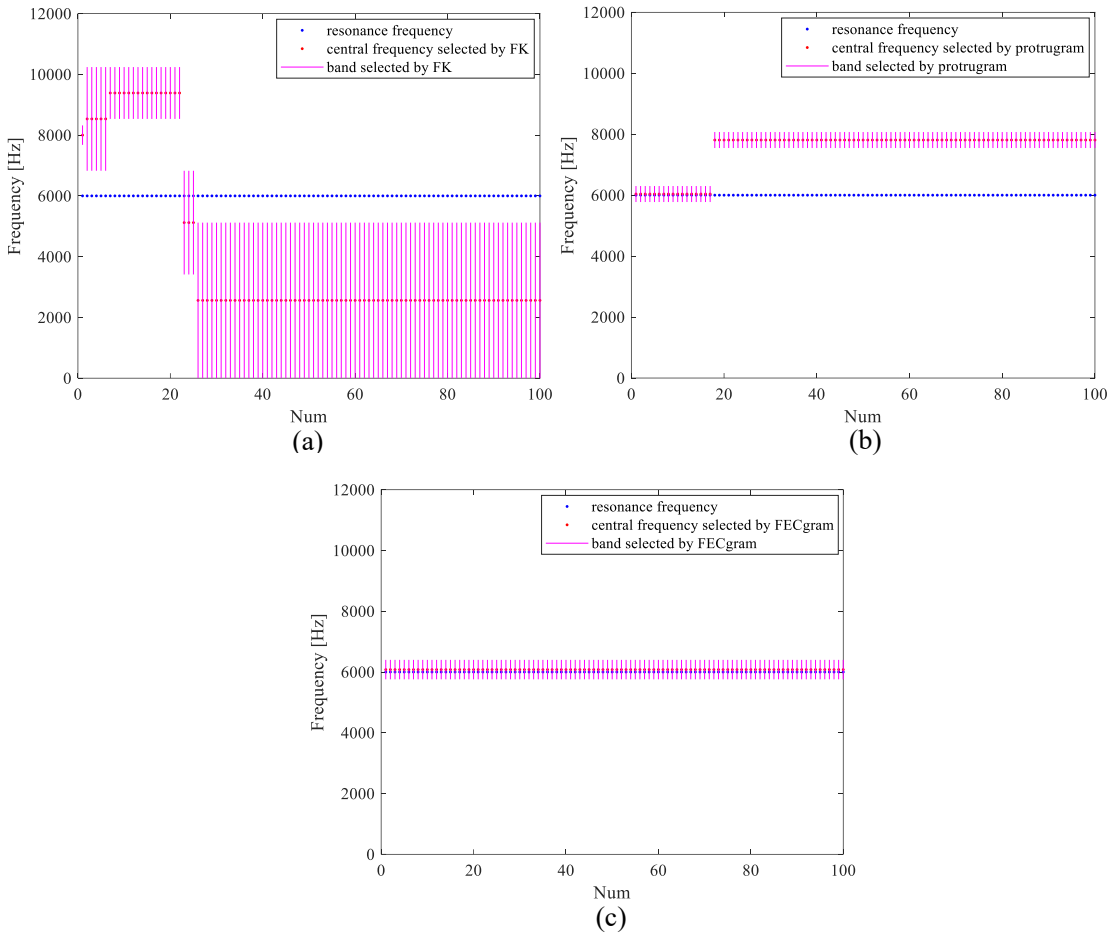


Figure 2.7. The selected IFBs under different levels of cyclostationary interferences using: (a) FK; (b) protrugram; (c) FECgram

including i) bearing fault with high-level manually generated fault-irrelevant impulsive interferences, ii) outer race faulty planet bearing driven by the planet carrier, and iii) inner race faulty planet bearing driven by the sun gear, are utilized to validate the superiority of the FECgram. Besides, the traditional FK and protrugram with a bandwidth of 3.5 times the FCF are employed for comparison.

2.5.1. Bearing fault with high-level impulsive interferences

The performance of FECgram under high-level impulsive interferences is validated through the Machinery Fault Simulator (MFS) Magnum at the University of Electronic Science and Technology of China (UESTC), China [131]. The layout of

the MFS is shown in Figure 2.8. It is driven by a motor, installed to the left-hand side of a healthy bearing. An MB ER-16K bearing seeded with an inner race fault by the SpectraQuest company is placed on the right-hand end. With nine rollers of diameter 7.94 mm and a pitch diameter of 38.5 mm, the test bearing provides a theoretical BPFI at 5.43 times the input shaft frequency. The load is provided by a 5 kg disc installed in the middle of the shaft. Several non-periodic impulses were generated by screwdriver manually to add non-periodic high-level impulsive interferences to the signal. The collected signal, measured at the sampling rate of 51.2 kHz, the shaft speed of 20 Hz, and synchronously with a phase reference on the input shaft, is shown in Figure 2.9 (a). The SES of the analyzed signal is displayed in Figure 2.9 (b) (the vertical red dashed lines denote the theoretical bearing FCF and its harmonics, and will be used throughout the following sections), demonstrating that it is hard to diagnose the bearing fault from the raw SES.

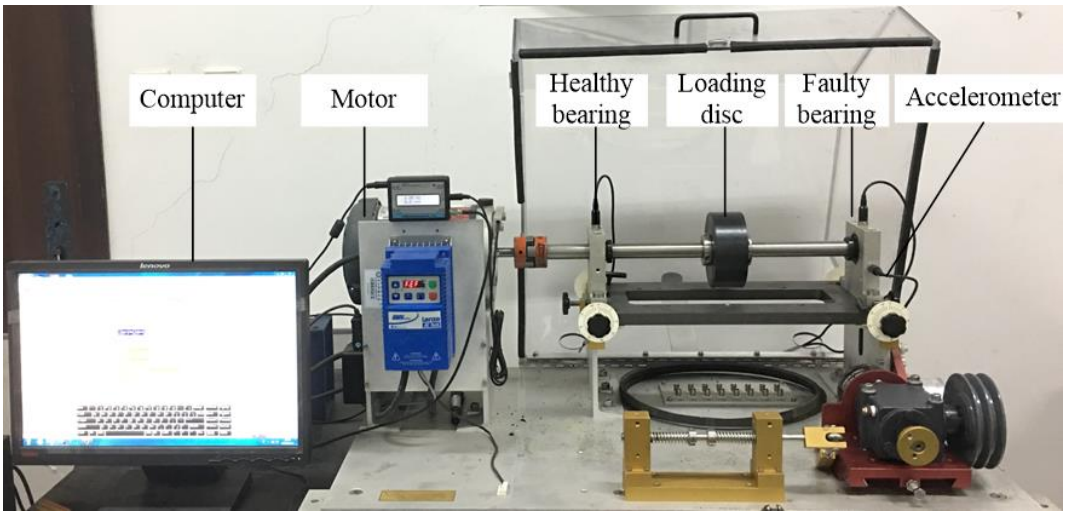


Figure 2.8. The layout of the UESTC MFS test rig [1]

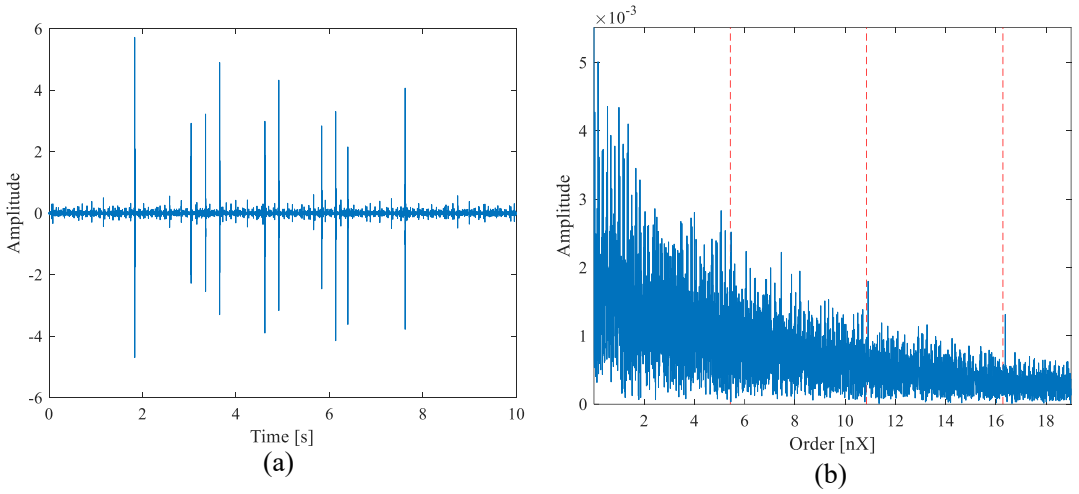


Figure 2.9. (a) Bearing signal with high-level impulsive interferences; and (b) its SES

The signal is order tracked at first, using the phase reference to transform the signal from time domain to angular domain, with which the effect of speed fluctuation can be eliminated. In all the squared envelope spectra shown below, the frequencies are presented in orders of the input shaft on the horizontal axes. Then the FK, protrugram and FECgram are applied for the selection of optimal IFBs. Subsequently, the IFBs with maximal selection criteria are employed to establish the squared envelope spectra respectively. The selected bands and their squared envelope spectra are displayed in Figure 2.10. Here it can be seen that the FK locates a demodulation band at (14.8 kHz, 800 Hz) (representing the center frequency of 14.8 kHz and bandwidth of 800 Hz), and no fault information can be extracted from the FK-based SES since no peaks at the BPFI and its harmonics appear. In contrast, both protrugram-based and FECgram-based squared envelope spectra provide reasonable diagnostic results, based on the demodulation bands at (761 Hz, 376 Hz) and (17.4 kHz, 400 Hz), respectively. Peaks at BPFI and its harmonics occur in both of these two spectra, also there are peaks at the orders of 3.43 and 7.43 in FECgram-based SES, representing the sidebands of the BPFI $(5.43 \pm 2 \times 1)$. It is worth mentioning that the peaks at the BPFI and its

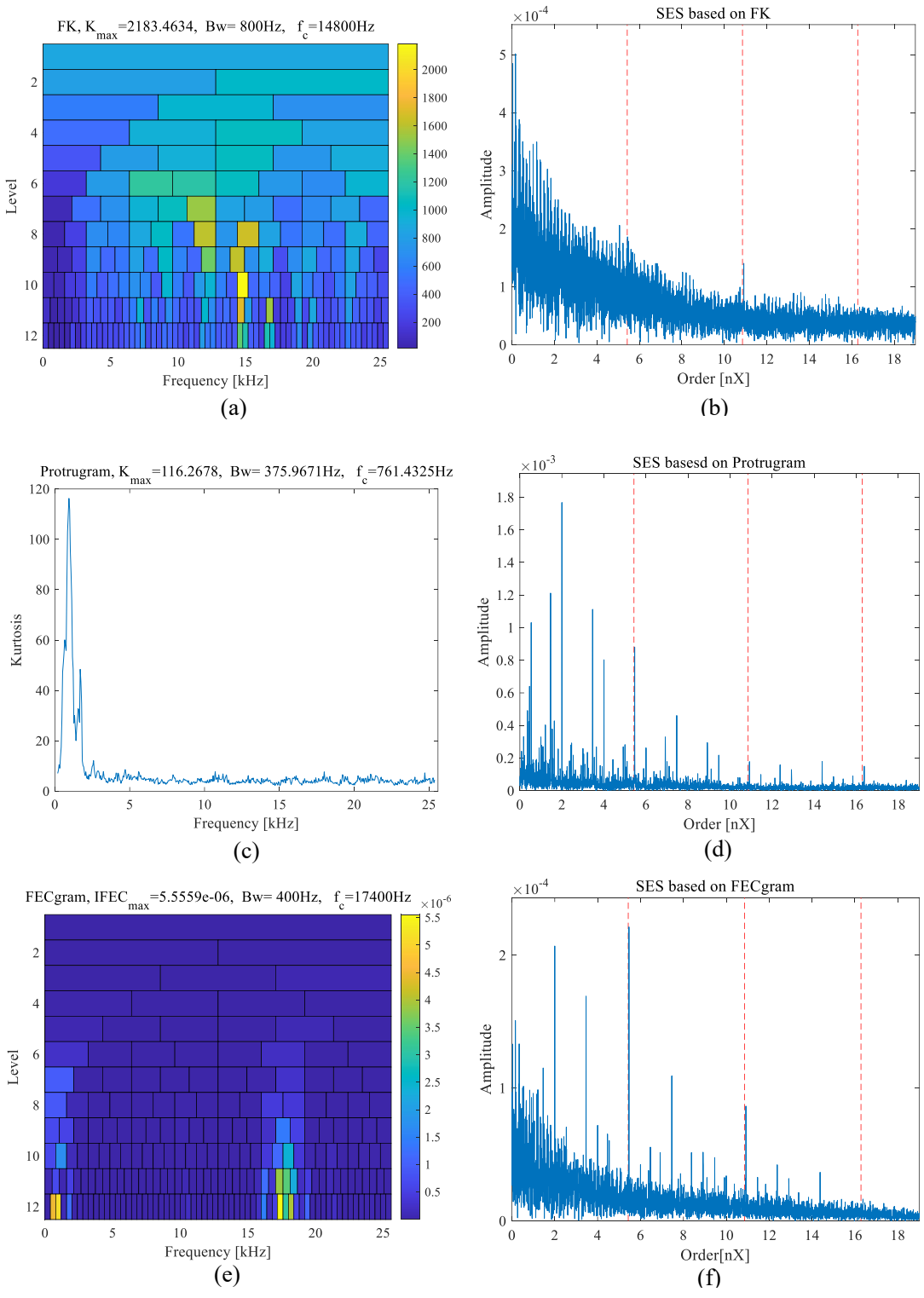


Figure 2.10. The IFB selection results under non-periodic high-level impulsive interferences using (a) FK; (c) protrugram; (e) FECgram; and the squared envelope spectra using (b) FK; (d) protrugram; (f) FEC gram

harmonics in the protrugram-based SES are not dominant and there are more obvious peaks at BPFI and its harmonics in the FECgram-based SES.

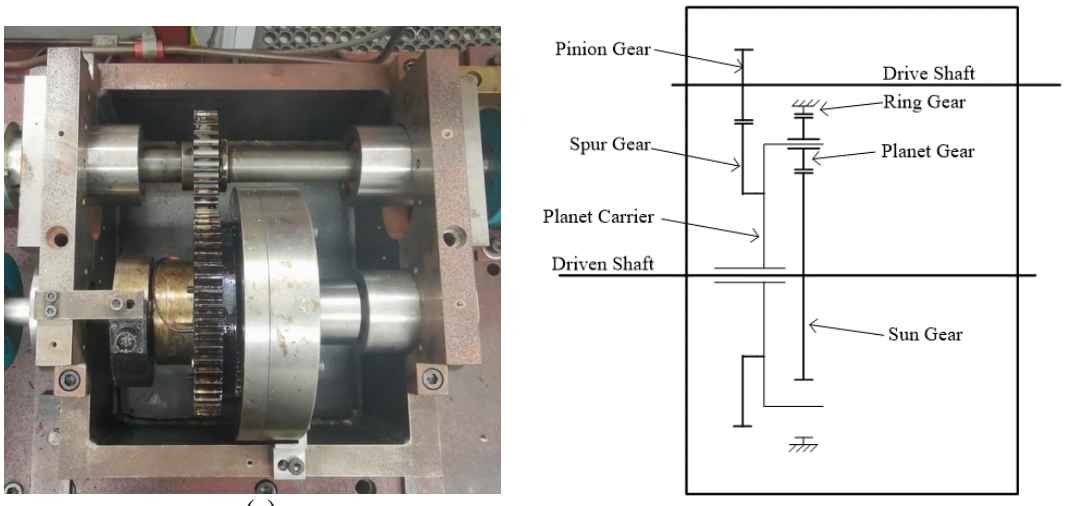
In this dataset, the presence of fault-irrelevant impulsive interferences brings great difficulty to the FK, misleading the diagnostic result of the faulty bearing. On the contrary, the protrugram and FECgram are less prone to be affected by the impulsive interferences, both of which diagnose the bearing fault correctly. The FECgram is likely to provide a better option, since the peaks at the BPFI and its harmonics are more dominant.

2.5.2. Outer race faulty planet bearing driven by the planet carrier

This section demonstrates another challenging experimental dataset collected from the planetary gearbox test rig at the University of New South Wales (UNSW), Australia [119]. The layout and schematic diagram of the rig are presented in Figure 2.11. As shown in Figure 2.11, the test rig consists of a parallel gear stage and a planetary gearbox stage, with the output shaft of parallel gear stage being connected to the planet carrier of the planetary gearbox stage. A planet bearing with a seeded outer race fault is installed in the planetary gearbox and specifications of this rig are shown in Table 2-2. Different from the ordinary fixed shaft bearing, the inner race of a planet bearing is fixed with the planet carrier, providing the inner race with a speed equal to the planet carrier's absolute speed. Simultaneously, the outer race is installed on the shaft hole of the planet gear, through which the outer race possesses the same speed with the planet gear's absolute speed. Together with the inner race and outer race, the rotating speed of the planet bearing should be the difference between the two speeds, equaling to the speed of planet gear relative to the planet carrier. Therefore, a more complicated calculation procedure of the FCF will be generated due to the planetary gearbox kinematics. The BPFO of the planet bearing in this planetary gearbox test rig can be calculated through

$$f_{BPFO} = \frac{n}{2}\left(1 - \frac{d}{D}\cos\theta\right) f_p^{(c)} = \frac{n}{2}\left(1 - \frac{d}{D}\cos\theta\right)\frac{Z_r}{Z_p}\frac{Z_{pg}}{Z_{sg}} f_i \qquad (2\text{-}16)$$

where the $f_p^{(c)}$ denotes the speed of planet gear relative to the planet carrier, n is the number of rollers, d is the roller diameter, D is the pitch diameter, θ is the contact angle, Z_r is the number of ring gear teeth, Z_p is the number of planet gear teeth, Z_{pg} is the number of pinion gear teeth, Z_{sg} is the number of spur gear teeth and f_i is the rotating frequency of the input shaft. From Equation (2-16), the BPFO of the planet bearing is 12.3 times f_i. And according to the kinematics of this planetary gearbox, there is a modulation at 2.66 times f_i (the frequency of planet gear relative to the planet carrier). Besides the unique structure, the planet bearings are surrounded by the gears and ring in the planetary gearbox as shown in Figure 2.12, thus the externally measured bearing signal is collected through complicated transfer paths and mixed with vibrations from other rotating components, which generally leads to great difficulties in the diagnosis of planet bearings [132].



(a) (b)

Figure 2.11. (a) The layout of the UNSW planetary test rig; and (b) the schematic diagram of the test rig

Table 2-2 The parameters of the UNSW planetary gearbox

Parallel gear stage		Planetary gear stage				Planet bearing			
Pinion gear teeth	Spur gear teeth	Number of planet gears	Planet gear teeth	Sun gear teeth	Ring gear teeth	Number of rollers	Contact angle (°)	Pitch diameter (mm)	Roller diameter (mm)
42	55	3	23	34	80	11	0	19	3

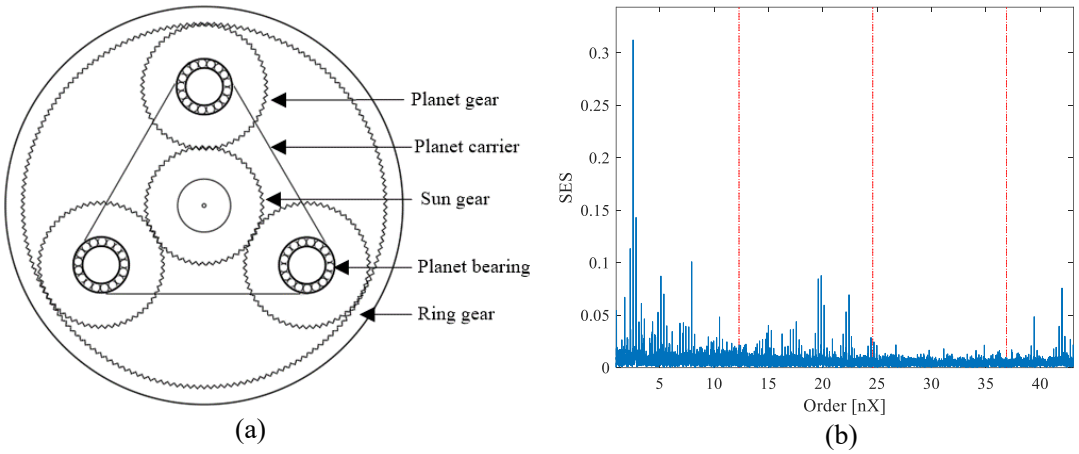


Figure 2.12. (a) The schematic diagram of the planetary gearbox; and (b) the SES of the outer race faulty planet bearing

To validate the effectiveness of the FECgram, a set of planet bearing data with sampling rate of 150 kHz and the input shaft speed of approximately 5.4 Hz are used. The signal to be analyzed is processed using the same procedure of Chapter 2.5.1, with the results being displayed in Figure 2.13. From Figure 2.13, it can be seen that the FK and protrugram locate IFBs at (26.367 kHz, 1.172 kHz) and (429 Hz, 233 Hz) respectively. The two selected IFBs, although far from each other, are neither suitable for the diagnosis of this bearing since no fault can be diagnosed according to the FK-based and protrugram-based squared envelope spectra. In contrast, the outer race fault of planet bearing can be verified through the FECgram-based SES as a result of the occurrence of peaks at the BPFO and its harmonics.

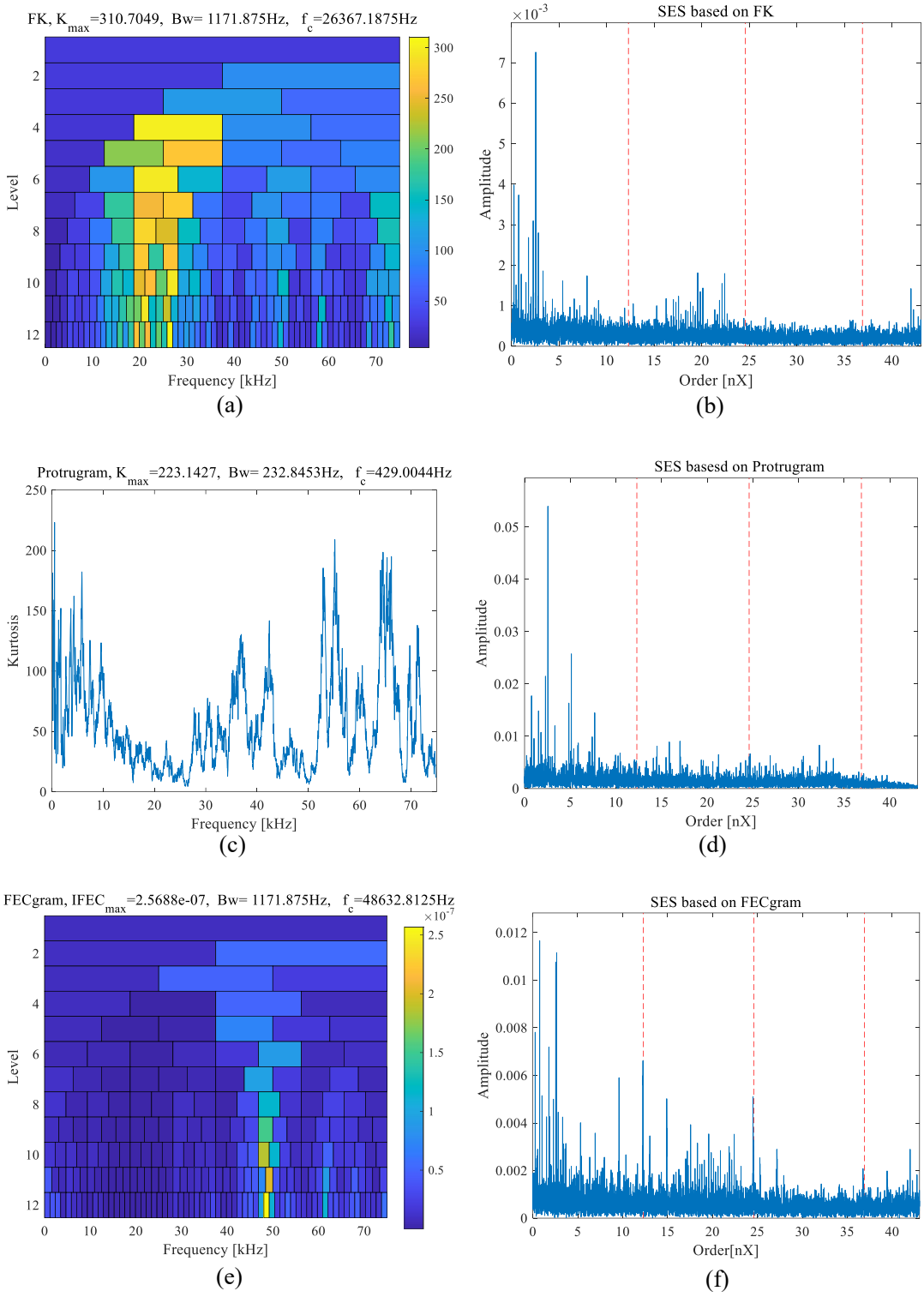


Figure 2.13. The IFB selection results of the outer race faulty planet bearing using (a) FK; (c) protrugram; (e) FECgram; and the squared envelope spectra using (b) FK; (d) protrugram; (f) FEC gram

2.5.3. Inner race faulty planet bearing driven by the sun gear

A challenging inner race faulty planet bearing dataset collected from the Drivetrain Diagnostics Simulator (DDS) of UESTC will be utilized here. The setup of the DDS test rig, which mainly consists of a two-stage planetary gearbox driven by the motor, one two-stage parallel shaft gearbox and a load, is shown in Figure 2.14 [133]. An inner race faulty planet bearing with the specifications listed in Table 2-3 is installed in the planetary gearbox for analysis. It should be noted that the location of the fault is on the inner race and the drive shaft of the planetary gearbox is the shaft of sun gear, which makes the measured signal different from the dataset in Chapter 2.5.2, hence this scenario is selected for the analysis in this section.

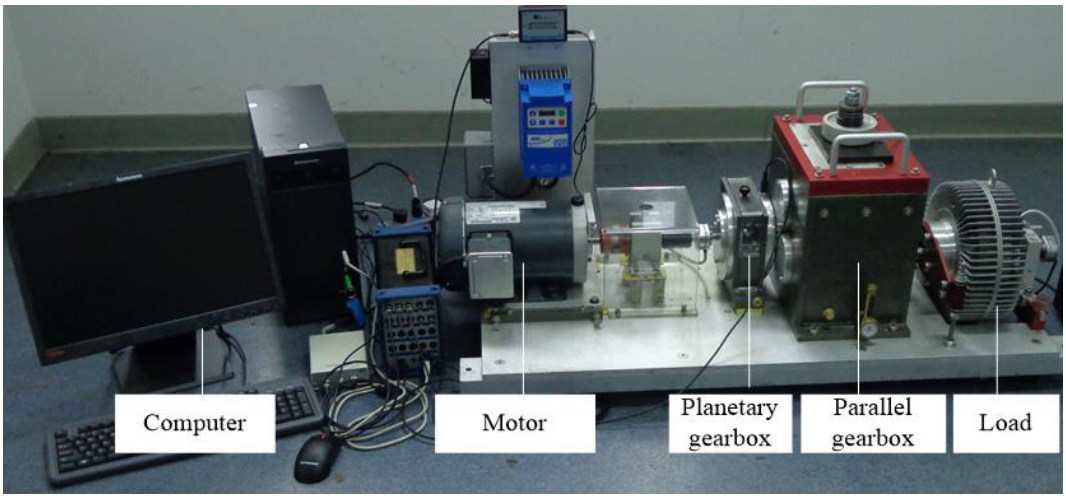


Figure 2.14. The layout of the UESTC DDS test rig

Table 2-3 The parameters of the UESTC DDS test rig

Planetary gear				Planet bearing			
Number of planet gears	Planet gear teeth	Sun gear teeth	Ring gear teeth	Number of rollers	Contact angle (°)	Pitch diameter (mm)	Roller diameter (mm)
4	36	28	100	10	0	14.5	2.4

A set of data collected at the sampling rate of 25.6 kHz and the input shaft speed of approximately 30 Hz is provided for analysis. Unlike the UNSW planet bearing

dataset, the gearbox is driven by the sun gear with no parallel gear stage installed, which results in a different formula to calculate the speed of planet gear relative to the planet carrier, and allowing the computation of the BPFI through

$$f_{BPFI} = \frac{n}{2}\left(1 + \frac{d}{D}\cos\theta\right) f_p^{(c)} = \frac{n}{2}\left(1 + \frac{d}{D}\cos\theta\right)\frac{(Z_r - Z_p)Z_s}{(Z_s + Z_r)Z_p} f_i \qquad (2\text{-}17)$$

where Z_s is the number of sun gear teeth. According to Equation (2-17), the BPFI of the UESTC planet bearing is 2.266 times f_i, with modulation at 0.389 times f_i. The SES of the raw signal is displayed in Figure 2.15, showing that no fault feature of the planet bearing can be extracted from this complicated signal. In addition, strong peaks at the order of 0.875 and 6.251 which represent the ring gear fault characteristic order and 2 times the sun gear fault characteristic order occur in Figure 2.15, indicating the possibility of the ring gear and sun gear faults in this planetary gearbox.

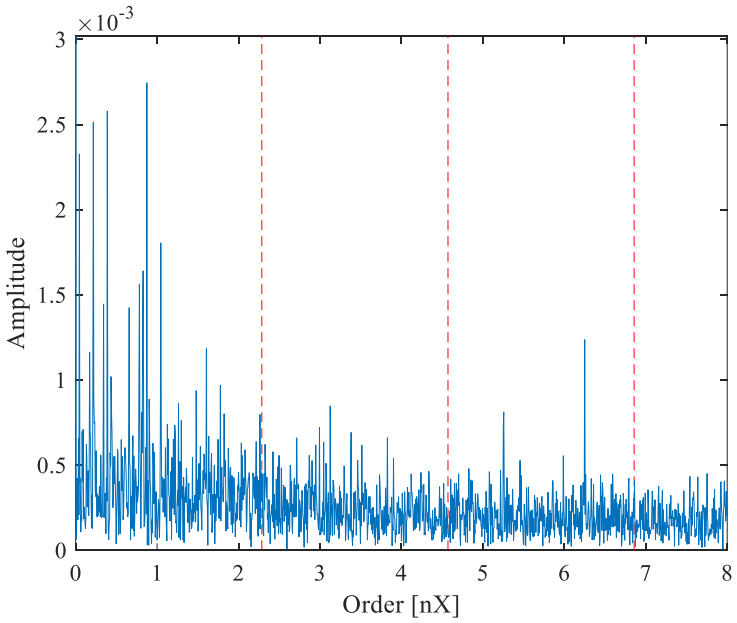


Figure 2.15. The SES of the inner race faulty planet bearing

A similar analysis procedure as presented in Chapter 2.5.1 is then applied with the results being shown in Figure 2.16. It can be seen that the FK-based SES fails to

diagnose the bearing fault, selecting a band where the BPFI and its harmonics are not included. The protrugram also loses its effectiveness in choosing a suitable band, since no peaks are found at the BPFI and its harmonics in the protrugram-based SES. A strong peak at the order of 0.218 occurs in this SES, revealing the existence of carrier rotating components. On the contrary, the FECgram can locate an optimal IFB for demodulation because the peaks at the BPFI and its harmonics are dominant in the FECgram-based SES.

