I, Zongchang Liu, hereby submit this original work as part of the requirements for the degree of Master of Science in Mechanical Engineering.

It is entitled:
A Systematic Framework for Unsupervised Feature Mining and Fault Detection for Wind Turbine Drivetrain Systems

Student's name: Zongchang Liu

This work and its defense approved by:

Committee chair: Jay Lee, Ph.D.
Committee member: Jay Kim, Ph.D.
Committee member: Allyn Phillips, Ph.D.
A Systematic Framework for Unsupervised Feature Mining and Fault Detection for Wind Turbine Drivetrain Systems

A thesis submitted to the
Division of Research and Advanced Studies
of the University of Cincinnati
in partial fulfillment of the requirements
for the Degree of
Master of Science
In the Department of Mechanical and Materials Engineering
of the College of Engineering and Applied Science
2016
by
Zongchang Liu

B.S.E. in Mechanical Engineering, University of Michigan, Ann Arbor (2012)
B.S.E. in Electrical and Computer Engineering, Shanghai Jiaotong University (2012)

Committee Chair: Dr. Jay Lee
Committee Member: Dr. Jay Kim
Committee Member: Dr. Allyn Phillips
ABSTRACT

The global installed capacity of wind turbines has been growing rapidly during the past decade. Along with the fast-growing number of wind turbines, the concerns for their maintenance and health management are also accumulating. The repair and maintenance for wind turbines are very expensive and time-consuming due to various reasons including logistics difficulties, distant locations, costly spare parts, and expensive labor force, etc. Prognostics and health management (PHM) technologies are of vital importance to wind turbines operation and maintenance since it can detect incipient faults in early time and predict the trend of their propagation so that the maintenance activities can be planned ahead of time to reduce the downtime and maintenance cost. Drivetrain systems are of the most concern in maintenance as they contribute the most downtime and repair costs. While there are various condition monitoring techniques available for drivetrain systems, vibration-based techniques have been most widely adopted due to its direct access to structure response and capability for early detection of incipient faults. However, there are impeding challenges of its application in wind turbine PHM: how to extract meaningful features from vibration signal when the rotating speed is unknown; how to detect and enhance incipient fault features under dynamic operation regimes and harsh environment; how to convert the multidimensional feature vectors into actionable health indicator to plan maintenance; and how to align these analytical techniques to enable smart, self-contained and unsupervised condition monitoring systems in big data environment.

This thesis presents a systematic framework for unsupervised feature mining and fault detection for drivetrain systems. It consists several novel techniques that address the critical issues for vibration-based condition monitoring: A novel method for instantaneous angular speed estimation under non-stationary operation conditions based on enhanced harmonics product
spectrum; Resonance band detection and incipient fault feature enhancement based on harmonics-targeting fast kurtogram; And a fleet-based data-driven fault detection method based on clustering and peer-to-fleet similarity assessment techniques. The contributions of the proposed framework to conventional vibration-based monitoring on drivetrain systems are mainly in two aspects, i.e., it can autonomously configure the signal processing techniques according to the signal characters and drivetrain design scheme, which makes the feature extraction process unsupervised and self-contained; the proposed fault detection modeling paradigm enables fleet level prognosis that has no requirement for context information of the health status of historical data.

The proposed framework is validated with two case studies, a full-scale wind turbine drivetrain test bed to validate the autonomous fault feature mining techniques, and a wind farm consisting of 48 wind turbines for fleet-based data-driven fault detection.
ACKNOLOGEMENT

I would like to show my special thanks to my academic advisor Professor Jay Lee who provided me research opportunity and guidance for my graduate study. I also would like to thank my committee member Prof. Jay Kim and Prof. Allyn Phillips for all their advice for my research. I also want to express my special thanks to Shanghai Electric and National Renewable Energy Laboratory for providing the data to develop and validate my research work. I am also thankful to former IMS researchers who provided me guidance and help throughout my graduate study: Dr. Edzel Lapira, Dr. Yan Chen, Dr. David Seigel, Dr. Wenyu Zhao, Dr. Shanhu Yang, Dr. Chuan Jiang, Dr. Mohammad Rezvani, Mr. Eric Huang and Ms. Xiaorui Tong. I also want to appreciate all our current IMS family members for all your help and support, including Ms. Wenjing Jin, Ms. Hung-An Kao, Mr. Hossein Davari, Mr. Chao Jin, Mr. Zhe Shi, Mr. Behrad Bagheri, Mr. Yuan Di, Mr. Matt Buzza, Ms. Laura Pahren, Mr. Aaron Shelly, Ms. Ellen Gamel. I would also like thank IMS staff Mr. Michael Lyons and Mr. Patrick Brown who are always supportive and helpful. At last, I want to thank my parents and my wife Huibin for all their support and encourage which provide me strength and faith to fulfil all this research work.
# Table of Contents

## CHAPTER 1 INTRODUCTION

1.1 Motivation .................................................................................................................. 1

1.2 Research Objectives, Tasks, and Methodology ......................................................... 6

1.3 Thesis Organization .................................................................................................... 8

## CHAPTER 2 REVIEW ON WIND TURBINE CONDITION MONITORING SYSTEMS AND ANALYTICAL TECHNIQUES

2.1 Operation, Maintenance, and Availability of Wind Turbines ................................. 10

2.1.1 The emerging trend of wind energy ........................................................................ 10

2.1.2 Cost of energy and life-cycle cost for wind energy .............................................. 12

2.1.3 Availability of wind turbines ............................................................................... 14

2.1.4 Benefits of prognostics and health management for wind turbines ................. 16

2.2 Failure Modes and Critical Components Analysis for Wind Turbines ................. 17

2.3 Review of Condition Monitoring Techniques for Drivetrain Systems ............... 23

2.4 Data Analytics Techniques for Drivetrain Vibration Signals ............................... 27

2.4.1 Review on signal processing techniques ................................................................. 27

2.4.2 Data-driven modeling techniques for drivetrain fault detection ..................... 30

2.5 Problem Statement for Condition Monitoring of Drivetrain Systems ............... 34

## CHAPTER 3 FRAMEWORK AND TECHNIQUES FOR UNSUPERVISED FAULT FEATURE MINING AND FAULT DETECTION

3.1 IAS Estimation with Enhanced Harmonic Product Spectrum ............................. 37

3.1.1 Development of enhanced harmonic product spectrum ..................................... 37

3.1.2 IAS estimation with relationship mining of HRSS fundamental frequencies .... 41

3.1.3 Tachometer signal synthesize and synchronized averaging ............................ 43

3.2 EHPS for Resonance Band Detection and Bearing Fault Feature Enhancement 46
List of Figures

Figure 2-1 Global installed wind capacity from 2000 to 2015 (Source: GWC 2015) .................. 11

Figure 2-2 Cost breakdown of wind energy (source: EWEA, 2009) ........................................ 13

Figure 2-3 Cost model for different maintenance strategies [24] ........................................... 17

Figure 2-4 Maintenance cost breakdown for an offshore wind farm in Denmark [27] ............. 18

Figure 2-5 Wind Turbine subsystems decomposition ................................................................. 19

Figure 2-6 Maintenance time and costs for different maintenance categories ......................... 20

Figure 2-7 Contribution of subsystems to downtime due to maintenance ............................... 21

Figure 2-8 Contribution of subsystems to maintenance costs .................................................. 21

Figure 2-9 Four-Quadrant chart for criticality analysis of subsystems ................................... 22

Figure 2-10 Configuration outline of a typical wind turbine vibration-based CMS .................. 26

Figure 2-11 Dynamic operation regime of wind turbine drivetrain ......................................... 36

Figure 2-12 Nonlinear relationship of monitoring parameters .................................................. 36

Figure 3-1 An adaptive scheme for EHPS calculation ............................................................... 40

Figure 3-2 Illustration of GMF identification by EHPS: (a): raw signal; (b) filtered frequency spectrum; (c) Effective HRSS by number of orders; (d) Effective HRSS and identified GMFs. ......................................................................................................................... 41

Figure 3-3: (a) Ratio relationship of effective FFs in the drivetrain vibration data set, and (b) distribution of estimated rotor speed ......................................................................................................................... 43
Figure 3-4: Procedure for HRSS relationships mining and tachometer signal synthesizing .......... 45

Figure 3-5: Illustration of EHPS based speed prediction and synchronous analysis .................. 45

Figure 3-6: Procedure for SK based fast kurtogram and Harmonic-targeting fast kurtogram ..... 48

Figure 3-6: Benchmark of Fast Kurogram and HTFK for resonance band detection: (a) Raw signal; (b) SK based Fast Kurtogram (color bar unit is spectral kurtosis) (c) EHPS based fast kurtogram (color bar unit is harmonic significance in EHPS at BPFI frequency) ............. 49

Figure 3-7: Illustration of envelope spectrum: (a) Envelope from SK based fast kurtogram; (b) GMF peaks are dominant at envelope spectrum; (c) Envelope from EHPS based fast kurtogram; (d) BPFI fault peaks are dominant at envelops spectrum ....................... 50

Figure 3-8: The procedure for unsupervised fault feature mining for drivetrain vibration signals .................................................................................................................................. 52

Figure 3-9: The step-by-step illustration for unsupervised fault feature mining for drivetrain vibration signals .................................................................................................................................. 54

Figure 3-10: Procedure for fleet-based data-driven fault detection ........................................ 58

Figure 4-1: Diagram of NREL dynamometer test facility and the drivetrain system being tested61

Figure 4-2: Gearbox scheme and nomenclature ..................................................................... 61

Figure 4-3: Drivetrain accelerometer locations and sensor nomenclature............................. 63

Figure 4-4: Intermediate steps for gear-meshing frequency identification: (a) raw signal; (b) Enhanced harmonic product spectrum; (c) Effective harmonics and identified GMF ....... 64

Figure 4-5: Results of instantaneous speed prediction ............................................................ 65
Figure 4-6: Gear-meshing frequency detection with predicted IS is more accurate

Figure 4-7: HSS GMF is obvious modulated by HS order, which suggest gear surface damage at HSS gear set

Figure 4-8: HSS GMF is obvious modulated by IS order, which suggest gear surface damage at ISS gear set

Figure 4-9: 2X GMF amplitude is five times higher than 1X GMF amplitude, which suggest misalignment on HS

Figure 4-10: (a) ISS pinion sideband with respect to ISS orders; (b) planetary gearbox sideband with respect to ring gear carrier frequency

Figure 4-11: (a) Gear surface damage features: (a) HSS pinion; (b) planet ring gear; (d) ISS pinion; and (d) ISS gear surface damage features

Figure 4-12: Procedure for unsupervised feature extraction for bearing defects

Figure 4-13: (a) Optimal resonance band for HSS downwind bearing BPFI frequency from HTFK, and (b) Envelop order spectrum of band-pass filtered signal

Figure 4-14: (a) ISS downwind bearing outer race fault: (a) Optimal resonance band from HTFK, and (b) Envelop order spectrum of band-pass filtered signal

Figure 4-15: (a) LSS downwind bearing cage fault: (a) optimal resonance frequency band for FTF harmonics, and (b) order spectrum of envelop analysis

Figure 4-16: Scheme of monitored wind turbines drivetrain, sensor locations, and DAQ configurations
Figure 4-17: Illustration of tachometer synthesizing and synchronous averaging: (a) raw signal; (b) frequency spectrum; (c) enhanced harmonic product spectrum; (d) synthesized tachometer signal; (e) time synchronous averaged signal; (f) order spectrum of synchronous averaged data ......................................................... 78

Figure 4-18: Procedure of fleet-based data-driven fault detection method ........................................ 80

Figure 4-19: Fitted relationship between rotor speed and: (a) HSS order magnitude; (b) HSS GMF order magnitude ........................................................................................................ 81

Figure 4-20: Health assessment results for all 48 turbines over time ................................................. 83

Figure 4-21: Health assessment results for all 48 turbines over time ............................................... 84

Figure 4-22: Order spectrum of turbine No. 38 in (a) normal condition and (b) defected condition ......................................................................................................................... 85
List of Tables

Table 2-1 Comparison of LCOE in 2010 and 2013 (Data Source: NREL 2010, 2013) ............... 14

Table 2-2 Summarization of condition monitoring techniques............................................. 24

Table 4-1 Description of gear elements and transmission ratio ........................................... 62

Table 4-2 Sensor nomenclature and location description ...................................................... 63

Table 4-3 Bearing characteristic frequencies and failure mode information ....................... 70

Table 4-4 Fault detection results and actual gearbox damages ............................................. 75

Table 4-5 Gearbox elements and tooth numbers .................................................................. 76
CHAPTER 1 INTRODUCTION

1.1 Motivation

As the concerns for global warming and atmospheric pollution becoming intense, the demand for alternative energy source increases inherently. In this context, wind energy market has been growing exponentially during the past decade. By the end of 2015, the cumulative installed capacity of wind energy has reached to 423GW globally [1]. The US Department of Energy (DoE) claims that it is technically feasible to meet its goal of using wind energy to supply 20% of the total energy requirements in US by 2030, but this will involve extensive researches in all aspects such as structural design, manufacturing, operation and maintenance (O&M), and construction [2].

