Vibration-based fault detection and severity assessment for fixed-axis gearboxes

by

Yuejian Chen

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mechanical Engineering
University of Alberta

Abstract

Gearboxes are widely used in modern power transmission systems such as wind turbines, helicopters, and railway vehicles. Due to inappropriate operation conditions or fatigue degradation, faults may develop in gears. If early faults cannot be detected and their severity assessed in a timely manner, faults will grow, and gearbox systems will fail eventually, which can lead to major economic loss or catastrophic accidents. Detection and the severity assessment of faults prior to the failure of gearbox systems can enable condition-based maintenance scheduling and thus prevent the sudden failure of gearbox systems and reduce maintenance costs. Therefore, it is of great significance to detect faults and assess their severity. Vibration-based signal analysis is a good option for the detection and severity assessment of gear tooth crack thanks to its advantages, including being easy to collect and sensitive to the tooth crack fault. To this end, the objective of this PhD research is to develop advanced vibration signal analyzing and processing techniques for tooth crack detection and severity assessment.

Specifically, the research objective is divided into three sub-objectives based on the operating condition of gearboxes. First, an improved singular value decomposition-based method is proposed for the tooth crack detection and severity assessment when the rotating speed of gearboxes is constant. The proposed method is more useful for the extraction of tooth crack induced periodic impulses than existing methods. Second, a sparse functional pooled autoregression model is proposed for more accurate modeling of

nonstationary baseline vibration from a gearbox under variable speed condition. Last, a time series model-based method is developed for the tooth crack detection and severity assessment under random speed variation.

The outcome of this research can help us better detect the gear tooth crack fault and assess its severity under either constant or variable speed conditions. Condition-based maintenance can then be better scheduled to prevent the sudden failure of gearbox systems and reduce maintenance costs. This research work has assumed only a single channel of vibration data is available for the fault detection and severity assessment. Future work will address multichannel scenarios. The fault detection and severity assessment under variable load conditions also deserve to be investigated in the future.

Preface

The material presented in this thesis is based on the original work done by Yuejian Chen. As detailed in the following, material from some chapters of this thesis has been published or submitted partially or as a whole in conference proceedings and journal articles under the supervision of Dr. Ming J. Zuo in conceptualization, review and editing, and providing resources. Dr. Xihui Liang, Dr. Stephan Schmidt, and Dr. Stephan Heyns are coauthors for some publications.

The work presented in chapter 2 has been partially published in a conference paper and as a whole in a journal paper as follows. In the journal paper, the coauthor Dr. Xihui Liang's contribution was on manuscript review and editing.

- Y. Chen and M. J. Zuo, 'Early gear tooth crack detection based on singular value decomposition', in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, Jun. 2019, pp. 1–6, doi: 10.1109/ICPHM.2019.8819417.
- Y. Chen, X. Liang, and M. J. Zuo, 'An improved singular value decomposition-based method for gear tooth crack detection and severity assessment', *J. Sound Vib.*, vol. 468, p. 115068, Mar. 2020, doi: 10.1016/j.jsv.2019.115068.

The work presented in chapter 3 has been published partially in two conference papers and as a whole in a journal paper as follows. The coauthor Dr. Xihui Liang's contribution was on manuscript review and editing. In the journal paper, one of the datasets used for

validating our approach was provided by Dr. Stephan Schmidt and Dr. Stephan Heyns at the University of Pretoria (UofP), South Africa.

- Y. Chen, X. Liang, and M. J. Zuo, 'An autoregression model for tooth crack detection considering load variations', presented at the Annual Conference of Canadian Machinery Vibration Association, Edmonton, Alberta, Canada, Oct. 2017, pp. 1–12.
- Y. Chen, X. Liang, and M. J. Zuo, 'Sparse time series modeling of the baseline vibration from a gearbox under time-varying speed condition', Mech. Syst. Signal Process., vol. 134, p. 106342, Dec. 2019, doi: 10.1016/j.ymssp.2019.106342.
- Y. Chen, X. Liang, and M. J. Zuo, 'Time series modeling of vibration signals from a gearbox under varying speed and load condition', in 2018 IEEE International Conference on Prognostics and Health Management (ICPHM), Jun. 2018, pp. 1–7, doi: 10.1109/ICPHM.2018.8449003.

The work presented in chapter 4 has been submitted as a journal paper as follows and is currently under review. The coauthor Dr. Stephan Schmidt's contributions were on software, validation, data curation, and manuscript review and editing. The coauthor Dr. Stephan Heyns's contributions were on supplying laboratory resources for the collection of gearbox vibration data, and manuscript review and editing.

- Y. Chen, S. Schmidt, P. S. Heyns, and M. J. Zuo, 'A time series model-based method for gear tooth crack detection and severity assessment under random speed variation', Mech. Syst. Signal Process., pp. 1–32, Submitted in Feb. 2020 (Under Review).

Acknowledgements

I wish to express my deepest gratitude to my doctoral supervisor Dr. Ming J. Zuo, for his help, guidance, good judgment, and criticisms. It is whole-heartedly appreciated that Dr. Zuo's great advice for my study proved monumental towards the success of this study. None of my research achievements would be possible without Dr. Zuo's supervision.

I would like to recognize the invaluable assistance that my supervisory committee members, Dr. Bob Koch and Dr. Albert Vette, provided during my study. Thanks Dr. Qing Zhao, and Dr. Chris K. Mechefske for reviewing my thesis and serving on my final defense committee. Thanks Dr. Zhigang Tian and Dr. Di Niu for severing on my PhD candidacy examination committee.

This work is dedicated to my family and my girlfriend, Miss Yizhao Li. Thanks them for their selfless support and care.

Table of Contents

List of Tables

List of Figures

List of Symbols

$\sum$	diagonal rectangular matrix storing the singular values of a matrix in descending order
$\mathbf{A}$	denotes an $m \times c$ real matrix
$\mathbf{a}$	matrix of AR model parameters
a_i	AR model parameter
$a_{i,j}$	projection parameter of FP-AR model
$a_{i,\ell,m}$	projection coefficient of VFP-AR model
$b_{i,m}$	updated projection coefficients
$b_{i,m}$	projection coefficients
$B_{i,r}(\theta_t)$	i^{th} (open) B-spline basis function
$B_{k,r}^{per}(\theta_t)$	k^{th} periodic B-spline basis function
c	number of columns of the matrix (e.g., Hankel)
$D_i(\omega_t)$	amplitude of the i^{th} impulse
$e(z)$	zero-mean white noise in z-domain
Env_t	analytical envelope of a time domain signal
Env_θ	analytical envelope of an angular domain signal
$f_i(\theta_t, \omega_t)$	i^{th} smoothing functions of phase and rotational speed
f_j	frequency of j^{th} structure vibration component;
f_m	gear meshing frequency
f_r	rotating frequency
f_s	sampling frequency
$\mathbf{G}(\theta_t, \omega_t)$	basis function set constructed by phase and rotational speed, size $p \times 1$
$\mathbf{G}_B(\theta_t)$	open B-spline basis function set constructed by phase, size $(K\text{-}r) \times 1$
$G_i(\omega_t)$	amplitude modulation function of gear meshing components induced by rotating speed

$G_j(\theta_t, \omega_t)$	j^{th} basis function constructed by phase and rotational speed
$\mathbf{G}_p(\omega_t)$	polynomial basis function set constructed by rotational speed, size $p_\omega \times 1$
$\mathbf{G}_{pB}(\theta_t)$	periodic B-spline basis function set constructed by phase, size $(K\text{-}r)\times 1$
$\mathbf{G}_{rB}(\theta_t)$	refined B-spline basis function set constructed by phase, size $(n+1)\times 1$
$H(z)$	transfer function of an system
$H^{-1}(z)$	inverse transfer function of a system
$H_{\text{bt}}()$	Hilbert transform
I	number of gear meshing harmonics
i_m	imaginary unit
J	number of structure vibration components
K	number of B-spline basis functions
k	location of impulses
l	number of SCs after SVD
l_t	load torque signal at time t
m	number of rows of the matrix (e.g., Hankel)
M	input data matrix constructed from multiple segments of signals for identifying an FP-AR model
Me	input data matrix constructed from a single segment of signals for identifying an FP-AR model
N	number of data points
n	duration of impulses
n_a	order of the AR terms
$\tilde{n}_a$	maximum lag order
n_p	number of basis functions constructed by phase and rotational speed
p	PMI value
p_s	starting point of the random speed profile
p_ω	order of Legendre polynomial basis functions
q	half-length of the tooth chordal thickness (e.g., circular tooth thickness)
q_0	depth of tooth crack
r	degree of the polynomial B-spline

$R_{Env}(0)$	total energy of the zero-mean envelope
$R_{Env}(\tau)$	autocorrelation of the zero-mean envelope
r_i	residual signal at the i^{th} point.
S	set of final SCs
$S(f, t)$	time-frequency spectrum
S_{AR}	set that specifies AR terms
$S_{F,i}$	set that specifies the functional space
S_{ini}	set of pre-selected initial SCs
S_{AR}^{ini}	initial consecutive AR set
$S_{F,i}^{\text{ini}}$	initial functional space set ith $a_i(\omega_t)$
S_F^{ini}	initial identical functional space set
t	time
T	transpose operation
T_e	number of teeth
T_h	threshold
$\mathbf{U}$	$=[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m]$
$\mathbf{u}_i$	i^{th} column vector of the unitary matrix U
$\mathbf{V}$	$=[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$
$\mathbf{v}_i$	i^{th} column vector of unitary matrix V
w	gear tooth face width
w_o	width of tooth crack
$\mathbf{x}$	$[x_1, x_2, \ldots, x_N]^T$, denotes a given one-dimensional signal (e.g., a measured vibration signal)
$X(z)$	vibration signal expressed in the z-domain and
$\mathbf{x}_{ci}$	i^{th} signal component from SVD
x_t	data points at time t of a time series $\mathbf{x}$
$Y(z)$	vibration signals in z-domain
y_m	one-step-ahead predication
y_r	real vibration response
y_s	standardized vibration signal
y_t	vibration at time t

z	denotes z-transform operator
α	decay rate of the impulses.
α_c	crack angle
ε_t	zero-mean white Gaussian noise at time t
θ_t	angular phase at time t
λ	tuning parameter
μ	coefficient of lognormal distribution
μ_s	sample mean
σ	standard deviation
σ_i	i^{th} singular values of a matrix
σ_s	sample standard deviation
τ	period of tooth crack fault induced impulses
ϕ_i	initial phase for i^{th} gear meshing harmonic
ψ_j	initial phase for j^{th} structure vibration component
$\omega_{\max}$	maximal speed among the speed profile
$\omega_{\min}$	minimal speed among the speed profile
ω_{nt}	normalized speed
ω_t	rotating speed of input shaft at time t

List of Acronyms

AR	autoregression
AAR	adaptive autoregression
AE	acoustic emission
AM	amplitude modulation
ARMED	autoregression minimum entropy deconvolution
ARX	autoregression model with exogenous excitation
BIC	Bayesian information criteria
CBM	condition-based maintenance
CM	condition monitoring
COT	computed order tracking
CPU	central processing unit
CV	cross-validation
CVMSE	cross-validation mean squared error
EHNR	envelope harmonic to noise ratio
EMD	empirical mode decomposition
FIR	finite impulse response
FM	frequency modulation
FP	functional pooled
FP-AR	functional pooled autoregression
FP-VAR	functional pooled vector autoregression
FS-TAR	functional series time-dependent autoregression
GMM	Gaussian mixture model
GSA	generalized synchronous average
HHT	Hilbert-Huang transform
HI	health indicator
HMM	hidden Markov model
IMF	intrinsic mode function
LASSO	least absolute shrinkage and selection operator

LDA	linear discriminate analysis
LTV	linear time-variant
MATLAB	matrix laboratory
MCKD	maximum correlated kurtosis deconvolution
MED	minimum entropy deconvolution
MSE	mean squared error
MSST	matching synchrosqueezing transform
NLL	negative log-likelihood
NPMI	normalized periodic modulation intensity
OLS	ordinary least squared
OSAP	one-step-ahead prediction
OT	order tracking
PMI	periodic modulation intensity
P-NARMAX	stochastic pooled nonlinear autoregressive moving average with exogenous excitation
P-NARX	stochastic pooled nonlinear autoregressive with exogenous excitation
RAM	random access memory
RLS	recursive least square
RMS	root mean square
RSS	residual sum of square
RSVD	reweighted singular value decomposition-based method
RUL	remaining useful life
SA	synchronous average
SC	singular component
SHM	structural health monitoring.
SSBAT	sinusoidal synthesis based adaptive tracking
STFT	include short-time Fourier transform
SV	singular value
SVD	singular value decomposition
SVM	support vector machine
TAR	time-dependent autoregression
TSMBM	time series model-based method
VAMSE	validation mean squared error
VFP-AR	vector functional pooled autoregression

VFP-AR-P	vector functional pooled autoregression model using the proposed refined B-splines
VFP-AR-R	vector functional pooled autoregression model using the periodic B-splines
VFP-ARX	vector functional pooled autoregression with exogenous excitation
VFP-VARX	vector functional pooled vector autoregression with exogenous excitation
VKF	Vold-Kalman filter
VS	variable speed
WT	Wavelet transform
WVD	Wigner Villa distribution

1

Introduction

This chapter contains 3 sections. Section 1.1 introduces the background of this thesis, including the basics, operational speed and fault modes of gearboxes, and fundamentals of fault detection and severity assessment; Section 1.2 contains a detailed literature review of the state-of-the-art of existing signal analyzing and processing methods; Section 1.3 provides the research objectives, topics, contributions, and organization of this thesis.

1.1 Background

1.1.1 Basics of gearboxes

Gearboxes are widely used in industrial applications, such as conveyors, automobiles, wind turbines, helicopters, milling machines, and railway vehicles. Their functions are to regulate rotating speed and torque of the power drive train. Fig. 1.1 shows two schematic examples of gearboxes used in railway vehicles [1] and wind turbines [2], respectively. In railway vehicles, the gearbox is used to transmit power from a traction motor, to decrease the speed of the traction motor, and to increase the torque on the wheel [3]. In wind turbines, the gearbox is used to transmit power from turbine blades to the generator and to increase the rotating speed of the generator rotor [4].

Based on the type of gear tooth, gearboxes can be classified into spur gears and helical gears [5]. Fig. 1.2 shows the spur gear and helical gear. Spur gears have teeth projecting

radially. The edges of each tooth are straight and aligned parallel to the axis of rotation. On the other hand, the edges of helical gear teeth are not parallel to the axis of rotation but are set at an angle. Spur gears are appropriate for low-speed applications, whereas helical gears are usually used in high-speed applications thanks to their smoother transmission. This thesis focuses on both spur gears and helical gears.

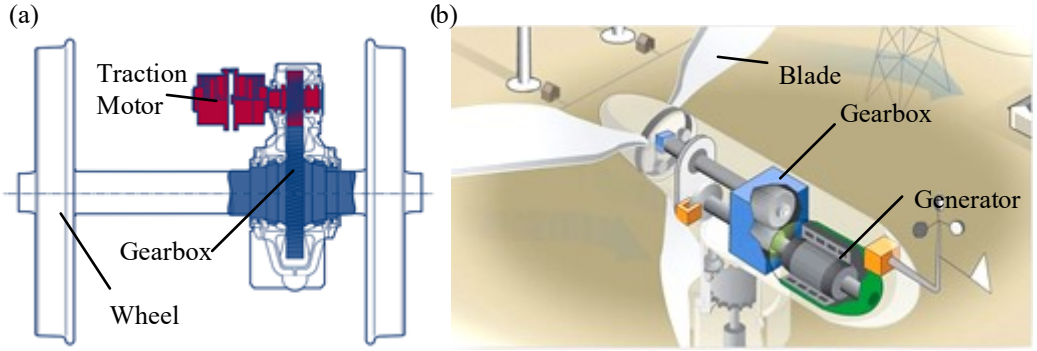

Fig. 1.1: Gearboxes used in (a) railway vehicle [1] and (b) wind turbine [2].

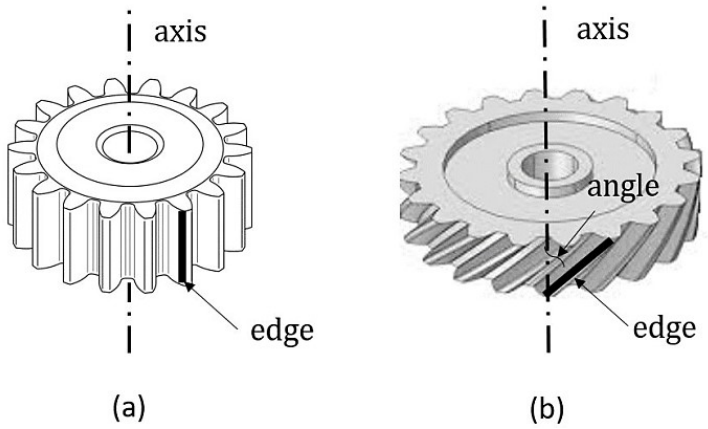

Fig. 1.2: Gear tooth-based classification (a) spur gear; (b) helical gear [5].

Based on the movement of axes, gearboxes can also be classified into fixed-axis gearboxes and planetary gearboxes. Fig. 1.3 (a) shows a typical fixed-axis gear set and Fig. 1.3 (b) a planetary gear set. The fixed-axis gear set consists of a gear and a pinion. The gear usually means the bigger ones in the gear set whereas the pinion is the smaller one [5]. The axes of the gear and pinion are fixed when the fixed-axis gear set operates. On the other hand,

the planetary gear set consists of a sun gear, a ring gear, and several planet gears held by a carrier. The axes of these planet gears are not fixed and will rotate when the planetary gear set operates. The planetary gearboxes have multiple gear pairs in mesh and hence can undertake high load torque. Therefore, they are widely used in heavy load applications such as wind turbines, helicopters, and milling machines. This thesis focuses on fixed-axis gearboxes. Planetary gearboxes are of future studies.

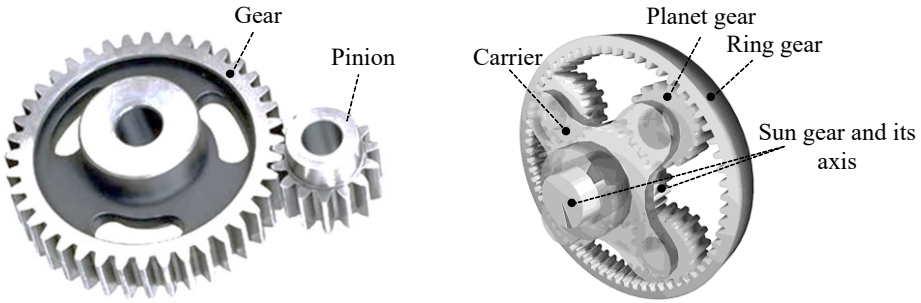

Fig. 1.3: Structure of a fixed-axis gear set (a) and a planetary gear set (b)

1.1.2 Rotating speed of gearboxes

The rotating speed of gearboxes can be either constant or variable. In railway traction, the gearbox operates under run-up, multiple levels of constant speed, and coast down conditions, as shown in Fig. 1.4. In rail transit, whether in urban subway systems or inter-city transits, the traction gearbox frequently operates under run-up and coast down conditions from stations to stations. The run-up and coast down conditions can be considered as deterministic speed profiles since they follow deterministic patterns.

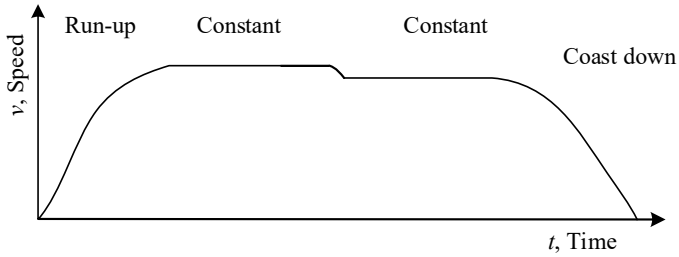

Fig. 1.4: Gearbox operating condition in railway traction

The majority of current wind turbines operate under variable speed (VS) [6]. In VS wind turbines, the gearbox operates under either constant or random speed conditions, depending on the wind speed. Fig. 1.5 shows the speed of the wind turbine gearbox relative to wind speed [7]. Four regions are involved. In region 1, the VS wind turbine is not working because the wind speed is too low. The VS wind turbine attempts to capture maximal wind energy when operating in region 2, typically from a wind speed of 5 m/s to 14 m/s [7], [8]. The speed of wind turbine gearbox is proportional to the wind speed in this region. A wind turbine operates most of its time in this region for maximizing power energy generation [7]. In region 3, the speed of wind turbine gearbox is maintained and limited to ensure that safe electrical and mechanical loads are not exceeded. The pitch angle is adjusted to regulate the input torque on the wind turbine blade and to maintain the speed of gearbox at a constant level. In region 4, the wind turbine is cut out through mechanical brake when the wind speed is too high. Fig. 1.6 further shows an illustrative example of the time waveform of wind turbine gearbox speed [7]. In region 3 when the wind speed is higher than rated speed, the gearbox operates under constant speed. In region 2 when the wind speed is higher than the cut-in speed but lower than the rated speed, the gearbox operates under random speed conditions due to the randomness of the wind speed. The speed also has autocorrelation effects due to the inertia of the drive train. The transient coast down occurs when the wind turbine changes from region 2 to region 1.

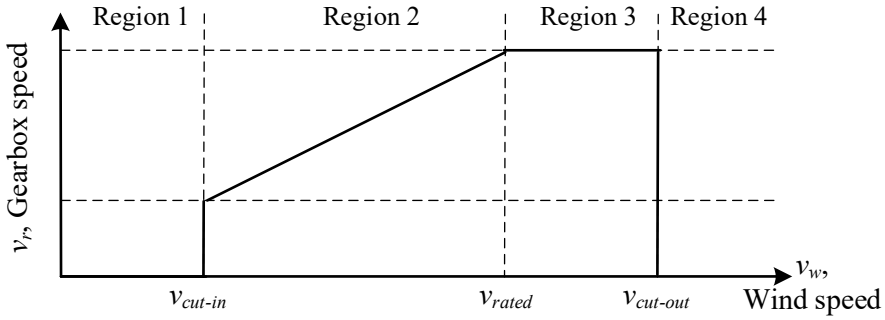

Fig. 1.5: Relationship of gearbox speed – wind speed in VS wind turbines [7]

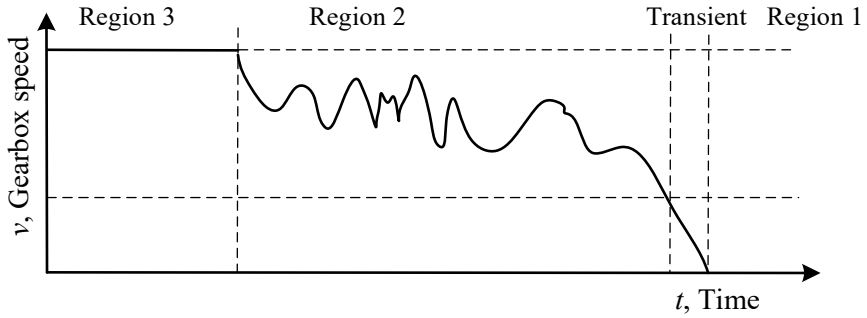

Fig. 1.6: An illustrative example of the time waveform of wind turbine gearbox speeds [7]

1.1.3 Gearboxes fault modes

Due to inappropriate operating conditions or simply fatigue [9], faults may develop in gears. A fault is defined as "*an abnormal condition or defect at the component, equipment, or sub-system level which may lead to a failure*" in document ISO 10303-226.

Faults need to be detected and their severity assessed in a timely manner. Otherwise, faults will grow, and gearbox systems will fail eventually, which can lead to major economic loss or catastrophic accidents. In Denmark 2008, a catastrophic failure of a Vestas wind turbine was reported, due to the failure of the main gear which led to a chain reaction of components failure [10]. Accident investigation of the gearbox revealed extensive damage in the main gear. The teeth of the high-speed gears had been almost completely ground down by the high-speed cog wheel. In the North Sea, United Kingdom, 2009, a helicopter crashed due to a catastrophic failure of the main rotor gearbox. Sixteen men on board were killed [11]. Accident investigations revealed that it was the crack fault which led to the catastrophic failure of the main rotor gearbox.

Generally, gear fault modes include tooth crack, tooth tip chipping, tooth wear, misalignment, eccentricity, etc [12]. Fig. 1.7 shows photos or sketches of 5 fault modes. Among these fault modes, tooth crack is one of the most frequent occurring fault modes

[13]. The tooth crack may lead to bending fatigue of the tooth, and the failure of the gearbox system.

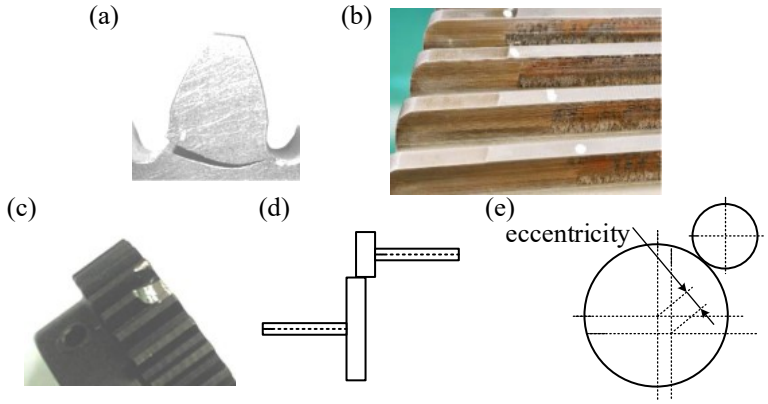

Fig. 1.7: Examples of fault modes (a) tooth crack (b) tooth wear [13], (c) tooth tip chipping, (d) misalignment, (e) eccentricity. The eccentricity is defined as the distance between the theoretical rotating center and real center.

The tooth crack consists of three distinctive stages: crack initiation, propagation and fracture [13]. Fig. 1.8 shows gear tooth crack at the initiation, propagation, and fracture stages. During the crack initiation stage, no gross yielding of the gear teeth occurs. But, local plastic deformation may occur in stress concentration regions. This stage ends when micro-cracks form inside grains. During the propagation stage, the crack grows across grain boundaries in a direction approximately perpendicular to the maximum tensile stress. Plastic deformation is confined to a small zone at the leading edge of the crack. The sudden fracture occurs during stage 3. It may be ductile, brittle, or mixed mode depending upon material toughness and magnitude of applied stress. At this stage, the tooth is completely removed from the gear. Note that in the above introduction, we focused only on crack which initiates at the root of the tooth and propagates towards the removal of the tooth. In reality, a crack may also initiate from other locations and propagate differently such as towards the gear rim, depending on the rim and web thicknesses of the gear, and

backup ratio (rim thickness divided by tooth height) [14], [15]. Fault diagnosis of these types of crack faults is of future study.

A common way to define the severity of gear tooth crack is based on the ratio between the length of the crack and the theoretical total length of a crack at the fracture stage [16]. Using this definition, the severity of tooth crack is 0% at the initiation stage and 100% at the fracture stage. At the propagation stage, the severity $\in (0, 100)\%$. For instance, the severity of the tooth crack as shown in Fig. 1.8(b) would be around 40%. This thesis focuses on the detection of tooth crack and assesses its severity.

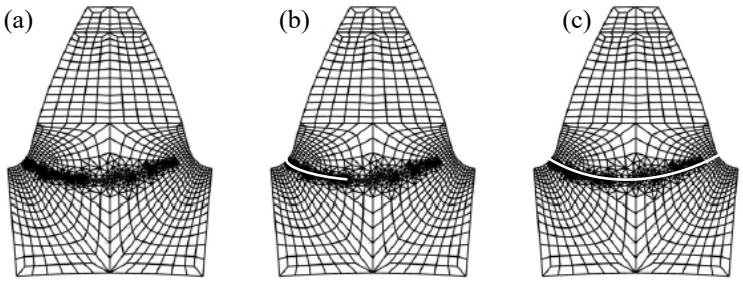

Fig. 1.8: Gear crack stages (a) initiation, (b) propagation, and (c) fracture

1.1.4 Fundamentals of fault detection and severity assessment

Fault detection and severity assessment prior to the failure of gearbox systems can prevent the catastrophic failure of gearbox systems, as well as enable condition-based maintenance and thus reduce maintenance costs. Therefore, it is of great significance to detect faults and assess their severity in the early stage of its development.

Fault detection and severity assessment belong to the condition monitoring (CM) discipline which is carried out in conjunction with condition-based maintenance (CBM) [17]. Fault detection refers to the identification of when a fault has occurred. Severity assessment refers to either quantitatively or qualitatively identify how severe the fault is. The severity can be either discrete levels or continuous. The CM includes not only fault

detection and severity assessment but also fault isolation and prognosis [18], [19]. Fault isolation refers to pinpointing the fault mode and location of the fault. Fault prognosis refers to the prediction of remaining useful life (RUL) of a component/system or estimates the probability that a component/system can still be functional for a certain future time interval [18]. Based on the CM information, CBM optimizes the operational and maintenance decision-making to ensure reliability and safety, improve availability, and reduce downtime and maintenance costs [20].

Accuracy and efficiency are two main aspects to evaluate how good a fault detection and severity assessment technique is. The accuracy refers to whether the fault detection and severity assessment results deviate from the real machine health status. The efficiency refers to how fast the fault detection and severity assessment technique returns a result.

Fault detection and severity assessment for the tooth crack fault can be achieved by offline methods, such as shutting down or even dismantling the gearbox for visual inspection. Such offline methods are usually accurate in detecting faults and assessing how severe the faults are. However, offline methods are inefficient because they not only disturb the normal operation of gearboxes but also bring an additional burden concerning operation and maintenance in terms of scheduling such inspection activities.

Online methods, on the other hand, are non-invasive, therefore being free from the aforementioned drawbacks of offline methods. Up to the present, online CM data that were widely employed for the gear tooth crack fault detection and severity assessment include vibration signals and acoustic emission (AE) [21]. The CM data is regularly collected in a certain interval (must be lower than half of the potential fault symptom to functional failure interval of the machine). By proper analyzing and/or processing these signals, one can accurately detect faults and assess their severity. The vibration refers to

the mechanical oscillation phenomenon. The vibration signals may include displacement, velocity, and acceleration acquired by laser displacement sensors, encoders, and accelerometers, respectively [18, Chapter 2]. Among these three variables, the acceleration is the most widely employed due to the maturity and low cost of piezoelectric accelerometers. Vibration signals are usually acquired at a rate ranging from 0-50kHz. The AE refers to the release of acoustic energy in the form of transitory elastic waves within a material via a dynamic deformation process [21]. The AE may also result from mechanical oscillations. Unlike vibration signals, AE is acquired in a high-frequency rate ranging from 50 kHz to 1 MHz, which means a large data size and more expensive data acquisition systems. Another limitation of the AE is the attenuation of the signal during propagation. Therefore, an AE sensor must be located as close to the fault source as possible [21], which may pose a practical constraint in applying AE. This thesis focuses on using the vibration signal for the fault detection and severity assessment of a tooth crack fault, given its advantages including easy installation, low cost, and smaller data size compared with acoustic emission.

In general, the accuracy and efficiency of the fault detection and severity assessment technique are negatively correlated. For instance, using a larger size of CM data can return a more accurate fault detection and severity assessment result, but will cost more computational time. Using a more complicated algorithm to process the CM data can return a more accurate fault detection and severity assessment result, but such complicated algorithm is relatively inefficient.

1.1.5 Fundamentals of vibration-based fault detection and severity assessment

Vibration-based fault detection and severity assessment refer to the detection of fault induced changes in the vibration signals and quantify such changes for severity assessment. Fig. 1.9 shows an example of vibration-based fault detection and severity assessment for a fixed-axis single-stage gearbox [23]. An accelerometer is mounted on the gearbox casing and used to collected vibration signals (shown as response $x(t)$). A cracked tooth will have reduced meshing stiffness when it is in meshing, and thus once the next healthy tooth starts to take over, an impulse (shock) will be excited [12]. This shock will be transmitted from the excitation spot to the accelerometer [24]–[26]. The transmission path effect may be characterized by the transfer function of the structures along the path [27]. In addition to the tooth crack induced impulses, the collected vibration signal also contains other vibration components which bring the need to process the vibration signal and to single out the signal that represents the fault induced shock.

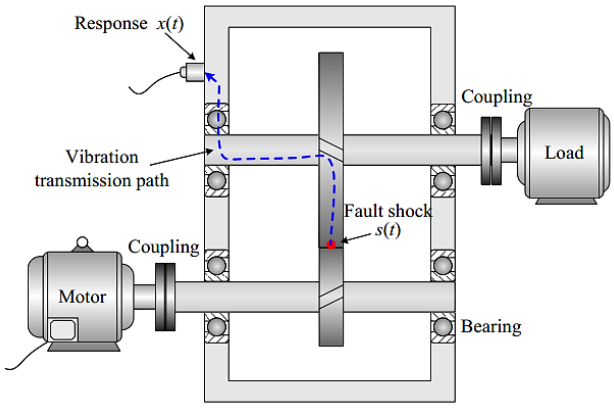

Fig. 1.9: An example of vibration-based fault detection and severity assessment for a fixed-axis single-stage gearbox. A transmission path is shown from the excitation source to the accelerometer [23].

Under constant rotating speed, the collected vibration signals mainly include gear meshing components, structural components, crack fault-induced periodic impulses, and

measurement noise [24]–[26]. The gear meshing components are induced from time-varying meshing stiffness of the gear transmission [5], [12]. These components exhibit discrete lines in the vibration spectrum with their amplitudes determined by gear meshing forces and transmission path effects. The structural components are generated by random excitations such as friction forces and tooth surface roughness [28]. These excitations transmit to the accelerometer along the transmission path, and thus the structural components exhibit broadband in the vibration spectrum. An impulse will be excited when the healthy tooth next to the cracked tooth starts to take over [12]. This impulse will be transmitted from the gear meshing location to the accelerometer. This impulse will be repeated once every revolution, therefore, termed crack fault-induced periodic impulses. The collected vibration signals are categorized as *cyclo-stationary* signals, given that the signal mean and variance vary cyclically [29]. The essence of tooth crack fault detection and severity assessment is the detection of crack fault-induced periodic impulses in the vibration signals and quantify their energy.

It is necessary to develop fault detection and severity assessment methods not only for gearboxes operating under constant speed conditions, but also for gearboxes under VS condition. As mentioned in Section 1.1.2, the traction gearbox in rail transit frequently operates under run-up and coast down conditions from stations to stations. A method capable of detecting faults and assessing severity under run-up and/or coast down conditions will be helpful in terms of providing timelier and more reliable fault detection and severity assessment results. Meanwhile, a wind turbine operates most of its time under random VS profiles [7]. A fault may occur, grow, and potentially lead to a catastrophic failure when the gearbox continuously operates under random VS profiles.

Therefore, a separate method is needed for the fault detection and severity assessment under random VS profiles.

When the rotating speed is varying, each of these components in the vibration signal will be modulated in amplitude and/or in frequency [28]. The gear meshing components and crack-induced periodic impulses will be modulated in both amplitude and frequency. The structural components will be modulated only in amplitude [30]. Lastly, the measurement noise can be assumed independent of speed variations [31]. The collected vibration signals are categorized as *nonstationary* signals, given that the signal mean and variance are time-varying and also does not repeat cyclically [28].

The VS profile is a principal factor that needs to be considered when developing fault detection and severity assessment techniques under VS conditions. The VS profile can be deterministic or random. If the vibration signal is collected under the run-up or coast down process during railway operation as mentioned in Section 1.1.2, the VS profile is deterministic for each segment of vibration signals. Note that the rotating speed needs to be non-zero. In other words, the data acquisition should start with non-zero speed during the run-up process or end with non-zero speed during the coast down process. Otherwise, the vibration signal does not contain any gearbox dynamic responses, and no fault-related features can be found in the vibration signal. Considering the vibration signals collected from wind turbine gearboxes, the VS profile is random. The VS profile could be hugely different among each segment of vibration signals. They could stay in a high/low level or vary rapidly. The fault detection and severity assessment technique are expected to detect and diagnose faults using a segment of vibration signal collected under each of the different VS profiles.

1.2 Literature review

This section contains a critical literature review on vibration signal analyzing and processing methods for gear tooth crack detection and severity assessment. This section is divided into two subsections based on the operating condition of the gearbox, namely constant speed and VS conditions.

1.2.1 Vibration-based fault detection and severity assessment under constant speed condition

In this subsection, we review methods for analyzing and processing the cyclo-stationary signals collected under constant speed conditions. Over the past decades, a wide variety of signal analyzing and processing methods have been developed for constant speed applications. The common goal of these methods is to detect the crack fault-induced periodic impulses in the vibration signal and quantify their energy. Based on the mathematical theory these methods relied on, we categorized them into general methods, resonance demodulation, time series model-based methods (TSMBMs), and singular value decomposition (SVD)-based methods. The general methods were mostly reported in the 1970s~2000s, whereas the rest of the three categories were reported during the past two decades.

1.2.1.1 General methods

The most widely employed general methods include various statistical indicators [32] and the synchronous average (SA) [33]. Typical statistical indicators include peak-to-peak, root mean square (RMS), kurtosis, skewness, and crest factor. In 2005, Samuel and Pines [34] reviewed various statistical health indicators for helicopter gearboxes. The SA method segments the vibration signal into blocks which have a length equals to the period of the a

gear rotating cycle, and then takes the average of them to obtain the periodic waveform [35]. The resulting signal is deemed deterministic. The residual signal is then obtained by subtracting the estimated deterministic part from the original signal, which constitutes the random part. Based on SA, researchers defined many health indicators (HIs) such as the energy ratio between the residual signal and the extracted periodic waveform [34], [36] and NA4 [37].

The statistical indicators are computationally fast and easy for implementation. However, these statistical indicators may suffer from a shortcoming of lacking robustness to noise [19]. As for SA, a problem is discontinuity at window boundaries [38]. The SA also requires accurate phase information.

1.2.1.2 Resonance demodulation

In 2002, Wang [25] proposed a resonance demodulation technique for early gear tooth crack detection under constant speed conditions. The technique first conducts SA to attenuate random noises. Then, the averaged waveform is band-pass filtered around the resonance band of the transmission path. The final step was to calculate the squared envelope of the band-passed signal to demodulate the crack-induced periodic impulses. Note that it is still necessary to remove the gear meshing components in the resonance band. On the selection of optimal band, kurtogram [39] was widely used, which calculates the kurtosis of different bandlimited signals and identifies the frequency bands with the maximum kurtosis. On the analysis of the envelope signal, Zakrajsek *et al.* [40] developed an NB4 indicator which is the quasi-normalized kurtosis of the analytical envelope. Xu *et al.* [41] defined an envelope harmonic to noise ratio (EHNR) to quantify the crack-induced periodic impulses.

The selection of an optimal band remains a challenge for the resonance demodulation technique. Most of the band selection methods like kurtogram rely on the extraction of a metric from the bandlimited signal. Extracting a metric only allows the frequency band which maximizes the metric to be identified, e.g. the impulsivity, and does not distinguish between the different events that contribute to the impulsivity [31].

1.2.1.3 Time series model-based methods

The TSMBMs are evolving for the fault detection and severity assessment of gearboxes [42], [43] during the past few decades. The advantages of TSMBMs include 1) being free from physics or analytical information, 2) requiring only a few signals for training, and 3) inherent accounting for uncertainties from measurement and operational conditions [44].

The identification of a baseline time series model to represent the vibration from a healthy state is a must for TSMBMs. Fault detection and severity assessment can be achieved based on analyzing the residuals obtained from applying the baseline time series model on the current vibration signals that are collected under an unknown health state, such as examining the whiteness [45], variance [45], or impulsiveness [42]. These methods belong to the category of *model residual-based methods*. Wang and Wong [42] used the autoregressive (AR) model for the fault detection and severity assessment of a gear tooth crack fault. The crack severity is assessed via examining the kurtosis (i.e., a measure of residual impulsiveness) of the baseline AR model. Endo and Randall [46] proposed applying the minimum entropy deconvolution (MED) on AR residual signals, forming a method called ARMED and demonstrated its effectiveness in detecting spalls and tooth cracks. The MED optimally determines a finite impulse response (FIR) filter whilst maximizing the kurtosis of the filtered residual signal. McDonald *et al.* [43] considered the periodic nature of the fault induced impulses and proposed a maximum correlated

kurtosis deconvolution (MCKD) method. The MCKD method optimally determines an FIR filter whilst maximizing the correlated kurtosis of the filtered residual signal. They further presented a multipoint optimal MED adjusted method with higher computational efficiency than MCKD [47]. Li *et al.* [48] proposed a sinusoidal synthesis based adaptive tracking (SSBAT) which incorporated a sinusoidal constraint of the adaptively estimated time series model. They further improved the SSBAT by incorporating MED into adaptive tracking [49]. These TSMBMs need the identification of the baseline model only.

We can also identify model(s) individually or globally for the vibration signals under various known fault severities. Fault detection and severity assessment are based on a fit measure (e.g., residual mean squared error (MSE)) of each model, or re-estimation of the severity variable such that the global model best fits (e.g., minimizing the residual MSE) the current testing signal from unknown fault severity. These methods are model residual-based and thus also belong to the category of *model residual-based methods*. Sakaris *et al.* [50] presented a vector functional pooled vector autoregression (VFP-ARX) model-based method for fault detection and localization of a truss structure. They identified a VFP-ARX model to globally represent the vibration signals under various fault locations. The fault is detected and localized via re-estimation of the location variable such that the VFP-ARX model best fits the current measured vibration signals. Kopsaftopoulos *et al.* [45] presented a VFP-ARX model-based method for fault detection, mode identification, and severity assessment of an aircraft skeleton structure. For each fault mode, they built a VFP-ARX model to represent the vibration signals under various fault locations and severities. Fault detection, localization, and severity assessment are based on the re-parameterized VFP-ARX models of any damage mode. Fault mode identification is based on evaluating

the whiteness of residuals from each VFP-ARX model. The fault mode is identified as the one which gives the whitest residual.

Fault detection and severity assessment can also be achieved using model parameters, hence termed *model parameter-based methods*. These methods require the identification of a model to represent the current vibration signals from an unknown health state. Fault can be detected by testing the statistical difference between the parameters of the current model and the parameters of the baseline model [51]. For fault severity assessment, the parameters of the current model can be treated as health condition features and used for severity trending or classification. Cheng *et al.* [52] identified AR models on intrinsic mode function (IMF) from the empirical mode decomposition (EMD). The AR model parameters are used as bearing health condition features. Man *et al.* [53] proposed an optimal sinusoidal model to represent the baseline vibration signals, as well as vibration signals collected when the gearbox has tooth crack. The crack fault detection and severity assessment are achieved via examining the energy of crack induced components in the optimal sinusoidal model. The results have shown a good trending of crack propagation. Yin *et al.* [54] further modified the optimal sinusoidal model by reducing the number of to-be-estimated parameters. Ma *et al.* [55] proposed a generalized nonlinear autoregressive - generalized autoregressive conditional heteroskedasticity model to represent the vibration of bearing. Model parameters are treated as features and are fed into a classifier for fault severity classification. These *model parameters-based methods* require the identification of a model during the testing stage, and hence generally inapplicable for online implementation.

Despite the significant amount of publications, the remaining challenges of TSMBMs include 1) the difficulty of selecting a correct model order [56] and 2) the computational complexity of model training, especially when the model order gets higher [44].

1.2.1.4 Singular value decomposition-based methods

The SVD-based methods have been widely used for periodic impulse extraction [19-25] thanks to their following merits. First, SVD-based methods are non-parametric. They do not require pre-defined basis functions [57]. Second, SVD-based methods have high computational efficiency. Last, the signal components (SCs) decomposed from SVD are free from phase shift and waveform distortion [58].

The idea behind most reported SVD-based methods for periodic impulse extraction is to preserve the SCs with significant singular values (SVs) for signal reconstruction. Zhao and Ye [59] studied an SVD-based method for headstock fault diagnosis. They introduced a difference spectrum to capture the abrupt change in the singular values, which can reflect the boundary between signal and noise. Following this idea, Qiao and Pan [58] proposed a correlation coefficient-based selection algorithm to extract abrupt information that represents the weak fault feature of rolling bearings. Golafshan and Sanliturk [60] applied an SVD- and Hankel-matrix-based de-noising method to the ball bearing time-domain vibration signals as well as to their frequency spectrums. The kurtosis ratio of the output de-noised signal to raw signal is calculated for selecting SCs. Jiang *et al.* [61] treated the ratios of neighboring SVs as features of rolling-element bearing vibration signals. Zhang *et al.* [62] proposed a time-varying SVD method to enhance periodic impulse extraction. The time-varying SVD method applies SVD to the signal within a sliding window to obtain a time-varying matrix of SVs that suppresses noises and enhances fault features.