2.5.4. Summary of the experimental results

The results obtained from the three challenging experimental datasets in bearing fault diagnosis can be briefly summarized as follows.

1) The FK fails to diagnose the bearing fault in all the three datasets, thus the FK is very sensitive to the impulsive and cyclostationary interferences.

3) The protrugram behaves poorly as it only diagnoses the bearing fault successfully in the first experimental validation. However, the peaks at the FCF and its harmonics are not as dominant as those in the FECgram-based SES in this experimental validation, indicating that the protrugram is not sensitive to the impulsive interference to a great extent but is still affected by this interference slightly. Also, the protrugram is heavily affected by the cyclostationary interference.

3) The FECgram performs consistently well and robustly in all cases, providing accurate diagnostic outcomes as a result of being not affected by the fault-irrelevant impulsive and cyclostationary interferences. The FECgram is, consequently, not sensitive to the impulsive and cyclostationary interferences.

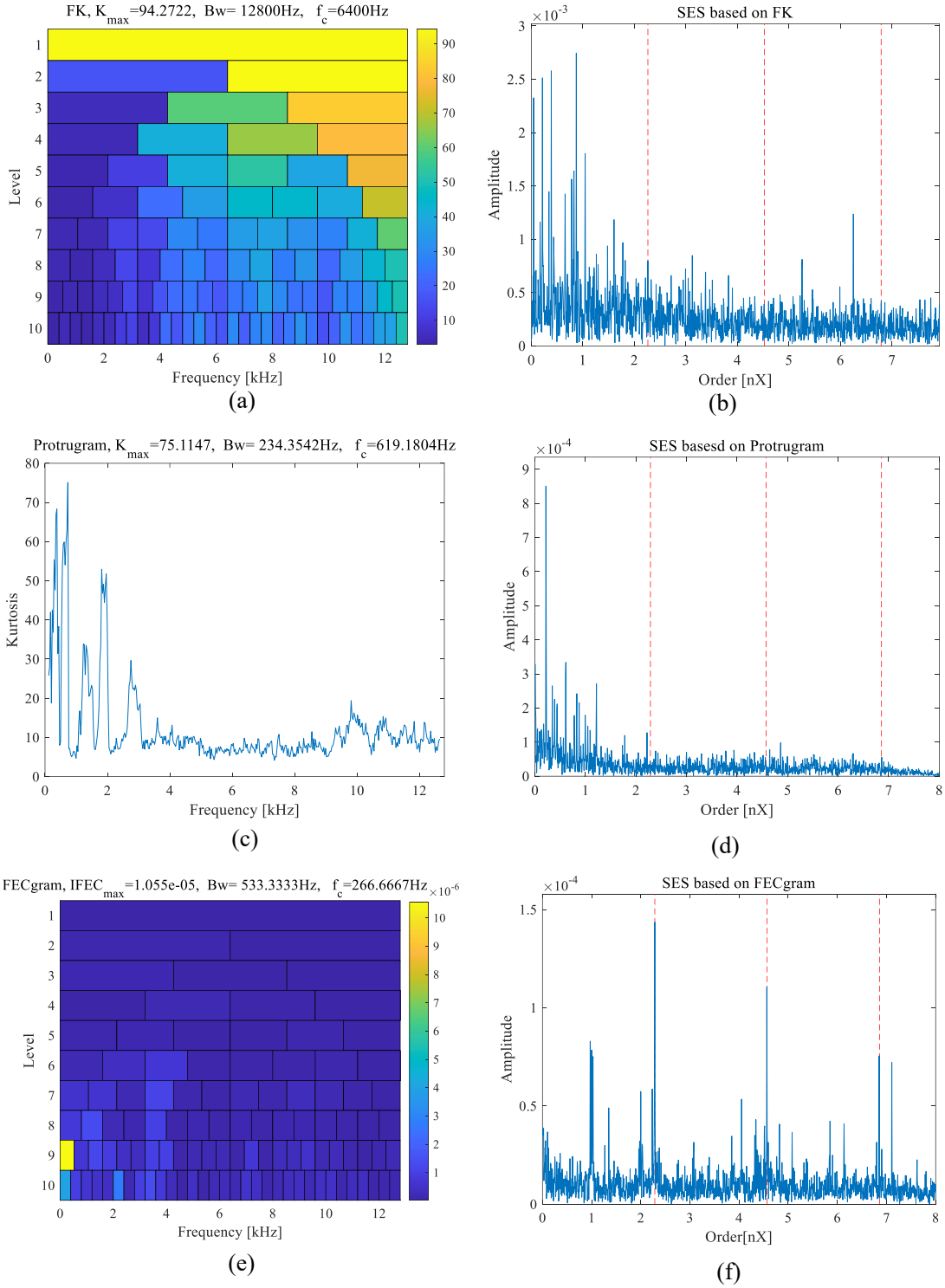


Figure 2.16. The IFB selection results of the inner race faulty planet bearing using (a) FK; (c) protrugram; (e) FECgram; and the squared envelope spectra using (b) FK; (d) protrugram; (f) FEC gram

2.6. Conclusion

The novelty and contribution of this chapter can be summarised as follows.

1) A novel indicator named FEC was proposed. By calculating the energy at the targeted FCFs based on correntropy, the FEC can solely represent the bearing fault feature and is not affected by the impulsive interference.

2) A novel method named FECgram was proposed to overcome the drawbacks of the traditional band selection methods. By taking advantage of the correntropy and the prior knowledge of the FCFs, the FECgram can select an IFB where the energy at targeted FCFs is strongest. It is, therefore, a robust method which is less affected by fault-irrelevant impulsive and cyclostationary interferences.

3) Successful diagnosis of three challenging faulty bearing datasets has been achieved through the proposed FECgram, demonstrating that the FECgram would be an effective approach for bearing diagnosis under most complicated operational occasions.

Chapter 3. A novel fault information-guided variational mode decomposition method for bearing diagnostics

3.1. Introduction

This chapter is a modified version of the journal paper titled “A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis”, which was published in Mechanical Systems and Signal Processing (164 (2022): 108216). The work completed in this chapter is dedicated to realizing objective 2 of this research.

Based on the literature review of Chapter 1.2.1, the VMD is an advanced AMD technique. However, its success depends highly on the pre-set mode number and bandwidth control parameter. Some improved-VMD methods target at addressing this issue, and the indicators used in these improved-VMD techniques can be roughly divided into three categories: impulsiveness, cyclostationarity, and disorder. These indicators can reflect the fault symptom to some extent as the faulty bearings can generate a series of repetitive transients, thus resulting in the increase of impulsiveness, cyclostationarity and disorder. However, these indicators are not solely sensitive to the bearing fault symptom. Specifically, the vibration interferences from other adjacent components and backgrounds can also increase these indicators, thus bringing challenges to the VMD and its improved method.

To address the above-mentioned issues, a novel bearing fault information-guided VMD method is proposed in this chapter to extract the weak bearing repetitive

transients. Two nested statistical models, namely the generalized Gaussian cyclostationary (GGCS) model and the generalized Gaussian stationary (GGS) model [120], are introduced to characterize the bearing fault vibrations. Then statistical indicator $I_{GGCS/GGS}$ and its statistical threshold are calculated using maximum likelihood estimation and the critical value of Chi-square distribution, through which several possible mode numbers are roughly selected. The $I_{GGCS/GGS}$ can reflect the cyclostationarity of a specific cycle under non-Gaussian scenarios. Therefore, by taking the advantage of $I_{GGCS/GGS}$ and bearing fault cycle, bearing fault cyclostationarity can be solely represented without being affected by fault-unrelated cyclostationary and impulsive noise. In addition, different from the manually setting thresholds used in many literatures [81, 83, 88], the statistical threshold based on the GGCS and GGS models is utilized in the proposed method to identify whether the bearing fault information is included without lack of mathematical basis. After that, the RFCA that can evaluate and quantify the fault information in the spectrum is defined and employed to optimize the bandwidth control parameter in the selected mode numbers. The BLIMF with the maximum bearing fault information together with Hilbert envelope spectrum (HES) are used for bearing diagnosis ultimately. The FIVMD is compared with the EMD, LMD and original VMD (the original VMD method proposed in Ref. [64] with mode number setting to be 5 and bandwidth control parameter setting to be 1000 in this chapter, since 5 and 1000 are the default values when using VMD method in Matlab [134]) in simulation and experimental datasets. Better performance of the proposed method in the simulation analysis and challenging experimental datasets, including early bearing fault signals in run-to-failure test datasets, signals with impulsive noise and planet bearing signals, demonstrates that the proposed FIVMD method can solve the parameter selecting problem of the

traditional VMD method and extract weak bearing repetitive transients under most complicated operational environments.

3.2. A brief review of the VMD method

The VMD is a recently proposed methodology that can adaptively decompose an original signal into a series of quasi-orthogonal BLIMFs [64]. Different from the IMF defined in EMD, the BLIMF in VMD is regarded as an amplitude-modulated-frequency-modulated (AM-FM) signal whose energy is compacted around its central frequency and can be expressed as

$$BLIMF_k(t) = A_k(t) \times \cos\left(\varphi_k(t)\right) \tag{3-1}$$

where k represents the k^{th} BLIMF, $A_k(t)$ and $\varphi_k(t)$ denote the instantaneous amplitude and phase, respectively.

To obtain the targeted BLIMFs, the decomposition procedure in VMD is transferred into a variational framework, by searching the optimal solution of the following constrained variational model using a non-iteratively sifting structure. Specifically, the decomposition procedure of VMD is to solve the following constrained variational problem

$$\begin{aligned} &\min_{\{BLIMF_k\},\{\omega_k\}} \left\{ \sum_K \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) * BLIMF_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ &s.t. \qquad \sum_K BLIMF_k\,(t) = S(t) \end{aligned} \tag{3-2}$$

where K is the number of BLIMFs (mode number) that should be set before decomposition, ω_k is the center frequency of k^{th} BLIMF, and the $S(t)$ is the signal to

be analyzed. The term $\left\|\partial_t\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*BLIMF_k(t)\right]e^{-j\omega_k t}\right\|_2^2$ represents the estimated bandwidth of k^{th} BLIMF. Therefore, this constrained variational problem can be described as determining K BLIMFs that can minimize the sum of all BLIMFs' bandwidths.

More specifically, the logic of the VMD decomposition procedure is:

1) Determine the mode number K and bandwidth control parameter α (also known as the balancing parameter). Then initialize the $BLIMF_k(t)$.

2) Calculate the analytic signal of $BLIMF_k(t)$ using Hilbert transform, which is equal to $\left(\sigma(t)+\frac{j}{\pi t}\right)*BLIMF_k(t)$.

3) Add $e^{-j\omega_k t}$ to the analytic signal $\left(\sigma(t)+\frac{j}{\pi t}\right)*BLIMF_k(t)$ so that the center frequency of the analytic signal is shifted to ω_k. Calculate the estimated bandwidth of the $BLIMF_k(t)$ through $\left\|\partial_t\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*BLIMF_k(t)\right]e^{-j\omega_k t}\right\|_2^2$.

4) Obtain the optimal selection of $BLIMF_k$ through solving the constrained variational problem given in Equation (3-2). To address this issue, the augmented Lagrange function is introduced and the new constrained problem becomes Equation (3-3).

$$L(\{BLIMF_k\},\{\omega_k\},\lambda)=\alpha\sum_K\left\|\partial_t\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*BLIMF_k(t)\right]e^{-j\omega_k t}\right\|_2^2 + \left\|S(t)-\sum_K BLIMF_k(t)\right\|_2^2+\langle\lambda(t),S(t)-\sum_K BLIMF_k(t)\rangle \tag{3-3}$$

where $\lambda(t)$ represents the Lagrangian multiplier and the term $\langle \lambda(t), S(t) - \sum_K BLIMF_k(t) \rangle$ can be rewritten as $\lambda(t)[S(t) - \sum_K BLIMF_k(t)]$.

5) Solve Equation (3-3) using the alternate direction method of multipliers (ADMM). The decomposed modes and center frequencies are transferred into the Fourier domain in the ADMM algorithm, which can be expressed as

$$\widehat{BLIMF}_k^{n+1}(\omega) \leftarrow \frac{\hat{S}(\omega) - \sum_{i<k} \widehat{BLIMF}_i^{n+1}(\omega) - \sum_{i>k} \widehat{BLIMF}_i^{n}(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2}, \quad (3\text{-}4)$$

and

$$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega \left| \widehat{BLIMF}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \widehat{BLIMF}_k^{n+1}(\omega) \right|^2 d\omega} \quad (3\text{-}5)$$

respectively. The updating of modes and center frequencies will continue until Equation (3-6) is satisfied

$$\frac{\sum_K \left\| \widehat{BLIMF}_k^{n+1} - \widehat{BLIMF}_k^{n} \right\|_2^2}{\left\| \widehat{BLIMF}_k^{n} \right\|_2^2} < \varepsilon \quad (3\text{-}6)$$

More details can be referred to Ref. [64].

3.3. The proposed FIVMD method

The superiority of VMD has been demonstrated when proper mode number and bandwidth control parameter are inputted [48]. Therefore, the key point in using VMD to extract the bearing fault signature is the determination of appropriate decomposition

parameters, i.e., mode number and bandwidth control parameter. If the mode number is under dimension, the bearing fault signature will still be entangled with other frequency components, while if it is over-dimension, the fault signature will be divided into different BLIMFs. Furthermore, even though an appropriate mode number is identified, a smaller bandwidth control parameter can result in a wider bandwidth of the BLIMF, leading to the appearance of large background noise, while a larger bandwidth control parameter can exclude some bearing fault information. Most of the existing methods improve the VMD based on kurtosis, kurtosis of the envelope spectrum, entropy, or their derivatives. These indicators, on the one hand, can reflect the fault information to some extent as the repetitive transients could increase impulsiveness, cyclostationarity and disorder of the signal. However, they are not solely sensitive to the bearing fault signature, which means that the abnormal impulses and vibration interferences from other adjacent components can also lead to the increase of these indicators. Therefore, these kinds of methods are prone to losing their effectiveness under the interferences from other components and external abnormal impulses. To address this issue, this chapter aims to establish a bearing fault information-guided VMD method that can solely extract the weak bearing repetitive transients under strong noise, thereby promoting the feasibility to diagnose the bearing fault in complicated operational environments. The proposed FIVMD method, based on the GGCS, GGS models and the RFCA, is presented in detail below.

3.3.1. Two statistical models for bearing vibration characterization

The impacts between bearing faulty and healthy parts can generate a series of repetitive transients, making the vibration of faulty bearing to be both impulsive and cyclostationary. Therefore, impulsiveness and cyclostationarity are two essential

properties to characterize the bearing fault signals. There are more studies dealing with impulsiveness than cyclostationarity due to its easy-understanding statistical foundation. Generalized Gaussian (GG) distribution is an appropriate approach to describe the impulsiveness of a signal, thus can be used to model the impulsiveness of a bearing fault signal [131]. The GG model, also called GGS model as a result of the hidden assumption of stationarity, can be expressed as

$$P_x[x(n);\beta_0,\eta_0] = \frac{\beta_0}{2\Gamma(1/\beta_0)\eta_0} e^{-\left(\frac{|x(n)|}{\eta_0}\right)^{\beta_0}} \quad (3\text{-}7)$$

where P denotes the probability density distribution, $\Gamma(.)$ represents the Gamma function, β_0 is the shape parameter and a smaller β_0 represents the increase of impulsiveness of the distribution, η_0 is the scale parameter, governing the variance of the signal.

The GGS model, although well characterizes the impulsiveness of the signal, neglects the cyclostationarity of bearing fault symptom, which renders the model to be not exactly identical to the fault signal. Therefore, Antoni and Borghesani [120] proposed the GGCS model by integrating impulsiveness and cyclostationarity into one model synchronously. The probability density function of the GGCS model is

$$p_x[x(n+k\times N);\beta_1,\eta_1(n)] = \frac{\beta_1}{2\Gamma(1/\beta_1)\eta_1(n)} e^{-\left(\frac{|x(n+k\times N)|}{\eta_1(n)}\right)^{\beta_1}} \quad (3\text{-}8)$$

where N is the sample number of each cyclic period, $n = 0, 1, \cdots, N-1$, $k = 0, 1, \cdots, K-1$ with K being the cycle number in the signal. The GGCS model takes the advantage of cyclostationarity and enables the distribution of the whole signal to

vary over time, but the corresponding points in each cyclic period follow the same GG distribution.

For a better interpretation of the two models, numerically generated signals with different shape parameters and scale parameters that follow the GGS and GGCS distribution are displayed in Figure 3.1. The shape parameters are set to be 2, 1, 0.5 respectively, the scale parameters are set to be 1 in the GGS model and $[1 + \sin(2\pi n/N)]/2$ in the GGCS model, the cyclic period is set to be 1024 and 8 cycles are used.

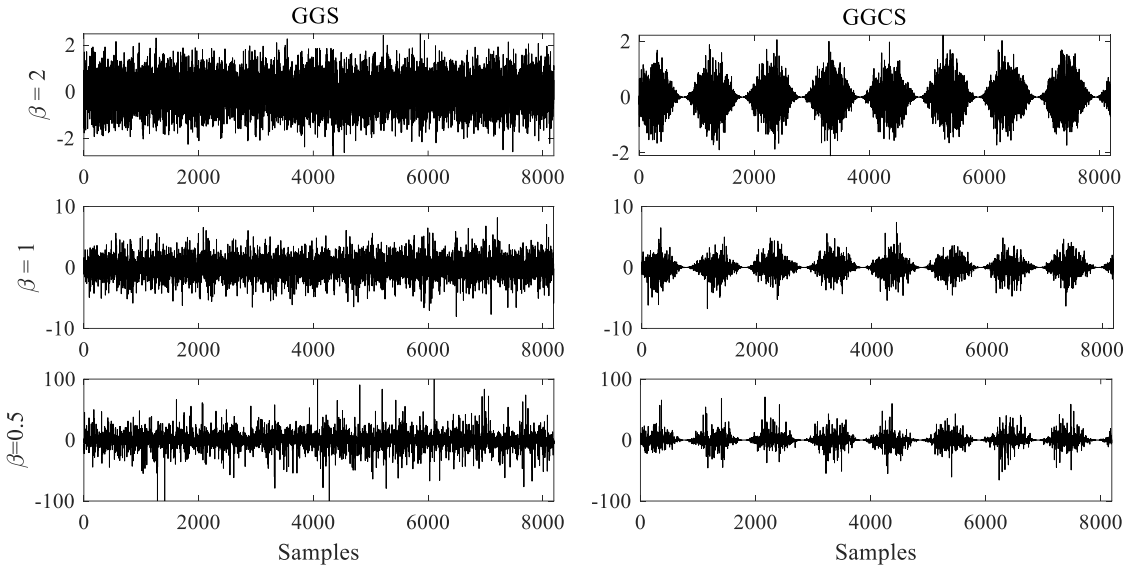


Figure 3.1. Simulated signals with different shape parameters and scale parameters following GGS and GGCS distribution

3.3.2. Criteria for determining VMD parameters

Two criteria based on three indicators, namely the $I_{GGCS/GGS}$, the statistical threshold of $I_{GGCS/GGS}$, and the RFCA, will be used in the FIVMD method. Detailed discussions are presented in the following subsections.

3.3.2.1. The log-likelihood ratio $I_{\mathrm{GGCS/GGS}}$

For the null hypothesis H_0 and the alternative hypothesis H_1, their log-likelihood ratio (written as I_{H_1/H_0}) estimates the extra information between the two hypotheses when the signal follows the alternative hypothesis H_1. Therefore, $I_{GGCS/GGS}$ represents the extra information between the GGCS and GGS models, measuring the cyclostationarity under non-Gausian hypothesis [120]. The definition of $I_{GGCS/GGS}$ is

$$I_{GGCS/GGS} = 2\frac{L_1 - L_0}{l} \tag{3-9}$$

where l is the length of the signal, $L_1$and L_0 are log-likelihood functions of the GGCS model and GGS model, which can be expressed as [120]

$$\begin{aligned} L_1 = l\ln\beta_1 - l\ln 2 - l\ln\Gamma\left(\frac{1}{\beta_1}\right) - K\sum_{n=0}^{N-1}\ln[\eta_1(n)] \\ -\sum_{n=0}^{N-1}\eta_1(n)^{-\beta_1}\sum_{k=0}^{K-1}|x(n+kN)|^{\beta_1} \end{aligned} \tag{3-10}$$

and

$$L_0 = l\ln\beta_0 - l\ln 2 - l\ln\Gamma\left(\frac{1}{\beta_0}\right) - l\ln\eta_0 - \sum_{n=0}^{l-1}|x(n)|^{\beta_0}{\eta_0}^{-\beta_0} \tag{3-11}$$

respectively. By taking the partial derivative of Equation (3-10) with respect to β_1 and $\eta_1(n)$ respectively, and searching the zero points of the resultant equations, β_1 and $\eta_1(n)$ can be obtained through the maximum likelihood estimation. Similarly, β_0 and

$\eta_0(n)$ can be calculated from Equation (3-11) using the maximum likelihood estimation.

For bearing fault diagnosis, if the cyclic cycle in GGCS model is set to be the bearing fault period, the $I_{GGCS/GGS}$ will possess the following merits:

1) It solely reflects the cyclostationarity of the given faulty bearing and is not affected by other frequency contents.

2) It is insensitive to impulsive noise since GG assumptions are contained in both GGCS model and GGS model, conveying no extra information of non-Gaussianity.

3.3.2.2. The statistical threshold of $I_{\mathrm{GGCS/GGS}}$

The $I_{GGCS/GGS}$ will be used to reveal the cyclostationarity of bearing fault in each decomposed mode in this chapter. In these decomposed modes, it is impossible to predict whether the bearing fault signature is included or not in advance. In other words, the GGCS assumption may be invalid in these modes. Therefore, it is important to set a threshold that can be used to decide whether the assumption is acceptable or not. Wilk's theorem [135] points out that two times the generalized likelihood ratio of two nested hypotheses follows a Chi-square distribution whose degree of freedom (DOF) is the dimensional difference between the two hypotheses, namely

$$2(L_1 - L_0) \sim \chi_d^2 \tag{3-12}$$

where d is the dimensional difference between the two hypotheses. By combining Equation (3-9) and Equation (3-12), it can be found that the $I_{GGCS/GGS}$ follows a Chi-square distribution with $(N-1)$ DOFs [120]:

$$I_{GGCS/GGS} \sim \frac{1}{l}\chi^2_{N-1} \tag{3-13}$$

Because the GGCS model has $(N+1)$ DOFs (β_1 , $\eta_1(n)$) and GGS model has 2 DOFs (β_0 , η_0). Therefore, the threshold of $I_{GGCS/GGS}$ in this chapter is used as the critical value of the $(N-1)$ DOFs' Chi-square distribution at a given significance level, which is usually set as 0.05 in hypothesis testing.

The possible mode number K_p is initially selected using $I_{GGCS/GGS}$ and its threshold at first, considering that the BLIMFs' property is strongly related to the mode number when bandwidth control parameters are located in the interval of [1000 4000] [85]. If the mode number K_p satisfies the following conditions: (1) there is at least one $I_{GGCS/GGS}$ exceeding the threshold using mode number K_p, and (2) the maximum $I_{GGCS/GGS}$ value using mode number K_p is bigger than those using $K_p - 1$ and $K_p + 1$, then the K_p will be selected as one possible option to access the optimal mode number K. The satisfaction of condition (1) guarantees the validation of GGCS hypothesis, indicating the bearing fault information can be obtained using mode number K_p. In addition, condition (2) ensures that the mode with the maximum $I_{GGCS/GGS}$ value using mode number K_p is the highest among adjacent mode numbers.

3.3.2.3. The RFCA

The amplitudes at the FCF and its harmonics are appropriate measures to quantify the fault information expressed by the spectrum. Motivated by the ratio of cyclic content proposed in Ref. [136], the RFCA is defined by Equation (3-14) in this chapter and used to determine the optimal bandwidth control parameter in the possible mode numbers.

$$RFCA = \frac{\sum_{i=1}^{I} \max_{n \in F_i} HES(n)}{\sum_{n=0}^{n=round((I+0.5)*FCF/df)} HES(n)} \tag{3-14}$$

where i is the i^{th} harmonic, I denotes the total harmonics to be analyzed and is set to be 3 in this chapter, df is the frequency resolution. F_1 represent a series of frequencies in the interval of $[\text{FCF} - 5 \times df, \text{FCF} + 5 \times df]$ in case of the deviations between theoretical FCF and actual FCF caused by the slippage of bearing [1], and F_i are similar frequencies of h^{th} harmonic. $HES(n)$ is the Hilbert envelope spectrum and is calculated through

$$HES(n) = \left|\text{FFT}\left|\text{Hilbert}\left(x(n)\right)\right|\right| \tag{3-15}$$

The higher value of RFCA indicates more bearing diagnostic information to be contained in the spectrum. For illustration purpose, the Hilbert envelope spectra of simulated constant Gaussian background noise with increasing bearing fault energies are plotted in Figure 3.2, and then the RFCAs are calculated accordingly. The simulated signal is generated by

$$x_b(t) = \sum_i O(t - iT - \tau_i) + n(t) \tag{3-16}$$

where i represents the i^{th} transient induced by the impact of faulty and healthy parts, T is the theoretical cyclic period (equals to 1/FCF), and τ_i denotes the deviation between theoretical and practical cyclic periods of the i^{th} transient, $n(t)$ is Gaussian white noise with a mean at 0 and a standard-deviation at 0.2, and $O(t)$ represents the oscillating attenuation function which can be written as

$$O(t) = C_0 e^{-2\pi\zeta f_n t} \sin\left(2\pi\sqrt{1-\zeta^2} f_n t + \varphi_0\right) \tag{3-17}$$

where C_0 is the initial displacement (by changing it the bearing fault signals with different energies can be obtained), ζ is the damping ratio, f_n is the natural frequency and φ_0 is the initial phase. The parameters of the simulated signal and corresponding calculated RFCAs are listed in Table 3-1.

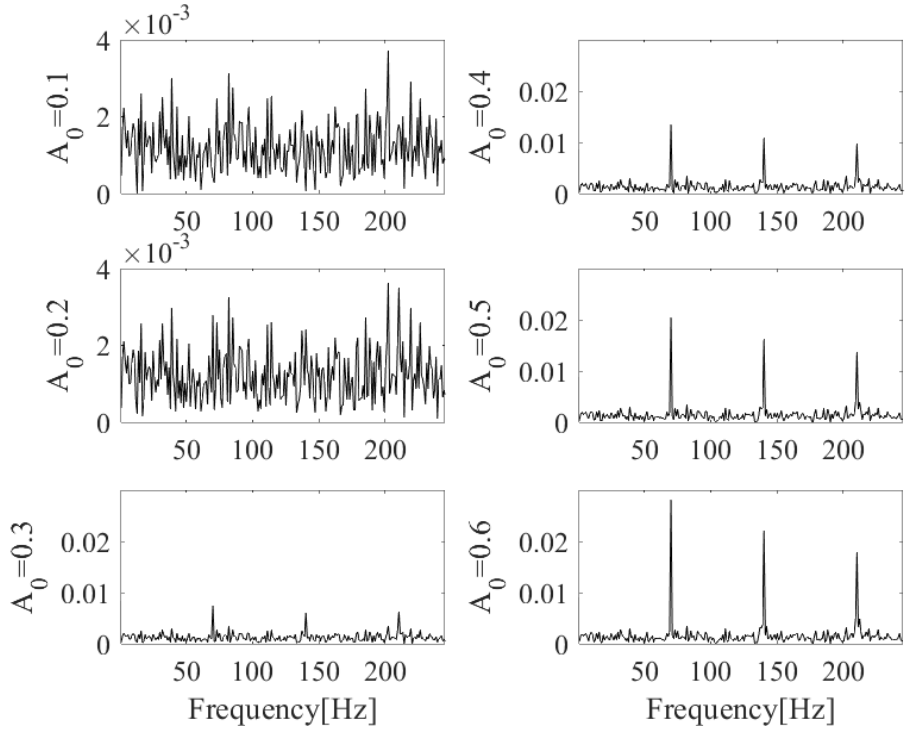


Figure 3.2. Hilbert envelope spectra of constant Gaussian background noise with increasing bearing fault energies

Table 3-1 Parameters of the simulated bearing signal

f_n	φ_0	FCF	ζ	τ_k	C_0	Corresponding RFCAs
4000	$\pi/2$	70	0.02	$\sim N(0, 0.01T)$ N represents Gaussian distribution	0.1, 0.2, 0.3, 0.4, 0.5, 0.6	0.0154, 0.0160, 0.0346, 0.0565, 0.0790, 0.1009

It can be concluded from Table 3-1 that the RFCAs increase with the growth of bearing fault energies. Accordingly, the RFCA is a rational indicator to measure the

bearing fault information. Another interesting attempt to validate the effectiveness of RFCA is to assume that the simulated signals can be characterized as the Laplacian distributions [137], then the scale parameters are able to auxiliary validate the performance of RFCA. Maximum likelihood estimation is used to estimate the location parameters and scale parameters of the simulated signals. All the location parameters are estimated to be 0 and the estimated scale parameters are 0.0089, 0.0177, 0.0266, 0.0354, 0.0443, 0.0533, respectively. The trend (monotonically increasing) of these estimated scale parameters is in line with the trend of the calculated RFCAs. For the Laplacian distribution, a higher scale parameter represents a higher variance and a fatter tail. In general, a heavier tail implies higher kurtosis than normal condition, indicating the increasing bearing fault information. Therefore, this trend demonstrates that the RFCA is an appropriate indicator to describe the fault information contained in the spectrum in another aspect. Thence, the RFCA will be used to finally determine the optimal parameters. The details are shown in the following section.

3.3.3. The FIVMD method

In this section, a novel fault information-guided VMD method is proposed to overcome the disadvantages of original VMD methods, in which the $I_{GGCS/GGS}$, the statistical threshold of $I_{GGCS/GGS}$ and RFCA are utilized as the criteria for parameter optimization. The procedure of the novel FIVMD method, as shown in Figure 3.3, is discussed below.

Step 1: Initialize the mode number k as 1 and set a maximum decomposition number K_{max}. According to Refs. [85, 138, 139] and by trial and error, the VMD functions

well when the bandwidth control parameter is set in the interval of [1000 4000]. Therefore, without loss of generality, the bandwidth control parameter α is initialized as the mean value 2500 in Step 1.

Step 2: Decompose the signal into k BLIMFs using the VMD method.

Step 3: Calculate the $I_{GGCS/GGS}$ and its statistical threshold which is equal to the critical value of Chi-square distribution with $N-1$ DOFs and 0.05 significance level of each BLIMF. In this chapter, the N is obtained through sampling frequency divided by the FCF. Considering a slight difference between the calculated N and the actual N, it is suggested to set the cyclic period from $N-5$ to $N+5$, and the cyclic period with the maximum log-likelihood value can be utilized as the final cyclic period.

Step 4: Search the maximum $I_{GGCS/GGS}$ value among the k BILMFs, and if (1) there is at least one $I_{GGCS/GGS}$ exceeding the threshold using mode number k, and (2) the maximum $I_{GGCS/GGS}$ value obtained by using mode number k is bigger than those using $k-1$ and $k+1$, then the k is saved as one possible option K_p to access the optimal mode number K. The satisfaction of these two conditions can guarantee that for the selected mode number K_p, the bearing fault information can be obtained and the maximum $I_{GGCS/GGS}$ in the selected mode is bigger than those in adjacent mode numbers. Furthermore, the incorporation of $I_{GGCS/GGS}$ makes the selection of possible mode number sensitive to bearing fault signature and less affected by other cyclic frequency contents and impulsive noise, as a result of solely reflecting the cyclostationarity of the given faulty bearing.