In spite of the inspiring facts about wind power industry, there are also fast-growing concerns for their O&M as the number of wind farms accumulates and the scale of wind turbines increases. As indicated in report [3], O&M cost is estimated to take 20-25% of total life-cycle cost (LCC) for offshore turbines and 18% for onshore turbines over the lifetime. In addition to the direct expenses of repair and maintenance, indirect costs associated with downtime and loss of production are also considerable. According to Wind System Magazine, 70% of total wind turbine maintenance costs are from unscheduled downtime. And for a 100 MW scale wind farm, only one percentage of availability increase can be worth between $300-500K of revenue per year. This view is also shared by the European Wind Energy Association (EWEA) to suggest that there is still much room to improve the cost of ownership for wind turbines [4].
Drivetrain system is one of the most critical subsystems for wind turbine. The average cost per failure of €230,000 for drivetrain gearbox is the highest among all other components. Generator and gearbox failures add up to 95% of all major replacement, and thus contribute the most unscheduled downtime and repair costs for wind turbines [5]. Therefore, it spurred the need for condition monitoring (CM) systems to detect incipient faults in drive train systems that can be fixed through minor repair. Among all CM approaches for wind turbine drivetrain systems, vibration-based condition monitoring is most widely adopted due to its sensitive and reliable fault detection capabilities. Conventional vibration-based condition monitoring relies on the experience of vibration analyst to look at the indicative features in the frequency spectrum of vibration signal, which is getting impractical nowadays since the wind turbine operators acquire an emerging number of turbines. Crabtree, C.J et al, conducted a survey on commercial CM systems for wind turbines, and concluded that autonomous, continuous, rapid, and reliable monitoring methods are very important in the big data environment of wind industry [6].

Prognostics and Health Management (PHM) is a multidisciplinary engineering field that aims to assess, predict, and optimize the health status of complex machinery systems. Being an important methodology category in PHM, data-driven methods are promising for monitoring wind turbine health due to its flexibility to different levels of signal availability and easiness of implementation. The general procedure for data-driven vibration-based CM is presented in Figure 1-1. The first step for this process is to identify the operation regime, including the rotating speed, load, and kinematics related components in the vibration signal. This step usually requires the assistance of tachometer to provide accurate estimation of instantaneous angular speed to interpret the harmonics in the vibration signal. The next step is to identify the fault related components in the vibration signal and extract quantitative features. This step usually
needs expert knowledge to define the fault indicative features, and involves advanced signal processing techniques for de-noising and weak feature enhancement. After the feature extraction process, each feature vector needs to be assigned with a label representing its health status to prepare the model training data. The feature vectors can be either just from a known good condition (unsupervised learning), or from both good and abnormal condition (supervised learning). The process of labeling for training data requires health context information, which can be either from maintenance record or system alarm data. The last step is to choose a data-driven modeling scheme to train a fault detection model with the selected training data. When implementing the model in condition monitoring, the presence and level of criticality of faults are assessed through performing an objective comparison with features extracted from both healthy and faulty conditions.

![Diagram](image1.png)

**Figure 1-1 General process for data-driven vibration-based CM approach**

However, the existing data-driven approaches have certain limitations at every steps when applied to vibration-based CM of wind turbines:

1) Tachometer signal is usually not available in wind turbine drivetrain condition monitoring system, and therefore identification of operation regimes is very difficult.
2) The existing CM systems depend on technicians to perform manual inspection on the signals. The number of turbines a technician can oversee is limited, and this challenge is growing rapidly with the emerging number of turbines they operate.

3) The incipient fault features are usually concealed by heavy background noise and other kinematics related harmonics in the vibration signal. These weak features are usually not detectable without advanced signal processing techniques.

4) There are situations where the health context information is not available for the labeling and selection of training data. And this will bring lots of extra efforts and uncertainties to the model training process.

5) The effectiveness and accuracy of data-driven fault detection model relies on the comprehensiveness of training data, which should cover all possible operation regimes of the monitored machine. False alarms usually occur when new operating regimes are encountered and are taken as fault condition. For wind turbine drivetrain systems, the vibration signals are collected very few times every day, and it will take tremendous time to accumulate data to train a model that is only applicable to a specific unit.

Due to these challenges, unit-specific modeling approaches have vital limitations to be feasible for wind turbine drivetrain health monitoring. The ultimate question for the existing CM systems for wind turbine drivetrain is: **How can the data-driven CM approach be self-contained and unsupervised?**

To enable a real smart and self-contained CM system for wind turbines, the following features are very important to have:
1) The system should be able to identify the operation regimes from vibration signal autonomously and accurately;

2) Be self-contained to monitor fault features and convert them to actionable information.

3) Automatically mining for weak signatures according to general applicable criteria.

4) Able to perform unsupervised selection and labeling for training data.

5) Utilize fleet data to train a fault detection model that is applicable to the whole fleet of similar turbines.

The challenges, needs, and gaps of the existing CM systems for wind turbine are summarized in Figure 1-2. To fill the technical gaps and enable smart monitoring systems for wind turbine drivetrain systems have motivated this research work on a systematic framework for unsupervised feature mining and fault detection.

![Figure 1-2 The present challenges, gaps, and goals for CM systems for wind turbine drivetrain systems](image-url)
1.2 Research Objectives, Tasks, and Methodology

This thesis presents research work that is aimed to address the aforementioned challenges by formulating a systematic framework to incorporate various novel methodologies, and enabling unsupervised feature mining and fault detection for wind turbine drivetrain systems. The objectives of this thesis are:

1) To develop a reliable and accurate instantaneous angular speed tracking method for drivetrain systems that are operating under highly dynamic regimes and harsh environment.

2) To develop robust signal processing schemes for fault feature detection and enhancement under low signal-to-noise ratio environment.

3) To formulate a data-driven fault detection approach for fleet-level health assessment of wind turbines applicable in condition of insufficient training data and no context information of health status.

4) To establish a systematic framework to integrate the feature mining and fault detection approaches to enable self-contained and unsupervised condition monitoring of wind turbine drivetrain systems.

5) To validate the proposed approaches with data obtained from real-world operating wind turbines, and instrumented full-scale test rigs.

The methodologies proposed in this research work to bridge the aforementioned technical gaps are summarized in Figure 1-3. And a brief description of each method is as follows:
1) An enhanced harmonic product spectrum (EHPC) is proposed to improve the conventional harmonic product spectrum in the following aspects: (1) more resistant to background noise; (2) adaptive to varying number of harmonic orders; and (3) autonomously determine the criteria for effective harmonic spectra;

2) Based on the enhanced harmonic product spectrum, this thesis proposed an instantaneous angular speed (IAS) estimation method. This method can effectively detect harmonic-related spectra structure (HRSS) in the vibration signal, and the gear-meshing frequencies are further determined based on the ratio relationship of the fundamental frequencies of HRSS.

3) A harmonics-targeting fast kurtogram (HTFK) is established for resonance band detection and bearing fault features enhancement. This method established a novel criterion of harmonics significance at target frequencies for resonance band detection. Therefore, it is able to customize the searching process according to the diagnostic information of bearing.

4) This thesis also proposed a fleet-based data-driven fault detection method based on clustering techniques and peer-to-fleet similarity assessment. As compared to unit-specific data-driven approach, this method can effectively detect incipient fault without context information of health status in training data.
The methodologies presented in this thesis to fill the gaps towards a smart monitoring system

<table>
<thead>
<tr>
<th>In real world we…</th>
<th>This thesis close the gaps with…</th>
<th>In ideal world we want…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Require tachometer signal to identify operation regimes</td>
<td>Unsupervised IAS tracking techniques</td>
<td>Autonomously identify operation regimes from the vibration signal</td>
</tr>
<tr>
<td>Need technicians to monitor fault indicative features</td>
<td>Autonomous scheme for feature extraction</td>
<td>Self-contained system to monitor fault features and convert them to actionable information</td>
</tr>
<tr>
<td>Have weak fault features deeply concealed by noise and other vibration sources</td>
<td>Harmonics-targeting scheme for fault enhancement</td>
<td>Automatically mining for weak signatures according to given criteria</td>
</tr>
<tr>
<td>Lose context health information to identify training data (baseline) for data-driven models</td>
<td>Identification of baseline data based on clustering techniques</td>
<td>Unsupervised selection for training data</td>
</tr>
<tr>
<td>Lack of complete operation regime data to train a model that is only applicable for the specific unit</td>
<td>Fleet-based data-driven fault detection method</td>
<td>A general model applicable to a fleet of similar turbines</td>
</tr>
</tbody>
</table>

Figure 1-3 The methodologies presented in this thesis to fill the gaps towards a smart monitoring system

The proposed methodologies have been validated with two case studies involving data collected from both real-world wind turbines and testing facilities. The first case study is on the NERL Round Robin test studies, where a full-scale wind turbine drivetrain system with 12 different failure modes is tested under various conditions. Another case study is on field data of a wind farm consisting of 48 wind turbines.

1.3 Thesis Organization

The remainder of the thesis is organized with four parts:

Chapter 2 provides a comprehensive review of wind turbine operation and maintenance issues, existing CM techniques, state-of-the-art signal processing techniques in CM of drivetrain systems, and advances of data-driven fault detection approaches.
Chapter 3 presents the theoretical background of the EHPS based IAS estimation method, HTFK for resonance band detection, weak feature enhancement, and fleet-based data-driven fault detection methodologies.

Chapter 4 applies the proposed methodologies in two case studies. The first case study is on the NREL Round Robin test studies, and is aimed to demonstrate the effectiveness of the proposed fault feature mining approaches under complicated multiple failure modes situation. While the second case study is more focused on validating the performance of the fleet-based data-driven fault detection method based on data collected from a fleet of wind turbines.

Chapter 5 is the conclusion of the thesis and the proposal for potential future research opportunities.
2.1 Operation, Maintenance, and Availability of Wind Turbines

2.1.1 The emerging trend of wind energy

In the past decades, and increased awareness of global warming and energy shortage has been raised in the world. By June 2016, a total of 192 countries have ratified the Kyoto Protocol (1997 treaty) to combat global warming. The legal binding of the Kyoto Protocol has encouraged the development and adoption of renewable energy, including wind, solar, and hydroelectricity. A new record renewable power capacity of more than 140GW was installed in 2015, and more than 40% was contributed by wind energy [7]. Wind energy is believed to be the renewable source with greatest potential to be scalable in future. According to the BP Statistical Review of World Energy, wind remains the largest source of renewable electricity at 52.2% of total renewable generation [8]. The global installed wind capacity has reached to 432.4GW in 2015, and 17% increase over the previous year [1].

As reported by the Energy Information Administration (EIA), the net production of wind energy in US has increased by 32.8% from March 2015-March 2016. The total installed capacity of wind energy in US has reached to 65.3GW, or 5.6% of total generation capacity in US by 2015 [9]. The US Department of Energy (DoE) claims that it is technically feasible to meet its goal of 20% the total energy requirements by 2030.

In Europe, wind energy has a compound annual growth rate of 10% from 2000 to 2014 [10]. The installed wind energy capacity in Europe has reached to 142GW, with 131GW on shore
and 11GW offshore. In 2015, wind energy has overtaken hydroelectricity as the third largest source of power generation in EU with 15.6% share of total power capacity, and produced 11.4% of total electricity consumption in EU [7]. The European Wind Energy Association has targeted to reach 230GW of installed wind power in Europe by the end of 2020, with 17% capacity from offshore [11].

The wind industry in China is also booming up in recent years. It has installed 30.5 GW wind energy in 2015, surpassing its own record of 21GW in 2014, and has contributed half of the global new wind capacity in 2015. The total installed capacity in China has reached to 145.1 GW, and has overtaken EU to become the global wind power leader [1].

![Global annual installed wind capacity 2000-2015](Image)

![Global cumulative installed wind capacity 2000-2015](Image)

Figure 2-1 Global installed wind capacity from 2000 to 2015 (Source: GWC 2015)
2.1.2 Cost of energy and life-cycle cost for wind energy

Levelized cost of energy (LCOE) is a metric to assess the cost of electricity generation and the plant-level impact of an electricity generation project. For wind power, the LCOE represents the sum of all costs of a fully operational wind power system over the lifetime of the project. There are five basic inputs for calculating LCOE of a wind farm project according to the formula given by [12]:

- Capital expenditures (CapEx): Initial investment for the installation of the wind farm
- Annual energy production (AEP): Annual net production of electricity in KWh.
- Operation expenditures (OpEx): Cost associated with operation, maintenance, and replacement of components of the wind farm.
- Fixed charge rate (FRC): Amount of revenue required to pay the carrying charges on the investment during the expected project life per year.
- Net Capacity Factor (CF): Ratio of its actual output to its potential output per year.

The following equation is used to calculate the LOCE for wind energy:

\[
LCOE = \frac{(\text{CapEx} \times \text{FRC}) + \text{OpEx}}{\text{AEP}}
\]

\[\text{AEP} = \text{MW}_{\text{net}} \times 8760 \times \text{CF}\]

The detailed cost breakdown for wind energy is presented in Figure 2-2. The key elements that determine the costs of wind energy include upfront investment cost for turbines, cost for turbine installation, cost of capital (discount rate), operation and maintenance cost, project development and planning cost, management and administration, and insurance. The
LCOE for wind energy varies due to different factors, including types of energy, capacity scale, countries, and more importantly onshore or offshore.