The SVs mainly reflect the energy of decomposed SCs. Therefore, most reported methods are essentially energy-based, which can highlight high-energy components in the measured signal but may ignore weak-energy features caused by early faults. Zhao and Jia [57] proposed a reweighted SVD (RSVD) method. This method applies SVD to decompose the raw signal into SCs, then uses a periodic modulation intensity (PMI) criterion innovatively to measure how influential the periodic impulses are in each SC. However, RSVD suffers from two issues: 1) it does not consider interference from non-fault related vibration components on the PMI, which leads to a high miss and/or false alarm rate; and 2) it directly conduct reweighted summation of all SCs with PMI values that exceed the threshold for signal reconstruction, which undermines periodic impulse extraction.

1.2.1.5 Summary

Table 1.1 summarizes the advantages and disadvantages of the reviewed signal analyzing and processing methods for cyclo-stationary signals collected when the gearbox operates under constant speed conditions. It is worth to mention that each category of the methods has its own advantages and disadvantages. There is no clear optimum. We may not conclude that one category of methods should be dominating the others. There is still room for further improvement in each category. Many researchers are making efforts on advancing each category.

Given the advantages of the SVD-based method as listed in Table 1.1, we are going to have a topic on further advancing the SVD-based method on the fault detection and severity assessment under constant speed condition.

Table 1.1: Summary of cyclo-stationary signal analyzing and processing methods

Categories	**Advantages**	**Disadvantages**
Statistical indicators	1) high computational efficiency 2) easy for implementation	1) lack of robustness to noise
Synchronous average	1) effectively suppress random noise 2) high computational efficiency 3) easy for implementation	1) affected by the noise-induced from discontinuities at window boundaries 2) require accurate rotating phase measurement
Resonance demodulation	1) high computational efficiency 2) easy for implementation	1) hard to select an optimal band
Time series model-based methods	1) being free from physics or analytical information 2) need only a few signals for model training 3) inherent accounting for uncertainties from measurement	1) difficult to select a correct model order 2) computational complexity of model training, especially when the model order gets higher
SVD-based methods	1) do not require pre-defined basis functions 2) high computational efficiency 3) free from phase shift and waveform distortion	1) may not extract the weak-energy features caused by early faults

As mentioned in Section 1.2.1.4, most reported methods are essentially energy-based, which may ignore weak-energy features caused by early faults. Zhao and Jia [57] proposed a reweighted SVD (RSVD) method which uses a criterion to measure how influential the periodic impulses are in each SC. This RSVD method addresses the problem of ignoring weak-energy features caused by early faults, but still suffers from two issues: 1) it does not consider interference from non-fault related vibration components on the PMI, which leads to a high miss and/or false alarm rate, and 2) it directly conduct reweighted summation of all SCs with PMI values that exceed the threshold for signal reconstruction, which undermines periodic impulse extraction.

To cope with these issues that RSVD face, we will propose an improved SVD-based method for the extraction of tooth crack induced periodic impulses, and thus for more accurate tooth crack fault detection and severity assessment. We will adopt an AR model-based approach [42] to remove the baseline vibration components from the raw signal. SVD is applied to a residual signal rather than the raw signal; interference from these baseline vibration components to the PMI is therefore eliminated. Second, we will select the SCs such that the PMI of the reconstructed signal is maximized. Periodic impulses extraction is therefore strengthened. Corresponding work will be detailed in Chapter 2.

1.2.2 Vibration-based fault detection and severity assessment under variable speed condition

In this subsection, we review methods for analyzing and processing the nonstationary signals collected under VS conditions. Based on the mathematical theory these methods relied on, we categorized them into general methods, discrepancy analysis, time-frequency analysis, and TSMBMs.

1.2.2.1 General methods

As mentioned in section 1.1.5, the gear meshing components are modulated in both amplitude and frequency under VS. Order tracking (OT) is a technique which demodulates the frequency modulation (FM). The OT transforms a measured signal from the time-domain to the angular-domain. Rotating phase measurement, using an encoder or tachometer, is required as additional information. In the 1990s, OT techniques included the hardware OT and computed OT [63]. Afterwards, Vold and Leuridan [64]–[66] proposed a Vold-Kalman filter (VKF) which can extract order waveforms in the time-domain without phase bias. Pan *et al.* [67] presented an adaptive version of the VKF. Recently, Borghesani *et al.* [68] proposed a velocity synchronous discrete Fourier

transform to directly assess the amplitude of synchronous gear meshing harmonics without an additional resampling step. Generally, OT cannot demodulate the amplitude modulation (AM) induced by VS. When demodulating the FM of gear meshing components, OT will induce additional FM to the structural components and crack-induced impulses which were originally only modulated in amplitude.

Researchers developed methods to demodulate the AM induced from VS condition, such that the nonstationary signals may be converted to cyclo-stationary. Schmidt *et al.* [69] proposed a procedure for the normalization of the AM caused by varying operating conditions for the fault detection and severity assessment of gearboxes. The key idea of the procedure is to divide the raw signal by its low-passed envelope. Urbanek *et al.* [70] defined a generalized angular deterministic signal which has FM induced from VS, and AM induced from both VS and variable load. They presented a demodulation method to convert the nonstationary signal into cyclo-stationary. These two works [69], [70] assumed an identical AM effect on each gear meshing component which was not realistic. Wang *et al.* [71] presented a multiscale filtering reconstruction method which adopts VKF to filter out each gear meshing component first and demodulates the AM effect on each gear meshing component. Unfortunately, these demodulation methods [69]–[71] not only demodulates operating condition effects but also attenuates fault induced impulses, which can weaken the fault induced features and hence undermine the accuracy of fault detection and severity assessment.

The SA was generalized by Abboud *et al.* [27] in 2016 for the VS case. The generalized synchronous average (GSA) method performs the cyclic average on samples pertaining to the same speed. The GSA assumed a stationary behaviour within a rotating speed range.

A drawback of the GSA is that it needs a long duration of signals for having an accurate estimation of deterministic components.

Various HIs were designed to be sensitive to crack induced periodic impulses but insensitive to operating condition variations. Sharma *et al.* [36] examined the effectiveness of various HIs, i.e., RMS, kurtosis, crest factor, FM0, FM4, M6A, NB4, energy ratio, NA4, energy operator, on nonstationary signals, and proposed two new HIs, namely, modified Rényi entropy, RMS ratio between residual signals and regular vibrations. Sharma *et al.* [72] proposed an RMS-based probability density function for the calculation of entropy-based HIs. Grylllia *et al.* [73] used cyclo-nonstationary HIs for the bearing health monitoring under VS condition. Villa *et al.* [74] found that the diagnostic features (e.g., root mean square) are dependent on the rotating speed and load. The change of such dependency can reflect the deterioration of a gearbox. Santos *et al.* [75] considered fault diagnosis of imbalance and misalignment faults under random speed and load variation. They extracted numerous features from the vibration signal and fed them along with speed and load information into a support vector machine (SVM). Although the features they extracted are sensitive to speed/load changes, SVM is expected to learn the dependency between these features and the speed/load and still return a good classification result. When applying the approach by Santos *et al.* [75] for gear tooth crack detection and severity assessment, we need to extract proper features that are sensitive to the gear tooth crack.

1.2.2.2 Discrepancy analysis

Recently, Schmidt *et al.* [76] presented a comprehensive novelty diagnostic method based on discrepancy analysis. The method is applicable to the fault diagnosis of a machine under distinct operating condition states such as idling, full load, and transient states

within a measurement [76]. The method first identifies the OC state of the gearbox by assuming that the speed does not vary significantly within a rotating cycle. Secondly, the method calculates a negative log-likelihood (NLL) based on machine health condition features and the OC state. The NLL can reflect the propagation of the localized gear fault well. The novelty diagnostic method may be adequate for smooth random speed variations. However, when speed varies rapidly, the vibration signal within a rotating cycle may be subjected to significant variations.

1.2.2.3 Time-frequency analysis

The time-frequency analysis studies a signal in the joint time-frequency domain thus can comprehensively reveal the characteristics of nonstationary vibration signals [77]. By time-frequency analysis, the constituent frequency components of the vibration signal and the transient time events in the signal such as fault-induced impulses can be revealed [78]. A large volume of time-frequency analysis methods has been presented for the fault detection and severity assessment of gearboxes. The most classical methods include short-time Fourier transform (STFT), Wavelet transform (WT), Wigner Villa distribution (WVD), and Hilbert-Huang transform (HHT). Two recent review papers were available by Feng *et al.* [78] and Yang *et al.* [79]. Notably, Wang *et al.* [80] applied the WT to gear residual signal from a gearbox under a time-varying load condition. Gear residual signal is obtained from conducting SA on raw signals. Antoniadou *et al.* [81] presented a time-frequency analysis method based on empirical mode decomposition (EMD) and Teager–Kaiser energy operator. The localized fault of the wind turbine gearbox was detected. Wang *et al.* [82] proposed a matching synchrosqueezing transform (MSST) and presented its application in fault detection of a wind turbine gearbox under VS. The MSST can qualitatively detect a slight tooth spall through visual observation of the time-frequency

spectrum. Feng *et al.* [83] constructed a time-frequency spectrum via the VKF and Hilbert transform. They analyzed vibration signals from a planetary gearbox under a coast down speed profile. A sun gear chipping fault was detected by observing the fault characteristic frequency on the time-frequency spectrum.

Although time-frequency analysis methods were widely investigated, each of these methods subject to certain limitations. For example, application of WT requires a proper selection of wavelet basis [78]; EMD subjects to mode mixing problem when the signals have instantaneous frequency trajectory crossings, and requires a higher sampling frequency of the vibration signal as cubic spline to fit data and hence it needs more samples for fine interpolation [78]. On the other hand, time-frequency analysis only qualitatively detects the fault by a subject matter expert and have not quantitatively assessed the fault severity. It is required to incorporate other feature extraction methods for quantitative fault severity assessment.

1.2.2.4 Time series model-based methods

As mentioned in Section 1.2.1.3, the identification of a baseline model to represent the vibration from a healthy state is a must for TSMBMs. The accuracy of baseline modeling directly determines the success of TSMBMs. Unlike modeling the cyclo-stationary signals, modeling the nonstationary signals require time-variant models.

Few time-variant time series models have been reported to represent the nonstationary baseline vibration from rotating machinery operating under VS conditions. Wyłomańska *et al.* [84] proposed a periodic AR for modeling the baseline vibration from a planetary gearbox of bucket wheel excavator. This periodic AR model is applicable to the case when the operating condition varies periodically. Zhan *et al.* [85] and Shao *et al.* [86] developed an adaptive AR model (AAR) to model the baseline vibration from a fixed-axis gearbox.

The Kalman filter was adopted to adaptively adjust the AR parameters in terms of minimizing the residual sum of squares. Heyns *et al.* [87] constructed a weighted ensemble of AR models for modeling the baseline vibration from a fixed-axis gearbox. A Bayesian probability scheme was adopted to adaptively determine the weight of each AR model. The constructed weighted ensemble of AR models is essentially a special type of AAR model. Makowski *et al.* [88] proposed to use the reflection coefficients of a Schur adaptive filter and their derivatives for fault detection of a rolling element bearing. Reflection coefficients are adaptively adjusted to minimize the mean squared prediction error. These adaptive models require a proper tuning of a convergence rate. Too high of a convergence rate results in overfitting, and too low of a convergence rate causes underfitting. Yang and Makis [89] used the AR model with exogenous excitation (ARX) to represent the baseline vibration from a fixed-axis gearbox under a time-varying load condition. They innovatively considered the load information, quantified from a torque sensor, as an exogenous added-up excitation to compensate time-varying load effects on the baseline vibration. Unfortunately, if we consider the VS as the exogenous part, time-varying harmonics with their frequencies different from the frequencies of the VS cannot be compensated [90]. Using ARX will have large modeling error for VS case.

Functional pooled (FP) models have been used in the context of SHM [45], [91], [92]. Kopsaftopoulos *et al.* [91] used a vector functional pooled autoregression (VFP-AR) model for globally representing the dynamics of a healthy aerospace structure under multiple flight states. Hois *et al.* [51] presented a functional pooled vector autoregression (FP-VAR) model-based method for fault detection of a composite cantilever beam under variable temperature conditions. Aravanis *et al.* [93] presented an FP-AR model with an exogenous excitation-based method for fault detection of a railway suspension under variable loading

conditions. In [45], [91], [92], vibration signals were collected under different constant levels of operating conditions and FP models were identified to represent the vibrations under different levels of operating conditions. Belonging to FP models, the FP-AR model configures the AR parameters as functions of operating conditions [94]. A more general model can be found in [95] incorporating moving average and external excitation terms. In addition, vector dependent FP [96] may be involved when multiple operating condition variables affect the vibration, and vector AR when modeling multiple channels of signals [97]. In statistics, the FP models belong to varying coefficient regression models [98]. In the system identification field, the FP models are also be referred as signal-dependent models [99].

Without loss of generality, we can consider the FP-AR model with its AR parameters dependent on VS. After configuring the AR parameters as functions of rotating speed, the FP models will have unique spectral content at each speed level. Therefore, the FP models can model the complex speed-dependent spectrum contents of the nonstationary baseline vibration.

Identifying an FP-AR model needs parameter estimation and model structure selection methods. Model parameters are estimated by the ordinary least squared (OLS) estimator [94]. Model structure refers to AR terms and functional spaces that describe the dependency between AR parameters and VS. Conventionally, the model structure is selected via the two-step Bayesian information criteria (BIC) based method [45], [50], [56]. This two-step BIC-based method firstly assumed consecutive AR terms ($S_{\mathrm{AR}} = \{1, 2, \ldots, n_a\}$, where S_{AR} is a set that specifies AR terms; and n_a is the maximum lag order of the AR term) to simplify the selection of AR terms into a one-dimensional optimization problem, i.e., determine an optimal $\hat{n}_a$. Such a one-dimensional optimization problem is solved by

evaluating enough range of n_a values and then determining the optimal $\hat{n}_a$ that minimizes BIC. Secondly, the two-step BIC-based method assumed identical sets of functional spaces $S_{F,i} = S_F$ for $i \in S_{AR}$ (where $S_{F,i}$ is a set that specifies the functional spaces), which simplifies the selection of many sets $S_{F,i}$ for $i \in S_{AR}$ into the selection of one set S_F. In this step, the backward deletion scheme is adopted [100]. The S_F is selected by initially starting from a high-dimensional set and then sequentially dropping each one of the basis functions until no reduction of the BIC is possible. However, the two-step BIC-based method limited the modeling accuracy of FP-AR. Firstly, much real-world time series, such as seismology waves, radar signals, and speech signals, are generated from a sparse AR process [101]. A sparse AR model uses scattered AR terms as predictors, rather than consecutive AR terms. The baseline vibration from a gearbox may also be generated from a sparse AR process. Thus, using an FP-AR with a sparse set S_{AR} to model the gearbox vibrations may be more reasonable than assuming consecutive set S_{AR}. Secondly, the dependency between different AR parameters and the VS for $i \in S_{AR}$ may be very different. They may require non-identical sets of functional spaces $S_{F,i}$ for $i \in S_{AR}$ to achieve accurate regressions. Assuming an identical set of functional spaces $S_{F,i}$ for $i \in S_{AR}$ may be too coarse.

Following the baseline vibration modeling, we consider the fault detection and severity assessment. As mentioned in Section 1.2.1.3, fault detection and severity assessment can be achieved based on analyzing the residuals obtained from applying the baseline model on the current vibration signals that are collected under an unknown health state, such as examining the whiteness, variance, or impulsiveness. Zhan *et al.* [85] used the kurtosis of the AAR model residual and normality hypothesis testing for crack fault detection and severity assessment. Shao *et al.* [86] used the Wilcoxon rank-sum test to exam the whiteness of the AAR model residual for crack fault detection and severity assessment.

Heyns *et al.* [87] used RMS of the residual of a weighted ensemble of AR models for crack fault detection and severity assessment. Yang and Makis [89] used an F-statistic to exam the variance of the ARX model residual for crack fault detection and severity assessment.

These baseline model residual-based methods may only apply to the cases when the testing signals were collected under deterministic speed profiles, for example, the run-up or coast down profiles in rail transit. Under random speed variations that typically occur in wind turbine gearboxes, the random profile within each data acquisition window could be hugely different. They could stay in a high/low level or vary rapidly. A CM system is expected to detect and diagnose faults using a segment of vibration signal collected under each of these different profiles. These various statistics calculated from the residual of the baseline model are still affected by the random speed profiles. Therefore, these methods, such as examining the whiteness, variance, or impulsiveness of the residual of baseline model, may be ineffective.

Another category of TSMBAs, namely, *model parameter-based methods*, is available for the fault detection and severity assessment. In the SHM field, Hois *et al.* [51] presented an FP-VAR model-based method for fault detection of a composite cantilever beam under variable temperature conditions. Fault detection is based on statistically comparing the model parameters from a testing signal with these parameters obtained from the training stage. Spiridonakos *et al.* [102] presented a functional series TAR (FS-TAR) model-based mode identification method for a pick-and-place mechanism. FS-TAR models are built for each fault mode. Fault mode identification is based on statistically comparing the model parameters from a testing signal with these parameters obtained from the training stage. Aravanis *et al.* [93] presented an FP-AR model with an exogenous excitation-based method for fault detection of a railway suspension under variable loading conditions.

Dimogianopoulos *et al.* [103] presented a two-stage stochastic pooled nonlinear autoregressive moving average with exogenous excitation (P-NARMAX) representations to model the relationships among available aircraft signals under varying operating conditions (e.g., flight angles). In the first stage, a constant coefficient P-NARMAX representation was identified to account for major nonlinearities in the relationship and to deliver the initial residuals. In the second stage, multiple stochastic pooled nonlinear autoregressive with exogenous excitation (P-NARX) models were identified for each health state using training data (assumed available). These health states included healthy and multiple faulty states. Online diagnostic of the aircraft is based on online measured aircraft signals and recursively estimating a P-NARX model via recursive least square (RLS). The parameters of the recursively estimated P-NARX model reflects the current health state, and the distance between the estimated P-NARX model parameters from the trained model parameters are computed as a fault diagnostic metric. However, these *model parameters-based methods* require the identification of a model during the testing stage, and hence generally inapplicable for online CM implementation of gearboxes due to the relatively large CM data size.

1.2.2.5 Time varying singular value decomposition-based method

The SVD-based methods were modified for the analysis of nonstationary signals by incorporating the short time windowing idea [62], [104]. Technically, the modified SVD-based method cuts the nonstationary signal into several segments using a sliding short time window and then applies SVD to decompose each segment. The sliding window will have a certain overlap. Then, the SV series are used as features for fault diagnosis or pattern recognition. However, such time varying SVD-based methods have the limitation caused by the sliding short time window, that is, there is a tradeoff between time resolution

and frequency resolution. If the duration of the window is too long, the method is unable to catch the nonstationary time varying characteristics of the signal; whereas if the duration of the window is too short, the method is unable to catch the frequency characteristics of the signal.

1.2.2.6 Summary

Table 1.2 summarizes the advantages and disadvantages of the reviewed signal analyzing and processing methods for nonstationary signals collected when the gearbox operates under VS condition. Again, each category of the methods has its own advantages and disadvantages. There is no clear optimum.

Given the advantages of time series model-based methods, we are going to have two topics on further advancing the time series model-based methods for fault detection and severity assessment under VS condition.

We first aim to improve the accuracy of baseline vibration modeling. As mentioned in Section 1.2.2.4, the structure of FP-AR model is conventionally selected via the two-step BIC based method [45], [50], [56]. The two-step BIC-based method assumed 1) consecutive AR terms, e.g., $S_{\mathrm{AR}}=\{1,2,\ldots, n_a\}$, and 2) identical sets of functional spaces for $i \in S_{AR}$, e.g., $S_{F,i} = S_F$. Although these two assumptions simplify the model selection and enable the two-step BIC-based method [45], [56], [91], they, however, limited the modeling accuracy of conventional FP-AR. To improve the modeling accuracy, this thesis will propose a sparse FP-AR model that uses a sparse AR set S_{AR} and non-identical functional spaces $S_{F,i}$. These two assumptions made in the two-step BIC based method will be relaxed. Corresponding work will be detailed in Chapter 3.

Table 1.2: Summary of nonstationary signal analyzing and processing methods

Categories	Advantages	Disadvantages
Statistical indicators	1) high computational efficiency 2) easy for implementation	1) lack of robustness to noise
Generalized synchronous average	1) effectively suppress random noise 2) high computational efficiency 3) easy for implementation	1) affected by the noise induced from discontinuities at window boundaries 2) needs a long duration of signals for having an accurate estimation of deterministic components
Order tracking	1) effectively demodulate FM effect induced from VS condition	1) cannot demodulate the AM induced by VS 2) will induce additional FM to the structural components and crack-induced impulses which were originally only modulated in amplitude
Amplitude demodulation	1) effectively demodulate AM effect induced from variable operating condition	1) may attenuate fault induced impulses
Discrepancy analysis	1) applicable for any kind of operating condition, e.g., idling, full load, and transient conditions	1) assumed insignificant speed changes within one rotating cycle, thus not suitable when speed varies rapidly 2) need lots of baseline data for training
Time-frequency analysis	1) comprehensively reveals the characteristics of nonstationary vibration signals	1) each of time-frequency analysis methods subject to certain limitations, such as mode mixing to EMD, cross term to WVD, selection of wavelets to WT
Time series model-based methods	1) being free from physics or analytical information 2) need only a few signals for training 3) inherent accounting for uncertainties from measurement	1) difficult to select the model structure; incorrect model structure will lead to inaccurate modeling of nonstationary vibration signals 2) computational complexity of model training, especially when the model order gets higher
Time varying SVD-based methods	1) do not require pre-defined basis functions 2) high computational efficiency 3) free from phase shift and waveform distortion	1) tradeoff between time resolution and frequency resolution

For the fault detection and severity assessment, we can use the residuals of the baseline model when the testing signals were collected under deterministic speed profiles. However, these methods are inapplicable to the cases when the testing signals are collected under random speed variations which typically occur in wind turbine gearboxes, as mentioned in Section 1.2.2.4.

We will aim to develop a TSMBM for the gear tooth crack detection and severity assessment under the random speed variation. We will identify models on the residuals of the baseline sparse FP-AR model under random speed variation. A VFP-AR model that considers rotating speed and rotating phase as covariates will be developed for representing the impulsive residual signals, given that the rotating speed affects the amplitude of tooth crack induced impulses and the rotating phase determines the location of these impulses. Multiple VFP-AR models will be built for each level of tooth crack severity. Fault detection and severity assessment can, therefore, be based on examining the residuals of each VFP-AR model. The final health state can be assessed as the severity state of the model which gives the minimal residual energy. Such fault detection and severity assessment scheme is computationally efficient and can be applied in an online scenario. Corresponding work will be detailed in Chapter 4.

1.3 Thesis objectives and outline

The overall objective of this thesis is to develop new vibration-based methods for more accurate gear tooth crack detection and severity assessment. Based on the state-of-the-art signal analyzing and processing methods as reviewed in Section 1.2, we have the following three specific sub-objectives:

1) Propose an improved SVD- based method for more effective extraction of tooth crack induced periodic impulses from cyclo-stationary vibration signals under constant speed applications.
2) Propose a new time series model for more accurate modeling of nonstationary baseline vibration from a gearbox under VS condition.
3) Develop a time series model-based approach for more accurate gear tooth crack detection and severity assessment under random speed variation.

Our basic assumptions in this thesis are as follows:

1) Only a single channel of vibration signal is available for the gear tooth crack fault detection and severity assessment.
2) Rotating speed and phase information can be accurately acquired by either tachometers or speed estimation techniques.
3) The load condition of the gearbox is either constant or correlated with the rotating speed.

To achieve the above three specific sub-objectives, we have completed three research topics as follows

In Topic #1, an improved SVD-based method is proposed for gear tooth crack detection and severity assessment under constant speed conditions. The shortcomings of the reported RSVD method [57] were addressed. First, the improved SVD-based method adopts an auto-regression model-based baseline removal approach. SVD is applied to decompose the residual signal rather than the raw signal. Interference from non-fault related vibration components on the PMI is therefore eliminated. Second, the improved method selects the SCs such that the PMI of the reconstructed signal is maximized. Periodic impulse extraction is therefore strengthened. An experimental study was

conducted. Results show that the improved method outperforms RSVD in terms of fault detection and severity assessment without creating a considerable computational burden. The materials of this topic have been published partially in a conference paper [105] and as a whole in a journal paper [106].

In topic #2, a sparse FP-AR model that uses sparse AR terms and non-identical functional spaces is proposed. The limitations of the conventional FP-AR model [91] (i.e., the FP-AR model with its structure selected by the two-step BIC-based method) were addressed. To obtain such a sparse FP-AR model, a new model selection procedure is developed by adopting the least absolute shrinkage and selection operator. The sparse FP-AR model has been validated using simulation signals from a simulation model for a fixed-axis gearbox and experimental signals from two independent fixed-axis gearbox test-rigs. The modeling accuracy was measured by mean squared errors and randomness tests of the modeling residuals, goodness-of-fit between the one-step-ahead prediction and real gear vibration, and time-frequency spectra. Results have shown that the proposed sparse FP-AR model has higher modeling accuracy than the conventional one. Meanwhile, the sparse FP-AR model was applied for detecting gear tooth crack faults under identical VS profile. Results have shown that the proposed method benefits the fixed-axis gearbox in early detection of faults and better assessment of fault progressions. The materials of this topic have been published partially in two conference papers [90], [107]and as a whole in a journal paper [95].

In topic #3, a time series model-based method is presented for gear tooth crack detection and severity assessment under random speed variation. Specifically, the rotating speed and phase are considered as covariates in a VFP-AR model for representing impulsive vibration signals. We proposed refined B-splines for mapping the dependency between AR

coefficients and the rotating phase. The performance of the presented time series model-based method has been validated using laboratory signals. The presented method can assess 93.8% of the tooth crack severity state correctly, which is better than the novelty diagnostic method (73.8%) [76] and SVM-based method (87.7%) [75]. The materials of this topic have been documented in a journal paper [109] which is under review.

The contributions of this thesis are summarized as follows

1) Proposed an improved SVD-based method for gear tooth crack detection and severity assessment under constant speed condition. Validated the improved SVD-based method using laboratory signals from a gearbox test-rig. Benchmarked the performance of the proposed model with the reported RSVD method [57].

2) Proposed a sparse FP-AR model for the representation of baseline vibration from gearboxes under VS conditions. Validated the sparse FP-AR model-based fault detection and severity assessment for gearboxes under VS condition. Simulation signal and laboratory signals from two independent test-rigs are used to validate the performance of the proposed method. Benchmarked the modeling accuracy of the proposed model with the conventional FP-AR model [91].

3) Presented a time series model-based method for the gear tooth crack detection and severity assessment under random speed variation. Rotating speed and rotating phase are considered as covariates in a VFP-AR model for the representation of impulsive vibration signals under random speed variation. Refined B-splines are proposed for the VFP-AR model when mapping the dependency between AR coefficients and the rotating phase. Validated the proposed VFP-AR model-based method using laboratory signals from a gearbox test-rig, and benchmarked its performance with the reported novelty diagnostic method [76], and SVM-based method [75].

The rest of this thesis is organized as follows. Chapter 2, Chapter 3, and Chapter 4 give details of the three research topics, respectively. Chapter 5 presents the summary and future works.

This thesis is written using the paper-based template which meets the formatting requirements of the University of Alberta.

2

An improved singular value decomposition-based method for gear tooth crack detection and severity assessment

This chapter focuses on the gear tooth crack fault detection and severity assessment under constant rotating speed conditions (the research topic #1 as introduced in Section 1.3). An improved singular value decomposition-based method is proposed in this chapter. In Chapters 3 and 4, the problem of vibration-based fault detection and severity assessment under variable speed conditions will be addressed. The materials of this chapter have been published partially in a conference paper [105] and as a whole in a journal paper [106].

2.1 Introduction

Tooth cracks are among the top gearbox failure modes [13]. Tooth crack detection and severity assessment in the early stage of cracking may prevent catastrophic failure and facilitate condition-based maintenance strategies that reduce downtime and maintenance costs, which has great significance for industries [110]. Tooth cracks induce periodic impulses in vibration signals [12]. Hence, tooth crack detection and severity assessment usually rely on periodic impulse extraction. This chapter presents an approach that extracts periodic impulses from vibration signals for early tooth crack detection and severity assessment. Gearbox operation conditions, such as rotating speed and load torque, are assumed to be constant.

During recent decades, a variety of signal processing techniques have been presented for the periodic impulse extraction, including various statistical indicators [85], [87], band-pass filter and envelope spectrum [28], Wavelet transform-based de-noising [111], empirical mode decomposition [112], spectral kurtosis [113], [114], spectral correlation [115], [116], minimum entropy deconvolution [43], [47], [49], compressed sensing [117]–[119], and singular value decomposition (SVD)-based methods [57], [60], [120]. Among these techniques, the SVD-based methods have been widely used for periodic impulse extraction [57]–[62], [120] thanks to their following merits. First, SVD-based methods are non-parametric. Unlike other techniques such as wavelet analysis and compressed sensing, they do not require predefined basis functions [57]. Second, SVD-based methods have high computational efficiency [60]. Last, the signal components (SCs) decomposed from SVD are free from phase shift and waveform distortion [58].

As reviewed in Section 1.2.1.4, the idea behind most reported SVD-based methods [59], [61], [62], for periodic impulse extraction is to preserve the SCs with significant singular values (SVs). These reported methods are essentially energy-based, which can highlight high-energy components in the measured signal but may ignore weak-energy features caused by early faults [57]. To address this issue, Zhao and Jia [57] proposed a reweighted SVD (RSVD) method. This method applies SVD to decompose the raw signal into SCs, then uses a periodic modulation intensity (PMI) criterion innovatively to measure how influential the periodic impulses are in each SC. The PMI essentially represents the energy ratio between periodic impulses and other components in a signal. SCs are selected when their PMI exceeds a predefined threshold. Selected SCs are then added together with assigned weights to reconstruct the extracted signal. Simulation and experimental studies

in Ref. [57] show that RSVD is more effective than other SV-based methods in selecting the correct SCs and highlighting weak and informative features.

However, RSVD [57] suffers from two issues: 1) it ignores that non-fault related vibration components (e.g., gear meshing and structure vibration components) will affect the sensitivity of PMI to periodic impulses. These non-fault related vibration components exist when the gearbox is healthy. When the gearbox has a tooth crack, vibration signals contain both periodic impulses and these non-fault related vibration components. Hereafter, we consider the non-fault related vibration components equivalent to baseline vibration components. We have found that the PMIs of SCs that are obtained from baseline vibration are also high, leading to a high false alarm or miss alarm rate; 2) the reweighted summation of SCs undermines periodic impulse extraction. The RSVD selects SCs with PMIs that exceed the threshold and conducts reweighted summation of these SCs for signal reconstruction. The reweighted summation of these SCs often results in a signal with a smaller PMI compared with the original SCs. The reconstructed signals reflect fault severity poorly. Therefore, RSVD has a room for improvement in terms of fault detection and severity assessment.

This chapter proposes an improved SVD-based method that overcomes the above issues in two aspects. First, the proposed method adopts an auto-regression (AR) model-based approach [42] to remove the baseline vibration components from the raw signal. SVD is applied to a residual signal rather than the raw signal; interference from these baseline vibration components to the PMI is therefore eliminated. Second, the improved method selects the SCs such that the PMI of the reconstructed signal is maximized. The reconstructed signal is still a weighted summation of optimal SCs, but not all SCs whose PMI exceeds the threshold. Compared with RSVD [57], this new signal selection scheme

ensures that the reconstructed signal has a high energy ratio between tooth crack induced impulses and other components.

The remainder of this chapter is organized as follows: Section 2.2 presents fundamentals, including SVD, AR modeling, and PMI theories; Section 2.3 proposes the improved SVD-based method; Sections 2.4 and 2.5 validate the proposed method using simulation signals and experimental signals, respectively; concluding remarks are given in Section 2.6.

2.2 Fundamentals

This section introduces fundamentals that are utilized later in this chapter, including SVD, removal of baseline vibration components via AR modeling, and PMI criterion.

2.2.1 Singular value decomposition

SVD of an $m \times c$ real matrix $\mathbf{A}$ can be formulated as a factorization in the following formula [57], [58]

$$\mathbf{A} = \mathbf{U} \sum \mathbf{V}^T \tag{2.1}$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m] \in \mathbf{R}^{m \times m}$, and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c] \in \mathbf{R}^{c \times c}$ are unitary matrices. The column vectors within $\mathbf{U}$ and $\mathbf{V}$ are the orthonormal eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$, respectively; $\sum$ is a diagonal rectangular matrix that stores the singular values σ of $\mathbf{A}$ in descending order, which is given by $\sum = [\text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_l), \mathbf{0}] \in \mathbf{R}^{m \times c}$, where $l = \min(m, c)$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_l$.

A given one-dimensional signal $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ (e.g., a measured vibration signal) should first be reshaped into a matrix. Many matrix forms are available in the literature, including the Toeplitz matrix, cycle matrix, and Hankel matrix [57]. Among them, the Hankel matrix is widely used due to its zero-phase shift property. For $\mathbf{x}$, the Hankel matrix can be written as follows [60]

$$\mathbf{A} = \begin{bmatrix} x_1 & x_2 & \cdots & x_c \\ x_2 & x_3 & \cdots & x_{c+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_{m+1} & \cdots & x_N \end{bmatrix} \tag{2.2}$$

where $m = N\text{-}c+1$. The parameter m is used to determine the number of decomposed SCs, which is usually less than c.

According to Eq. (2.1), the Hankel matrix $\mathbf{A}$ can be expressed as the superposition of m sub-matrix $\mathbf{A}_i$, which satisfies

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_m^T \end{bmatrix} \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_m \mathbf{u}_m \mathbf{v}_m^T \\ &= \mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_m \end{aligned} \tag{2.3}$$

where $\mathbf{u}_i \in \mathbf{R}^{m\times 1}$ is the i^{th} column vector of the unitary matrix $\mathbf{U}$, $\mathbf{v}_i \in \mathbf{R}^{c\times 1}$ is the i^{th} column vector of the unitary matrix $\mathbf{V}$, and σ-matrix is the diagonal rectangular matrix $\sum$ storing the singular values σ of $\mathbf{A}$ in descending order. In Eq. (2.3), we assume that the σ-matrix is right augmented by 0 column vectors, and the final expression is valid as it is.

Each sub-matrix $\mathbf{A}_i$ in Eq. (2.3) corresponds to an SC $\mathbf{x}_{ci}$ in the raw signal $\mathbf{x}$. According to the construction of the Hankel matrix given in Eq. (2.2), $\mathbf{x}_{ci}$ can be obtained from $\mathbf{A}_i$ by averaging along the anti-diagonals, as shown in Fig. 2.1.

$$\mathbf{A}_i = \begin{bmatrix} x_{ci,1} & x_{ci,2} & x_{ci,3} & \cdots & x_{ci,c} \\ x_{ci,2} & x_{ci,3} & x_{ci,4} & \cdots & x_{ci,c+1} \\ x_{ci,3} & x_{ci,4} & x_{ci,5} & \cdots & x_{ci,c+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{ci,m} & x_{ci,m+1} & x_{ci,m+2} & \cdots & x_{ci,N} \end{bmatrix}$$

Fig. 2.1: Extraction of $\boldsymbol{x}_{ci}$ from $\mathbf{A}_i$ by anti-diagonals averaging method

The SVD-based signal decomposition can be understood as a filter bank approach. The obtained SCs are outputs of filters with different frequency bands. The frequency bands are unknown ahead of time and adaptive to the signal. Also, the frequency bands of different filters may have overlap, which means different SCs may contain the same frequency components.

When it comes to extracting the periodic impulses, the carrier frequency of the periodic impulses equals to the natural frequency of the transmission path [12], [24], [25] . Thus, these decomposed SCs with their frequency bands included or close to the natural frequency will contain the periodic impulses. Note that there may be more than one SCs which contain periodic impulses.

2.2.2 Removal of baseline vibration components via auto-regression modeling

Baseline vibration components are to be removed to reduce their interference to the PMI. Many methods are available for fulfilling this purpose, including the time series model-based approach [42], frequency domain editing, and time domain subtraction [16]. In this chapter, we use the time series model-based approach given that it does not require any prior information like gear meshing frequency, and is free from requiring rotating phase alignment that is required by time domain subtraction methods. The time series model-based approach identifies a time series model, e.g., an auto-regression (AR) model, to represent the baseline vibration from a gearbox; constructs an inverse filter based on the baseline model; and uses the inverse filter to remove the baseline vibration from signals collected when the health condition of the gearbox is to be determined. A “residual signal” is then obtained. Note that the baseline data is better to be collected at multiple time points to account for any variations (e.g., environmental temperature).

The AR model is a typical time series model, which is expressed by the formula [42]

$$x_t = \sum_{i=1}^{n_a} a_i x_{t-i} + \varepsilon_t \tag{2.4}$$

where x_t and x_{t-i} denote the data points at time t and time t-i, respectively, of a time series (e.g., a measured vibration signal); n_a specifies the order of the AR terms; ε_t is a zero-mean white Gaussian noise at time t; a_i represents the AR parameters. Identifying a suitable AR model requires parameter estimation and model order selection. Model parameters are estimated using the ordinary least squared estimator [108]. The model order is selected by minimizing Bayesian information criteria [115].

Applying z-transform [121] to Eq. (2.4), we obtain

$$X(z) = \frac{1}{1 - \sum_{i=1}^{n_a} a_i z^{-i}} e(z) = H(z) e(z) \tag{2.5}$$

where $X(z)$ is the vibration signal expressed in the z-domain and $e(z)$ is the zero-mean white noise in the z-domain. According to "linear time-invariant (LTI) system" theory, X(z) is modeled as the output of an LTI system with transfer function $H(z)$ excited by a zero-mean white noise $e(z)$.

Based on Eq. (2.5), the inverse filter is of the following transfer function [121]:

$$H^{-1}(z) = 1 - \sum_{i=1}^{n_a} a_i z^{-i} \tag{2.6}$$

This inverse filter, when applied to a collected vibration signal, can remove the baseline vibration components to obtain a residual signal.

2.2.3 Periodic modulation intensity

According to [57], the value of PMI represents the energy ratio between tooth crack induced impulses and other components in a given signal x_t (e.g., a measured vibration signal or SC). The analytical envelope Env_t of the x_t is obtained as

$$\mathrm{Env}_t = abs\left[x_t + i_m H_{\mathrm{bt}}\left(x_t\right)\right] \tag{2.7}$$

where i_m is the imaginary unit, and H_{bt} represents the Hilbert transform. The mean of Env_t is then removed from the time series Env_t to acquire the zero-mean envelope. Finally, the PMI value p is calculated from this zero-mean envelope as follows.

$$p = \frac{R_{\mathrm{Env}}\left(\tau\right)}{R_{\mathrm{Env}}\left(0\right) - R_{\mathrm{Env}}\left(\tau\right)} \tag{2.8}$$

where τ is the period of tooth crack fault induced impulses; $R_{\mathrm{Env}}(\tau)$ is the autocorrelation of the zero-mean envelope at τ; $R_{\mathrm{Env}}(0)$ is the autocorrelation at $\tau = 0$ of the zero-mean envelope, which is the total energy of the zero-mean envelope.

2.3 Improved singular value decomposition-based method

The improved SVD-based method consists of the following steps for tooth crack fault detection and severity assessment. Fig. 2.2 shows a flowchart of the improved SVD-based method.

Step 1: Removal of baseline vibration components via AR modeling. In this step, an AR model is first identified using baseline vibration signals. Specifically, baseline vibration signals are used to select the AR model order and estimate the AR model parameters. The inverse filter is then constructed. For signals collected when the health condition of the gearbox is to be determined, applying the inverse filter removes the baseline vibration

components and yields a residual signal. With the baseline vibration component removed, interference from baseline vibration components to the PMI can be eliminated.

Step 2: SVD-based decomposition of the residual signal. SVD decomposes the residual signal into m SCs. The selection of m is critical to the success of SVD [122]. Miao *et al.* [122] empirically assigned m as 100 for decomposing vibration signals from rotating machinery after considering the tradeoff between computational efficiency and effectiveness of SVD. We follow their research to select m as 100.

Step 3: Information evaluation. The PMI that measures how influential the periodic impulses are is calculated for each SC.

Step 4: Fault detection. If any SC has the PMI value that exceeds a predefined threshold T_h, a tooth crack fault is detected. Otherwise, no fault exists. In such case, Step 5 is skipped, and the process ends. We determine T_h by decomposing residual signals from baseline vibration signals. By extracting the maximum PMI of SCs from multiple segments of the residual signal, a statistical distribution for the maximum PMI can be fitted. We can set T_h at a certain percentile of the obtained statistical distribution. Too low the percentile will lead to a high false alarm rate, whereas too high the selected percentile will lead to a high miss alarm rate. For instance, a 98$^{\text{th}}$ percentile of the distribution means that the allowable limit of the false alarm rate is <2%. The determination of T_h may vary for different applications, and it is up to the users to determine which percentile should be selected for their application. In this thesis, we set T_h as the 98$^{\text{th}}$ percentile of the distribution for illustration purposes.

Step 5: Signal reconstruction. We proposed a new signal reconstruction scheme. The scheme first pre-selects the SC(s) with a PMI value that exceeds threshold T_h. Suppose p_i is the PMI value of i^{th} SC, and $S_{\text{ini}} = \{i \mid p_i > T_h, i \in (1, 2, \ldots, m)\}$ represents the set for pre-selected

initial SCs. The new signal reconstruction scheme determines a subset $S \in S_{ini}$ such that the PMI value p_d of the reconstructed signal $\mathbf{x}_d = \sum_{i \in S} p_i \mathbf{x}_{ci} \Big/ \sum_{i \in S} p_i$ is maximized:

$$S = \arg\max_{S \in S_{\text{ini}}} \{ p_d \} \tag{2.9}$$

The above integer optimization problem can be solved using the forward selection scheme [100]. The reconstructed signal is then used for severity assessment. Generally, the higher the energy ratio between tooth crack induced impulses and other components, the more severe the fault [12].

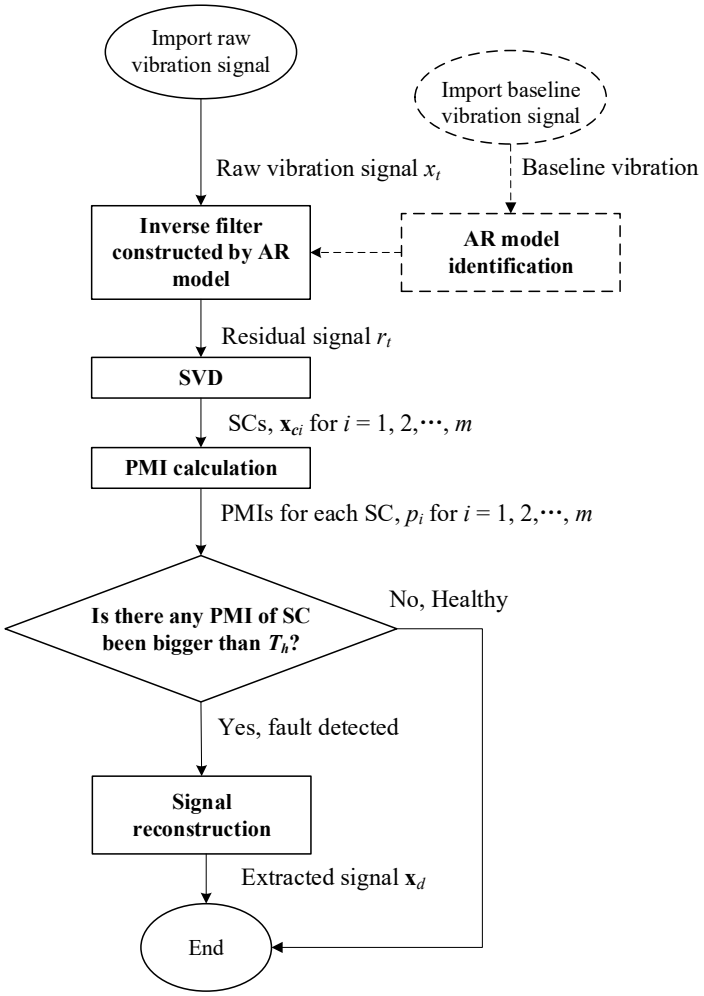

Fig. 2.2: Flowchart of the improved SVD-based method

Compared with RSVD [57], step #1 is newly adopted. Step #5 is also newly proposed and replaces the direct reweighted summation of all SCs with a PMI that exceeds threshold T_h.

When it comes to the online implementation of the proposed SVD-based method, users need to follow the steps 1~5 described above. However, the baseline AR model in step #1 and the fault detection threshold T_h in step #4 need to be determined ahead of the online implementation. In other words, these parameters need to be determined offline.