Step 5: Compare the K_p with K_{max}, if $K_p < K_{max}$, go back to Step 2 and set $k = k + 1$. If $K_p \geq K_{max}$, save all possible mode numbers $K_{p1}, K_{p2,} \ldots, K_{pi}$ and move to Step 6.

Step 6: All the possible mode numbers with different bandwidth control parameters α_p (set to be 1000:200:4000) are regarded as possible pairs of optimal mode number and bandwidth control parameter, then calculate the RFCA of each BLIMF using these pairs of K_p and α_p.

Step 7: Among these possible pairs, a pair will be chosen as the final optimal K and α which maximize the RFCA, via

$$opt_{K,\alpha} = \arg\max\{\mathrm{RFCA}(K, \alpha)\} \tag{3-18}$$

Then decompose the signal according to the optimal K and α, and the BLIMF with the maximum RFCA will be used for further analysis.

3.4. Simulation analysis

By incorporating the $I_{GGCS/GGS}$, the statistical threshold of $I_{GGCS/GGS}$ and the RFCA, the proposed FIVMD can extract the BLIMF with the strongest bearing fault information and is less affected by other cyclic frequency contents and impulsive noise. To validate the performance of the FIVMD, the signal that simulates the vibrations of the rotating machine derived from Ref. [140] is adopted for analysis. The simulated signal is generated by

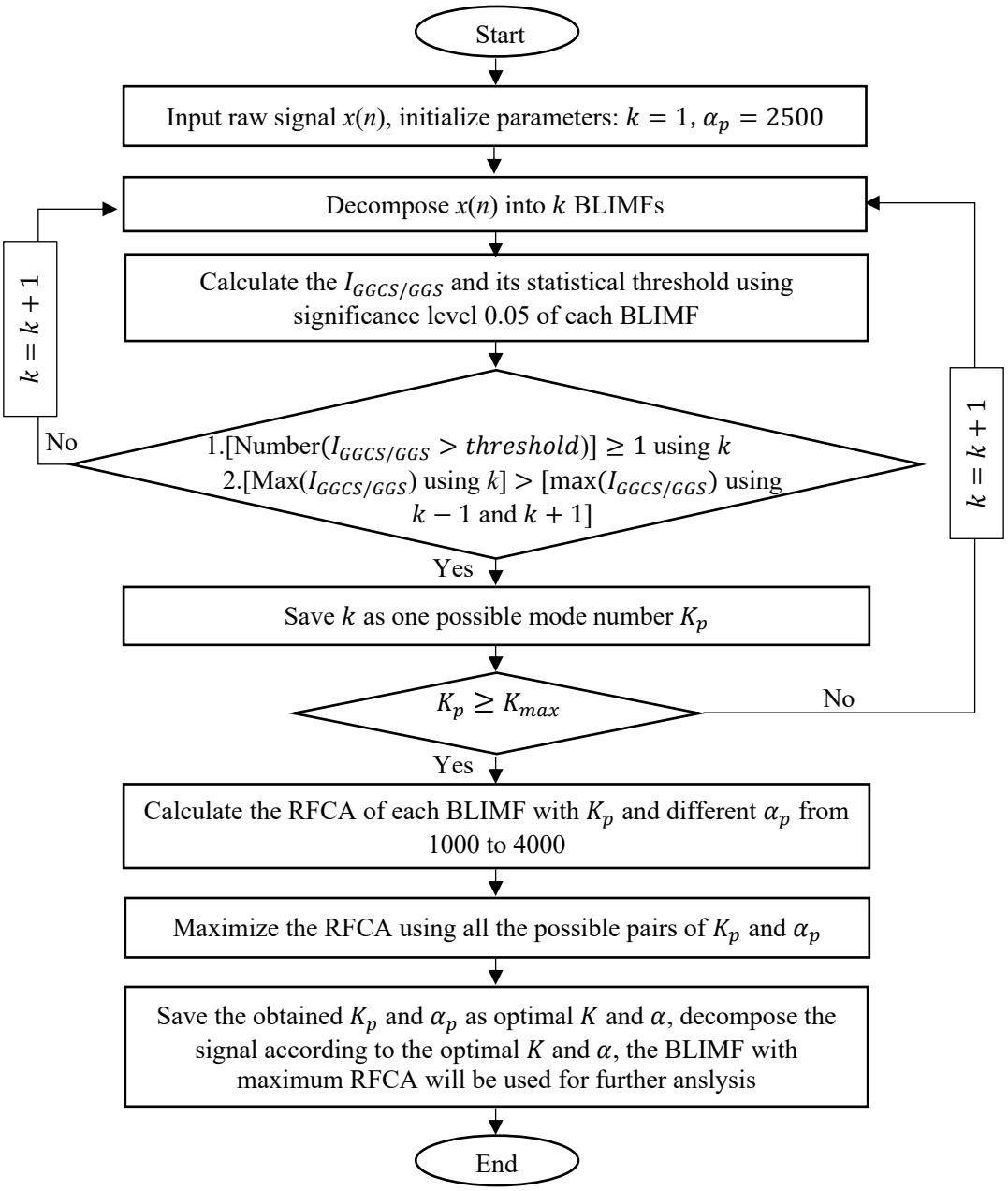


Figure 3.3. Flowchart of the FIVMD method

$$\begin{cases} X(t) = x_1(t) + x_2(t) + x_3(t) + x_4(t) + n(t) \\ x_1(t) = \sum_l A_l \cos(2\pi l f_o t + \theta_l) \\ x_2(t) = \sum_k B_k \cos(2\pi k Z f_o t + \phi_k) \\ x_3(t) = \sum_i C_i e^{-2\pi\zeta f_n(t-iT-\tau_i)} \sin\left(2\pi\sqrt{1-\zeta^2} f_n(t-iT-\tau_i) + \varphi_0\right) \\ x_4(t) = \sum_j D_j e^{-2\pi\zeta' f_n' t} \sin\left(2\pi\sqrt{1-\zeta'^2} f_n' t\right) \end{cases} \quad (3\text{-}19)$$

where $x_1(t)$ denotes the shaft vibration, with f_o, A_l and θ_l being the shaft rotating frequency, initial amplitude and phase of the shaft, respectively. $x_2(t)$ represents the gear meshing vibration with Z, B_k and ϕ_k denoting the tooth number, initial amplitude and phase of the gear meshing, respectively. $x_3(t)$ is the bearing fault vibration, which shares the same meaning as in Equation (3-16) and Equation(3-17). $x_4(t)$ denotes the abnormal impulses which may be generated by the external environment and are simulated as the impulse response functions with different natural frequencies and initial amplitudes. Besides, the added abnormal impulses are assumed to consist of three impulses whose initial amplitudes are set to be three random variables that follow the uniform distribution between 0 and 4, considering their randomness. $n(t)$ is Gaussian white noise which is employed to simulate the background noise 0 dB. Except for the initial amplitude setting as a constant 0.3, residual parameters utilized for $x_3(t)$ remain the same as Table 3-1. Other parameters used in this simulation analysis are listed in Table 3-2. The time-domain waveform, the spectrum and the HES of the simulated signal are displayed in Figure 3.4. It can be found that the shaft rotating and gear meshing frequencies are dominant in the spectrum, showing obvious peaks at 20 Hz, 640 Hz and 1280 Hz. While the bearing fault symptom is submerged by the interferences since no peaks are visible at the FCF and its harmonics in the spectrum and envelope spectrum.

Table 3-2 Parameters of the simulated vibration for the rotating machine

$x_1(t)$			$x_2(t)$					$x_4(t)$			
f_o	A_1	θ_1	Z	B_1	B_2	ϕ_1	ϕ_2	ζ'	${f_{n1}}'$	${f_{n2}}'$	${f_{n3}}'$
20	0.1	$\pi/2$	32	0.2	0.08	$\pi/2$	$\pi/2$	0.015	6000	7000	9000

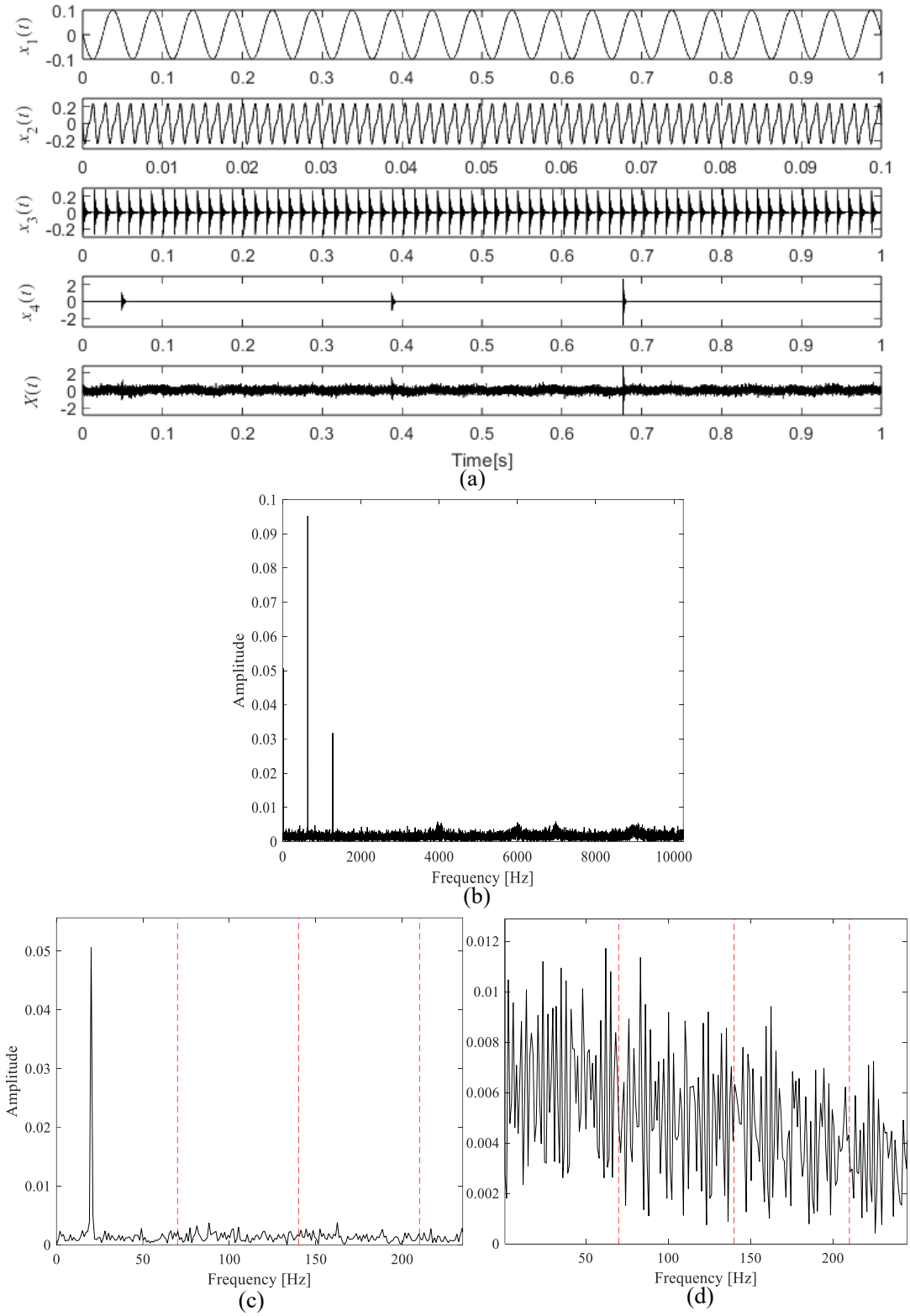


Figure 3.4. The simulated signal: (a) time domain; (b) the spectrum; (c) the zoomed spectrum (0-3.5×FCF); (d) the zoomed HES (0-3.5×FCF)

Then the simulated signal is analyzed by the proposed FIVMD method. The possible mode numbers selected by $I_{GGCS/GGS}$ and its statistical threshold are 6 and 9, which are displayed in Figure 3.5 with the RFCAs of all possible pairs of mode numbers and bandwidth control parameters. As shown in Figure 3.5, mode number 6 and bandwidth control parameter 3800 can maximize the RFCA, thus they are selected as the optimal parameters by FIVMD. The spectrum of all decomposed BLIMFs using the optimal mode number and bandwidth control parameter is presented in Figure 3.6. It can be seen that 20 Hz, 640 Hz (and 1280 Hz), 4000 Hz, 6000 Hz, 7000 Hz and 9000Hz are dominant in these 6 BLIMFs, which are well in line with the shaft rotating frequency, gear meshing frequencies, bearing's natural frequency, and abnormal impulses' nature frequencies, respectively, indicating that the proposed FIVMD method can decompose the multi-component signal into appropriate BLIMFs.

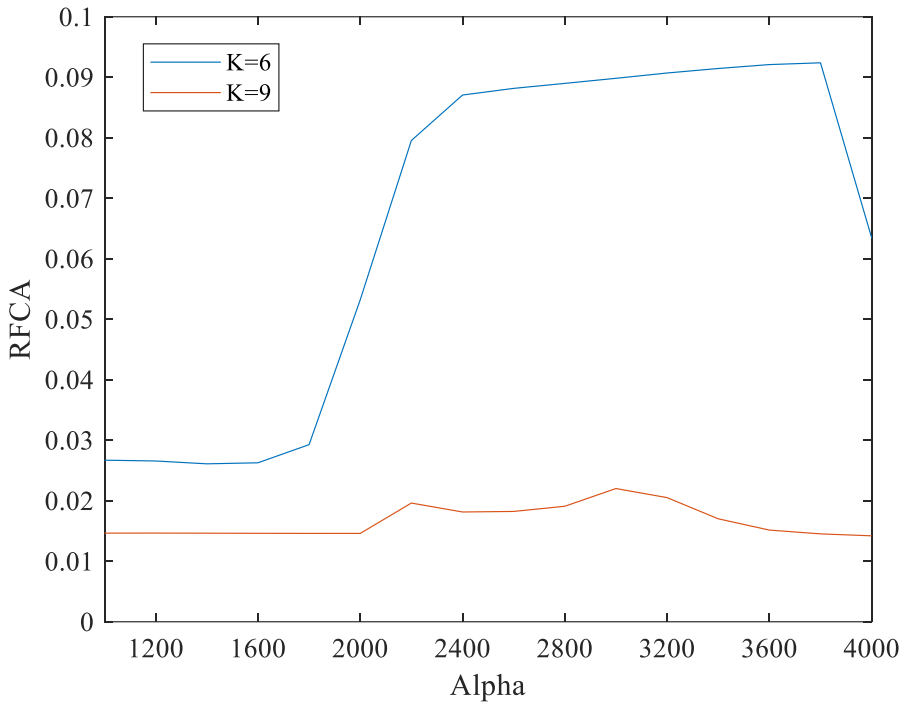


Figure 3.5. The RFCAs of the possible pairs of mode numbers and bandwidth control parameters

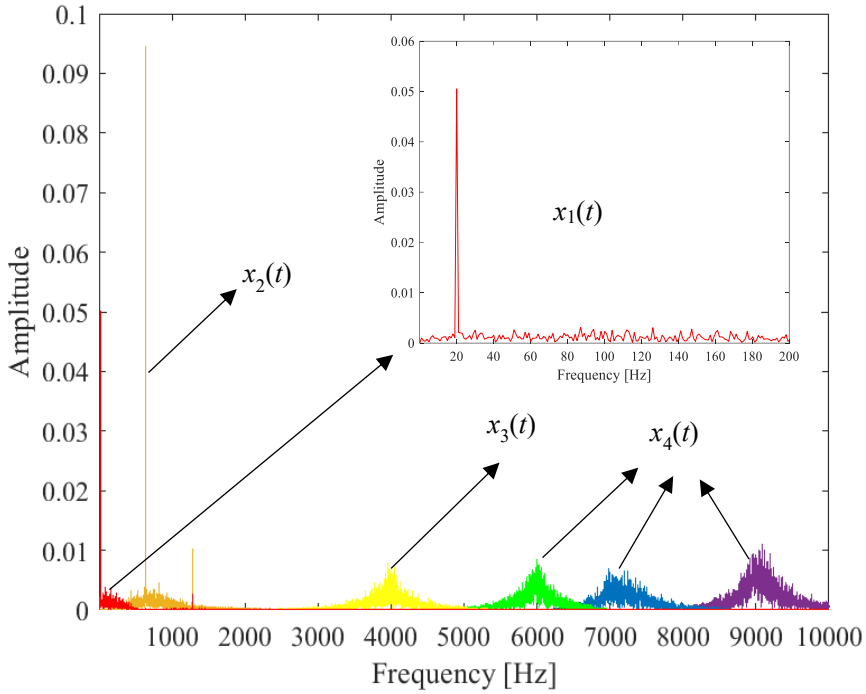


Figure 3.6. The spectrum of all decomposed BLIMFs using the optimal mode number and bandwidth control parameter selected by the FIVMD

Furthermore, the HES of the BLIMF with the maximum RFCA using FIVMD is presented in Figure 3.7 (a), obvious peaks are visible at the bearing FCF and its harmonics, demonstrating that the proposed FIVMD can extract the weak bearing respective transients under complex interferences. To illustrate the superiority of the proposed FIVMD over other traditional methods, the HES of the BLIMF with maximum RFCA using the original VMD, the Hilbert envelope spectra of the first 6 IMFs using EMD, and the Hilbert envelope spectra of the first 6 PFs using LMD, are plotted in Figure 3.7 (b), Figure 3.7 (c), Figure 3.7(d) respectively. As shown in Figure 3.7, the EMD and LMD methods can not extract the bearing fault signature as no bearing fault-related peaks can be found in the envelope spectra. While employing the HES of original VMD can diagnose the bearing fault, but the bearing fault-related absolute amplitudes are lower and the noise level is higher than those in the HES of FIVMD. For further quantifying the bearing fault information, the RFCAs of Figure 3.7 (a), Figure 3.7 (b), Figure 3.7 (c), Figure 3.7(d) are calculated, which are 0.092385,

0.042134, 0.019230 (RFCA of IMF5, RFCAs of all IMFs are calculated and the RFCA of IMF5 is the highest) and 0.01865 (RFCA of PF3, RFCAs of all PFs are calculated and the RFCA of PF3 is the highest), respectively. The proposed FIVMD method, therefore, can extract a BLIMF that contains the most bearing fault information. To summarize, the original signal, EMD-decomposed IMFs and LMD-decomposed PFs are unable to diagnose the bearing fault, while both the original VMD-decomposed and FIVMD-decomposed BLIMFs can extract the weak bearing respective transients under complex interferences, however, the FIVMD method provides a better option as more bearing fault information is included.

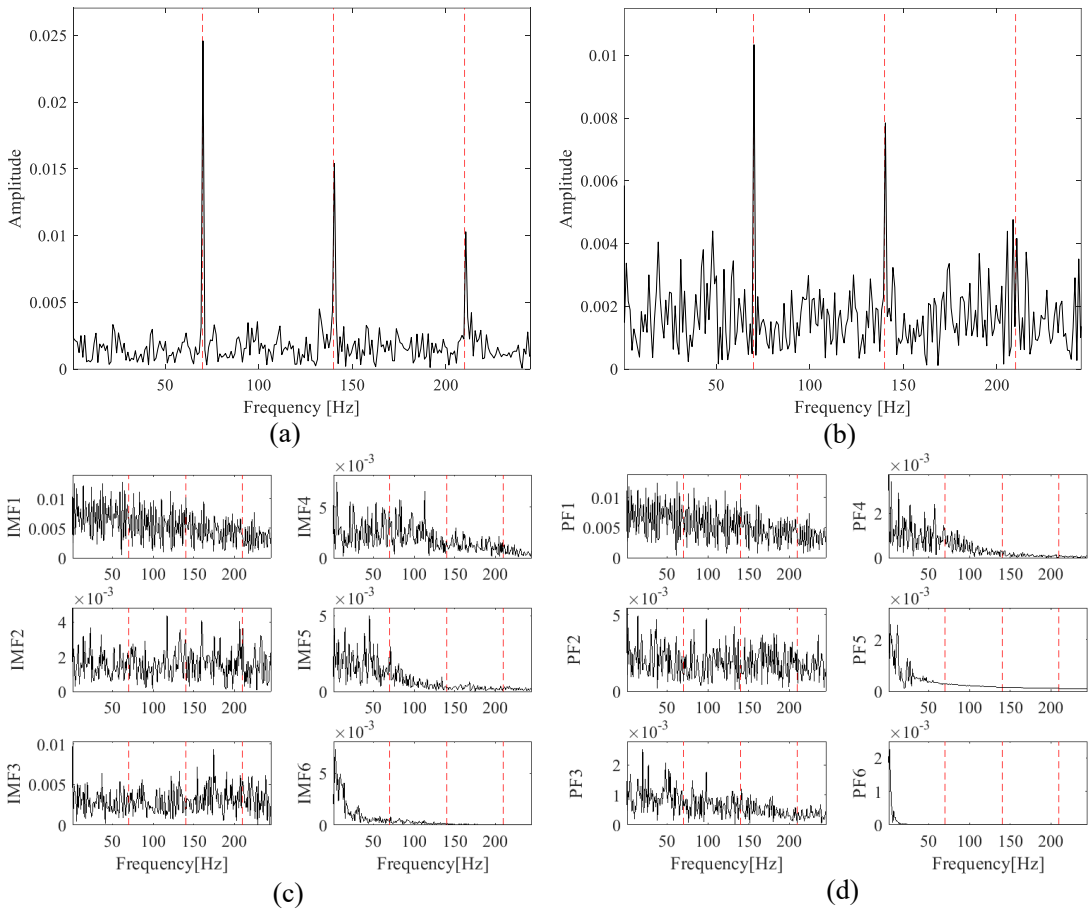


Figure 3.7. Analysis of the simulated signal: (a) The HES of the BLIMF with maximum RFCA using the FIVMD; (b) the HES of the BLIMF with maximum RFCA using the original VMD; (c) the Hilbert envelope spectra of the first 6 IMFs using EMD; (d) the Hilbert envelope spectra of the first 6 PFs using LMD

3.5. Experimental validation

In this section, experimental datasets are utilized to demonstrate the effectiveness of the proposed FIVMD. Three categories of challenging datasets, including the early bearing fault signals from a run-to-failure test, the bearing fault signals under impulsive noise and the faulty planet bearing signals, are employed to validate the superiority of the FIVMD. Furthermore, the EMD, LMD and original VMD are used for comparison.

3.5.1. Early bearing fault diagnosis from a run-to-failure test

The performance of the FIVMD for early bearing fault diagnosis is validated using the run-to-failure dataset from Intelligent Maintenance Systems (IMS) Center [141]. The layout of the test rig is displayed in Figure 3.8. The test rig consists of four test bearings on one shaft which is driven by the alternating current motor via rub belts. The dataset of Bearing 1 (specifications listed in Table 3-3) is used for analysis. This dataset is composed of 984 individual files which are 1-second vibrations collected every 10 minutes at the sampling frequency of 20 kHz. The experiment was run at a constant speed of 2000 revolution-per-minute for about 7 days, and the outer race fault was observed at the end of the test, more information can be referred to Ref. [141].

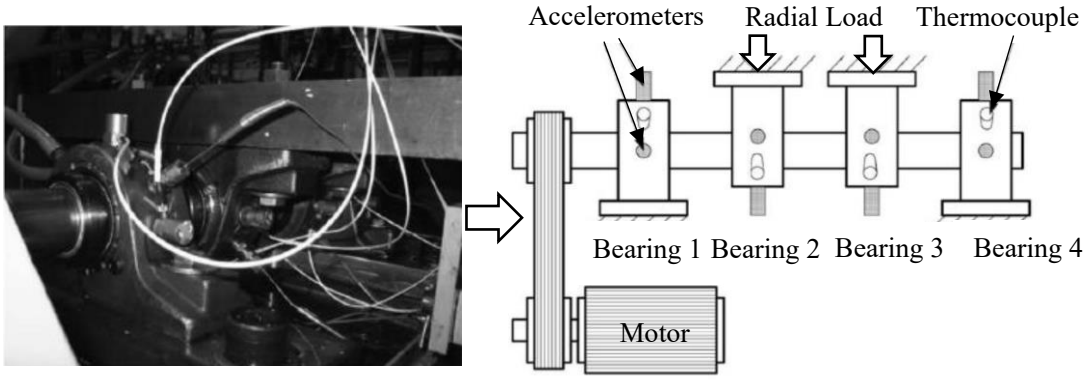


Figure 3.8. The layout of the IMS run-to-failure bearing test rig

Table 3-3 Specifications of the bearing in IMS's run-to-failure test

Bearing Type	Number of rollers	Contact angle	Pitch diameter	Roller diameter	Ball pass frequency of outer race
ZA-2115 of Rexnord	16	15.17°	2.815 in	0.331 in	236.4 Hz

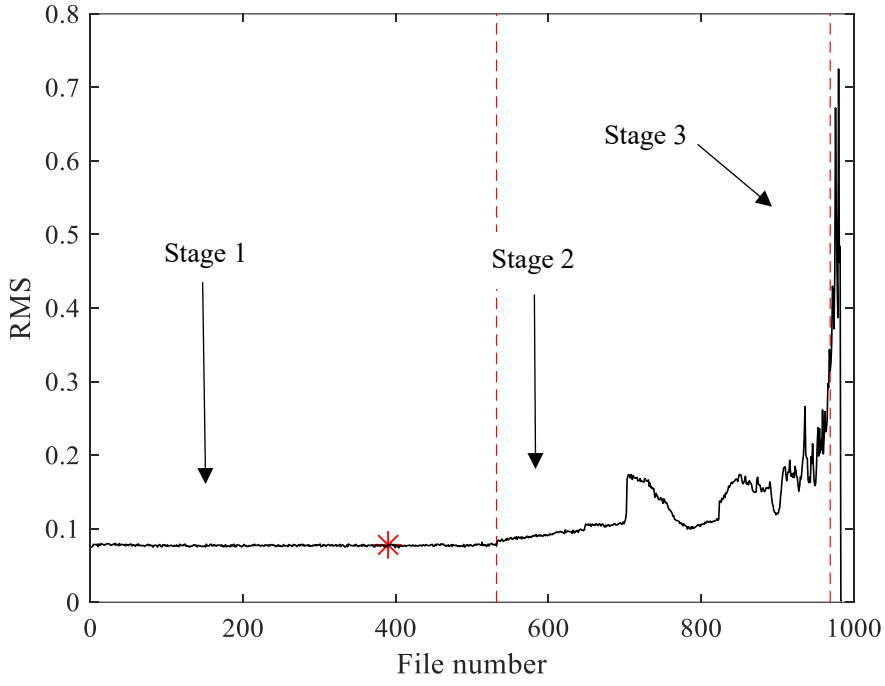


Figure 3.9. The RMS of the test bearing

The RMS of the 984 individual files is calculated and plotted in Figure 3.9. According to Figure 3.9 and Ref. [120], the entire life cycle of the test bearing can be divided into three stages (stage 1: 1st - 533rd files; stage 2: 534th - 969th files; stage 3: 970th - 984th files). To verify the effectiveness of these decomposition methods for ordinary

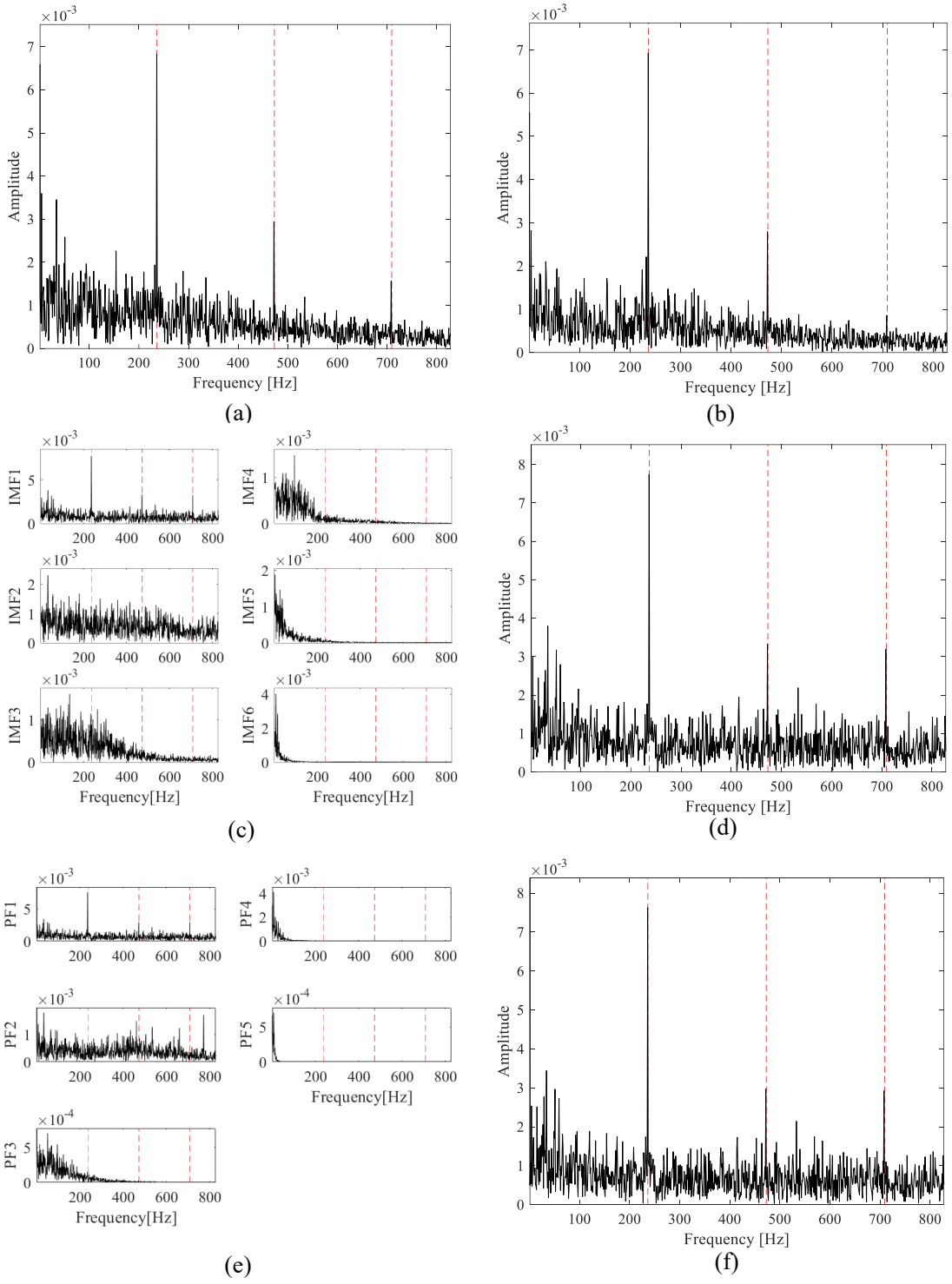


Figure 3.10. Analysis of file 534: (a) The HES of the BLIMF with maximum RFCA using the FIVMD; (b) the HES of the BLIMF with maximum RFCA using the original VMD; (c) the Hilbert envelope spectra of the first 6 IMFs using EMD; (d) HES of IMF1 using EMD; (e) the Hilbert envelope spectra of the first 5 PFs using LMD; (f) HES of PF1 using LMD

bearing fault diagnosis, file 534 is utilized. The signal is analyzed by the FIVMD method at first, through which the optimal mode number and bandwidth control

parameter are determined as 4 and 3200, respectively. Then the RFCAs of all BLIMFs using these optimal parameters are calculated, and the HES of the BLIMF with maximum RFCA, which is the first mode with RFCA of 0.022418 in this case, is plotted in Figure 3.10 (a). Simultaneously, the HES of the BLIMF with maximum RFCA using the original VMD, the Hilbert envelope spectra of the first 6 IMFs using EMD, the HES of the IMF with maximum RFCA (IMF1 in this case), the Hilbert envelope spectra of the first 5 PFs using LMD (in the following discussions, if the decomposed sub-signals exceed 6, then the first 6 sub-signals will be plotted; while if the sub-signals are less than 6, then all the sub-signals will be plotted), the HES of the PF with maximum RFCA (PF1 in this case) are displayed in Figure 3.10 (b), Figure 3.10 (c), Figure 3.10 (d), Figure 3.10 (e), and Figure 3.10 (f), respectively. From Figure 3.10, obvious peaks can be noticed at the bearing FCF and its harmonics in all the Hilbert envelope spectra, revealing that the proposed FIVMD as well as other decomposition methods can all extract the bearing respective transients at this time.