According to International Renewable Energy Agency, the average LCOE for typical onshore wind farms is between 0.06 – 0.14 USD/kWh, and 0.13-0.19 USD/kWh for offshore wind farms in Europe [10]. The O&M costs are a significant part of the overall LCOE of wind energy, and count for 20% - 25% of the total LCOE of current wind power systems [13]. The O&M costs of onshore wind farms average between 0.01 and 0.025 USD/kWh, and of offshore wind farms between 0.027 and 0.048 USD/kWh. The O&M costs for offshore wind farms are significantly higher than onshore wind farms due to logistics difficulties in marine environment.

The National Renewable Energy Laboratory launched a report in 2013 on the cost of wind energy in US. It was reported that the LCOE was 0.066 USD/kWh of onshore wind farms, and 0.215 USD/kWh of offshore wind turbines. The O&M cost for onshore wind farms was 0.015 USD/kWh, and accounted for 27% of total LCOE. And for offshore wind farms, the numbers were 0.039 USD/kWh and 19% [14].

![Cost breakdown of wind energy](source: EWEA, 2009)
Based on the data provided in the Cost of Wind Energy Review in 2010 and 2013 [14-15], a comparison of the LCOE change for wind energy is provided in Table 2-1. There have been two distinct trends in the development of wind energy: one was the increase of design capacity of wind turbines; and the other one was the decrease of installed capital cost. However, as the designed capacity of wind turbines increased, the O&M costs also increased significantly due to higher value of spare parts and greater difficulties of repair and maintenance. Hence the improvement on LCOE of wind energy was minimal due to the bottleneck of O&M practice of wind turbines.

<table>
<thead>
<tr>
<th>Year</th>
<th>2010 LCOE ($/kWh)</th>
<th>2013 LCOE ($/kWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turbine Type</td>
<td>1.5-MW onshore</td>
<td>3.6-MW offshore</td>
</tr>
<tr>
<td>Installed Capital Cost</td>
<td>0.061</td>
<td>0.194</td>
</tr>
<tr>
<td>O&amp;M Cost</td>
<td>0.10</td>
<td>0.031</td>
</tr>
<tr>
<td>Total LCOE</td>
<td>0.071</td>
<td>0.225</td>
</tr>
</tbody>
</table>

Table 2-1 Comparison of LCOE in 2010 and 2013 (Data Source: NREL 2010, 2013)

Therefore, there is a great potential for improvement on the O&M cost of wind energy to make its LCOE more competitive than other energy sources.

2.1.3 Availability of wind turbines

As indicated in Equation 1, the LCOE of wind energy can be improved from two aspects: to reduce the total O&M cost, and to increase the total capacity factor. The opportunity for
improvement on the O&M costs will be discussed in details in the later sections. The improvement of capacity factor can be achieved by improving the availability of wind turbines. Different institutions and different manufacturers define availability rates differently. The most common definition is to use the amount of energy actually produced relative to a situation where the turbine is ready to run at all times.

The availability of onshore wind turbines has been reviewed in [16]–[18]. Onshore wind farms have availability typically between 95%-99%, offshore wind farms are between 65%-97%. The average availability of 80.3% for UK Round I offshore wind farms is way short of expectations [19]. The average availability of 80% for the pilot offshore wind farms in China is also unsatisfying [20]. Onshore wind farms in China have availability between 96%-98% [21]. In a recent study of 350 offshore wind turbines throughout Europe in [5], the availability is estimated to be 98.2%. The availability of wind farms reported in US for onshore wind turbines is also around 98% [13].

As compared with conventional power plant, a concern for wind energy is the variability of power production. This issue is of great concern for grid safety, and is also the preliminary challenge for its feasibility to be scalable. A lot of research has been done to improve the energy-based availability and supply stability in the past years. The researches main focus on health prognostics, maintenance scheduling optimization, wind power prediction, control optimization, and grid dispatching optimization, etc.
2.1.4 Benefits of prognostics and health management for wind turbines

Prognostics and health management technologies are promising to increase the availability and reduce the O&M costs of wind turbines. As the capacity scale and asset value of wind turbines increase, condition monitoring systems have been widely adopted by wind farm operators. A lot of studies have been done to verify the cost benefits of condition monitoring systems (CMS) for wind turbines [22–26], and the benefits can be summarized as follows:

1) Condition monitoring systems can detect and prevent incipient failures in critical components, which otherwise can develop to major breakdown of turbines. The difference between minor repair and major replacement in downtime and costs are significant.

2) The early detection of incipient faults in wind turbines can prevent waiting time caused by weather window waiting, spare part logistics, and maintenance resource mobilization. The maintenance scheduling can be optimized before the turbines break down.

3) Condition monitoring systems are basic to enable predictive maintenance of wind turbines. And thus can reduce the redundancy of preventive maintenance and high repair costs of reactive maintenance (Figure 2-3).

4) Condition monitoring systems can help to optimize the planning for maintenance resources by predicting their demands according to real-time information of wind turbines’ health conditions.
Majority of studies suggest positive economic benefits of CMS. The simulation study performed in [24] reported that the economic benefit of condition monitoring system for the LCC of a single 3MW wind turbine is 190,000 €. And the cost of a CMS will breakeven after its first detection and prevention of a major failure on the gearbox [25].

2.2 Failure Modes and Critical Components Analysis for Wind Turbines

The maintenance cost is known to be an important part of the LCC of wind turbines and LOCE of wind energy. In a study for an offshore wind farm in the coast of Esbjerg in Denmark, the maintenance cost is estimated to count for 40% of the LCC [27]. The cost breakdown of maintenance is presented in Fig. 2-4, where the main drivers of costs are identified as corrective maintenance and serial failures.
To reduce the downtime and cost associated with machine breakdown, it is very important to understand which subsystems contribute the most impact. Criticality analysis for wind turbines need to be performed to prioritize the instrumentation of condition monitoring systems for wind turbine. Failure data of wind turbines are mainly collected through surveys of published papers in this thesis.

To begin with, the decomposition of subsystems for wind turbine is illustrated in Figure 2-5. The important subsystems of wind turbine in the order of their contribution to capital costs are tower, rotor blades, gearbox, power convertor, transformer, and generator, etc. However, from the perspective of contribution to maintenance costs, gearbox and generator comes first in the list. The criticality of each component is analyzed according to its failure rate versus the impact of failure.
The criticality analysis that is presented below is based on data from literature [5]. The data is collected from around 350 offshore wind turbines through Europe, and consists over 1768 turbine years of operation data. The turbines in this dataset are between 3-10 years of age, and are from 10 wind farms. The nominal capacities of the turbines range from 2-4MW.

The maintenance categories are grouped according to the Reliawind categories form in which failures are classified as minor repair, major repair, or major replacement. A maintenance task with total material cost of less than €1000 is regarded as minor repair, between €1000 and €10,000 a major repair, and greater than €10,000 a major replacement. Therefore, the minor
repair can be regarded as preventive maintenance, and the major replacement are corrective maintenance for major failures.

The average downtime for the 350 turbines is 102.98 hours per turbine per year, and the overall availability is 98.8%. The contributions of different maintenance categories are presented in Figure 2-6. Minor repair and major replacement has contributed similar percentage of the downtime (38% and 43%). This observation shows that intensive preventive maintenance will not necessarily lead to reduced amount of failures and corrective maintenance since faults may be initiated and propagate in between preventive maintenance actions. Gearbox and generator contributes over 56% of the total maintenance time, and over 95% of the major replacement downtime (Figure 2-7).

![Figure 2-6 Maintenance time and costs for different maintenance categories](image-url)
The cost of production loss due to downtime is also considered in addition to the material costs when calculating the maintenance costs. However, other costs such as labor, transportation, and logistics are not considered due to lack of data. The production loss is calculated by assuming a net capacity factor of 38% [14], electricity price of 0.211€/kWh (Eurostat, 2016), and average nominal capacity of 3MW. Following these assumptions, the average maintenance cost for the 350 turbines is 6,5095 euros per turbine per year. Major replacement contributes 72% of the total maintenance cost, while gearbox and generator count for 92% of the major replacement cost and 79% of the total maintenance cost (Figure 2-8).

The maintenance data are further used to make a bubble plot as illustrated in Figure 2-9. The x-axis is maintenance time of the subsystems in hours/turbine/year, the y-axis is failure rate in failures/turbine/year, and the sizes of the bubbles represents maintenance cost in €/turbine/year. The chart is divided into four quadrants, with each quadrants satisfying a type of maintenance strategy [28]. Subsystems fall into the first quadrant have both high failure rate and
downtime, which should not exist in a well-designed system. Subsystems that fall into this quadrant are suggested for redesign of product specifications or operation behaviors. Those belong to the second quadrant have high failure rate but short mean time to repair (MTTR), and corrective maintenance is preferred for these subsystems with optimized inventory management of spare parts. Most of subsystems fall into the third quadrant, where preventive maintenance and regular inspections are appropriate for their health management. The subsystems in quadrant 4 have lower failure rate, but will cause much greater impact in terms of repair time and costs, and hence are suitable for predictive maintenance strategy.

According to the Four-Quadrant chart analysis for wind turbine subsystems, gearbox and generator are appropriate for predictive maintenance. These two subsystems are also the most important components for drivetrain systems. Various sensor and monitoring systems have been instrumented around these components, and there have been intensive researches focusing on prognosis and diagnosis techniques for the drivetrain systems.

![Figure 2-9 Four-Quadrant chart for criticality analysis of subsystems](image-url)
The following conclusions are addressed from the analysis in section 2.1 and 2.2:

1) O&M costs and reliability concerns have become the bottleneck for the improvement of LOCE of wind energy.

2) Majority of the downtime and maintenance costs come from major replacement of components, which can be avoided if early detection and maintenance are applicable.

3) Drivetrain subsystems including gearbox and generator contribute the most downtime and maintenance costs. The prognosis and health management of drivetrain systems are important to improve the reliability and O&M costs for wind turbines.

4) Condition monitoring and predictive maintenance strategy is appropriate for drivetrain gearbox and generator according to the criticality analysis.

2.3 Review of Condition Monitoring Techniques for Drivetrain Systems

Due to the wide acknowledged importance of drivetrain systems, there has been a number of testing techniques applied to monitor the local parameters related to their performance and health status. The condition monitoring techniques can be classified into two main categories, namely intrusive test and non-intrusive test techniques. Their mechanisms have been reviewed in [29]–[33], and their features are summarized in Table 2-2. The techniques that belong to intrusive test category are highlighted by light grey.
<table>
<thead>
<tr>
<th>Signal Scheme</th>
<th>Applicable Components</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vibration</td>
<td>• Gearbox</td>
<td>• Reliable</td>
<td>• High cost</td>
</tr>
<tr>
<td></td>
<td>• Generators</td>
<td>• Standardized (ISO 10816)</td>
<td>• Sensor fault</td>
</tr>
<tr>
<td></td>
<td>• Bearing</td>
<td>• Direct accessibility</td>
<td>• Not applicable to low-frequency faults</td>
</tr>
<tr>
<td></td>
<td>• Shaft</td>
<td>• Early detection of failures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Nacelle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>• Gearbox</td>
<td>• Low cost</td>
<td>• Late detection</td>
</tr>
<tr>
<td></td>
<td>• Generators</td>
<td>• Reliable</td>
<td>• Influenced by other factors</td>
</tr>
<tr>
<td></td>
<td>• Bearing</td>
<td>• Standardized (IEEE 841)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Torque</td>
<td>• Rotor</td>
<td>• Direct measurement</td>
<td>• High cost</td>
</tr>
<tr>
<td></td>
<td>• Gearbox</td>
<td></td>
<td>• Influenced by other factors</td>
</tr>
<tr>
<td></td>
<td>• Blades</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oil debris</td>
<td>• Gearbox</td>
<td>• Early detection</td>
<td>• Limited to closed-loop lubrication system</td>
</tr>
<tr>
<td></td>
<td>• Bearing</td>
<td>• Accurate diagnosis</td>
<td>• Very expensive for online monitoring</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reliable RUL prediction</td>
<td></td>
</tr>
<tr>
<td>Acoustic Emission</td>
<td>• Gearbox</td>
<td>• Early detection</td>
<td>• Very high sampling rate</td>
</tr>
<tr>
<td></td>
<td>• Bearing</td>
<td>• Applicable for low frequency faults</td>
<td>• High cost</td>
</tr>
<tr>
<td>Visual inspection</td>
<td>• Gearbox</td>
<td>• Low-cost</td>
<td>• Limited to visible damages</td>
</tr>
<tr>
<td></td>
<td>• Bearing</td>
<td>• Applicable to wide categories of components</td>
<td>• Not online monitoring</td>
</tr>
<tr>
<td></td>
<td>• Generator</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cooling/lubrication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endoscope check</td>
<td>• Gearbox</td>
<td>• Reliable</td>
<td>• Time consuming</td>
</tr>
<tr>
<td></td>
<td>• Bearing</td>
<td>• Accurate information of damage level and failure mode</td>
<td>• Not online monitoring</td>
</tr>
<tr>
<td>Ultrasonic</td>
<td>• Rotor</td>
<td>• Quick and reliable</td>
<td>• Not applicable to fault diagnosis</td>
</tr>
<tr>
<td></td>
<td>• Blades</td>
<td>• Low-cost</td>
<td>• Not online monitoring</td>
</tr>
<tr>
<td></td>
<td>• Tower structure</td>
<td>• Sensitive to undersurface faults</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Less affected by other factors</td>
<td></td>
</tr>
<tr>
<td>Thermography</td>
<td>• Gearbox</td>
<td>• Applicable to wide categories of components</td>
<td>• Not sensitive</td>
</tr>
<tr>
<td></td>
<td>• Generator</td>
<td>• Fault diagnosis and localization</td>
<td>• Influenced by other factors</td>
</tr>
<tr>
<td></td>
<td>• Convertor</td>
<td></td>
<td>• Not online monitoring</td>
</tr>
<tr>
<td></td>
<td>• Shaft</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Bering</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2-2 Summarization of condition monitoring techniques
Two types of wind turbine condition monitoring techniques are widely instrumented in the commercial market. The first one is based on Supervisory Control and Data Acquisition (SCADA) systems, which is the basic informatics system for large-scale wind turbines. And the second is purpose-designed CMS with add-on sensors and additional instrumentation to wind turbines.

