Wind Turbine Gearbox Diagnostics Using Artificial Intelligence

PhD Thesis

Sofia Koukoura

Wind and Marine Energy Systems
Department of Electronic and Electrical Engineering
University of Strathclyde, Glasgow

December 17, 2019

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Abstract

To meet the latest strike price, the cost of energy from wind turbines needs to decrease. One of the biggest cost contributors to wind energy is the operation and maintenance cost. If this cost is driven down, the cost of energy of wind will substantially decrease and the reliability of the wind turbine assets need to increase. For that reason, condition monitoring systems are installed in modern wind turbines. These systems collect data and in abnormal conditions trigger alarms that are an indication of a fault. Maintenance actions can be scheduled accordingly that way, and faulty components can be replaced before catastrophic failures and large downtimes occur.

Therefore, the aim of this thesis is to utilise vibration and performance data collected from wind turbine gearboxes, in order to perform fault detection and diagnosis. The data is collected at various times prior to gearbox component failures and advanced signal processing techniques are applied to reveal fault signatures. Machine learning models are trained based on features extracted from vibration spectra and operational data separately, but also a combination of these two types of data is investigated. The output is fault detection and isolation on gearbox component level. Both unit-specific and fleet-based methods are examines. The models are trained on specific turbines, but the generalization to other turbines is also examined.

The above will provide a flexible but robust framework for the early detection of emerging wind turbine faults. This will lead to minimisation of the wind turbine downtime and increase of the wind turbines reliability and income through operational enhancement.

Contents

Contents

Acronyms

AI Artificial Intelligence.

ANN Artificial Neural Network.

CBM Condition Based Maintenance.

CM Condition Monitoring.

CMS Condition Monitoring Systems.

EMD Empirical Mode Decomposition.

MAE Mean Absolute Error.

O&M Operation and Maintenance.

OEM Original Equipment Manufacturer.

PCA Principal Component Analysis.

PHM Prognostics and Health Management.

RMSE Root Mean Squared Error.

SCADA Supervisory Control and Data Acquisition.

SK Spectral Kurtosis.

SOM Self-Organising Map.

SVM Support Vector Machine.

TSA Time Synchronous Averaging.

List of Figures

List of Figures

List of Tables

Preface/Acknowledgements

I would like to begin by gratefully acknowledging the EPSRC grant EP/G037728/1 which funded this work. I would also like to thank the Wind and Marine Energy Systems CDT management and administrative stuff, who gave me the opportunity to become part of this wonderful team.

This PhD would not have been possible without my supervisors James Carroll and Alasdair McDonald who have given me substantial guidance and support throughout my PhD. Moreover, I would like to thank Shawn Sheng and his team at NREL who have hosted me for two wonderful summers. I would also like to thank Simon Nevermann Holte, Jacob Juhl Christensen, Mikael Sonne Hansen and the whole fleet diagnostics team at Orsted, for giving me the chance to implement my thesis methods into real case study scenarios. Special gratitude to Athanasios Kolios for all his help, optimism and opportunities he gave me.

I'm always grateful for my parents Michael and Marina, who have invested in my education and always encouraged me to be the best I can be. My sister Christina has also always been a role model and I dedicate this PhD to her new wonderful family. Thanks to all the new friends I've made in Glasgow, especially the CDT students who taught me what it means to be a team member. Also thanks to old friends, especially Alex and Chris, who made moving to a new country seem like a fun adventure. Becky, Elena, Estefania and Eva have also been a vital part of my windy support group and we went through this stressful process together with lots of laughs and whatsapp messages. Last but not least, this busy last year of my PhD would have been very hard without my Yorkshire flat family, Jonathan, James and Sara, who put a smile on my face everyday. Special thanks to Jonathan for all our writing and stress relief sessions.

Chapter 0. Preface/Acknowledgements

Chapter 1

Introduction

1.1 Thesis Background

Renewable energy generation capacity must be increased if the ambitious targets of climate protection and energy security are to be achieved. The EU has committed to meeting 32% of its total energy demand from renewable sources by 2030. Wind energy harvesting has become a strong and fast growing renewable energy technology worldwide due to recent technological advances and commercial growth (Figure 1.1). For the first time renewable source generation is expected to come online below market prices and without additional subsidy on bills [9]. To ensure wind energy meets the latest strike price, ways of reducing the cost of energy cost must be investigated.