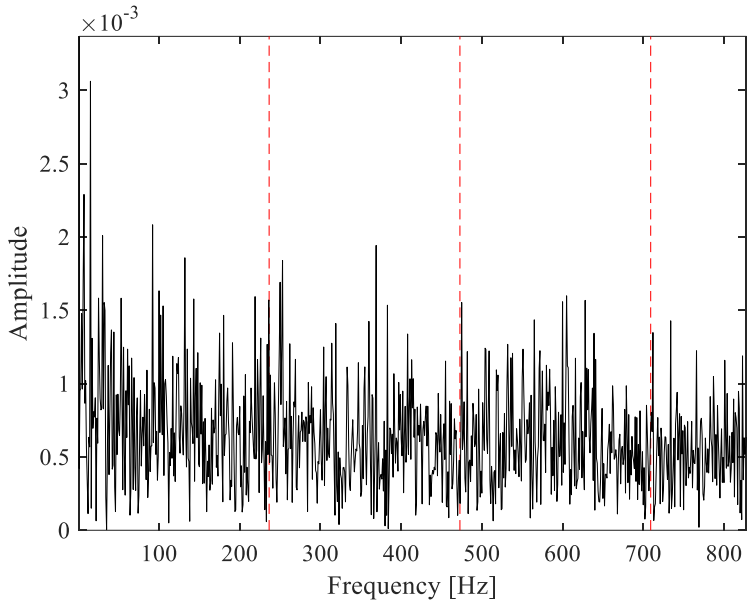


Figure 3.11. The HES of the original signal for file 390

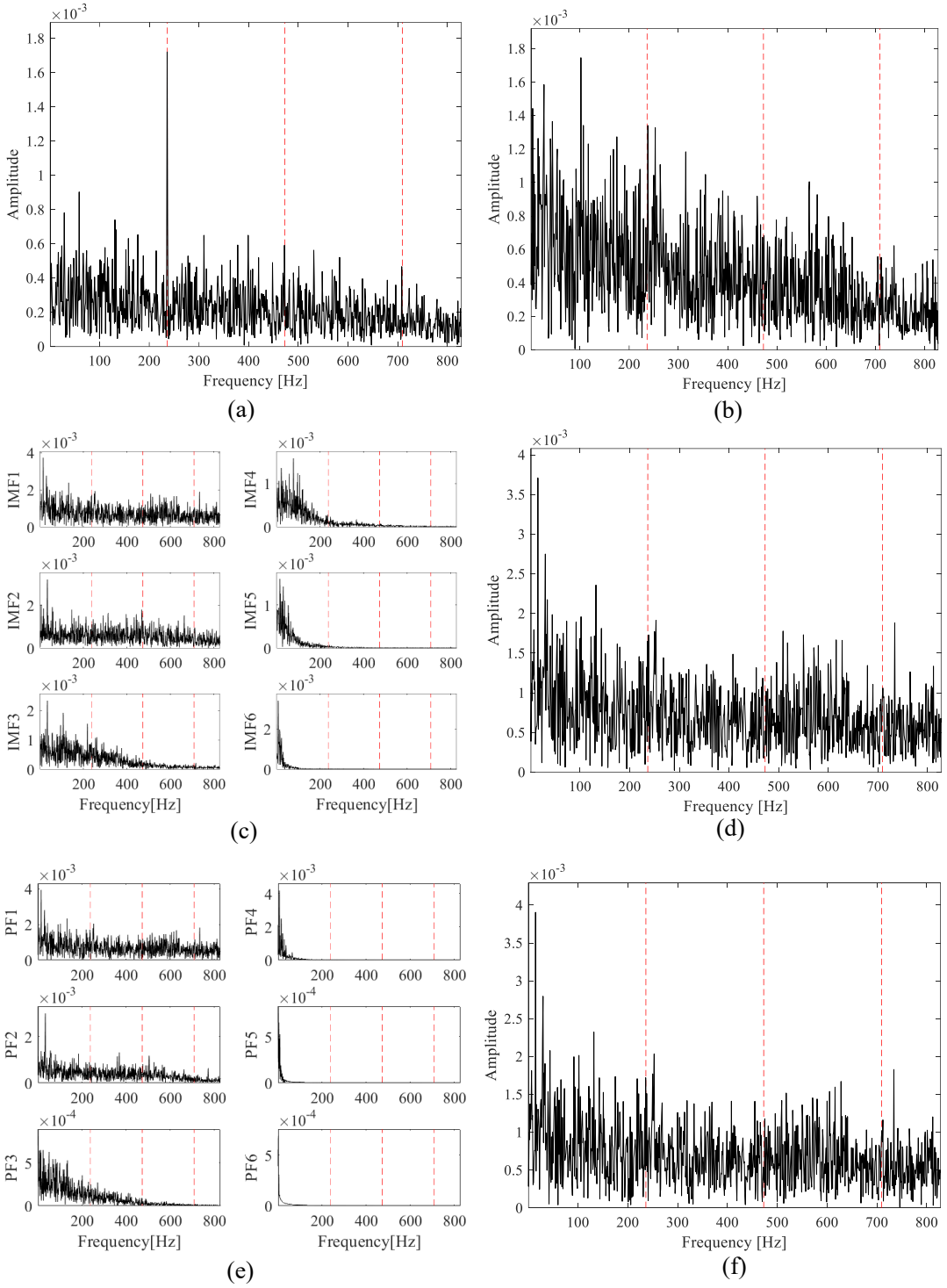


Figure 3.12. Analysis of file 390: (a) The HES of the BLIMF with maximum RFCA using the FIVMD; (b) the HES of the BLIMF with maximum RFCA using the original VMD; (c) the Hilbert envelope spectra of the first 6 IMFs using EMD; (d) HES of IMF1 using EMD; (e) the Hilbert envelope spectra of the first 6 PFs using LMD; (f) HES of PF1 using LMD

It is of great significance to diagnose the bearing fault at an early stage to avoid performance degradation of the whole machine and unexpected accidents, thus file 390 which was saved one day earlier than file 534 is employed for analysis. The HES of the original signal is displayed in Figure 3.11 at first, representing no diagnostic information of this bearing. After that, the Hilbert envelope spectra based on the FIVMD, the original VMD, the EMD and LMD are plotted in Figure 3.12, which are processed through the same procedure as for file 534. It can be found from Figure 3.12 that only the FIVMD-based HES can extract the weak bearing respective transients, identifying the bearing fault at an early stage. While other decomposition methods are overwhelmed by the heavy background noise and thus are unable to accurately diagnose the early bearing fault.

3.5.2. Bearing fault diagnosis under impulsive noise

The dataset used in Chapter 3.5.2 is provided by the Case Western Reserve University (CWRU) [142], which has become one of the benchmark datasets for bearing diagnosis. As shown in Figure 3.13, the CWRU bearing test rig is mainly composed of three parts, including one 2hp electric motor on the left, one torque transducer in the middle and one dynamometer on the right. Two test bearings installed on the fan end and drive end were introduced with localized faults for analysis. The vibrations were collected for about 10 seconds with sampling frequencies at 12k and 48k Hz, respectively. Specifications of the bearings are listed in Table 3-4, more details can be found in Ref. [142].

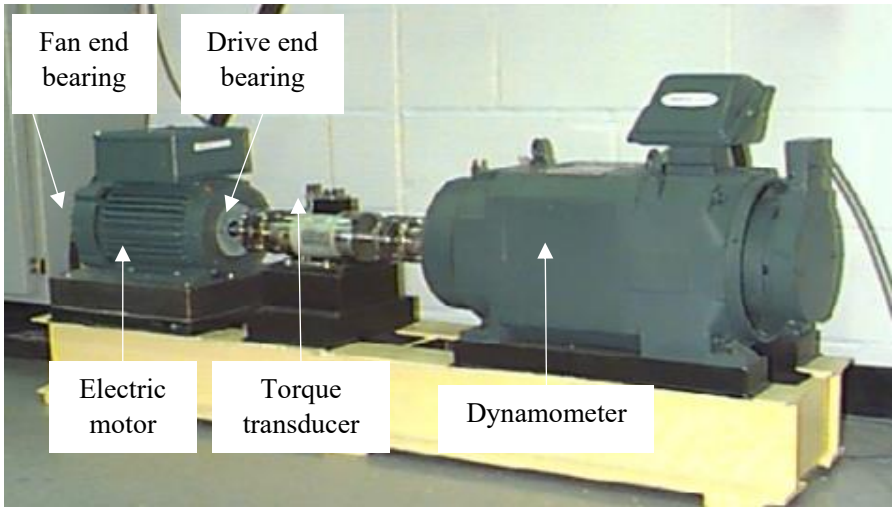
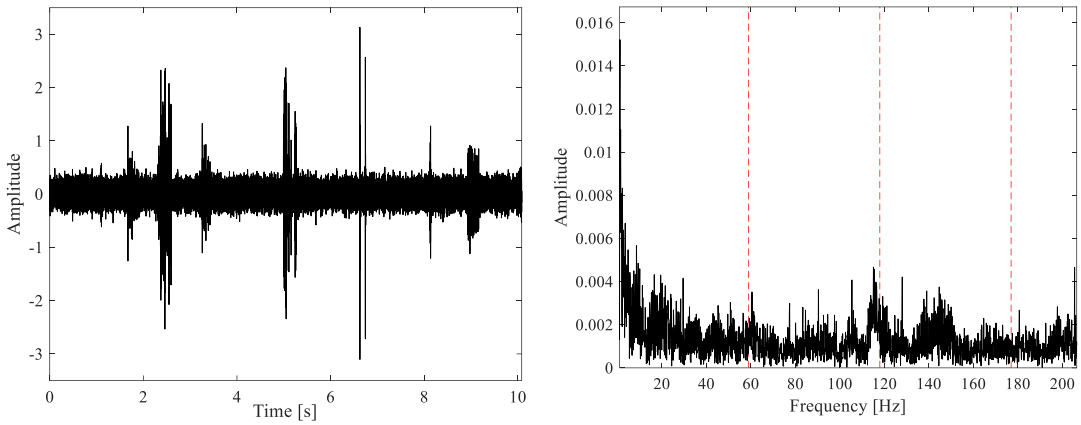


Figure 3.13. The layout of the CWRU bearing test rig [142]

Table 3-4 Specifications of the bearings in CWRU

Position on rig	Bearing Type	Number of rollers	Contact angle	Pitch diameter	Roller diameter
Drive end	SKF 6205-2RS	9	0º	39.04 mm	7.94 mm
Fan end	SKF 6203-2RS	8	0º	28.4988 mm	6.7462 mm



(a) (b)

Figure 3.14. (a) The time-domain signal of file 291 FE; (b) the HES of file 291 FE

The recorded file 291FE with a rolling element fault is used to validate the performance of the FIVMD under impulsive noise since several strong non-periodic impulses were found in this vibration signal in Ref. [56]. The signal in the time domain

and its HES are plotted in Figure 3.14, showing that the bearing fault is not diagnosable by using the HES of the original signal. Then the signal is analyzed using the same procedure presented in Chapter 3.5.1, with the Hilbert envelope spectra displayed in Figure 3.15. It is easily observed from Figure 3.15 that the peaks at bearing FCF and its harmonics are visible in all the Hilbert envelope spectra. But the FCF and its harmonics in the original VMD-based HES are not dominant, the amplitudes of the shaft rotating frequency and other components are higher. Besides, the noise levels are higher and the bearing fault-related absolute amplitudes are lower in the EMD-based HES and LMD-based HES than those in the FIVMD-based HES. To further quantify the diagnostic performance of these methods, the RFCAs of Figure 3.15 (a), Figure 3.15 (b), Figure 3.15 (d), and Figure 3.15 (f) are calculated, which are 0.031981, 0.012642, 0.010772, 0.012476, respectively. Therefore, the diagnostic performance of FIVMD is better than all the other methods under impulsive noise. In addition, by using the FIVMD, other bearing fault datasets with impulsive noise such as the 203DE, 284DE, 284FE, and 285DE [56, 143] can be similarly diagnosed.

3.5.3. Faulty planet bearing diagnosis

The dataset used in this section was collected from the DDS test rig. Details of this test rig and specifications of the gearbox can refer to Chapter 2.5.3. A planet bearing with seeded inner race fault was installed on the planetary gearbox to evaluate the performance of FIVMD. The vibration was collected at the motor speed of 30 Hz and the sampling frequency of 25600 Hz. According to the kinematics of the planet bearing, the BPFI of this planet bearing is 2.266 times the input shaft frequency, which is 67.98 Hz on this occasion. The planet bearings are compassed by the planet gears

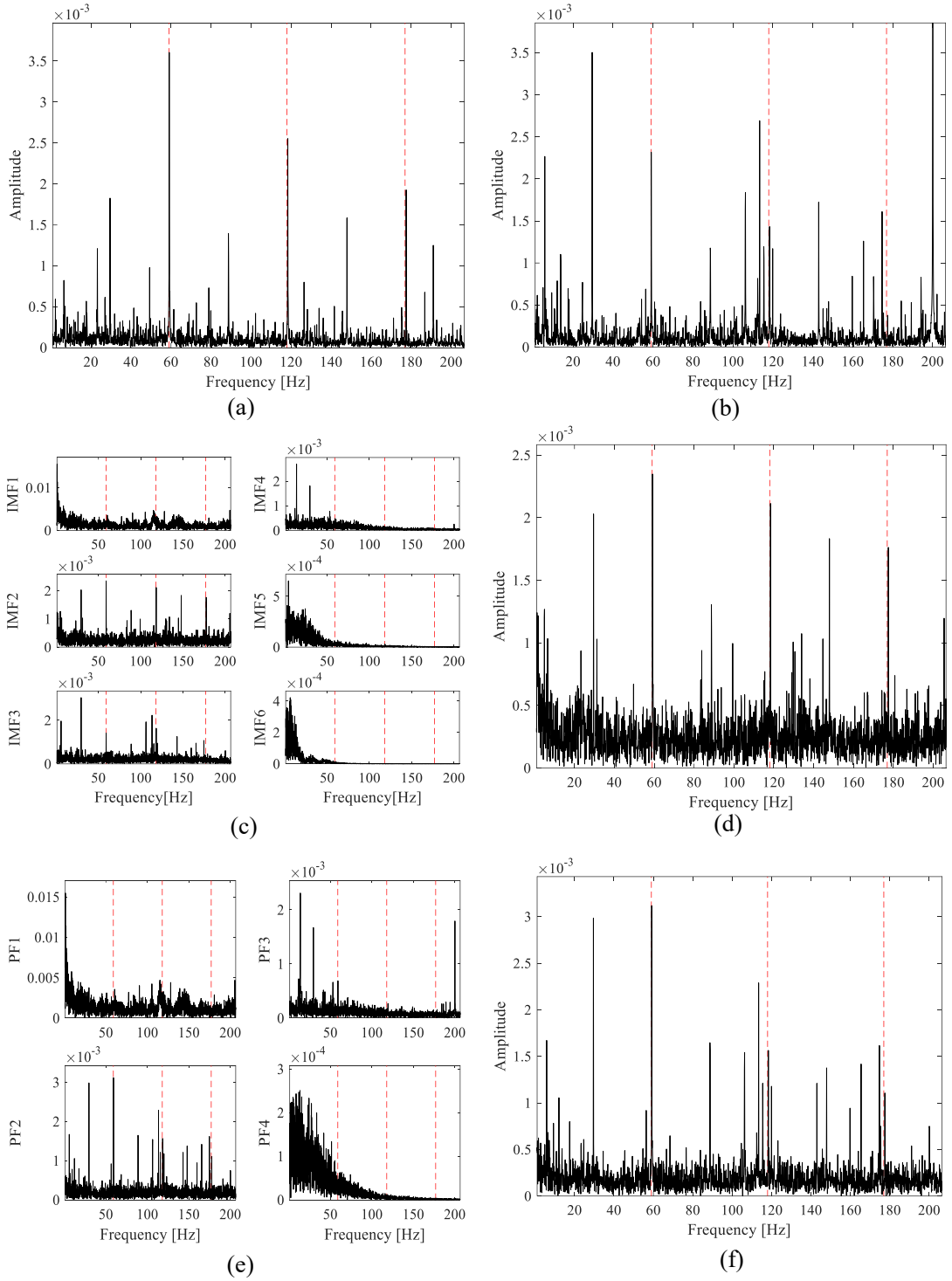


Figure 3.15. Analysis of 291 FE dataset: (a) The HES of the BLIMF with maximum RFCA using the FIVMD; (b) the HES of the BLIMF with maximum RFCA using the original VMD; (c) the Hilbert envelope spectra of the first 6 IMFs using EMD; (d) HES of IMF2 using EMD; (e) the Hilbert envelope spectra of the first 4 PFs using LMD; (f) HES of PF2 using LMD

and ring, and the planet bearings are fixed with the planet carrier, which makes it also rotate with the planet carrier. The unique kinematics of the planetary gearbox results

in the external measured vibrations being collected through time-varying transfer paths and mingled interferences from other components, enabling the diagnosis of planet bearings to be difficult.

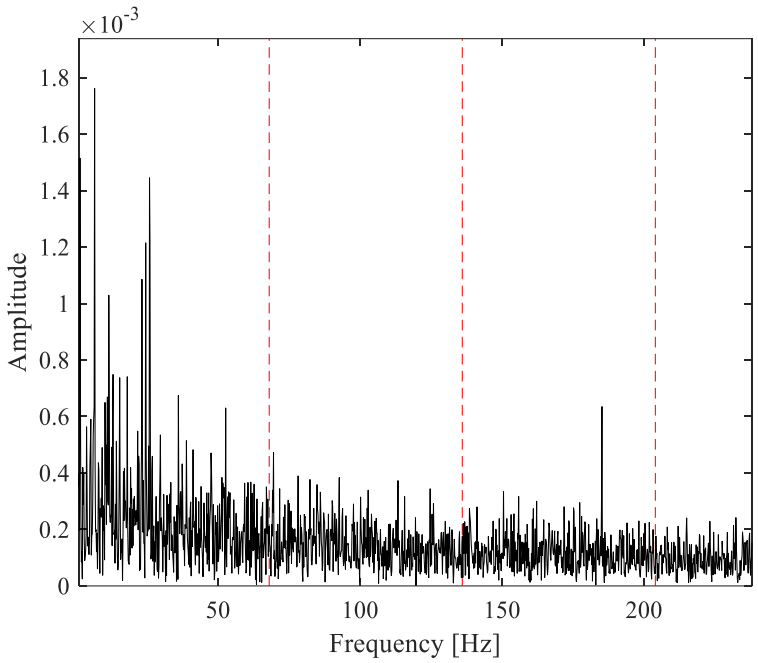


Figure 3.16. The HES of the planet bearing

The HES of the original signal is displayed in Figure 3.16, demonstrating that the planet bearing fault can not be diagnosed using the HES of the original signal. Note that although the test rig and operating speed used in Chapter 2.5.3 are the same as those in this section, these signals are not collected at the same time, leading to slight differences between Figure 2.15 and Figure 3.16 (Another possible reason is that Figure 2.15 is the order spectrum while Figure 3.16 is the frequency spectrum). Then the signal is processed using the same steps shown in Chapter 3.5.1, with the Hilbert envelope spectra plotted in Figure 3.17. It can be found from Figure 3.17 that the bearing fault can not be identified by the original VMD-based HES. Small fault-related peaks are visible in the EMD-based HES. However, they are very weak

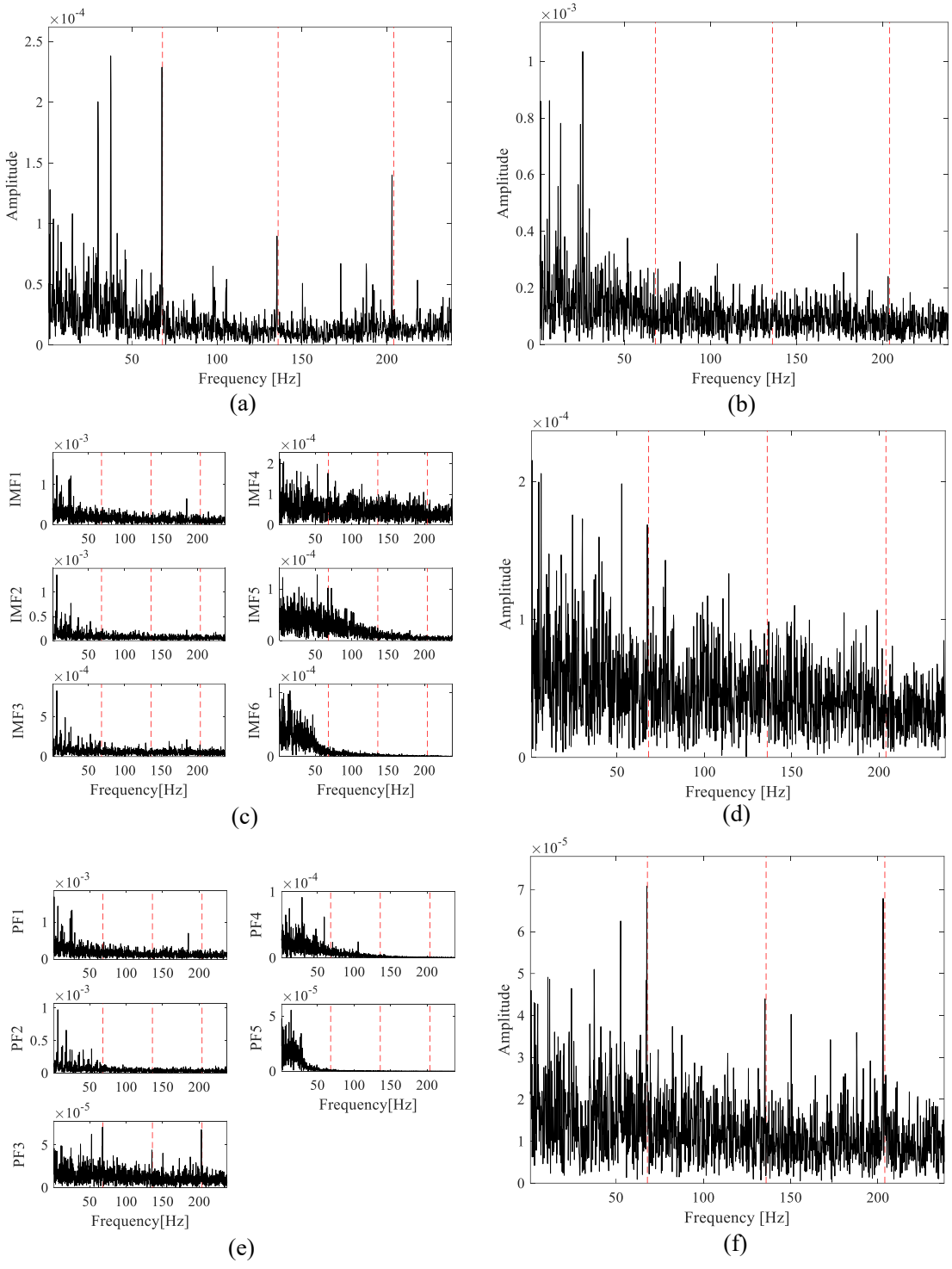


Figure 3.17. Analysis of planet bearing: (a) The HES of the BLIMF with maximum RFCA using the FIVMD; (b) the HES of the BLIMF with maximum RFCA using the original VMD; (c) the Hilbert envelope spectra of the first 6 IMFs using EMD; (d) HES of IMF4 using EMD; (e) the Hilbert envelope spectra of the first 5 PFs using LMD; (f) HES of PF3 using LMD

compared with the background noise and there are no visible peaks at the harmonics of the FCF. On the contrary, the peaks at the FCF and its harmonics in FIVMD-based

HES and LMD-based HES are dominant, but the absolute amplitudes of the bearing fault-related components are higher in FIVMD-based HES. Moreover, the noise level is lower in the FIVMD-based HES. To further quantify the diagnostic performance of these methods, the RFCAs of Figure 3.17 (a), Figure 3.17 (b), Figure 3.17 (d), and Figure 3.17 (f) are calculated to be 0.015191, 0.004146, 0.004476, 0.009154, respectively. This also confirms that the performance of FIVMD-based HES is better than the other methods. In summary, the FIVMD is superior in extracting the weak bearing fault repetitive transients under vibration interferences from other components. Furthermore, by using the FIVMD, other planet bearing faulty datasets under different input speeds such as 40Hz and 50Hz can be similarly diagnosed and their RFCAs are listed in Table 3-5.

3.5.4. Summary of results

The results obtained from simulation analysis and experimental verification are summarized as follows.

In the simulated signal, only the original VMD and FIVMD can diagnose the bearing fault, since no peaks at FCF and its harmonics are visible when using the other methods. But the noise level in the original VMD-bassed HES is much higher than that in the FIVMD-based HES. For the early bearing fault diagnosis, the FIVMD is the only method that can extract the bearing fault feature and diagnose the bearing fault appropriately. As for the signal under impulsive noise, all the Hilbert envelope spectra except for the HES of the original signal show visible peaks at bearing fault-related frequencies, but the amplitudes of these frequencies in the FIVMD-based HES are the highest with less background noise. For the planet bearing fault diagnosis, the fault bearing signal is inevitably entangled with vibrations from other components,

resulting in the incapability of original signal-based HES and original VMD-based HES. There are bearing fault-related peaks appearing in the EMD-based, LMD-based and FIVMD-based Hilbert envelop spectra, however, the amplitudes in the EMD-based HES are very small and no peaks are visible at FCF's harmonics. In addition, the absolute amplitudes of the bearing fault-related components are higher and the noise level is lower in the FIVMD-based HES than those in the LMD-based HES. In short, the performance of the FIVMD-based HES is the highest in all the circumstances, showing strong peaks at the FCF and its harmonics with lower background noise.

Table 3-5 The RFCAs of the simulated and experimental datasets

Name	Condition	FIVMD	Original VMD	EMD	LMD	Orignal signal
Simulated signal	Simulation	0.092385	0.042134	0.019230	0.018651	0.016941
File 390	Early fault diagnosis	0.013821	0.006694	0.006010	0.006263	0.007201
203 DE	Impulsive noise	0.007600	0.006751	0.005009	0.004655	0.004964
284 DE	Impulsive noise	0.035800	0.034643	0.006286	0.025916	0.007692
284 FE	Impulsive noise	0.032478	0.023383	0.017232	0.022539	0.004442
285 DE	Impulsive noise	0.034026	0.030794	0.008569	0.027842	0.006222
291 FE	Impulsive noise	0.031981	0.012642	0.010772	0.012476	0.002382
Planet bearing	Planetary gearbox (30 Hz)	0.015192	0.004146	0.004476	0.009154	0.003635
Planet bearing	Planetary gearbox (40 Hz)	0.016378	0.003094	0.006407	0.008929	0.002381
Planet bearing	Planetary gearbox (50 Hz)	0.012284	0.002482	0.004872	0.005517	0.002505

To further quantify the diagnostic performance of these methods, the RFCAs calculated from the simulated and experimental datasets are summarized in Table 3-5. It can be concluded that the RFCAs based on the FIVMD are the highest in all the

conditions, demonstrating a better diagnostic performance of FIWMD than the other methods used in this chapter.

3.6. Conclusion

The novelty and contribution of this chapter can be summarised as follows.

1) The $I_{GGCS/GGS}$ and the statistical threshold of $I_{GGCS/GGS}$ was introduced, and the RFCA was proposed to optimally determine the mode number and bandwidth control parameter.

2) A novel FIVMD method has been proposed to address the limitations of the traditional VMD method. The novel FIVMD method enables the decomposed BLIMF to be sensitive to the bearing fault signature and less affected by the abnormal impulses and vibrations from other components.

3) Successful diagnosis in challenging datasets, including the early bearing fault signals, the bearing fault signals under impulsive noise and the faulty planet bearing signals, has been achieved using the proposed method. This confirms that the FIVMD is a promising method for bearing diagnosis under complicated operating conditions.

Chapter 4. A novel spectral correlation and wasserstein distance-based health indicator for bearing prognostics

4.1. Introduction

This chapter is a modified version of the journal paper titled “Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network”, which was published in IEEE Transactions on Industrial Informatics (19.2 (2023): 1301-1311). The work completed in this chapter is dedicated to realizing objectives 3 and 4 of this research.

Based on the literature review of Chapter 1.2.5, the construction of HI is vital for bearing prognostics, and there are usually two challenges to build a proper HI with high predictive accuracy. First, the external background or adjacent internal components generated noise are always entangled with bearing vibrations. Second, there are spurious fluctuations that are irrelevant to bearing failure progression. These phenomena can overwhelm the bearing degradation information and thus decrease the prediction accuracy. In addition, the final predicted RUL is obtained by mapping the HI and RUL using the prediction model. Thus, defining an appropriate prediction model is also essential for RUL prediction. To address the issues mentioned above, this chapter proposes a systematic data-driven prognostic scheme for bearings based on a novel HI and GRU.

The rest of this chapter is organized as follows: Chapter 4.2 focuses on constructing the novel HI, i.e., LR-WDSC. Chapter 4.3 includes the integrated BO and GRU (BO-

GRU) network and the proposed prognostic scheme. Chapter 4.4 presents the experimental data and comparison results of the proposed method. Chapter 4.5 is the conclusion of this chapter.

4.2. The construction of novel HI: LR-WDSC

This section aims to construct an appropriate HI to reveal the degradation characteristics of bearings. The novel HI is established based on the SC, WD, and LR, with the construction procedure being introduced below.

4.2.1. The spectral correlation

The bearing fault vibrations are usually treated as a series of periodic impulses. However, there are random slips at the load angle when operating, which makes the bearing fault vibrations not exactly periodic but periodical time-variance. This characteristic implies cyclostationarity, inspiring the researchers to analyze the bearing signals using cyclostationary analysis. First-order and second-order cyclostationary signals are the most common cyclostationary signals. The bearing fault signals are second-order cyclostationary signals whose second-order autocorrelation function is periodic in time [28], i.e.,

$$R_{xx}(t,\tau) = R_{xx}(t+T,\tau) \tag{4-1}$$

$$R_{xx}(t,\tau) = \mathbb{E}\left\{x\left(t-\frac{\tau}{2}\right)x\left(t+\frac{\tau}{2}\right)\right\} \tag{4-2}$$

where $x(t)$ is the time signal, t denotes time, τ indicates the time-lag, T is the periodic period, and $\mathbb{E}\{\cdot\}$ represents the average ensemble operator. Among various second-order cyclostationary analysis techniques, the SC is one of the most popular methods since the SC can reflect all information of the second-order cyclostationary signal. The SC is a bi-spectral that performs a two-dimensional Fourier transform of Equation (4 - 2) and can be expressed as

$$S_{xx}(\alpha, f) = \lim_{W\to\infty} \frac{1}{W} \int_R \int_{-W/2}^{W/2} R_{xx}(t, \tau)\, e^{-j2\pi(f\tau+\alpha t)} dt d\tau \qquad (4\text{-}3)$$

where α is the cyclic frequency transformed from time t, indicating the modulation frequency, and f is the frequency transformed from time-lag τ, indicating the carrier frequency. In the case of bearing fault signal, the autocorrelation function is periodic in t and transient in τ, resulting in the discreteness in α direction and continuity in f direction of the corresponding SC. This can be expressed as

$$S_{xx}(\alpha, f) = \begin{cases} S_{xx}(f), & \alpha = i \times \mathrm{FCF} \\ 0, & \text{others} \end{cases} \qquad (4\text{-}4)$$

where $i = 0, 1, 2, 3, \ldots$, and $S_{xx}(f)$ is known as the cyclic power-spectrum.