2.4 Experimental study

This section presents an experimental study with comparative analysis for evaluating the fault detection and severity assessment effectiveness and computational cost of the improved SVD-based method.

2.4.1 Experimental signals

This experiment was designed and carried out by the candidate with the help of other members of Dr. Zuo's research group. The experiment was completed in December 2018 and details of the experiment were documented in a technical report [123].

A gearbox test rig in the Reliability Research Lab at the University of Alberta, Canada, was used. Fig. 2.3(a) shows the gearbox test rig which consists of a drive motor, bevel gearbox, 1st stage planetary gearbox, 2nd stage planetary gearbox, 1st stage speed-up gearbox, 2nd stage speed-up gearbox, and load motor. We choose the 2nd stage speed-up gearbox as the target of our experiment. Fig. 2.3(b) shows the schematic of the target gearbox which has two spur gear mesh pairs. The gear and pinion of the first mesh pair have 48 and 18 teeth, respectively, whereas the gear and pinion of the second mesh pair have 64 and 24 teeth, respectively. The pinion of the first mesh pair was selected as the object gear with its

corresponding specification listed in Table 2.1. Fig. 2.3 (c) shows the locations of four accelerometers (1000 mV/g sensitivity) labeled as 1, 2, 3, and 4.

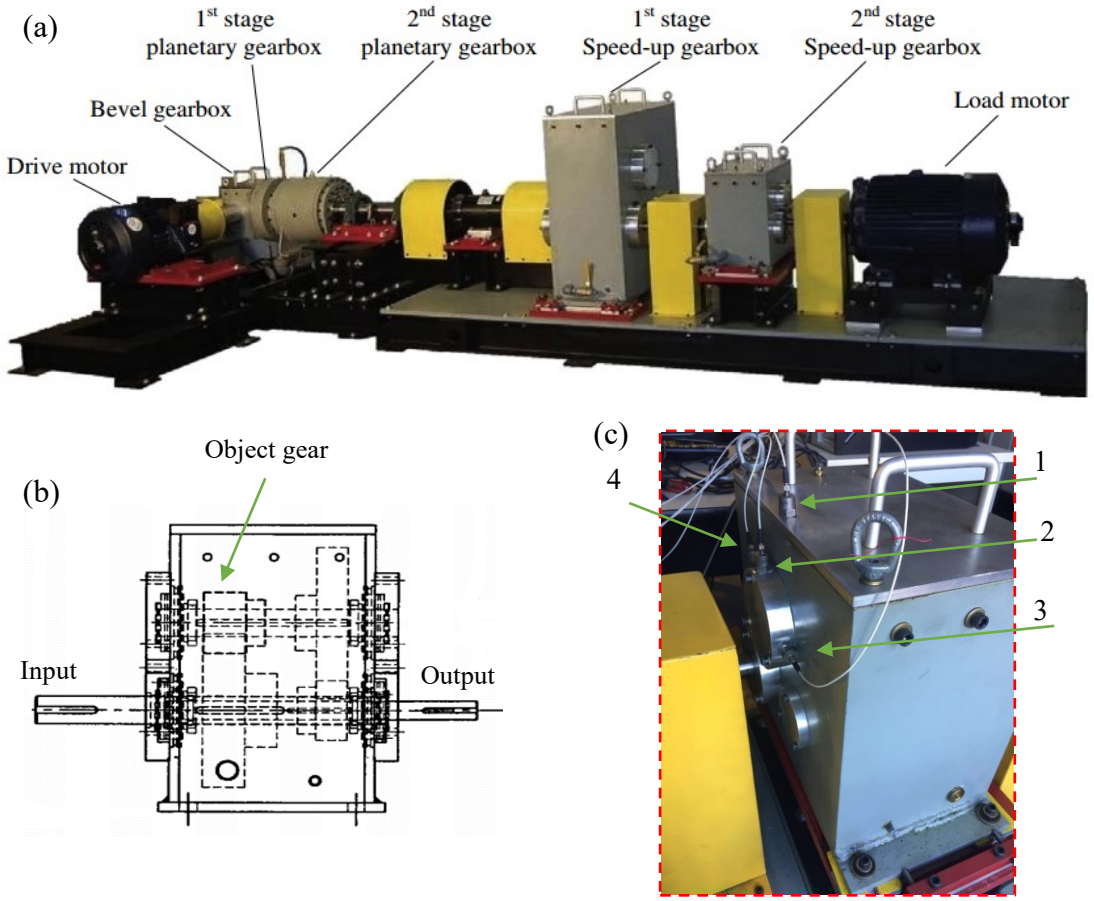

Fig. 2.3: Experiment setup: (a) gearbox test rig, (b) schematic of the 2nd stage speed-up gearbox, (c) four sensor locations [123]

Table 2.1: Object gear specifications [123]

Parameters	**Values**
Face width (Inches)	2.0
Diametral pitch (DP)	6
Pressure angle (Deg)	20
Pitch diameter (Inches)	3.0
Num of teeth	18

Artificial cracks were seeded at the tooth root. Fig. 2.4 shows the artificial crack that is expected to propagate along the gear face width and crack depth simultaneously after its initiation [124]. In Fig. 2.4, w denotes gear tooth face width, w_o denotes the width of tooth crack, q denotes a half-length of the tooth chordal thickness (e.g., circular tooth thickness), q_0 denotes the depth of tooth crack, and α_c is the crack angle. An electric discharge machine was used to induce such a crack. The cutter has a thickness of 0.1 mm. The spatial shape of the resulting artificial crack is a cubic of size 0.1 mm $\times w_o \times q_0$, and the tip of the crack is not as sharp as shown in Fig. 2.4. Five levels of tooth cracks as listed in Table 2.2 were considered, as well as the healthy condition. Fig. 2.5 shows the photos of the manually damaged gears. For more details on the experiment design, see [123].

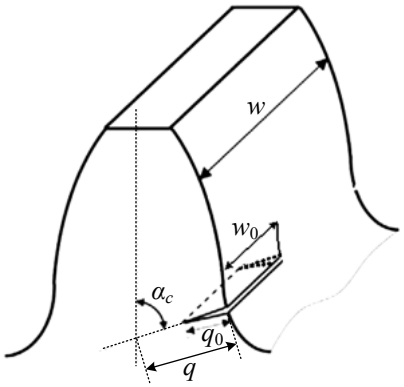

Fig. 2.4: Gear tooth with root crack [123]

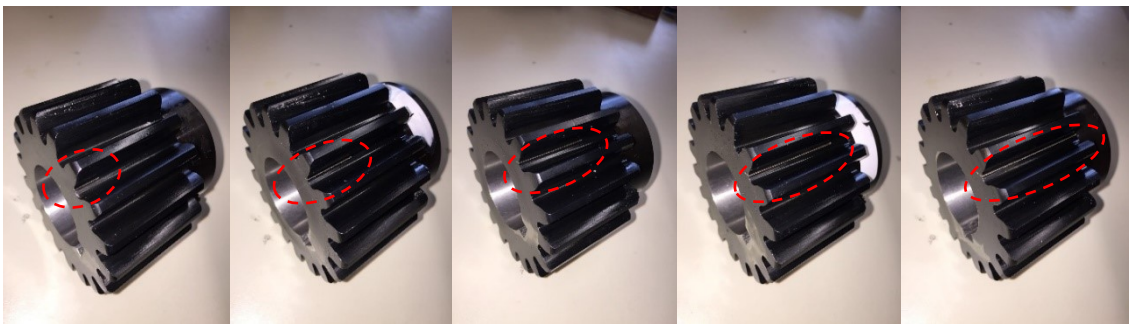

Fig. 2.5: Photos of the manually damaged gears, from left to right corresponding to F1 to F5 [123]

Table 2.2: Tooth crack fault configurations [123]

Fault level	Parameters
F1	w_0 = 0.2w; q_0 = 0.2q; α_c = 60°
F2	w_0 = 0.4w; q_0 = 0.4q; α_c = 60°
F3	w_0 = 0.6w; q_0 = 0.6q; α_c = 60°
F4	w_0 = 0.8w; q_0 = 0.8q; α_c = 60°
F5	w_0 = 1.0w; q_0 = 1.0q; α_c = 60°

Experimental signals were collected under three categories of operating conditions, namely constant condition, variable speed condition, and variable load condition. Under the constant condition, both the speed and load are fixed. Under the variable speed condition, the speed is time varying and load fixed. Under the variable load condition, the speed is fixed and load time varying. For more details on the operating conditions, see [123]. In this chapter, only the experimental signals collected under the constant condition will be used. The experimental signals collected under variable speed condition will be used in Chapter 3.

Under the constant condition, the object gear runs at a constant speed of 2.953 Hz and a constant load of 80 N.m. Sampling frequency f_s needs to be higher than 20 times of the gear meshing frequency to cover the 10th order harmonic of the gear meshing frequency [125]. As the gear meshing frequency f_m = 53.154 Hz (i.e., 2.953 Hz × 18), we set f_s = 6.4 kHz which is sufficiently high. Experimental signals under the healthy condition and the F2, F3, F4, F5 levels of tooth cracks were used in this chapter. Without loss of generality, the signals under F1 level of tooth crack are omitted given that the crack induced signature is too week. The duration of the signal should last 10 cycles of the object gear. Therefore, we set the duration as 3s. In total, 50 segments of vibration signals were collected under each health condition, and each segment has 19.2 k data points. For more details of the data collected from this experiment, see [123].

Without loss of generality, we analyze the vibration signals from the vertical accelerometer installed on the bearing cap (Sensor #2). Fig. 2.6 shows the experimental signals under each health condition. As shown in Fig. 2.6(b), the observable high-energy components are primarily the gear meshing frequency and its harmonics. The periodic impulses are not observable and may be located anywhere in the frequency spectrum. We cannot detect the faults by observing the time waveform and frequency spectrum. Tooth crack induced features were submerged into the vibrations.

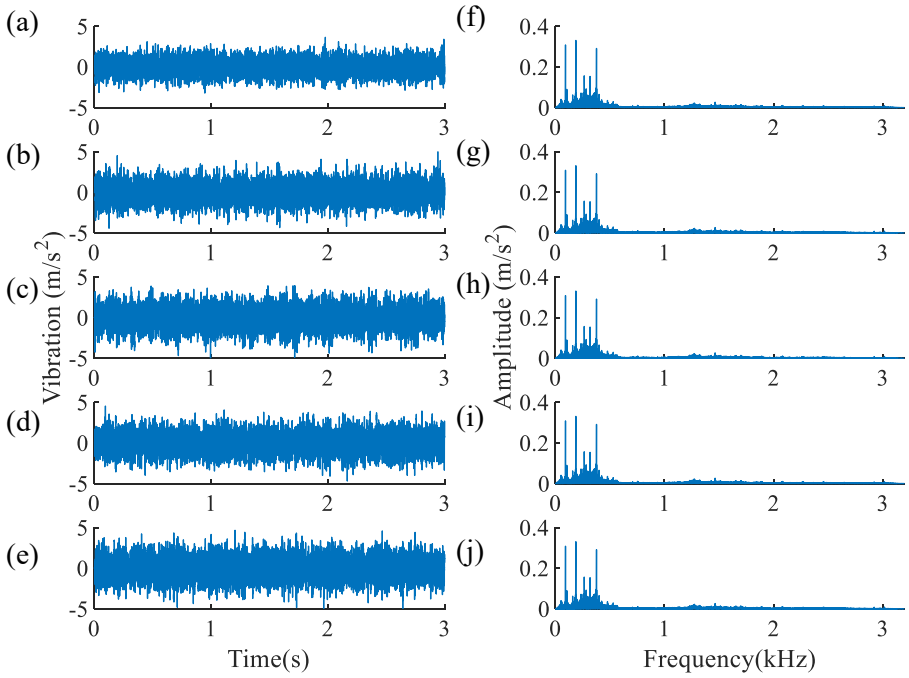

Fig. 2.6: Experimental signals: (a)-(e) time-domain waveforms under the Baseline (B), F2, F3, F4, and F5 conditions, respectively; (f) -(j) frequency spectrums under the B, F2, F3, F4, and F5 conditions, respectively

2.4.2 Performance on fault detection

We implemented the improved SVD-based method to process experimental signals. During Step #1, an AR model with n_a = 249 was identified using a baseline vibration signal. Fig. 2.7 shows the obtained residual signals in time and frequency domains. By comparing the residuals and raw signals in the frequency domain (i.e., Fig. 2.7(f~j) versus Fig. 2.6(f~j),

respectively), we can see that the spikes (mostly in the frequency range of 0-1 kHz) of the raw signals have been significantly attentuated after the AR inverse filter. The amplitude of these specturm spikes is around 0.3 m/s² in the raw signals (i.e., Fig. 2.6 (f~j)), but has been reduced to around 0.05 m/s² in the residual signals (i.e., Fig. 2.7 (f~j)) after the AR inverse filter, becoming close to the amplitude (0.03 m/s²) of the noise. The frequency spectrum of the residuals (i.e., Fig. 2.7(f~j)) becomes distributed in borad frequency band, rather than concentrated in the 0-1 kHz region. These indicate that the baseline vibration components were mostly removed. We cannot detect the faults or differentiate the severities by observing residual signals in time and frequency domains. Further steps as given in Section 2.3 need to be implemented.

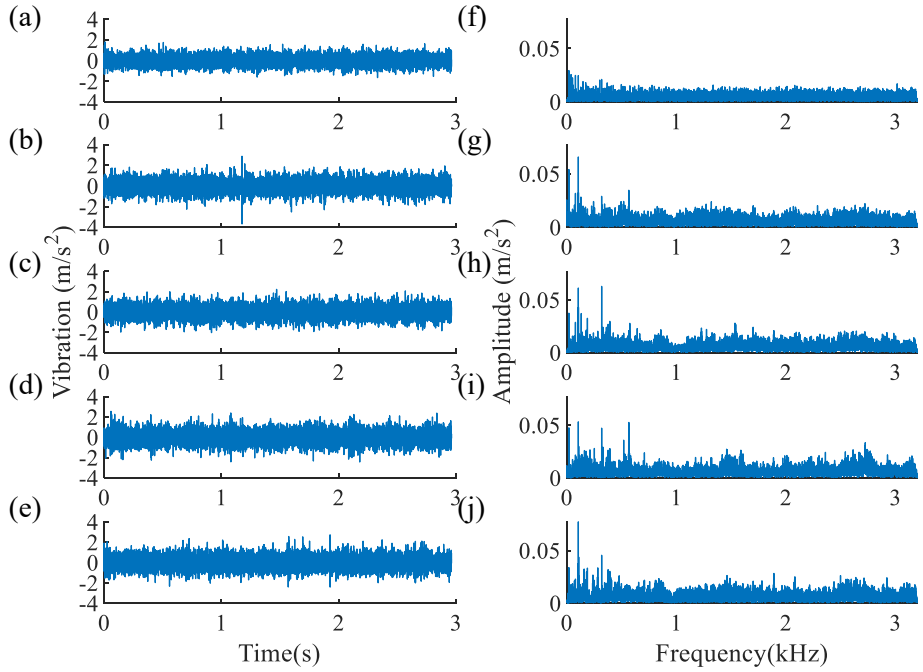

Fig. 2.7: Residuals of experimental signals: (a)-(e) time-domain waveforms under the B, F2, F3, F4, and F5 conditions, respectively; (f)-(j) frequency spectrums under B, F2, F3, F4, and F5 conditions, respectively.

During Steps #2~3, SVD was applied to decompose each residual signal and obtain m = 100 SCs. Fig. 2.8 shows the first 20 SCs with high PMI values. The remaining 80 SCs were

not plotted due to page limits. The amplitude of the time waveform is normalized. During Step #4, a fault was detected if any SC had a PMI value that exceeded the predefined threshold T_h. The T_h is set as 0.4090, which will be explained later. The T_h is shown as a dashed line in Fig. 2.8. The improved method detected 4 levels of faults successfully, including the earliest fault F2 and did not falsely detect the baseline. As shown in Fig. 2.8(b), two SCs from the F2 condition have PMIs that exceeded T_h. Such number of SCs are 2, 5, 3 for F3, F4, and F5, respectively. It is interesting to note that the number of SCs which have higher PMIs than the threshold T_h does not monotonically increase when the fault severity increases. In the future beyond this thesis, further investigations are to be conducted to discover the underlying reason.

For comparison, the RSVD presented by Zhao and Jia [57] was implemented. RSVD applies SVD directly on the raw signals. Fig. 2.9 shows the first 20 SCs with high PMI values obtained by applying SVD on raw signals. The RSVD used a threshold of $T_h = 1$ for fault detection based on analyzing the vibration signals of rotating machinery [57], as the dashed line in Fig. 2.9 indicates. To this end, no fault was detected for any of the 4 fault levels. Alternatively, after processing multiple segments of baseline vibration signals, the threshold T_h was set as 0.6253 for fault detection. Using $T_h = 0.6253$ as the dashed line in Fig. 2.9 indicates, the RSVD detected F3 and F5 fault levels. In conclusion, using either $T_h = 1$ as used by Zhao and Jia [57] or $T_h = 0.6253$ works comparably worse than the improved method for fault detection. We were unable to find a T_h that meets the fault detection requirements when implementing RSVD.

Next, we examined the repeatability of the improved method when analyzing multiple experimental signals. The improved method was used to process 50 segments of experimental signals under each health condition.

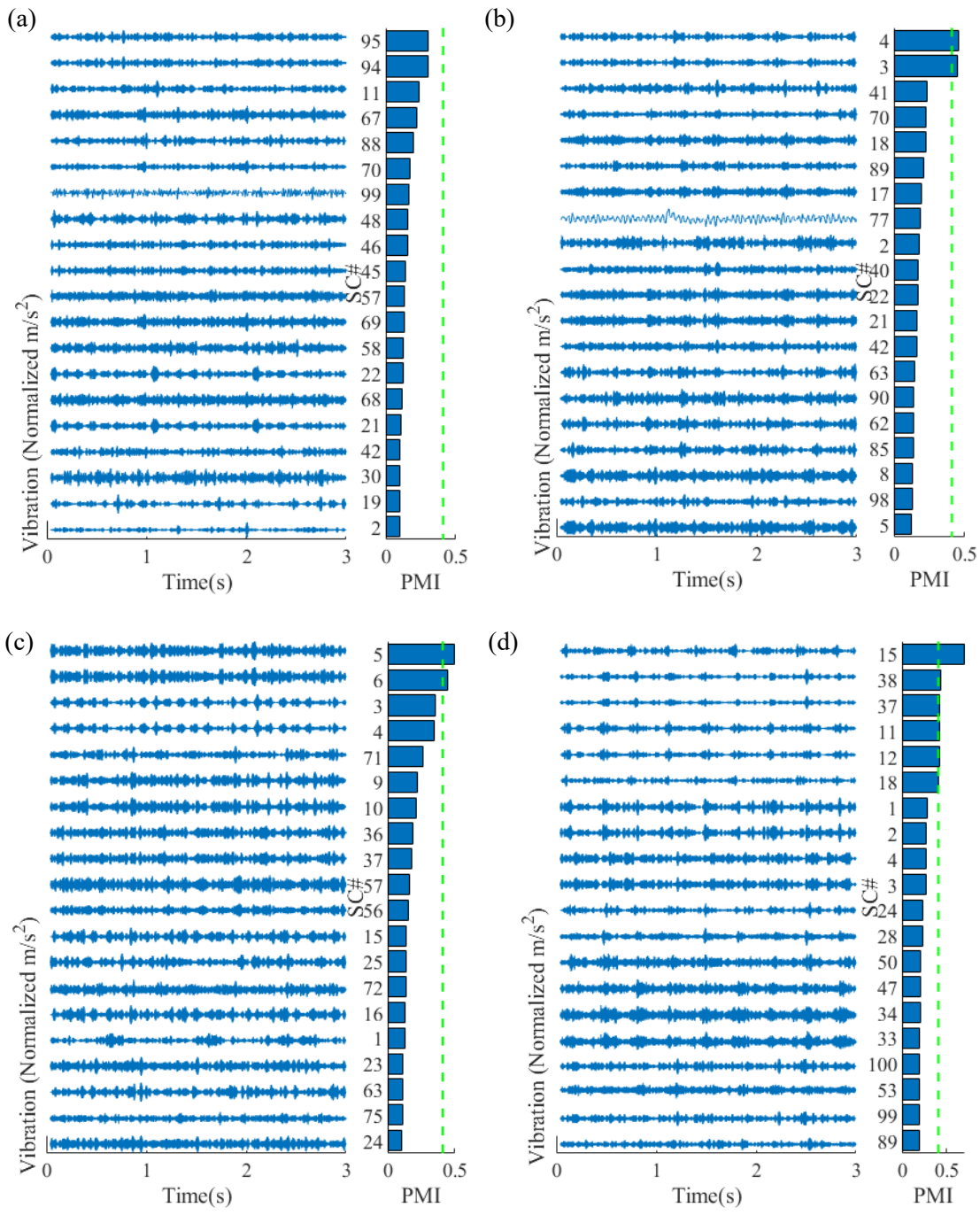

Fig. 2.8: Signal components decomposed from the residuals of experimental signals (a) B, (b) F2, (c) F3, (d) F4, and (e) F5.

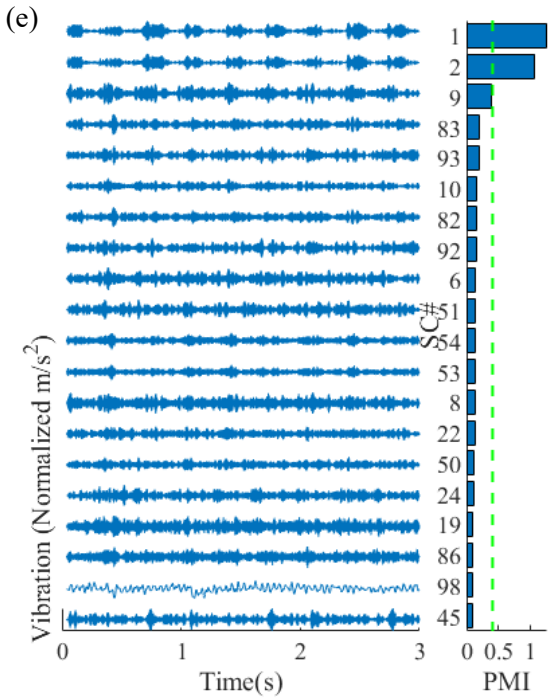

Fig. 2.8 (Continued): Signal components decomposed from the residuals of experimental signals (a) B, (b) F2, (c) F3, (d) F4, and (e) F5.

The threshold T_h for fault detection was determined by processing multiple segments of baseline vibration signals, as mentioned in Section 2.3. The maximum PMI of SCs decomposed by the improved method was extracted for each segment under baseline condition, as shown in Fig. 2.10 (a). We can observe that the distributions of the PMIs are skewed to the right (positive side). Based on how PMI is calculated as given in Eq. (2.8), $R_{Env}(\tau) \in [-R_{Env}(0), R_{Env}(0)]$ which leads to the PMI value $p \in [-0.5, +\infty]$. The upper boundary of ∞ and the lower boundaries of -0.5 of the PMI explain why the distribution of the PMIs is skewed to the right. The T_h was set to be 0.4090, which is the 98th percentiles of the fitted lognormal distribution. As for the RSVD method, the maximum PMI is shown in Fig. 2.10(b) and T_h was set to be 0.6253. Note that the obtained threshold is applicable to the studied gearbox only. When applying our improved SVD-based method to another gearbox, it is necessary to collect sufficient vibration signals; complete the baseline

removal (Step #1), SVD decomposition (Step #2), and information evaluation (Step #3); fit a distribution; obtain a threshold for the gearbox.

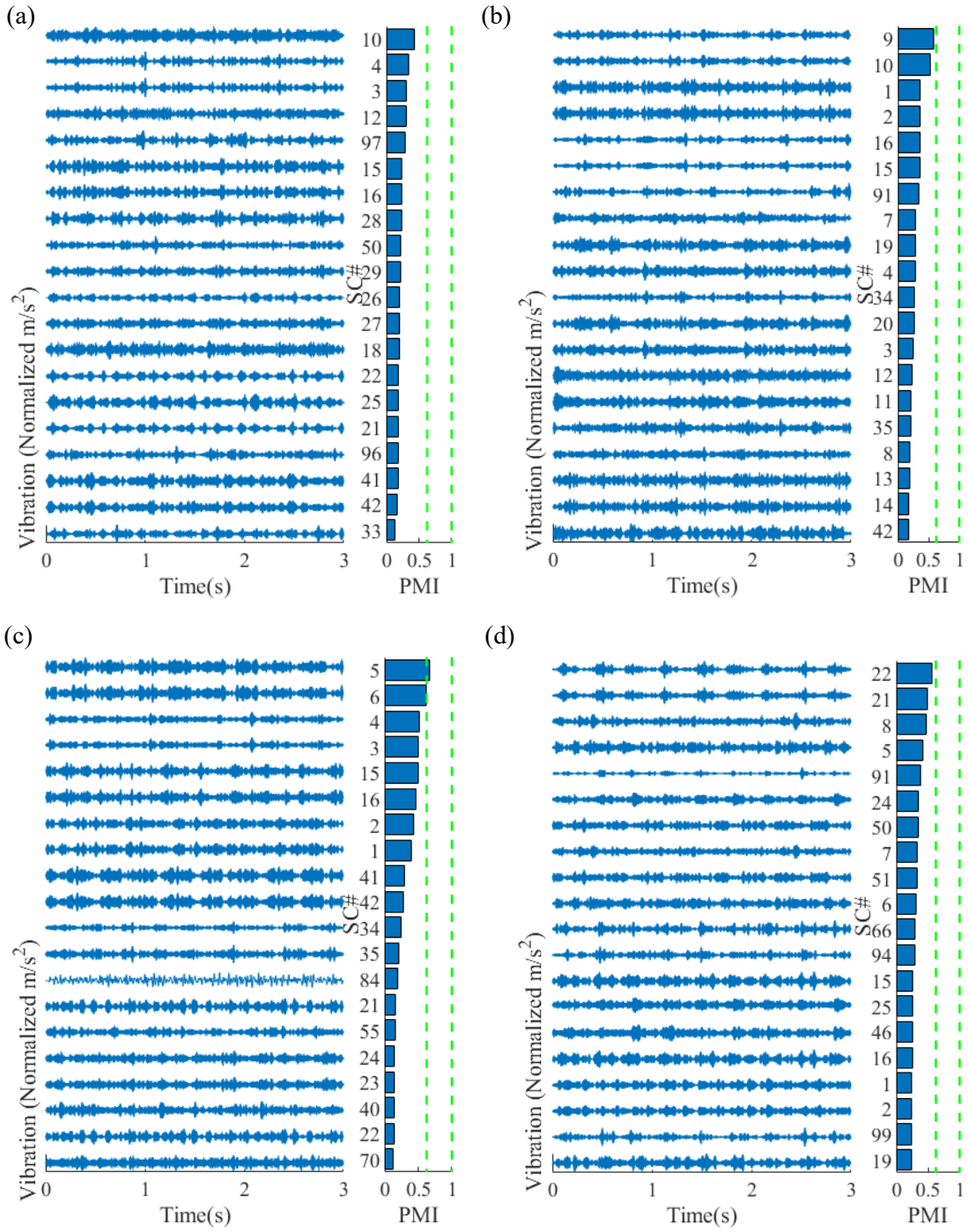

Fig. 2.9: Signal components decomposed from raw experimental signals (a) B, (b) F2, (c) F3, (d) F4, and (e) F5.

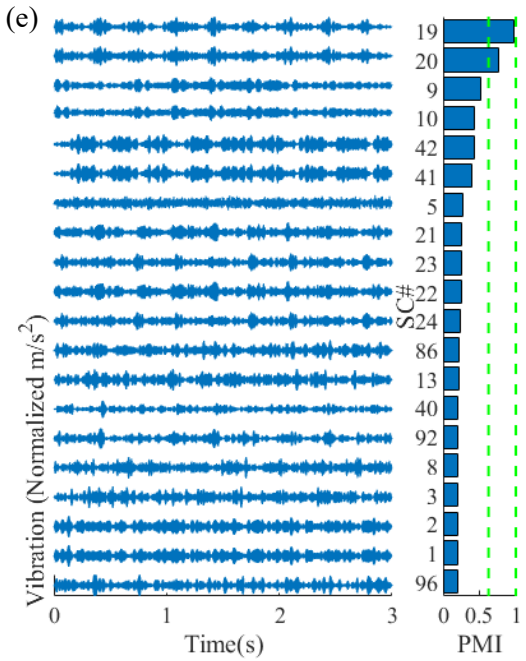

Fig. 2.9(Continued): Signal components decomposed from raw experimental signals (a) B, (b) F2, (c) F3, (d) F4, and (e) F5.

With the threshold T_h determined, the fault detection rate was obtained as listed in Table 2.3. The improved method detects F2 at a rate of 20%, F3 at 46%, F4 at 96%, and F5 at 100%. A 2% false alarm rate was incurred since we have set the fault detection threshold as the 98% percentile of the statistical distribution. By contrast, RSVD missed the 4 fault levels when using T_h = 1 whereas detects the F2~F5 at a lower rate when using T_h = 0.6253. RSVD performed worse than the improved method when analyzing the experimental signals.

Table 2.3: Fault detection rate

Fault levels	**B**	**F2**	**F3**	**F4**	**F5**
Improved method	2%	20%	46%	96%	100%
RSVD using T_h = 1 [57]	0%	0%	0%	0%	0%
RSVD using T_h = 0.6065	0%	2%	34%	18%	50%

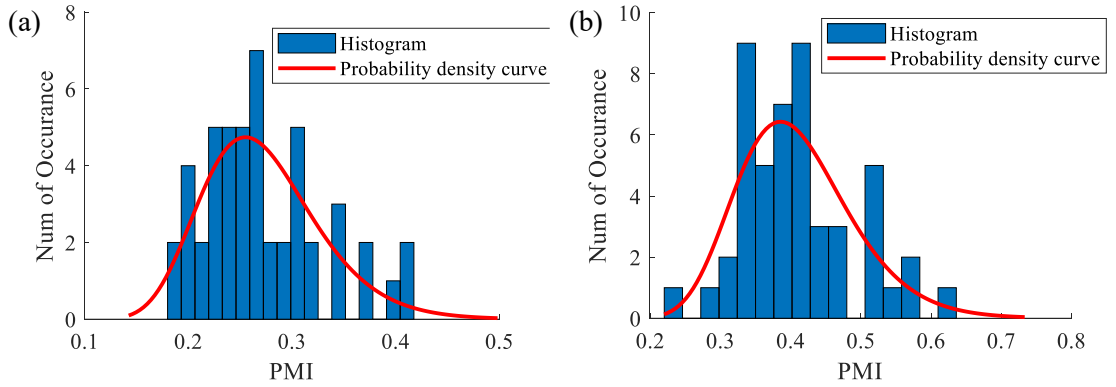

Fig. 2.10: Maximum PMI of SCs that were decomposed from baseline experimental signals: (a) decomposed from residual signals; probability density curve ~ Lognormal(μ = 1.3224, σ = 0.2086); (b) decomposed from raw signals; probability density curve ~ Lognormal(μ = -0.9125, σ = 0.2008).

In the following, we provide our interpretation of why the RSVD presented by [57] performed worse than the improved method when processing our vibration signals. As shown in Fig. 2.9(a), SC #10 under the baseline condition does not show explicit periodic impulses but does have a relatively high PMI of 0.43. It is apparent that the PMI is sensitive not only to periodic impulses, but also to other sources. To study how the sensitivity of PMI to periodic impulses was affected, SC #10 decomposed from the raw baseline vibration signal was extracted. Fig. 2.11(a) shows the time-waveform of the extracted signal, and Fig. 2.11(b) the frequency spectrum. The obvious frequencies in Fig. 2.11(b) are the gear meshing harmonics $2\times f_m$ = 106.31Hz and $8\times f_m$ = 425.23 Hz. No periodic impulses are visible. Fig. 2.11(c) and (d) show the envelope in the time and frequency domain, respectively. Taking the envelope demodulates SC #10. The envelope has a dominated component $6\times f_m$ (i.e., $8\times f_m$-$2\times f_m$), as shown in Fig. 2.11(d). The normalized autocorrelation (i.e., $R_{Env}(t)/R_{Env}(0)$) of the zero-mean envelope, as shown in Fig. 2.11(e), consequently shows multiple peaks with an occurrence period of 0.0031s = $1/(6\times f_m)$. The normalized autocorrelation equals to 0.3007 at the period of tooth crack fault induced impulses, as indicated by the dashed line in Fig. 2.11(e, f). The PMI value is

calculated as 0.3007/(1-0.3007) = 0.4300 based on Eq. (2.8). The above analysis reveals that baseline vibration components can interfere the PMI. Such interference disables RSVD from differentiating the baseline condition and faulty conditions. By contrast, the improved method removes baseline vibration components. Interference of baseline vibration components on the PMI is eliminated. Thus, the improved method outperforms RSVD on fault detection when processing our vibration signals.

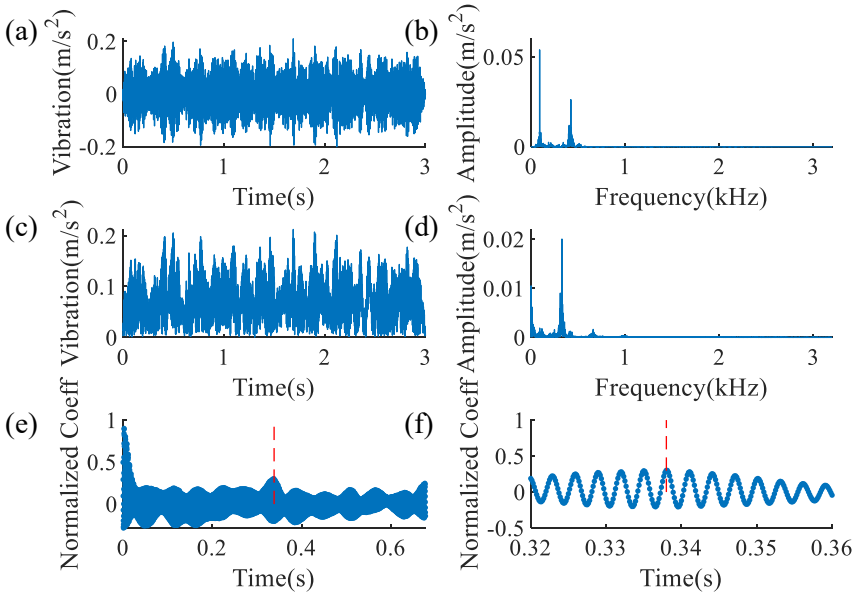

Fig. 2.11: SC #10 decomposed from the baseline raw experimental signal: (a) time waveform, (b) frequency spectrum, (c) analytical envelope, (d) frequency spectrum of the envelope, (e) autocorrelation of the zero-mean envelope residual, and (f) zoom-in plot of (e) in [0.32 0.36]s.

2.4.3 Performance on severity assessment

Severity assessment involves Steps #1-#5 of the improved method. Severity is assessed when a fault is detected. The signal is reconstructed via the new scheme and used for severity assessment. Take the F4 condition as an example of the signal reconstruction. Pre-selection of the SCs results in an initial set of S_{ini} = {15, 38, 37, 11, 12, 18}. Then, the forward selection scheme returns a set of S = {15} such that the PMI of the reconstructed signal is maximized. The reconstructed signal is the same as SC #15 and has a PMI p_d of

0.7084. Fig. 2.12 shows the reconstructed signals for each fault condition in the time and frequency domain. PMI values of these reconstructed signals show an ascending trend that reflects fault severities. The fault induces periodic impulses with an occurrence period of ~0.338 s (i.e., $\tau = 1/f_r$ where f_r is the rotating frequency). The frequency spectrum of the extracted signal under F2 concentrated on a natural frequency region around 0.5 kHz, whereas the frequency spectrum of F2 on 0.3 kHz, and the frequency spectrum of F4 and F5 on 0.1 kHz. Such variation of the natural frequency regions may occur because the impulses excitation under different damage level has different frequency band. In addition, our study is concerned about the natural frequency of the system, included gears, shaft, bearing, pillow block, and shell. A gearbox could have a natural frequency lower than 1kHz from the observations in Mbarek *et al.* 2017 [126] and Guo *et al.* 2014 [127]. Besides, the studied gearbox is comparably big in size, and its components are big in size and thus with smaller natural frequencies.

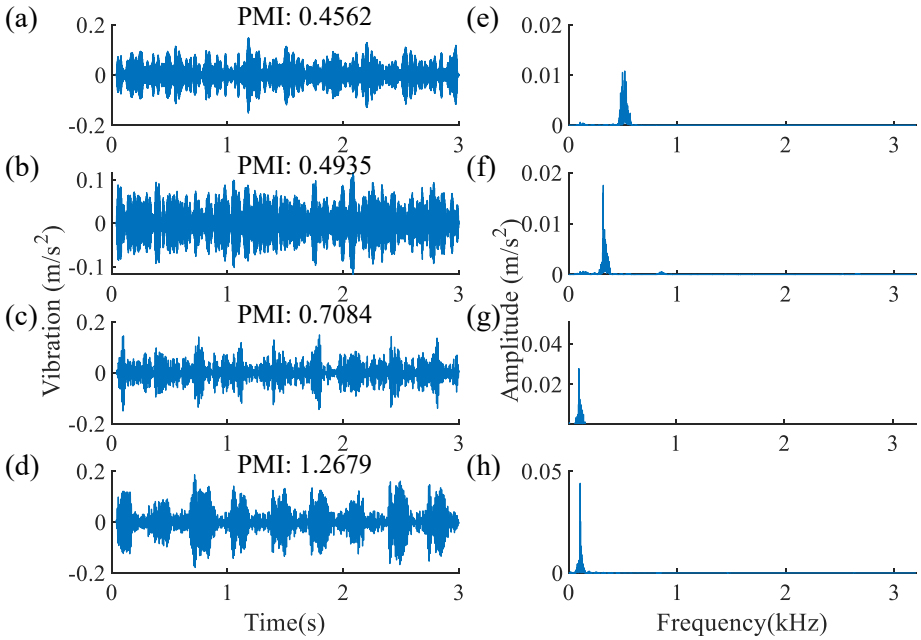

Fig. 2.12: Reconstructed signals via the improved method: (a)-(d) time waveforms under the F2, F3, F4, and F5 conditions, respectively; (e)-(h) frequency spectrums under the F2, F3, F4, and F5 conditions, respectively.

For comparison, the RSVD reported in [57] was implemented. RSVD conducted a reweighted summation of all SCs with PMI value exceeds the threshold for signal reconstruction. Here, we conducted the reweighted summation on the SCs decomposed from residual signals because RSVD did not detect faults when decomposing the raw signals. Take the F4 condition as an example to show the reweighted summation. The reconstructed signal is a reweighted summation of SC# 15, 38, 37, 11, 12, 18, and has a PMI of 0.2437. Fig. 2.13 shows the reconstructed signals under each tooth crack condition in time and frequency domains. PMI values of these reconstructed signals do not show a monotonous ascending trend. The PMI value of F4 case is the smallest. This implies weak severity separability. The frequency spectrum of the extracted signal under the F2 condition concentrated on two natural frequency regions around 0.1 kHz and 0.5 kHz, whereas the frequency spectrum of F3 on 0.3 kHz; the frequency spectrum of F4 on three natural frequency regions around 0.1 kHz, 0.5 kHz, and 3 kHz; and the frequency spectrum of the extracted signal under F5 on 0.1 kHz. RSVD aggregated the signal components on natural frequencies and constructed a signal with a small PMI. The reconstructed signals are poorly reflective of fault severities.

The repeatability of the method was also studied. Fig. 2.14(a) shows the boxplot of PMI value for signals reconstructed via the improved method. Again, the PMI at various fault levels shows a good ascending trend that reflects the fault severities. Under the influence of noise, PMIs overlapped between neighbour fault levels. To numerically show the separability of the 4 fault levels, we trained a linear discriminate analysis (LDA) model [100] with the PMI as input and the class label as output. The number of extracted signals under the F2, F3, and F4 health conditions did not equal to 50 due to the fault detection rate. If we directly train an LDA model for classification, these health conditions with

fewer samples tend to be misclassified more often than the F5 health condition with more samples [128]. Thus, we resampled the PMI of the extracted signals under such class (similar to the idea of bootstrap [100]) to equalize the sample numbers. The data was then randomly partitioned into training and testing datasets where the training set holds 75%, and the testing set 25%. Here, we choose the 75%-25% data partition ratio for illustration purposes. Users may choose other ratios to achieve the best classification accuracy. The MATLAB function *fitcdiscr* was used to train the LDA model. Fig. 2.15(a) shows the testing data along with LDA decision boundaries. Fig. 2.15(b) shows the confusion matrix that shows the classification results. Using the reconstructed signals from the improved method, we can classify 69.6% of the fault levels correctly. The confusion matrix also gives the information of the miss alarm and false alarm rates. The miss alarm rate for each severity level is given in the bottom row, whereas the false alarm rate is given on the very right-hand column. For instance, the miss alarm rate of severity level F3 is 43.8%. The false alarm rate of severity level F4 is 25%. Note that 100% minus the false alarm rate gives the confidence level for an assessed severity level. For instance, the confidence level of the severity level F3 is 75%.

For comparison, Fig. 2.14(b) shows the boxplot of PMI values of signals reconstructed via RSVD. Due to the reweighted summation, the PMI of the extracted signal under the F3 condition was smaller. An LDA model was trained to numerically show the separability, using the same data resampling and partition methods. Fig. 2.15(c) shows the testing data and LDA decision boundaries. Fig. 2.15 (d) shows the confusion matrix that contains the classification results. The reconstructed signal classifies 51.8% of the fault levels correctly, which is 17% worse than the improved method.

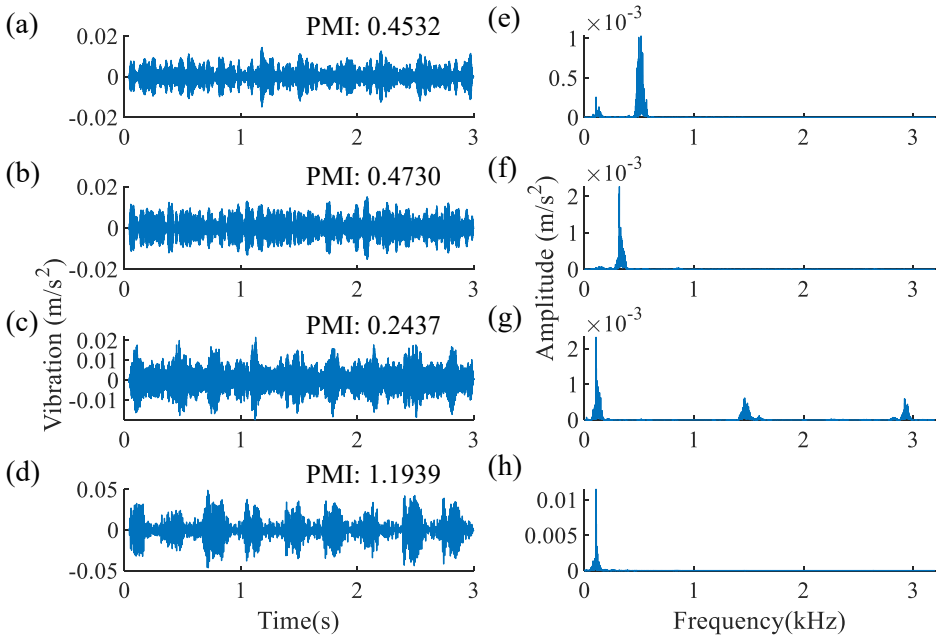

Fig. 2.13: Reconstructed signals via reweighted summation: (a)-(d) time waveforms under the F2, F3, F4, and F5 conditions, respectively; (e)-(h) frequency spectrums under the F2, F3, F4, and F5 conditions, respectively.

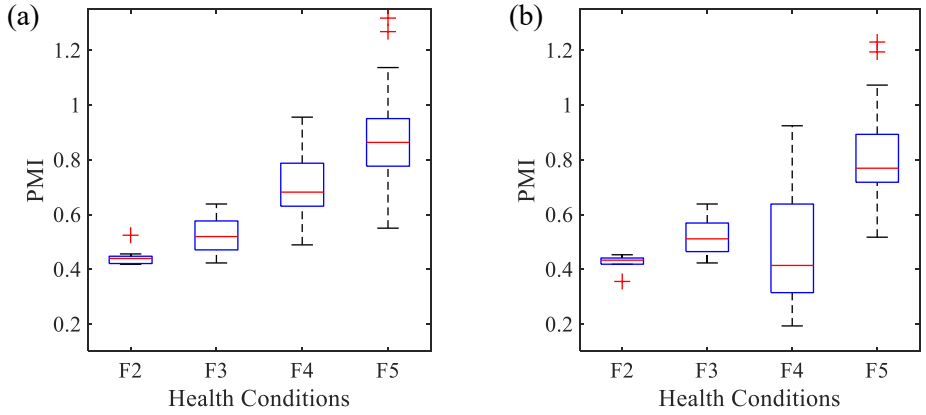

Fig. 2.14: PMI value of the signals reconstructed via (a) the improved method; (b) the RSVD method. The central mark on the box denotes the median value of the repeated measurement, and the edges of the box represent the 25th and 75th percentiles. Whiskers extend to the most extreme data points, and outliers are plotted as red crosses.