SCADA system is basic for wind turbine operation and safety. It incorporates a lot of sensors widely distributed to all kinds of critical subsystems. A SCADA system monitors signals and alarms in relative low frequencies to reduce the data transmission bandwidth. It provides a cheap and across-the-board solution for monitoring the status of operation efficiency and safety status of wind turbines. The signals in a SCADA that are relative to drivetrain systems include reactive power output, rotor speed, bearing and gearbox temperatures, generator winding temperatures, gearbox lubrication oil temperatures, nacelle ambient temperatures, generator current and voltage, and nacelle vibration statistics, etc. However, it cannot replace the purpose-designed CMS due to the following limitations:

1) SCADA data are collected in low frequency, which are not sensitive enough to incipient fault and transient characteristics of drivetrain systems.

2) It can provide fault location information, but very limited diagnostic features.

3) An abnormal SCADA data change usually indicate very late-stage fault status, and is not giving enough prognostic lead time for maintenance planning.

Purpose-designed CMS is usually an optional product provided by the OEMs to reinforce the condition monitoring for drivetrain systems. The most widely adopted CMS in commercial wind energy market is vibration-based CMS. There is a number of standards and guidelines developed for vibration-based condition motoring of wind turbines, include the ICE 61400-25-6,
ISO 1086-21, GL Renewables Certification 2013, and ISO 13373-2. Figure 2-10 provides a typical configuration outline for vibration-based CMSs. The following locations are usually considered for accelerometers instrumentation:

- Rotor bearing
- Gearbox input shaft bearing
- Gearbox planetary stage
- Gearbox intermediate stage
- Gearbox output high-speed shaft
- Gearbox output low-speed shaft
- Generator input shaft bearing
- Generator output shaft bearing

Figure 2-10 Configuration outline of a typical wind turbine vibration-based CMS
Vibration-based condition monitoring systems have certain advantages over other CMS techniques including direct accessibility to local components, reliable, sensitive to incipient fault, and indicative of diagnostic information. Wind turbine drivetrain systems are different from traditional rotating machines as they are operating in highly dynamic conditions and remote locations. Therefore, more advanced signal processing and feature extraction techniques are required for reliable fault detection and diagnosis of drivetrain systems.

2.4 Data Analytics Techniques for Drivetrain Vibration Signals

There have been intensive researches conducted to develop effective and robust data analytics techniques for vibration-based condition monitoring of wind turbine drivetrain systems. These researches have three main focus as follows:

- Research and development of advanced signal processing techniques for non-stationary operations, noise reduction, and incipient fault feature enhancement.
- Study of fault characteristics and indicative features for different failure modes.
- Physical-based and data-driven modeling techniques for fault prognosis and diagnosis.

2.4.1 Review on signal processing techniques

The performance of CM to a great extent depends on the quality of monitored features, while the quality of features is defined as their level of discrimination for different health status. Features extracted from vibration signals are affected by various factors, including dynamic operating regimes, harsh working environment, low signal-to-noise ratio, and lack of rotating
speed information. Therefore, a lot of efforts have been put on the research and development of advanced signal processing technologies to detect and enhance fault characteristic features in non-stationary and noisy signals. The related researches are mainly from the following three aspects: estimate instantaneous angular speed in non-stationary and tachometer-less conditions; detect and enhance weak fault characteristic features from high noise level signals; and autonomously mining the pre-defined features with minimum expert knowledge and human involvement.

Both time domain and frequency domain analysis have been adopted in practical CMSs for wind turbines. Time domain features are usually obtained from the statistical analysis of the vibration signal, such as the root-mean-square (RMS) value, kurtosis, crest factor, and peak-to-peak value. Joel Igba et al proposed a method for drivetrain fault detection based on a correlation model of rotation speed and RMS (and peak) values of vibration signals [34]. Frequency domain analysis is very important for extracting diagnostic information in vibration signals. The primary challenge for conventional FFT in drivetrain vibration signal analysis come from the extremely non-stationary and dynamic operation conditions of wind turbines. The variation of rotating speed will cause the fault related harmonics peaks smeared into multiple frequencies. Therefore, synchronous sampling analysis and time synchronous average (TSA) are usually adapted to locate and enhance the frequency components of interest [35]. Instantaneous angular speed (IAS) provides vital information for diagnosis and is essential for time synchronous analysis [36]–[38]. IAS is usually estimated from tachometer or optical encoder signals with order tracking techniques [39]. However, tachometer signal is not always available in drivetrain CMSs due to difficulties of instrumentation and cost of additional DAQ channels. When tachometer signal is not available, Siegel et al proposed a Hilbert Transform based method to generate synthesized
tachometer signal that recovers the phase information of the vibration signature to be analyzed [40]. This method, however, still needs accurate estimation of the rotational speed range to design the narrow band pass filter around the calculated bearing fault characteristic frequencies, and is only applicable to stationary speed systems. In the most recent advances of IAS tracking, Quentin Leclère et al proposed a method based on multi-order probabilistic approach that incorporates the design specifications of the drivetrain gearbox to measure the gear-meshing components in time-frequency domain. This method has been validated in a drivetrain test bed with wide fluctuating speed, and the results shows significant reduction of estimation error to less than 0.5% [41]. The limitation of this method is that it still requires the input of speed fluctuation range to generate the priori probability density function of IAS. The limitations for the state-of-art synchronous analysis techniques are summarized as follows:

- They all require prior knowledge of rotating speed for IAS tracking, which is still not qualified as unsupervised techniques.

On the other hand, de-noising and weak fault signature enhancement techniques have been deeply analyzed for bearing fault detection and diagnosis. Signal processing techniques that are developed for such purpose usually fall into two categories, i.e., to suppress the noise in a certain transform domain by thresholding; or to reconstruct the signal component of interest in other transform domain. Different types of wavelet transforms are de-noising tools that belong to the former category. It is effective for signals with local Gaussian noise, or when signal and noise are located in different scales. On the other hand, empirical mode decomposition (EMD) is an ideal tool in the latter category for processing non-stationary and nonlinear signals. The most widely used techniques in commercial CMSs for bearing fault detection is envelope analysis, which also falls into the latter category. It requires detection of the structure resonance band
where fault related features are enhanced. Band-pass filter can be applied at the resonance frequency band to improve the signal-to-noise ratio of the related harmonics of bearing fault. To achieve unsupervised searching for resonance band, J. Antoni proposed an autonomous method that uses spectral kurtosis as the indicator for bearing fault related components [42]. However, this method only target for the most impulsive excitations without considering its characters, and hence may result in extracting impulsive noise in a harsh working environment. Moreover, this method tends to extract the feature of the most severe stage of failure, while neglecting other failures in a combined failure mode condition. The challenges for incipient fault features enhancement in bearing fault detection are summarized as follows:

- The detection of optimal resonance band should be robust to non-Gaussian noise and can be adaptively customized to different fault characteristic frequencies.

### 2.4.2 Data-driven modeling techniques for drivetrain fault detection

For wind turbine drivetrain systems, a robust physical model is difficult to establish due to its structural complexity and non-stationary operating conditions. As alternatives, data-driven modeling techniques can be used to learn the behavior of drivetrain systems from its historical operation data. It has been widely used in drivetrain condition monitoring since it requires minimal prior knowledge of the system. The general process for data-driven modeling method is shown in Figure 2-11. While there are different modeling techniques available in this family, their major differences are in the model training and testing procedure.
Since this thesis will primarily deal with condition monitoring for wind turbines, the data-driven fault detection methods will just be reviewed for their applications in wind turbine condition monitoring (WTCM). These methods are categorized according to the schemes of how a model is developed and a fault metric is computed from relevant features.

2.4.2.1 Pattern recognition

Pattern recognition refers to the assignment of a label to a vector of input data according to some certain similarity assessment criteria. The model training is usually performed in supervised process, where samples of different labels are fed into the model to learn the discrimination criteria between classes. The criteria can be developed based on distance, membership functions, kennel mapping functions, and classifiers, etc.

Self-organizing Map (SOM) is a type of machine learning techniques for pattern recognition. It can aggregate the samples into different spatial regimes with kernel functions to map the topological relationship of the input data [43]. SOM is suitable for both supervised and unsupervised learning schemes. W. Zhao el al. investigated the performance for health assessment of wind turbine drivetrain systems in [44]. Edzel et al. used SOM for power curve performance monitoring in [45].
Fuzzy logic techniques perform induction rules of clustering and classification based on membership functions. The advantages of fuzzy logic algorithms are their capabilities to interpret expert knowledge and reasoning logics quantitatively when they are inexact. A wind turbine fault detection system based on SCADA using normal behavior models and fuzzy logic was presented in [46]. The combination of fuzzy logic and neural network in wind turbine fault diagnosis can be found in [47]. Another recent research study by Li et al. focused on improving the fuzzy synthetic condition assessment of a WT generator system [48].

Support vector machine is gaining popularity for its capability of using only important samples (support vectors) to clustering, classification, and regression. SVM has been extensively studied for fault detection and diagnosis, and to name only a few, clustering binary tree SVM has been applied for drivetrain fault detection in [49], and SVM classifiers have been investigated for SCADA based WTCM in [50], [51].

Different kinds of artificial neural network, e.g., back-propagation neural network (BPNN), multi-layer perceptron (MLP), and neural network ensemble (NNE), etc., have also been investigated in wind turbine fault detection. The characteristic of the model deviation can be interpreted in a way that the signal deviating from the normal behavior (model output) can be identified. Examples of their applications for drivetrain systems based on SCADA and vibration signals can be found in [52]–[55].

2.4.2.2 Statistical pattern techniques

Statistical pattern cognition (SPC) techniques are usually used for fault detection when the actual reasoning and expert knowledge is not known, but changes in one or a combination of stochastic parameters are indicating system abnormality. SPC can detect the control limit for monitored signals according to statistical significance under a given confidence interval from the
historical data. These techniques have been widely adopted in wind turbine power curve monitoring due to its robustness and sensitivity [45], [56]–[58]. Statistical test on system model residual has seen a large number of applications in fault detection. Principle component analysis (PCA) is usually adopted in multivariate statistics, and the Hotelling’s $T^2$ and square prediction error are often used to quantify the system variations. For implementations of these techniques, readers are referred to [59]–[61].

2.4.2.3 Regression techniques

Regression techniques are implemented when a mathematical model for relationship between variables is needed. The mathematical relationship between the predictors and output is usually established with data of a known good condition, so that the residual between the model output and the actual measurement can be used as the indicator of condition change. Now that the regression models require the input variables to be independent, principle component analysis (PCA) is usually performed prior to the regression process. Regression techniques are also able to identify a control limit based on the residual and confidence level of training data, so as to determine if a sample actually keep to the former relationship in fault detection process. Linear Hinges Model (LHM) was used in power curve performance monitoring in [62]. The LHM is a piece-wise linear regression model that combines the properties of nonparametric models as the noise filtering feature of an existing scatterplot smoothing algorithm. J. Igba et al. used regression techniques to model the relationship between rotor speed and gearbox temperatures and peak vibration value for incipient detection in drivetrain systems [34].
2.5 Problem Statement for Condition Monitoring of Drivetrain Systems

Vibration-based condition monitoring is recognized as the most direct and effective method for wind turbine drivetrain prognosis. The economic benefits and recent advances of condition monitoring systems for wind turbines have been analyzed comprehensively in previous sections. However, there are several unique challenges towards effective vibration-based condition monitoring to drivetrain systems:

1) Instantaneous angular speed (IAS) is essential for extracting diagnostic features and performing synchronous analysis for vibration signals. However, tachometer signal is not always available due to the installation accessibility and cost of data acquisition channels.

2) Rotating speed parameter in SCADA system is sampled in so wide time interval to provide accurate instant speed information. Thus will bring significant uncertainties in detection of fault characteristic features.

3) The operation of wind turbine is under very dynamic and unpredictable conditions (Figure 2-11 and 2-12). The speed and load variation can be significant in very short period of time. IAS tracking in such highly nonstationary condition is very difficult.

4) Existing IAS tracking from vibration signal can be classified into two families: signal harmonic tracking in time-frequency spectrum; and narrow-band filtering around one selected harmonic followed by instantaneous phase demodulation. Due to the structure complicity of drivetrain systems, it is difficult to associate the harmonics to the correct periodic kinematics. Therefore, prior knowledge of speed fluctuation range is necessary for these IAS tracking methods.
5) Wind turbines are usually operating in remote and harsh environment, and thus background noise can cause large impact on vibration signal. Signal de-noising and feature enhancement methods are required for incipient fault detection.