A large proportion of the total cost of energy from wind in large wind farms is

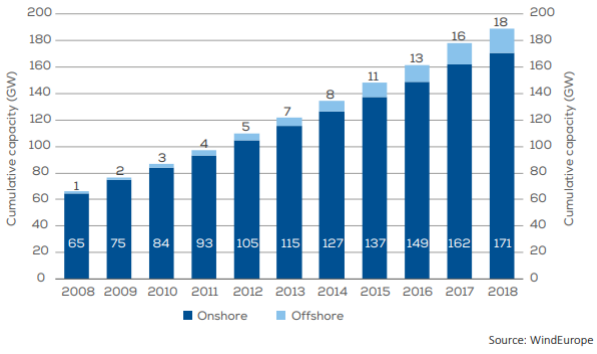


Figure 1.1: Installed wind capacity over the years

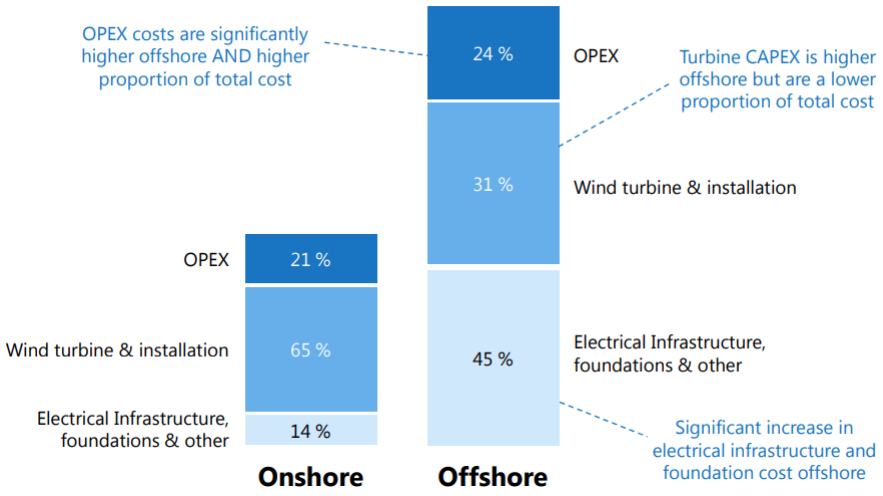


Figure 1.2: Cost breakdown wind farms. Source [1]

composed of Operation and Maintenance (O&M) costs. An indicative breakdown of the costs is shown in Figure 1.2 [1]. The purpose of maintenance actions is to achieve a desired level of performance from a component or a system. Conventional onshore O&M activities consist of a combination of preventive and corrective maintenance operations, but this approach has room for improvement [2]. In case of failure, a maintenance action is carried out which aims to restore an asset to a condition in which it can perform its intended function. This means initialising an expedition for repairs and essentially having an additional visit to the wind turbine over any planned maintenance visits. This type of maintenance practice can be impractical especially for locations that are difficult to access, such as offshore wind farms. Also, faults require a wide range of responses; it could either be from an inspection to a manual restart of a wind turbine which could take several days or even weeks. Therefore, unexpected failures of components directly translates to wind turbine downtime, loss of reliability and revenue reduction.

O&M costs and wind turbine downtime can be reduced if incipient machinery faults are successfully detected before they become catastrophic failures. This can be achieved through Condition Based Maintenance (CBM) preventative practice which monitors the actual condition of the wind turbine components. A CBM strategy constitutes maintenance tasks being carried out in response to the deterioration in the condition or performance of an asset or component as indicated by a Condition Monitoring (CM)

process [10]. In wind turbines, this process can be through Supervisory Control and Data Acquisition (SCADA) systems and Condition Monitoring Systems (CMS).