It is interesting to note that the diagonal numbers are the same in the confusion matrices obtained from the two SVD-based methods, except for F4. This is because the reweighted summation undermines the PMI of F4 the most, which was shown in Fig 2.13. In the future

beyond this thesis, further investigations are to be conducted to reveal the underlying reason why the reweighted summation undermines the PMI of F4 the most.

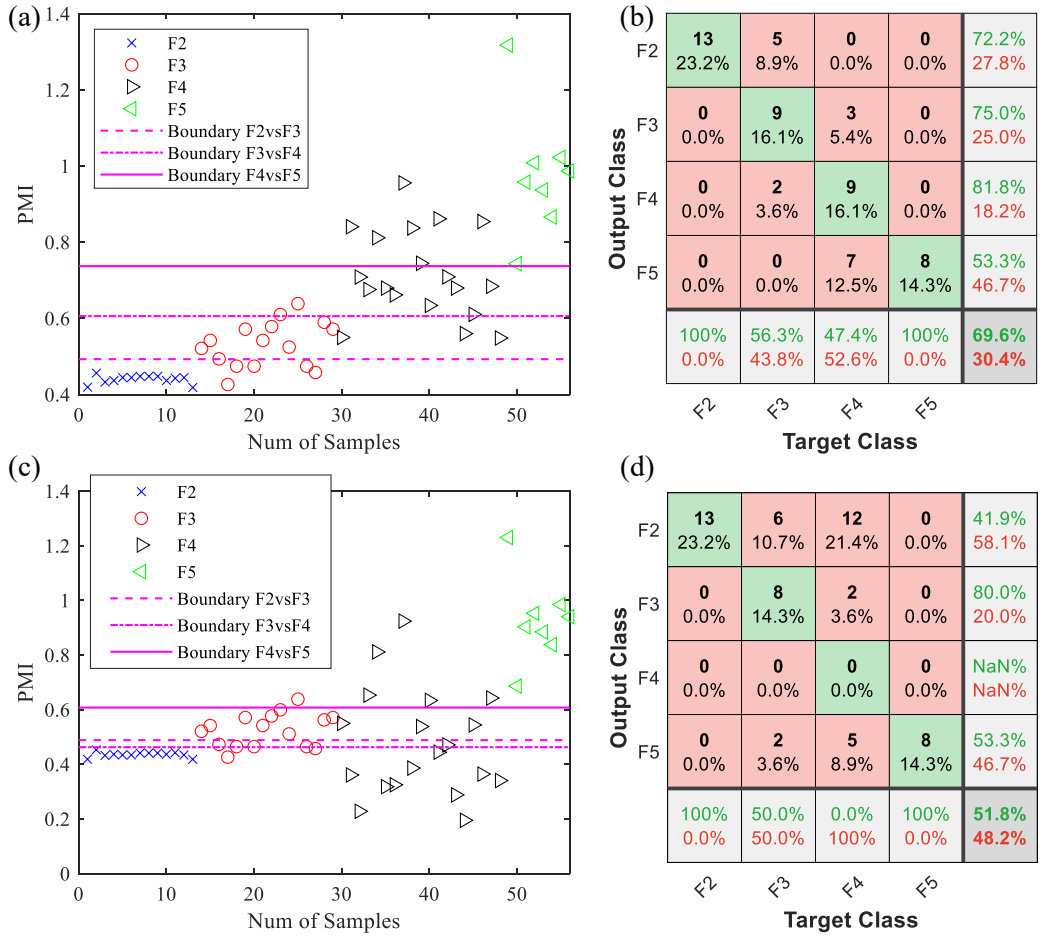

Fig. 2.15: Fault severity classification results: (a) testing data from the improved method with LDA decision boundaries; (b) confusion matrix for the testing data from the improved method; (c) testing data from the RSVD method with LDA decision boundaries; (d) confusion matrix for the testing data from RSVD.

The improved method classifies 4 fault severity levels with still a relatively low accuracy (i.e., 69.6% for experimental signals, as shown in Fig. 2.15(b)) which is not good enough for industrial applications. A reason for this outcome is that this chapter solely relied on the PMI criterion of the reconstructed signal for fault severity classification. To improve severity classification accuracy, one may need to extract more fault sensitive features (e.g., kurtosis and root mean squared) from the reconstructed signals, select more effective

features, and use an advanced classifier (e.g., support vector machine). Improving fault severity classification accuracy further will be left as one of the future research topics.

2.4.4 Computational cost

The computational cost was evaluated. Each vibration signal lasted 3 s and has 19,200 data points. Table 2.4 lists the central processing unit time consumed to process a vibration signal. Step #1 takes 0.06 s, whereas Step #5 takes less than 0.01 s. The difference between the improved method and RSVD involves Step #1 and Step #5. These computational costs can be ignored in light of the time ~50 s consumed by Step #2. Thus, the improved method does not create a considerable computational burden.

The improved SVD-based fault detection and severity assessment method is relatively less efficient. In many applications where incipient faults do not immediately lead to a catastrophic failure of the gearbox system, updating the fault information every minute is acceptable. The improved method was implemented in MATLAB 2018a on a desktop computer with two Intel 2.4 GHz processors and 16 GB of RAM. Computational efficiency can be further improved by implementing the algorithm on a more powerful computer.

Table 2.4: Central processing unit time. Algorithms were coded in MATLAB 2018a on a desktop computer with two Intel 2.4GHz processors and 16GB of RAM

Algorithms	**Time (s)**
Step #1: AR inverse filter	0.058
Step #2: SVD	49.968
Step #3: Evaluate PMI	0.030
Step #4: Fault detection	<0.001
Step #5: Reconstruct the signal	<0.001

2.5 Conclusions

This chapter proposed an improved SVD-based method for tooth crack fault detection and severity assessment. First, the improved method adopted an AR model-based baseline removal approach to generate residual signals. Then, SVD was applied to decompose the residual signals rather than raw signals. Interference from non-fault related vibration components on the PMI is therefore eliminated. Second, the improved method selected SCs such that the PMI of the reconstructed signal was maximized, rather than directly conducting a reweighted summation on all SCs with PMI values exceed the threshold. An experimental study has been conducted. Repeatability of the improved method has been studied. The results have shown that the improved method outperforms RSVD on fault detection and severity assessment and, meanwhile, does not create a considerable computational burden.

Note that the proposed method is inapplicable for some applications where baseline vibration signals are not available. But, in many applications, it is not hard to get baseline vibration signals. For example, when a new gearbox is installed, vibration signals can be collected and saved as a baseline. The improved SVD-based method is useful for these applications with available baseline vibration signals.

This chapter has addressed the problem of vibration-based fault detection and severity assessment under constant speed conditions. The proposed method can extract the crack fault-induced periodic impulses from the cyclo-stationary vibration signals. Fault detection and severity assessment are therefore based on the calculation of feature(s) on the extracted signal and feeding them into a classifier for fault detection severity level classification.

Future studies beyond this thesis include 1) investigate the reason why the number of SCs which have higher PMIs than the threshold T_h does not monotonically increase when the fault severity increases; 2) investigate why conducting the reweighted summation undermines the PMI of F3 the most; 3) improve the fault severity classification accuracy by extracting more fault sensitive features (e.g., kurtosis and root mean squared) from the reconstructed signals, selecting more effective features, and using an advanced classifier (e.g., support vector machine).

As introduced in Section 1.1.2, the rotating speed of gearboxes can be variable, such as those gearboxes used in railway traction and wind turbines. In Chapters 3 and 4, the problem of vibration-based fault detection and severity assessment under variable speed conditions will be addressed.

3

Sparse time series modeling of the baseline vibration from a gearbox under variable speed condition

The identification of a baseline model to represent the vibration from a healthy state is a must for the time series model-based methods (TSMBMs). The accuracy of baseline modeling directly determines the success of TSMBMs. In this chapter, we aim to improve the accuracy of baseline vibration modeling under variable speed condition (the research topic #2 as introduced in Section 1.3). In chapter #4, we will develop a time series model-based method for the gear tooth crack detection and severity assessment under random speed variation. The materials of this chapter have been published partially in two conference papers [90], [107] and as a whole in a journal paper [108].

3.1 Introduction

Vibration-based condition monitoring and health assessment for gearboxes are significant in preventing catastrophic failure, enabling condition-based maintenance and thus minimizing downtime and/or maintenance costs.

Gearboxes often operate under variable speed (VS) conditions. For example, wind turbine gearboxes operate under VS due to the randomness of the wind [8]. Fans in demand ventilation systems operate under VS to reduce operating costs [129]. Under VS, the

vibration signals become nonstationary. Nonstationary signal analyzing and processing tools are thus in demand for condition monitoring and health assessment.

Typical nonstationary signal analyzing and processing methods include order tracking [130], adaptive filtering [131], generalized synchronous average (GSA) method [27], cyclo-nonstationary tools [132], wavelet analysis [80], [81], empirical mode decomposition [112], and time-series model-based method (TSMBM) [85]–[87], [89]. Among these methods, TSMBM is promising in processing vibration signals for assessing the health condition of a gearbox [12], [34], [110]. Advantages of TSMBM include being free from physics or analytical information, and inherent accounting for uncertainties from measurement and operational [44]. When applied to gearboxes health assessment, *model residual-based method* [44], [85]–[87], [89] has the following steps:

a) Identify a time series model to accurately represent the baseline vibration from a gearbox (i.e., vibration generated when the gearbox is healthy);
b) Construct an inverse filter based on the baseline model and use the inverse filter to remove the baseline vibration from the signals that are collected when the health state of the gearbox is to be determined. The so-called residual signal can be obtained, which has a higher signal to noise ratio than the raw signal;
c) Assess the health condition by analyzing the residual signal, such as using statistic indicators [133].

Among these steps, step (a) is of vital importance. Inaccurate modeling of the baseline vibration will cause insufficient or excessive removal of the baseline vibration. Consequently, the condition monitoring system with TSMBM embedded may have high miss-alarm or false-alarm rate. Therefore, the maintenance cost cannot be reduced,

catastrophic failures cannot be prevented, and downtime cannot be minimized. This chapter focuses on accurate time series modeling of the baseline vibration.

As reviewed in Section 1.2.2.4, time-variant time series models like have been reported to represent the nonstationary baseline vibration from rotating machinery operating under VS conditions, including the periodic AR [84], adaptive AR model (AAR) [85], [86], weighted ensemble of AR models [87], Schur adaptive filter [88], AR model with exogenous excitation [89], and functional pooled AR model [45], [91], [92]. Unfortunately, each of these models has its drawbacks.

In this chapter, we investigate the FP-AR model with its AR parameters dependent on VS. After configuring the AR parameters as functions of VS, the FP models will have unique spectral content at each speed level. Therefore, the FP models are capable of modeling the complex speed-dependent spectrum contents of the nonstationary baseline vibration. Since the rotating speed of gearboxes is measurable via a tachometer or estimation method [134], stochastic functional models for non-measurable operating conditions [135] are beyond our scope. The functional series time-dependent AR model (FS-TAR) [56] is another widely employed approach for nonstationary signal modeling. The FS-TAR model has the same model structure with the FP-AR model, but with its AR coefficients dependent on time. The FS-TAR is beyond our scope because a TSMBM based on FS-TAR will require the speed profile of the testing signal being exactly the same as the speed profile during model training, which is hard to fulfill.

Identifying an FP-AR model needs parameter estimation and model structure selection methods. The least squared and maximum likelihood estimators are available for estimating model parameters [136]. The least squared estimator can also be computed recursively, and thus termed recursive least squared estimator. Both least squared and

maximum likelihood estimators are asymptotically Gaussian distributed with mean coinciding to the true value. The maximum likelihood estimator achieves lower model parameter estimation variance than the least squared estimator. On the other hand, the maximum likelihood maximum likelihood estimator has a higher computational cost than the least squared method. The least squared estimator is often selected in large data size case considering the computational [94]. Model structure refers to AR terms and functional spaces that describe the dependency between AR parameters and VS. Conventionally, the model structure is selected via the two-step Bayesian information criteria (BIC) based method [45], [56], [91]. However, the two-step BIC-based method limited the modeling accuracy of FP-AR. Firstly, much real-world time series, such as seismology waves, radar signals, and speech signals, are generated from a sparse AR process [101]. A sparse AR model uses scattered AR terms as predictors, rather than consecutive AR terms. The baseline vibration from a gearbox may also be generated from a sparse AR process. Thus, an FP-AR with a sparse set S_{AR} may be more accurate to model the gearbox vibrations than assuming consecutive set S_{AR}. Secondly, the dependency between different AR parameters and the VS for $i \in S_{AR}$ may be very different. They may require non-identical sets of functional spaces $S_{F,i}$ for $i \in S_{AR}$ to achieve accurate regressions. Assuming an identical set of functional spaces $S_{F,i}$ for $i \in S_{AR}$ may be too coarse.

To overcome the limitations of the conventional FP-AR model (i.e., the FP-AR model with its structure selected by the two-step BIC-based method), a sparse FP-AR model is proposed that uses a sparse AR set S_{AR} and non-identical functional spaces $S_{F,i}$. These two assumptions made in the two-step BIC based method are relaxed. To obtain such a sparse

FP-AR model, we develop a new model selection procedure by adopting the least absolute shrinkage and selection operator (LASSO) [137].

The main contributions of this study include:

1) Proposed a sparse FP-AR model for the representation of baseline vibration from gearboxes under VS condition.
2) Validations of the sparse FP-AR model-based fault detection and severity assessment for gearboxes under deterministic VS. Simulation signal and laboratory signals from two independent test-rigs are used to validate the performance of the proposed method.

The rest of this chapter is organized as follows: Section 3.2 proposes the sparse FP-AR model. A new model selection procedure is detailed. Section 3.3 validates the proposed sparse FP-AR model using simulation signals and two independent experimental signals. Conclusion remarks are given in Section 3.4.

3.2 Sparse FP-AR model

This section begins with the basics of the sparse FP-AR model, introduces LASSO, and develops a new model selection procedure, and finally lists the model validation procedure.

3.2.1 Model basics

The time-invariant AR model has the following difference equation according to [138]

$$y_t = \sum_{i \in S_{AR}} a_i y_{t-i} + \varepsilon_t \tag{3.1}$$

where y_t and y_{t-i} denote the vibration at time t and t-i, respectively; S_{AR} is a set that specifies the AR terms; a_i stands for the AR parameters which are constant; ε_t is a zero-mean white noise at time t.

The FP-AR model extends the time-invariant AR model to time-variant (i.e., operating condition-variant) as follows [94]

$$y_t = \sum_{i \in S_{AR}} a_i(\omega_t) y_{t-i} + \varepsilon_t \tag{3.2}$$

where ω_t denotes the rotational speed at time t; a_i is a function of ω_t.

Applying z-transform [121] to Eq. (3.2), we obtain

$$Y(z) = \frac{1}{1 - \sum_{i \in S_{AR}} a_i(\omega_t) z^{-i}} e(z) = H(z) e(z) \tag{3.3}$$

where $Y(z)$ is the nonstationary gear meshing vibration signal expressed in z-domain; $e(z)$ is the zero-mean white noise in z-domain. In the viewpoint of "linear time-variant (LTV) system" theory, the $Y(z)$ is modeled as the output of an LTV system $H(z)$ excited by a zero-mean white noise $e(z)$.

Based on Eq.(3.3), the transfer function of the inverse filter is [121]:

$$H^{-1}(z) = 1 - \sum_{i \in S_{AR}} a_i(\omega_t) z^{-i} \tag{3.4}$$

This inverse filter is rotating speed dependent and with the finite impulse response.

The dependency $a_i(\omega_t)$ is modeled using the basis function expansions [56]. Reported basis functions include discrete cosine transform functions, Legendre polynomials, and harr functions, normalized B-spines [56]. Given that the modeling accuracy of the functional series expansion method depends on the choice of functional spaces rather than the type of basis function [56], we use the Legendre polynomials in our method. Therefore, the dependency $a_i(\omega_t)$ is of the form:

$$a_i(\omega_t) = \sum_{j \in S_{F,i}} a_{i,j} \omega_t^{\,j} \tag{3.5}$$

where $a_{i,j}$ stands for the AR parameters of projection; $S_{F,i}$ is a set that specifies the functional spaces.

Remarks: i) When other operating conditions vary independently, FP-AR model can be extended to the VFP-AR model in which the AR parameters are functions of vectors. Take the load l_t as an example, VFP-AR is of the form

$$y_t = \sum_{i \in S_{AR}} a_i\left(\omega_t, l_t\right) y_{t-i} + \varepsilon_t \tag{3.6}$$

where the dependency $a_i(\omega_t, l_t)$ is projected into bivariate basis functions of ω_t, l_t; $\mathbf{G}(\omega_t, l_t)$ which is resulted from a Kronecker product $\otimes$ of univariate basis functions [45] $\mathbf{G}(\omega_t)$ and $\mathbf{G}(l_t)$ as follows

$$a_i\left(\omega_t, l_t\right) = \sum_{j \in S_{F,j}} a_{i,j} G_j\left(\omega_t, l_t\right) \tag{3.7a}$$

$$\mathbf{G}\left(\omega_t, l_t\right) = \mathbf{G}\left(\omega_t\right) \otimes \mathbf{G}\left(l_t\right) \tag{3.7b}$$

ii) When the vibration signal y and the VS ω_t are sampled in the angular θ domain [89], the model can be easily adopted by changing the t into θ.

The conventional FP-AR model assumed 1) consecutive AR terms, e.g., $S_{\mathrm{AR}}=\{1,2,\ldots, n_a\}$, and 2) identical sets of functional spaces for $i \in S_{AR}$, e.g., $S_{F,i} = S_F$. Although these two assumptions simplify the model selection and enable the two-step BIC-based method [45], [56], [91], they however limited the modeling accuracy of conventional FP-AR.

To improve the conventional FP-AR model, we propose a sparse FP-AR model that uses a sparse set S_{AR} and non-identical sets of functional spaces $S_{F,i}$ for $i \in S_{AR}$. The sparse set S_{AR} implies that the AR terms are spread out, rather than being consecutive. The non-identical

set of functional spaces $S_{F,i}$ for $i \in S_{AR}$ implies that the dependency $a_i(\omega_t)$ is modeled using different basis functions for different $i \in S_{AR}$, rather than using the same basis functions.

3.2.2 LASSO

The idea behind the LASSO algorithm is to minimize the residual sum of squares while penalizing the sum of the absolute value of the coefficients [137]. The LASSO algorithm forces insignificant coefficients to be exactly zero and hence performs model selection.

Suppose that we sampled baseline vibration y_t and rotating speed ω_t for t=1, 2, …, N from a gearbox. Given an initial consecutive AR set S_{AR}^{ini} with a maximum lag order of $\tilde{n}_a$ and an initial identical set S_F^{ini} that specifies the functional spaces, we can construct the following matrix M that is used as input for identifying an FP-AR model

$$\mathbf{M} = \begin{bmatrix} y_1 & y_2 & \cdots & y_{\tilde{n}_a} & \omega_{\tilde{n}_a+1} & y_{\tilde{n}_a+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{t-\tilde{n}_a} & y_{t-\tilde{n}_a+1} & \cdots & y_{t-1} & \omega_t & y_t \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{N-\tilde{n}_a} & y_{N-\tilde{n}_a+1} & \cdots & y_{N-1} & \omega_N & y_N \end{bmatrix} \tag{3.8}$$

Each row is of the form $\begin{bmatrix} y_{t-\tilde{n}_a} & y_{t-\tilde{n}_a+1} & \cdots & y_{t-1} & \omega_t & y_t \end{bmatrix}$. The first $\tilde{n}_a$ +1 columns of the matrix contain predictors and the last column responses.

Remark: Eq. (3.8) specifies a matrix constructed from one segment of continuously sampled data. For the case of using s number of segments for model identification, s number of matrixes can be constructed from each segment and then augmented to construct a matrix $\mathbf{M}_e$, that is

$$\mathbf{M}_e = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_s \end{bmatrix} \tag{3.9}$$

Based on the predictors and responses specified in Eq. (3.8), the LASSO estimates $\hat{\mathbf{a}} = \left[\hat{a}_{i,j} \right]$ for $i \in S_{AR}^{ini}$ and $j \in S_F^{ini}$ as follows [137]:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \left\{ \sum_{t=\tilde{n}_a+1}^{N} \left(y_t - \sum_{i \in S_{AR}^{ini}} \sum_{j \in S_F^{ini}} a_{i,j} \omega_t^{\,j} y_{t-i} \right)^2 + \lambda \sum_{i \in S_{AR}^{ini}} \sum_{j \in S_F^{ini}} \left| a_{i,j} \right| \right\} \tag{3.10}$$

where $\lambda \geq 0$ is a tuning parameter. The squared summation part is the l_2 norm of the one-step-ahead prediction error. The sum of the absolute value of the coefficients is the penalized term. Such penalized optimization problem can be solved via the alternating direction method of multipliers algorithm [139]. The LASSO returns a sparse FP-AR model that uses a sparse AR set and non-identical sets of functional spaces.

The selection of tuning parameter λ is critical. The λ controls the amount of shrinkage that is applied to the estimates. When $\lambda = 0$, the LASSO will reduce to the OLS estimator. Too large λ will force too many coefficients in $[a_{i,j}]$ for $i \in S_{AR}^{ini}$ and $j \in S_F^{ini}$ to zero, whereas too small λ will force a limited number of coefficients to zero. The λ can be selected by either the K-fold cross-validation or validation set approach [140]. The K-fold cross-validation can "shake" the training data effectively and thus select a good value of λ but needs relatively higher computational cost. The validation set approach consumes less computational cost than the K-fold cross-validation approach but needs more data for validation and easy to be overfitted to the chosen validation data. We use the 10-fold cross-validation (CV) [140] to select a good value of λ given that the computational cost during offline model training is not critical.

Applying 10-fold CV into time series is argued inappropriate because 1) we should not use future data to predict the past and 2) serial correlation in the data [140]. However, a recent study [140] has justified the use of standard 10-fold CV without modification when

autoregressive models are applied. The two reasons are proved to be largely irrelevant. Therefore, we apply the 10-fold CV without modification. In total, 4 steps are followed:

1) The matrix **M** is randomly partitioned into 10 equal size subsamples.
2) Among the 10 subsamples, a single subsample is retained as the validation data, and the remaining 9 subsamples are used as training data.
3) Step 2 is repeated 10 times, with each of the 10 subsamples used exactly once as the validation data.
4) The 10 results from the folds can not only be averaged to produce an estimation of mean CV mean squared error (CVMSE) but also be combined to produce a standard deviation for CVMSE.

3.2.3 A new model selection procedure

To obtain the proposed sparse FP-AR model, we develop a new model selection procedure by adopting LASSO. The new model selection procedure has 3 steps, as shown in Fig. 3.1. Each step is detailed as follows:

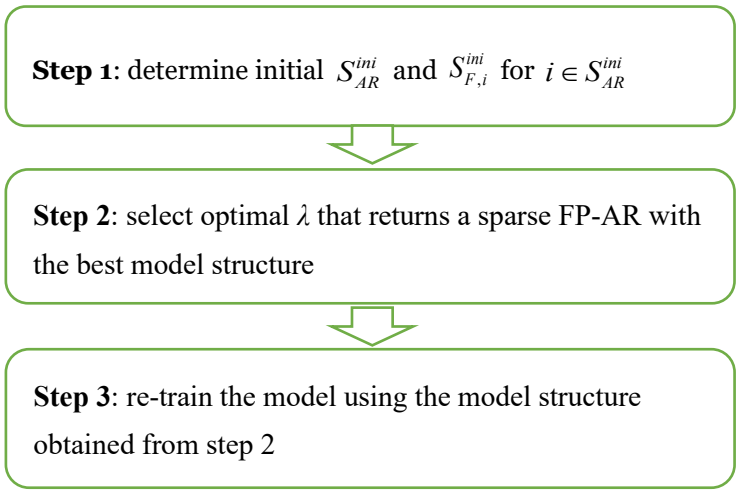

Fig. 3.1: Flowchart of the new model selection procedure

Step 1: determine initial S_{AR}^{ini} and $S_{F,i}^{ini}$. We consider conservative initial sets so that the model terms of the optimal sparse FP-AR model can be selected. In this step, the initial S_{AR}^{ini} is assumed as a consecutive set, e.g., S_{AR}^{ini} ={1,2,..., $\tilde{n}_a$ }, and the sets of functional spaces $S_{F,i}^{ini}$ are given as identical for $i \in S_{AR}$, e.g., $S_{F,i}^{ini} = S_F^{ini}$. These assumptions are the same as what made in the conventional FP-AR model [91]. With these assumptions, firstly, the initial S_F^{ini} is empirically determined to be high-dimensional, for example S_F^{ini} = {0, 1, 2, ..., 7} (i.e., basis functions = {1, ω_t, ω_t^2, ..., ω_t^7}). We suggest that an order of polynomial from 5th to 8th is adequate. Secondly, using S_F^{ini}, the maximum order of AR terms $\tilde{n}_a$ is selected using BIC:

$$\mathrm{BIC} = N \ln\left(\frac{\mathbf{e}^T\mathbf{e}}{N}\right) + \left(\ln N\right) n_a \tag{3.11}$$

where $\mathbf{e} = \left\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N\right\}^T$ is the residual vector, and T denotes the transpose operation. This BIC-based method evaluates n_a in enough large region, for example, n_a =1, 2, ..., 1000, calculates the corresponding BIC values and finds the $\hat{n}_a$ which minimizes BIC. Then $\tilde{n}_a$ is determined as 1.5 times of $\hat{n}_a$ so that the S_{AR}^{ini} can be conservative.

Step 2: select the tuning parameter λ that returns an optimal sparse FP-AR with the best sparse model structure. In this step, we empirically configure the candidates of tuning parameter λ as [0, 1×10^{-8}, 1×10^{-7}, ..., 1×10^{-1}]. For each candidate of tuning parameter λ, a sparse FP-AR model is obtained, along with its mean and standard deviation of CVMSE. The tuning parameter λ is selected as the largest one which has a mean CVMSE within one standard deviation of the minimum mean CVMSE. For instance, as illustrated in Fig. 3.2, the λ_3 gives the minimum mean CVMSE. The mean CVMSE of λ_2 is within one standard error of the minimum CVMSE, whereas the mean CVMSE of λ_1 does not. Hence, λ_2 is selected as the final value which gives the optimal sparse FP-AR model. The rationale

behind is that the larger the tuning parameter λ is, the more coefficients are forced to 0 and hence the sparse FP-AR model has a smaller number of non-zero coefficients. By selecting the largest λ that has not undermined the CVMSE too much, we can control the number of non-zero coefficients and hence avoid overfitting.

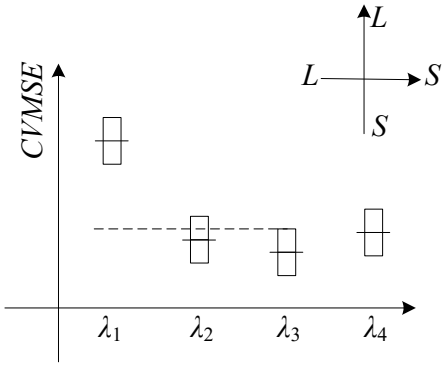

Fig. 3.2: Illustration of selecting the tuning parameter λ. In this Fig, the central line of the box is the mean CVMSE; the edges of the box are the mean CVMSE plus/minus one standard deviation; *L* denotes large; and *S* denotes small.

Step 3: re-train the model using the structure obtained from step 2. This step updates the model parameters using the OLS estimator. Although LASSO can estimate the model parameters and perform model selection simultaneously [137], we find incorporating this step can further improve the modeling accuracy.

3.2.4 Validation of modeling accuracy

The modeling accuracy is validated using a testing signal. Upon the identification of a sparse FP-AR model for representing the baseline vibration of a gearbox, the inverse filter is constructed based on Eq. (3.4) and then applied to process the testing signal. We call the residual obtained from the testing signal as "residual-of-testing".

The modeling accuracy of sparse FP-AR model is firstly measured by examining the mean squared error (MSE) of the residual-of-testing which is calculated as

$$\mathrm{MSE} = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_t^{\,2} \tag{3.12}$$

where ε_t is the residual at the t^{th} point. A model with lower MSE is more accurate in modeling the baseline vibration than those with higher MSE [36, Section 7].

Secondly, the randomness of the residual-of-testing is examined. The more random the residual-of-testing is, the more accurate the model. Ljung–Box test [141] is conducted to quantify the randomness. Autocorrelations of a random time series equal to zero. The Ljung-Box test statistically tests whether any of a group of autocorrelations of a time series is different from zero. The *p*-value is reported as the probability of being random.

Thirdly, modeling accuracy is judged by one-step-ahead predictions (OSAPs) and its goodness-of-fit with the real gear vibration signal (i.e., simulation signal minus noise). The higher the fit is, the more accurate a model predicts the real vibration. The fit value is a normalized mean square error as:

$$\mathrm{fit} = 1 - \sum_{t=1}^{N}\left(y_{r,t} - y_{m,t}\right)^2 \Big/ \sum_{t=1}^{N}\left(y_{r,t} - \bar{y}_r\right)^2 \tag{3.13}$$

where y_r is the real vibration response; $\bar{y}_r$ is the mean value of the real vibration response y_r; y_m is the OSAP. Note that for experimental signals, the real gearbox vibration is unknown, and the goodness-of-fit measure is inapplicable.

Lastly, we present the parametric spectrum $S(f, t)$ of FP-AR models by frozen-time formula [56]

$$S\left(f,t\right) = \left|1 + \sum_{i \in S_{AR}} a_i\left(\omega_t\right) e^{-j2\pi \frac{f}{f_s} i}\right|^{-2} \tag{3.14}$$

where j is the imaginary part; f_s is the sampling frequency; f designating the frequency response from 0Hz to $f_s/2$. The non-parametric spectrum of the testing signal is also

obtained using MATLAB *spectrogram* function. We can judge the modeling accuracy by comparing the agreement between the parametric spectrum and non-parametric one. If the parametric spectrum deviates from the non-parametric spectrum, then the modeling accuracy is low. Otherwise, the modeling accuracy is high.

3.3 Performance evaluation

This section evaluates the proposed sparse FP-AR model using a simulation signal and two independent experimental datasets. All algorithms are coded in MATLAB 2018a and implemented on a desktop with two Intel 2.4 GHz processors and 16 GB of RAM.

3.3.1 Simulation signal

Based on a simulation model in Ref. [26] for generating vibration signal from a fixed-axis gearbox under constant speed condition, we develop a new simulation model that considers the amplitude modulation (AM) and frequency modulation (FM) effects induced from VS, and incorporates the vibration components from structures. For a one-stage fixed-axis gearbox in which the input gear has T number of teeth, the new simulation model describes the nonstationary vibration signal y_t under VS as follows:

$$y_t = \underbrace{\sum_{i=1}^{I} G_i\left(\omega_t\right)\cos\left(i2\pi T_e \int_0^t \omega_t dt + \varphi_i\right)}_{gear\ meshing\ vibration} + \underbrace{\sum_{j=1}^{J} S_j\left(\omega_t\right)\cos\left(2\pi f_j t + \psi_j\right)}_{structure\ vibration} + \underbrace{\sum_{k=1}^{K} D_k\left(\omega_t\right) S_d\left(t - \Omega^{-1}\left(k\right)\right)}_{defect\ impulses} + \underbrace{\varepsilon_t}_{noise} \tag{3.15}$$

The first term represents gear meshing component and its harmonics, where I is the number of gear meshing harmonics; t is time; T_e is the number of teeth; ϕ_i is an initial phase for i^{th} gear meshing harmonic; ω_t is the rotating speed of input shaft at time t. From [142], gear meshing harmonics less than 10 times of meshing frequency are significant.

Thus, the value of I is recommended to be $\left\{ I \in \mathrm{Z}^{+} \,|\, 1 \le I \le 10 \right\}$. Under VS, additional AM $G_i\left(\omega_t\right)$ acts on the gear meshing component and its harmonics. Many references [143]–[145] find that the frequencies of some vibration components rarely affected by the variation of rotation speed can be observed in baseline vibration. The amplitude of these vibration components is still modulated by the speed. Ref. [144] explained the structures, e.g., casing and/or bearing cap, as the source of these vibration components. Hence, the second term is added to account for the vibration components from structures, where f_j is the frequency of j^{th} structure vibration component; ψ_j is the initial phase for j^{th} structure vibration component; and J is the number of structure vibration components. Following by the observations in Ref. [57], the value of J is recommended to be $\left\{ J \in \mathrm{Z}^{+} \,|\, 1 \le J \le 5 \right\}$. The third term represents the periodic impulses induced by localized defects (e.g., tooth crack), where $D_i(\omega_t)$ is the amplitude of the i^{th} impulse; K is an integer which denotes the number of defect impulses also the number of rotating cycles; $\Omega^{-1}(k)$ is the inverse function of $\Omega(t) = \int_0^t \omega_t dt = k$ which determines the time of k^{th} defect impulses occurs. The defect impulses are described by an exponentially decaying sinusoid of the following form:

$$S_d\left(t\right) = e^{-\alpha t} \sin\left(2\pi f_r t\right) \tag{3.16}$$

where f_r specifies the resonance frequency excited by the impact; α is the decay rate of the impulses. The last term represents measurement noise, where ε_t is the zero-mean Gaussian random environmental noise. The parameters are configured as listed in Table 3.1.

We can obtain the angular domain signal either using hardware order tracking or computed order tracking. But we choose to analyze time domain signals directly. This is because conducting order tracking is unable to completely demodulate the frequency modulation of the nonstationary signals. For vibration components of non-rotating parts, additional frequency modulation will be induced by order tracking [27].

Table 3.1: Parameters configuration for the simulation model

Parameter	Value(s)	Parameter	Value(s)
T_e	37	ψ_1	$\pi/2$
t	0~7.5 s	ψ_2	$\pi/2$
Δt	1ms	$S_1(\omega_t)$	$0.12\omega_t^2+0.5\omega_t+0.3$
I	4	$S_2(\omega_t)$	$0.2\omega_t^2+0.4\omega_t+0.2$
ϕ_1	$\pi/2$	f_1	140 Hz
ϕ_2	$\pi/3$	f_2	180 Hz
ϕ_3	$\pi2/3$	f_r	140Hz
ϕ_4	$\pi2/3$	α	100
$G_1(\omega_t)$	$0.1\omega_t^2+0.3\omega_t+0.2$	$D_i(\omega_t)$	Healthy: 0
$G_2(\omega_t)$	$0.1\omega_t^2+0.3\omega_t+0.4$		Faulty #1 (F1): $2+\omega_t$
$G_3(\omega_t)$	$0.2\omega_t^2+0.6\omega_t+0.4$		Faulty #2(F2): $3+1.5\omega_t$
$G_4(\omega_t)$	$0.15\omega_t^2+0.5\omega_t+0.5$		Faulty #3(F3): $4+2\omega_t$
ε_t	$\sim\square(0, 1\ (m/s^2)^2)$		Faulty #4(F4): $5+2.5\omega_t$
J	2		Faulty #5(F5): $6+3\omega_t$

We vary the speed profiles continuously with a large difference between maximum and minimum values which can reflect a machine operating between various speed conditions. Fig. 3.3 shows two segments of baseline vibration, as well as their corresponding speed profiles. A gearbox may experience any kind of speed profiles and it is impractical to consider all kinds of profiles. We selected a typical testing profile widely used in [76], [146]. The selected testing profile is sinusoidal-like and covers a wide range of speed variations. In general, the amplitude of the baseline vibration has a positive correlation with the rotating speed profile. The left-side baseline vibration and its corresponding speed profile are used for training, whereas the right-side for testing. Moreover, the left side of Fig. 3.4 shows five levels of tooth crack induced impulses with an ascending amplitude $D_i(\omega_t)$ as specified in Table 3.1, whereas the right side of Fig. 3.4 shows the corresponding faulty vibrations which are a summation of baseline vibration and tooth crack induced impulses. The rotating speed of these faulty vibrations follows the profile in Fig. 3.3 (c). We cannot

detect the faults by visually observing the time waveform of faulty vibrations. Tooth crack induced impulses are submerged into the vibrations. We will apply the TSMBM based on a sparse FP-AR model to detect the tooth crack faults.

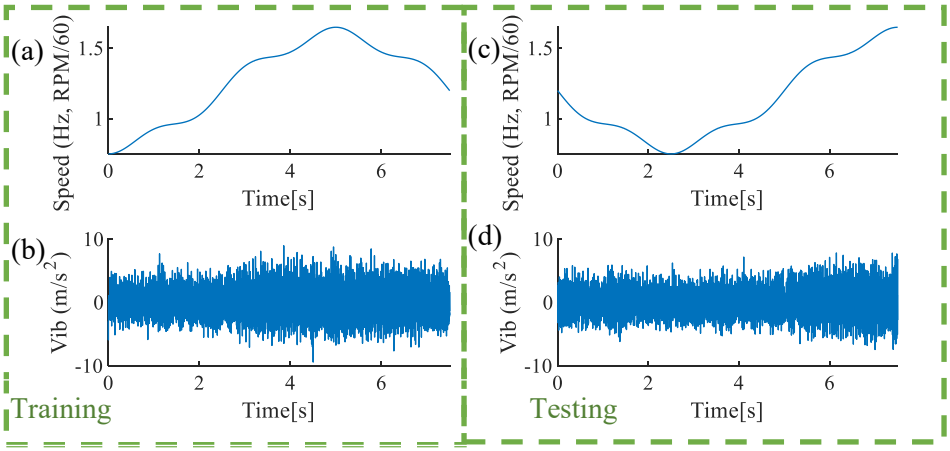

Fig. 3.3: Baseline vibration (simulation): (a) speed profile of training signal, (b) time waveform of training vibration, (c) speed profile of testing signal, (d) time waveform of testing vibration.

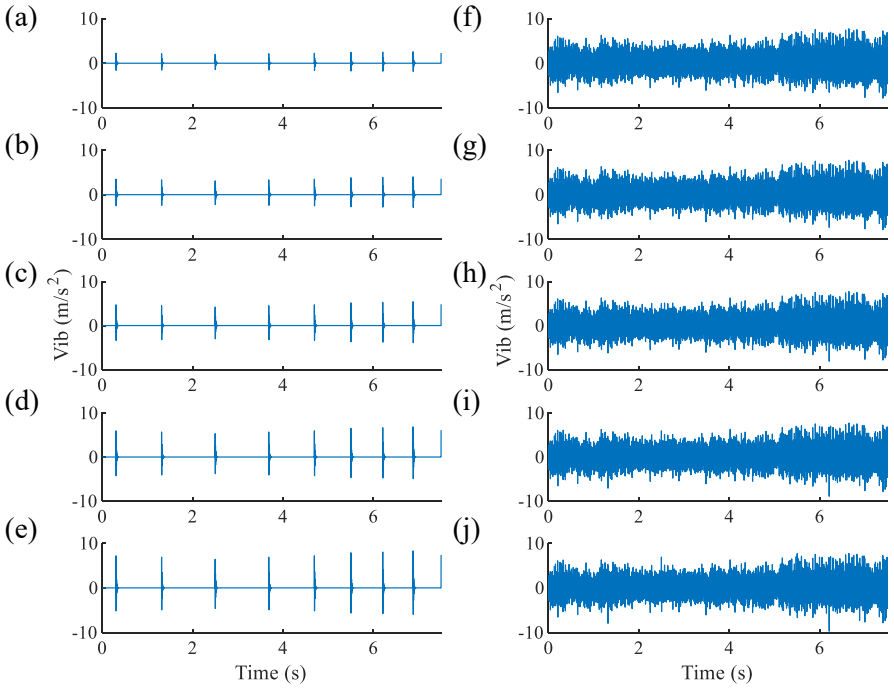

Fig. 3.4: Faulty vibrations (simulation): (a) ~ (e) time waveform of tooth crack induced impulses, with ascending amplitudes of $D_i(\omega_t)$ as specified in Table 3.1; (f) ~ (j) time waveform of faulty vibration signal, corresponding to the fault level F1, F2, ..., F5, respectively.

Before the modeling process, normalize and standardize operations are applied to the rotating speed and vibration, respectively. The ω_t is normalized as follows

$$\omega_{nt} = \frac{\omega_t - \omega_{\min}}{\omega_{\max} - \omega_{\min}} \tag{3.17}$$

where ω_{nt} is the normalized speed; $\omega_{\max}$ is the maximal speed among the speed profile, and $\omega_{\min}$ is the minimal speed among the speed profile. Meanwhile, the vibration signal y is standardized,

$$y_S=(y-\mu_S)/\sigma_S \tag{3.18}$$

where μ_s is the sample mean of y and σ_s the sample standard deviation of y.

3.3.1.1 Sparse FP-AR modeling

During the sparse FP-AR modeling, the training and testing signals as shown in Fig. 3.3 are used. The initial set of functional spaces is configured as $S_{F,i}^{ini} = S_F^{ini} = \{0, 1, 2, ..., 5\}$ (i.e., basis functions = $\{1, \omega_t, \omega_t^2, ..., \omega_t^5\}$). The initial set of AR terms S_{AR}^{ini} is determined as {1, 2, ..., 147}. The λ is configured as $[0, 1\times10^{-8}, 1\times10^{-7}, ..., 1\times10^{-1}]$. Fig. 3.5 shows the CVMSE under each λ. The minimum CVMSE appears when $\lambda = 1\times10^{-5}$ and the largest $\lambda = 1\times10^{-4}$ which returns a model which has a CVMSE within one standard error of the minimum CVMSE. Therefore, $\lambda = 1\times10^{-4}$ is chosen. Table 3.2 lists a portion of the identified model structure and model parameters. Many parameters have been forced to 0, thus returns a spare FP-AR model. In total, the sparse FP-AR model has 539 non-zero coefficients spread out in the $\hat{\mathbf{a}} = \left[\hat{a}_{i,j} \right]$ for $i \in \{1, 2, ..., 147\}$ and $j \in \{0, 1, 2, ..., 5\}$.

3.3.1.2 Conventional FP-AR modeling

For comparison purposes, the conventional FP-AR model is identified. Model parameters are estimated using the ordinary least square algorithm, and the optimal model order and

functional spaces are selected using the two-step BIC based method [45]. During the model AR order selection, the initial set of functional spaces is configured as $S_{F,i}^{ini} = S_F^{ini}$ = {0, 1, 2, ..., 5} (i.e., basis functions = {1, ω_t, ω_t^2, ..., ω_t^5}). The BIC values for each AR order n_a= {1, 2, ..., 300} are shown in Fig. 3.6(a). The optimal model order is chosen as n_a = 98 when minimizing the BIC. Then, the optimal functional spaces is a 3-dimensional spaces {1, ω_t^4, ω_t^5}, as shown in Fig. 3.6(b). Functional basses {ω_t, ω_t^2, ω_t^3} are consecutively dropped based on the backward deletion scheme. In total, the conventional FP-AR model has 294 coefficients $\hat{\mathbf{a}} = \left[\hat{a}_{i,j} \right]$ for $i \in \{1, 2, ..., 98\}$ and $j \in \{0, 4, 5\}$.

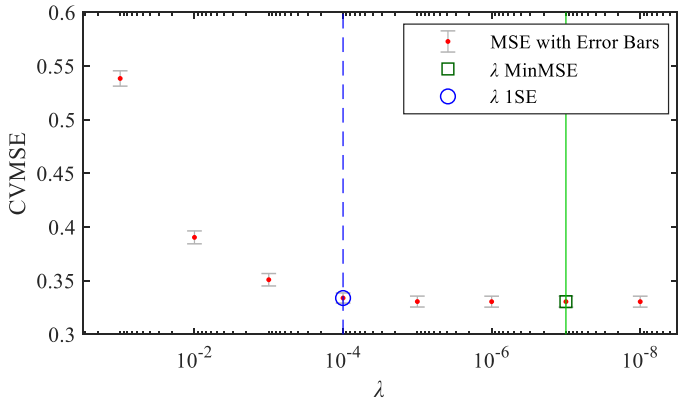

Fig. 3.5: CVMSE of lasso Fit from simulation study. The CVMSE when λ =0 was not shown due to the log x-axis was 0.3305±0.0051 which is the same as the CVMSE when $\lambda = 1\times10^{-8}$.