6) The wind turbine operating regimes are under both speed-varying and load-varying manners, thus resulting in extremely nonlinear characteristics of monitoring parameters (Figure 2-12). The conventional way of setting constant thresholds for each monitoring parameters is not able to provide optimal action decisions.

7) The existing methods for vibration signal analysis mostly rely on manual inspection of its frequency and order spectrum. The number of turbines that a technician is able to oversee is very limited. In the big data environment of wind industry, there is a need for self-contained and unsupervised smart monitoring techniques to perform fault prognosis and diagnosis in autonomous and effective ways.

8) Some data-driven prognosis methods were proposed to address the challenges mentioned above. Successful implementation of such methods can be found in [35], [55], [63]. However, data-driven modeling techniques are demanding to the quality and comprehensiveness of training data. CMSs acquire data very few times a day, and thus will take tremendous time to accumulate training data of complete operating regimes.

9) On the other hand, data-driven fault detection techniques need context information of health status of training data to identify baseline, which are sometimes not accessible due to lost track of maintenance records.
10) The baseline model trained from data of an individual turbine is only applicable to that specific turbine, and is not able to give consistent performance when utilizing to fleet of turbines.

Figure 2-12 Dynamic operation regime of wind turbine drivetrain

Figure 2-13 Nonlinear relationship of monitoring parameters
CHAPTER 3 FRAMEWORK AND TECHNIQUES FOR UNSUPERVISED FEATURE MINING AND FAULT DETECTION

3.1 IAS Estimation with Enhanced Harmonic Product Spectrum

3.1.1 Development of enhanced harmonic product spectrum

There are commonly three types of components contained in vibration signals of wind turbine gearbox.

- Time dependent periodical components produced by subsystems that are operating in stationary frequencies and irrelevant to drivetrain rotating speed.
- Phase dependent, or drivetrain kinematics dependent periodic components, with fundamental frequency in constant ratio with the gearbox input speed.
- Non-periodic components that are usually caused by background noise and random impulsive excitations.

The existence of those three types of components in vibration signals has made their frequency spectrums extremely complicated. Usually, only the harmonics that are related to gearbox kinematics are useful for IAS tracking. Those harmonics are generally associated with gear meshing, shaft unbalance, and bearing defects. The gear-meshing harmonics are often selected for shaft speed tracking due to its significance in typical gearbox vibration signals.
The purpose for enhanced harmonic product spectrum is to remove the time dependent periodic and non-periodic components in the vibration signal to make its time-frequency map less complicated, and the gearbox kinematics harmonic-related spectral structure (HRSS) more obvious. Harmonic product spectrum (HPS) is an effective tool for harmonics detection, which has been widely used for voice identification analysis [64]. This method has demonstrated advantages in detecting latent harmonics and further estimating its fundamental frequency. The mathematical expression of HPS are presented below:

Let \( x(t) \) be a real valued signal, the HPS in the frequency domain is defined as follows:

\[
H(\omega) = F(\omega) \cdot F(2\omega) \cdots F(K\omega) = \prod_{k=1}^{K} F(k\omega) \tag{3}
\]

where \( F(\omega) \) is the amplitude spectrum of the analyzed signal \( x(t) \), and \( K \) denotes the order of harmonics taken into consideration.

The definition of HPS ensures that, for a fundamental frequency \( \omega \), its HPS will be enhanced as the amplitudes of its harmonics are multiplied together. Therefore, the HRSS will be enhanced, while other components will be eliminated in HPS. A preliminary work of spectral whitening of the signal is required to remove the background noise and ensure the robustness in detecting true HRSS. The following adaptive filter is applied to the original frequency spectrum to remove local noise:

\[
F'(\omega) = \begin{cases} 
\frac{F(\omega)}{N(\omega \pm \omega_b)}, & F(\omega) > 1.96 \times N(\omega \pm \omega_b) \\
0, & \text{otherwise}
\end{cases} \tag{4}
\]

where \( N(\omega \pm \omega_b) \) is the standard deviation of local background noise of a moving window centered at \( \omega \) with width of \( \omega_b \).
Therefore, the enhanced harmonic product spectrum (EHPS) is further defined as:

\[
H'(\omega) = (F'(\omega) \cdot F'(2\omega) \cdots F'(k\omega))^{1/k} = \left( \prod_{k=1}^{K} F'(k\omega) \right)^{1/K}
\]  

(5)

The EHPS has improved the original definition of HPS in two aspects, i.e., reducing the effect of varying level of local background noise; and forcing the EHPS to a dimensionless and normalized representation to improve its universality.

The magnitude of EHPS spectra is also dependent on the order \( k \), namely the number of harmonics with the same fundamental frequency being multiplied. It is not surprising that the vibration signal may contain multiple HRSS with different number of orders of harmonics, and therefore it is not feasible to determine an optimal number of \( k \) that can fit all cases. To make the EHPS more general and unsupervised, the following scheme is proposed to calculate EHPS in an adaptive manner:
The EHPS will be zero if the fundamental frequency $\omega$ doesn’t have harmonics up to the minimum order of $k_0$, and an effective HRSS is defined by two criteria as follows:

1) The magnitude of EHPS is significant enough (greater than 1.96 times standard deviation of overall EHPS by 95% confidence interval)

2) The order $k$ is larger than a given threshold, which can be determined by elbow method that has been used in determining the number of clusters in a data set.

A demonstration of EHPS is illustrated in Figure 3-2, where a more concise and effective representation of frequency spectrum is retained to ensure the robustness of gear-meshing harmonics tracking.
Figure 3-2 Illustration of GMF identification by EHPS: (a) raw signal; (b) filtered frequency spectrum; (c) Effective HRSS by number of orders; (d) Effective HRSS and identified GMFs.

3.1.2 IAS estimation with relationship mining of HRSS fundamental frequencies

After performing the EHPS analysis to the vibration signal, the drivetrain kinematics related components are retained. Those harmonics are from gear-meshing excitations, shaft unbalance, and other shaft speed related components. Bearing fault may be another source of amplitude-modulated vibration due to the periodical contact force change at the defect spot. This amplitude modulation is difficult to be observed at incipient failure stage, unless amplified under a certain resonance frequency band.

Given the fact that there are various sources of harmonics related to shaft speed, it is still challenging to identify the correct harmonics associated with gear-meshing phenomenon.
However, there are usually multiple stages of gears in wind turbine drivetrain systems, and thus there will be several harmonics associated to different gears in the time-frequency spectrum. The fundamental frequencies of those harmonics have a fixed relationship defined by the transmission gain ratio of the gearbox. With this relationship determined from the design configurations of the gearbox, there will be a unique pair for HRSS in the EHPS that match with this particular ratio relationship.

The proposed method is validated with data collected from a fleet of real-world wind turbines, and has a total number of 8446 effective samples of vibration data when rotor speed is greater than 5 rpm. The drivetrain systems being tested have three stages of gears, and the relative transmission ratio between the second and third stage is 5.2857. For each sample of vibration data, the effective HRSS are identified in EHPS, and the pair-wise ratio relationships of their fundamental frequencies (FF) are calculated for all effective HRSS. The probability density function of the FF ratio relationships is shown in Figure 3-3 (a). The two outstanding peaks of pdf are located at 5.29 and 2.64, which are corresponding to the ratio of $\frac{GMF_{3rd}}{GMF_{2nd}}$ and $\frac{GMF_{3rd}}{2 \times GMF_{2nd}}$. This relationship is consistent and precise despite of the variations of rotating speed. Among the 8446 effective vibration signals, 67.6% of the signals have been detected to have harmonics with the given ratio relationship, and have been successfully predicted of their input speed. The distribution of predicted rotor speed is shown in Figure 3-3(b).
Figure 3-3: (a) Ratio relationship of effective FFs in the drivetrain vibration data set, and (b) distribution of estimated rotor speed.

3.1.3 Tachometer signal synthesize and synchronized averaging

Time Synchronous Average (TSA) is an essential signal processing tools for periodic response feature enhancement. During the TSA process, the raw vibration signals are resampled synchronously with respect to the rotating frequency of a given shaft, making the vibration signature of integer order to the reference shaft speed enhanced from other periodical signatures and noise. Grabill et al. used the Fourier Transform of the time synchronous averaged vibration signal to detect shaft imbalance and gear problems, where obvious sidebands around gear meshing orders were observed [65]. A review of different TSA algorithms is referred to [66], in which detailed explanation of TSA process and validation results with field data has been discussed in step-by-step manners.

Through mining the frequency relationships of the effective HRSS, it is feasible to provide a robust estimation of the shaft and gear-meshing frequencies. Furthermore, a narrow band-pass
filter followed by Hilbert Transform can be applied to the detected GMFs to generate the synthesized tachometer signal [35]. Due to the reasonably high accuracy of gear meshing frequency estimation, the bandwidth was chosen to be 1Hz plus or minus the calculated GMF. The filtered signal is further processed with Hilbert Transform to get the analytical signal. The mathematical expression of Hilbert Transform is shown in Eq. 4, which is defined as a convolution of $1/(\pi t)$ in time domain. The analytical signal has a format as Eq. 5, with $x(t)$ and $y(t)$ representing the original signal and transformed signal respectively.

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau$$  \hspace{1cm} (6)$$

$$z(t) = x(t) + jy(t)$$  \hspace{1cm} (7)$$

The corresponding envelop and phase of the analytical signal are provided in Eq. 8 and Eq. 9, and to further unwrap its phase information will generate the synthesized signal equivalent to a tachometer signal with number of pluses per revolution equal to number of teeth of the gear on the corresponding shaft. The detailed procedure for tachometer synthesize is described in Figure 3-4, and the illustration of EHPS based IAS prediction and synchronous averaging is presented in Figure 3-5.

$$a(t) = [(x(t))^2 + (y(t))^2]^{1/2}$$  \hspace{1cm} (8)$$

$$\phi(t) = \arctan \left( \frac{y(t)}{x(t)} \right)$$  \hspace{1cm} (9)$$
Family of effective HRSS fundamental frequencies (refer to procedure in Fig. 3-1)

Calculate the pair-wise ration relationship between the FFs

Searching for FF pairs that matches the gear meshing relationships

Narrow band-pass filter at the selected gear meshing frequency

Hilbert transform for filtered waveform to get the analytical signal

Get the phase information of the analytical signal and unwrap

Synthesized Tachometer Signal

Figure 3-4: Procedure for HRSS relationships mining and tachometer signal synthesizing

Figure 3-5: Illustration of EHPS based speed prediction and synchronous analysis
3.2 EHPS for Resonance Band Detection and Bearing Fault Feature Enhancement

3.2.1 Weak signature enhancement for bearing defects

Roller element bearings are critical components of wind turbine drivetrain systems. According to the gearbox failure database of NREL, which contains 289 of gearbox failure incidents, around 70% of failures, come from bearings, especially those mounted at high-speed shaft (HSS). Hence the detection of incipient bearing defects is essential for drivetrain condition monitoring and predictive maintenance.

When a rolling element interacts the defective surface, a shock is introduced and the high frequency resonance will be excited [67]. As the interaction surface periodically moves between the loading surface and unloading surface, the amplitude of the shock excitation also varies. And thus will make an amplitude modulation phenomenon on the vibration signal [68]. However, those impulses are usually very weak at the early stage of defect, and are easily masked by background noise and other vibration components.

For the purpose of weak bearing defect signature enhancement, previous research work proposed that high frequency resonance band contains the impulsive signals of bearing defect, and by band-pass filtering this band will highlight the fault related transients. There are a number of research studies dedicated in proper detection of resonance band. As discussed in [69], the structural resonances of the system can be found through modal testing or benchmarking spectrums of a known good and defected condition. Further studies to make the method cost
effective and autonomous can be found in [42], [70]–[73]. Those methods introduced different optimizing criteria for resonance band detection, including spectral kurtosis, envelope spectrum kurtosis, and signal sparsity. Although the above techniques have been proven effective in laboratory tests, challenges still exist when they are applied to the diagnosis and prognosis of bearings in industrial applications due to the following limitations:

- For bearing operating under harsh working condition, or the structure of the monitored system is complicated, the optimizing criteria proposed by previous research work may lead to detection of resonance band containing excitations of not interested sources.
- For a system that contains multiple defected bearings, those methods will only detect the worst defect or the most impulsive case.

The reason behind the limitations of previous research work on resonance band detection is that, none of them consider the relationship between the optimizing criteria and the fault characteristic frequencies.

3.2.2 A harmonic-targeting scheme for resonance band detection

To improve the unsupervised resonance band detection over the abovementioned limitations, this thesis proposed a novel method of Harmonic-targeting Fast Kurtogram (HTFK), which is based on EHPS and optimization criteria of harmonic significance at the fault characteristic frequency of interest. Instead of using spectral kurtosis as the criterion, HTFK uses the magnitude of EHPS at a target fundamental frequency as the indicator for level of dominance.
of the fault features. Therefore, the optimal frequency band is customized for detecting each specific failure mode of interest, and is more robust against random non-Gaussian noise.