From Equation (4-3), the SC can be explained as the bi-spectral to reveal the elementary waves' strength at all combinations of (α, f) that x is modulated and carried. Therefore, the SC is capable of unwrapping the complicated time series into frequency and cyclic frequency axis, reflecting every detail of modulation. Hence, the SC is a promising technique to separate the bearing fault signatures from the interferences, even for those spectral supports that are overlapping, rendering the possibility to reveal bearing fault information hidden by high-level noise and will be adopted in this chapter to construct the novel HI.

4.2.2. The wasserstein distance

The WD is a measure to reflect the distance between two probability distributions [144]. It is also known as the Earth mover's distance, representing the minimum transformation cost from one probability distribution to another. Compared with

classical probability measures, such as the KL divergence and Jensen-Shannon (JS) divergence, the WD reveals not only a metric of probability distribution distance but also an approach to transform one probability distribution into another. In addition, the WD can deal with situations where the support sets of two distributions have no or little overlap, while the JS divergence is a constant, and the KL divergence may be meaningless in these cases [145]. Therefore, the WD can be recognized as a superior measure to KL and JS, and will be utilized to construct a novel HI.

Let (χ, ρ) be the Polish metric space for two probability measures $\mathbb{p}_x$ and $\mathbb{p}_y$ on χ, the p^{th} WD of $\mathbb{p}_x$ and $\mathbb{p}_y$ is defined as [146]

$$WD_p(\mathbb{p}_x, \mathbb{p}_y) = (\inf_{\gamma \in \Gamma(\mathbb{p}_x, \mathbb{p}_y)} \int_{\chi \times \chi} \rho(x,y)^p d\gamma(x,y))^{\frac{1}{p}} \quad (4\text{-}5)$$

where $\Gamma(\mathbb{p}_x, \mathbb{p}_y)$ indicates all possible joint probability measures with marginals $\mathbb{p}_x$ and $\mathbb{p}_y$ on $\chi \times \chi$, namely $\int \gamma(x,y)dx = \mathbb{p}_y(y)$ and $\int \gamma(x,y)dy = \mathbb{p}_x(x)$. $\rho(x,y)$ represents the distance function. When $p = 1$, the WD is also termed as the Earth Mover's distance, signifying the minimum cost to transport one distribution to another. It is the most commonly used version of WD, and $p = 1$ is utilized in this paper. The distance function is correspondingly set to be $\rho(x,y) = \|x - y\|$. The formulation of the WD used in this paper can therefore be written as

$$WD_1(\mathbb{p}_x, \mathbb{p}_y) = \inf_{\gamma \in \Gamma(\mathbb{p}_x, \mathbb{p}_y)} \int_{\chi \times \chi} \|x - y\| d\gamma(x,y) \quad (4\text{-}6)$$

In particular, when $p = 1$ and $\chi = R$, the WD can be calculated with an explicit formula using the inverse of the cumulative distribution functions ($F_{\mathbb{p}_x}^{-1}$and $F_{\mathbb{p}_y}^{-1}$) or cumulative distribution functions ($F_{\mathbb{p}_x}$ and $F_{\mathbb{p}_y}$) of $\mathbb{p}_x$ and $\mathbb{p}_y$ through

$$\begin{aligned} WD_1(\mathbb{p}_x, \mathbb{p}_y) &= \int_0^1 \left| F_{\mathbb{p}_x}^{-1}(a) - F_{\mathbb{p}_y}^{-1}(a) \right| da \\ &= \int_R \left| F_{\mathbb{p}_x}(b) - F_{\mathbb{p}_y}(b) \right| db \end{aligned} \tag{4-7}$$

4.2.3. The construction of LR-WDSC

To construct the novel HI, the WD of SC will be calculated and accumulated first (named WDSC), and then a LR technique will be proposed to reduce the fluctuations over the gradual deterioration of bearings. From the modeling perspective, these fluctuations of HI are not the characteristics of bearing degradation and can further complicate the model, leading to an inaccurate RUL prediction. The LR is executed through Equation (4-8) and Equation (4-9), viz

$$\begin{aligned} &LR - WDSC(i) = \\ &\begin{cases} WDSC(i), \forall WDSC(i-1) \le WDSC(i) \le WDSC(i+1) \\ WDSC(i-1) + rate, \quad \text{others} \end{cases} \end{aligned} \tag{4-8}$$

$$rate = \frac{WDSC(i+N) - WDSC(i)}{N} \tag{4-9}$$

where N represents the number from $WDSC(i)$ to the next indicator that is greater than $WDSC(i)$ for the first time. Adopting the above theoretical basis, a novel indicator named LR-WDSC is constructed, following the procedure shown in Algorithm 4-1.

Algorithm 4-1: Construction of the LR-WDSC

Input: The bearing run-to-failure dataset $\{y_i\}_{i=1}^{N}$ collected at N moments.

Output: *N* LR-WDSCs of the run-to-failure dataset.

Initialize: Select one bearing healthy signal as reference x, define the other bearing signal that needs to be analyzed as y_i.

Procedure:

Step 1: Calculate the autocorrelation function and the SC of x and y_i using Equation (4 - 2) and Equation (4 - 3).

Step 2: Slice the SC of x and y_i along α axis, obtain a series of cyclic power-spectra denoted as $\{S_{xx}(f)_{0k},\ S_{xx}(f)_{1k},\ S_{xx}(f)_{2k}, \dots, S_{xx}(f)_{mk}, \dots\}$ and $\{S_{y_iy_i}(f)_{0k},\ S_{y_iy_i}(f)_{1k},\ S_{y_iy_i}(f)_{2k}, \dots, S_{y_iy_i}(f)_{mk}, \dots\}$, where k represents the resolution in α axis, $m = 0, 1, 2, \dots, \lfloor F_s/k \rfloor$ with F_s being the sampling frequency.

Step 3: Calculate the WD between all cyclic power-spectra with the same α, namely calculate the WD between $S_{xx}(f)_{mk}$ and $S_{y_iy_i}(f)_{mk}$ from $m = 0$ to $m = \lfloor F_s/k \rfloor$.

Step 4: Obtain WDSC by accumulating the WD calculated in Step 3, viz

$$WDSC(x, y_i) = \sum_m WD\left(S_{xx}(f)_{mk}, S_{y_iy_i}(f)_{mk}\right)$$

Step 5: Calculate *N* WDSCs of x and y_i from $i = 1$ to $i = N$.

Step 6: Construct the LR-WDSC using the LR technique shown in Equation (4-8) and Equation (4-9) to process N WDSCs obtained in Step 5.

For illustration purposes, an example for constructing the LR-WDSC is given in Figure 4.1. In Algorithm 4-1, Step 3 guarantees the expression of the distance between two cyclic power-spectra at all α, which can thus reveal the differences between x and y_i at every cyclic frequency. Then the accumulation in Step 4 enables the concentration of the differences for all cyclic frequencies. In SC, the cyclic frequency is transformed from time t and indicates the modulation. Therefore, the LR-WDSC is an indicator to reveal the deviations of all modulations, which are one of the essential phenomena of bearing failure since the occurrence and progression of bearing failure are always accompanied by the growth of modulation. The integration of LR further

increases the monotonicity of the proposed HI, with its advantages further being explored in Chapter 4.4.

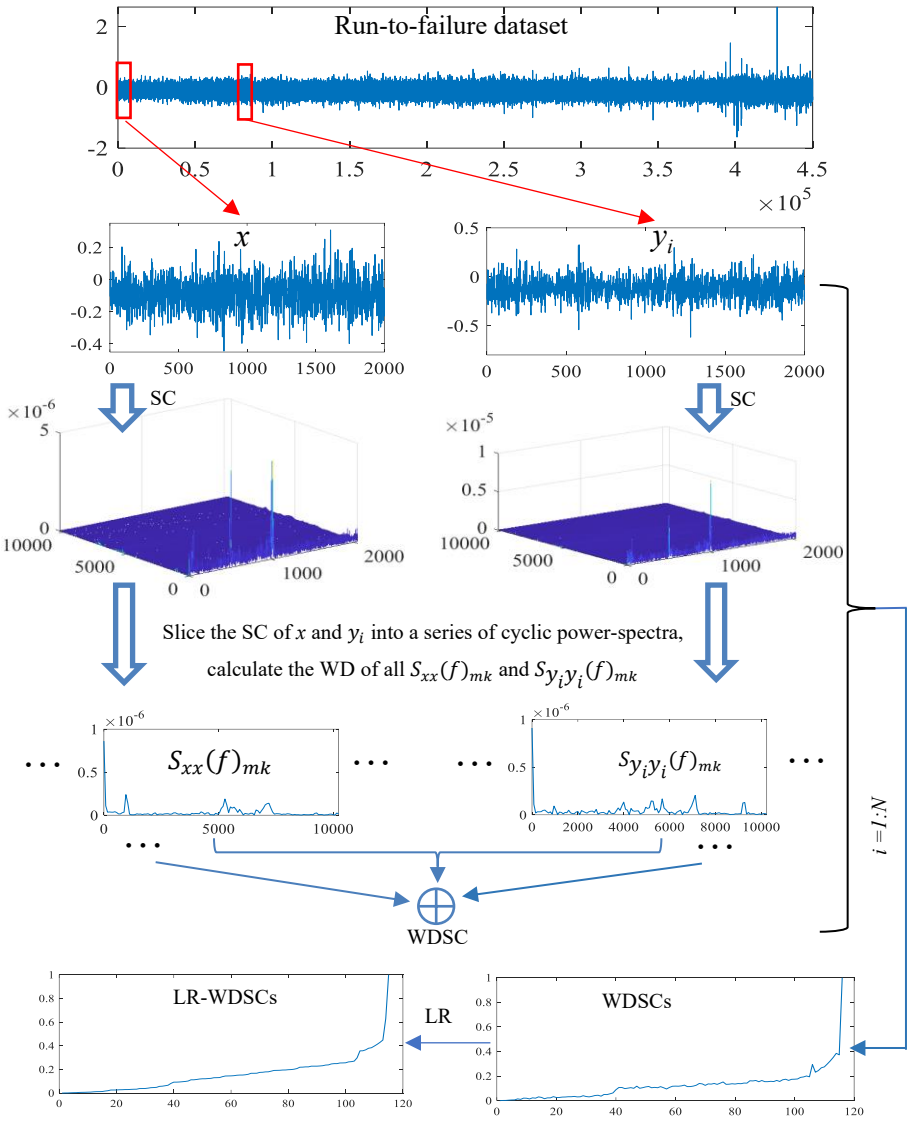


Figure 4.1. Illustration of how to calculate the LR-WDSC

4.3. The proposed prognostic scheme

After the construction of HI, the bearing life span needs to be divided into two or more stages. In this chapter, the bearing life is divided into healthy and unhealthy stages, using an adaptive 3σ criterion-based technique [147]. The division procedure can be summarized as follows: for the given HI series (denoted as $HI_1, HI_2, HI_3, \ldots, HI_N$),

compute the mean value μ_j and standard deviation σ_j of $\{HI_1, HI_2, \dots, HI_j\}$ from $j = 5$. The first point at the unhealthy stage is detected when the following condition is satisfied for the first time

$$\left|HI_{j+i} - \mu_j\right| > 3\sigma_j \tag{4-10}$$

where $i = 1, 2, \dots, I_{max}$. Once the entrance of the unhealthy stage is detected, RUL prediction is activated.

4.3.1. The GRU network

The GRU is an improved variant of RNNs that can remember the previous information and adapt it to calculate the current states. Due to the high capability in solving the gradient vanishing or exploding issues, the GRU has become an efficient and popular method in machine learning field in recent years. The output of the GRU is obtained through two gates, namely the update gate and the reset gate. In this chapter, the gate mechanism of the GRU with optimized hyperparameters is introduced to capture the characteristics of time series and predict the RUL. A typical structure of GRU is depicted in Figure 4.2. As presented in Figure 4.2, the transmission process of the GRU can be concluded as

$$Z_t = \sigma(\omega_{xz} x_t + \omega_{hz} h_{t-1} + b_z) \tag{4-11}$$

where Z_t is the update gate that determines the amount of historical information to be brought into the future state. ω_{xz}, ω_{hz}, and b_z are the corresponding weight matrices and bias vector. x_t denotes the t^{th} input vector, and σ denotes the sigmoid function.

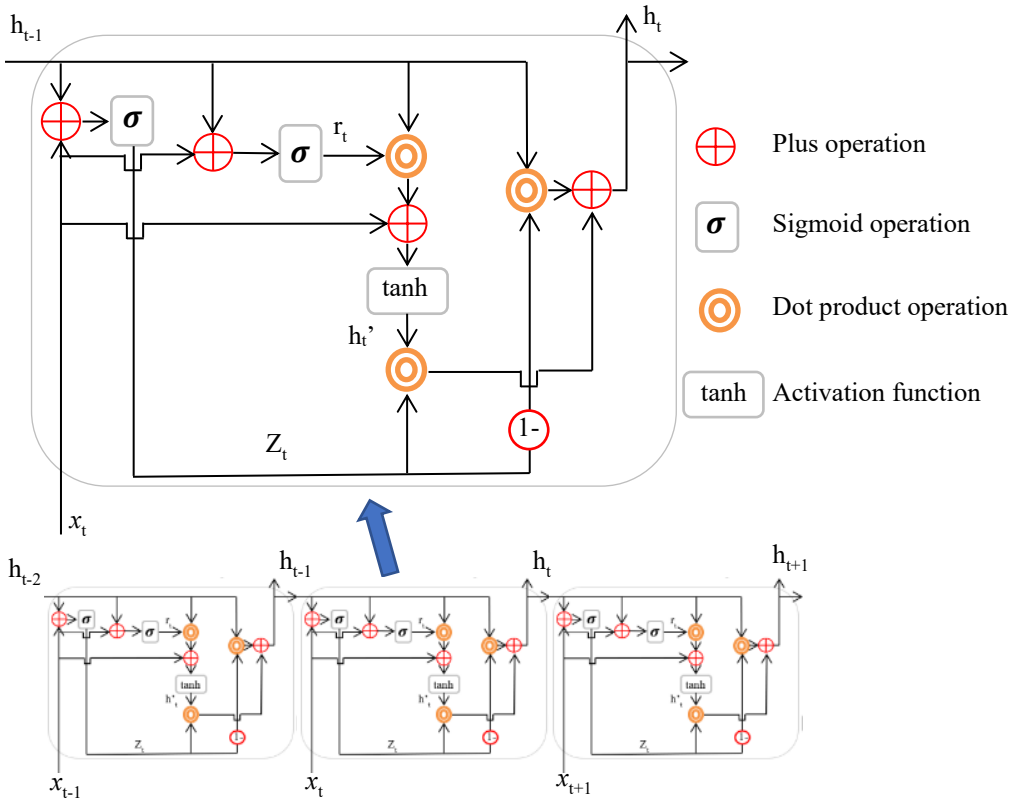


Figure 4.2. The typical structure of GRU network

The reset gate involves the approach about how to integrate the historical information into the current state. Similar to Z_t, the reset gate can be calculated through

$$r_t = \sigma(\omega_{xr} x_t + \omega_{hr} h_{t-1} + b_r) \tag{4-12}$$

where ω_{xr}, ω_{hr}, and b_r are the corresponding weight matrices and bias vector.

The candidate hidden state $h_t{}'$ is based on the reset gate and can be obtained by

$$h_t{}' = \tanh(\omega_{xh'} x_t + \omega_{hh'} (r_t \odot h_{t-1}) + b_{h'}) \tag{4-13}$$

where tanh denotes the hyperbolic tangent function, and $\odot$ represents the dot product operation. ω_{xh}, $\omega_{hh'}$ and $b_{h'}$ are the corresponding weight matrices and bias vector.

The hidden state h_t at the current state can be determined by the update gate Z_t, candidate hidden state $h_t{}'$ and hidden state at $t-1$ by

$$h_t = (1 - Z_t) \odot h_{t-1} + Z_t \odot h_t{}' \tag{4-14}$$

4.3.2. The proposed bearing prognosis algorithm

Despite its merits in time series prediction, the final prediction accuracy of the GRU depends heavily on the selection of hyperparameters. To date, most of the hyperparameters tuning is achieved manually, which brings enormous workload and uncertainty. The artificially selected procedure is highly manual-experience-dependent, time-consuming, and not universal for different cases, which renders low generalization ability of the prediction model. To address this issue, the BO is employed for adaptively hyperparameters tuning.

The BO is a global optimization algorithm to search the extrema of a given objective function $f(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents a vector of hyperparameters to be tuned. Compared with other optimization methods, such as genetic algorithm and particle swarm optimization, the BO adopts the Gaussian process (GP) to update the posterior distribution of the objective function by continuously adding samples until the posterior distribution fits the actual distribution. The updating procedure of the BO can be summarized as follows.

The fundamental assumption of BO is that the objective function is GP prior, i.e., $f(\boldsymbol{\theta}) \sim \boldsymbol{\mathcal{GP}}(\mu(\boldsymbol{\theta}), k(\boldsymbol{\theta}, \boldsymbol{\theta}'))$, where $\mu(\boldsymbol{\theta})$ and $k(\boldsymbol{\theta}, \boldsymbol{\theta}')$ represent the mean and covariance, respectively. For simplification, the prior mean $\mu(\boldsymbol{\theta})$ is assumed to be 0. Denote the iterative observations from 1 to t as $\{\boldsymbol{\theta}_{1:t}, \boldsymbol{y}_{1:t}\}$, where $\boldsymbol{y}_{1:t} = f(\boldsymbol{\theta}_{1:t})$.

The $\boldsymbol{y}_{1:t}$, therefore, follows a multivariate Gaussian distribution $\boldsymbol{y}_{1:t} \sim \boldsymbol{\mathcal{N}}(0, \mathbf{K})$ with $\mathbf{K}$ being expressed as

$$\mathbf{K} = \begin{bmatrix} k(\boldsymbol{\theta}_1, \boldsymbol{\theta}_1) & \dots & k(\boldsymbol{\theta}_1, \boldsymbol{\theta}_t) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{\theta}_t, \boldsymbol{\theta}_1) & \dots & k(\boldsymbol{\theta}_t, \boldsymbol{\theta}_t) \end{bmatrix} \tag{4-15}$$

The $\mathbf{K}$ is called the kernel matrix. Denote the next evaluation point as $\boldsymbol{\theta}_{t+1}$ and the corresponding $\boldsymbol{y}$ value as $\boldsymbol{y}_{t+1}$. For Gaussian prior, $\boldsymbol{y}_{1:t}$ and $\boldsymbol{y}_{t+1}$ are jointly Gaussian that can be expressed as

$$\begin{vmatrix} \boldsymbol{y}_{1:t} \\ \boldsymbol{y}_{t+1} \end{vmatrix} \sim \boldsymbol{\mathcal{N}} \left(0, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^{\boldsymbol{T}} & k(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_{t+1}) \end{bmatrix} \right) \tag{4-16}$$

where

$$\mathbf{k} = [k(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_1) \;\; k(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_2) \;\; \dots \;\; k(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_t)] \tag{4-17}$$

And the posterior predictive distribution at $t+1$ can be accessed through

$$P\left(\boldsymbol{y}_{t+1} \middle| \boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_{1:t}, \boldsymbol{y}_{1:t} \right) \sim \boldsymbol{\mathcal{N}}(\mu(\boldsymbol{\theta}_{t+1}), \sigma^2(\boldsymbol{\theta}_{t+1})) \tag{4-18}$$

where

$$\mu(\boldsymbol{\theta}_{t+1}) = \mathbf{k}^{\boldsymbol{T}} \mathbf{K}^{-\mathbf{1}} \boldsymbol{y}_{1:t} \tag{4-19}$$

$$\sigma^2(\boldsymbol{\theta}_{t+1}) = k(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_{t+1}) - \mathbf{k}^{\boldsymbol{T}} \mathbf{K}^{-\mathbf{1}} \mathbf{k} \tag{4-20}$$

According to Equation (4-15) - Equation (4-20), the posterior distribution $\boldsymbol{y}_{t+1}$ can be updated by the $\mu(\boldsymbol{\theta}_{t+1})$ and $\sigma^2(\boldsymbol{\theta}_{t+1})$, which is merely determined by the selection of covariance function $k(\boldsymbol{\theta}, \boldsymbol{\theta}')$.

By integrating the LR-WDSC and BO-GRU network, this chapter proposed a systematic framework for bearing prognosis. The calculating procedure is demonstrated in Algorithm 4-2.

Algorithm 4-2: A data-driven prognostic scheme for bearings

Input: The bearing run-to-failure dataset $\{y_i\}_{i=1}^{N}$ collected at N moments.

Output: The predicted RUL.

Initialize: Select one bearing healthy signal as reference x.

Procedure:

Step 1: Calculate the WDSC between x and y_i from $i = 1$ to $i = N$. The results are denoted as $\{WDSC_i\}_{i=1}^{N}$.

Step 2: Utilize the 3σ criterion presented in Equation (4-10) to divide the $\{WDSC_i\}_{i=1}^{N}$ into healthy and unhealthy stages.

Step 3: Calculate the LR-WDSC using the LR technique to process the WDSC at the unhealthy stage. The results are denoted as $\{LR - WDSC_j\}_{j=m}^{N}$, where m is the first point of the unhealthy stage.

Step 4: Set the training error as the objective function and employ the BO to determine the essential hyperparameters of the GRU network adaptively.

Step 5: Use the $\{LR - WDSC_j\}_{j=m}^{N}$ and the corresponding RUL to train the BO-GUR network and predict the RUL.

4.4. Experimental validation

4.4.1. The experimental setup

The bearing run-to-failure datasets used in this chapter are Xi'an Jiaotong University and Changxing Sumyoung Technology Co., Ltd. (XJTU-SY) bearing datasets [148]. Fifteen bearings' run-to-failure experiments were conducted using the accelerated degradation test rig shown in Figure 4.3. Three operating conditions, as shown in Table 4-1, were set in these experiments, and there are five experiments in each condition. The data is collected by the VibraQuest package developed by

SpectraQuest, Inc. To collect the data with complete degradation processes, all the experiments were started from the healthy condition and stopped when the maximum amplitude of the vibration exceeding ten times the maximum amplitude of the healthy condition. Two accelerometers were installed horizontally and vertically on the bearing house to collect the vibrations. These vibration signals were collected every 60-seconds with a sampling frequency of 25.6 kHz and a sampling time of 1.28 seconds. At the end of these experiments, bearing failures, including the inner race, outer race, ball, cage, and compound failures, were recorded. Some of the tested bearings at the final stage are shown in Figure 4.4. More details can refer to [148]. In addition, the experimental verification is conducted on a server with Intel (R) Xeon (R) Gold 6238R CPU @ 2.20GHz. MATLAB 2021a is used as the programming tool to code the proposed algorithms.

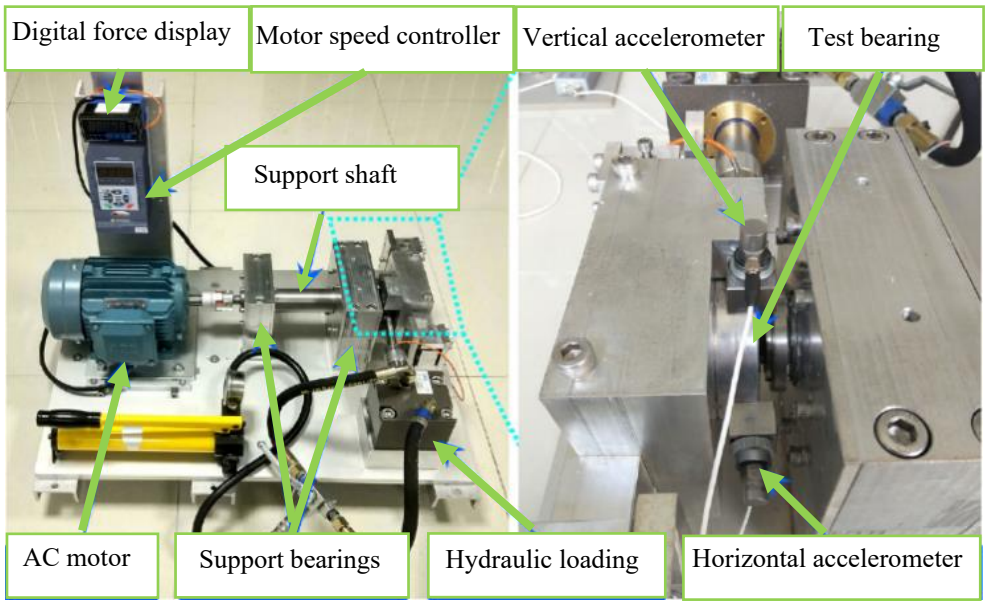


Figure 4.3. The XJTU-SY accelerated degradation test rig [148]

(a) (b)

Figure 4.4. The failure bearings: (a) Inner race wear (b) Outer race fracture [148]

Table 4-1 Operating conditions of the test bearings

Operating conditions	Radial force	Rotating speed	Bearing dataset
Condition 1	12 kN	35 Hz	Bearing 1-1 to Bearing1-5
Condition 2	11 kN	37.5 Hz	Bearing 2-1 to Bearing2-5
Condition 3	10 kN	40 Hz	Bearing 3-1 to Bearing3-5

4.4.2. LR-WDSC extraction and its monotonicity

The LR-WDSC needs to be extracted from the run-to-failure data before implementing prediction. It should be noted that fourteen run-to-failure data are used in this chapter, although fifteen data were recorded. Dataset Bearing 1-4 is not considered because it exceeds the failure threshold once it enters the unhealthy stage. For the other data sets, the first bearing vibration of each dataset is chosen as the reference. Then the LR-WDSC of the corresponding dataset is calculated following the framework of Algorithm 1. Subsequently, the LR-WDSC is normalized from 0 to 1 using the normalized scheme shown in Equation (4-21). For illustration, the normalized LR-WDSC and the RMS from three different operating conditions are displayed in Figure 4.5.

$$HI_j = [HI_j - \min(HI)] / [\max(HI) - \min(HI)] \tag{4-21}$$

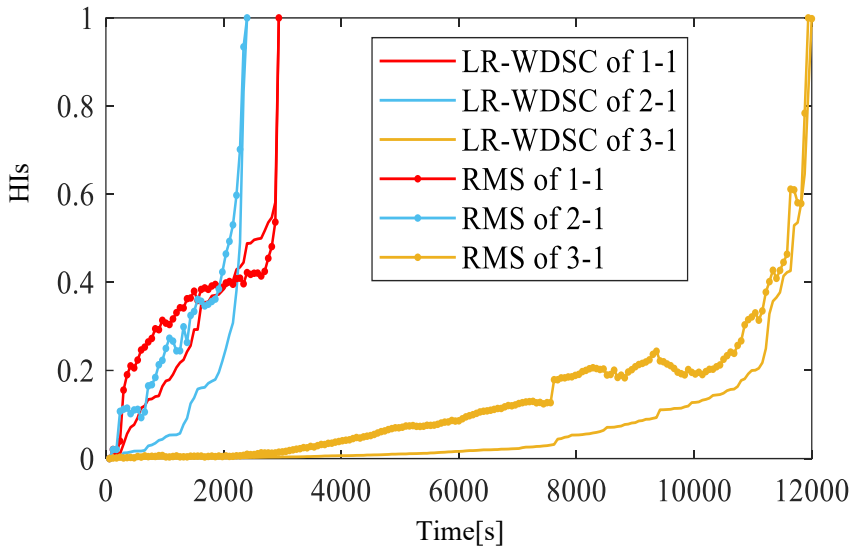


Figure 4.5. The LR-WDSC and RMS at three different operating conditions (same colored curves indicate the same operating condition)

From Figure 4.5, it can be found that the proposed LR-WDSC shows high monotonicity, which is the inherent property of the HI and can be calculated through

$$Mon(HI) = \frac{1}{L-1} |\text{no.}PD - \text{no.}ND| \tag{4-22}$$

where the L is the length of HI, $\text{no.}PD$ and $\text{no.}ND$ represent the numbers of positive and negative differences, in which the difference equals to $HI_{l+1} - HI_l$. The monotonicity ranges from 0 to 1, and a higher value indicates a better performance of the HI. To validate the superiority of LR-WDSC quantitatively, the monotonicity of the LR-WDSC and RMS is calculated, with results being presented in Figure 4.6. In addition, the monotonicity of the kurtosis, ICS2 [119], and entropy of the test bearings are also plotted. The three HIs mentioned above are also selected since they are representative HIs of impulsiveness, cyclostationarity, and disorder, which are typical symptoms of bearing failure [104]. From Figure 4.6, it can be found that the LR-

WDSC shows the highest monotonicity in all experiments. The monotonicity of HI is of great importance to the RUL prediction since the fluctuations of HI are not bearing degradation's characteristics and can further affect the prediction model. Therefore, the proposed LR-WDSC is expected to improve the prediction accuracy, and details are shown in the next section.

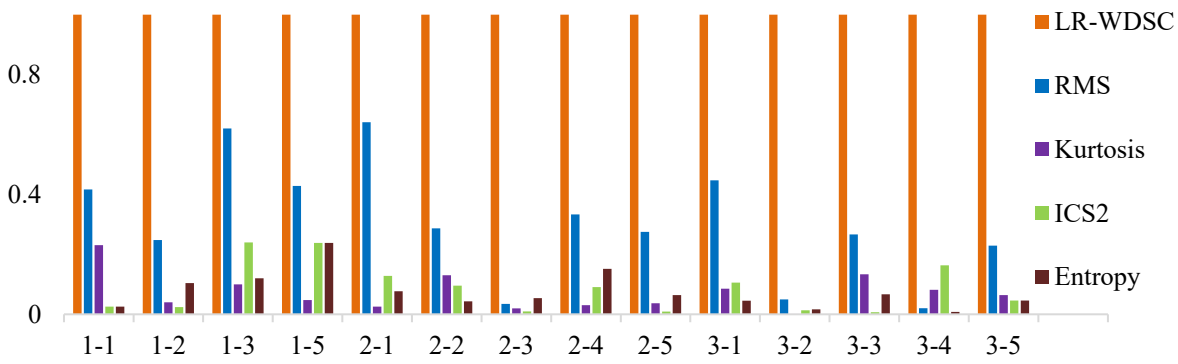


Figure 4.6. The monotonicity of different HIs

4.4.3. Ablation studies

The superiority of the proposed method is demonstrated through ablation studies in this section. The proposed bearing RUL prediction algorithm comprises three key components: the LR-WDSC, the BO, and the GRU. Ablation studies are conducted to verify the function of each element.