Table 3.2: A portion of estimated parameters a_{ij} for representing simulation signals

	t-1	t-2	t-3	t-4	t-5	t-6	t-7	...	t-147
1	-0.3286	-0.0891	0.0567	0.1477	0.0776	-0.0175	-0.0714	...	-0.0061
ω_t	-0.3621	0.3563	0.9424	0.3903	-0.0631	0	0.1905	...	0
ω_t^2	0	0	-0.6720	-0.7566	-0.2965	0.0039	0.2373	...	0
ω_t^3	0.5262	0	-0.0830	-0.0500	0	0.0559	0	...	0.1039
ω_t^4	0.0296	0	0	0	0	0	-0.0683	...	0
ω_t^5	0.3397	0	-0.1512	0.2720	0.3416	-0.0422	-0.3637	...	-0.0928

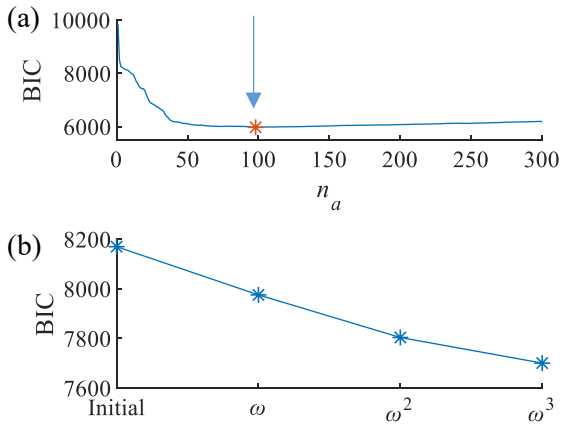

Fig. 3.6: FP-AR model structure selection for the simulation gearbox vibration signal: (a) AR order selection and (b) BIC values obtained by backward deletion scheme

3.3.1.3 Performance comparison

We evaluate the performance of sparse and conventional FP-AR models in terms of 1) modeling accuracy and 2) the benefits that improved model accuracy can bring to the fault detection and severity assessment of a gearbox under VS.

The time waveform of the residuals-of-testing from both models is shown in Fig. 3.7 (b,c). For comparison, Fig. 3.7 (a) shows the raw testing signal. We can see that the amplitude of residuals-of-testing from both models has decreased due to the removal of components in baseline vibration. Meanwhile, the mean and variance of the residuals-of-testing are likely time-invariant. In this case, the residuals-of-testing has become stationary.

Table 3.3 listed the results of sparse FP-AR in comparison with the conventional FP-AR. Firstly, the sparse FP-AR model has an MSE of 0.3075, smaller than 0.3346 that the conventional FP-AR has, and closer to the irreducible error 0.1880. Secondly, the residual-of-testing of the sparse FP-AR model has a p-value of 0.2942 from Ljung–Box test. But, the conventional FP-AR model has 0.0998. The sparse FP-AR model has a goodness-of-fit of 51.21% which is higher than 44.27% that the conventional FP-AR has. Fig. 3.8 shows

the OSAPs of both two models along with the real gear vibration signal. The OSAP from the sparse FP-AR model fits better than the conventional one. These imply that the residual-of-testing of the sparse FP-AR model is more random than its counterpart. Meanwhile, recall that sparse FP-AR has 539 non-zero coefficients and its counterpart 294. We conclude that the conventional FP-AR model tends to under-fit due to the smaller number of non-zero coefficients. The sparse FP-AR model outperforms the conventional FP-AR model in terms of modeling accuracy. Lastly, the sparse FP-AR model paid about 20 times more than the conventional one in the training stage, and slightly more in the testing stage. The higher computational cost in training originates from solving the penalized objective function and performing 10-fold CV. The higher computational cost in testing originates from the higher number of non-zero coefficients in the sparse FP-AR model (539) than its counterpart (294).

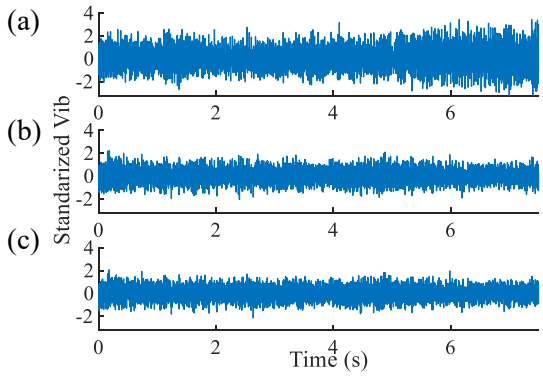

Fig. 3.7: Residuals-of-testing in comparing with the raw testing signal, simulation study: (a) raw testing signal; (b) residual-of-testing from conventional FP-AR model; (c) residual-of-testing from sparse FP-AR model

Table 3.3: Performance of sparse and conventional FP-AR models from the simulation study

Models	MSE of irreducible error	MSE of the residual-of-testing	Ljung–Box test, *p*-value	Goodness-of-fit, (%)	CPU time in training, (s)	CPU time in testing, (s)
Con FP-AR	0.1880	0.3346	0.0998	44.27	84.0	0.10
Spa FP-AR		0.3075	0.2942	51.21	1785.4	0.13

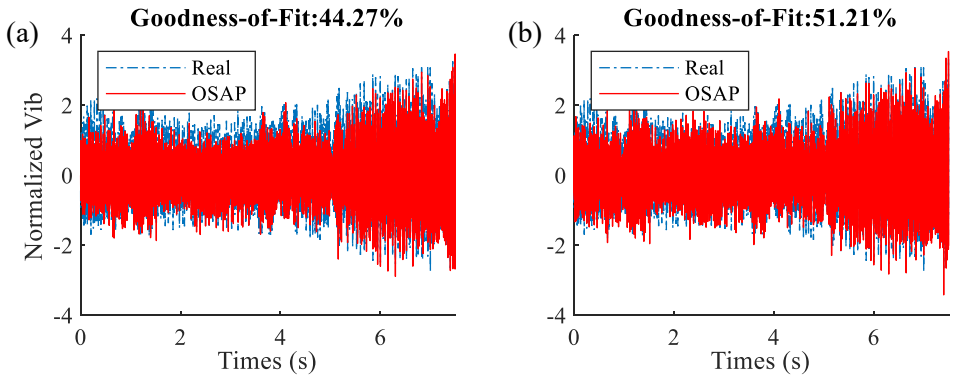

Fig. 3.8: One-step-ahead predication from (a) conventional FP-AR and (b) sparse FP-AR models. Simulation study.

Fig. 3.9 shows the frozen-time spectrums (b, c) obtained from both FP-AR models and the non-parametric spectrum (a). The frozen-time spectrum obtained from the conventional FP-AR model failed to align with the non-parametric spectrum in the region of [0 s 5 s; 0 kHz 0.13 kHz] as highlighted in Fig. 3.9 (b). On the other hand, the frozen-time spectrum obtained from the sparse FP-AR model aligns well with the non-parametric spectrum. The sparse FP-AR model has higher modeling accuracy than the conventional one.

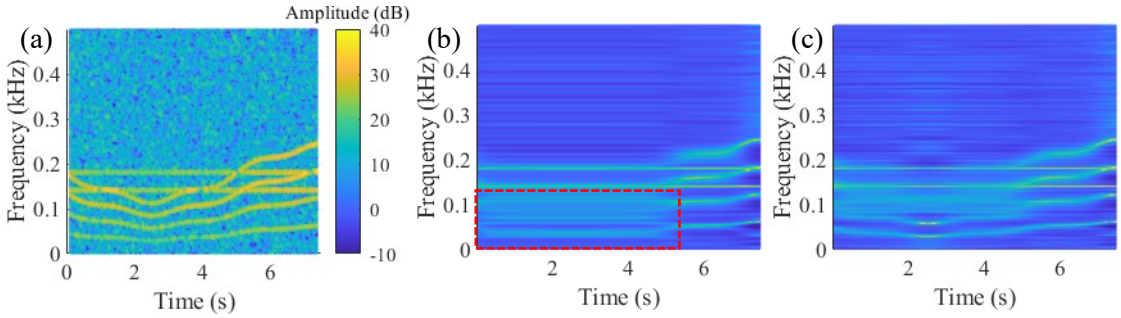

Fig. 3.9: Time-frequency spectrums of simulation signal: (a) non-parametric spectrum of testing signal obtained by MATLAB *spectrogram* function; (b) Frozen-time spectrum obtained from the conventional FP-AR model; (c) Frozen-time spectrum obtained from the sparse FP-AR model. z-axis scales are the same for three plots.

To show the benefits that improved modeling accuracy can bring to the fault detection and severity assessment of a gearbox under VS, we apply the inverse filter constructed from the sparse FP-AR model for detecting gear tooth crack faults. Fig. 3.10 shows the obtained residuals by filtering faulty vibration signals (shown in Fig. 3.4) using inverse filters

constructed from both sparse and conventional FP-AR models. We can visually observe that the residuals show an ascending "peskiness" when the fault level increases. Meanwhile, residuals from sparse FP-AR models are "peakier" than the residuals from conventional one. Next, we will calculate a periodic modulation intensity (PMI) [57] indicator to numerically measure how peaky each residual is.

The value of PMI represents the energy ratio between tooth crack induced impulses and other components [57]. The residuals shown in Fig. 3.10 are first converted to angular-domain through computed order tracking (COT) [63]. Then, PMI is calculated following Eqs. (2.7-2.8).

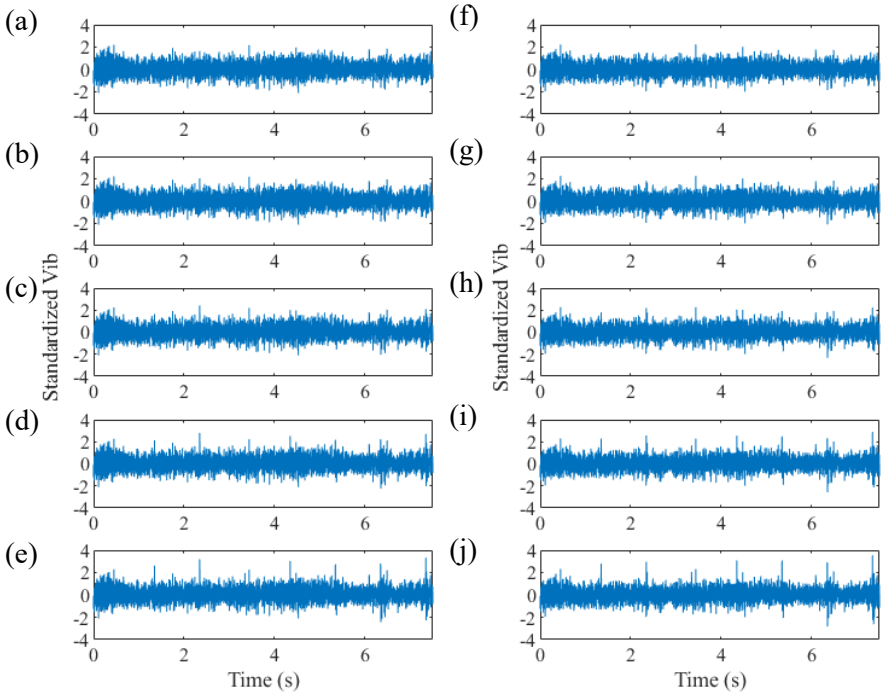

Fig. 3.10: Residuals of faulty vibrations (simulation) through inverse filtration: (a)~(e) corresponds to F1, F2, ..., F5, respectively, and are from conventional FP-AR model. (f)~(j) corresponds to F1, F2, ..., F5, respectively, and are from sparse FP-AR model.

Fig. 3.11 shows the normalized PMI (NPMI, i.e., PMI value of the residual of faulty vibrations divided by the PMI of the residual of baseline vibration) for detecting five levels of tooth crack faults. The NPMI from the raw signal failed to detect each fault levels. Thanks to the removal of baseline vibration, the NPMI calculated from res-con (i.e., the residual of conventional FP-AR model) ascends with the increasing of fault severities, especially for F3 to F5. Thanks to more accurate baseline vibration modeling and hence more accurate removal of baseline vibration, the NPMI calculated from the res-spa (i.e., residual of the sparse FP-AR model) shows a higher NPMI than the conventional one. The slope of the NPMI curve from res-spa is steeper than the conventional one, which implies that res-spa better assesses the severities of faults. Noticeably, the res-spa can detect F2, whereas the res-con failed. We conclude that an improved modeling accuracy ensures the residual to contain more fault induced features, and hence benefits the health state assessment of a gearbox under VS in early detection of faults and better assess the severities of faults.

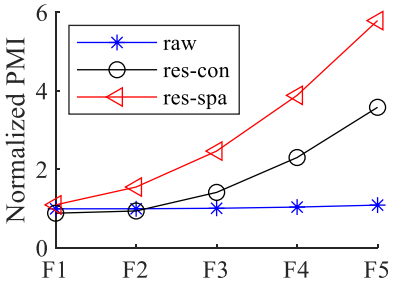

Fig. 3.11: Normalized PMI for detecting five levels of tooth crack faults, obtained from the simulation study. (raw: raw signal; res-con: the residual of conventional FP-AR model; res-spa: the residual of conventional FP-AR model.)

3.3.2 UofA dataset

The descriptions of this dataset have been given in Section 2.4.1. The experimental signals collected under the variable speed condition will be used in this chapter.

The speed profiles were configured to be similar to the profile in the simulation signal. The load motor generates a constant load level of 80Nm during the experiment. Fig. 3.12 shows two segments of baseline vibration, as well as their corresponding speed profiles acquired by an encoder. We set the sampling frequency f_s = 6.4 kHz. The left-side baseline vibration and its corresponding speed profile are used as training data, whereas the right-side as testing data. The duration of the testing signal is 15 s which covers more than 30 cycles of object gear rotation.

The left side of Fig. 3.13 shows speed profiles, whereas the right side the time waveforms of faulty vibrations collected under five levels (i.e., F1~F5) of tooth crack fault. The speed profiles are generally the same. We cannot detect the faults by visually observing the time waveform of faulty vibrations. Tooth crack induced features are submerged into the vibrations. Signal processing methods need to be employed for fault detection and severity assessment.

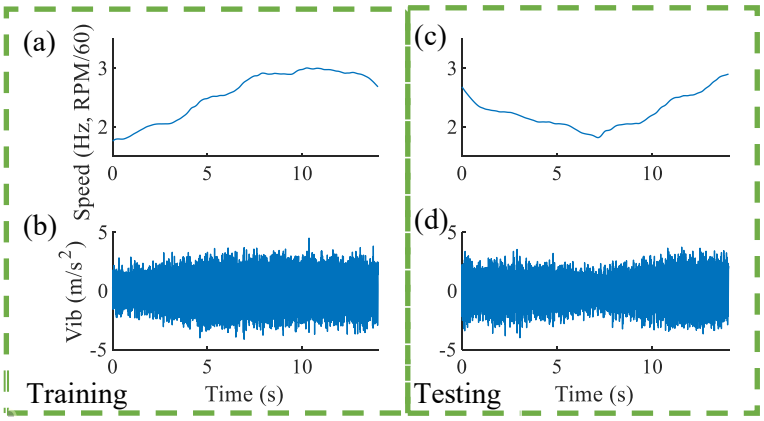

Fig. 3.12: Baseline vibration signals from UofA experiment: (a) speed profile of training signal, (b) time waveform of training vibration signal, (c) speed profile of testing signal, (d) time waveform of testing vibration signal.

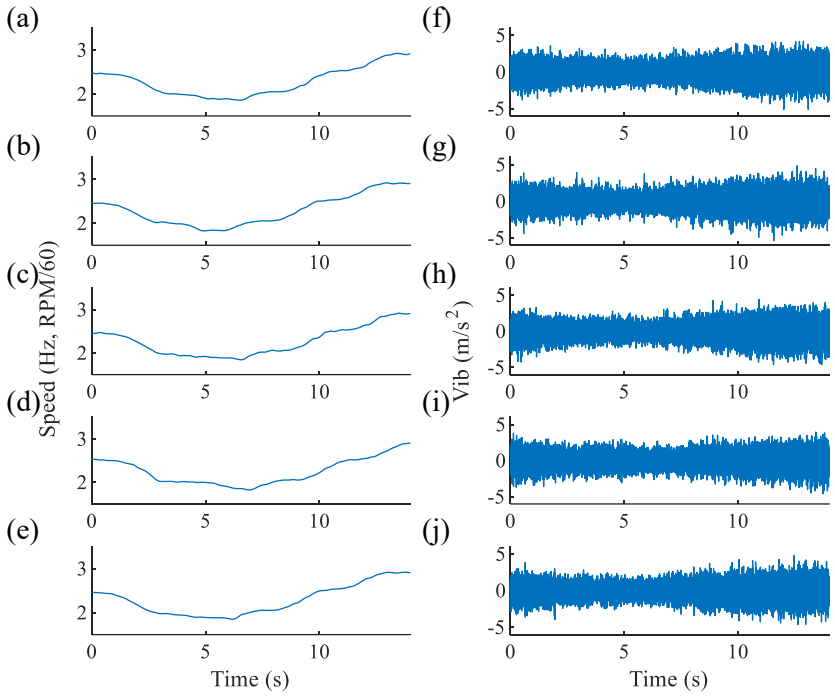

Fig. 3.13: Faulty vibration signals from UofA experiment: (a) ~ (e) speed profiles; (f) ~ (j) time waveform of faulty vibration signals, corresponding to the fault level F1, F2, ..., F5, respectively.

3.3.2.1 Sparse FP-AR modeling

During the sparse FP-AR modeling, the training and testing signals as shown in Fig. 3.12 are used. The initial set of functional spaces is configured as $S_{F,i}^{ini} = S_F^{ini} = \{0, 1, 2, ..., 7\}$ (i.e., basis functions = $\{1, \omega_t, \omega_t^2, ..., \omega_t^7\}$). The initial set of AR terms S_{AR}^{ini} is determined as {1, 2, ..., 600} by BIC. The λ is configured as $[0, 1\times10^{-8}, 1\times10^{-7}, ..., 1\times10^{-1}]$. Fig. 3.14 shows the CVMSE under each λ. The minimum CVMSE appears while $\lambda=1\times10^{-8}$ whereas the $\lambda = 1\times10^{-5}$ returns a model which has a CVMSE within one standard error of the minimum CVMSE. Therefore, $\lambda = 1\times10^{-5}$ is chosen. Table 3.4 lists a portion of the identified model structure and model parameters. Many parameters have been forced to 0, thus results to a spare FP-AR model. In total, the sparse FP-AR model has 3512 non-zero coefficients spread out in the $\hat{\mathbf{a}} = \left[\hat{a}_{i,j}\right]$ for $i \in \{1, 2, ..., 600\}$ and $j \in \{0, 1, 2, ..., 7\}$.

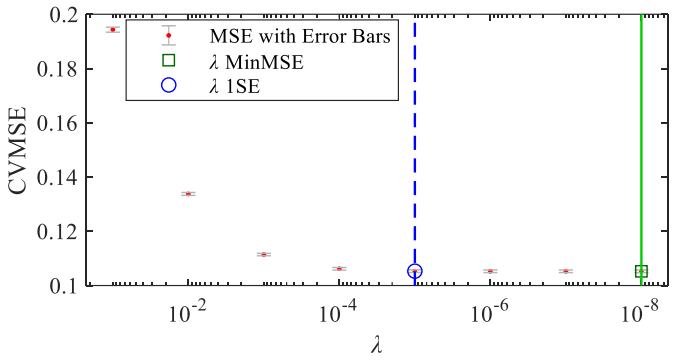

Fig. 3.14: CVMSE of lasso Fit. The CVMSE when λ =0 was not shown due to the log x-axis, and was 0.1052±0.0005 which is the same as the CVMSE when $\lambda = 1\times10^{-8}$.

Table 3.4: A portion of estimated parameters a_{ij} for representing UofA experimental signal

	t-1	t-2	t-3	t-4	t-5	t-6	t-7	...	t-600
1	-1.1703	0.5790	-0.3906	0.0595	0.1966	0.1476	-0.0102	...	0
ω_t	0.0604	0	0.1009	-0.0872	0	0.0427	0.0047	...	0
ω_t^2	0	-0.0205	0.0471	-0.1321	0	0.0892	0.0815	...	0.0212
ω_t^3	0.2180	-0.0870	0.0206	-0.0315	0	0.0144	0.0187	...	0
ω_t^4	0	-0.0107	0.0182	-0.0018	-0.1373	0	0	...	0
ω_t^5	0	0	0	0	-0.0204	0	0	...	0
ω_t^6	0	-0.0210	0.1134	0	0	0	0	...	0
ω_t^7	0.0166	-0.0393	0	0.0074	0	0.0286	0.0033	...	-0.0090

3.3.2.2 Conventional FP-AR modeling

During the model AR order selection, the initial set of functional spaces is configured as $S_{F,i}^{ini} = S_F^{ini} = \{0, 1, 2, ..., 7\}$ (i.e., basis functions = $\{1, \omega_t, \omega_t^2, ..., \omega_t^7\}$). The BIC values for each AR order $n_a = \{1, 2, ..., 800\}$ are shown in Fig. 3.15(a). The optimal model order is chosen as n_a = 400 when minimizing the BIC. The optimal functional space is a 3-dimensional spaces $\{1, \omega_t^6\}$, as shown in Fig. 3.15(b). Functional basses $\{\omega_t, \omega_t^2, \omega_t^3\ \omega_t^4, \omega_t^5, \omega_t^7\}$ are consecutively dropped based on the backward deletion scheme. In total, the conventional FP-AR model has 800 coefficients $\hat{\mathbf{a}} = \left[\hat{a}_{i,j}\right]$ for $i \in \{1, 2, ..., 400\}$ and $j \in \{0, 6\}$.

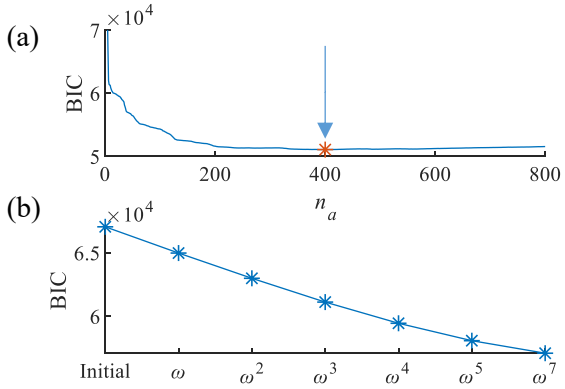

Fig. 3.15: FP-AR model structure selection for the UofA gearbox vibration signal: (a) AR order selection; and (b) BIC values obtained by backward deletion scheme

3.3.2.3 Performance comparison

Same as the simulation study, we evaluate the performance of sparse and conventional FP-AR models in terms of 1) modeling accuracy and 2) the benefits that improved model accuracy can bring to the fault detection and severity assessment of a gearbox under VS.

Fig. 3.16 shows the time waveforms of the residuals-of-testing from both models. The amplitude of residuals-of-testing from both models has decreased comparing with raw signals. From Fig. 3.16(b, c), the mean and variance of the residuals-of-testing are observed to be time-invariant. Thus, the residuals-of-testing has become stationary based on the wider sense definition of a stationary signal.

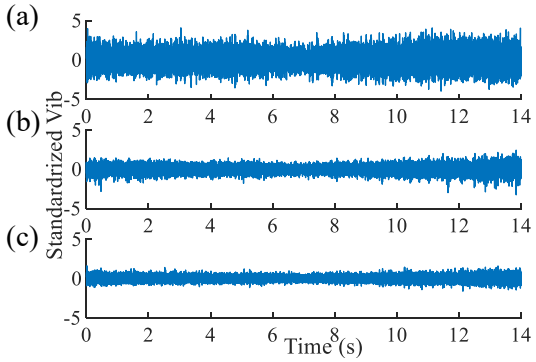

Fig. 3.16: Residuals-of-testing, UofA signals: (a) raw testing signal; (b) residual-of-testing from conventional FP-AR model; (c) residual-of-testing from sparse FP-AR model

Table 3.5 lists the results of sparse FP-AR in comparison with the conventional FP-AR. These results are in good agreement with the simulation study. First, the sparse FP-AR model has an MSE of 0.0710, smaller than the 0.1265 that the conventional FP-AR has. In experimental signals, the environmental noise is unknown hence the irreducible error is unknown. Second, the residual-of-testing of sparse FP-AR model has a *p*-value of 5.489×10^{-3} from Ljung–Box test, whereas its counterpart has a *p*-value of 6.165×10^{-7}. The OSAPs are not shown since the environmental noise is unknown and hence the real gear vibration signal unknown. These results imply that the residual-of-testing of sparse FP-AR model is more random than its counterpart. Meanwhile, the sparse FP-AR model has 3512 non-zero coefficients, whereas the conventional FP-AR model has 800. We conclude that the conventional FP-AR model tends to underfit due to the smaller number of non-zero coefficients. The sparse FP-AR model outperforms the conventional FP-AR model in terms of modeling accuracy. Lastly, the sparse FP-AR model paid about 10 times more than the conventional one in the training stage, and 3 times more in the testing stage.

Table 3.5: Performance of sparse and conventional FP-AR models from UofA signals

Models	MSE of residual,	Ljung–Box test, *p*-value,	CPU time in training, (h)	CPU time in testing, (s)
Con FP-AR	0.1265	6.165×10^{-7}	2.8	3.1
Spa FP-AR	0.0710	5.489×10^{-3}	21.2	9.2

Fig. 3.17 shows the frozen-time spectrums (b, c) obtained from both FP-AR models and non-parametric spectrum (a). The frozen-time spectrum obtained from conventional FP-AR model tends to be stationary (i.e., time-invariant) in the region of [2 13] s, which does not align with the non-parametric spectrum. On the other hand, the frozen-time spectrum obtained from sparse FP-AR model aligns well with the non-parametric spectrum. The sparse FP-AR model has higher modeling accuracy than the conventional one.

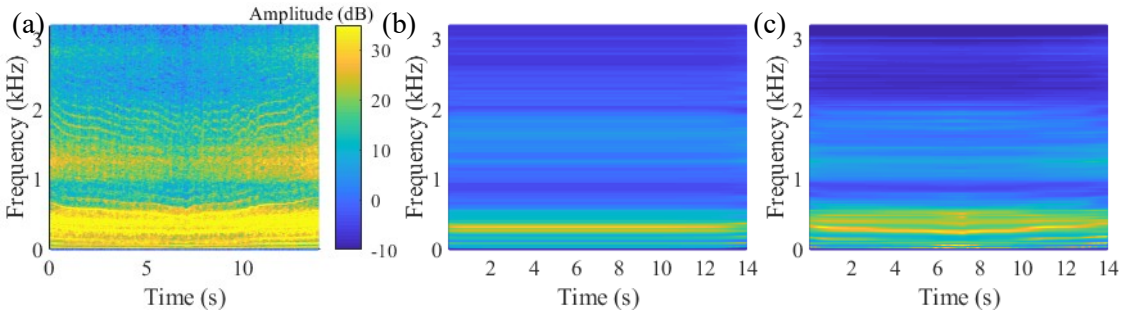

Fig. 3.17: Time-frequency spectrums of UofA signal: (a) non-parametric spectrum of testing signal obtained by MATLAB *spectrogram* function; (b) Frozen-time spectrum obtained from the conventional FP-AR model; (c) Frozen-time spectrum obtained from the sparse FP-AR model. z-axis scales are the same for three plots.

We apply the inverse filter constructed from sparse FP-AR model for detecting gear tooth crack faults. Fig. 3.18 shows the obtained residuals by filtering faulty vibration signals (shown in Fig. 3.13) using inverse filters constructed from both sparse and conventional FP-AR models. Residuals from two FP-AR models seem similar. Next, we calculate PMI to numerically compare these residuals.

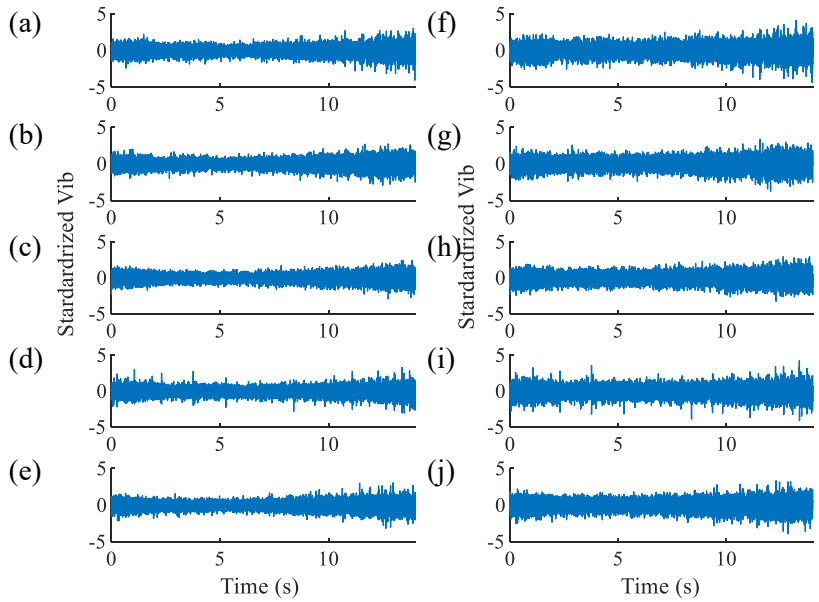

Fig. 3.18: Residuals of faulty vibration signals (UofA signal) through inverse filtration where (a)~(e) corresponds to F1, F2, ..., F5, respectively, and are from conventional FP-AR model. (f)~(j) corresponds to F1, F2, ..., F5, respectively, and are from sparse FP-AR model.

Fig. 3.19 shows the NPMI for detecting five levels of tooth crack faults. The raw signal failed to detect each fault levels. The NPMI calculated from res-con ascends in general with the increasing of fault severities, especially for F3 to F5. The NPMI calculated from the res-spa shows a higher NPMI than the conventional one. The ascending trend of res-spa is also monotonic, whereas the res-con decreases from F2 to F3 and from F4 to F5, which implies that res-spa better assesses the severities of faults. The result is in good agreement with the simulation study.

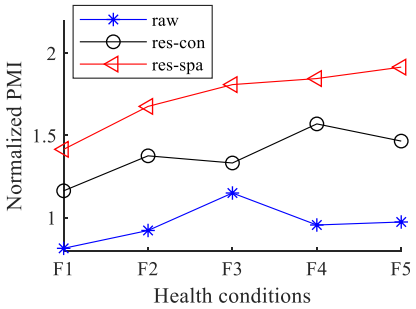

Fig. 3.19: Normalized PMI for detecting five ascending levels of tooth crack faults, obtained from UofA signal

3.3.3 UofP dataset

This experimental dataset is provided by the University of Pretoria (UofP), South Africa and details about the experiments and the data collected are given in [76], [134]. We will use this independent experimental dataset to show the applicability of the proposed FP-AR model.

Fig. 3.20 shows the experimental setup that contains one step-down and two step-up helical gearboxes, an electrical motor and an alternator. The electrical motor drives the whole drive train, whereas the alternator generates load torque. The step-up helical gearbox (labelled as 6) is selected as the test gearbox, which has one mesh pair. The gear and pinion of this gearbox have 37 teeth and 20 teeth, respectively. The input shaft is

connected to the gear, whereas the output shaft connected to the pinion of the helical gearbox.

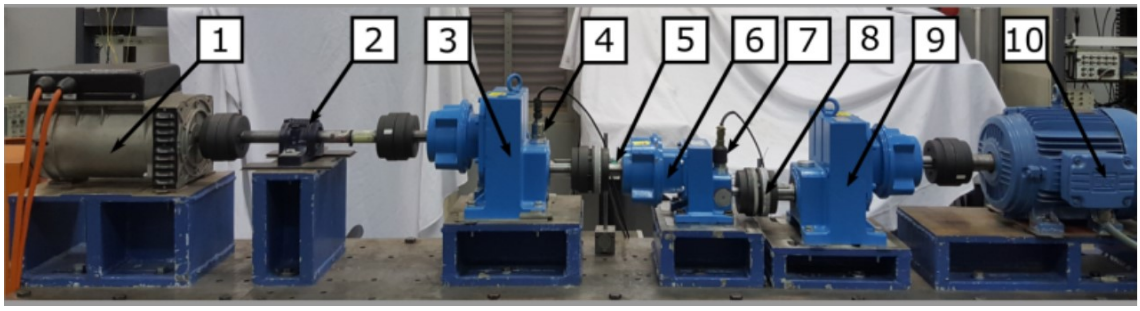

Fig. 3.20: UofP experimental setup: 1, Alternator; 2, Support bearing; 3, Helical gearbox; 4, Accelerometer; 5, Proximity probe applied to key of shaft; 6, Helical gearbox (Test gearbox); 7, Accelerometer (with a sensitivity of 500 mV/g); 8, Zebra-strip shaft encoder and optical switch (88 pulses per revolution); 9, Helical gearbox; 10, Electrical motor. [76], [134]

In addition to the components shown in Fig. 3.20, another tri-axle accelerometer was mounted on the target gearbox (labelled as 6 in Fig. 3.20). Fig. 3.21 shows the tri-axle accelerometer mounted on the bearing housing of the input shaft.

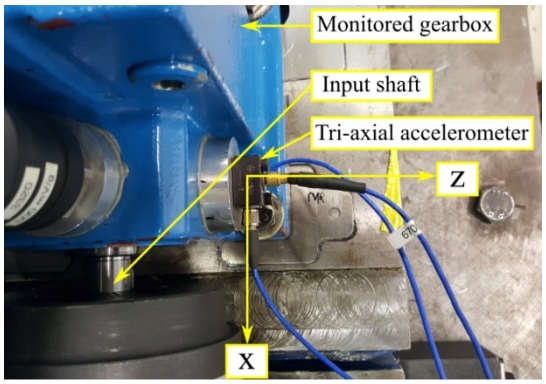

Fig. 3.21: The 100 mV/g tri-axial accelerometer. The input shaft of the test gearbox is seen and the positive *x*-direction points to the electrical motor. [76], [134]

At first, 100 data files were acquired from a healthy gearbox. Each data file lasted 20 s. The sampling frequency was 25.6 kHz. When collecting each data file, the electrical motor (labelled as 10 in Fig. 3.20) drove the transmission train such that the rotating speed of the object gearbox followed a sinusoidal-like profile with a period equal to 10s. The alternator (labelled as 1 in Fig. 3.20) generated a load torque positively correlated to the

speed. Thereafter, the gearbox was disassembled so that the gear can be artificially damaged. An initial tooth crack as shown in Fig. 3.22 (left) was machined with the spark erosion technique and had the following dimensions: ~0.3 mm high to the tooth root, 50% of the tooth width deep, and through the entire face of the tooth. The gearbox was assembled with the artificially damaged gear. A run-to-failure experiment was then conducted which lasted around 21 days of continuous running. At last, the tooth was missing, as shown in Fig. 3.22 (right). During the run-to-failure process, 1400 data files were collected with an equal time interval. Each data file was collected under the same speed and load conditions as the data files collected under the healthy state.

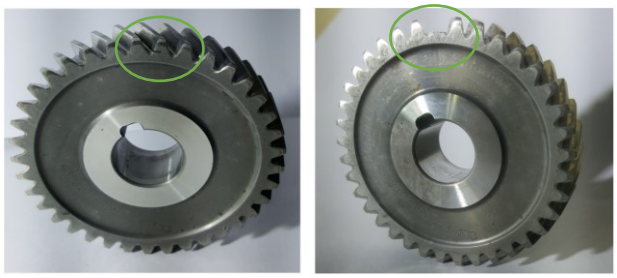

Fig. 3.22: The gear before and after the run-to-failure experiment. [76], [134]

In this chapter, the signal from the vibration sensor labelled as 7 will be used. The vibration signal was low passed using an FIR filter with a cutoff frequency f_c = 3.2 kHz, and then downsampled from f_s =25.6 kHz to f_s =6.4 kHz. The resulting sampling frequency fulfills the principle that f_s needs to be higher than 20 times of the gear meshing frequency to cover the 10th order harmonic of the gear meshing frequency. Fig. 3.23 shows two segments of baseline vibration, as well as their corresponding speed profiles of the input shaft acquired by the zebra-strip shaft encoder. The left-side baseline vibration and its corresponding speed profile are used as training data, whereas the right-side as testing data. Except for this baseline vibration, 30 additional baseline data files are selected. Each data file follows the speed signal as shown in Fig. 3.23(c). The duration of the training and

testing signal is 6.5 s, truncated from the 20 s long signal. This duration lasted more than 15 cycles of object gear rotation.

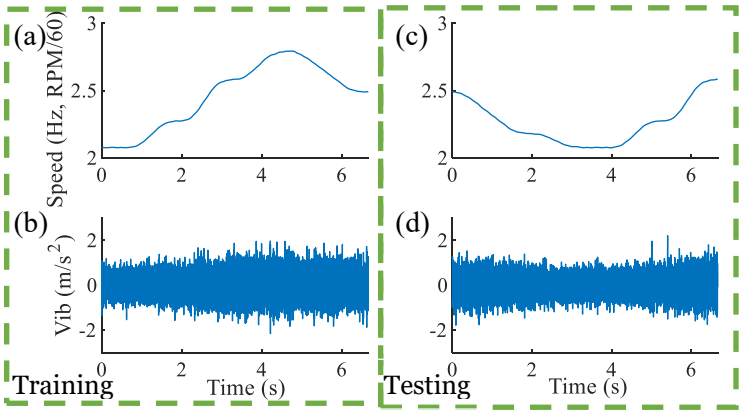

Fig. 3.23: Baseline vibration signals from UofP experiment: (a) speed profile of training signal, (b) time waveform of training vibration signal, (c) speed profile of testing signal, (d) time waveform of testing vibration signal.

To reduce the data size, 170 faulty data files are selected from the total 1400 data files. The selected data files have a time interval of 3 h between consecutive data files, and will be used in this chapter. Each data file contains 6.5 s of vibration signal as well as the speed signal. The right side of Fig. 3.24 shows the time waveforms of faulty vibrations from data file #30, #100, and #170, whereas the left side their corresponding speed profile. We can visually observe that the faulty vibration from data file #170 is peakier than the other two.

3.3.3.1 Sparse FP-AR modeling

During the sparse FP-AR modeling, the training and testing signals as shown in Fig. 3.23 are used. The initial set of functional spaces is configured as $S_{F,i}^{ini} = S_F^{ini} = \{0, 1, 2, ..., 7\}$ (i.e., basis functions = $\{1, \omega_t, \omega_t^2, ..., \omega_t^7\}$). The initial set of AR terms S_{AR}^{ini} is determined as {1, 2, ..., 963} by BIC. The λ is configured as $[0, 1\times10^{-8}, 1\times10^{-7}, ..., 1\times10^{-1}]$. Fig. 3.25 shows the CVMSE under each λ. The minimum CVMSE appears when $\lambda=1\times10^{-4}$. Meanwhile, $\lambda = 1\times10^{-3}$ returns a model which has a CVMSE exceeds one standard error of the minimum CVMSE.

Therefore, λ = 1×10^{-4} is chosen. Table 3.6 lists a portion of the identified model structure and model parameters. Many parameters have been forced to zero, resulting in a spare FP-AR model. In total, the sparse FP-AR model has 3598 non-zero coefficients spread out in the $\hat{\mathbf{a}} = \left[\hat{a}_{i,j} \right]$ for $i \in \{1, 2, ..., 963\}$ and $j \in \{0, 1, 2, ..., 7\}$.

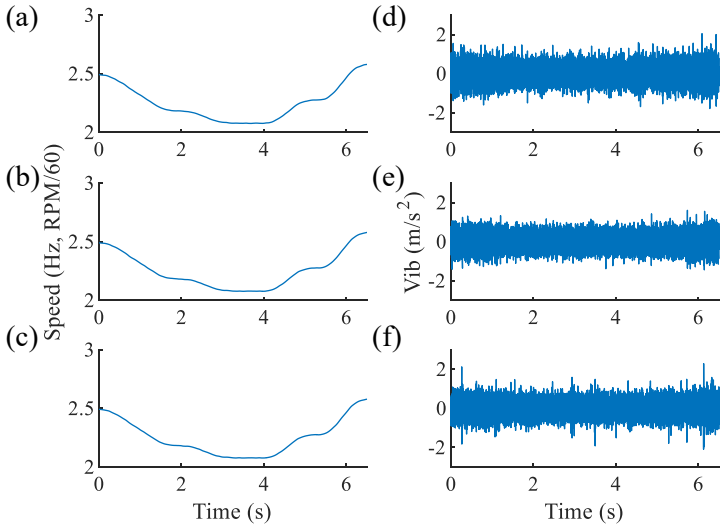

Fig. 3.24: Faulty vibration signals from UofP experiment: (a) ~ (c) speed profiles; (d) ~ (f) time waveform of faulty vibration signals, corresponding to faulty data file #30, #100, and #170, respectively.

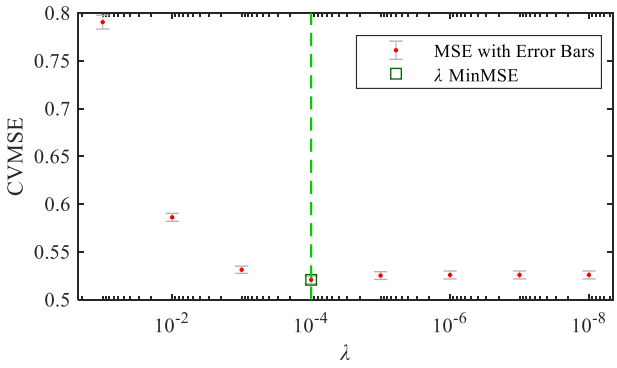

Fig. 3.25: CVMSE of lasso fit from UofP experimental signal. The CVMSE when λ =0 was not shown due to the log x-axis and was 0.5259±0.0042 which is the same as the CVMSE when λ = 1×10^{-8}.

Table 3.6: A portion of estimated parameters a_{ij} for representing UofP baseline vibration

	t-1	t-2	t-3	t-4	t-5	t-6	t-7	...	t-963
1	-0.1023	-0.1829	0.1631	0.5552	-0.2451	-0.2679	0	...	0.0173
ω_t	0.0956	0.1075	0.1426	-0.1328	-0.2054	0.2097	0.1183	...	0
ω_t^2	0.1205	-0.0423	-0.0870	0	0.1294	-0.0119	-0.1311	...	-0.0416
ω_t^3	0	-0.0607	-0.2454	0.0107	0.6043	-0.0042	-0.2012	...	0
ω_t^4	0	0	0	0.0207	0	0	0	...	0
ω_t^5	0	0	0	0	0	0	0	...	0
ω_t^6	0	0	0	0	0	0	0	...	0
ω_t^7	0.0367	0.0210	0.1822	0.1568	-0.2605	-0.0803	0.1251	...	0.0378

3.3.3.2 Conventional FP-AR modeling

During the model AR order selection, the initial set of functional spaces is configured as $S_{F,i}^{ini} = S_F^{ini} = \{0, 1, 2, ..., 7\}$ (i.e., basis functions = $\{1, \omega_t, \omega_t^2, ..., \omega_t^7\}$). The BIC values for each AR order $n_a = \{1, 2, ..., 1000\}$ are shown in Fig. 3.26(a). The optimal model order is chosen as $n_a = 642$ when minimizing the BIC. Then, the optimal functional space is a 1-dimensional space {1}, as shown in Fig. 3.26(b). Functional basis $\{\omega_t, \omega_t^2, \omega_t^3, \omega_t^7, \omega_t^6, \omega_t^5, \omega_t^4\}$ are consecutively dropped based on the backward deletion scheme. Note here conventional FP-AR reduces to a classical AR model. In total, the conventional FP-AR model has 642 coefficients $\hat{\mathbf{a}} = \left[\hat{a}_{i,j}\right]$ for $i \in \{1, 2, ..., 642\}$ and $j \in \{0\}$.

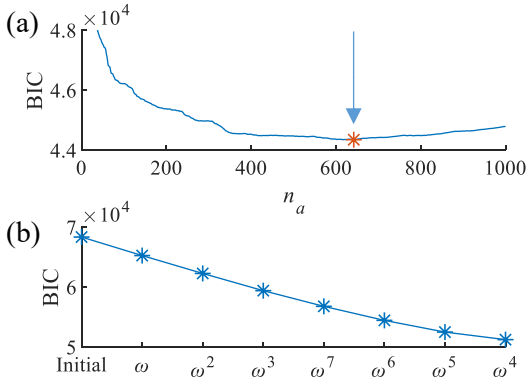

Fig. 3.26: FP-AR model structure selection for the UofP gearbox vibration signal: (a) AR order selection; and (b) BIC values obtained by backward deletion scheme

3.3.3.3 Performance comparison

Fig. 3.27 shows the time waveform of the residuals-of-testing from both models. The amplitude of residuals-of-testing from both models has decreased in comparison with the raw testing signal. From Fig. 3.27 (b, c), the mean and variance of the residuals-of-testing are observed to be time-invariant. Thus, the residuals-of-testing has become stationary based on the wider sense definition of a stationary signal.