The procedure of HTFK is presented in Figure 3-6 with comparison to the conventional SK based Fast Kurtogram. In order to benchmark the proposed method with Fast Kurtogram, a simulation signal was generated supposing the bearing shaft rotating frequency is 10Hz (\(f_r=10\)), and the signal consists of non-fault impulses caused by tooth impacts at 16 times the shaft speed (i.e. \(f_g=16*f_r=160\)Hz) with a constant amplitude of 1.5, a localized defect occurs on its inner-race, whose characteristic frequency (i.e. BPFI) is 82Hz, impulses produced by random external knocks on the bearing housing, and background white noise.

![Diagram](image)

**Figure 3-6: Procedure for SK based fast kurtogram and Harmonic-targeting fast kurtogram**
The results of resonance band detection from Fast Kurtogram and proposed HTFK method are shown in Figure 3-7, and the envelope spectrums of the band-pass filtered signal at detected resonance band are shown in Figure 3-8. The dominance of the gear-meshing excitation has misled the SK based Fast Kurtogram to detect resonance band containing impulses caused by tooth knock impacts, while HTFK has accurately detected the resonance frequency band of bearing fault defect.

![Figure 3-7: Benchmark of Fast Kurtogram and HTFK for resonance band detection: (a) Raw signal; (b) SK based Fast Kurtogram (color bar unit is spectral kurtosis) (c) EHPS based fast kurtogram (color bar unit is harmonic significance in EHPS at BPFI frequency).](image)
The proposed Harmonics-targeting Fast Kurtogram has advantages over the Spectral Kurtosis Fast Krutogram in the following two aspects:

1) More resistant to random excitations and background noise

2) Able to perform customized searching criteria for different diagnostic frequencies and thus can detect optimal resonance band for each individual fault features in mixed failure mode conditions.
3.3 The Framework for Unsupervised Fault Feature Mining for Wind Turbine Drivetrain Systems

The methods that have been discussed in section 3.1 and 3.2 provide basis for unsupervised fault feature mining for wind turbine drivetrain systems in the following two aspects:

1) It enables the detection of structure kinematics related harmonics for IAS tracking in unsupervised manners.

2) It makes the detection of failure related harmonics more robust against noise, and more flexible for harmonics of interest. These methods provide basic guidelines to interpret the components in the vibration signal in unsupervised ways.

The flowchart of how the previously described techniques are integrated to complete the loop of unsupervised feature mining process is provided in Figure 3-9, and a detailed step-by-step diagram illustration for the procedure is provided in Figure 3-10.
The procedures of the proposed framework are explained in details as follows:

**Step 1:** Acquire the raw vibration signals and perform data quality check.

**Step 2:** Determine the effective HRSS in the EHPS of vibration signal according to the procedure described in Figure 3-1.
Step 3: Search for fundamental frequencies pairs that match the relationship of transmission ratio for each stage. With gear meshing frequencies determined, calculate shaft speed and generate synthesized tachometer signal following steps illustrated in Figure 3-4.

Step 4: Perform synchronous sampling to transform the raw signal to order spectrum, and perform time synchronous averaging with respect to each shaft. Afterwards extract gears and shafts related fault features in order spectrum.

Step 5: Use the shaft speed information acquired from Step 3 to calculate the bearing fault characteristic frequencies based on bearing design parameters.

Step 6: For each individual characteristic frequency, perform HTFK to detect resonance frequency band (Figure 3-6), and perform envelope analysis at the filtered signal to extract bearing defect features.

Step 7: Apply the gear, shaft, and bearing defect features to prognosis and diagnosis models for health assessment.
3.4 Unsupervised Modeling Scheme for Fleet-based Data-driven Fault Detection

In addition to extracting health related features, another challenge is to convert the features into prognosis information for decision making of maintenance. The conventional approach is to set thresholds for each feature parameters to trigger alarms of different criticality. However, this approach has certain limitations from the following three aspects:
• It requires tremendous expert knowledge to set appropriate threshold for each monitoring parameters. The threshold needs to be justified either from statistical analysis of historical data, or years of operation experiences.

• As the dimension of feature vectors can reach to hundreds, it is difficult and impractical to set appropriate threshold for each individual parameters.

• Due to the nonlinear relationship between operating regime and the monitored parameters, it is hardly possible to set a constant threshold without considering their correlations.

The third limitation exists for wind turbine drivetrain system due to its extremely dynamic operation conditions. As illustrated in Figure 3-1, the simultaneous change of load and rotating speed between cut-in and rated wind speed at the power curve has made it extremely difficult to model the quantitative relationship between the health features and rotating speed.

Figure 3-11: Illustration of the nonlinear relationship between rotor speed and a monitored feature
Data-driven fault detection methods have been introduced to convert the multidimensional features into a single decision variable that represents the health condition of the monitored drivetrain [28]. The majority of data-driven fault detection approaches for wind turbines fall into two groups: unit-specific [29, 30] and fleet-based [17, 31] methods.

Unit-specific data-driven prognosis modeling techniques utilize historical data of a single turbine for model training. An abstractive representation of either the statistical pattern, or spatial distribution of features under healthy status is obtained from the modeling training process. This procedure usually requires comprehensive historical data under known statuses that covers all possible working regimes of wind turbines. In the monitoring process, the figure of merit for drivetrain health is calculated by performing an objective comparison between the testing feature vector and baseline model. Supervised learning process is feasible when data of different failure modes is available, so that the optimal threshold for fault alarm can be set.

Unit-specific data-driven fault detection model are effective when sufficient historical data and the context information of their health status are available. However, it has certain limitations when applying to wind turbine drivetrain systems:

- The CMSs acquire data very few times every day, and the control of data acquisition is not linked to the operation regime of wind turbines. Hence it will take very long time to accumulate historical data of one signal turbine that covers all its operation regimes.
- The context information of health status is not always available, in which situation training of unsupervised health prognosis model will not be feasible. Conventional method assumes that data acquired from the beginning of life is healthy, which is not always true. To validate such assumption is even more difficult when CMSs are instrumented from the middle of life.
• The unit-specific model is only applicable to the single wind turbine, and usually is not applicable to fleet level monitoring. The unit-specific prognosis model fails to take the variations among units into consideration, and thus its threshold for fault alarm will not be optimal.

Fleet-based prognosis method can better address the above-mentioned challenges by leveraging data collected from a fleet of similar wind turbines. E. Lapira proposed a framework for fleet-based fault detection for a network of similar machines based on clustering approach. It incorporates a two-step process that consists of fleet clustering and local cluster fault detection [75]. In the local cluster detection process, an instantaneous comparison approach is proposed for fault detection under non-stationary regimes. The first step of this approach is regime similarity (RS) evaluation of the operation regime parameters \( Z = \{ z_j \} \) between the testing samples and baseline samples:

\[
RS_j = d(z_j, z_k) \quad k = 1, \ldots, N_y; \quad j \neq k
\]  

(10)

Let \( y_j \) denotes the health related parameters, and \( M_j \) is the class of units that are close to unit \( j \) based on \( RS_j \). The health value of tested unit \( j \), or \( CV_j \) is calculated by its average distances with baselines of similar regimes \( (RS_j) \) :

\[
CV_j = \frac{1}{|M_j|} \sum_{k \in M_j} d(y_j, y_k)
\]  

(11)

Furthermore, the health value can be adjusted using \( RS_j \) as a weighting factor:

\[
wCV_j = \frac{1}{|M_j|} \sum_{k \in M_j} d(y_j, y_k) \times RS_{jk}
\]  

(12)
The distance function \( d(\cdot) \) can either be a simple distance metric between the health future vectors (Euclidean, correlation, cityblock), or distance between distributions (GMM, SOM).

The peer-to-peer comparison concept proposed by E. Lapira has improved the conventional unit-specific data-driven fault detection methods in that selection of baseline data is no longer restricted to health context information. However, it still has some limitations such as:

1) The peer-to-peer similarity evaluation is time consuming, and has not taken the correlation relationship between the features into consideration.
2) The weighted health value is greatly dependent on \( RS_{jk} \) and the size of \( M_j \).

Therefore, CV of samples that have different size of \( M_j \) is not in the same ground.

To address these challenges, the following fleet-based data-driven fault detection scheme is proposed based on peer-to-fleet similarity evaluation:

The basic assumption for the peer-to-fleet based fault detection is that the majority of units in the fleet are in healthy condition, so that units under faulty condition can be identified as the cluster of minorities. The first step for fleet-based data-driven health assessment is to identify outliers (failure samples) and normality (healthy samples) in the fleet data. This procedure can be
either based on clustering approach (e.g. DBSCAN, Linkage) or regression (e.g. support vector regression, principle component regression). Afterwards, the normality identified from the fleet data is taken as fleet baseline, and is used to train a global fault detection model using data-driven approaches. If the data-driven approach is based on objective comparison, the distribution of inner-cluster similarity of the global baseline data can be used to determine the control limit for alarm.

For drivetrain system, rotating speed is the most important operation regime parameters. Therefore, the normality identification process is performed for each feature against rotating speed, and samples are classified as either outliers (abnormal condition) or normality (baseline condition). Afterwards, a collection of baseline data is identified and can be further used as input for health prognosis model training. The threshold for fault detection can be determined at the 95% confidence level boundaries of the distribution of inner-cluster similarity. Since the clusters contain data from all units in the fleet and have taken into consideration the inner-cluster variations, the threshold should be generally optimal for the fleet.

The proposed fleet-based fault detection modeling approach has improved the procedure of unit-specific fault detection approach from the following two aspects:

1) It has made the baseline data selection process unsupervised and no longer limited by the context information of their health status.

2) The threshold for fault detection can be determined in unsupervised process, and is globally optimal for the fleet.
CHAPTER 4 CASE STUDIES

In this chapter, the approach developed for unsupervised feature mining and fault detection is validated with two case studies. The first case study is on the wind turbine drivetrain test data provided in the NREL Gearbox Reliability Collaborative Round Robin Test studies. The data is collected from a full-scale drivetrain test rig, where 12 different failure components exist at the same time. The second case study is on real-world data from a fleet of 48 wind turbines.

The focus of the first case study is to validate the performance of the unsupervised feature mining method under an extremely complicated multiple failure mode condition. While the second case study is focused on validation of the fleet-based data-driven fault detection method.

4.1 Case study on Round Robin Test Studies Data

4.1.1 Description of the test bed and data acquisition configurations

The Gear Reliability Collaborative Round Robin study is initiated by the National Renewable Energy Laboratory (NREL) [29]. The purposes for the study are to get a deeper understanding of wind turbine gearbox fault signatures and investigate the performance of various fault diagnosis techniques. The concurrence of twelve damages at various components and different levels of severity had made it considerably difficult to diagnosis each individual failure mode, thus making the data set valuable to benchmark different signal processing methods. The drivetrain system was tested on the NREL 2.5MW dynamometer test facility
(Figure 4-1), and had three data sets collected from tests under two constant speeds (20Hz and 30Hz) and two constant loads (25% and 50% rated loads) [76]. The complete nacelle and drivetrain was hard fixed on the floor, and a field controller was connected to the test bed to provide system start-up and safety response.

Figure 4-1: Diagram of NREL dynamometer test facility and the drivetrain system being tested [76]

Figure 4-2: Gearbox scheme and nomenclature [76]
The drivetrain system being tested had a three stage planetary gearbox as shown in Figure 4-2. It was composed of a low speed planetary stage gearbox and two parallel stages at the intermediate speed and high-speed stages. The overall transmission ratio of the gearbox was 1:81.491. The tooth number and transmission ratio at each stage of the gearbox are shown in Table 4-1.

<table>
<thead>
<tr>
<th>Gear Element</th>
<th>No. of Teeth</th>
<th>Mate Teeth</th>
<th>Transmission ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ring gear</td>
<td>99</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Planet gear</td>
<td>39</td>
<td>99</td>
<td>5.71</td>
</tr>
<tr>
<td>Sun gear</td>
<td>21</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Intermediate gear</td>
<td>82</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Intermediate pinion</td>
<td>23</td>
<td>82</td>
<td>3.57</td>
</tr>
<tr>
<td>HSS gear</td>
<td>88</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>HSS pinion</td>
<td>22</td>
<td>88</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Table 4-1 Description of gear elements and transmission ratio

The data acquisition system (DAS) was composed of 12 accelerometers, a torque sensor on the low-speed shaft, and an optical encoder on the high-speed shaft. The data was collected at 40KHz in every channel using a National Instrument PXI-4472B high speed DAS. The locations of the accelerometers are illustrated in Figure 4-3 and described in Table 4-1. The locations of the sensors were chosen according to the actual locations in typical commercial CMS systems.
Table 4-2 Sensor nomenclature and location description

<table>
<thead>
<tr>
<th>Sensor Label</th>
<th>Location Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN1</td>
<td>Main bearing radial</td>
</tr>
<tr>
<td>AN2</td>
<td>Main bearing axial</td>
</tr>
<tr>
<td>AN3</td>
<td>Ring gear radial 6 o’clock</td>
</tr>
<tr>
<td>AN4</td>
<td>Ring gear radial 12 o’clock</td>
</tr>
<tr>
<td>AN5</td>
<td>LSS radial</td>
</tr>
<tr>
<td>AN6</td>
<td>ISS radial</td>
</tr>
<tr>
<td>AN7</td>
<td>HSS radial</td>
</tr>
<tr>
<td>AN8</td>
<td>HSS upwind bearing radial</td>
</tr>
<tr>
<td>AN9</td>
<td>HSS downwind bearing radial</td>
</tr>
<tr>
<td>AN10</td>
<td>Carrier downwind radial</td>
</tr>
<tr>
<td>AN11</td>
<td>Generator drive end radial</td>
</tr>
<tr>
<td>AN12</td>
<td>Generator non-drive end axial</td>
</tr>
</tbody>
</table>

After the tests, the drivetrain gearbox was disassembled in a rebuild shop for a detailed failure analysis, where 12 damaged items in total were identified. In addition, baseline spectrum data collected from a healthy test gearbox under high speed and high load condition are also available for further benchmark studies. The data were provided to NREL partners for a blind study to investigate failure modes existed on the drivetrain being tested.
4.1.2 Instantaneous speed estimation

The raw signal of the speed sensor that was used to measure the instantaneous speed (IS) of the high-speed shaft (HSS) is a periodic wave that fluctuates around the speed control settings. The range of the fluctuation was more than 10 rpm, and would cause uncertainties for synchronous analysis and accurate feature extraction. Therefore, the proposed EHPS based IAS estimation method was used to predict the actual rotating speed of HSS.