In recent years, condition monitoring techniques have been incorporated in wind turbines and their efficiency has been improved. The main condition monitoring techniques along with the signal processing methods used for diagnosis and their applications in wind power are presented in [11], [12], [13]. The most popular technologies are vibration analysis and oil analysis, while other techniques include thermography, accoustic monitoring and electrical signals.

Wind turbine SCADA system collects information extensively from key wind turbine subassemblies. The data are low frequency (usually 10 min average data) recorded from a range of instrumentation (anemometers, thermocouples vibration transducers etc). The challenges with SCADA data is that the values of SCADA data (e.g. vibration and temperature) vary over wide ranges under varying operational conditions, making it hard to detect incipient faults [14]. Also the large volume of data poses a barrier in the accurate interpretation. Therefore, advanced data mining techniques and physics of failure models are used in order to improve the reliability of SCADA systems. Examples can be found in [15], [16].

Wind turbines also have CMSs installed, with their state of the art review found in [17]. According to this survey, the majority of CMSs include high frequency vibration monitoring of the drive-train. In recent years, CMSs have been increasingly installed in wind farms, but their application is not yet completely understood as well as in other industries. In contrast to other types of machines, wind turbines operate under variable speed conditions and are subject to aerodynamic loads with stochastic characteristics. Due to the non stationary signals, traditional frequency domain signal processing techniques and consequently prognostic algorithms are not easily implemented [12]. A large amount of the available CMSs in wind power systems rely on the experience of operators. This can become impractical in large wind farms. Thus, there is definitely benefit from automating the diagnostic and prognostic process and incorporating data mining techniques.

Reliability is an important factor regarding wind farm economic sustainability. Ac-

cording to initial failure statistics datasets for onshore wind turbines [18], [19], [20], electrical and electronic subassemblies fail more frequently than mechanical ones, but the mechanical subassemblies experience longer downtimes. It should be noted that these datasets where collected during 1993-2006 and most wind turbines examined have lower power rating than modern ones. The findings of this survey are comparable to the most recent study which also showed that that onshore downtime is dominated by gearbox and generator failures [21]. When it comes to offshore, the largest downtime reason is a gearbox failure and a typical downtime period can be 41 days [22]. The offshore downtime is usually larger due to lead time for needed equipment (such as crane vessels), logistics as well as wind and wave conditions. Data from 350 offshore wind turbines collected over a 5 year period shows that the gearbox is a component with one of the highest downtime, demand for required technicians and by far the highest repair cost [3], as shown in Figure 1.4. A similar reliability analysis [23] of 3 years of available data from an offshore wind farm of 36 3MW wind turbines suggests that the gearbox has the highest downtime. A comprehensive reliability review was carried out in [2], where all wind turbine failure statistics datasets are compared and analysed, to give meaningful conclusions for all types of wind turbines. According to the findings of this study, the gearbox shows the highest downtime, up to 56%, as shown in Figure 1.3.