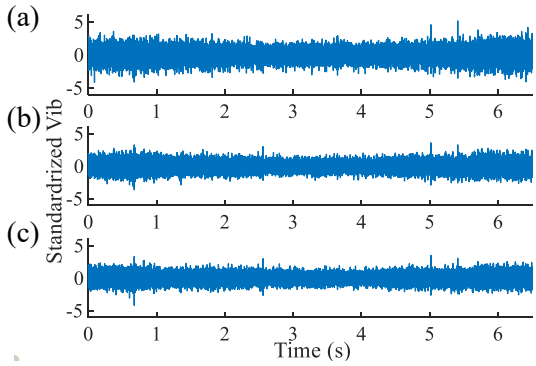

Fig. 3.27: Residuals-of-testing, UofP signals: (a) raw testing signal; (b) residual-of-testing from conventional FP-AR model; (c) residual-of-testing from sparse FP-AR model

Table 3.7 listed the results of sparse FP-AR in comparison with the conventional FP-AR. These results are in good agreement with previous studies. First, the sparse FP-AR model has an MSE of 0.3971, smaller than 0.4547 that the conventional FP-AR has. These MSEs are larger than that from simulation and UofA signals, which implies that independent signals behave differently. Second, the residual-of-testing of sparse FP-AR model has a *p*-value of 0.0052 from Ljung–Box test, whereas the conventional FP-AR model has 8.2800×10^{-12}. Meanwhile, the sparse FP-AR model has 3598 non-zero coefficients, whereas the conventional FP-AR model has 642. We conclude that the conventional FP-AR model tends to underfit due to the smaller number of non-zero coefficients, and the sparse FP-AR model outperforms the conventional FP-AR model in terms of modeling

accuracy. Lastly, the sparse FP-AR model consumes about 9 times more CPU time than the conventional one in the training stage, and 3 times more in the testing stage.

Table 3.7: Performance of sparse and conventional FP-AR models from UofP signals

Models	MSE of residual	Ljung–Box test, *p*-value,	CPU time in training, (h)	CPU time in testing, (s)
Con FP-AR	0.4547	8.2800×10^{-12}	1.5	2.8
Spa FP-AR	0.3971	0.0052	12.5	8.9

Fig. 3.28 shows the frozen-time spectrums (b, c) obtained from both FP-AR models and non-parametric spectrum (a). Since conventional FP-AR reduced to an AR model, the frozen-time spectrum is stationary, which does not align with the non-parametric spectrum. On the other hand, the frozen-time spectrum obtained from sparse FP-AR model aligns well with the non-parametric spectrum by tracking the time-varying components. The sparse FP-AR model has higher modeling accuracy than the conventional one.

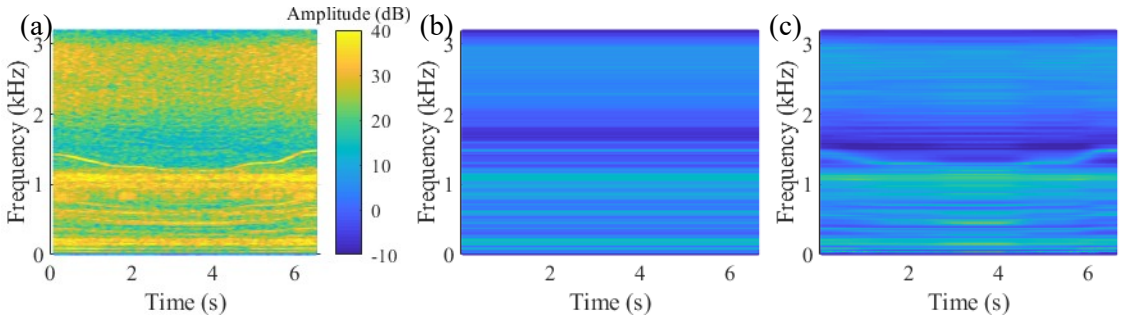

Fig. 3.28: Time-frequency spectrums of UofP signal: (a) non-parametric spectrum of testing signal obtained by MATLAB *spectrogram* function; (b) Frozen-time spectrum obtained from the conventional FP-AR model; (c) Frozen-time spectrum obtained from the sparse FP-AR model. z-axis scales are the same for three plots.

We apply the inverse filter constructed from sparse FP-AR model for detecting the progression of gear tooth crack faults. Fig. 3.29 shows the obtained residuals by filtering faulty vibration signals (shown in Fig. 3.24) using inverse filters constructed from both

sparse and conventional FP-AR models. Residuals from sparse FP-AR models seem "peakier" than the residuals from conventional one.

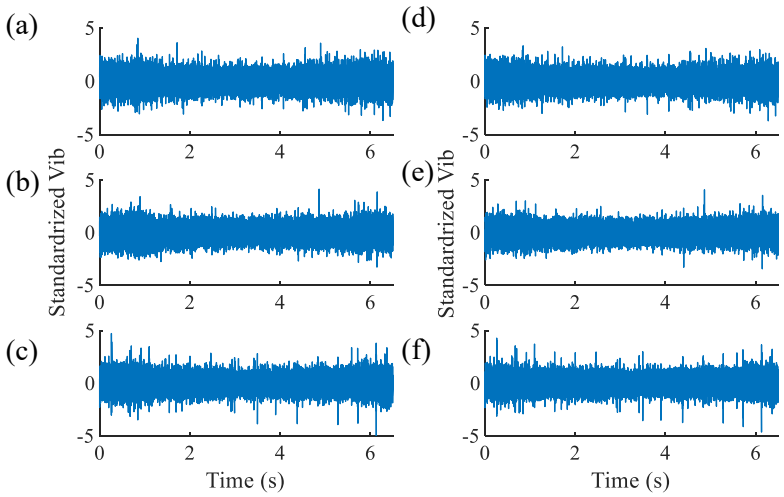

Fig. 3.29: Residuals of faulty vibration signals (UofP signal) through inverse filtration where (a)~(c) corresponds to data file #30, #100, and #170, respectively, and are from conventional FP-AR model. (d)~(f) corresponds to data file #30, #100, and #170, respectively, and are from sparse FP-AR model

Fig. 3.30 shows the NPMI for assessing the progression of tooth crack faults. Fig. 3.30(a) shows the NPMI for 30 data files of baseline vibration, whereas Fig. 3.30(b) shows the NPMI for 170 data files of faulty vibrations that are collected with a time interval of 3 h, along with smoothing curves obtained by MATLAB *spaps*() function. The NPMI from baseline vibrations oscillates around 1. For detecting tooth crack faults, we can treat the maximum NPMI from baseline vibrations as threshold and detect tooth crack fault whenever the NPMI exceeds such threshold. For the NPMI from the raw signal, many of the NPMIs of faulty data files #1~160 locate below the threshold (1.22). These NPMIs of faulty data files #160~170 that has severe faults locate above the threshold. On the other hand, the NPMIs from res-con and res-spa of faulty vibrations are above their threshold (1.14). Both of them can detect tooth crack faults. This is not surprising because the UofP experiment started from an initial crack with a depth of 50% tooth width and through the

entire face of the tooth, which is considerably severe compared with the tooth crack fault in UofA dataset. Noticeably, thanks to more accurate baseline vibration modeling and hence more accurate removal of baseline vibration, the NPMI from res-spa is 0.12 higher than its counterpart on average by observing the smoothing curve. This implies that the sparse FP-AR model extracts more fault induced features than conventional FP-AR model, and hence better assesses the progression of tooth crack faults.

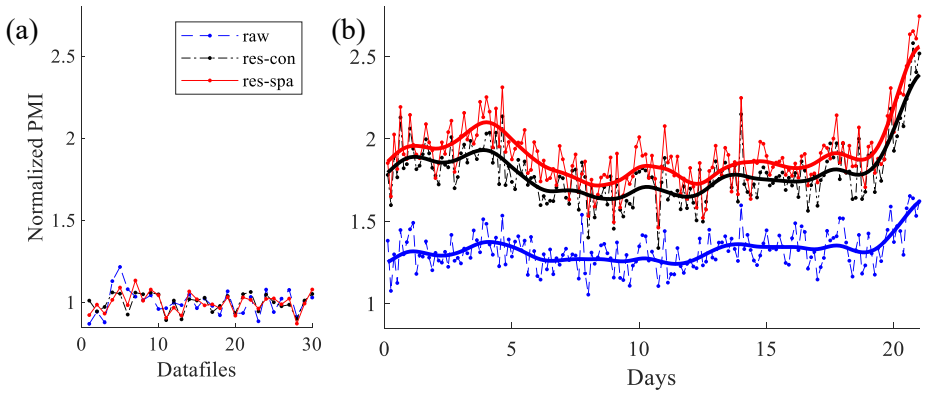

Fig. 3.30: Normalized PMI for detecting the progression of tooth crack faults, obtained from UofP signal. (a) healthy data files; (b) faulty days. Smoothing curves were obtained for each NPMI under faulty cases.

The proposed sparse FP-AR model spent about 25 h in training using the UofA signal that has 89600 data points, and about 13 h using the UofP signal that has 44800 data points. Since this training process is completed offline, the length of time required is not very critical. Besides, the computational efficiency can be improved by implementing the algorithm in a more powerful industrial computer. The computational time in testing stage primarily determines whether the TSMBM based on sparse FP-AR model can be implemented in online condition monitoring systems. The inverse filter constructed by the sparse FP-AR model spent 9.2 s for processing the UofA signal that has 89600 data points, and 8.9s for processing the UofP signal that has 44800 data points. Therefore, the TSMBM based on sparse FP-AR model has a potential for the implementation in online condition

monitoring systems. More importantly, the significant improvement on fault detection severity assessment provided by the sparse FP-AR model could easily outweigh the additional offline CPU time.

3.4 Discussions

When it comes to the online implementation of the TSMBM, users first need to apply the inverse filter to process the newly collected signals and then calculate indicator(s) from the residual to assess the health condition of the gearbox. The baseline sparse FP-AR model needs to be trained ahead of the online implementation. In other words, the model training is completed offline, whereas the testing phase is implemented online for gearbox health monitoring.

To reliably estimate the model parameters (i.e., reduce the variance of estimated model parameters), the number of data points needs to be at least 10 times greater than the number of parameters to be estimated, according to Knofczynski and Mundfrom [147]. The parameters of the proposed sparse FP-AR model are reliability estimated. In the simulation study, the number of training data points is 7500 which is 25 times the number of parameters (i.e., 297). In experimental studies, this ratio of data point over parameters is 25.5 when analyzing UofA signals and 12.5 when analyzing UofP signals.

The TSMBM works theoretically different compared with the generalized synchronous average (GSA) method [27]. TSMBM works well with just a few training data, whereas the GSA requires relatively more training data for the removal of baseline vibration components.

3.5 Conclusions

In this chapter, a sparse FP-AR model is proposed. The sparse FP-AR model uses a sparse AR set and non-identical functional spaces. A new model selection procedure by adopting the LASSO has been developed for obtaining such a sparse FP-AR model. The sparse FP-AR model has been validated using simulation signals from a simulation model for a fixed-axis gearbox and experimental signals from two independent fixed-axis gearbox test-rigs. Results show that although the proposed sparse FP-AR model consumes more computational time than the conventional FP-AR model, it has higher modeling accuracy. We can use the residuals of the sparse FP-AR model for the fault detection and severity assessment when the testing signals were collected under deterministic speed profiles. The improved modeling accuracy benefits the health condition assessment of a gearbox under VS in detecting earlier faults and better assessing the progression of faults.

This chapter has focused on improving the accuracy of baseline vibration modeling. The residuals of the accurate baseline model can be employed for the fault detection and severity assessment when the testing signals were collected under deterministic speed profiles. However, such baseline model residual-based methods are inapplicable to the cases when the testing signals were collected under random speed variations which typically occurs in wind turbine gearboxes, as explained in Section 1.2.2.4. Therefore, in the next chapter, we will develop a time series model-based method for the gear tooth crack detection and severity assessment under the random speed variation.

4

A time series model-based method for gear tooth crack detection and severity assessment under random speed variation

In industries (e.g., wind power), gearboxes operate under random speed variations. A condition monitoring system is expected to detect faults and assess their severity using vibration signals collected under different speed profiles. This chapter presents a time series model-based method for gear tooth crack detection and severity assessment under random speed variation (the research topic #3 as introduced in Section 1.3). The materials of this topic have been documented in a journal paper [109] which is under review.

4.1 Introduction

Gearboxes often operate under variable conditions. Vibration-based condition monitoring of gearboxes needs to account for such variation of operating conditions (OCs) since the OC changes the vibration pattern.

A few studies have been reported for the condition monitoring of gearboxes under random speed variations. Wang *et al.* [148] proposed a matching synchrosqueezing transform (MSST) and presented its application in fault detection of a wind turbine gearbox under random speed variation. The MSST can qualitatively detect a slight tooth spall through visual observation of the time-frequency spectrum. Villa *et al.* [74] found that the diagnostic features (e.g., root mean square) are dependent on the rotating speed and load.

The change of such dependency can reflect the deterioration of a gearbox. Santos *et al.* [75] considered fault diagnosis of imbalance and misalignment faults under random speed and load variation. The authors extracted numerous features from the vibration signal and fed them along with speed and load information into a support vector machine (SVM). Although the features they extracted are sensitive to speed/load changes, SVM is expected to learn the dependency between these features and the speed/load and still return a good classification result. When applying the approach by Santos *et al.* [75] for gear tooth crack detection and severity assessment, we need to extract proper features that are sensitive to the gear tooth crack. Recently, Schmidt *et al.* [76] presented a comprehensive novelty diagnostic method. The method first identifies the OC state of the gearbox by assuming that the speed does not vary significantly within a rotating cycle. Secondly, the method calculates a negative log-likelihood (NLL) based on machine health condition features and the OC state. The NLL can reflect the propagation of the localized gear fault well. The novelty diagnostic method may be adequate for smooth random speed variations. However, when speed varies rapidly, the vibration signal within a rotating cycle may be subjected to significant variations.

Nowadays, time-series model-based methods (TSMBMs) are widely employed in the field of structural health monitoring (i.e., fault detection, localization, mode identification, and severity assessment). Advantages of TSMBMs include 1) being free from physics or analytical information, 2) need only a few signals for training, and 3) inherent accounting for uncertainties from measurement and operational conditions [44], [108]. There is a large volume of literature on time series models, including review papers [44], [56].

For fault detection, TSMBM is based on the identification of a baseline time series model for representing the vibration signal under a healthy state and various OCs.

Kopsaftopoulos *et al.* [91] used a vector functional pooled autoregression (VFP-AR) model for globally representing the dynamics of a healthy aerospace structure under multiple flight states. Hois *et al.* [51] presented a functional pooled vector autoregression model-based method for fault detection of a composite cantilever beam under variable temperature conditions. Aravanis *et al.* [149] presented a functional pooled autoregression model with an exogenous excitation-based method for fault detection of a railway suspension under variable loading conditions. The method can detect the suspension stiffness reduction without the measurement of loading conditions during the testing stage. Functional pooled (FP) models require measured OC variables (e.g., temperature, loading). For time-varying structures without a measurable OC variable, time-dependent AR (TAR) models may be employed. Spiridonakos *et al.* [136] proposed an adaptable functional series TAR for representing the vibration signals from a healthy pick-and-place mechanism consisting of two coaxially aligned linear motors. Functional series are adaptively estimated from the vibration signals. Ma *et al.* [150] presented a kernel recursive extended least-squares time-dependent autoregression and moving average method for representing the vibration signals from a healthy beam with a moving steel mass sliding on. The method recursively selects the optimal kernel function basis.

For fault localization, mode identification, and severity assessment, TSMBM is based on the identification of time series model(s) individually or globally for the vibration signals under various locations, modes, and/or severities. Sakaris *et al.* [50] presented a vector functional pooled vector autoregression (VFP-ARX) model-based method for fault detection and localization of a truss structure. They identified a VFP-ARX model to globally represent the vibration signals under various fault locations. The fault is detected and localized via re-estimation of the location variable such that the VFP-ARX model best

fits the current measured vibration signals. Kopsaftopoulos *et al.* [45] presented a VFP-ARX model-based method for fault detection, mode identification, and severity assessment of an aircraft skeleton structure. For each fault mode, they built a VFP-ARX model to represent the vibration signals under various fault locations and severities. Fault detection, localization, and severity assessment are based on the re-parameterized VFP-ARX models of any damage mode. Fault mode identification is based on evaluating the whiteness of residuals from each VFP-ARX model. The fault mode is identified as the one which gives the whitest residual. Spiridonakos *et al.* [102] presented a functional series TAR (FS-TAR) model-based mode identification method for a pick-and-place mechanism. FS-TAR models are built for each fault mode. Fault mode identification is based on statistically comparing the model parameters from a testing signal with these parameters obtained from the training stage.

In condition monitoring of gearboxes, TSMBM is limited to the multiple levels of constant speed condition [85], [89] and cyclic speed variation [84], [108]. TSMBM has immense potential for fault detection and severity assessment of gearboxes under random speed variations. Therefore, this chapter aims to develop a time series model-based method for the gear tooth crack detection and severity assessment under the random speed variation, as well as to benchmark its performance with the novelty diagnostic method [76] and SVM-based method [75].

We first adopt the baseline removal technique [108] to pre-process the raw vibration signal and obtain the impulsive vibration signal (Hereafter, we denote the signal after baseline removal as the impulsive vibration signal.) with enhanced crack-induced signal (impulses) to noise ratio. The rotating speed and phase of a gearbox are measurable using an encoder. Therefore, we select the VFP-AR model [45], [91] for the representation of

impulsive vibration signals under random speed conditions. The rotating speed and phase are considered as covariates in a VFP-AR model, given that the rotating speed affects the amplitude of tooth crack induced impulses, and the rotating phase determines the location of these impulses. We propose refined bases of spline (B-splines) for mapping the dependency between AR coefficients and the rotating phase. The refined B-splines are deduced from the periodic B-splines [151] by removing unnecessary bases outside of the location of impulses. The refined B-splines are less complex than the periodic B-splines and hence can avoid overfitting.

We consider the tooth crack detection and severity assessment as a multi-state classification problem (i.e., classify the healthy and different levels of fault), given that the International Organization for Standardization 10816 divides the health condition of rotating machinery into discrete levels [152]. VFP-AR models are built for each level of tooth crack severity (included the healthy state). Fault detection and severity assessment are therefore based on examining the mean squared error (MSE) of VFP-AR model residuals. The final health state is assessed as the severity state of the VFP-AR model which gives the minimal MSE. The effectiveness of the VFP-AR model-based method was validated using a laboratory dataset collected at the University of Pretoria, South Africa. The performance of the VFP-AR model-based method is compared with the novelty diagnostic method [76] and SVM-based method [75].

The main contributions of this chapter include:

1) Rotating speed and rotating phase are considered as covariates in a VFP-AR model for the representation of impulsive vibration signals under random speed variation.
2) Refined B-splines are proposed for the VFP-AR model when mapping the dependency between AR coefficients and the rotating phase.

3) The performance of the proposed VFP-AR model-based method has improved accuracy than the reported novelty diagnostic method [76] and SVM-based method [75].

The rest of this chapter is organized as follows: Section 4.2 presents the VFP-AR model-based gear tooth crack detection and severity assessment method. Section 4.3 validates the presented method using a laboratory dataset. Section 4.4 discusses some further considerations of the presented VFP-AR model-based method. Conclusions are drawn in Section 4.5.

4.2 VFP-AR model-based fault detection and severity assessment method

This section includes the basics of the VFP-AR model, the proposed refined B-splines, a sparse model identification algorithm, modeling accuracy measures, and the overall scheme of the VFP-AR model-based fault detection and severity assessment method.

4.2.1 VFP-AR model basics

The VFP-AR model was reported in [45], [91] for representing the dynamics of structures under variable OCs. For the representation of impulsive vibration signals under random speed conditions, we consider the rotating speed and rotating phase as covariates in a VFP-AR model, because the rotating speed affects the amplitude of tooth crack induced impulses, and the rotating phase determines the location. The difference equation of such VFP-AR model is as follows

$$y_t = \sum_{i\in\{1,\cdots,n_a\}} f_i\left(\theta_t, \omega_t\right) y_{t-i} + \varepsilon_t \tag{4.1}$$

where y_t and y_{t-i} denote the impulsive vibration signal at time t and t-i, respectively; θ_t $\{\theta_t \in \mathbf{R} \mid 0 \le \theta_t \le 2\pi\}$ denotes the phase at time t; ω_t denotes the rotational speed at time t; n_a is

the order of autoregressive terms; ε_t is the white noise at time t. The $f_i(\theta_t, \omega_t)$ is the i^{th} smoothing function of phase and rotational speed and is approximated by basis expansion approach as follows [100]

$$f_i\left(\theta_t, \omega_t\right) = \sum_{j=1}^{n_p} a_{i,j} G_j\left(\theta_t, \omega_t\right) \tag{4.2}$$

where $a_{i,j}$ stands for the projection parameters; $G_j(\theta_t, \omega_t)$ j =1,2,... n_p is j^{th} basis function constructed by phase and rotational speed; n_p is the number of basis functions. Let $\mathbf{G}(\theta_t, \omega_t) = [G_1(\theta_t, \omega_t)\ ,\ G_2(\theta_t, \omega_t)\ ,\ldots,\ G_p(\theta_t, \omega_t)]^T$. $\mathbf{G}(\theta_t, \omega_t)$ is of size $n_p \times 1$ and results from the Kronecker product (denoted by $\otimes$) of the univariate basis functions [45]

$$\mathbf{G}\left(\theta_t, \omega_t\right) = \mathbf{G}_{rB}\left(\theta_t\right) \otimes \mathbf{G}_p\left(\omega_t\right) \tag{4.3}$$

We propose refined B-splines (to be detailed in Section 4.2.2) and augment the refined B-splines with a constant 1 to obtain $\mathbf{G}_{rB}(\theta_t)$. For $\mathbf{G}_p(\omega_t)$, we select Legendre polynomial basis functions. These give

$$\mathbf{G}_{rB}\left(\theta_t\right) = \left[1, B_{k,r}^{per}\left(\theta_t\right), \cdots, B_{k+n,r}^{per}\left(\theta_t\right)\right]^T \tag{4.4a}$$

$$\mathbf{G}_p\left(\omega_t\right) = \left[\omega_t^0, \omega_t^1, \cdots, \omega_t^{p_\omega}\right]^T \tag{4.4b}$$

where r is the degree of the polynomial spline, k and n specify the location of impulses, $B_{k+\ell,r}^{per}\left(\theta_t\right)$ for $\ell = 0,1,\cdots,n$ are basis functions from periodic B-splines (to be detailed in Section 4.2.2), and p_ω is the order of Legendre polynomial basis functions.

The rationale of augmenting the refined B-splines by a constant 1 is to account for deterministic components if any (e.g., the remained baseline components due to imperfect baseline removal.) in the impulsive vibration signal. As will be detailed in Section 4.2.2, the $B_{k+\ell,r}^{per}\left(\theta_t\right)$ for $\ell = 0,1,\cdots,n$ are significant at the rotating phase θ_t when the cracked

tooth is in mesh, whereas $B_{k+\ell,r}^{per}(\theta_t)=0$ for $\ell=0,1,\cdots,n$ at the rotating phase θ_t when healthy teeth are in mesh. Without augmentation of a constant 1, $f_i(\theta_t, \omega_t)$ = 0 when $B_{k+\ell,r}^{per}(\theta_t)=0$ for $\ell=0,1,\cdots,n$. The VFP-AR model will be unable to represent any deterministic components.

Substituting Eqs. (4.3) and (4.4) into Eq. (4.2), we get a detailed basis expansion form for $f_i(\theta_t, \omega_t)$ as

$$f_i(\theta_t,\omega_t)=\sum_{\ell=0}^{n}\sum_{m=0}^{p_\omega}a_{i,\ell,m}B_{k+\ell,r}^{per}(\theta_t)\omega_t^{\,m}+\sum_{m=0}^{p_\omega}b_{i,m}\omega_t^{\,m} \tag{4.5}$$

where $a_{i,\ell,m}$ and $b_{i,m}$ are the updated projection coefficients, which are equivalent to $a_{i,j}$ in Eq. (4.2) for j =1, 2, ..., n_p and n_p = (n+2)(p_ω+1). In Eq. (4.5), the right-hand side term is resulted from augmenting the refined B-splines with a constant 1.

Applying z-transform [121] to Eq. (4.1), we obtain

$$Y(z)=\frac{1}{1-\sum_{i\in\{1,\cdots,n_a\}}f_i(\theta_t,\omega_t)z^{-i}}e(z)=H(z)e(z) \tag{4.6}$$

where Y(z) is the vibration signals y_t expressed in z-domain; and $e(z)$ is the zero-mean white noise in z-domain. In the viewpoint of the linear time-variant (LTV) system theory, the Y(z) is modeled as the output of an LTV system $H(z)$ excited by a zero-mean white noise $e(z)$.

Based on Eq. (4.6), the inverse filter is of the following transfer function [121]:

$$H^{-1}(z)=1-\sum_{i\in\{1,\cdots,n_a\}}f_i(\theta_t,\omega_t)z^{-i} \tag{4.7}$$

This inverse filter is rotating speed and phase-dependent and with a finite impulse response (FIR).

4.2.2 Refined B-splines

The periodic B-splines reported in [151] have bases uniformly placed across the rotating phase $\theta_t\{\theta_t \in \mathbf{R} \mid 0 \le \theta_t \le 2\pi\}$. We propose to refine the periodic B-splines by removing these bases outside of the location of tooth crack induced impulses. For example, a gear has K teeth and the k^{th} tooth has a crack. The impulses, therefore, locate at the rotating phase when the k^{th} cracked tooth and few subsequent teeth (e.g., $(k+1)^{th}$,..., $(k+n)^{th}$) are in mesh. In other words, impulses locate only at the rotating phase θ_t region $[2\pi k/K, 2\pi(k+n)/K]$. We propose to remove those bases outside of this region. After these removals, the refined B-splines are less complex than the periodic B-splines and hence can avoid overfitting. A comparative study will be conducted in Section 4.3.2 to prove the superiority of the refined B-splines over periodic B-splines.

We now show the mathematical details of the refined B-splines. To start with, the open B-splines given in [151] for θ_t gives

$$\mathbf{G}_B(\theta_t) = \left[B_{1,r}(\theta_t), B_{2,r}(\theta_t), \cdots, B_{K,r}(\theta_t) \right]^T \tag{4.8}$$

where r is the degree of the polynomial spline, K is the number of basis functions, and $B_{i,r}(\theta_t)$ is obtained by means of Cox-de Boor recursion formula [153]

$$B_{\ell,0}(\theta_t) = \begin{cases} 1 & if \;\; \theta_\ell \le \theta_t < \theta_{\ell+1} \\ 0 & otherwise \end{cases} \tag{4.9}$$

$$B_{\ell,r}(\theta_t) = \frac{\theta_t - \theta_\ell}{\theta_{\ell+r} - \theta_\ell} B_{\ell,r-1}(\theta_t) + \frac{\theta_{\ell+r+1} - \theta_t}{\theta_{\ell+r+1} - \theta_\ell} B_{\ell+1,r-1}(\theta_t), \;\; \ell = 1,\ldots,K \tag{4.10}$$

where $0 = \theta_1 = \theta_2 = \cdots = \theta_{r+1} < \theta_{r+2} \cdots < \theta_K = \theta_{K+1} = \cdots = \theta_{K+r} = 2\pi$ is a uniform knot sequence with $K+r$ interior knots. A knot vector is said to be open if its first and last knots

have multiplicity r+1 [151]. A knot vector is said to be uniform if the knots are equispaced [151].

To construct a periodic B-splines, the uniform knot sequence for the open B-splines should be extended to both ends by r steps (i.e., $r \times 2\pi/K$) as

$$-r \times \frac{2\pi}{K} = \theta_1 = \theta_2 = \cdots = \theta_{r+1} < \theta_{r+2} < \cdots < \theta_{K+2r} = \theta_{K+2r+1} = \cdots = \theta_{K+3r} = 2\pi + r \times \frac{2\pi}{K} \quad (4.11)$$

The first and last r number of bases in an open B-splines $\mathbf{G}_B(\theta_t)$ are then discarded. Based on the remaining K number of open B-spline bases, we can construct periodic B-splines by adding together r number of bases laying at the beginning and end [151]:

$$\mathbf{G}_{pB}(\theta_t) = \left[B_{1,r}^{per}(\theta_t), \cdots, B_{K-r,r}^{per}(\theta_t) \right]^T \text{ with } \begin{cases} B_{\ell,r}^{per}(\theta_t) = B_{\ell,r}(\theta_t) + B_{K-\ell,r}(\theta_t) & \ell = 1, \cdots, r \\ B_{\ell,r}^{per}(\theta_t) = B_{\ell,r}(\theta_t) & \text{otherwise} \end{cases} \quad (4.12)$$

Note that the dimension of $\mathbf{G}_{pB}(\theta_t)$ becomes K-r.

We now refine the B-splines given in Eq. (4.12) by removing those bases located outside of the impulses in the periodic B-splines $\boldsymbol{G}_{pB}(\theta_t)$. Suppose k $\{k \in \mathbf{N} | 1 \le k \le K\text{-}r\}$ and n $\{n \in \mathbf{N} | 0 \le n \le K\text{-}r\}$ specifies the location of impulses, the refined B-splines become

$$\begin{aligned} \tilde{\boldsymbol{G}}_{rB}(\theta_t) &= \left[0, \cdots, 0, B_{k,r}^{per}(\theta_t), \cdots, B_{k+n,r}^{per}(\theta_t), 0, \cdots, 0 \right]^T \\ &= \left[B_{k,r}^{per}(\theta_t), \cdots, B_{k+n,r}^{per}(\theta_t) \right]^T \end{aligned} \quad (4.13)$$

where $\tilde{\boldsymbol{G}}_{rB}(\theta_t)$ if of size (n+1)×1. Note that the kept bases in Eq. (4.13) are consecutive. In Eq. (4.13), we have assumed (k+n) < (K-r). When (k+n) ≥ (K-r), bases located in the end and/or beginning of the periodic B-splines (e.g., $B_{K-r,r}^{per}(\theta_t), B_{1,r}^{per}(\theta_t), \ldots, B_{(k+n)-(K-r),r}^{per}(\theta_t)$) may be kept according to the periodic nature.

Fig. 4.1 shows examples of cubic (r = 3) B-splines with θ_t $\{\theta_t \in \mathbf{R} \mid 0 \le \theta_t \le 2\pi\}$, K = 10, k = 5, and n = 1. Fig. 4.1(a) shows open B-splines which has 10 bases. The beginning θ_t = 0 and end θ_t = 2π points are not continuously connected, and hence the open B-splines are

non-periodic. Fig. 4.1(b) shows the periodic B-splines which have 7 bases uniformly placed across $\theta_t \in [0, 2\pi]$. The beginning $\theta_t = 0$ and end $\theta_t = 2\pi$ points are continuously connected. Therefore, it has a periodic property. Fig. 4.1(c) shows the refined B-splines which are constructed from the periodic B-splines shown in Fig. 4.1(b) by removing those dashed line bases and only keeping two solid line bases. It is evident that the refined B-splines have a fewer number of bases and are therefore less complex than the periodic B-splines.

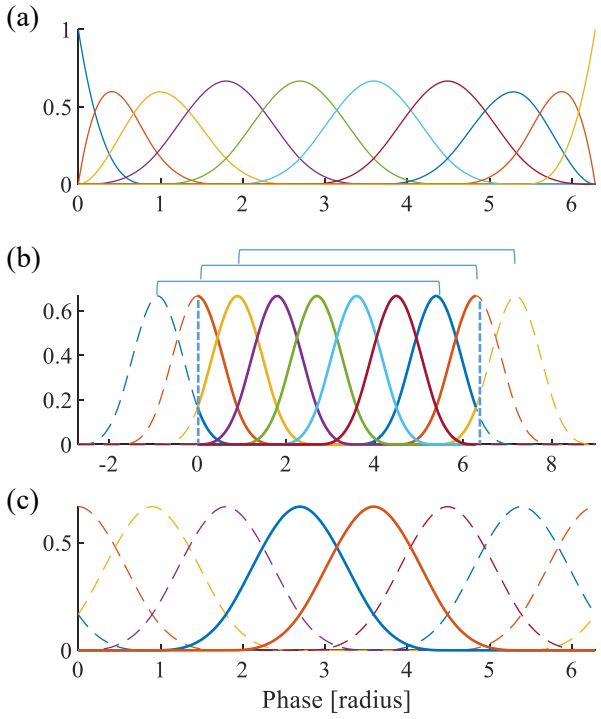

Fig. 4.1: B-splines: (a) open; (b) periodic; (c) refined.

4.2.3 VFP-AR model identification

Suppose we have available training and validation datasets from a gearbox under each known health state. Both training and validation datasets include a segment of impulsive vibration signals y_t, as well as the synchronous acquired rotating phase θ_t and speed ω_t for $t = 1, 2, \ldots, N$. The speed ω_t should cover a wide range of variation. These datasets will be

used for identifying a VFP-AR model for the representation of the vibration signal under each known health state.

Given a VFP-AR model with its hyperparameters n_a, K, k, n, r, and p_ω, ordinary least squared (OLS) estimator returns parameters $\mathbf{a}$ = [$a_{i,\ell,m}$ $b_{i,m}$] for $i \in \{1,\cdots,n_a\}$, $\ell \in \{0,\cdots,n\}$, and $m \in \{0,\cdots,p_\omega\}$ using the training dataset as

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \{RSS\}$$

$$= \arg\min_{\mathbf{a}} \left\{ \sum_{t=n_a+1}^{N} \left(y_t - \sum_{i\in\{1,\cdots,n_a\}} \left(\sum_{\ell\in\{0,\cdots,n\}} \sum_{m\in\{0,\cdots,p_\omega\}} a_{i,\ell,m} B_{k+\ell,r}^{per}(\theta_t)\omega_t^m + \sum_{m\in\{0,\cdots,p_\omega\}} b_{i,m}\omega_t^m \right) y_{t-i} \right)^2 \right\} \tag{4.14}$$

where the squared summation part is the l_2 norm of the predication error, denoted as the residual sum of squares (*RSS*).

We adopt the least absolute shrinkage and selection operator (LASSO) estimator [108], [137] for estimating a sparse VFP-AR model.

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \left\{ RSS + \lambda \sum_{i\in\{1,\cdots,n_a\}} \left(\sum_{\ell\in\{0,\cdots,n\}} \sum_{m\in\{0,\cdots,p_\omega\}} \left|a_{i,\ell,m}\right| + \sum_{m\in\{0,\cdots,p_\omega\}} \left|b_{i,m}\right| \right) \right\} \tag{4.15}$$

where $\lambda \geq 0$ is a tuning parameter. The sum of the absolute value of the coefficients is the penalty term. Such an optimization problem can be solved via the alternating direction method of multipliers algorithm [139]. The λ controls the amount of shrinkage that is applied to the estimates. The selection of λ is critical. When $\lambda = 0$, the LASSO will reduce to the OLS estimator. Too large a λ value will force too many coefficients in [$a_{i,\ell,m}$, $b_{i,m}$] to zero, whereas too small a λ value will force a limited number of coefficients to zero. Considering a reasonable amount of computational cost, we use the validation set approach [100] to select the λ. Given a candidate set, the λ is selected as the one which gives the minimal validation mean squared error (VAMSE) of the residual.

Hyperparameters of the VFP-AR model include n_a, K, k, n, r, and p_ω. We empirically determine the $p_\omega = 5$, $r = 3$, and K as the number of target gear tooth plus r. The k and n are determined by estimating a sparse VFP-AR model with a given n_a and the periodic B-splines, and examining these maximal occurred periodic B-spline bases. For the determination of n_a, the validation set approach [100] is again used and the n_a determined as the minimal one in a candidate set which does not significantly reduce the VAMSE.

The detailed procedure for model identification is summarized in Algorithm 4.1.

Algorithm 4.1. VFP-AR model identification

1:	**Initialize p_ω=5, r = 3, and K = the number of target gear teeth plus r**
2:	**Determination of k and n with using a given (small) n_a and the periodic B-splines**
2.1:	For a candidate set of λ
2.2:	⋮ Adopt LASSO to return a sparse VFP-AR
2.3:	⋮ Re-estimate the sparse model parameters via OLS
2.4:	⋮ Apply the estimated sparse model on the validation signal and obtain a VAMSE
2.5:	End For
2.6:	Find the λ which gives the minimal VAMSE and store the corresponding model
2.7:	Determine the k and n based on these maximal occurred periodic B-spline bases.
3:	**Determination of n_a**
3.1:	For a candidate set of n_a
3.2:	⋮ Repeat the loop given by steps 2.1~2.6.
3.3:	End For
3.4:	Determine the n_a as the minimal one that does not significantly reduce VAMSE
4:	**Output the final sparse VFP-AR model**

4.2.4 Model accuracy measures

The modeling accuracy is measured using additional validation datasets from a gearbox under each known health state. These validation datasets are non-identical to the training and validation datasets used in Section 4.2.3. Each validation dataset includes impulsive vibration signal y_t, as well as its corresponding rotating phase θ_t and speed ω_t for $t = 1$,

2, …, N. Upon the identification of sparse VFP-AR model, the inverse filter is constructed based on Eq. (4.3) and then applied to process these validation datasets. We call the residual obtained from processing the validation signal as “residual-of- validation”.

The modeling accuracy of the VFP-AR model is measured by examining three criteria [56], [100], [108]: the mean squared error (MSE) of the residual-of-validation, the randomness of the residual-of-validation via Ljung–Box test, and comparison between the frozen time spectrum of VFP-AR models and the non-parametric spectrum.

4.2.5 Scheme of the VFP-AR model-based method

The VFP-AR model-based fault detection and severity assessment method has two phases, namely the training phase and testing phase, as shown in Fig. 4.2. The training phase needs to be completed offline, whereas the testing phase can be implemented in online gearbox health monitoring. During the training phase, the training and validation signals collected under each severity level of gear tooth crack and a wide range of the speed variation are used. The vibration signals will be preprocessed via the baseline removal technique [108] to obtain the impulsive vibration signal y_t with enhanced crack-induced signal (impulses) to noise ratio. Using these impulsive vibration signals y_t, along with rotating phase θ_t, and speed ω_t, VFP-AR models are trained following the model identification algorithm in Section 4.2.3.

During the testing phase, these trained VFP-AR models will be used for fault detection and severity assessment. In the testing phase, impulsive vibration signal y_t, along with rotating phase θ_t, and speed ω_t are obtained under unknown health state of the gearbox and random speed profiles. Inverse filters from each of the VFP-AR models are applied to process the testing signal and to obtain residuals. The final health state is classified as the state with an inverse filter that gives the minimal MSE.

Note that for a new fault severity level that is different from those that exist in training datasets, during the testing phase it will be classified to be the closest state which is in the training data.

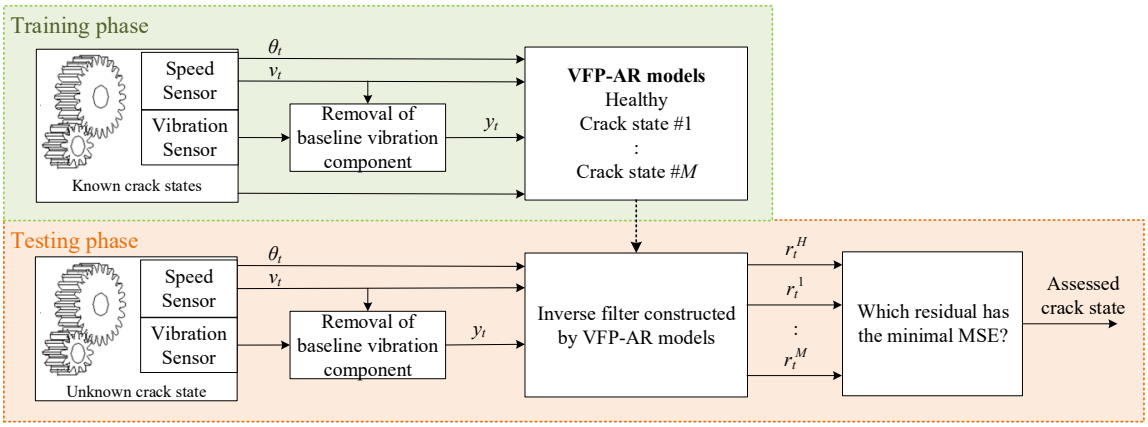

Fig. 4.2: Schematic of the VFP-AR model-based fault detection and severity assessment method

4.3 Experimental study

An experimental study with comparative analysis is conducted to evaluate the proposed VFP-AR model-based fault detection and severity assessment method.

4.3.1 Experimental signals

This experimental dataset is provided by the University of Pretoria, South Africa and details about the experiment and the data collected are given in Section 3.3.3. In this chapter, the signal from the *x*-direction of the tri-axle accelerometer (as shown in Fig. 3.21; with a sensitivity of 100 mV/g) will be used, as well as the signal from the zebra-tape shaft encoder.

In this chapter, we consider 4 health states, namely, healthy (H), initial crack (F1), intermediate crack (F2), and missing tooth (F3). The F1 state corresponds to the gear with the initial 50% tooth crack. The F2 state corresponds to the gear that had run 17 days after

the initial 50% tooth crack. The F3 state corresponds to the gear that ran right before the end of the run-to-failure experiment.

Under each health state, training, validation, and testing signals were prepared and pre-processed. In total, 43 data files under each health state were used in which one served for training (training signal), two for validation (one used for model identification and the other for measuring modeling accuracy), and the rest 40 for testing fault detection and severity assessment results. The 43 data files under each health state should be similar to one another. Therefore, the training and validation data files were arbitrarily selected from the 43 data files under each health state. The preprocessing included low pass filtering, downsampling, and baseline components removal [108]. Firstly, these vibration signals were low passed using an FIR filter with a cutoff frequency f_c = 1.6 kHz. Then, these vibration signals were downsampled from f_s = 25.6 kHz to f_s =3.2 kHz. The chosen f_s fulfilled the principle that f_s needs to be higher than 20times of the gear meshing frequency to cover the 10th order harmonic of the gear meshing frequency. Lastly, using an additional healthy data file, we identified a sparse FP-AR model with S_{AR}^{ini} = {1, 2, ..., 600} and $S_{F,i}^{ini}$ = {0, 1, ..., 7} for the baseline components removal [108]. The raw vibration signals from each health condition were filtered with the inverse filter constructed by the sparse FP-AR model to obtain impulsive vibration signals.

For the training and validation signals, we used the first half of the vibration signal in a data file only. Such vibration signal lasted 10 s and experienced a full cycle of the speed variation. Fig. 4.3 shows the impulsive vibration signals under each health state, along with their rotating speed and phase. These signals were used for model validation. The training and other validation signals were similar to these signals shown in Fig. 4.3 and hence are not plotted. We can observe an increased number of spikes in the time waveform

of impulsive vibration signals from health state H, to F1, F2, and F3. The amplitude of impulsive vibration signals is also modulated by the rotating speed.