To quantitatively evaluate the performance of RUL prediction, the root mean square error (RMSE) and mean absolute error (MAE) obtained through Equation (4-23) and Equation (4-24) between the predicted RUL and calculated RUL are adopted in these ablation studies.

$$\text{RMSE} = \sqrt{\sum_{j=1}^{L} \left(RUL_j - \widehat{RUL_j}\right)^2 / L} \tag{4-23}$$

$$\text{MAE} = \sum_{j=1}^{L} \left|RUL_j - \widehat{RUL_j}\right| / L \tag{4-24}$$

4.4.3.1. Ablation study of different HIs

The effectiveness of the LR-WDSC is validated in this sub-section. In the bearing RUL prediction algorithm, the HI is indispensable. Thus, in this ablation study, the LR-WDSC is replaced by other typical HIs rather than being deleted directly.

The RUL prediction is executed using the proposed prognostic scheme. The leave-one-out strategy is adopted for predicting RUL. More specifically, one bearing run-to-failure dataset is used as the test set, while the other 13 bearing run-to-failure datasets are utilized for training the GRU network. The LR-WDSC is employed as the feature, and the corresponding RUL calculated through Equation (4-25) is adopted as the label in all datasets.

$$RUL_j = (L - j)/(L - 1) \tag{4-25}$$

where j represents the j^{th} vibration data at the unhealthy stage, and the L is the total number of vibrations at the unhealthy stage. Hyperparameters, including the number of hidden units and initial learning rate, are adaptively acquired using the BO technique, in which the objective function is set to be the RMSE of the training process. It should be noted that only one layer is used here because the data size is not massive, and one layer has achieved high accuracy. The layers can also be optimized

using the BO when necessary. For a better understanding, a block diagram is displayed in Figure 4.7.

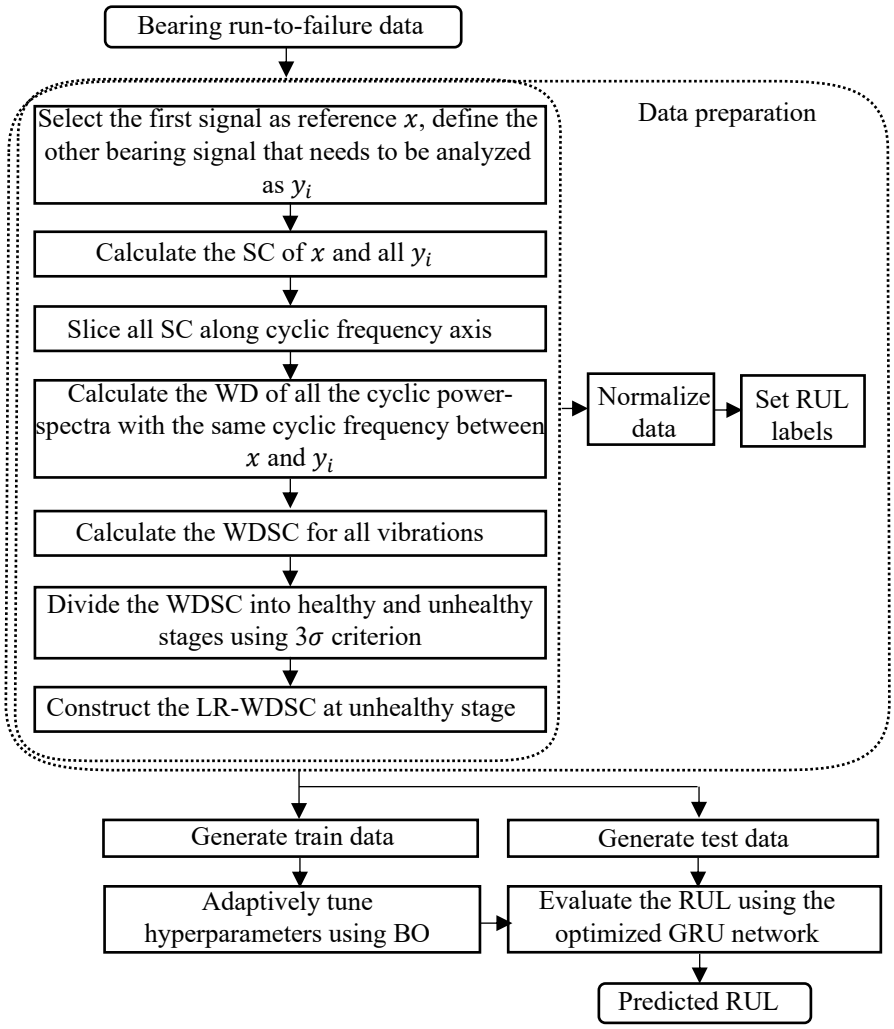


Figure 4.7. The diagram for XJTU-SY bearing RUL prediction

To quantify the performance of the proposed method, the MAE and RMSE are calculated for each run-to-failure bearing dataset and are summarized in Table 4-2. In addition, three detailed prediction results of the proposed method under different operating conditions are displayed in Figure 4.8.

To validate the superiority of the proposed HI, the RMS, kurtosis, ICS2, and entropy are employed to replace the LR-WDSC, remaining other processing procedures unchanged. Note that the first vibrations at the unhealthy stage of kurtosis, ICS2, and

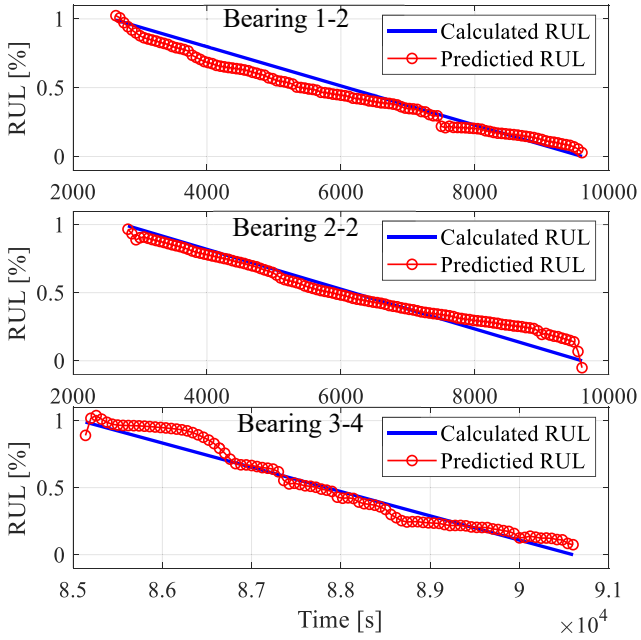


Figure 4.8. The prediction results of the proposed prognostic method

entropy utilize the same values as the first vibrations of RMS since significant fluctuations are observed in some experiments of kurtosis, ICS2, and entropy, invaliding the 3σ-based dividing technique. The MAE and RMSE of the RMS-based, kurtosis-based, ICS2-based, and entropy-based prognostic methods are listed in Table 4-2. The detailed prediction results of the same run-to-failure bearing datasets of RMS are selected as the representative, being plotted in Figure 4.9. It can be concluded from Table 4-2, Figure 4.8, and Figure 4.9 that the proposed LR-WDSC-based method performs better than the RMS-based, kurtosis-based, ICS2-based, and entropy-based prognostic methods in all experiments, which shows the lower values in both MAE and RMSE. The LR-WDSC, which can thoroughly reveal the probability distribution variations over deterioration without losing any information in modulation, thus contributes to achieving higher prediction accuracy than the RMS, kurtosis, ICS2, and entropy in all the experiments.

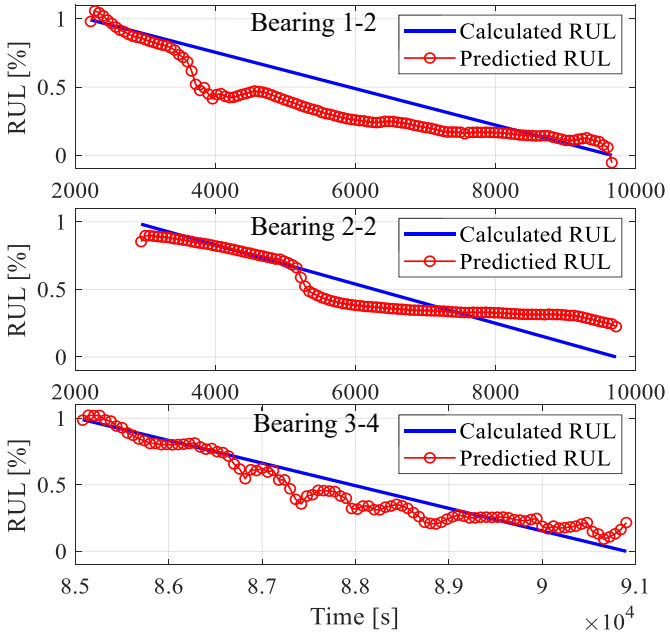


Figure 4.9. The prediction results of the RMS-based prognostic method

Table 4-2 Prognostic performance of different HIs for XJTU-SY run-to-failure datasets

	LR-WDSC		**RMS**		**Kurtosis**		**ICS2**		**Entropy**	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1-1	0.081	0.105	0.167	0.205	0.127	0.150	0.188	0.234	0.111	0.140
1-2	0.058	0.066	0.135	0.164	0.132	0.162	0.185	0.240	0.152	0.183
1-3	0.084	0.098	0.172	0.192	0.178	0.205	0.139	0.185	0.200	0.231
1-5	0.165	0.198	0.280	0.313	0.294	0.352	0.175	0.226	0.236	0.264
2-1	0.146	0.165	0.233	0.270	0.243	0.308	0.234	0.262	0.212	0.263
2-2	0.056	0.076	0.098	0.130	0.165	0.190	0.126	0.157	0.160	0.190
2-3	0.056	0.072	0.073	0.086	0.144	0.171	0.212	0.238	0.258	0.295
2-4	0.073	0.084	0.240	0.274	0.236	0.267	0.240	0.271	0.258	0.294
2-5	0.075	0.095	0.155	0.175	0.179	0.201	0.256	0.317	0.255	0.294
3-1	0.147	0.171	0.225	0.258	0.212	0.293	0.246	0.275	0.153	0.176
3-2	0.200	0.248	0.295	0.340	0.281	0.320	0.346	0.408	0.298	0.355
3-3	0.161	0.189	0.194	0.239	0.193	0.231	0.220	0.282	0.229	0.261
3-4	0.058	0.076	0.073	0.091	0.172	0.228	0.090	0.111	0.173	0.180
3-5	0.116	0.136	0.253	0.296	0.171	0.210	0.160	0.185	0.259	0.296
Average	0.105	0.127	0.185	0.217	0.195	0.235	0.201	0.242	0.211	0.244

4.4.3.2. Ablation study of BO

To verify the effectiveness of the BO, the BO is deleted from the proposed bearing RUL prediction algorithm in this sub-section. For the BO-deleted method (denoted as M1), the number of hidden units and the initial learning rate are manually set to be 32 and 0.05 for all the datasets since most of the selected hyperparameters are around these two values in the proposed algorithm. The RUL prediction results of M1 are calculated, with MAE and RMSE being listed in Table 4-3. In addition, the computation time of the BO-deleted method and the proposed method (denoted as M) is also recorded in Table 4-3.

Table 4-3 Prognostic performance with and without BO for XJTU-SY run-to-failure datasets

	M1			**M**
	MAE	RMSE	Time (s)	Time(s)
1-1	0.182	0.236	4.034	318.6
1-2	0.081	0.098	3.925	331.1
1-3	0.217	0.228	3.962	345.0
1-5	0.243	0.279	4.087	319.9
2-1	0.236	0.267	4.019	353.9
2-2	0.087	0.142	4.004	363.0
2-3	0.409	0.479	4.011	331.9
2-4	0.213	0.294	3.977	302.8
2-5	0.161	0.174	4.221	341.2
3-1	0.371	0.438	3.962	304.8
3-2	0.246	0.266	2.288	171.8
3-3	0.231	0.269	3.973	347.8
3-4	0.138	0.176	3.955	363.0
3-5	0.150	0.183	3.955	301.2

From the prediction accuracy perspective, it can be found that: a) for M1, a few datasets achieve low prediction error, such as dataset 1-2 and dataset 2-2, but some datasets present high prediction error, such as dataset 2-3 and dataset 3-1; b) overall, the M1 shows higher prediction error than the proposed method in all the datasets.

Therefore, a) the hyperparameters that are effective in one condition may not work for another, and it would be better to fine-tune the hyperparameters in different conditions; b) BO does contribute to improving prediction accuracy as it can adaptively tune the rational hyperparameters for different cases.

As for the computation time, it can be found that the time for executing the RUL prediction of one dataset using M1 and M is around 4s and 300s, respectively. More computation time is required for the proposed method from this perspective. However, only one iteration is implemented in M1. In contrast, sixty iterations are executed to search the rational hyperparameters in M. And it should be noted that the recorded time of M1 is the final execution time, ignoring the time of manual hyperparameter tunning at the preparation stage. In practice, the manual hyperparameter tuning is based on the researchers' experience and needs numerous trials and errors. In addition, different cases usually need to re-tune the hyperparameters. It is believed this process usually takes far more than 300s. On the contrary, the recorded time of the proposed method is the total time from the preparation to the final stage, without human effort on trials and errors for one dataset and re-tunning for other datasets. From this perspective, the BO contributes to the human-labor-saving and experience-reliance-less of the proposed method.

4.4.3.3. Ablation study of different deep learning networks

The effectiveness of the prediction network is validated in this sub-section. The prediction model is indispensable for the bearing RUL prediction, thus the GRU is replaced by other prediction models rather than deleted directly. Typical learning methods, including CNN, long short-term memory (LSTM), and deep belief networks

Table 4-4 Prognostic performance of different deep learning networks for XJTU-SY run-to-failure datasets

	M2			M3			M4		
	MAE	RMSE	Time(s)	MAE	RMSE	Time(s)	MAE	RMSE	Time(s)
1-1	0.183	0.216	5018	0.146	0.196	1627	0.200	0.237	3303
1-2	0.138	0.159	4929	0.091	0.109	1739	0.140	0.160	3242
1-3	0.057	0.072	4914	0.099	0.131	1501	0.067	0.084	3070
1-5	0.050	0.060	5233	0.224	0.244	2949	0.089	0.101	2072
2-1	0.083	0.100	5347	0.163	0.184	1765	0.109	0.133	3668
2-2	0.091	0.106	4804	0.070	0.093	2136	0.084	0.113	4276
2-3	0.092	0.109	4359	0.074	0.092	1339	0.092	0.109	2821
2-4	0.194	0.233	5053	0.137	0.149	1581	0.192	0.222	4890
2-5	0.080	0.105	4591	0.102	0.135	1751	0.098	0.110	2735
3-1	0.196	0.218	4596	0.201	0.229	1596	0.181	0.202	3692
3-2	0.106	0.117	2695	0.202	0.250	738.4	0.119	0.142	1171
3-3	0.257	0.295	5119	0.124	0.151	1652	0.256	0.292	2488
3-4	0.087	0.107	4780	0.079	0.107	1441	0.097	0.116	3729
3-5	0.195	0.231	4867	0.115	0.140	1649	0.207	0.244	1999

(DBN), are selected for comparison. The GRU is replaced by CNN, LSTM, and DBN with five layers, while other procedures remain unchanged. The prediction results and the computation time are listed in Table 4-4. From Table 4-4, it can be found that: a) the CNN-based method (denoted as M2) achieves lower prediction error than M at datasets 1-3, 1-5, 2-1, and 3-2; b) the LSTM-based method (denoted as M3) shows lower prediction error than M at dataset 3-3 and similar prediction error at datasets 3-2, 3-5; c) the DBN-based method (denoted as M4) presents lower prediction error than the proposed method in datasets 1-3, 1-5, 2-1, and 3-2; d) the computation time for M, M2, M3, and M4 is around 300s, 5000s, 1600s, and 3000s, respectively. It can be concluded that the CNN-based, LSTM-based, and DBN-based methods can realize the lower prediction error in some datasets, but the proposed method shows better performance in more datasets, which indicates the proposed method is more reliable. In addition, less computation time is required by the proposed method. Therefore,

utilizing the GRU architecture contributes to the proposed method being superior with high prediction accuracy and less computation time.

4.4.4. Comparisons with different state-of-the-art RUL prediction algorithms

To comprehensively demonstrate the superiority of the proposed method, four state-of-the-art RUL estimation methods, i.e., the Bidirectional LSTM (BLSTM)-based method [149], Multiscale Convolutional Neural Network (MSCNN)-based method [150], Gated Convolutional Unit Transformer (GCU-Transformer) [151], and deep multi-scale window-based transformer (DMW-Transformer) [152] are used for comparison. Table 4-5 from [152] and [153] summarizes the prognostic performance

Table 4-5 Prognostic performance of different methods for XJTU-SY run-to-failure datasets [152] [153]

	BLSTM [149]		**MSCNN [150]**		**GCU-Transformer [151]**		**DMW-Transformer [152]**	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1-1	0.191	0.228	0.213	0.242	0.169	0.201	0.084	0.113
1-2	0.231	0.305	0.229	0.262	0.142	0.168	0.128	0.154
1-3	0.106	0.130	0.155	0.184	0.165	0.241	0.103	0.131
1-5	0.314	0.362	0.181	0.215	0.193	0.213	0.123	0.157
2-1	0.129	0.152	0.126	0.148	0.303	0.356	0.138	0.195
2-2	0.094	0.134	0.194	0.232	0.116	0.141	0.088	0.120
2-3	0.170	0.216	0.164	0.199	0.160	0.197	0.117	0.176
2-4	0.267	0.311	0.195	0.231	0.125	0.161	0.064	0.101
2-5	0.278	0.308	0.090	0.108	0.110	0.149	0.189	0.233
3-1	0.297	0.351	0.214	0.247	0.148	0.169	0.139	0.168
3-3	0.162	0.188	0.156	0.191	0.203	0.260	0.065	0.083
3-4	0.135	0.175	0.139	0.165	0.116	0.133	0.168	0.235
3-5	0.251	0.305	0.225	0.267	0.122	0.140	0.119	0.153
Average	0.202	0.243	0.175	0.207	0.159	0.195	0.117	0.155

of these four methods for the XJTU-SY run-to-failure bearing datasets. It can be observed from the comparisons between Table 4-2 and Table 4-5 that the lowest MAE and RMSE can be obtained by the proposed method in most operating conditions and fault locations. Eight datasets' MAE and RMSE are less than or close to 0.1 when using the proposed method, and most of the remaining datasets represent MAE and RMSE being around 0.15. For BLSTM-based method, it shows lower MAE than the proposed method in dataset 2-1 and lower RMSE in datasets 2-1, 3-3. For MSCNN-based method, it shows lower MAE than the proposed method in datasets 2-1, 3-3 and lower RMSE in dataset 2-1. For GCU-Transformer-based method, it shows lower RMSE than the proposed method in dataset 3-1. For DWM-Transformer-based method, it shows lower MAE than the proposed method in datasets 1-5, 2-1, 2-4, 3-1, 3-3 and lower RMSE in datasets 1-5, 3-1, 3-3. However, in the rest of the datasets, the proposed method shows lower MAE and RMSE, and the overall average MAE and RMSE are the lowest, demonstrating that the predicted RUL of the proposed method is closer to the actual RUL. Therefore, the proposed algorithm is a superior and reliable method for bearing prognostics.

4.5. Conclusion

The novelty and contribution of this chapter can be summarised as follows.

1) The SC of the bearing run-to-failure data is adopted to reveal comprehensive information on cyclostationarity. The bearing fault signatures can be easily separated from the interferences with SC, thus rendering the possibility to reveal bearing fault information. Then the WD between the cyclic power-spectra of healthy reference and

run-to-failure data is calculated and accumulated, thoroughly reflecting the changes in the probability distribution of all cyclic power-spectra.

2) A LR tool is then proposed to remove the spurious fluctuations. Based on the SC, WD, and LR, a novel HI named LR-WDSC is proposed, providing a measure with high monotonicity and can capture the probability distribution variations at all cyclic frequencies over deterioration without losing any information in modulation, even for the weak information buried by the internal or external interferences.

3) An integrated BO-GRU network is established to predict the RUL accurately. A systematic bearing prognostic scheme from HI construction to RUL prediction has been developed to promote the capability of accurate and robust RUL prediction. Results demonstrate that the proposed method is a promising prognostic method with high prediction accuracy, which renders the possibility of the proposed method to be applied in industrial applications.

Chapter 5. A novel distribution-similarity-based multi-scale fused health indicator for bearing prognostics

5.1. Introduction

This chapter is a modified version of one paper titled “Data-driven bearing health management using a novel multi-scale fused feature and gated recurrent unit”, which is being peer-reviewed. The work completed in this chapter is dedicated to realizing objectives 3 and 4 of this research.

Although also devoted to constructing a proper HI to deal with the entangled fault-irrelevant vibrations and spurious fluctuations, this chapter is distinguished from Chapter 4 [154] as it is derived from another theoretical basis, namely the features in the selected frequency band. Rather than constructing HI based on SC, extracting features in the selected frequencies/frequency bands is another effective and commonly employed approach in bearing condition monitoring to capture the failure information from massive noise. In Ref. [155], the power value in one chosen sensitive frequency band was utilized to track the bearing deterioration. The SK and its derived version, including the skewness-SK, mean-SK, std-SK, peak to peak-SK, area under curve-SK, were utilized for high-speed wind turbine bearing severity tracking in Ref. [156]. The information entropy at specific frequencies was employed as HI in Ref. [157], through which the amplitude changes of bearings were captured. Dybata [158] proposed two novel HIs using the local amplitudes in the spectrum and power spectrum filtered by bands around the characteristic frequencies to monitor bearing

degradation. These HIs, extracted from one or several frequencies/frequency bands, can reflect the bearing degradation characteristics to some extent. However, it was found that vibration spectra present different forms at different stages [159]. The information extracted in different frequency zones is associated with each other, showing similarities and differences simultaneously. Therefore, extracting the features in one or limited frequency regions inevitably loses information. To comprehensively reveal the degradation characteristics in different frequency ranges and scales, this chapter develops a library of multi-scale frequency distribution-similarity-based features under the framework of 1/3-binary tree filter structure and WD. Moreover, the high dimensional features in the library are fused by the multiobjective grasshopper optimization algorithm (MOGOA) to simultaneously optimize the monotonicity, robustness, and trendability, thus the fluctuations that are irrelative to bearing failure progression can be mitigated to a great extent. In addition, a GRU network integrated with the GOA is established for RUL prediction, through which some critical parameters are adaptively determined.

The rest of this chapter is organized as follows: Chapter 5.2 provides the theoretical basis and processes to construct the novel composite multi-scale wasserstein distance (CMSWD). Chapter 5.3 includes the establishment of the prognostic framework. Experimental validation is conducted in Chapter 5.4, and the conclusion is presented in Chapter 5.5.

5.2. Construction of the novel HI: CMSWD

This section focuses on developing a proper HI that can comprehensively reveal the underlying failure characteristics of bearing during deterioration with a good

degradation trend. This novel HI is based on the WD, 1/3-binary tree filters, and MOGOA, whose construction procedure is described below.

5.2.1. The WDgram

The related theoretical background of the WD has been studied previously in this thesis, and details can refer to Chapter 4.2.2. The distribution variations over the deterioration can be captured by calculating the WD between the bearing degradation signals and healthy reference. It is, however, on an overall frequency scale. To comprehensively reflect the distribution variations at different scales without losing information, the signals are divided into a series of sub-signals using 1/3-binary tree filter strategy. The theoretical background of the 1/3-binary tree filter has also been studied previously and can refer to Chapter 2.2.1.2. By implementing multiple parallel filtering of the 1/3-binary tree, a series of multi-scale sub-signals $C_{lev,n}$, as shown in Figure 5.1, are obtained. The WD of these multi-scale sub-signals $C_{lev,n}$ between the signal to be analyzed and healthy reference, coined as the multi-scale wasserstein distance (MSWD) in this chapter, can be calculated according to Equation (5-1).

$$MSWD_{lev,n}\left(\mathbb{P}_{A_{lev,n}}, \mathbb{P}_{R_{lev,n}}\right) = \int \left| F_{\mathbb{P}_{A_{lev,n}}}(b) - F_{\mathbb{P}_{R_{lev,n}}}(b) \right| \mathrm{d}b \qquad (5\text{-}1)$$

where $A_{lev,n}$ and $R_{lev,n}$ are the multi-scale sub-signals of the analyzed signal and reference. $\mathbb{P}_{A_{lev,n}}$ and $\mathbb{P}_{R_{lev,n}}$ are the discrete probability distributions of $A_{lev,n}$ and $R_{lev,n}$. $F_{\mathbb{P}_{A_{lev,n}}}$ and $F_{\mathbb{P}_{R_{lev,n}}}$ are the cumulative distribution functions of $A_{lev,n}$ and $R_{lev,n}$.

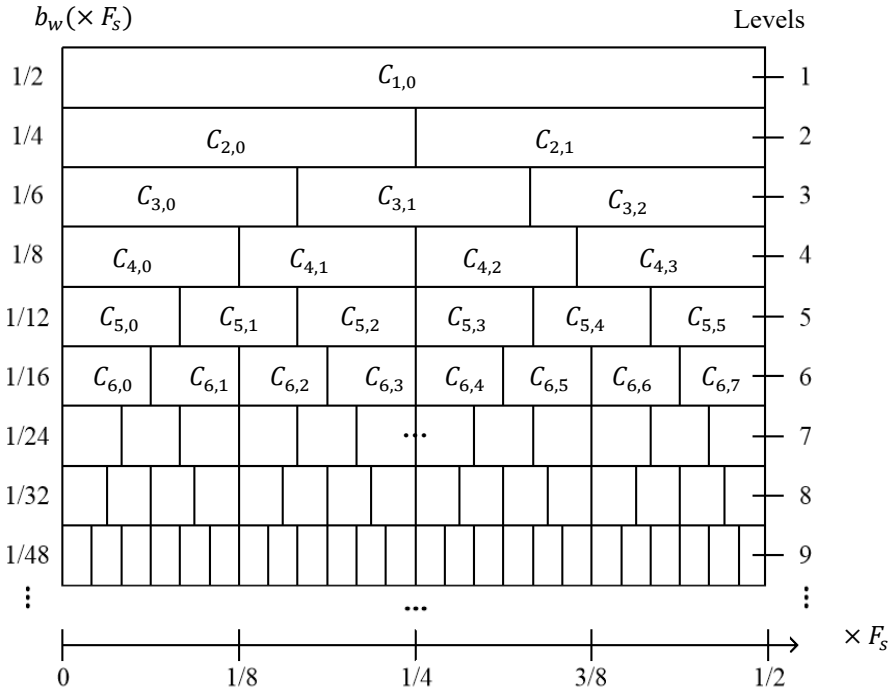


Figure 5.1. Multi-scale sub-signals using 1/3-binary tree

By calculating the MSWD for all the sub-signals, a graph representing the distribution similarity in different frequency ranges and scales, named WDgram, can be generated. The superiority of the construction of WDgram is threefold: 1) The information located in different frequency zones is associated with each other, as shown in Figure 5.2, showing similarities and differences simultaneously. Thus, most traditional HIs cannot reveal a complete picture of bearing failure characteristics. Instead of characterizing the distribution on an overall scale, the WDgram can depict not only the global but also the local properties. The WDgram, therefore, can comprehensively reveal the degradation characteristics at different scales without losing information. 2) Compared with commonly used properties, including energy, impulsiveness, cyclostationarity, and disorder, the distribution similarity at different scales is introduced to characterize the bearing deterioration, enriching the scopes of HI for bearing prognostics. 3) Some MSWDs show better trends than the original one, which

can be used to construct a composite HI with a better degradation trend for accurate bearing RUL prediction.

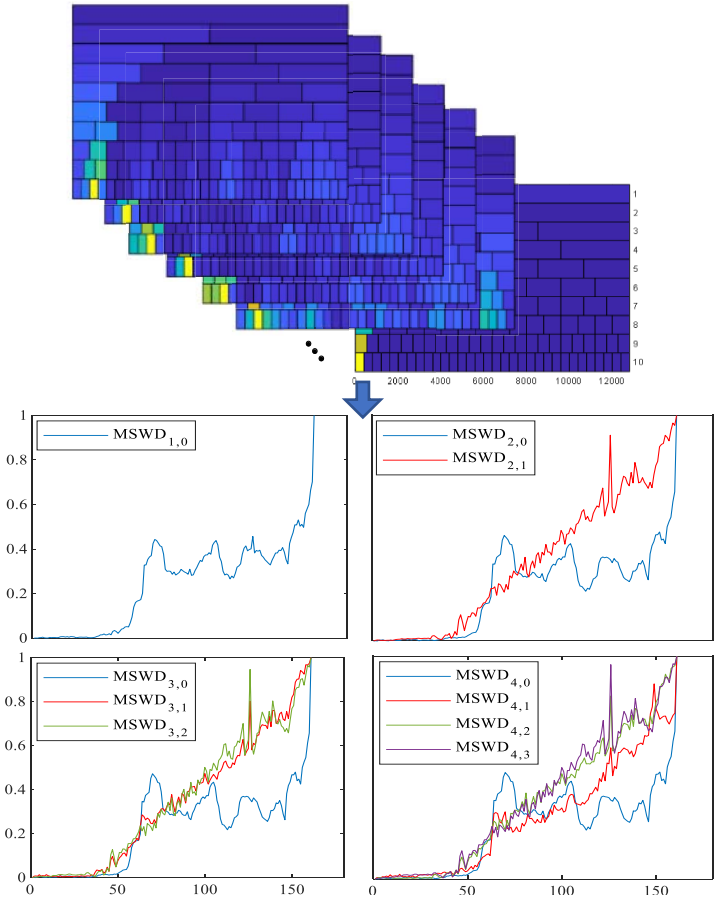


Figure 5.2. A portion of MSWDs extracted from the WDgram

5.2.2. Construction of the novel composite HI

The establishment of the WDgram enables a comprehensive description of the bearing failure characteristics. The extracted MSWDs, however, are high-dimensional HIs. Note that not all the MSWDs equally contribute to higher prediction accuracy and high-dimensional HIs can bring excessive computational burden, resulting in dimensional disaster. Therefore, the optimization algorithm is introduced in this chapter to reduce the dimensions and construct a composite HI with superior

prognostic properties. The feature-level fusion model to build the novel CMSWD is defined as

$$CMSWD = \sum_{i=1}^{K} \omega(i).D_i.MSWD_{lev,n}(i) \quad (5\text{-}2)$$

where K denotes the number of MSWDs to be fused. $D(i) = 1$ or -1 represents the adjustment factor for degradation trend, with 1 representing the increasing trend and -1 the decreasing trend. $\omega(i)$ is the weight factor to be optimized.

Monotonicity, robustness, and trendability, as revealed in [19], are critical metrics for prognostic performance evaluation. Therefore, these metrics are employed as objective functions for optimization. The monotonicity and robustness, denoted as O_{Mon} and O_{Rob}, are the inherent property that can be calculated using the sequence information directly. The monotonicity represents the irreversible property during the deterioration of bearing, i.e., the property that a failed component cannot self-reinstate without manual maintenance and is calculated through Equation (5-3). The robustness measures the property that a HI is not affected by the noise and stochasticity during bearing degradation and is calculated through Equation (5-4). The trendability, denoted as O_{Tre}, is a metric to present the correlation between the HI and time, which is calculated through Equation (5-5).

$$O_{Mon}(CMSWD) = \frac{1}{L-1}|\text{no.}PD - \text{no.}ND| \quad (5\text{-}3)$$

$$O_{Rob}(CMSWD) = \frac{1}{L}\sum_{l=1}^{L} e^{-\left|\frac{CMSWD_l - CMSWD_l{}^T}{CMSWD_l}\right|} \quad (5\text{-}4)$$

$$O_{Tre}(CMSWD,T) = \frac{L(\sum_{l=1}^{L} CMSWD_l t_l) - (\sum_{l=1}^{L} CMSWD_l)(\sum_{l=1}^{L} t_l)}{\sqrt{\left[L\sum_{l=1}^{L} CMSWD_l{}^2 - (\sum_{l=1}^{L} CMSWD_l)^2\right]\left[L\sum_{l=1}^{L} t_l{}^2 - (\sum_{l=1}^{L} t_l)^2\right]}} \quad (5\text{-}5)$$

where $CMSWD = \{CMSWD_l\}_{l=1:L}$ represents the CMSWD sequence of length L, no.PD and no.ND denote the numbers of positive and negative differences, in which the difference equals to $CMSWD_{l+1} - CMSWD_l$. $CMSWD^T$ represents the mean trend value of the $CMSWD$, acquired through the moving average smoothing method in this chapter. t_l is the time value of the l^{th} sample.