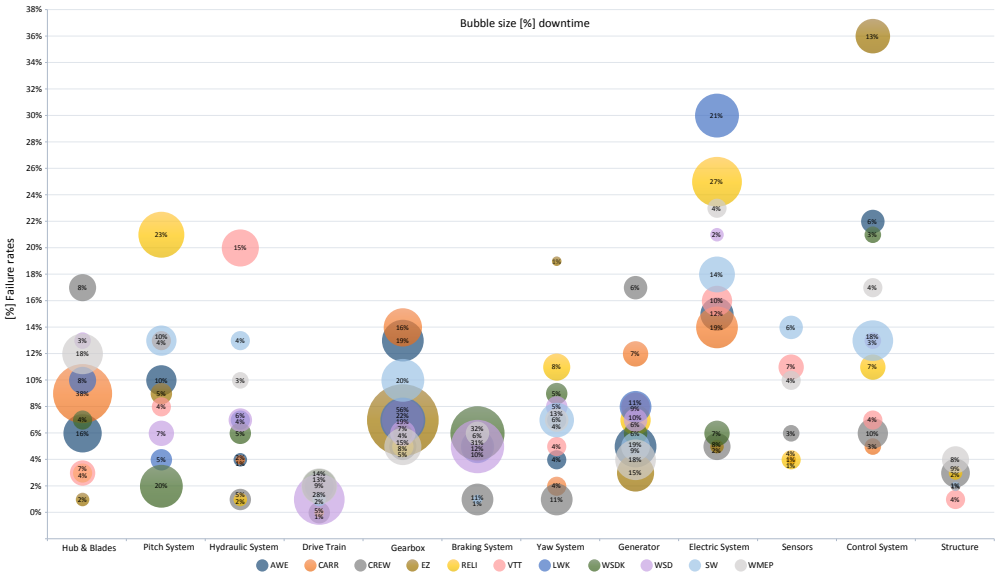


Figure 1.3: Failure rates against downtime for wind turbine assemblies. [2]

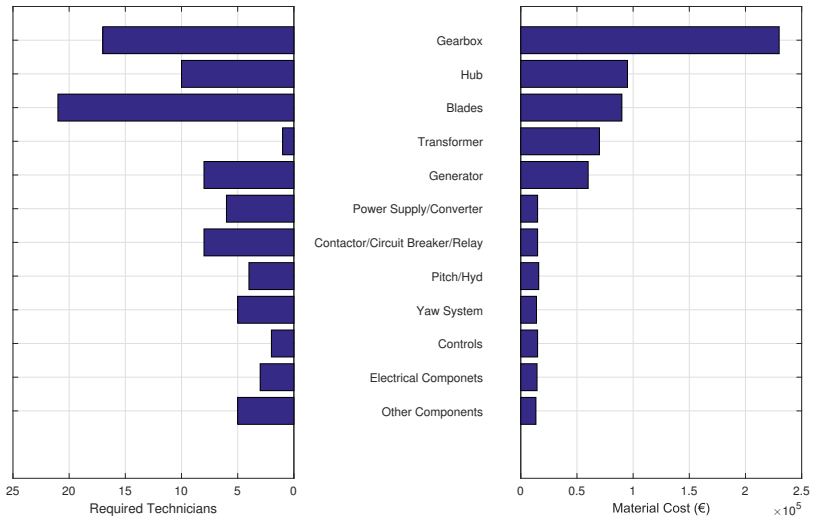


Figure 1.4: Pareto chart of average repair cost and average number of technicians for each sub-assembly/component, regarding major replacements in offshore wind. Adapted from [3].

It is therefore imperative that condition monitoring strategies for wind turbine gearboxes are improved. As the installed wind capacity grows, so does the volume of condition monitoring data. In the era of Big Data and Internet of Things, there

are plenty of Artificial Intelligence (AI) techniques that can be utilised by the wind industry to improve asset management.

1.2 Research Question

Based on the problem statement that the repair cost and downtime of the wind turbine gearbox is too high, the objective of this thesis is to answer the following research question:

"How can incipient wind turbine gearbox faults be detected based on a combination of condition monitoring data and the use of AI before catastrophic failure occurs?"

To answer the primary research question a number of other smaller secondary research questions must first be answered. These secondary research questions are answered throughout the chapters of this thesis. The beginning of each chapter will set out the secondary research question to be answered in that chapter. The conclusion of each chapter will answer the secondary research questions.

1.2.1 Approach Taken

To answer the primary research question outlined in the previous Section 1.2, the steps as shown in Figure 1.5 must be taken.

The methodology is component failure specific. Initially, a specific turbine type and component failure are chosen. The next step is data-processing, on data that could potentially be related to this failure. In this context, operational SCADA and high frequency vibration data from the gearbox are analysed. A signal processing step is often necessary in order to remove noise and further reveal fault signatures. Afterwards, feature extraction of health indicators progressively before failure is performed. These indicators are used as inputs in machine learning models that learn the behaviour characteristics of failure trends. The output and final step of these models is anomaly detection and diagnosis.

The novelty in this research mainly lies in the framework provided, but more detail is given in Section 1.3. Each of the steps are analysed in the chapters throughout the

thesis and the respective novelties are highlighted.

The programming language used is MATLAB 2019a, with the signal processing and statistics and machine learning toolboxes. For the data extraction process, SQL and Python were used.