AN7 accelerometer was used for IAS estimation since it was closest to HSS. During the IAS estimation process, a moving window with size of 0.2s and 50% overlap was used to segment the raw vibration signal. The EHPS method was applied to each segment to detect the effective harmonics in its frequency spectrum. According to the gear elements parameters in Table 4-1, the gear-meshing frequencies of the intermediate speed and high-speed stages had a fixed ratio of 3.826, and their harmonics could be identified from the EHPS. The process of gear-meshing frequency detection is illustrated in Figure 4-4.

![Figure 4-4: Intermediate steps for gear-meshing frequency identification: (a) raw signal; (b) Enhanced harmonic product spectrum; (c) Effective harmonics and identified GMF](image)
Afterwards, the identified gear-meshing frequency was used to further estimate the instantaneous angular velocity of HSS. A narrow-band pass filter with bandwidth of 1Hz was applied to the HSS pinion gear-meshing frequency, and phase information of the filtered signal was calculated with Hilbert Transform to get the synthesized tachometer signal. The zero-crossing time of synthesized tachometer signal was then calculated to estimate the IAS.

The raw speed signal and the predicted instantaneous speed are shown in Figure 4-5. The fluctuation ranges of predicted instantaneous speed is about 1 rpm, which is more stable and precise than the raw speed measurement. Furthermore, the accuracy of GMF peak detection with the predicted and measured IS are benchmarked in Figure 4-6. The GMF identified with predicted IS is right at the center of the peak, while the GMF located by measured IS is about 1 Hz away. Therefore, the predicted IS is more accurate and should be able to provide more accurate diagnostic information.

![Figure 4-5: Results of instantaneous speed prediction](image-url)
4.1.3 Analysis on gear fault related features

Three gear-related faults were detected from the post-test analysis in the rebuild shop, and are summarized as follows:

- Severe scuffing at high-speed shaft (HSS) gear set.
- Severe polishing and scuffing at intermediate speed shaft (ISS) gear and pinion.
- Moderate scuffing and corrosion at the planetary rind gears.

All the three gear related faults mentioned above haven been obviously detected from the synchronous analysis in the order spectrum. The gear meshing frequencies of HSS gear sets are very obviously modulated by both HSS and ISS orders as shown in Figure 4-7 and Figure 4-8, which indicates gear wear defect at both HSS and ISS gear sets. Moreover, the 2X GMF magnitude is five times higher than the 1X GMF amplitude, and has obvious sidebands at integer numbers of HSS order (Figure 4-9). These features suggest misalignment at HSS.
Figure 4-7: HSS GMF is obvious modulated by HS order, which suggest gear surface damage at HSS gear set.

Figure 4-8: HSS GMF is obvious modulated by IS order, which suggest gear surface damage at ISS gear set.

Figure 4-9: 2X GMF amplitude is five times higher than 1X GMF amplitude, which suggest misalignment on HS.
As shown in Figure 4-10 (a), the gear meshing frequencies of ISS gear set also have sidebands with respect to ISS shaft, and are especially obvious at the 2X GMF. Therefore, it suggests surface damage on the ISS pinion. Similarly, the 2X GMF of planetary gearbox shows relatively large sidebands up to three 3X orders of ring gear carrier frequency (0.037 order), which indicates ring gear surface damage (Figure 4-10 (b)).

![Graph showing gear meshing frequencies and sidebands](image)

Figure 4-10: (a) ISS pinion sideband with respect to ISS orders; (b) planetary gearbox sideband with respect to ring gear carrier frequency

After investigating the gear related fault features by visual check in the order spectrum, the proposed autonomous features extraction scheme is further applied to all the test data to validate its robustness and consistence. The sideband ratio features are selected as indicators for gear surface damage, and are extracted for all the test data to benchmark with the baseline condition. Up to 3X GMF and three orders of their sidebands are used to calculate the gear meshing amplitude and sideband ratio. Same features are extracted from baseline data for a fair
comparison. Sideband ratio features for HSS pinion, ISS gear, ISS pinion, and planetary ring gear from both test and baseline cases under high speed and high load condition are shown in Figure 4-11. All the four features in test condition are consistently and significantly higher than baseline condition, and therefore has validated the robustness of the proposed method to monitor gear fault related features.

Figure 4-11: (a) Gear surface damage features: (a) HSS pinion; (b) planet ring gear; (d) ISS pinion; and (d) ISS gear surface damage features

4.1.4 Analysis of bearing fault related features

The examination report of bearing fault analysis has detected five bearing related faults identified in the rebuild shop. The bearing characteristic frequencies of cage defect (FTF), roller
element defect (RED), outer race defect (BPFO), and inner race defect (BPFI), and information about detected failure modes are summarized in Table 4-3 [77], [78].

<table>
<thead>
<tr>
<th>Bearing location</th>
<th>Characteristic frequencies</th>
<th>Relative order</th>
<th>Failure mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSS downwind</td>
<td>FTF</td>
<td>0.425</td>
<td>Mild overheating on inner race.</td>
</tr>
<tr>
<td></td>
<td>RED</td>
<td>6.234</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFO</td>
<td>8.490</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFI</td>
<td>11.49</td>
<td></td>
</tr>
<tr>
<td>ISS upwind</td>
<td>FTF</td>
<td>0.145</td>
<td>Moderate level of dent, corrosion and deformation on inner race surface.</td>
</tr>
<tr>
<td></td>
<td>RED</td>
<td>1.563</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFO</td>
<td>1.793</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFI</td>
<td>2.457</td>
<td></td>
</tr>
<tr>
<td>ISS downwind</td>
<td>FTF</td>
<td>0.114</td>
<td>Severe plastic deformation and dents on outer race surface</td>
</tr>
<tr>
<td></td>
<td>RED</td>
<td>2.600</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFO</td>
<td>3.525</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFI</td>
<td>4.225</td>
<td></td>
</tr>
<tr>
<td>Planet carrier</td>
<td>FTF</td>
<td>0.007</td>
<td>Severe corrosion on carrier and outer race</td>
</tr>
<tr>
<td>upwind bearing</td>
<td>RED</td>
<td>0.199</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFO</td>
<td>0.294</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFI</td>
<td>0.332</td>
<td></td>
</tr>
<tr>
<td>LSS downwind</td>
<td>FTF</td>
<td>0.033</td>
<td>Severe abrasion at locknut</td>
</tr>
<tr>
<td>bearing</td>
<td>RED</td>
<td>1.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFO</td>
<td>1.346</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPFI</td>
<td>1.529</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-3 Bearing characteristic frequencies and failure mode information

The primary challenge for bearing fault feature analysis is to detect and enhance the weak signature of bearing defect masked in the vibration signal. The harmonic-targeting fast kurtogram (HTFK) has been analyzed in previous sections to detect the optimal resonance band and enhance bearing defect signature. Based on HTFK, the process for unsupervised bearing fault feature detection is shown in Figure 4-12.
In the bearing defect analysis for NREL Round Robin tests data, the first step is to estimate the HSS frequency based on GMF harmonics following the same process illustrated in section 4.1.3. Afterwards, the HTFK is applied to determine the optimal resonance frequency bands for the characteristic frequencies in Table 4-2. Furthermore, apply envelope analysis to the band-pass filtered data, and calculate the harmonic significance level at characteristic frequencies. The maximum order of harmonics is set to 3 when calculating the EHPS in HTFK.

The analysis results for HSS downwind bearing fault feature detection are presented in Figure 4-13. The envelope order spectrums after applying band-pass filter at the optimal resonance band to raw vibration signal have dominant peaks at up to 3X orders of BPFI frequency, and are clearly modulated by the HSS frequency. This phenomenon suggests severe inner race surface damage on the HSS downwind bearing.
Similarly, the outer race defect on ISS downwind bearing can also be obviously detected from the envelope order spectrum in Figure 4-14. The order spectrum has obvious peaks at up to 3X orders of BPFO frequency.

It is worth to mention that no partners in the Gearbox Reliability Collaborative group detected the locknut abrasion fault at the LSS downwind bearing. The final report of the Round
Robin tests indicated that this fault is not feasible to be detected by vibration signal [76]. The reason is that the cage defect characteristic frequency is very low due to the low rotating speed of LSS. Therefore, the excitation of its structural resonance is not significant, and amplitude of fault characteristic component is deeply buried in the background noise and other dominant vibration components.

However, by setting a relatively small bandwidth when searching for cage defect harmonics in HTFK, the optimal frequency band for cage defect response is identified successfully. As illustrated in Figure 4-15, the envelope order spectrum has high peaks at cage defect frequencies with obvious harmonics up to the forth order.

Figure 4-15: (a) LSS downwind bearing cage fault: (a) optimal resonance frequency band for FTF harmonics, and (b) order spectrum of envelop analysis.

4.1.5 Summary of results

Among the 12 damaged items listed in Table 4-3, the sixteen Round Robin partners has detected 7/12 of the actual damages by vibration analysis [78]. The proposed feature mining
method has detected 11/12 actual damages according to the evidence listed in Table 4-4. The three additional damages that have been detected are, LSS shaft scuffing, HSS shaft misalignment, and LSS downwind bearing locknut abrasion. The reasons for the improvement of fault detection results are from more accurate estimation of IAS and fault location in synchronous analysis, and more robust enhancement of bearing fault features.

<table>
<thead>
<tr>
<th>Component</th>
<th>Failure Mode</th>
<th>Optimal Sensor</th>
<th>Feature/Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSS gear set</td>
<td>Scuffing</td>
<td>AN7</td>
<td>Average sideband ratio of 4.94 at HSS order</td>
</tr>
<tr>
<td>ISS gear set</td>
<td>Polishing and scuffing</td>
<td>AN9</td>
<td>Average side band of 8.62 at IS pinion shaft order</td>
</tr>
<tr>
<td>Planetary ring gear</td>
<td>Scuffing and corrosion</td>
<td>AN3</td>
<td>Average side band of 3.01 at ring gear order (.037)</td>
</tr>
<tr>
<td>HSS downwind bearings</td>
<td>Mild overheating on inner race</td>
<td>AN7</td>
<td>BPFI peaks detected up to 3X order, and modulated by HSS shaft order</td>
</tr>
<tr>
<td>ISS upwind</td>
<td>Moderate level of dent, corrosion</td>
<td>AN8</td>
<td>BPFI peaks detected up to 3X order, and modulated by ISS shaft order</td>
</tr>
<tr>
<td>ISS downwind</td>
<td>Severe plastic deformation and dents on outer race surface</td>
<td>AN6</td>
<td>BPFO peaks detected up to 3X order.</td>
</tr>
<tr>
<td>Planet carrier upwind bearing</td>
<td>Severe corrosion on carrier and outer race</td>
<td>AN3</td>
<td>BPFO peaks detected.</td>
</tr>
<tr>
<td>LSS downwind bearing</td>
<td>Severe abrasion at locknut</td>
<td>AN6</td>
<td>Up to 4X harmonics of FTF (.033) can be detected on the envelop spectrum</td>
</tr>
<tr>
<td>HSS shaft</td>
<td>Misalignment</td>
<td>AN7</td>
<td>Very high peaks can be observed at every integer order of HSS, significant high peak at 7 order, and HSS pinion frequency modulated by HS frequency.</td>
</tr>
<tr>
<td>LSS Shaft</td>
<td>scuffing, unbalance</td>
<td>AN5</td>
<td>HSS GMF and ISS GMF both modulated by LS order.</td>
</tr>
</tbody>
</table>
### Table 4-4 Fault detection results and actual gearbox damages

<table>
<thead>
<tr>
<th>Component</th>
<th>Condition</th>
<th>Status</th>
<th>Actual Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun pinion thrust rings</td>
<td>Fretting corrosion</td>
<td>NA</td>
<td>Not detected</td>
</tr>
<tr>
<td>Oil transfer ring for planet carrier</td>
<td>Polishing</td>
<td>NA</td>
<td>HS GMF sidebands at planet carrier frequency</td>
</tr>
</tbody>
</table>

| Sideband ratio of 0.3 detected at LS order (.07) |

**4.2 Case Study on Field Data from Real-world Wind Turbines**

#### 4.2.1 Description of data and monitored drivetrain systems

This case study is based on field data from 48 real-world wind turbines in the same wind farm. For each wind turbine, a three stage planetary gearbox is used to connect the blade rotor to the generator, and the scheme of its configuration is shown in Figure 4-16. The first stage consists of a sun gear, planetary gears, a ring gear, and a planet carrier. The second and third stages are parallel stages with two pairs of transmission gears. The overall gain ratio is 118.4, which is the speed of generator input shaft over the speed of blade rotor shaft. The detailed information of gear parameters is listed in Table 4-5.
Figure 4-16: Scheme of monitored wind turbines drivetrain, sensor locations, and DAQ configurations

<table>
<thead>
<tr>
<th>Stage</th>
<th>Gear</th>
<th>No. Of Tooth</th>
<th>Rotating Order</th>
<th>Gear Meshing Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>Ring</td>
<td>92</td>
<td>1</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Planet</td>
<td>36</td>
<td>1 (spin)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.56 (rotate)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sun</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>Input</td>
<td>92</td>
<td>5.6</td>
<td>515.2</td>
</tr>
<tr>
<td></td>
<td>Output</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>Input</td>
<td>111</td>
<td>24.53</td>
<td>2722.8</td>
</tr>
<tr>
<td></td>
<td>Output</td>
<td>23</td>
<td>118.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-5 Gearbox elements and tooth numbers

The wind turbines have CMS systems that consist of eight accelerometers (Figure 4-16), with three different sampling rates at 2560, 7680 and 12800 Hz. All the accelerometers take the same number of samples of 8192 at each time the DAQ system is triggered. The accelerometers collect data asynchronously, and each day they are triggered to collect data for three times. The vibration data analyzed has a total span of 6 months, during which no significant downtime has
occurred. However, critical defects on two of the gearboxes are reported from maintenance record and manual inspections.