To construct a HI with high values of the aforementioned properties, the MOGOA is utilized due to its high accuracy and search efficiency compared with other commonly used multiobjective optimization methods, such as non-dominated sorting genetic algorithm and multi-objective particle swarm optimization [160]. The GOA is the basis of MOGOA, which is a metaheuristic bionic optimization algorithm that simulates the grasshoppers' swarming behavior to search for the optimal [161]. The primary characteristics of the swarming behavior for larval are small steps with slow motion, and for adults are long steps with abrupt movement. Therefore, the bionic principle of GOA is to map the small steps of larvae to local exploitation with slow motion and the long steps of adults to global exploration with rapid movement. To simulate grasshoppers' swarming behavior, the mathematical model shown below can be utilized.

$$P_i = S_i + \hat{T} \quad (5\text{-}6)$$

$$S_i = \sum_{j=1;j\neq i}^{K} s(|P_j - P_i|)\frac{P_j - P_i}{|P_j - P_i|} \tag{5-7}$$

$$s(r) = fe^{\frac{-r}{d}} - e^{-r} \tag{5-8}$$

where P_i is the i^{th} grasshopper's position, $\hat{T}$ represents the target position (best position searched until now), S_i denotes the grasshoppers' social interaction. K is the total number of grasshoppers, $|P_j - P_i|$ represents the distance between the j^{th} and i^{th} grasshoppers, $s(.)$ means the social force. f is the intensity of attraction, d indicates the length scale of attraction. Different combinations of f and d can result in different kinds of interaction for grasshoppers. The suggested f and d are 0.5 and 1.5 in [161], and the same values are adopted in this chapter.

Employing the aforementioned model, however, is found to not converge to a specified point when arriving at the comfort area. To this end, an improved model [161] is established as

$$P_i = c\left(\sum_{j=1;j\neq i}^{K} c\frac{u_b - l_b}{2} s(|P_j - P_i|)\frac{P_j - P_i}{|P_j - P_i|}\right) + \hat{T} \tag{5-9}$$

$$c = c_{max} - g\frac{c_{max} - c_{min}}{G} \tag{5-10}$$

where u_b and l_b are the upper and lower bounds for optimization, respectively. g and G are the values of current iteration and total iteration, respectively. c denotes the decreasing coefficient, which linearly decreases from c_{max} to c_{min}. Note that two c are used in Equation (5-9), the external c and internal c contribute to reducing the search coverage and social interaction over the increasing iteration, respectively.

The MOGOA estimates the optimal solution by incorporating an archive and target selection technique into GOA. The main difference between the GOA and MOGOA is the approach to updating the target. The target is selected in GOA using the best solution acquired until now, while more than one solution can be obtained in MOGOA. To search the target, Pareto optimal solutions are calculated and stored in an archive in MOGOA. Then the roulette wheel is integrated to choose the target from the archive, with the purpose to improve the distribution among all solutions. The probability of selecting the target on the archive is determined by

$$Pro_i = 1/N_i \tag{5-11}$$

where N_i denotes the number of neighboring solutions of the i^{th} solution in the archive.

By estimating the $\omega = \{\omega_i\}_{i=1:K}$ through Equation (5-3) to Equation (5-11) to solve the multiobjective optimization problem defined in Equation (5-12), the CMSWD determined in Equation (5-2) can be constructed. Note that higher values of O_{Mon}, O_{Rob}, and O_{Tre} indicate better prognostic performance, therefore higher values are desired in the optimization, which is equivalent to minimizing $-O_{Mon}$, $-O_{Rob}$, and $-O_{Tre}$ as defined in Equation (5-12).

$$\begin{aligned} &\min_{\omega}\{-O_{Mon}, -O_{Rob}, -O_{Tre}\} \\ &s.t. \quad 0 \le \omega_i \le 1 \end{aligned} \tag{5-12}$$

5.3. The proposed bearing prognostic approach

Following the construction of CMSWD, the bearing life is divided into healthy and unhealthy stages, which adopts an adaptive 3σ criterion-based method with details shown in [147]. The RUL prediction is subsequently activated once the unhealthy stage is detected. The prediction is realized through an integrated GOA and GRU (GOA-GRU) network, with details being shown below.

Theoretical background of GRU has been studied previously, and details can refer to Chapter 4.3.1. Despite the virtues of GRU for time series RUL prediction, its accuracy relies significantly on the hyperparameters setting. To date, the tuning of the hyperparameters for GRU, as well as other machine learning methods, is mainly realized artificially. This manual tuning procedure is usually time-consuming, highly-experience-reliant, and generally not universal for different operating conditions, making the prediction model low generalizability. To this end, an integrated GOA-GRU network, in which some crucial hyperparameters are adaptively tuned by GOA, is established for achieving accurate RUL prediction. Together with the novel HI, a data-driven prognostic approach for the bearing is proposed, with pseudocodes shown in Algorithm 5-1. In addition, the framework of the proposed prognostics system is plotted in Figure 5.3 for illustration purpose.

Algorithm 5-1: A Data-driven Prognostic Approach for Bearings

Input: Run-to-failure series $S_a = \{S_{a_i}\}_{i=1:L}$ with L vibration signals.

Output: Bearings' predicted RUL.

Steps:

1. Define the first signal as healthy reference S_r.
2. **for** i = 1:L **do**
3. Decompose the healthy reference S_r into multi-scale sub-signals $R_{lev,n}$ using a 1/3-binary tree filter strategy
4. Decompose run-to-failure vibration S_{a_i} into multi-scale sub-signals $(A_{lev,n})_i$ using 1/3-binary tree filter strategy.
5. Calculate the $(MSWD_{lev,n})_i$ between all multi-scale sub-signals $(A_{lev,n})_i$ and $R_{lev,n}$ so that generate WDgram based on Equation (5-1).
6. **end for**
7. Obtain the MSWDs for all $S_a = \{S_{a_i}\}_{i=1:L}$, the sequences are denoted as $\{(MSWD_{lev,n})_i\}_{i=1:L}$.
8. Fuse the MSWD at all scales using the MOGOA to optimize the monotonicity, robustness, and trendability simultaneously based on Equation (5 - 2) to Equation (5-12), so that a novel composite CMSWD can be constructed.
9. Divide the $\{CMSWD_i\}_{i=1}^{L}$ into healthy and unhealthy stages based on an adaptive 3σ criterion technique, the HIs at the unhealthy stage $\{CMSWD_i\}_{i=m}^{L}$ are used for prediction, where m is the first vibration at the unhealthy stage.
10. Adaptively tune hyperparameters in GRU by integrating the GOA to optimize the training error.
11. Employ the $\{CMSWD_i\}_{i=m}^{L}$ and corresponding RUL to train the GOA-GRU.
12. Predict the RUL based on the trained GOA-GRU.

5.4. Experimental validation

5.4.1. HI construction results

In this section, the proposed bearing prognostic approach is validated through the bearing run-to-failure datasets provided by XJTU-SY, details can refer to Chapter 4.4.1. To execute the proposed RUL prediction algorithm, the CMSWD needs to be constructed and preprocessed firstly. For the XJTU-SY bearing datasets, analysis of Bearing 1-4 and Bearing 3-2 are not included in this chapter since Bearing 1-4 exceeds the failure threshold immediately when the unhealthy stage is detected, and there are

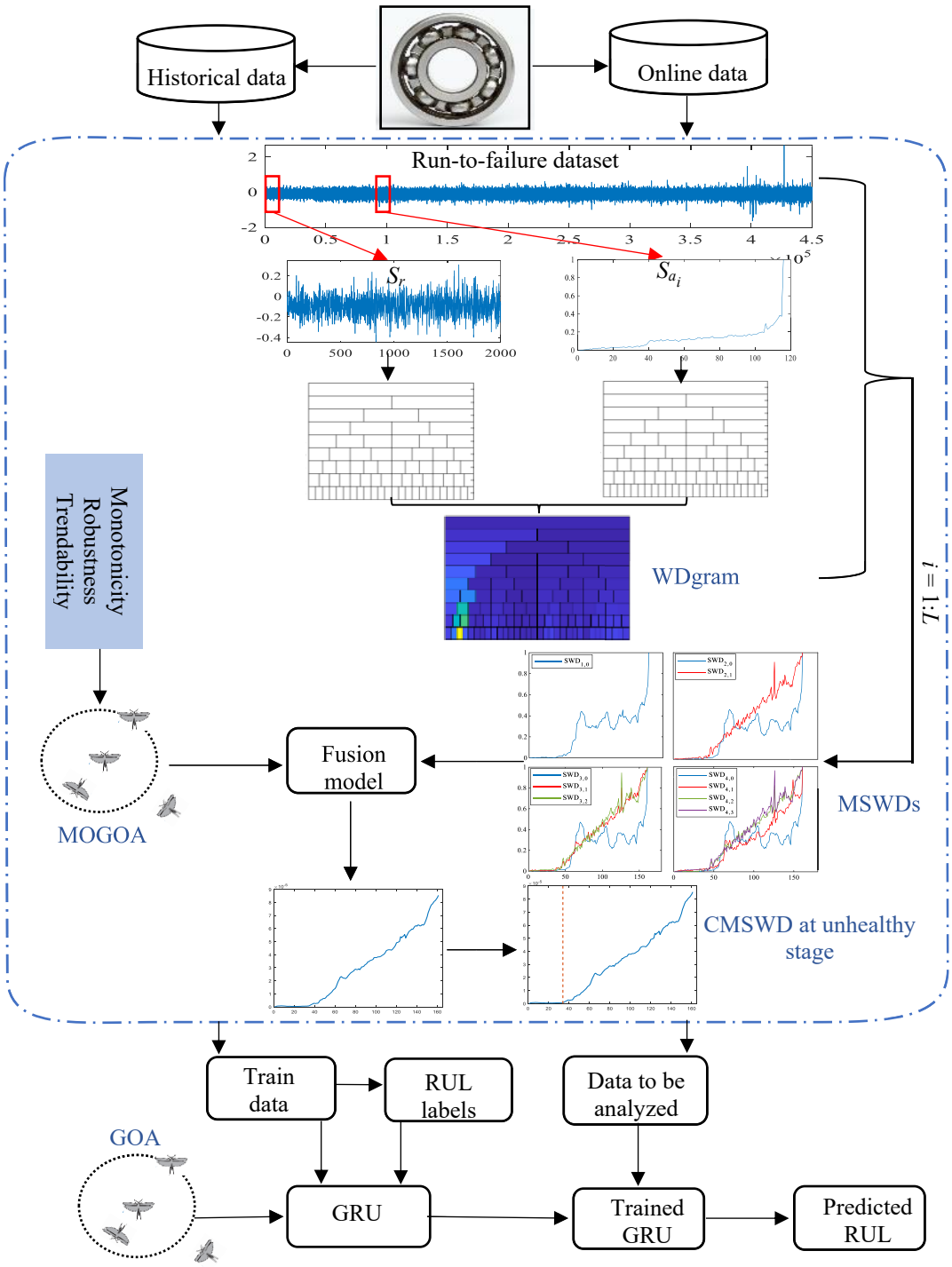


Figure 5.3. The framework of the proposed bearing prognostic approach

significant variations in the collected signals of bearing 3-2 [153]. For the rest thirteen datasets, the CMSWD of each run-to-failure dataset is calculated following step 1 to step 6 in Algorithm 5-1. In the CMSWD calculation framework, the grasshopper number in MOGOA is set to be 500, iteration number and archive size are both set to

be 100. The CMSWD is subsequently normalized using the max-min normalization strategy for prediction. Due to limited space, the normalized CMSWD, its original version (i.e., WD), and a portion of randomly selected MSWDs of one run-to-failure dataset (dataset 2-2) are drawn in Figure 5.4 for illustration purpose. In addition, the monotonicity, robustness, and trendability of the CMSWD and the comparisons with WD are displayed in Figure 5.5. Comparing the values of monotonicity, robustness, and trendability for the proposed CMSWD and the traditional WD shown in Figure 5.5, it can be concluded that the proposed CMSWD demonstrates higher values than the WD for all prognostic metrics in all run-to-failure datasets. Therefore, the proposed CMSWD outperforms the traditional WD regarding monotonicity, robustness, and trendability across various operating conditions. In other words, the proposed CMSWD is a superior HI that can better depict the underlying degradation characteristics for bearings.

5.4.2. RUL prediction results and comparisons

To validate the effectiveness of the proposed methodology, RUL prediction is executed under the framework of Algorithm 5-1. Leave-one-out strategy is utilized for prediction. Some vital hyperparameters, including the GRU layers (from 1 to 3), the number of hidden units in each GRU layer (from 16 to 512), optimizer (sgdm, rmsprop, adam), and learning rate (from 0.00001 to 1), are adaptively determined by GOA. To quantitatively evaluate the RUL prediction performance, the MAE and RMSE are employed.

The MAE and RMSE of each bearing run-to-failure dataset are calculated and listed in Table 5-1 utilizing Algorithm 5-1. To demonstrate the superiority of the proposed

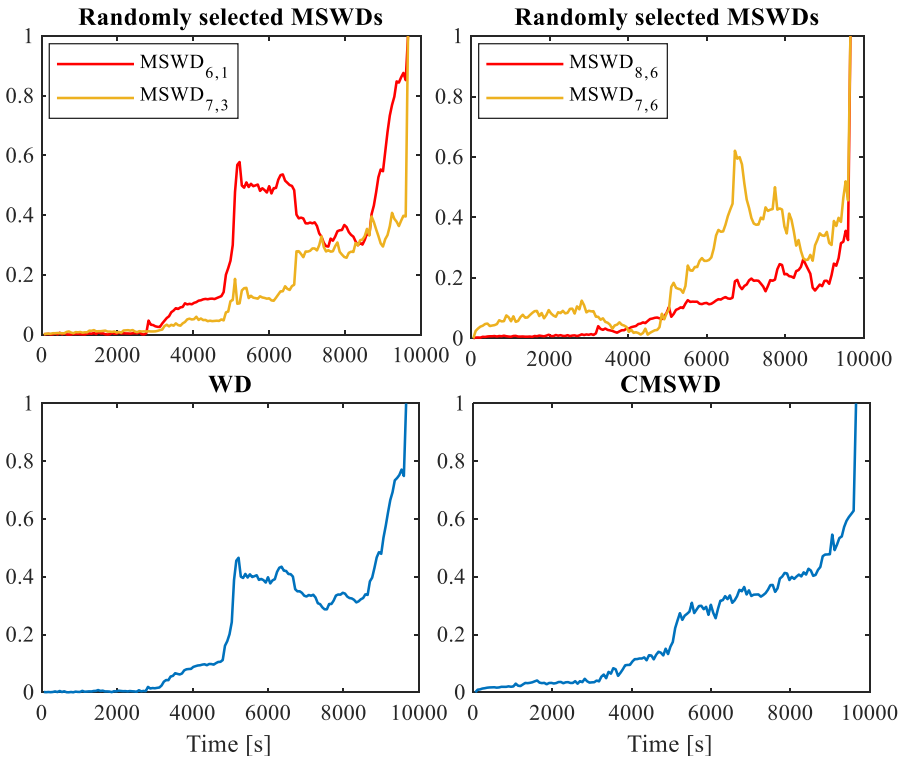


Figure 5.4. Visualization of a portion of randomly selected MSWDs, WD, and CMSWD for dataset 2-2

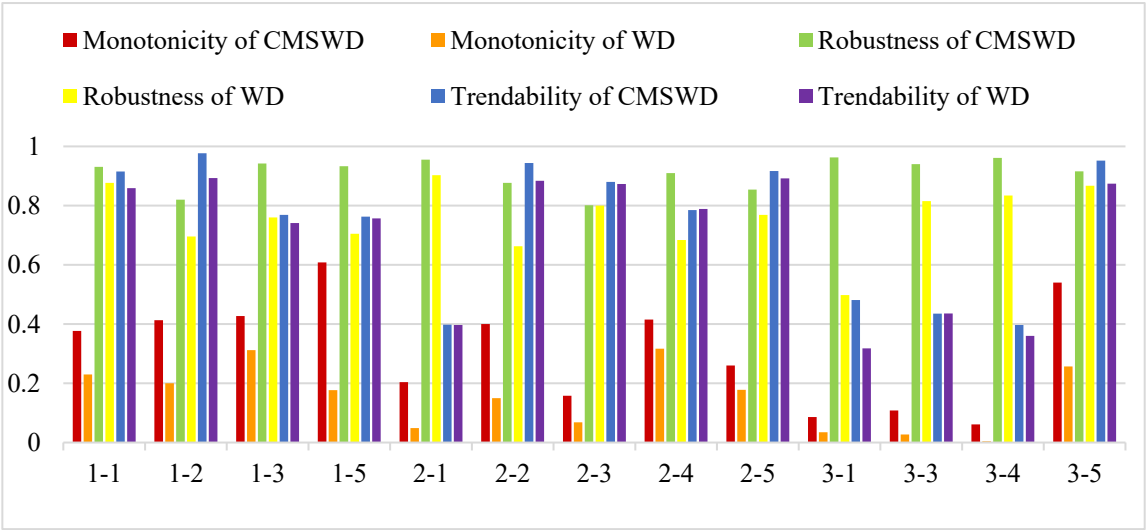


Figure 5.5. Properties of CMSWD and WD for XJTU-SY run-to-failure datasets

method, different HIs that characterize the bearing degradation from different aspects, including the WD, RMS, kurtosis, ICS2, and entropy, are utilized to replace the CMSWD, keeping other procedures in Algorithm 5-1 unchanged. The MAE and RMSE of the bearing run-to-failure datasets based on other HIs are also displayed in

Table 5-1. Note that the prediction errors using the RMS, kurtosis, ICS2, and entropy in Table 5-1 are different from those in chapter 4.4.3.1 due to the differences in the prediction models' architectures of these two chapters.

In addition, the detailed prediction results using the proposed method for the datasets with the lowest MAE for each operating condition, including datasets 1-2, 2-2, 3-4, are drawn in Figure 5.6. The detailed results of the WD-based and RMS-based methods of the same datasets are selected as representatives and depicted in Figure 5.7 for comparison. From Table 5-1, Figure 5.6, and Figure 5.7, it can be concluded that the proposed method shows better prognostic performance than the WD-based, RMS-based, kurtosis-based, ICS2-based, and entropy-based prognostic methods, obtaining lower prediction errors for all bearing datasets obtained under different operating conditions.

Table 5-1 RUL prediction errors of different HIs for XJTU-SY run-to-failure datasets

	CMSWD		**WD**		**RMS**		**Kurtosis**		**ICS2**		**Entropy**	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1-1	0.119	0.144	0.147	0.157	0.134	0.176	0.152	0.175	0.201	0.250	0.214	0.272
1-2	0.054	0.065	0.098	0.119	0.115	0.144	0.169	0.206	0.105	0.125	0.208	0.230
1-3	0.157	0.186	0.175	0.205	0.171	0.198	0.216	0.251	0.194	0.248	0.191	0.243
1-5	0.147	0.165	0.284	0.308	0.232	0.270	0.345	0.418	0.217	0.269	0.302	0.365
2-1	0.137	0.150	0.145	0.166	0.189	0.212	0.272	0.344	0.410	0.444	0.286	0.327
2-2	0.051	0.067	0.081	0.101	0.068	0.091	0.060	0.080	0.074	0.093	0.152	0.197
2-3	0.079	0.098	0.100	0.109	0.152	0.198	0.305	0.360	0.334	0.378	0.224	0.264
2-4	0.077	0.091	0.261	0.287	0.334	0.374	0.257	0.290	0.268	0.292	0.232	0.275
2-5	0.093	0.108	0.094	0.117	0.127	0.155	0.269	0.310	0.097	0.124	0.198	0.224
3-1	0.072	0.091	0.107	0.128	0.156	0.184	0.153	0.206	0.241	0.321	0.158	0.214
3-3	0.101	0.118	0.157	0.191	0.199	0.247	0.330	0.373	0.294	0.341	0.185	0.212
3-4	0.067	0.082	0.104	0.118	0.080	0.094	0.136	0.160	0.124	0.159	0.115	0.134
3-5	0.092	0.113	0.187	0.235	0.218	0.258	0.178	0.211	0.197	0.223	0.132	0.154
Average	0.096	0.114	0.149	0.172	0.167	0.200	0.219	0.260	0.212	0.251	0.200	0.239

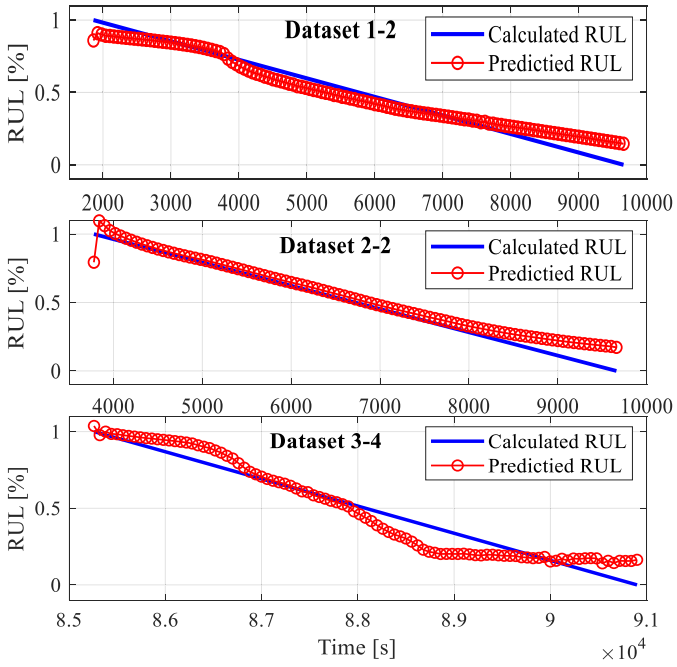


Figure 5.6. Detailed prediction results of the proposed prognostic approach

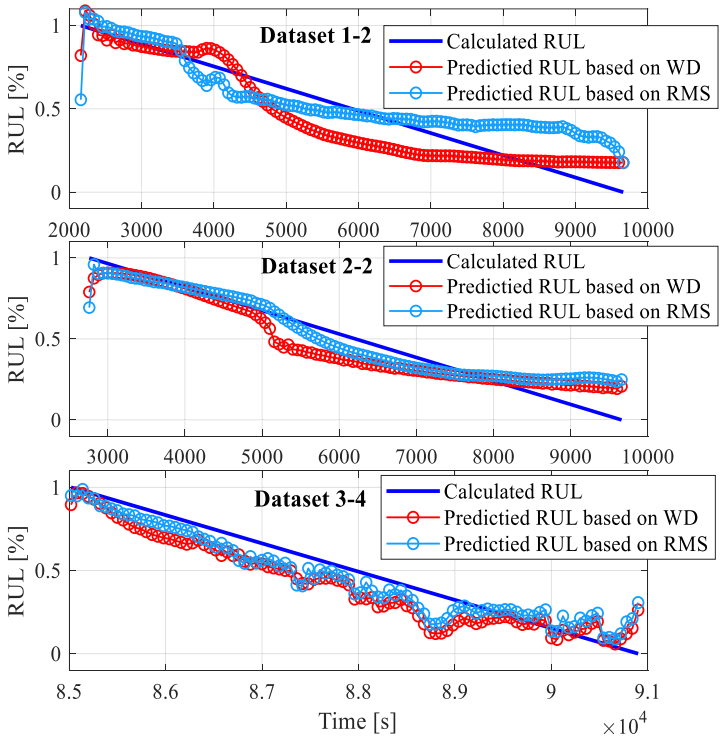


Figure 5.7. Detailed prediction results of the WD-based and RMS-based prognostic approaches

Furthermore, to comprehensively illustrate the prediction performance of the proposed method, the prediction results of the same XJTU-SY datasets using four

different state-of-the-art prognostic methods, including the BLSTM-based method [149], MSCNN-based method [150], GCU-Transformer [151], DMW-Transformer [152], which were used in Chapter 4.4.4, are again utilized in this section. From Table 4-5 and Table 5-1, it can be observed that for BLSTM-based method, it shows lower MAE than the proposed method in datasets 1-3, 2-1 and lower RMSE in dataset 1-3. For MSCNN-based method, it shows lower MAE than the proposed method in datasets 1-3, 2-1, 2-5, lower RMSE in datasets 1-3, 2-1, and the same RMSE in dataset 2-5. For DWM-Transformer-based method, it shows lower MAE than the proposed method in datasets 1-1, 1-3, 1-5, 2-4, 3-3 and lower RMSE in datasets 1-1, 1-3, 1-5, 3-3. However, lower prediction errors are achieved in the rest of the datasets, and the overall average MAE and RMSE are the lowest using the proposed method. The proposed prognostic approach, therefore, is a superior methodology for bearing RUL prediction.

5.4.3. Discussion on the results of LR-WDSC-based method and CMSWD-based method

Both Chapter 4 and Chapter 5 propose bearing RUL prediction methodologies. Compared the prognostic performance of the LR-WDSC-based method and CMSWD-based method in Table 4-2 and Table 5-1, it can be found that the LR-WDSC-based method achieves lower MAE and RMSE than the CMSWD-based method in datasets 1-1, 1-3, 2-3, 2-4, 2-5, 3-4, and attains an overall average MAE of 0.105 and RMSE of 0.127. While the CMSWD-based method achieves an overall average MAE of 0.096 and RMSE of 0.114. Therefore, the CMSWD-based method achieves a slightly lower prediction error than the LR-WDSC-based method from this aspect. However, these two methods are developed on a different theoretical basis.

The LR-WDSC-based method utilizes the WD and SC to comprehensively reflect the degradation status and employ the LR to reduce the fluctuation. While the CMSWD-based method utilizes the WD in different frequency ranges and scales to comprehensively depict the degradation status and employs a MOGOA-based fusion model to optimize the monotonicity, robustness, and trendability. From this aspect, the CMSWD-based algorithm is more complex than the LR-WDSC-based method because of the optimization process. In addition, the layers used in the CMSWD-based method are between one layer to three layers rather than one layer utilized in the LR-WDSC-based method. Thus, it is difficult to draw a simple conclusion about the better method as the CMSWD-based method has a slightly lower prediction error but a more complex calculation. It is suggested to select the appropriate method according to its real applications and requirements.

5.5. Conclusion

The novelty and contribution of this chapter can be summarized as follows.

1) The vibration signals are decomposed into multi-scale frequency sub-signals using a bank of filters with a 1/3-binary tree structure. Then the WD between the analyzed signal and healthy reference is calculated to generate a novel graph tool, named WDgram, for bearing condition monitoring. By taking advantage of the WDgram, the distribution variations can be represented in different frequency scales and ranges, thus rendering the possibility to comprehensively reveal the underlying failure characteristics at different scales without losing information buried by noise.

2) The high dimensional multi-scale distribution-similarity-based features at the WDgram are fused by the MOGOA to simultaneously optimize the monotonicity,

robustness, and trendability and thus construct a novel composite HI. The novel HI, named CMSWD, presents a good degradation trend and is fluctuations-affected-less, which contributes to accurate RUL prediction subsequently.

3) An adaptive integrated prediction model GOA-GRU is established for accurate RUL estimation. Some important hyperparameters are adaptively determined by the GOA. Results demonstrate that the proposed method is a promising prognostic method with high prediction accuracy.

Chapter 6. Conclusions and future work

6.1. Conclusion

This thesis promotes the capability of an accurate and robust condition monitoring framework for bearings by proposing compelling and innovative algorithms. Specifically, two novel diagnostic and two prognostic methodologies were proposed in Chapter 2, Chapter 3, Chapter 4, and Chapter 5 respectively to achieve this aim.

In Chapter 2, a novel correntropy-based IFB selection method has been proposed to address the limitations of traditional FK and protrugram. Being a classical yet efficient approach, the FK plays an important role in bearing fault diagnosis. However, the IFB selection results are significantly affected by the fault-irrelevant impulsive interferences. Some improved IFB selection methods, such as the protrugram, are able to suppress the influence of impulsiveness to some extent, but still surrender under high-level external impulsive interferences. Furthermore, both FK and protrugram are unable to address the issue of locating IFB properly under fault-irrelevant cyclostationary interferences, which remains a challenging topic to be studied. To bridge the research gap, a novel method named FECgram has been proposed. By taking advantage of the correntropy and the prior knowledge of the FCFs, the FECgram can select an optimal IFB where the energy at targeted FCFs is strongest. It is, therefore, a robust method which is less affected by fault-irrelevant impulsive and cyclostationary interferences. Furthermore, successful diagnosis of three challenging faulty bearing datasets has been achieved through the proposed FECgram. Both simulation and experimental studies demonstrated the promising outcomes of the

FECgram, which would be an effective approach for bearing diagnosis under complicated operational occasions.

In Chapter 3, a novel fault information-guided VMD method has been proposed to address the limitations of the traditional VMD method. The novel FIVMD method can optimally determine the mode number and bandwidth control parameter, which enables the decomposed BLIMF to be sensitive to the bearing fault signature and less affected by the abnormal impulses and vibrations from other components. The sensitivity to the bearing fault signature, which was achieved through the integration of the bearing fault cyclic period, enables early bearing fault diagnosis competently, while the insensitivity to abnormal impulses and vibrations from other components guarantee the proposed method to work effectively under strong interferences. The superiority of the proposed FIVMD method has been validated using both simulation and experimental datasets. Successful diagnosis in challenging datasets, including the early bearing fault signals, the bearing fault signals under impulsive noise and the faulty planet bearing signals, has been achieved. This confirms that the FIVMD is a promising method for bearing diagnosis under complicated operating conditions.

In Chapter 4, a systematic bearing prognostic scheme from HI construction to RUL prediction has been developed to promote the capability of accurate RUL prediction. Specifically, a novel HI named LR-WDSC has been constructed to capture the probability distribution variations at all cyclic frequencies over deterioration without losing any hidden bearing fault signature. In addition, the proposed HI eliminates random fluctuations that are irrelevant to the characteristics of degradation. Furthermore, an integrated BO-GRU network was established to adaptively determine some hyperparameters for accurate RUL prediction, improving the generalization

ability of the proposed method. Higher prediction accuracy compared with different HIs, prediction models, and state-of-the-art prognostic methods has been obtained by the proposed method. This demonstrates that the proposed method is a promising prognostic method with high prediction accuracy, which renders the possibility of the proposed method to be applied in industrial applications.

In Chapter 5, a prognostic approach has been proposed to facilitate accurate RUL prediction in bearing health management. In the proposed prognostic framework, a novel HI named CMSWD was constructed to characterize bearing degradation. The CMSWD was developed on the basis of WDgram, which generates high dimensional multi-scale distribution-similarity features, and can comprehensively reveal the underlying failure characteristics at different scales without losing information. In addition, the high dimensional features were subsequently fused by the MOGOA, enabling the proposed CMSWD to achieve high monotonicity, robustness, and trendability, which alleviates the fluctuations that are irrelative to bearing degradation. Furthermore, an adaptive integrated GOA-GRU prediction model was established for accurate RUL prediction, which was human-labor saving and could be used for different operating conditions. The performance of the proposed method was validated by bearing run-to-failure datasets under different operating conditions and fault locations. High prediction accuracy was obtained using the proposed method. The proposed bearing prognostic approach, therefore, is a prognostic method with high accuracy, which may be a promising tool for RUL prediction in modern industrial systems.