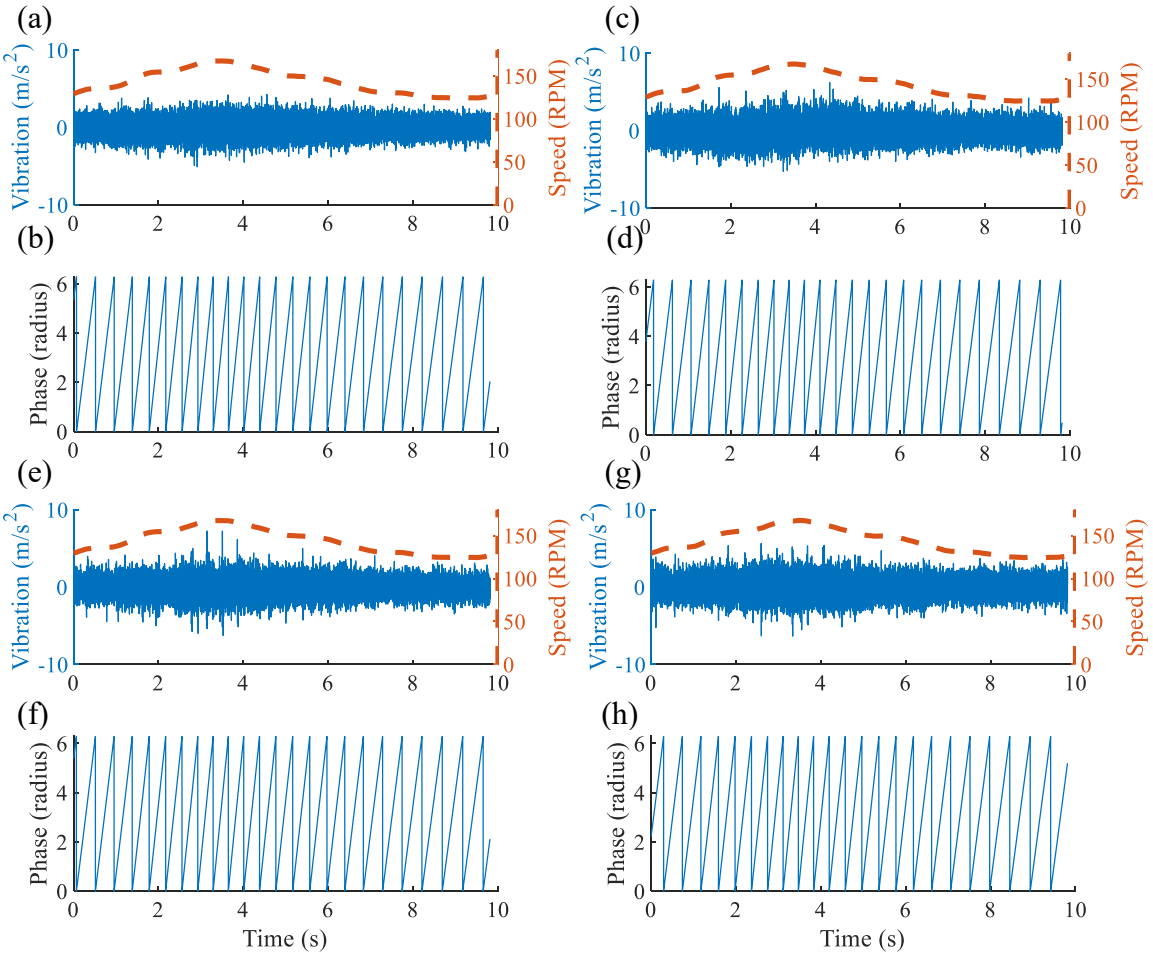

Fig. 4.3: Impulsive vibration signals along with their rotating speed and phase: (a, b) H; (c, d) F1; (e, f) F2; (g, h) F3.

The testing signals were further segmented, such that each testing signal has a different speed profile. Specifically, the testing signal only lasted 5 s and has a starting point p_s randomly sampled from [0.25, 0.5, ..., 10] s. Each segment of the testing signal lasted at least 10 cycles of the object gear rotation. In total, we have 40 signals under each health state for testing the fault detection and severity assessment. Fig. 4.4 shows an example of segmentation using the starting point equals 7.5 s. Fig. 4.5 shows a part (5 segments under

each health state; starting point p_s = [0.25, 2, 3.75, 5, 7.25] s) of resulted testing signals for testing the fault detection and severity assessment.

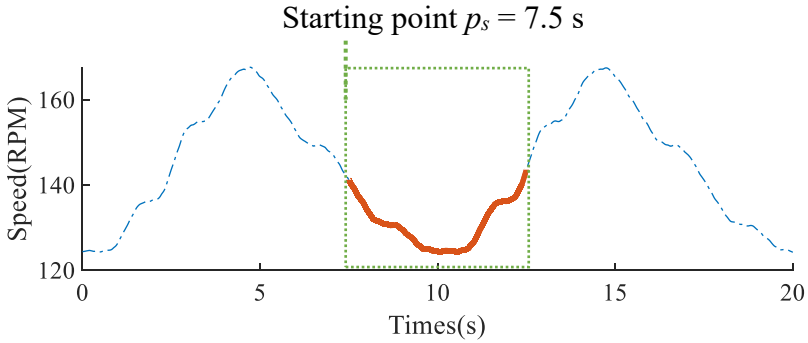

Fig. 4.4: An example of segmentation of signal for testing the fault detection and severity assessment. Dashed rectangle denotes a data segmentation window.

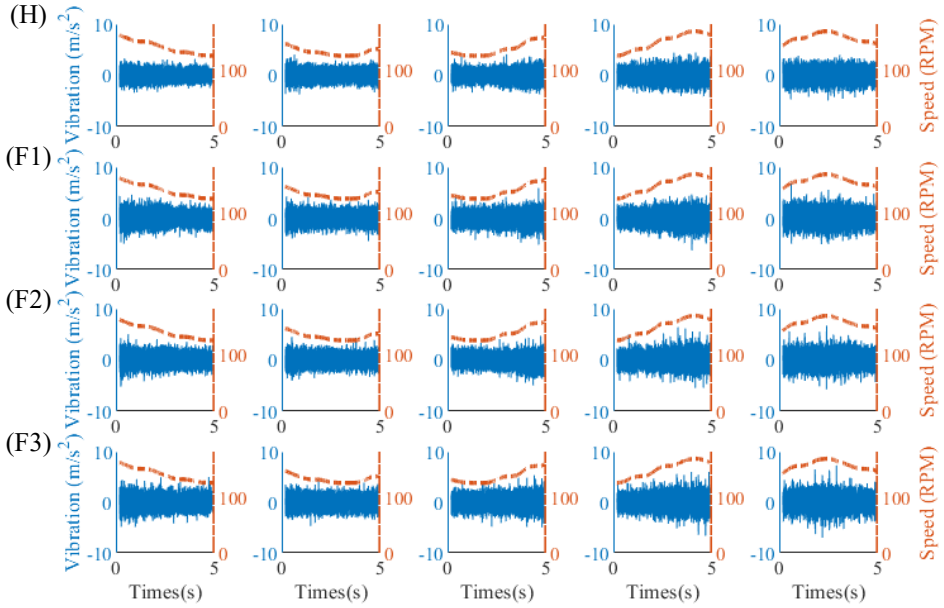

Fig. 4.5: Signals for testing the fault detection and severity assessment. Each row belongs to a health state and different columns are different realizations of the random speed profile.

4.3.2 VFP-AR model-based method

In this section, we evaluate the presented VFP-AR model-based method for fault detection and severity assessment. To prove the superiority of the refined B-splines over periodic B-splines, we evaluate the performance of using the periodic B-splines as well. Hereafter,

VFP-AR-R stands for the VFP-AR model using the proposed refined B-splines and VFP-AR-P for the VFP-AR model using the periodic B-splines.

The k and n need to be determined during the identification of VFP-AR-R models. We first estimated a VFP-AR-P model with a small n_a = 5 and K = 40 (i.e., the number of teeth 37 plus r = 3). The k and n were determined by examining the occurrence of periodic B-spline bases. To reduce the computational cost, the candidate set for λ was firstly coarsely configured as [0, 1×10^{-9}, 1×10^{-8}, ..., 1×10^{-1}], and then a refined candidate set for λ was created based on the λ in the coarse set which gives a minimal validation MSE. For instance, if the $\lambda = 1\times10^{-6}$ in the coarse set gives a minimal VAMSE, the refined candidate set becomes [2×10^{-7}, 3×10^{-7}, ..., 1×10^{-6}, 2×10^{-6}, 3×10^{-6}, ..., 9×10^{-6}]. The final λ was chosen as the one in the refined set which gives a minimal validation MSE.

Fig. 4.6 shows the occurrence of periodic B-spline bases in the VFP-AR-P model (n_a = 5, K = 40) for health states F1, F2, and F3. Since the vibration signal under the H state does not have crack induced impulses, its corresponding VFP-AR model does not need to consider phase. In other words, the VFP-AR model for H state reduced to an FP-AR model. From Fig. 4.6, the k and n are determined as (24, 1) for F1 state; (24, 2) for F2 state; (24, 3) for F3 state. These values were determined based on a threshold T_h specified as follows

$$T_h = \mu_m + 3\sigma \tag{4.16}$$

where μ_m is the median of occurrence of periodic B-spline bases; and σ is the standard deviation of occurrence. These bases with their occurrence higher than the threshold T_h were selected. For instance, bases #24 and #25 were selected for F1 case. Therefore, the corresponding k and n were determined as (24, 1). It is interesting to note that the increase of the fault severity leads to an increase of the n. This phenomenon indicates that a more severe crack fault will tend to induce an impulse with a longer duration.

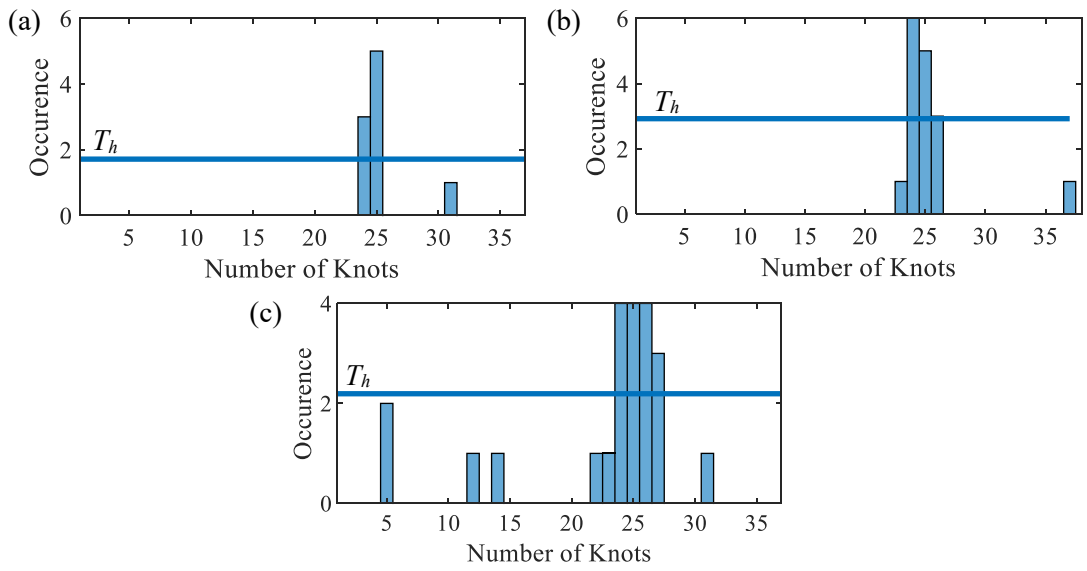

Fig. 4.6: Determination of k and n for (a) F1; (b) F2; (c) F3.

The n_a is to be determined after obtaining the k and n. We configured a candidate set {5, 10, ..., 140} for n_a. Fig. 4.7 shows the VAMSE of VFP-AR-R models for each health state. The n_a is empirically determined as the minimal one in the candidate set which does not significantly reduce the VAMSE. This is based on the consideration of reducing model complicity whilst only sacrificing minor and acceptable modeling accuracy. To this end, the n_a is determined (50, 40, 35, 35) for health states (H, F1, F2, F3), respectively. For instance, the reduction of VAMSE in H health state becomes insignificant when n_a gets higher than 50. Therefore, the n_a is determined as 50 for health state H.

For comparison, we identify VFP-AR-P models as well. The previously identified FP-AR model for H state was kept. The candidate set for λ was configured as [1×10^{-3}, 2×10^{-3}, ..., 3×10^{-2}, 3.2×10^{-2}, 3.4×10^{-2},..., 6×10^{-2}], and the final λ was chosen as the one in the candidate set which gives a minimal validation MSE. We determined the n_a from a candidate set of {5, 10, ..., 140}. Fig. 4.8 shows the VAMSE for each health state. The n_a is determined as (35, 35, 35) for health states (F1, F2, F3), respectively.

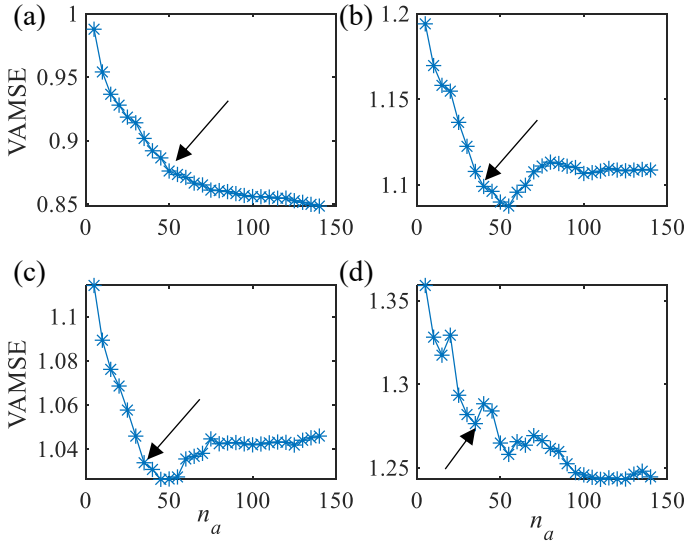

Fig. 4.7: VAMSE for the determination of n_a of VFP-AR-R models: (a)H; (b)F1; (c)F2; (d)F3.

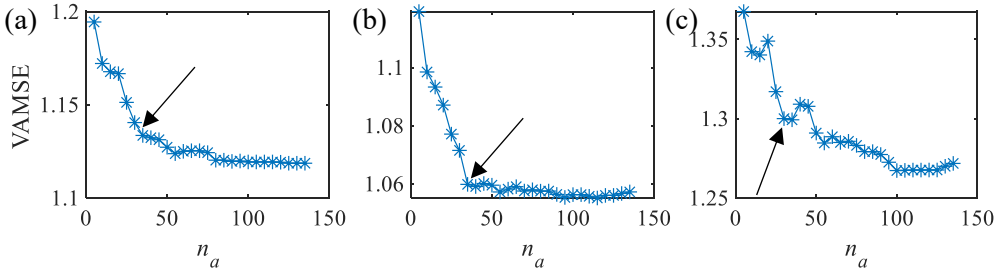

Fig. 4.8: VAMSE for the determination of n_a of VFP-AR-P models: (a)F1; (b)F2; (c)F3.

Upon the identification of both VFP-AR-R and VFP-AR-P models, the inverse filter was constructed based on Eq. (4.3) and then applied to process the additional validation signals for measuring the modeling accuracy. Table 4.1 lists the MSE and randomness of the residuals. The *p*-values from Ljung–Box tests were reported, which means the probability of being random. It is evident that VFP-AR-R models return a residual with smaller MSE and more random than VFP-AR-P models. For the F2 state, the residuals from both two models seem not random, which may due to the variation between the training and validation signals and testing signal.

Computational costs were evaluated and listed in Table 4.1. The VFP-AR-R model spent less than 6.3 h in training. The training data points are 64000. For four VFP-AR-R models,

the maximum time required for training will be less than 25.2 h. The VFP-AR-P models spent around 9.1h for each model and less than 36 h for four models. Since this training process is completed offline, the length of time required is not very critical [108]. Besides, the computational efficiency can be improved by implementing the algorithm in a more powerful industrial computer [108]. The computational time in the testing stage primarily determines whether the method can be implemented in online condition monitoring systems [108]. The inverse filter constructed by the VFP-AR-R model spent less than 2.6 s for processing the testing signal that has 64000 data points. For four VFP-AR-R models, the maximum time required for testing will be 10.4 s. Therefore, the VFP-AR-R model-based method has immense potential for the implementation of online condition monitoring systems. Note that the inverse filter constructed by the VFP-AR-P models spent almost 3 times of computational cost than the VFP-AR-R models.

Table 4.1: Modeling accuracy and computational cost of VFP-AR models. Algorithms were coded in MATLAB 2019a on a desktop with two Intel 2.4GHz processors and 16GB of RAM.

Health state	**H**	**F1**		**F2**		**F3**	
B-splines	**-**	**periodic**	**refined**	**periodic**	**refined**	**periodic**	**refined**
MSE $((m/s^2)^2)$	0.881	1.204	1.172	1.135	1.114	1.203	1.191
Randomness test	1.55×10^{-5}	2.85×10^{-7}	1.56×10^{-5}	1.11×10^{-16}	0	1.65×10^{-4}	6.93×10^{-3}
Time 1* (h)	1.5	8.9	4.9	7.1	5.2	9.1	6.3
Time 2# (s)	2.2	6.2	2.4	6.2	2.4	6.4	2.6

* CPU time in training; # CPU time in testing (i.e., inverse filtration)

Fig. 4.9 shows the non-parametric spectrum (a, b, c, d) of the validation signals, as well as the frozen-time spectrums of VFP-AR-R models (e, f, g, h) and VFP-AR-P models (i, j, k). We can see tooth crack induced impulses as vertical lines in these spectrums. The vertical lines in the frozen-time spectrums of VFP-AR-R models behave discretely, which are in good agreement with the discrete lines in non-parametric spectrums. On the other hand,

the vertical lines are continuous in the frozen-time spectrums of VFP-AR-P models. It is evident that we can better represent the tooth crack induced impulses using VFP-AR-R models. Apart from tooth crack induced impulses, the frozen-time spectrums of VFP-AR-R models are more sensitive to speed variations, which are in better agreement with the non-parametric spectrums than the VFP-AR-P models.

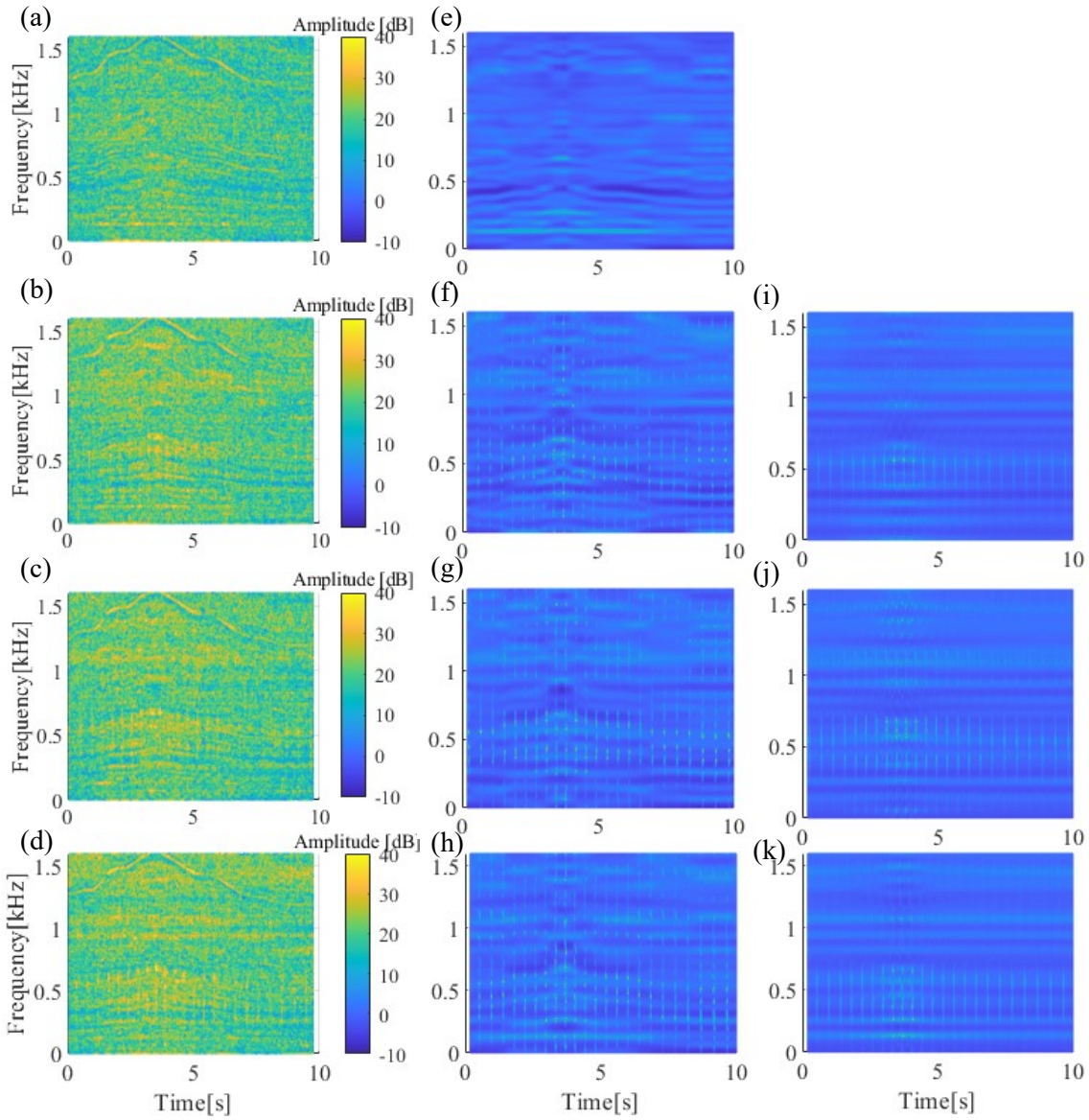

Fig. 4.9: Time-frequency spectrums: (a~d) STFT spectrum of the testing signal obtained by MATLAB *spectrogram* function; (e~h) Frozen-time spectrum of VFP-AR-R models; (i~k) Frozen-time spectrum of VFP-AR-P models. Note that the VFP-AR model for H state does not need to consider phase, and hence the same if using periodic or refined B-splines. z-axis scales are the same for all spectrums

VFP-AR models under known health state were applied to these testing signals for the fault detection and severity assessment. For each testing signal, 4 model residuals were obtained, and their MSE calculated. The health state is classified as the state of the residual which gives minimal MSE. Fig. 4.10 shows the MSE of four residuals obtained by processing the testing signals shown in Fig. 4.5. The proposed refined B-splines were used. We can see two misclassifications. For comparison, Fig. 4.11 shows results using the VFP-AR-P models. We can see three misclassifications.

Fig. 4.12 shows the classification results when processing the 40 testing signals under each health state. Using the VFP-AR-R models, we can classify 93.8% correctly. Using the VFP-AR-P models, we can classify 90.0% correctly. The results not only show the effectiveness of the VFP-AR model-based method but also prove the superiority of the proposed refined B-splines over the periodic B-splines.

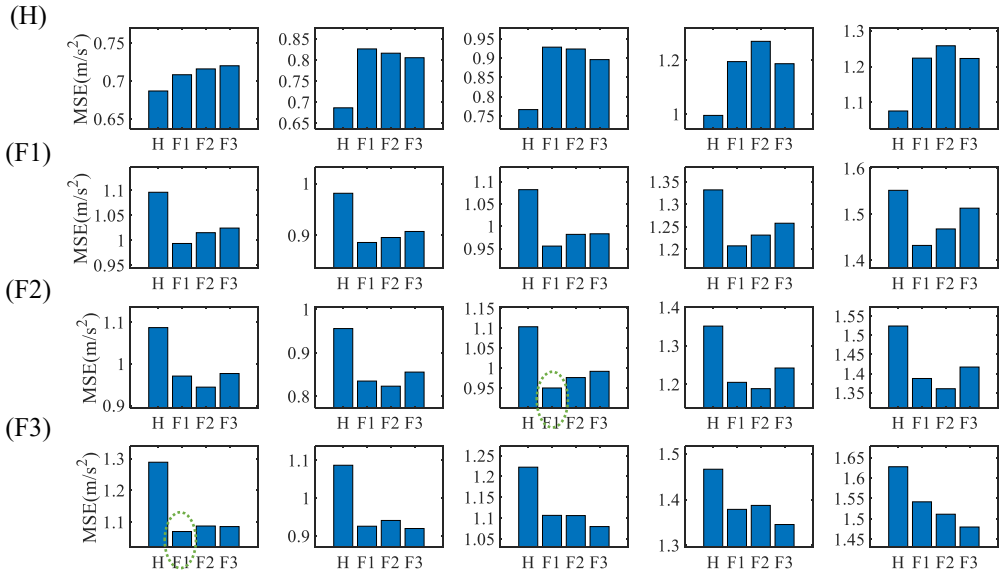

Fig. 4.10: MSE of residuals from four VFP-AR-R models. Each row belongs to a true health state given on the left-hand side and different columns are different realizations of the random speed profile. The dashed circle denotes misclassification.

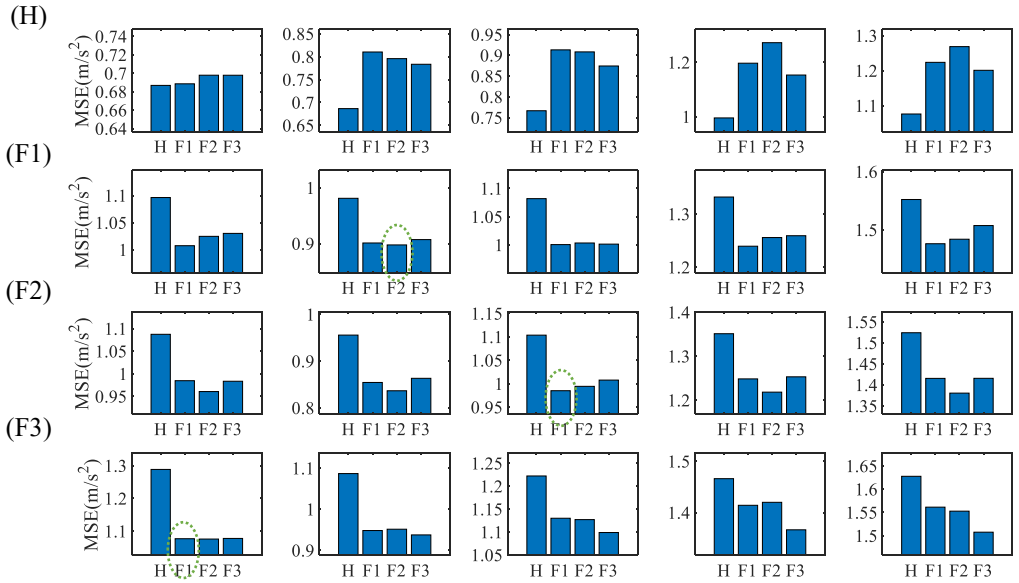

Fig. 4.11: MSE of residuals from four VFP-AR-P models. Each row belongs to a true health state given on the left-hand side and different columns are different realizations of the random speed profile. The dashed circle denotes misclassification.

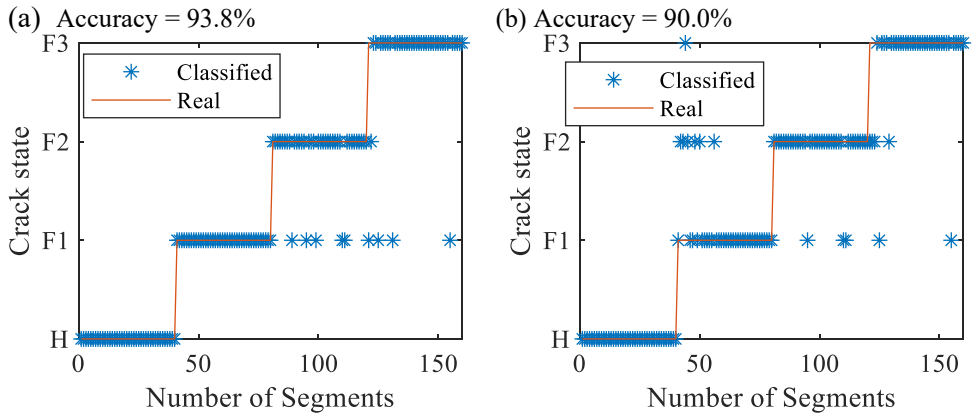

Fig. 4.12: Classification result: (a) used the proposed refined B-splines; (b) used the periodic B-splines.

4.3.3 Comparison with the novelty diagnosis method[1]

The novelty diagnostic method presented by Schmidt *et al.* [76] was applied for comparison and will be briefly summarized in this section. The novelty diagnostic method first identifies the OC state of the gearbox, with an assumption that speed does not vary

[1] The NLL feature extraction work presented in this subsection was completed by Dr. Stephan Schmidt and Dr. Stephan Heyns at the University of Pretoria (UofP), South Africa.

significantly within a rotating cycle. The average rotational speed in one gear rotation is calculated as one of the OC features. The spectrogram of each vibration signal in the healthy dataset is calculated, and then the data around the gear mesh frequency and its five harmonics are extracted as additional OC features. The window length of the spectrogram is set to one gear rotation. Afterwards, these OC features are scaled, and their principal components extracted as the final OC features. Subsequently, the OC features are modelled with a Gaussian Mixture Model (GMM). This is slightly different from Ref. [76], where a Hidden Markov Model (HMM) was used. Secondly, a GMM model is used for modelling the healthy response of the system in specific operating condition states, whereafter a discrepancy measure, namely the NLL, is calculated and used for fault detection. The GMM is better suited than HMM for discrepancy analysis under time-varying operating conditions [154], [155], which was the motivation for using it in this work. The NLL is further time synchronous averaged, and the variance of the averaged NLL is extracted to serve as a health condition indicator for reflecting the propagation of the localized gear fault.

The novelty diagnostic method needs to train a healthy model on the machine condition features associated with each OC state. For training purposes, 60 healthy data files were used, which are non-identical to the testing signals #2 for examining the fault detection and severity assessment.

Using the trained healthy model, the variance of the time synchronous averaged NLL was calculated for the same testing signals as used in Section 4.3.2. Fig. 4.13 shows the variance of the time sync averaged NLL. We can see a clear increasing trend when the tooth crack becomes severe. Overlaps are also clear, especially between H, F1, and F2.

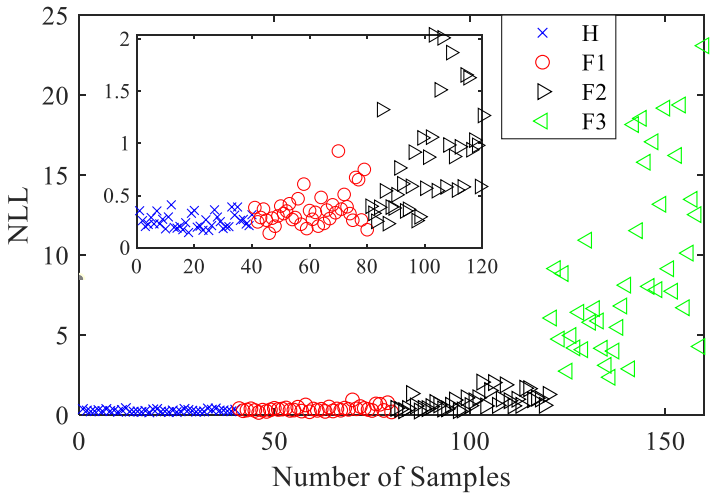

Fig. 4.13: Variance of the time synchronous averaged NLL

To quantify the classification accuracy, we trained an SVM using MATLAB function *fitcecoc* [156]. The hyperparameters of the SVM included the types of kernel functions (included Gaussian kernel, polynomial, and linear), box constraint value (also referred as the cost parameter), kernel-scale value, with or without training data standardization, and one-verse-one or one-verse-all algorithms [100]. These five hyperparameters were optimized by minimizing 10-fold cross-validation error [100]. The folds on the training data were randomly partitioned. Fig. 4.14 shows the training accuracy when all the 160 health indicators shown in Fig. 4.13 were used. The training accuracy equals to 73.8%, not to mention the testing accuracy. Therefore, it is clear that the novelty diagnostic method performs comparably worse than the proposed time series model-based method. This may due to the assumption that the rotating speed does not vary significantly within a rotating cycle. It is also worth to note that the novelty diagnostic method utilizes only the baseline data for training, whereas the proposed time series model-based method needs data from multiple health states including the baseline data for training. In cases when only baseline data is available, the novelty diagnostic method is expected to perform better.

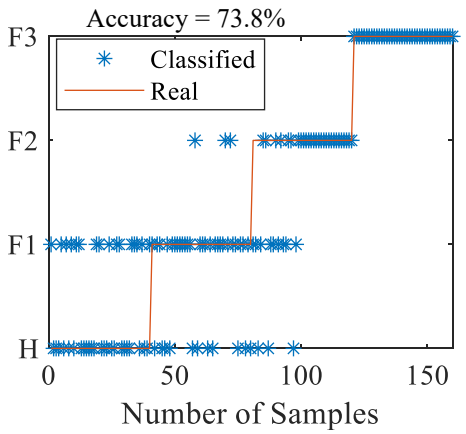

Fig. 4.14: Classification result of the novelty diagnosis method.

4.3.4 Comparison with the SVM-based method

The SVM-based method presented by Santos *et al.* [75] was applied for comparison. Fig. 4.15 shows the scheme of the SVM-based method. The method includes feature extraction, feature selection, and SVM classifier. For gear tooth crack detection and severity assessment, we extracted proper features that are sensitive to gear tooth crack. Specifically, we considered 11 features. They are the variance of the time sync average NLL(as described in Section 4.3.3), the mean rotating speed within a testing signal, root mean square (RMS) [87], periodic modulation intensity [57], crest factor [157], signal-to-noise ratio [28], kurtosis, modified Rényi entropy [36], RMS ratio between residual signals and regular vibrations [36], Normality hypothesis test statistics [86], and randomness test statistics [45]. We adopted the forward regression scheme [100] for feature selection. The SVM model was trained using the MATLAB function *fitcecoc* and its hyperparameters were optimized by minimizing 10-fold cross-validation error.

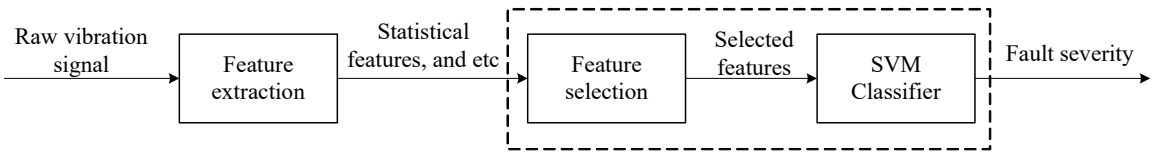

Fig. 4.15: Scheme of the SVM-based method

The SVM-based method was applied to the same testing signals as used in Section 4.3.2. These testing signals were further randomly partitioned into three subsets, training, validation, and testing. Features were extracted from these signals. Forward feature selection was achieved by minimizing the validation error. Fig. 4.16 shows the classification accuracy applied to randomly partitioned testing subset. We can see the most accurate case is 87.7%. From the results, the SVM-based method performs comparably worse than the time series model-based method. The inferior performance may because the features are still sensitive to speed/load changes, which disables the SVM classifier. Another advantage of the time-series model-based method is the requirement of only a few signals for model training. Comparably, the SVM-based method performs even worse than 87.7% if only a small labelled data size is available for training, e.g., the 10-45-45 partition case.

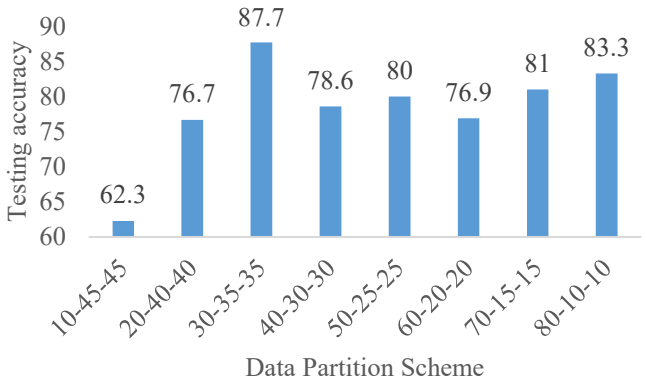

Fig. 4.16: Classification result of the SVM-based method using different data partition scheme. For instance, 20-40-40 denotes 20% for training, 40% for validation, and 40% for testing.

4.4 Discussions

In this chapter, we have assumed that the tooth crack occurs in the same tooth in a gear. In industrial applications, the crack fault may occur in other teeth. To address this issue, we need to identify the location of crack before conducting the fault detection and severity assessment. The crack identification can be done by simply estimating a shifting variable

of the rotating phase which representing the angular distance between the crack in the newly commissioned gearbox and the training gearbox. Estimating the shifting variable is a one-dimensional minimization problem where the objective function is VAMSE of the residuals from VFP-AR models.

The misclassification is conservative when using the time series model-based method. In other words, miss-classification occurs only from severe crack state to less severe state, as shown in Fig. 4.12(a). In the viewpoint of maintenance and operation, we may tolerate an aggressively miss-classification rather than a conservative one. To classify the health state more aggressively, we can design a new classification rule rather than using the minimal MSE scheme. For instance, if the minimal MSE occurs in the residual from the F1 state model, but the MSE in the residual from the F2 state is not significantly greater than the minimal MSE, we can still classify the state as F2. The development of new classification rules is of future studies.

4.5 Conclusions

This chapter presented a time series model-based method for the gear tooth crack fault detection and severity assessment under random speed variation. Specifically, we have considered the rotating speed and phase as covariates in a VFP-AR model for representing impulsive vibration signals. We have proposed the refined B-splines for mapping the dependency between AR coefficients and the rotating phase. The performance of the presented method has been validated using laboratory signals. The presented method can classify 93.8% of the health state correctly, which is comparably better than the novelty diagnostic method (73.8%) and SVM-based method (87.7%).

5

Summary and future works

This chapter summarizes this thesis study and suggests some future works that are worth to be investigated.

5.1 Summary

Fault detection and severity assessment before the failure of gearbox systems have great significance in terms of preventing the sudden failure, as well as enabling condition-based maintenance and thus reducing maintenance costs or minimizing downtime. Vibration-based condition monitoring has been widely employed for gearboxes, thanks to the accessible collection of vibration signals and the effectiveness in identifying fault information from vibration signals. In this thesis, we focused on the vibration-based gear tooth crack fault detection and severity assessment. The objective of this thesis was to develop new vibration analyzing and processing methods for the gear tooth crack detection and severity assessment. With the new methods, more accurate gear tooth crack detection and severity assessment results were obtained. This thesis study considered both constant speed and variable speed (VS) conditions. The load condition of the gearbox was assumed to be either constant or correlated with the rotating speed.

We first proposed an improved singular value decomposition (SVD)-based method for the gear tooth crack detection and severity assessment under constant speed operation. The

SVD-based method has the advantages of being non-parametric, computationally efficient, and being free from phase shift and waveform distortion. However, most reported SVD-based methods are essentially energy-based, which may ignore weak-energy features caused by early faults. A recently reported reweighted SVD (RSVD) method used a criterion to measure how influential the periodic impulses are in each signal component (SC). This RSVD method addressed the problem of ignoring weak-energy features caused by early faults. But, the RSVD but still suffers from two issues: 1) it does not consider interference from non-fault related vibration components on the periodic modulation intensity (PMI), which leads to a high miss and/or false alarm rate; and 2) it directly conduct reweighted summation of all SCs with PMI values that exceed the threshold for signal reconstruction, which undermines periodic impulse extraction. In our improved SVD-based method, we addressed these issues that RSVD face. Specifically, the improved SVD-based method incorporates an auto-regression model-based baseline removal approach. The SVD is applied to decompose the residual signal rather than the raw signal. Interference from non-fault related vibration components on the PMI is therefore eliminated. Second, the improved SVD-based method selects the SCs such that the PMI of the reconstructed signal is maximized. Periodic impulse extraction is consequently strengthened. An experimental study was conducted. Results showed that the improved method outperforms RSVD in terms of fault detection and severity assessment without creating a considerable computational burden.

We then investigated the time series model-based methods (TSMBMs) for the gear tooth crack detection and severity assessment under the VS condition. In TSMBMs, the identification of a baseline model to represent the vibration from a healthy state is a must. The accuracy of baseline modeling directly determines the success of TSMBMs. To

accurately model nonstationary signals, time-variant time series models are in demand. Conventional functional pooled autoregression (FP-AR) model is a good option. However, conventional FP-AR assumed 1) consecutive autoregression (AR) terms and 2) identical functional space that describes the dependency between AR parameters and rotating speed, which limited its modeling accuracy. To improve the modeling accuracy, we proposed a sparse FP-AR model for the representation of baseline vibration signals under VS condition. The sparse FP-AR model uses a sparse AR set and non-identical functional spaces. A new model selection procedure by adopting the least absolute shrinkage and selection operator has been developed for obtaining such a sparse FP-AR model. The sparse FP-AR model has been validated using simulation signals from a simulation model for a fixed-axis gearbox and experimental signals from two independent fixed-axis gearbox test-rigs. Results show that although the proposed sparse FP-AR model consumes more computational time than the conventional FP-AR model, it has higher modeling accuracy. With the sparse FP-AR model, the *model residual-based method* can be employed for the fault detection and severity assessment when the testing signals were collected under deterministic speed profiles. The improved modeling accuracy benefits the health condition assessment of a gearbox under VS in detecting earlier faults and better assessing the progression of faults.

With the sparse FP-AR model, the *model residual-based method* can be employed for the fault detection and severity assessment when the testing signals were collected under deterministic speed profiles. However, the *model residual-based method* is inapplicable to the cases when the testing signals were collected under random speed variations which typically occurs in wind turbine gearboxes. Under random speed variations, the speed profile within each data acquisition window could be hugely different. They could stay in

a high/low level or vary rapidly. A condition monitoring system is expected to detect and diagnose faults using a segment of vibration signal collected under each of these different profiles. These various statistics, such as examining the whiteness, variance, or impulsiveness of the residual of the baseline model, calculated from the residual of the baseline model are still affected by the random speed profiles. Therefore, *model residual-based methods* may be ineffective when the testing signals were collected under random speed variations.

We presented a time series model-based method for the gear tooth crack fault detection and severity assessment under random speed variation. Specifically, we identified time series models on the residuals of the baseline sparse FP-AR model under random speed variation. We denote the residuals of the baseline sparse FP-AR model as the impulsive vibration signal which has enhanced crack-induced signal (impulses) to noise ratio. We considered the rotating speed and phase as covariates in a vector functional pooled autoregression (VFP-AR) model for representing impulsive vibration signals, given that the rotating speed affects the amplitude of tooth crack induced impulses, and the rotating phase determines the location of these impulses. We proposed the refined B-splines for mapping the dependency between AR coefficients and the rotating phase. The refined B-splines are deduced from the periodic B-splines by removing unnecessary bases outside of the location of impulses. The refined B-splines are less complex than the periodic B-splines and hence can avoid overfitting. Multiple VFP-AR models will be built for each level of tooth crack severity. Fault detection and severity assessment can, therefore, be based on examining the residuals of each model. The final health state can be assessed as the severity state of the model which gives the minimal residual energy. The performance of the presented method has been validated using laboratory signals. The presented

method can classify 93.8% of the health state correctly, which is comparably better than the novelty diagnostic method (73.8%) and the SVM-based method (87.7%).

5.2 Future works

Based on the scope of this thesis, we suggest the following three topics for future study.

(1) Further advancing SVD-based methods for more accurate fault detection and severity assessment

As mentioned in Section 2.5, two remaining questions exist in the improved SVD-based method, which are 1) why the number of SCs which have higher PMIs than the threshold T_h does not monotonically increase when the fault severity increases? and 2) why conducting the reweighted summation undermines the PMI of F3 the most? In the future, we will investigate these two questions and find the reasons behind. Meanwhile, the improved SVD-based method classifies 4 fault severity levels with a relatively low accuracy (i.e., 69.6%) which is not good enough for industrial applications. Another future work of this thesis is to improve severity classification accuracy. A feasible technical route is to extract more fault sensitive features (e.g., kurtosis and root mean squared) from the reconstructed signals, select more effective features, and use an advanced classifier (e.g., support vector machine).

(2) Multivariate baseline vibration modeling

In this thesis, we analyzed single-channel vibration signals for the gear tooth crack fault detection and severity assessment. A more general case is analyzing multichannel vibration signals which were collected from different sensor locations and/or directions [158]–[160]. Considering multichannel vibration signals offer two advantages: (1) Multichannel vibration signals provide more information from different locations and/or

directions and hence have much more chances of capturing fault induced features in comparison with the single-channel vibration signal; (2) Multichannel vibration signals are usually collected simultaneously, which may reveal the correlation of the random noise between channels and, therefore, the removal of the correlated random noise become possible.

A future work of this thesis is to develop a multivariate time series model for representing multichannel nonstationary baseline vibration signals from a gearbox. The potential multivariate time series model would combine the advantages of sparse time series modeling [108] and the utilization of multichannel vibration signals. The former has shown to have higher modeling accuracy than conventional non-sparse time series modeling, whereas the latter enables the removal of the correlated random noise between channels. The hypothesis is that a multivariate time series model-based fault detection scheme will have a higher fault detection rate than using the sparse single variate FP-AR model [108] and conventional (non-sparse) multivariate functional pooled vector autoregressive (FP-VAR) model [160].

(3) Baseline vibration modeling considering rotational speed and phase measurement uncertainty

In this thesis, we have assumed an accurate measurement of the rotational speed and phase information. However, the measurement uncertainties of rotating speed and phase are inevitable. When using tachometers, factors like torsional vibration [161] and quantification error will affect the measurement of rotating speed and phase. When using the speed estimation method, estimation uncertainties are usually involved and maybe even higher than these uncertainties of tachometer measurement [162]. Merely ignoring the measurement uncertainty would result in a biased time series model [163].

A future work of this thesis is to consider the measurement uncertainties of rotating speed and phase when building either FP-AR models or VFP-AR models. The hypothesis is that the model with measurement uncertainties considered will have higher modeling accuracy, and thus benefit the fault detection and severity assessment.