4.2.2 Instantaneous speed estimation and feature extraction

In this case study, neither tachometer signal nor reference rotating speed is provided to support the signal processing of vibration data. Even though data from the Supervised Control and Data Acquisition System (SCADA) can provide reference information of shaft speed, its sampling interval is too large (10 min) that the speed variation in between can be very significant. Moreover, the CMS and SCADA systems are powered by different processors, which have different system time and will make the data not aligned correctly.

Therefore, it is necessary to accurately predict the instantaneous speed from vibration signal in order to extract meaningful features for fault detection. The proposed rotating speed estimation and tachometer synthesizing method are performed at the beginning of signal processing steps. The raw data collected from output shaft of HSS are shown in Figure 4-17 (a), and obvious peaks of GMFs can be observed by visual check from its frequency spectrum. Unlike the test data in the first case study, these vibration data have very short sampling time, and therefore speed variation is negligible. The high-speed shaft (HSS) is selected as the reference for speed estimation, and therefore the GMFs of the third stage gears are used for tachometer synthesizing.
Figure 4-17: Illustration of tachometer synthesizing and synchronous averaging: (a) raw signal; (b) frequency spectrum; (c) enhanced harmonic product spectrum; (d) synthesized tachometer signal; (e) time synchronous averaged signal; (f) order spectrum of synchronous averaged data

Figure 4-17 illustrates the detailed intermediate results for each step of the signal processing steps described in previous sections. The GMFs of the 2\textsuperscript{nd} and 3\textsuperscript{rd} stage of the gearbox are obvious in the harmonic product spectrum, and the transmission ratio between 3\textsuperscript{rd} stage and 2\textsuperscript{nd} stage gear set is precisely at 5.286. A narrow band-pass filter is applied to the 3\textsuperscript{rd} stage GMF, and the unwrapped phase information of Hilbert Transform analytical signal is used as the synthesized tachometer (in Figure 4-17 (d)). The synthesized tachometer signal is further applied for time-synchronous average (TSA) analysis of the raw vibration signal. And the TSA signal has obvious 23 impulses representing gear-meshing knocks of one complete rotating cycle. Figure 4-
17(f) is the frequency spectrum of the averaged signal, where the gear-meshing harmonics and their sidebands are obviously enhanced. The illustrated process is then applied to all the data set to extract the following gear related features:

- RMS value of raw signal
- 1-3X shaft order amplitude
- 1-4X GMF amplitude of TSA signal.
- Sideband energy ratio of GMF up to 2X shaft order.
- RMS value of TSA residual signal.
- Kurtosis of TSA signal
- Corresponding shaft speed

4.2.3 Fleet-based data-driven fault detection method

Generally, the next step after the feature extraction process is to use selected feature set to train pattern recognition or machine learning models for fault detection. However, as there is no context information of the wind turbines’ health status, the selection of appropriate training data is a great challenge in this case. It is technically feasible to manually select baseline data from visual inspection of their frequency spectrums. However, this data set contains more than 20,000 samples of data from 48 wind turbines, and therefore to manually determine baseline condition for each wind turbine is too costly. The fleet-based data-driven fault detection method is used to address those challenges.

The process for the proposed fleet-based data-driven fault detection approach has been discussed in the previous sections, and is further explained for this specific case application in
Figure 4-18. The first step is to identify the baseline data from the collection of fleet historical data. The basic assumptions for this method are as follows:

1) Majority of the wind turbines in the fleet are in healthy and normal operation status.
2) The wind turbines exhibit statistically similar feature set under healthy condition.

For drivetrain systems, rotating speed is the most important operation regime parameter. Therefore, a clustering process is performed for each feature against rotating speed, and the feature vectors are classified as either outliers (abnormal condition) or normality (baseline condition). In this case study, DBSCAN is first implemented to remove the obvious outliers. Afterwards, the remaining data are used to fit a support vector regression model, where the center and boundary of the distributions of features at each region of rotating speed can be identified. The advantage of support vector regression is its effectiveness to model nonlinear relationship and to determine the boundary (support vector) of the residuals. The boundary to discriminate normal samples and abnormal samples can be determined as the 95% confidence level of the feature vectors distribution within the same local cluster.
With features extracted from all 48 turbines, Figure 4-19 presents the scatter plot of rotor speed versus sum of 1-3X order of HSS and sum of 1-4X HSS gear mesh magnitude. The vibration features have obvious tendency of clustering, and with obvious nonlinear relationship between local clusters distributions and rotor speed. The nonlinear relationship between features and rotor speed are fitted with support vector regression model, where the excepted value and upper limit boundary of 95% confidence interval can be determined adaptively with respect to shaft speed. The samples that fall outside of the boundary are removed, and therefore providing a collection of baseline data that represents the normal behavior of the turbine fleet.

The identified fleet baseline data are then used to train a data-driven model for fault detection. In this case study, Self-organizing Map (SOM) is investigated as a candidate data-driven method. The Self-organizing Map is a type of machine learning algorithm for pattern recognition and classification [79], [80]. When being used for health assessment under unsupervised learning situation, it can automatically recognize the spatial distribution of baseline data, and gives an abstractive representation of the cluster with \( m \) neurons. Each neuron is a weight vector that has the same dimension \( n \) as input vectors (health features) and the weight
vector is updated recursively during the training period in a competitive learning scheme. At each step \( t \), the Euclidean distances between an input vector \( (x) \) and all neurons are calculated, and the neuron closest to the input vector is named Best Matching Unit (BMU, \( w_c \)). Afterwards, the values of weight vectors are updated, so that neurons close to the input vector come further topologically closer to the input vector. The update procedure is performed in a competitive way as Eq. 13-14:

\[
\begin{align*}
    w_j &= [w_{j1}, w_{j2}, w_{j3}, ... w_{jn}]^T, j = 1, 2, ..., m, \\
    w_j(t + 1) &= w_j(t) + \alpha(t) \ast h_{j,w_c}(t) (x - w_j(t)) 
\end{align*}
\]  

(13)

(14)

where \( h_{j,w_c} \) is the Gaussian kernel function around the BMU, and \( \alpha(t) \) denotes step size, which monotonically decreases with step iterations.

During the degradation assessment process, the Euclidean distance between the input vector and its BMU, or Minimum Quantization Error (MQE), is used as the health indicator that represents the extent of similarity between present condition and baseline (Eq. 15). Therefore, the larger the MQE, the more critical the condition is. The threshold for MQE can be determined from statistical point of view. Since the kernel function in Eq. 14 is selected to be Gaussian, the distribution of MQE for training data can also be approximated as Gaussian, so that a control limit can be determined by putting 95% confidence level of variance for MQE of training data. In addition to the MQE value, the contribution \( C_i \) of each feature can be further calculated by Eq. 15 to identify the source of variation.

\[
    MQE = \sqrt{e_1^2 + e_2^2 + \cdots + e_n^2} 
\]

(15)

\[
    C_i = \frac{e_i^2}{MQE^2} 
\]

(16)
4.2.4 Fault detection results

The health assessment results for all 48 turbines are illustrated in Figure 4-20, where the color map unit is the ratio of MQE over its threshold. Hence, a color value of over one indicates an early warning of defects. There are two outstanding areas corresponding to turbine 38 and 43 that have obvious high risks. A degradation trend for turbine 38 from September to October can be observed as the color getting darker with time. However, there are no obvious defects observed from other turbines, which indicted that the false alarm rate of the proposed method is reasonably low.

Further analysis of the features contribution at a selected time when the MQE value of turbine 38 exceeds the threshold has narrowed down the source of defect to high-speed shaft (HSS). The contribution of HSS related features have a dominant contribution as shown in Figure 4-21.

Figure 4-20: Health assessment results for all 48 turbines over time
A detailed analysis in the order spectrum of the vibration data collected from turbine 38 on a suspected defect data (Sep. 10 2014) is performed in comparison with its order spectrum under normal condition. The comparison of the two order spectrums is shown in Figure 4-22. The order spectrum under defected condition has obvious higher magnitude at 1X HSS order. Moreover, its first order GMF has dominant peak, and is strongly modulated by HSS order. These observations suggest that the high-speed shaft has unbalance or misalignment at sever level. These findings also match with the maintenance record, where a major repair has been performed in end of October 2014.
Though there is no tachometer signal and context information of health status, the proposed method has demonstrated its capability to accurately extract health related features and effectively detect incipient faults with fleet-based data-driven models. The verification of suspected turbine damages has been confirmed through a visual inspection on the order spectrum.
CHAPTER 5 CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

Research on wind turbine reliability and life-cycle cost has brought the challenges of operation and maintenance to awareness. The O&M cost has become the bottleneck in the efforts to reduce cost of wind energy, and drivetrain systems have contributed the vast majority of maintenance costs and downtime. Detection and enhancement of incipient fault features in vibration signals of drivetrain system is vitally important to enable a predictive condition-based maintenance strategy for wind turbines. Lack of rotating speed information, complicated working regimes, harsh environment, and large volume of vibration data, and missing context health information has brought significant challenges to fault prognosis and diagnosis of wind turbines.

In this thesis, a systematic framework for unsupervised feature mining and fault detection for wind turbine vibration-based condition monitoring is presented. In this framework, an unsupervised instantaneous speed estimation method based on enhanced harmonic product spectrum is proposed to accurately predict instantaneous angular speed under non-stationary and high-noise conditions. Furthermore, a harmonic-targeting fast kurtogram is developed to enable unsupervised detection of optimal resonance band to enhance bearing fault features. And a fleet-based data-driven approach is introduced to enable unsupervised model training for fault detection in a fleet of similar turbines. It addresses the challenge of missing health context of training data in data-driven fault detection methods, and bridges the gap between unit-specific
and global applicable models. The above mentioned methodologies are very important to enable smart, self-contained, and unsupervised condition monitoring systems for wind turbines.

Two case studies are used to demonstrate the performance of the proposed framework. A full-scale wind turbine drivetrain test bed in a combined failure mode situation is analyzed. The proposed unsupervised fault mining techniques are able to diagnosis 11 out of 12 damaged components that are present at the same time. A fleet of 48 wind turbines from the same wind farm in Southwest China has also been analyzed. The proposed fleet-based data-driven fault detection approach successfully established a global fault detection model without using context health information as reference in model training process, and the model has detected two damaged turbines in the fleet, which are later verified with manual inspections and maintenance record. The two case studies have demonstrated the effectiveness and robustness of the proposed framework in both laboratory test environment and real-world implementation.

5.2 Future Work

The area of self-contained and smart monitoring for wind turbine drivetrain systems will become a new research spotlight in future as wind farm operators get more wind turbines and face the challenges of big data. The future goal in condition monitoring of wind turbines is to continue to minimize the efforts required from operators and maximize the number of turbines that a technician is able to oversee. Continuous efforts are still needed to develop advanced signal processing techniques for vibration analysis under non-stationary and harsh working environment. Though the proposed instantaneous speed prediction method is well adopted for drivetrain systems due to the limited rate of speed variation of wind turbines, more efforts are
needed to make it applicable to other rotating machineries where rapid speed variation is expected.

Furthermore, automation of prognosis and diagnosis systems will also be an important area for wind turbine predictive maintenance. The schemes of data-driven models need to be further improved from the aspects of computation efficiency and prediction accuracy. Fleet-based data-driven approach can be further modified to enable self-learning capability to discover causal relationships from the signals to provide more insight for fault diagnosis and root cause analysis. More research work on methodologies and analytical tools are nonetheless needed to enable self-awareness and self-comparison for health management of large scale of wind turbine fleets.
Bibliography