To conclude, two innovative bearing diagnostic methodologies were proposed to enhance the weak bearing fault signature under complicated operating conditions.

Two bearing prognostic methodologies were proposed to better reflect bearing fault signature over degradation and predict the RUL with high accuracy. All of these proposed methods are under the framework of facilitating advanced bearing condition monitoring techniques, which can contribute to the research community and industrial applications.

6.2. Future work

Some potential future work is listed below.

1) The current work for the bearing is most focused on localized faults. Distributed faults are also important fault types during operation. Conducting research in this field can enrich the theoretical system of bearing condition monitoring. In addition, there are some similarities in mechanism between other rotating components and bearings. The potential and promising applications of the proposed methods to other rotating machinery, such as gears, will be involved and explored in further work.

2) The current work proposed effective vibration-based algorithms for bearing condition monitoring. It should be mentioned that determining the status of bearings using only vibration signal sometimes may be unilateral and result in misinterpretation. Therefore, multi-sensor fusion is gaining increasing attention in the CBM of modern industry. The future work will involve analyzing more types of signals, such as the temperature signal, acoustic signal [162], and thermal image [163], to establish a more comprehensive framework for bearing condition monitoring.

3) Conduct more experiments to verify the generalization ability of the proposed methods. This could include datasets under time-varying speed conditions, time-varying load conditions, different fault severity conditions, and more complex structures. By using more datasets, the robustness and generality of the proposed method can be further proved and lead to the potential for applications in more complicated real-world scenarios.

References

[1] Q. Ni, J.C. Ji, K. Feng, B. Halkon, A novel correntropy-based band selection method for the fault diagnosis of bearings under fault-irrelevant impulsive and cyclostationary interferences, Mechanical Systems and Signal Processing, 153 (2021) 107498.

[2] A. Kumar, G. Vashishtha, C.P. Gandhi, Y. Zhou, A. Glowacz, J. Xiang, Novel Convolutional Neural Network (NCNN) for the Diagnosis of Bearing Defects in Rotary Machinery, IEEE Transactions on Instrumentation and Measurement, 70 (2021) 1-10.

[3] A. Glowacz, Ventilation Diagnosis of Angle Grinder Using Thermal Imaging, Sensors, 21 (2021) 2853.

[4] P. Zhang, Y. Du, T.G. Habetler, B. Lu, A survey of condition monitoring and protection methods for medium-voltage induction motors, IEEE Transactions on Industry Applications, 47 (2010) 34-46.

[5] National Renewable Energy Laboratory, "Wind Turbine Gearbox Damage Distribution Statistics 2016". http://grd.nrel.gov/#/stats (accessed August, 2022).

[6] H. Cao, L. Niu, S. Xi, X. Chen, Mechanical model development of rolling bearing-rotor systems: A review, Mechanical Systems Signal Processing, 102 (2018) 37-58.

[7] Q. Ni, K. Feng, K. Wang, B. Yang, Y. Wang, A case study of sample entropy analysis to the fault detection of bearing in wind turbine, Case Studies in Engineering Failure Analysis, 9 (2017) 99-111.

[8] K. Feng, K. Wang, Q. Ni, M.J. Zuo, D. Wei, A phase angle based diagnostic scheme to planetary gear faults diagnostics under non-stationary operational conditions, Journal of Sound Vibration, 408 (2017) 190-209.

[9] J. Zheng, Z. Dong, H. Pan, Q. Ni, T. Liu, J. Zhang, Composite multi-scale weighted permutation entropy and extreme learning machine based intelligent fault diagnosis for rolling bearing, Measurement, 143 (2019) 69-80.

[10] K. Feng, J.C. Ji, Q. Ni, A novel adaptive bandwidth selection method for Vold–Kalman filtering and its application in wind turbine planetary gearbox diagnostics, Structural Health Monitoring, 22 (2022) 1027-1048.

[11] Y. Lei, Intelligent fault diagnosis and remaining useful life prediction of rotating machinery, Butterworth-Heinemann, 2016.

[12] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing, 20 (2006) 1483-1510.

[13] Z. Gao, C. Cecati, S.X. Ding, A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches, IEEE Transactions on Industrial Electronics, 62 (2015) 3757-3767.

[14] R.B. Randall, Vibration-based condition monitoring: industrial, automotive and aerospace applications, John Wiley & Sons, 2021.

[15] J. Antoni, Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions, Journal of Sound and Vibration, 304 (2007) 497-529.

[16] W.A. Smith, Z. Fan, Z. Peng, H. Li, R.B. Randall, Optimised Spectral Kurtosis for bearing diagnostics under electromagnetic interference, Mechanical Systems and Signal Processing, 75 (2016) 371-394.

[17] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, Mechanical Systems and Signal Processing, 21 (2007) 108-124.

[18] J. Zheng, J. Cheng, Y. Yang, Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis, Mechanical Systems and Signal Processing, 40 (2013) 136-153.

[19] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: A systematic review from data acquisition to RUL prediction, Mechanical Systems and Signal Processing, 104 (2018) 799-834.

[20] P.R.N. Childs, Mechanical Design Engineering Handbook (Second Edition), Butterworth-Heinemann, 2019.

[21] A. Klausen, Condition monitoring of rolling element bearings during low and variable speed conditions, University of Agde, 2019.

[22] R. Stewart, Application of signal processing techniques to machinery health monitoring, IHalsted Press, 1983.

[23] J. Brändlein, P. Eschmann, L. Hasbargen, K. Weigand, Ball and roller bearings: theory, design and application, John Wiley & Sons, 1999.

[24] I. Howard, A Review of Rolling Element Bearing Vibration 'Detection, Diagnosis and Prognosis', Defence Technical information center, 1994.

[25] K. Feng, J.C. Ji, Q. Ni, M. Beer, A review of vibration-based gear wear monitoring and prediction techniques, Mechanical Systems and Signal Processing, 182 (2023) 109605.

[26] K. Feng, Q. Ni, M. Beer, H. Du, C. Li, A novel similarity-based status characterization methodology for gear surface wear propagation monitoring, Tribology International, 174 (2022) 107765.

[27] Rolling bearings—Damage and failures—Terms, characteristics and causes, International Organization for Standardization, 2017.

[28] R.B. Randall, J. Antoni, Rolling element bearing diagnostics—A tutorial, Mechanical Systems and Signal Processing, 25 (2011) 485-520.

[29] N. Sawalhi, Diagnostics, prognostics and fault simulation for rolling element bearings, The University of New South Wales, 2007.

[30] H. Ahmed, A.K. Nandi, Condition monitoring with vibration signals: Compressive sampling and learning algorithms for rotating machines, John Wiley & Sons, 2020.

[31] R. Alfredson, J. Mathew, Time domain methods for monitoring the condition of rolling element bearings, Nasa Sti/recon Technical Report A, 86 (1985) 102-107.

[32] D. Dyer, R. Stewart, Detection of rolling element bearing damage by statistical vibration analysis, ASME. Journal of Mechanical Design, 100 (1978) 229-235.

[33] R.B. Randall, Frequency analysis, Naerum (DK): Bruel & Kjaer, 1987.

[34] M.S. Darlow, R.H. Badgley, G. Hogg, Application of high-frequency resonance techniques for bearing diagnostics in helicopter gearboxes, Mechanical Technology Inc Latham NY, 1974.

[35] B.P. Bogert, The quefrency alanysis of time series for echoes; Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking, Proc. Symposium Time Series Analysis, 1963, pp. 209-243.

[36] S. Gade, K. Gram-Hansen, Non-stationary signal analysis using wavelet transform, short-time fourier transform and wigner-ville distribution, Bruel & Kjaer Technical Review, 2 (1996) 1-28.

[37] Z. Feng, M. Liang, F. Chu, Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples, Mechanical Systems and Signal Processing, 38 (2013) 165-205.

[38] S. Wang, X. Chen, I.W. Selesnick, Y. Guo, C. Tong, X. Zhang, Matching synchrosqueezing transform: A useful tool for characterizing signals with fast varying instantaneous frequency and application to machine fault diagnosis, Mechanical Systems and Signal Processing, 100 (2018) 242-288.

[39] R. Randall, N. Sawalhi, M. Coats, A comparison of methods for separation of deterministic and random signals, International Journal of Condition Monitoring, 1 (2011) 11-19.

[40] S. Braun, The extraction of periodic waveforms by time domain averaging, Acta Acustica united with Acustica, 32 (1975) 69-77.

[41] D. Ho, R.B. Randall, Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals, Mechanical Systems and Signal Processing, 14 (2000) 763-788.

[42] W. Wang, A.K. Wong, Autoregressive Model-Based Gear Fault Diagnosis, Journal of Vibration and Acoustics, 124 (2002) 172-179.

[43] J. Antoni, R.B. Randall, Unsupervised noise cancellation for vibration signals: part I—evaluation of adaptive algorithms, Mechanical Systems and Signal Processing, 18 (2004) 89-101.

[44] J. Antoni, R.B. Randall, Unsupervised noise cancellation for vibration signals: part II—a novel frequency-domain algorithm, Mechanical Systems and Signal Processing, 18 (2004) 103-117.

[45] R.B. Randall, N. Sawalhi, A new method for separating discrete components from a signal, Sound and Vibration, 45 (2011) 6.

[46] P. Flandrin, G. Rilling, P. Goncalves, Empirical mode decomposition as a filter bank, IEEE Signal Processing Letters, 11 (2004) 112-114.

[47] Y. Wang, R. Markert, Filter bank property of variational mode decomposition and its applications, Signal Processing, 120 (2016) 509-521.

[48] Y. Wang, R. Markert, J. Xiang, W. Zheng, Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system, Mechanical Systems and Signal Processing, 60-61 (2015) 243-251.

[49] J. Antoni, R.B. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mechanical Systems and Signal Processing, 20 (2006) 308-331.

[50] R.F. Dwyer, A technique for improving detection and estimation of signals contaminated by under ice noise, The Journal of the Acoustical Society of America, 74 (1983) 124-130.

[51] J. Antoni, The spectral kurtosis: a useful tool for characterising non-stationary signals, Mechanical systems and signal processing, 20 (2006) 282-307.

[52] Y. Lei, J. Lin, Z. He, Y. Zi, Application of an improved kurtogram method for fault diagnosis of rolling element bearings, Mechanical systems and signal processing, 25 (2011) 1738-1749.

[53] C. Li, D. Cabrera, J.V. de Oliveira, R.-V. Sanchez, M. Cerrada, G. Zurita, Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram, Mechanical Systems and Signal Processing, 76-77 (2016) 157-173.

[54] X. Zhang, J. Kang, L. Xiao, J. Zhao, H. Teng, A New Improved Kurtogram and Its Application to Bearing Fault Diagnosis, Shock and Vibration, 2015 (2015) 1-22.

[55] P. Borghesani, P. Pennacchi, S. Chatterton, The relationship between kurtosis- and envelope-based indexes for the diagnostic of rolling element bearings, Mechanical Systems and Signal Processing, 43 (2014) 25-43.

[56] A. Moshrefzadeh, A. Fasana, The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis, Mechanical Systems and Signal Processing, 105 (2018) 294-318.

[57] G. Yu, C. Li, J. Zhang, A new statistical modeling and detection method for rolling element bearing faults based on alpha–stable distribution, Mechanical Systems and Signal Processing, 41 (2013) 155-175.

[58] T. Barszcz, A. JabŁoński, A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram, Mechanical Systems and Signal Processing, 25 (2011) 431-451.

[59] J. Antoni, The infogram: Entropic evidence of the signature of repetitive transients, Mechanical Systems and Signal Processing, 74 (2016) 73-94.

[60] A. Raad, J. Antoni, M. Sidahmed, Indicators of cyclostationarity: Theory and application to gear fault monitoring, Mechanical Systems and Signal Processing, 22 (2008) 574-587.

[61] P. Borghesani, J. Antoni, CS2 analysis in presence of non-Gaussian background noise – Effect on traditional estimators and resilience of log-envelope indicators, Mechanical Systems and Signal Processing, 90 (2017) 378-398.

[62] N.E. Huang, Z. Shen, S.R. Long, The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, 454 (1998) 903–995.

[63] J.S. Smith, The local mean decomposition and its application to EEG perception data, Journal of the Royal Society Interface, 2 (2005) 443–454.

[64] K. Dragomiretskiy, D. Zosso, Variational Mode Decomposition, IEEE Transactions on Signal Processing, 62 (2014) 531-544.

[65] Y. Lei, J. Lin, Z. He, M.J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing, 35 (2013) 108-126.

[66] D. Yu, J. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mechanical Systems and Signal Processing, 19 (2005) 259-270.

[67] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Advances in Adaptive Data Analysis, 01 (2009) 1-41.

[68] I. Daubechies, J. Lu, H.-T. Wu, Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool, Applied and Computational Harmonic Analysis, 30 (2011) 243-261.

[69] J. Gilles, Empirical Wavelet Transform, IEEE Transactions on Signal Processing, 61 (2013) 3999-4010.

[70] Y. Lei, Z. He, Y. Zi, EEMD method and WNN for fault diagnosis of locomotive roller bearings, Expert Systems with Applications, 38 (2011) 7334-7341.

[71] X. Xue, J. Zhou, Y. Xu, W. Zhu, C. Li, An adaptively fast ensemble empirical mode decomposition method and its applications to rolling element bearing fault diagnosis, Mechanical Systems and Signal Processing, 62-63 (2015) 444-459.

[72] J. Cheng, Y. Yang, Y. Yang, A rotating machinery fault diagnosis method based on local mean decomposition, Digital Signal Processing, 22 (2012) 356-366.

[73] Z. Liu, Z. He, W. Guo, Z. Tang, A hybrid fault diagnosis method based on second generation wavelet de-noising and local mean decomposition for rotating machinery, ISA Transactions, 61 (2016) 211-220.

[74] J.P. Amezquita-Sanchez, H. Adeli, Synchrosqueezed wavelet transform-fractality model for locating, detecting, and quantifying damage in smart highrise building structures, Smart Materials and Structures, 24 (2015) 065034.

[75] W. Juan, G. Hongli, L. Shichao, Z. Li, H. Xiang, L. Weixiong, Fault diagnosis of ball bearings using Synchrosqueezed wavelet transforms and SVM, 2015 Prognostics and System Health Management Conference (PHM), 2015, pp. 1-6.

[76] H. Cao, F. Fan, K. Zhou, Z. He, Wheel-bearing fault diagnosis of trains using empirical wavelet transform, Measurement, 82 (2016) 439-449.

[77] J. Chen, J. Pan, Z. Li, Y. Zi, X. Chen, Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals, Renewable Energy, 89 (2016) 80-92.

[78] Z. Li, Y. Jiang, Q. Guo, C. Hu, Z. Peng, Multi-dimensional variational mode decomposition for bearing-crack detection in wind turbines with large driving-speed variations, Renewable Energy, 116 (2018) 55-73.

[79] X. Jiang, S. Li, C. Cheng, A Novel Method for Adaptive Multiresonance Bands Detection Based on VMD and Using MTEO to Enhance Rolling Element Bearing Fault Diagnosis, Shock and Vibration, 2016 (2016) 8361289.

[80] M.F. Isham, M.S. Leong, M. Lim, M. Zakaria, A review on variational mode decomposition for rotating machinery diagnosis, MATEC Web of Conferences, EDP Sciences, 255 (2019) 02017.

[81] Z. Li, J. Chen, Y. Zi, J. Pan, Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive, Mechanical systems and signal processing, 85 (2017) 512-529.

[82] J. Lian, Z. Liu, H. Wang, X. Dong, Adaptive variational mode decomposition method for signal processing based on mode characteristic, Mechanical Systems and Signal Processing, 107 (2018) 53-77.

[83] A. Dibaj, M.M. Ettefagh, R. Hassannejad, M.B. Ehghaghi, Fine-tuned variational mode decomposition for fault diagnosis of rotary machinery, Structural Health Monitoring, 19 (2019) 1453-1470.

[84] P. Shi, W. Yang, Precise feature extraction from wind turbine condition monitoring signals by using optimised variational mode decomposition, IET Renewable Power Generation, 11 (2017) 245-252.

[85] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, Z. Zhu, A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines, Mechanical Systems and Signal Processing, 116 (2019) 668-692.

[86] J. Zhu, C. Wang, Z. Hu, F. Kong, X. Liu, Adaptive variational mode decomposition based on artificial fish swarm algorithm for fault diagnosis of rolling bearings, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 231 (2015) 635-654.

[87] X. Zhang, Q. Miao, H. Zhang, L. Wang, A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery, Mechanical Systems and Signal Processing, 108 (2018) 58-72.

[88] Y. Miao, M. Zhao, J. Lin, Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition, ISA Transactions, 84 (2019) 82-95.

[89] X. Yan, M. Jia, Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings, Mechanical Systems and Signal Processing, 122 (2019) 56-86.

[90] W. Ahmad, S.A. Khan, J. Kim, A Hybrid Prognostics Technique for Rolling Element Bearings Using Adaptive Predictive Models, IEEE Transactions on Industrial Electronics, 65 (2018) 1577-1584.

[91] W. Ding, J. Li, W. Mao, Z. Meng, Z. Shen, Rolling bearing remaining useful life prediction based on dilated causal convolutional DenseNet and an exponential model, Reliability Engineering & System Safety, 232 (2023) 109072.

[92] Y. Li, X. Huang, P. Ding, C. Zhao, Wiener-based remaining useful life prediction of rolling bearings using improved Kalman filtering and adaptive modification, Measurement, 182 (2021) 109706.

[93] L. Liao, Discovering Prognostic Features Using Genetic Programming in Remaining Useful Life Prediction, IEEE Transactions on Industrial Electronics, 61 (2014) 2464-2472.

[94] L. Quagliato, D. Kim, N. Lee, S. Hwang, J. Domblesky, N. Kim, Run-out based crossed roller bearing life prediction by utilization of accelerated testing approach and FE numerical models, International Journal of Mechanical Sciences, 130 (2017) 99-110.

[95] M. Kordestani, M. Saif, M.E. Orchard, R. Razavi-Far, K. Khorasani, Failure Prognosis and Applications—A Survey of Recent Literature, IEEE Transactions on Reliability, 70 (2021) 728-748.

[96] Y. Ding, M. Jia, Q. Miao, P. Huang, Remaining useful life estimation using deep metric transfer learning for kernel regression, Reliability Engineering & System Safety, 212 (2021) 107583.

[97] M. Ma, Z. Mao, Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction, IEEE Transactions on Industrial Informatics, 17 (2021) 1658-1667.

[98] D. Yao, B. Li, H. Liu, J. Yang, L. Jia, Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit, Measurement, 175 (2021) 109166.

[99] Y. Cao, Y. Ding, M. Jia, R. Tian, A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings, Reliability Engineering & System Safety, 215 (2021) 107813.

[100] B. Wang, Y. Lei, N. Li, W. Wang, Multiscale convolutional attention network for predicting remaining useful life of machinery, IEEE Transactions on Industrial Electronics, 68 (2020) 7496-7504.

[101] H. Cheng, X. Kong, G. Chen, Q. Wang, R. Wang, Transferable convolutional neural network based remaining useful life prediction of bearing under multiple failure behaviors, Measurement, 168 (2021) 108286.

[102] Y. Wang, Y. Peng, Y. Zi, X. Jin, K. Tsui, A Two-Stage Data-Driven-Based Prognostic Approach for Bearing Degradation Problem, IEEE Transactions on Industrial Informatics, 12 (2016) 924-932.

[103] Z. Pan, Z. Meng, Z. Chen, W. Gao, Y. Shi, A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings, Mechanical Systems and Signal Processing, 144 (2020) 106899.

[104] Q. Ni, J.C. Ji, K. Feng, B. Halkon, A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis, Mechanical Systems and Signal Processing, 164 (2022) 108216.

[105] R. Liu, B. Yang, A.G. Hauptmann, Simultaneous Bearing Fault Recognition and Remaining Useful Life Prediction Using Joint-Loss Convolutional Neural Network, IEEE Transactions on Industrial Informatics, 16 (2020) 87-96.

[106] K. Feng, W.A. Smith, P. Borghesani, R.B. Randall, Z. Peng, Use of cyclostationary properties of vibration signals to identify gear wear mechanisms and track wear evolution, Mechanical Systems and Signal Processing, 150 (2021) 107258.

[107] P. Boškoski, M. Gašperin, D. Petelin, Đ. Juričić, Bearing fault prognostics using Rényi entropy based features and Gaussian process models, Mechanical Systems and Signal Processing, 52-53 (2015) 327-337.

[108] P. Wen, S. Zhao, S. Chen, Y. Li, A generalized remaining useful life prediction method for complex systems based on composite health indicator, Reliability Engineering & System Safety, 205 (2021) 107241.

[109] T. Xie, X. Huang, S.K. Choi, Intelligent Mechanical Fault Diagnosis Using Multisensor Fusion and Convolution Neural Network, IEEE Transactions on Industrial Informatics, 18 (2022) 3213-3223.

[110] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, Neurocomputing, 240 (2017) 98-109.

[111] S.A. Aye, P.S. Heyns, An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission, Mechanical Systems and Signal Processing, 84 (2017) 485-498.

[112] J. Liu, V. Vitelli, E. Zio, R. Seraoui, A Novel Dynamic-Weighted Probabilistic Support Vector Regression-Based Ensemble for Prognostics of Time Series Data, IEEE Transactions on Reliability, 64 (2015) 1203-1213.

[113] W. Caesarendra, A. Widodo, P.H. Thom, B. Yang, J.D. Setiawan, Combined Probability Approach and Indirect Data-Driven Method for Bearing Degradation Prognostics, IEEE Transactions on Reliability, 60 (2011) 14-20.

[114] Z. Xu, J.H. Saleh, Machine learning for reliability engineering and safety applications: Review of current status and future opportunities, Reliability Engineering & System Safety, 211 (2021) 107530.

[115] N. Ding, H. Li, Q. Xin, B. Wu, D. Jiang, Multi-source domain generalization for degradation monitoring of journal bearings under unseen conditions, Reliability Engineering & System Safety, 230 (2023) 108966.

[116] L. Xiao, Z. Liu, Y. Zhang, Y. Zheng, C. Cheng, Degradation assessment of bearings with trend-reconstruct-based features selection and gated recurrent unit network, Measurement, 165 (2020) 108064.

[117] Y. Qin, D. Chen, S. Xiang, C. Zhu, Gated Dual Attention Unit Neural Networks for Remaining Useful Life Prediction of Rolling Bearings, IEEE Transactions on Industrial Informatics, 17 (2021) 6438-6447.

[118] J. Zhou, Y. Qin, D. Chen, F. Liu, Q. Qian, Remaining useful life prediction of bearings by a new reinforced memory GRU network, Advanced Engineering Informatics, 53 (2022) 101682.

[119] W.A. Smith, P. Borghesani, Q. Ni, K. Wang, Z. Peng, Optimal demodulation-band selection for envelope-based diagnostics: A comparative study of traditional and novel tools, Mechanical Systems and Signal Processing, 134 (2019) 106303.

[120] J. Antoni, P. Borghesani, A statistical methodology for the design of condition indicators, Mechanical Systems and Signal Processing, 114 (2019) 290-327.

[121] I. Santamaria, P.P. Pokharel, J.C. Principe, Generalized correlation function: definition, properties, and application to blind equalization, IEEE Transactions on Signal Processing, 54 (2006) 2187-2197.

[122] L. Weifeng, P.P. Pokharel, J.C. Principe, Correntropy: A Localized Similarity Measure, The 2006 IEEE International Joint Conference on Neural Network Proceedings, 2006, pp. 4919-4924.

[123] W. Liu, P.P. Pokharel, J.C. Principe, Correntropy: Properties and Applications in Non-Gaussian Signal Processing, IEEE Transactions on Signal Processing, 55 (2007) 5286-5298.

[124] F. Jin, T. Qiu, Adaptive time delay estimation based on the maximum correntropy criterion, Digital Signal Processing, 88 (2019) 23-32.

[125] J. Peng, Q. Du, Robust Joint Sparse Representation Based on Maximum Correntropy Criterion for Hyperspectral Image Classification, IEEE Transactions on Geoscience and Remote Sensing, 55 (2017) 7152-7164.

[126] X. Zhao, Y. Qin, C. He, L. Jia, L. Kou, Rolling Element Bearing Fault Diagnosis under Impulsive Noise Environment Based on Cyclic Correntropy Spectrum, Entropy, 21 (2019) 50.

[127] Y. Fu, L. Jia, Y. Qin, J. Yang, D. Fu, Fast EEMD Based AM-Correntropy Matrix and Its Application on Roller Bearing Fault Diagnosis, Entropy, 18 (2016) 242.

[128] K. Feng, J.C. Ji, Y. Li, Q. Ni, H. Wu, J. Zheng, A novel cyclic-correntropy based indicator for gear wear monitoring, Tribology International, 171 (2022) 107528.

[129] P. Pennacchi, P. Borghesani, R. Ricci, S. Chatterton, An experimental based assessment of the deviation of the bearing characteristic frequencies, the 6th International Conference Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques, 2011, pp. 25–26.

[130] R.B. Randall, J. Antoni, S. Chobsaard, The Relationship between Spectral Correlation and Envelope Analysis in the Diagnostics of Bearing Faults and Other Cyclostationary Machine Signals, Mechanical Systems and Signal Processing, 15 (2001) 945-962.

[131] Q. Ni, K. Wang, J. Zheng, Rolling Element Bearings Fault Diagnosis Based on a Novel Optimal Frequency Band Selection Scheme, IEEE Access, 7 (2019) 80748-80766.

[132] Z. Feng, H. Ma, M.J. Zuo, Vibration signal models for fault diagnosis of planet bearings, Journal of Sound and Vibration, 370 (2016) 372-393.

[133] K. Feng, K. Wang, M. Zhang, Q. Ni, M.J. Zuo, A diagnostic signal selection scheme for planetary gearbox vibration monitoring under non-stationary operational conditions, Measurement Science and Technology, 28 (2017) 035003.

[134] Math Works, "The Math Works variational mode decomposition website". https://au.mathworks.com/help/signal/ref/vmd.html?searchHighlight=vmd%26s_tid=srchtitle (accessed August, 2022).

[135] S.S. Wilks, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, The annals of mathematical statistics, 9 (1938): 60-62.

[136] P. Borghesani, P. Pennacchi, S. Chatterton, The relationship between kurtosis- and envelope-based indexes for the diagnostic of rolling element bearings, Mechanical Systems and Signal Processing, 43 (2014) 25-43.

[137] S. Mohanty, K.K. Gupta, K.S. Raju, Effect of unitary sample shifted Laplacian and rectangular distributions in bearing fault identifications of induction motor, IET Science, Measurement & Technology, 11 (2017): 516-524.

[138] X. Jiang, C. Shen, J. Shi, Z. Zhu, Initial center frequency-guided VMD for fault diagnosis of rotating machines, Journal of Sound and Vibration, 435 (2018) 36-55.

[139] H. Wang, X. Jiang, W. Guo, J. Shi, Z. Zhu, An Enhanced VMD with the Guidance of Envelope Negentropy Spectrum for Bearing Fault Diagnosis, Complexity, 2020 (2020) 5162916.

[140] M. Zhao, X. Jia, A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery, Mechanical Systems and Signal Processing, 94 (2017) 129-147.

[141] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, Journal of Sound and Vibration, 289 (2006) 1066-1090.

[142] Case western reserve university, " Case western reserve university bearing data center website". http://engineering.case.edu/bearingdatacenter/home (accessed August, 2022).

[143] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, Mechanical Systems and Signal Processing, 64-65 (2015) 100-131.

[144] S.S. Vallender, Calculation of the Wasserstein Distance Between Probability Distributions on the Line, 18 (1974) 784-786.

[145] B. Han, X. Zhang, J. Wang, Z. An, S. Jia, G. Zhang, Hybrid distance-guided adversarial network for intelligent fault diagnosis under different working conditions, Measurement, 176 (2021) 109197.

[146] A. Frohmader, H. Volkmer, 1-Wasserstein distance on the standard simplex, Algebraic Statistics, 12 (2021) 43-56.

[147] N. Li, Y. Lei, J. Lin, S.X. Ding, An Improved Exponential Model for Predicting Remaining Useful Life of Rolling Element Bearings, IEEE Transactions on Industrial Electronics, 62 (2015) 7762-7773.

[148] B. Wang, Y. Lei, N. Li, N. Li, A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings, IEEE Transactions on Reliability, 69 (2020) 401-412.

[149] C. Huang, H. Huang, Y. Li, A Bidirectional LSTM Prognostics Method Under Multiple Operational Conditions, IEEE Transactions on Industrial Electronics, 66 (2019) 8792-8802.

[150] J. Zhu, N. Chen, W. Peng, Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network, IEEE Transactions on Industrial Electronics, 66 (2019) 3208-3216.

[151] Y. Mo, Q. Wu, X. Li, B. Huang, Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit, Journal of Intelligent Manufacturing, 32 (2021) 1997-2006.

[152] W. Zou, Z. Lu, Z. Hu, L. Mao, Remaining Useful Life Estimation of Bearing Using Deep Multi-Scale Window-Based Transformer, IEEE Transactions on Instrumentation and Measurement, 72 (2023) 3514211.

[153] C. Huang, H. Huang, Y. Li, W. Peng, A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing, Journal of Manufacturing Systems, 61 (2021) 757-772.

[154] Q. Ni, J. Ji, K. Feng, Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network, IEEE Transactions on Industrial Informatics, 19 (2023) 1301-1311.

[155] B. Wu, W. Li, M. Qiu, Remaining Useful Life Prediction of Bearing with Vibration Signals Based on a Novel Indicator, Shock and Vibration, 2017 (2017) 8927937.

[156] L. Saidi, J. Ben Ali, E. Bechhoefer, M. Benbouzid, Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR, Applied Acoustics, 120 (2017) 1-8.

[157] D. An, J.H. Choi, N.H. Kim, Remaining useful life prediction of rolling element bearings using degradation feature based on amplitude decrease at specific frequencies, Structural Health Monitoring, 17 (2017) 1095-1109.

[158] J. Dybała, Diagnosing of rolling-element bearings using amplitude level-based decomposition of machine vibration signal, Measurement, 126 (2018) 143-155.

[159] X. Wang, L. Cui, H. Wang, H. Jiang, A generalized health indicator for performance degradation assessment of rolling element bearings based on graph spectrum reconstruction and spectrum characterization, Measurement, 176 (2021) 109165.

[160] S.Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, I. Aljarah, Grasshopper optimization algorithm for multi-objective optimization problems, Applied Intelligence, 48 (2018) 805-820.

[161] S. Saremi, S. Mirjalili, A. Lewis, Grasshopper Optimisation Algorithm: Theory and application, Advances in Engineering Software, 105 (2017) 30-47.

[162] A. Glowacz, R. Tadeusiewicz, S. Legutko, W. Caesarendra, M. Irfan, H. Liu, F. Brumercik, M. Gutten, M. Sulowicz, J.A. Antonino Daviu, T. Sarkodie-Gyan, P. Fracz, A. Kumar, J. Xiang, Fault diagnosis of angle grinders and electric impact drills using acoustic signals, Applied Acoustics, 179 (2021) 108070.

[163] A. Glowacz, Thermographic Fault Diagnosis of Ventilation in BLDC Motors, Sensors, 21 (2021) 7245.