(4) Fault detection and severity assessment under variable load condition

In this thesis, we have assumed that the load condition of the gearbox is either constant or correlated with the rotating speed. In industry, a gearbox may run under variable load and constant rotating speed conditions. For example, the gearbox in railway traction may experience variable load due to the variable wind resistance. Meanwhile, the operation speed of the train is controlled to be constant.

Different from the variable speed condition, the variable load torque modulates the amplitude of vibration signals only. Given this fact, we can design a time series model with its phase of zeros (i.e., zeros of the transfer function) constrained, but the amplitude of zeros dependent on variable load. Such a time series model is less complex than conventional FP-AR models, and still captures the characteristics of vibration signals under variable load torque conditions.

(5) Fault diagnosis of gearboxes

In this thesis, we have focused on the detection and severity assessment of only the gear tooth crack fault which initiates at the root of the tooth and propagates towards the removal of the tooth. One of the future works is to consider various fault modes in a gearbox, and to develop methods for the gearbox fault detection, severity assessment, and fault mode identification.

In particular, other types of localized faults in gearbox systems, such as gear tooth crack which initiates at different locations, tooth tip chipping and bearing ball crack, also generate periodic impulses in the vibration signal. The methods presented in this thesis can extract the periodic impulses, and therefore are theoretically applicable for the detection and severity assessment of other types of localized faults. However, the fault mode needs to be identified before assessing the fault severity. Fault mode can be identified based on further distinguishing the patterns of the impulses, such as its duration, amplitude, and carrier frequency. Once the fault been detected and its mode identified, fault severity can be assessed based on the energy of impulse.

References

[1] 'Gear units for rail vehicles', *voith.com*. https://d2euiryrvxi8z1.cloudfront.net/asset/445934742530/0803620164a2a492be6b4140bde3bb7d (accessed Apr. 08, 2020).

[2] 'The Inside of a Wind Turbine', *Energy.gov*. https://www.energy.gov/eere/wind/inside-wind-turbine (accessed Apr. 08, 2020).

[3] H. Henao, S. H. Kia, and G. A. Capolino, 'Torsional-Vibration Assessment and Gear-Fault Diagnosis in Railway Traction System', *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1707–1717, May 2011, doi: 10.1109/TIE.2011.2106094.

[4] A. Romero, Y. Lage, S. Soua, B. Wang, and T.-H. Gan, 'Vestas V90-3MW Wind Turbine Gearbox Health Assessment Using a Vibration-Based Condition Monitoring System', *Shock Vib.*, vol. 2016, p. e6423587, Sep. 2016, doi: 10.1155/2016/6423587.

[5] Y. Fang, 'Dynamics of a Spur Gear Pair under Stochastic Internal and External Excitations', Doctoral Dissertation, University of Alberta (Canada), Edmonton, Alberta, Canada, 2019.

[6] P. W. Carlin, A. S. Laxson, and E. B. Muljadi, 'The History and State of the Art of Variable-Speed Wind Turbine Technology', *Wind Energy*, vol. 6, no. 2, pp. 129–159, Apr. 2003, doi: 10.1002/we.77.

[7] J. M. Ha, H. Oh, J. Park, and B. D. Youn, 'Classification of operating conditions of wind turbines for a class-wise condition monitoring strategy', *Renew. Energy*, vol. 103, pp. 594–605, Apr. 2017, doi: 10.1016/j.renene.2016.10.071.

[8] K. E. Johnson, L. Y. Pao, M. J. Balas, and L. J. Fingersh, 'Control of variable-speed wind turbines: standard and adaptive techniques for maximizing energy capture', *IEEE Control Syst.*, vol. 26, no. 3, pp. 70–81, Jun. 2006, doi: 10.1109/MCS.2006.1636311.

[9] F. Chaari, W. Baccar, M. S. Abbes, and M. Haddar, 'Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission', *Eur. J. Mech. - ASolids*, vol. 27, no. 4, pp. 691–705, Jul. 2008, doi: 10.1016/j.euromechsol.2007.11.005.

[10] R. D.- December 1 and 2008 Safety Structural Failure Denmark, 'Final report on investigation of a catastrophic turbine failures', Feb. 2008. http://www.windaction.org/posts/20929-final-report-on-investigation-of-a-catastrophic-turbine-failures-february-22-and-23-2008#.XkLjWehKg2x (accessed Feb. 11, 2020).

[11] Air Accidents Investigation Board, 'S5/2009 Eurocopter AS332L2 Super Puma, G-REDL', Jul. 2009. [Online]. Available: https://assets.publishing.service.gov.uk/media/548ac196e5274a42900002d6/S5-2009_Eurocopter_AS332L2_Super_Puma__G-REDL_08-09.pdf.

[12] X. Liang, M. J. Zuo, and Z. Feng, 'Dynamic modeling of gearbox faults: A review', *Mech. Syst. Signal Process.*, vol. 98, pp. 852–876, Jan. 2018, doi: 10.1016/j.ymssp.2017.05.024.

[13] S. Sheng, R. Errichello, and M. Mark, 'Wind Turbine Gearbox Failure Modes - A Brief', presented at the ASME/STLE 2011 International Joint Tribology Conference, Los Angeles, California, 2011.

[14] D. G. Lewicki, 'Gear Crack Propagation Path Studies-Guidelines for Ultra-Safe Design', *J. Am. Helicopter Soc.*, vol. 47, no. 1, pp. 64–72, Jan. 2002, doi: info:doi/10.4050/JAHS.47.64.

[15] H. Ma, J. Zeng, R. Feng, X. Pang, Q. Wang, and B. Wen, 'Review on dynamics of cracked gear systems', *Eng. Fail. Anal.*, vol. 55, pp. 224–245, Sep. 2015, doi: 10.1016/j.engfailanal.2015.06.004.

[16] S. Wu, M. J. Zuo, and A. Parey, 'Simulation of spur gear dynamics and estimation of fault growth', *J. Sound Vib.*, vol. 317, no. 3, pp. 608–624, Nov. 2008, doi: 10.1016/j.jsv.2008.03.038.

[17] Φ. Κοψαυτόπουλος, 'Advanced functional and sequential statistical time series methods for damage diagnosis in mechanical structures', PhD Thesis, University of Patras, 2012.

[18] M. S. Kan, A. C. C. Tan, and J. Mathew, 'A review on prognostic techniques for non-stationary and non-linear rotating systems', *Mech. Syst. Signal Process.*, vol. 62–63, pp. 1–20, Oct. 2015, doi: 10.1016/j.ymssp.2015.02.016.

[19] R. B. Randall, *Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications*. John Wiley & Sons, 2011.

[20] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, 'Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications', *Mech. Syst. Signal Process.*, vol. 42, no. 1, pp. 314–334, Jan. 2014, doi: 10.1016/j.ymssp.2013.06.004.

[21] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, 'Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges', *Energies*, vol. 7, no. 4, pp. 2595–2630, Apr. 2014, doi: 10.3390/en7042595.

[22] D. Abboud, 'Vibration-based condition monitoring of rotating machines in nonstationary regime', 2015.

[23] M. Zhao, J. Lin, Y. Miao, and X. Xu, 'Feature Mining and Health Assessment for Gearboxes Using Run-Up/Coast-Down Signals', *Sensors*, vol. 16, no. 11, p. 1837, Nov. 2016, doi: 10.3390/s16111837.

[24] R. B. Randall, 'A New Method of Modeling Gear Faults', *J. Mech. Des.*, vol. 104, no. 2, p. 259, 1982, doi: 10.1115/1.3256334.

[25] W. Wang, 'Early detection of gear tooth cracking using the resonance demodulation technique', *Mech. Syst. Signal Process.*, vol. 15, no. 5, pp. 887–903, Sep. 2001, doi: 10.1006/mssp.2001.1416.

[26] P. D. McFadden, 'Examination of a technique for the early detection of failure in gears by signal processing of the time domain average of the meshing vibration', *Mech. Syst. Signal Process.*, vol. 1, no. 2, pp. 173–183, Apr. 1987, doi: 10.1016/0888-3270(87)90069-0.

[27] D. Abboud, J. Antoni, S. Sieg-Zieba, and M. Eltabach, 'Deterministic-random separation in nonstationary regime', *J. Sound Vib.*, vol. 362, pp. 305–326, Feb. 2016, doi: 10.1016/j.jsv.2015.09.029.

[28] D. Abboud, J. Antoni, S. Sieg-Zieba, and M. Eltabach, 'Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment', *Mech. Syst. Signal Process.*, vol. 84, pp. 200–226, Feb. 2017, doi: 10.1016/j.ymssp.2016.06.033.

[29] A. Napolitano, 'Cyclostationarity: New trends and applications', *Signal Process.*, vol. 120, no. Supplement C, pp. 385–408, Mar. 2016, doi: 10.1016/j.sigpro.2015.09.011.

[30] D. Abboud, J. Antoni, M. Eltabach, and S. Sieg-Zieba, 'Angle\time cyclostationarity for the analysis of rolling element bearing vibrations', *Measurement*, vol. 75, pp. 29–39, Nov. 2015, doi: 10.1016/j.measurement.2015.07.017.

[31] S. Schmidt, A. Mauricio, P. S. Heyns, and K. C. Gryllias, 'A methodology for identifying information rich frequency bands for diagnostics of mechanical components-of-interest under time-varying operating conditions', *Mech. Syst. Signal Process.*, vol. 142, p. 106739, Aug. 2020, doi: 10.1016/j.ymssp.2020.106739.

[32] International Organization for Standardizatio, *Mechanical vibration-Evaluation of machine vibration by measurements on non-rotating parts.*, Second Edition. ISO 10816-7, 2009.

[33] S. Braun, 'The synchronous (time domain) average revisited', *Mech. Syst. Signal Process.*, vol. 25, no. 4, pp. 1087–1102, May 2011, doi: 10.1016/j.ymssp.2010.07.016.

[34] P. D. Samuel and D. J. Pines, 'A review of vibration-based techniques for helicopter transmission diagnostics', *J. Sound Vib.*, vol. 282, no. 1, pp. 475–508, Apr. 2005, doi: 10.1016/j.jsv.2004.02.058.

[35] S. Braun, 'Extraction of periodic waveforms by time domain averaging', *Acustica*, vol. 32, no. 2, pp. 69–77, 1975.

[36] V. Sharma and A. Parey, 'Gear crack detection using modified TSA and proposed fault indicators for fluctuating speed conditions', *Measurement*, vol. 90, pp. 560–575, Aug. 2016, doi: 10.1016/j.measurement.2016.04.076.

[37] P. J. Dempsey and S. Sheng, 'Investigation of data fusion applied to health monitoring of wind turbine drivetrain components', *Wind Energy*, vol. 16, no. 4, pp. 479–489, 2013, doi: 10.1002/we.1512.

[38] P. D. McFadden, 'A technique for calculating the time domain averages of the vibration of the individual planet gears and the sun gear in an epicyclic gearbox', *J. Sound Vib.*, vol. 144, no. 1, pp. 163–172, Jan. 1991, doi: 10.1016/0022-460X(91)90739-7.

[39] J. Antoni, 'Fast computation of the kurtogram for the detection of transient faults', *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 108–124, Jan. 2007, doi: 10.1016/j.ymssp.2005.12.002.

[40] J. J. Zakajsek, 'A Review of Transmission Diagnostics Research at NASA Lewis Research Center', Dec. 1994. Accessed: Feb. 17, 2020. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=19950010492.

[41] X. Xu, M. Zhao, J. Lin, and Y. Lei, 'Envelope harmonic-to-noise ratio for periodic impulses detection and its application to bearing diagnosis', *Measurement*, vol. 91, pp. 385–397, Sep. 2016, doi: 10.1016/j.measurement.2016.05.073.

[42] W. Wang and A. K. Wong, 'Autoregressive Model-Based Gear Fault Diagnosis', *J. Vib. Acoust.*, vol. 124, no. 2, pp. 172–179, Mar. 2002, doi: 10.1115/1.1456905.

[43] G. L. McDonald, Q. Zhao, and M. J. Zuo, 'Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection', *Mech. Syst. Signal Process.*, vol. 33, pp. 237–255, Nov. 2012, doi: 10.1016/j.ymssp.2012.06.010.

[44] S. D. Fassois and F. P. Kopsaftopoulos, 'Statistical Time Series Methods for Vibration Based Structural Health Monitoring', in *New Trends in Structural Health*

Monitoring, vol. 542, W. Ostachowicz and J. A. Güemes, Eds. Vienna: Springer Vienna, 2013, pp. 209–264.

[45] F. P. Kopsaftopoulos and S. D. Fassois, 'A functional model based statistical time series method for vibration based damage detection, localization, and magnitude estimation', *Mech. Syst. Signal Process.*, vol. 39, no. 1, pp. 143–161, Aug. 2013, doi: 10.1016/j.ymssp.2012.08.023.

[46] H. Endo and R. B. Randall, 'Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter', *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 906–919, Feb. 2007, doi: 10.1016/j.ymssp.2006.02.005.

[47] G. L. McDonald and Q. Zhao, 'Multipoint Optimal Minimum Entropy Deconvolution and Convolution Fix: Application to vibration fault detection', *Mech. Syst. Signal Process.*, vol. 82, pp. 461–477, Jan. 2017, doi: 10.1016/j.ymssp.2016.05.036.

[48] G. Li, G. L. McDonald, and Q. Zhao, 'Sinusoidal synthesis based adaptive tracking for rotating machinery fault detection', *Mech. Syst. Signal Process.*, vol. 83, no. Supplement C, pp. 356–370, Jan. 2017, doi: 10.1016/j.ymssp.2016.06.019.

[49] G. Li and Q. Zhao, 'Minimum entropy deconvolution optimized sinusoidal synthesis and its application to vibration based fault detection', *J. Sound Vib.*, vol. 390, pp. 218–231, Mar. 2017, doi: 10.1016/j.jsv.2016.11.033.

[50] C. S. Sakaris, J. S. Sakellariou, and S. D. Fassois, 'A time series generalized functional model based method for vibration-based damage precise localization in structures consisting of 1D, 2D, and 3D elements', *Mech. Syst. Signal Process.*, vol. 74, pp. 199–213, Jun. 2016, doi: 10.1016/j.ymssp.2015.07.014.

[51] J. D. Hios and S. D. Fassois, 'A global statistical model based approach for vibration response-only damage detection under various temperatures: A proof-of-concept study', *Mech. Syst. Signal Process.*, vol. 49, no. 1, pp. 77–94, Dec. 2014, doi: 10.1016/j.ymssp.2014.02.005.

[52] C. Junsheng, Y. Dejie, and Y. Yu, 'A fault diagnosis approach for roller bearings based on EMD method and AR model', *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 350–362, Feb. 2006, doi: 10.1016/j.ymssp.2004.11.002.

[53] Z. Man, W. Wang, S. Khoo, and J. Yin, 'Optimal sinusoidal modelling of gear mesh vibration signals for gear diagnosis and prognosis', *Mech. Syst. Signal Process.*, vol. 33, no. Supplement C, pp. 256–274, Nov. 2012, doi: 10.1016/j.ymssp.2012.07.004.

[54] J. Yin, W. Wang, Z. Man, and S. Khoo, 'Statistical modeling of gear vibration signals and its application to detecting and diagnosing gear faults', *Inf. Sci.*, vol. 259, no. Supplement C, pp. 295–303, Feb. 2014, doi: 10.1016/j.ins.2013.03.029.

[55] J. Ma, F. Xu, K. Huang, and R. Huang, 'GNAR-GARCH model and its application in feature extraction for rolling bearing fault diagnosis', *Mech. Syst. Signal Process.*, vol. 93, no. Supplement C, pp. 175–203, Sep. 2017, doi: 10.1016/j.ymssp.2017.01.043.

[56] M. D. Spiridonakos and S. D. Fassois, 'Non-stationary random vibration modelling and analysis via functional series time-dependent ARMA (FS-TARMA) models – A critical survey', *Mech. Syst. Signal Process.*, vol. 47, no. 1–2, pp. 175–224, Aug. 2014, doi: 10.1016/j.ymssp.2013.06.024.

[57] M. Zhao and X. Jia, 'A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery', *Mech. Syst. Signal Process.*, vol. 94, pp. 129–147, Sep. 2017, doi: 10.1016/j.ymssp.2017.02.036.

[58] Z. Qiao and Z. Pan, 'SVD principle analysis and fault diagnosis for bearings based on the correlation coefficient', *Meas. Sci. Technol.*, vol. 26, no. 8, p. 085014, 2015, doi: 10.1088/0957-0233/26/8/085014.

[59] X. Zhao and B. Ye, 'Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock', *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1617–1631, Jul. 2011, doi: 10.1016/j.ymssp.2011.01.003.

[60] R. Golafshan and K. Yuce Sanliturk, 'SVD and Hankel matrix based de-noising approach for ball bearing fault detection and its assessment using artificial faults', *Mech. Syst. Signal Process.*, vol. 70–71, pp. 36–50, Mar. 2016, doi: 10.1016/j.ymssp.2015.08.012.

[61] H. Jiang, J. Chen, G. Dong, T. Liu, and G. Chen, 'Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis', *Mech. Syst. Signal Process.*, vol. 52–53, pp. 338–359, Feb. 2015, doi: 10.1016/j.ymssp.2014.07.019.

[62] S. Zhang, S. Lu, Q. He, and F. Kong, 'Time-varying singular value decomposition for periodic transient identification in bearing fault diagnosis', *J. Sound Vib.*, vol. 379, pp. 213–231, Sep. 2016, doi: 10.1016/j.jsv.2016.05.035.

[63] K. R. Fyfe and E. D. S. Munck, 'Analysis of computed order tracking', *Mech. Syst. Signal Process.*, vol. 11, no. 2, pp. 187–205, Mar. 1997, doi: 10.1006/mssp.1996.0056.

[64] H. Vold and J. Leuridan, 'High Resolution Order Tracking at Extreme Slew Rates, Using Kalman Tracking Filters', SAE International, Warrendale, PA, SAE Technical Paper 931288, May 1993. doi: 10.4271/931288.

[65] M.-Ch. Pan and Y.-F. Lin, 'Further exploration of Vold–Kalman-filtering order tracking with shaft-speed information—I: Theoretical part, numerical implementation and parameter investigations', *Mech. Syst. Signal Process.*, vol. 20, no. 5, pp. 1134–1154, Jul. 2006, doi: 10.1016/j.ymssp.2005.01.005.

[66] M.-Ch. Pan and Y.-F. Lin, 'Further exploration of Vold–Kalman-filtering order tracking with shaft-speed information—II: Engineering applications', *Mech. Syst. Signal Process.*, vol. 20, no. 6, pp. 1410–1428, Aug. 2006, doi: 10.1016/j.ymssp.2005.01.007.

[67] M.-C. Pan and C.-X. Wu, 'Adaptive Vold–Kalman filtering order tracking', *Mech. Syst. Signal Process.*, vol. 21, no. 8, pp. 2957–2969, Nov. 2007, doi: 10.1016/j.ymssp.2007.06.002.

[68] P. Borghesani, P. Pennacchi, S. Chatterton, and R. Ricci, 'The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery', *Mech. Syst. Signal Process.*, vol. 44, no. 1, pp. 118–133, Feb. 2014, doi: 10.1016/j.ymssp.2013.03.026.

[69] S. Schmidt and P. S. Heyns, 'Normalisation of the amplitude modulation caused by time-varying operating conditions for condition monitoring', *Measurement*, vol. 149, p. 106964, Jan. 2020, doi: 10.1016/j.measurement.2019.106964.

[70] J. Urbanek, T. Barszcz, M. Strączkiewicz, and A. Jablonski, 'Normalization of vibration signals generated under highly varying speed and load with application to signal separation', *Mech. Syst. Signal Process.*, vol. 82, pp. 13–31, Jan. 2017, doi: 10.1016/j.ymssp.2016.04.017.

[71] J. Wang, F. Cheng, W. Qiao, and L. Qu, 'Multiscale Filtering Reconstruction for Wind Turbine Gearbox Fault Diagnosis Under Varying-Speed and Noisy Conditions', *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4268–4278, May 2018, doi: 10.1109/TIE.2017.2767520.

[72] V. Sharma and A. Parey, !'Gearbox fault diagnosis using RMS based probability density function and entropy measures for fluctuating speed conditions', *Struct. Health Monit.*, vol. 16, no. 6, pp. 682–695, Nov. 2017, doi: 10.1177/1475921716679802.

[73] K. Gryllias, S. Moschini, and J. Antoni, 'Application of Cyclo-Nonstationary Indicators for Bearing Monitoring Under Varying Operating Conditions', *J. Eng. Gas Turbines Power*, vol. 140, no. 1, Jan. 2018, doi: 10.1115/1.4037638.

[74] L. F. Villa, A. Reñones, J. R. Perán, and L. J. de Miguel, 'Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load', *Mech. Syst. Signal Process.*, vol. 29, pp. 436–446, May 2012, doi: 10.1016/j.ymssp.2011.12.013.

[75] P. Santos, L. F. Villa, A. Reñones, A. Bustillo, and J. Maudes, 'An SVM-Based Solution for Fault Detection in Wind Turbines', *Sensors*, vol. 15, no. 3, pp. 5627–5648, Mar. 2015, doi: 10.3390/s150305627.

[76] S. Schmidt, P. S. Heyns, and J. P. de Villiers, 'A novelty detection diagnostic methodology for gearboxes operating under fluctuating operating conditions using probabilistic techniques', *Mech. Syst. Signal Process.*, vol. 100, pp. 152–166, Feb. 2018, doi: 10.1016/j.ymssp.2017.07.032.

[77] L. Liu, 'Vibration Signal Analysis for Planetary Gearbox Fault Diagnosis', Doctoral Dissertation, University of Alberta (Canada), Edmonton, Alberta, Canada, 2018.

[78] Z. Feng, M. Liang, and F. Chu, 'Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples', *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 165–205, Jul. 2013, doi: 10.1016/j.ymssp.2013.01.017.

[79] Y. Yang, Z. Peng, W. Zhang, and G. Meng, 'Parameterised time-frequency analysis methods and their engineering applications: A review of recent advances', *Mech. Syst. Signal Process.*, vol. 119, pp. 182–221, Mar. 2019, doi: 10.1016/j.ymssp.2018.07.039.

[80] X. Wang, V. Makis, and M. Yang, 'A wavelet approach to fault diagnosis of a gearbox under varying load conditions', *J. Sound Vib.*, vol. 329, no. 9, pp. 1570–1585, Apr. 2010, doi: 10.1016/j.jsv.2009.11.010.

[81] I. Antoniadou, G. Manson, W. J. Staszewski, T. Barszcz, and K. Worden, 'A time–frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions', *Mech. Syst. Signal Process.*, vol. 64, no. Supplement C, pp. 188–216, Dec. 2015, doi: 10.1016/j.ymssp.2015.03.003.

[82] S. Wang, X. Chen, I. W. Selesnick, Y. Guo, C. Tong, and X. Zhang, 'Matching synchrosqueezing transform: A useful tool for characterizing signals with fast varying instantaneous frequency and application to machine fault diagnosis', *Mech. Syst. Signal Process.*, vol. 100, pp. 242–288, Feb. 2018, doi: 10.1016/j.ymssp.2017.07.009.

[83] Z. Feng, W. Zhu, and D. Zhang, 'Time-Frequency demodulation analysis via Vold-Kalman filter for wind turbine planetary gearbox fault diagnosis under nonstationary speeds', *Mech. Syst. Signal Process.*, vol. 128, pp. 93–109, Aug. 2019, doi: 10.1016/j.ymssp.2019.03.036.

[84] A. Wyłomańska, J. Obuchowski, R. Zimroz, and H. Hurd, 'Periodic Autoregressive Modeling of Vibration Time Series From Planetary Gearbox Used in Bucket Wheel Excavator', in *Cyclostationarity: Theory and Methods*, F. Chaari, J. Leśkow, A.

Napolitano, and A. Sanchez-Ramirez, Eds. Springer International Publishing, 2014, pp. 171–186.

[85] Y. Zhan and C. K. Mechefske, 'Robust detection of gearbox deterioration using compromised autoregressive modeling and Kolmogorov–Smirnov test statistic—Part I: Compromised autoregressive modeling with the aid of hypothesis tests and simulation analysis', *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 1953–1982, Jul. 2007, doi: 10.1016/j.ymssp.2006.11.005.

[86] Y. Shao and C. K. Mechefske, 'Gearbox vibration monitoring using extended Kalman filters and hypothesis tests', *J. Sound Vib.*, vol. 325, no. 3, pp. 629–648, Aug. 2009, doi: 10.1016/j.jsv.2009.03.029.

[87] T. Heyns, S. J. Godsill, J. P. de Villiers, and P. S. Heyns, 'Statistical gear health analysis which is robust to fluctuating loads and operating speeds', *Mech. Syst. Signal Process.*, vol. 27, pp. 651–666, Feb. 2012, doi: 10.1016/j.ymssp.2011.09.007.

[88] R. Makowski and R. Zimroz, 'A procedure for weighted summation of the derivatives of reflection coefficients in adaptive Schur filter with application to fault detection in rolling element bearings', *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 65–77, Jul. 2013, doi: 10.1016/j.ymssp.2012.05.005.

[89] M. Yang and V. Makis, 'ARX model-based gearbox fault detection and localization under varying load conditions', *J. Sound Vib.*, vol. 329, no. 24, pp. 5209–5221, Nov. 2010, doi: 10.1016/j.jsv.2010.07.001.

[90] Y. Chen, X. Liang, and M. J. Zuo, 'Time series modeling of vibration signals from a gearbox under varying speed and load condition', in *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, Jun. 2018, pp. 1–7, doi: 10.1109/ICPHM.2018.8449003.

[91] F. Kopsaftopoulos, R. Nardari, Y.-H. Li, and F.-K. Chang, 'A stochastic global identification framework for aerospace structures operating under varying flight states', *Mech. Syst. Signal Process.*, vol. 98, pp. 425–447, Jan. 2018, doi: 10.1016/j.ymssp.2017.05.001.

[92] J. S. Sakellariou and S. D. Fassois, 'Functionally Pooled models for the global identification of stochastic systems under different pseudo-static operating conditions', *Mech. Syst. Signal Process.*, vol. 72–73, pp. 785–807, May 2016, doi: 10.1016/j.ymssp.2015.10.018.

[93] T.-C. I. Aravanis, J. S. Sakellariou, and S. D. Fassois, 'A stochastic Functional Model based method for random vibration based robust fault detection under variable non–measurable operating conditions with application to railway vehicle suspensions', *J. Sound Vib.*, p. 115006, Oct. 2019, doi: 10.1016/j.jsv.2019.115006.

[94] J. S. Sakellariou and S. D. Fassois, 'A functional pooling framework for the identification of systems under multiple operating conditions', in *2007 Mediterranean Conference on Control Automation*, Jun. 2007, pp. 1–6, doi: 10.1109/MED.2007.4433918.

[95] J. S. Sakellariou and S. D. Fassois, 'Global identification of stochastic dynamical systems under different pseudo-static operating conditions: The functionally pooled ARMAX case', *Mech. Syst. Signal Process.*, vol. 82, pp. 32–55, Jan. 2017, doi: 10.1016/j.ymssp.2016.05.002.

[96] F. P. Kopsaftopoulos and S. D. Fassois, 'Vector-dependent Functionally Pooled ARX Models for the Identification of Systems Under Multiple Operating Conditions', *IFAC Proc. Vol.*, vol. 45, no. 16, pp. 310–315, Jul. 2012, doi: 10.3182/20120711-3-BE-2027.00261.

[97] C. S. Sakaris, J. S. Sakellariou, and S. D. Fassois, 'Random-vibration-based damage detection and precise localization on a lab–scale aircraft stabilizer structure via the Generalized Functional Model Based Method', *Struct. Health Monit.*, vol. 16, no. 5, pp. 594–610, Sep. 2017, doi: 10.1177/1475921717707903.

[98] B. U. Park, E. Mammen, Y. K. Lee, and E. R. Lee, 'Varying Coefficient Regression Models: A Review and New Developments', *Int. Stat. Rev.*, vol. 83, no. 1, pp. 36–64, Apr. 2015, doi: 10.1111/insr.12029.

[99] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. John Wiley & Sons, 2013.

[100] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: data mining, inference, and prediction.*, 2nd ed. 2009.

[101] L. Fenga and D. N. Politis, 'LASSO order selection for sparse autoregression: a bootstrap approach', *J. Stat. Comput. Simul.*, vol. 87, no. 14, pp. 2668–2688, Sep. 2017, doi: 10.1080/00949655.2017.1341885.

[102] M. D. Spiridonakos and S. D. Fassois, 'An FS-TAR based method for vibration-response-based fault diagnosis in stochastic time-varying structures: Experimental application to a pick-and-place mechanism', *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 206–222, Jul. 2013, doi: 10.1016/j.ymssp.2012.05.014.

[103] D. G. Dimogianopoulos, J. D. Hios, and S. D. Fassois, 'FDI for Aircraft Systems Using Stochastic Pooled-NARMAX Representations: Design and Assessment', *IEEE Trans. Control Syst. Technol.*, vol. 17, no. 6, pp. 1385–1397, Nov. 2009, doi: 10.1109/TCST.2008.2005836.

[104] P. Ghaderyan, A. Abbasi, and A. Ebrahimi, 'Time-varying singular value decomposition analysis of electrodermal activity: A novel method of cognitive load estimation', *Measurement*, vol. 126, pp. 102–109, Oct. 2018, doi: 10.1016/j.measurement.2018.05.015.

[105] Y. Chen and M. J. Zuo, 'Early gear tooth crack detection based on singular value decomposition', in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, Jun. 2019, pp. 1–6, doi: 10.1109/ICPHM.2019.8819417.

[106] Y. Chen, X. Liang, and M. J. Zuo, 'An improved singular value decomposition-based method for gear tooth crack detection and severity assessment', *J. Sound Vib.*, vol. 468, p. 115068, Mar. 2020, doi: 10.1016/j.jsv.2019.115068.

[107] Y. Chen, X. Liang, and M. J. Zuo, 'An autoregression model for tooth crack detection considering load variations', Edmonton, Alberta, Canada, Oct. 2017, pp. 1–12.

[108] Y. Chen, X. Liang, and M. J. Zuo, 'Sparse time series modeling of the baseline vibration from a gearbox under time-varying speed condition', *Mech. Syst. Signal Process.*, vol. 134, p. 106342, Dec. 2019, doi: 10.1016/j.ymssp.2019.106342.

[109] Y. Chen, S. Schmidt, P. S. Heyns, and M. J. Zuo, 'A time series model-based method for gear tooth crack detection and severity assessment under random speed variation', *Mech. Syst. Signal Process.*, pp. 1–32, Feb. 2020, doi: Under Review.

[110] A. K. S. Jardine, D. Lin, and D. Banjevic, 'A review on machinery diagnostics and prognostics implementing condition-based maintenance', *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006, doi: 10.1016/j.ymssp.2005.09.012.

[111] R. Yan, R. X. Gao, and X. Chen, 'Wavelets for fault diagnosis of rotary machines: A review with applications', *Signal Process.*, vol. 96, pp. 1–15, Mar. 2014, doi: 10.1016/j.sigpro.2013.04.015.

[112] Y. Lei, J. Lin, Z. He, and M. J. Zuo, 'A review on empirical mode decomposition in fault diagnosis of rotating machinery', *Mech. Syst. Signal Process.*, vol. 35, no. 1, pp. 108–126, Feb. 2013, doi: 10.1016/j.ymssp.2012.09.015.

[113] J. Antoni, 'The spectral kurtosis: a useful tool for characterising non-stationary signals', *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 282–307, Feb. 2006, doi: 10.1016/j.ymssp.2004.09.001.

[114] Y. Liao, P. Sun, B. Wang, and L. Qu, 'Extraction of repetitive transients with frequency domain multipoint kurtosis for bearing fault diagnosis', *Meas. Sci. Technol.*, vol. 29, no. 5, p. 055012, 2018, doi: 10.1088/1361-6501/aaae99.

[115] D. Zhu, Y. Zhang, and Q. Zhu, 'Fault Diagnosis Method for Rolling Element Bearings Under Variable Speed Based on TKEO and Fast-SC', *J. Fail. Anal. Prev.*, vol. 18, no. 1, pp. 2–7, Feb. 2018, doi: 10.1007/s11668-018-0383-y.

[116] D. Abboud, S. Baudin, J. Antoni, D. Rémond, M. Eltabach, and O. Sauvage, 'The spectral analysis of cyclo-non-stationary signals', *Mech. Syst. Signal Process.*, vol. 75, no. Supplement C, pp. 280–300, Jun. 2016, doi: 10.1016/j.ymssp.2015.09.034.

[117] X. Chen, Z. Du, J. Li, X. Li, and H. Zhang, 'Compressed sensing based on dictionary learning for extracting impulse components', *Signal Process.*, vol. 96, pp. 94–109, Mar. 2014, doi: 10.1016/j.sigpro.2013.04.018.

[118] W. He, Y. Ding, Y. Zi, and I. W. Selesnick, 'Repetitive transients extraction algorithm for detecting bearing faults', *Mech. Syst. Signal Process.*, vol. 84, pp. 227–244, Feb. 2017, doi: 10.1016/j.ymssp.2016.06.035.

[119] J. Jiao, M. Zhao, J. Lin, and K. Liang, 'Hierarchical discriminating sparse coding for weak fault feature extraction of rolling bearings', *Reliab. Eng. Syst. Saf.*, Feb. 2018, doi: 10.1016/j.ress.2018.02.010.

[120] K. Zheng, T. Li, B. Zhang, Y. Zhang, J. Luo, and X. Zhou, 'Incipient Fault Feature Extraction of Rolling Bearings Using Autocorrelation Function Impulse Harmonic to Noise Ratio Index Based SVD and Teager Energy Operator', *Appl. Sci.*, vol. 7, no. 11, p. 1117, Oct. 2017, doi: 10.3390/app7111117.

[121] K. D. Rao, *Signals and Systems*. Birkhäuser Basel, 2018.

[122] Y. Miao, M. Zhao, and J. Lin, 'Periodicity-Impulsiveness Spectrum Based on Singular Value Negentropy and Its Application for Identification of Optimal Frequency Band', *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3127–3138, Apr. 2019, doi: 10.1109/TIE.2018.2844792.

[123] Y. Chen *et al.*, 'Experiment Design and Data Collection on a Fixed-axis Gearbox under Time-Varying Operation Conditions', Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada, Technical report, 2018.

[124] Z. Chen and Y. Shao, 'Dynamic simulation of spur gear with tooth root crack propagating along tooth width and crack depth', *Eng. Fail. Anal.*, vol. 18, no. 8, pp. 2149–2164, Dec. 2011, doi: 10.1016/j.engfailanal.2011.07.006.

[125] X. Fan, Y. Wang, W. Li, S. Wu, and M. Zuo, 'Experiment Design of Gear Crack Growth', Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada, Aug. 2005.

[126] A. Mbarek, A. Hammami, A. Fernandez Del Rincon, F. Chaari, F. Viadero Rueda, and M. Haddar, !'Effect of load and meshing stiffness variation on modal properties of planetary gear', *Appl. Acoust.*, Aug. 2017, doi: 10.1016/j.apacoust.2017.08.010.

[127] Y. Guo, T. Eritenel, T. M. Ericson, and R. G. Parker, 'Vibro-acoustic propagation of gear dynamics in a gear-bearing-housing system', *J. Sound Vib.*, vol. 333, no. 22, pp. 5762–5785, Oct. 2014, doi: 10.1016/j.jsv.2014.05.055.

[128] Y. Sun, A. K. C. Wong, and M. S. Kamel, 'Classification of Imbalanced Data: A Review', *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 04, pp. 687–719, Jun. 2009, doi: 10.1142/S0218001409007326.

[129] J. McBain and M. Timusk, 'Fault detection in variable speed machinery: Statistical parameterization', *J. Sound Vib.*, vol. 327, no. 3, pp. 623–646, Nov. 2009, doi: 10.1016/j.jsv.2009.07.025.

[130] R. Uma Maheswari and R. Umamaheswari, 'Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train – A contemporary survey', *Mech. Syst. Signal Process.*, vol. 85, no. Supplement C, pp. 296–311, Feb. 2017, doi: 10.1016/j.ymssp.2016.07.046.

[131] R. Zimroz and W. Bartelmus, 'Application of Adaptive Filtering for Weak Impulsive Signal Recovery for Bearings Local Damage Detection in Complex Mining Mechanical Systems Working under Condition of Varying Load', *Solid State Phenom.*, vol. 180, pp. 250–257, 2012, doi: 10.4028/www.scientific.net/SSP.180.250.

[132] R. B. Randall, 'Vibration-based diagnostics of gearboxes under variable speed and load conditions', *Meccanica*, vol. 51, no. 12, pp. 3227–3239, Dec. 2016, doi: 10.1007/s11012-016-0583-z.

[133] B. Assaad, M. Eltabach, and J. Antoni, 'Vibration based condition monitoring of a multistage epicyclic gearbox in lifting cranes', *Mech. Syst. Signal Process.*, vol. 42, no. 1, pp. 351–367, Jan. 2014, doi: 10.1016/j.ymssp.2013.06.032.

[134] S. Schmidt, P. S. Heyns, and J. P. de Villiers, 'A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process', *Mech. Syst. Signal Process.*, vol. 100, pp. 630–646, Feb. 2018, doi: 10.1016/j.ymssp.2017.07.053.

[135] T.-C. I. Aravanis, J. S. Sakellariou, and S. D. Fassois, 'Vibration based fault detection under variable non-measurable operating conditions via a stochastic Functional Model method and application to railway vehicle suspensions', *Conference*, p. 14, 2017.

[136] M. D. Spiridonakos and S. D. Fassois, 'Adaptable functional series TARMA models for non-stationary signal representation and their application to mechanical random vibration modeling', *Signal Process.*, vol. 96, Part A, pp. 63–79, Mar. 2014, doi: 10.1016/j.sigpro.2013.05.012.

[137] R. Tibshirani, 'Regression Shrinkage and Selection via the Lasso', *J. R. Stat. Soc. Ser. B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[138] J. D. Cryer and K.-S. Chan, *Time Series Analysis - With Applications in R*. Springer, 2008.

[139] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 'Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers', *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jul. 2011, doi: 10.1561/2200000016.

[140] C. Bergmeir, R. J. Hyndman, and B. Koo, 'A note on the validity of cross-validation for evaluating autoregressive time series prediction', *Comput. Stat. Data Anal.*, vol. 120, pp. 70–83, Apr. 2018, doi: 10.1016/j.csda.2017.11.003.

[141] G. M. Ljung and G. E. P. Box, 'On a measure of lack of fit in time series models', *Biometrika*, vol. 65, no. 2, pp. 297–303, Aug. 1978, doi: 10.1093/biomet/65.2.297.

[142] X.-H. Liang, Z.-L. Liu, J. Pan, and M. J. Zuo, 'Spur Gear Tooth Pitting Propagation Assessment Using Model-based Analysis', *Chin. J. Mech. Eng.*, vol. 30, no. 6, pp. 1369–1382, Nov. 2017, doi: 10.1007/s10033-017-0196-z.

[143] F. K. CHOY, Y. F. RUAN, J. ZAKRAJSEK, and F. B. OSWALD, 'Modal simulation of gear box vibration with experimental correlation', *J. Propuls. Power*, vol. 9, no. 2, pp. 301–306, 1993, doi: 10.2514/3.23622.

[144] X. Chen and Z. Feng, 'Time-Frequency Analysis of Torsional Vibration Signals in Resonance Region for Planetary Gearbox Fault Diagnosis Under Variable Speed Conditions', *IEEE Access*, vol. 5, pp. 21918–21926, 2017, doi: 10.1109/ACCESS.2017.2763172.

[145] Q. Leclère, H. André, and J. Antoni, 'A multi-order probabilistic approach for Instantaneous Angular Speed tracking debriefing of the CMMNO14′ diagnosis contest', *Mech. Syst. Signal Process.*, vol. 81, pp. 375–386, Dec. 2016, doi: 10.1016/j.ymssp.2016.02.053.

[146] J. Park, Y. Kim, K. Na, and B. D. Youn, 'Variance of energy residual (VER): An efficient method for planetary gear fault detection under variable-speed conditions', *J. Sound Vib.*, vol. 453, pp. 253–267, Aug. 2019, doi: 10.1016/j.jsv.2019.04.017.

[147] G. T. Knofczynski and D. Mundfrom, 'Sample Sizes When Using Multiple Linear Regression for Prediction', *Educ. Psychol. Meas.*, vol. 68, no. 3, pp. 431–442, Jun. 2008, doi: 10.1177/0013164407310131.

[148] S. Wang, X. Chen, I. W. Selesnick, Y. Guo, C. Tong, and X. Zhang, 'Matching synchrosqueezing transform: A useful tool for characterizing signals with fast varying instantaneous frequency and application to machine fault diagnosis', *Mech. Syst. Signal Process.*, vol. 100, no. Supplement C, pp. 242–288, Feb. 2018, doi: 10.1016/j.ymssp.2017.07.009.

[149] T.-C. I. Aravanis, J. S. Sakellariou, and S. D. Fassois, 'A stochastic Functional Model based method for random vibration based robust fault detection under variable non–measurable operating conditions with application to railway vehicle suspensions', *J. Sound Vib.*, p. 115006, Oct. 2019, doi: 10.1016/j.jsv.2019.115006.

[150] Z.-S. Ma *et al.*, 'Parametric output-only identification of time-varying structures using a kernel recursive extended least squares TARMA approach', *Mech. Syst. Signal Process.*, vol. 98, pp. 684–701, Jan. 2018, doi: 10.1016/j.ymssp.2017.05.013.

[151] M. Montardini, F. Remonato, and G. Sangalli, 'Isogeometric Methods for Free Boundary Problems', p. 24.

[152] International Standard Organization, 'ISO 10816: Mechanica vibration - Evaluation of machine vibration by measurements on non-rotating parts'. 2010, Accessed: Feb. 26, 2020. [Online]. Available: https://global.ihs.com/doc_detail.cfm?document_name=BS%20ISO%2010816%2D1&item_s_key=00537900&rid=GS.

[153] C. de Boor, 'A Practical Guide to Splines', 1978, doi: 10.1007/978-1-4612-6333-3.

[154] S. H. Schmidt and K. Gryllias, 'Discrepancy analysis for gearbox condition monitoring: A comparison of different healthy data models', in *ResearchGate*, Sun International's Lost City, South Africa, 2018, pp. 181–188, Accessed: Feb. 26, 2020. [Online]. Available: https://www.researchgate.net/publication/325256884_Discrepancy_analysis_for

_gearbox_condition_monitoring_A_comparison_of_different_healthy_data_models.

[155] S. Schmidt, P. S. Heyns, and K. C. Gryllias, 'A discrepancy analysis methodology for rolling element bearing diagnostics under variable speed conditions', *Mech. Syst. Signal Process.*, vol. 116, pp. 40–61, Feb. 2019, doi: 10.1016/j.ymssp.2018.06.026.

[156] T. Han, D. Jiang, Q. Zhao, L. Wang, and K. Yin, 'Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery', *Trans. Inst. Meas. Control*, vol. 40, no. 8, pp. 2681–2693, May 2018, doi: 10.1177/0142331217708242.

[157] G. A. Skrimpas, T. Ursin, C. Sweeney, K. Marhadi, N. Mijatovic, and J. Holboell, 'Residual signal feature extraction for gearbox planetary stage fault detection', *Wind Energy*, p. n/a-n/a, Jan. 2017, doi: 10.1002/we.2099.

[158] B. Liu and V. Makis, 'Gearbox failure diagnosis based on vector autoregressive modelling of vibration data and dynamic principal component analysis', *IMA J. Manag. Math.*, vol. 19, no. 1, pp. 39–50, Jan. 2008, doi: 10.1093/imaman/dpm002.

[159] H. Lütkepohl, *New introduction to multiple time series analysis*. Berlin: New York : Springer, 2005.

[160] C. S. Sakaris, J. S. Sakellariou, and S. D. Fassois, 'Vibration-based damage precise localization in three-dimensional structures: Single versus multiple response measurements', *Struct. Health Monit.*, vol. 14, no. 3, pp. 300–314, May 2015, doi: 10.1177/1475921714568407.

[161] M. Zhao, X. Jia, J. Lin, Y. Lei, and J. Lee, 'Instantaneous speed jitter detection via encoder signal and its application for the diagnosis of planetary gearbox', *Mech. Syst. Signal Process.*, vol. 98, pp. 16–31, Jan. 2018, doi: 10.1016/j.ymssp.2017.04.033.

[162] M. Rao, Q. Li, D. Wei, and M. J. Zuo, 'A deep bi-directional long short-term memory model for automatic rotating speed extraction from raw vibration signals', *Measurement*, p. 107719, Mar. 2020, doi: 10.1016/j.measurement.2020.107719.

[163] M. Wang, P. Zhao, and X. Kang, 'Structure identification for varying coefficient models with measurement errors based on kernel smoothing', *Stat. Pap.*, pp. 1–17, May 2018, doi: 10.1007/s00362-018-1009-x.