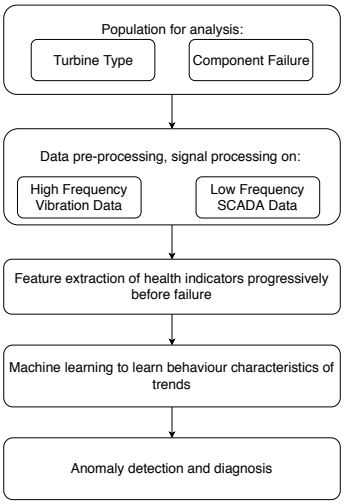


Figure 1.5: PhD methodology approach.

1.2.2 Definitions

Some definitions that are used throughout the thesis are further clarified. "Fault detection" is used to describe the detection of an incipient fault (fault that has already started developing but is not catastrophic yet and therefore the wind turbine continues to operate). This detection is performed by using sensor readings. "Failure prediction" is also used in this thesis to describe early fault detection, before it becomes a catastrophic failure and the system (wind turbine gearbox) ceases to perform its required action. It should not be confused with prediction of future states. "Diagnostics" goes beyond detection and provides insight into possible causes for the problem, aiding in course of action decision making.

1.2.3 Structure

Each of the chapters consists of a small introduction to the work carried out in the chapter, along with the methodology followed, the results obtained and conclusions drawn. There is a separate chapter for the literature review of the whole thesis but there are also some research gaps and literature elaborated specifically in each chapter. The end of each chapter will have a conclusion sub-section but there is also an overall conclusion at the end of the thesis. The references are all provide at the end of the thesis.

1.3 Novelty of Research

From the initial literature review carried out for defining this work, it was determined that this research would be novel because little or no past work was carried out (or at least published on) in the following areas:

- Failure rates and reliability of offshore wind turbine drivetrain components
- Failure rates of modern multi MW both geared and direct drive wind turbines
- Study on the fault development of sufficient real gearbox wind turbine data progressively before a real failure occurred
- Diagnostics of operating double planetary stage wind turbine gearboxes
- Comparison of the state-of-the-art signal processing techniques for wind turbine gearbox signals on the same failure examples
- Development of a full pattern recognition diagnostic framework for wind turbine gearboxes using machine learning techniques
- Case studies on generalization from one wind turbine machine learning model to another

1.4 Research Output

Peer reviewed journals and conferences published or submitted:

1. **A comparison of wind turbine gearbox vibration analysis algorithms based on feature extraction and classification** Sofia Koukoura, James Carroll, Alasdair McDonald, Stephan Weiss, IET Renewable Power Generation, DOI: 10.1049/iet-rpg.2018.5313

2. **Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques.** James Carroll, Sofia Koukoura, Alasdair McDonald, Anastasis Charalambous, Stephan Weiss, Stephen McArthur, Wind Energy 11/2018;, DOI:10.1002/we.2290

3. **Wind turbine gearbox planet bearing failure prediction using vibration data.** Sofia Koukoura, James Carroll, Alasdair McDonald, Stephan Weiss, Journal of Physics Conference Series 10/2018; 1104:012016., DOI:10.1088/1742-6596/1104/1/012016

4. **Wind turbine gearbox temperature monitoring using SCADA data** Sofia Koukoura, Lindy Williams, James Carroll, Alasdair McDonald, Applied Energy (Submitted)

5. **Prediction of wind turbine generator failure using 2-stage cluster-classification methodology** Alan Turnbull, James Carroll, Alasdair McDonald, Sofia Koukoura, Wind Energy (Published)

6. **Wind turbine intelligent gear fault identification.** Sofia Koukoura, James Carroll, Alasdair McDonald, Annual Conference of the Prognostics and Health Management Society, St Petersburg, Florida; 10/2017

7. **Wind turbine gearbox vibration signal signature and fault development through time.** Sofia Koukoura, James Carroll, Alasdair McDonald, Stephan Weiss, 25th European Signal Processing Conference, EUSIPCO 2017. N.J.: IEEE, p. 1380-1384 5 p.

8. **An insight into planet bearing fault prediction using SCADA Data.** Sofia Koukoura, James Carroll, Alasdair McDonald, European Conference of the Prognostics and Health Management Society, Utrecht, Netherlands; 07/2018

9. **On the use of AI based vibration condition monitoring of wind turbine gearboxes**
   Sofia Koukoura, James Carroll, Alasdair McDonald, Journal of Physics Conference Series (Accepted)

10. **The effect of operational parameters on vibration signals of wind turbine gearboxes** Sofia Koukoura, Eric Bechoefer, James Carroll, Alasdair McDonald, Proceedings of the 38th International Conference on Ocean, Offshore Arctic Engineering

11. **A diagnostic framework for wind turbine gearboxes using machine learning.** Sofia Koukoura, James Carroll, Alasdair McDonald, Annual Conference of the Prognostics and Health Management Society, Scottsdale, Arizona; 09/2019

Invited Presentations:

1. **Diagnostics and prognostics of wind turbine gearboxes using high frequency vibration data**, Sofia Koukoura, ETP Annual Conference 2017

2. **Failure and RUL prediction of wind turbine gearboxes using condition monitoring data and machine learning** Sofia Koukoura, NREL Drivetrain Reliability Collaborative 2018

3. **Working on SCADA data** Sofia Koukoura, Elena Gonzalez, EAWE PhD Seminar 2018, Vrije Universiteit Brussel

4. **On the use of SCADA and vibration data for wind turbine gearbox failure prediction**, Sofia Koukoura, ETP Annual Conference 2018

5. **Using high frequency currents and instantaneous speed signals for**

Wind turbine drivetrain diagnostics Sofia Koukoura, NREL Drivetrain Reliability Collaborative 2019

6. **Wind turbine drivetrain monitoring using fleet vibration data** Sofia Koukoura, Wind Energy Science Conference, Cork, Ireland 2019

Chapter 2

Literature Review

2.1 Wind Turbine Drivetrains

Utility-scale wind turbines are designed to operate in remote locations, where strong winds are available. The turbine nacelle is placed on top of a tower which is normally more than 60 meters in height. The drivetrain of a wind turbine is a set of components that transmits the kinetic energy captured by the rotor blades to the electric generator. The configuration of this set of components is closely linked to the type of electrical generator and control system, so it is common to include in the description of the drivetrain arrangement, the information about the electrical generator (i.e., synchronous or asynchronous) and converter (partial or total power) [4]. Regarding drivetrain configuration, in 2017, geared wind turbines with DFIGs (Doubly Fed Induction Generators) continue to be the preferred solution in the global market [5]. This configuration requires a gearbox between the electric generator and the wind turbine rotor in order to adapt the slowly rotating high torque power from the wind turbine rotor to high speed low torque power of the asynchronous generator. That way the mechanical energy from the blades is converted into electrical energy in the generator. In such a configuration, some of the loads transmitted from the rotor to the generator will go through the gearbox [24]. Nevertheless, this arrangement is increasingly losing share in favour of arrangements with full-power converters (both direct drive and hybrid arrangements), as nominal power of new wind turbines increases. This trend is especially prominent

in the European market.

An alternative arrangement with increasing market impact is the direct drive configuration, which consists of a synchronous electric generator directly coupled to the rotor blades and a full power converter to adapt the frequency to the typical operating conditions of the main electric grid (i.e., 50 or 60 Hz).

References [25], [26] have classified wind turbine drivetrain configurations in six types;

- Type A (Type 1). Fixed-speed generator. No power converter or other speed regulation techniques are used in this configuration. The operating speed of an asynchronous electric generator (i.e., high-speed generator, which in wind technology operates usually at around 1500 rpm) is constrained by the rotational speed of the rotor blades (according to a certain multiplication factor fixed by the gearbox ratio and the number of poles of the generator).

- Type B (Type 2). The speed of the asynchronous generator is controlled by a variable resistance that enables modifying the circulating current in the rotor of the electrical generator. This solution provides higher control flexibility than Type A. However, the electrical losses are relatively high and the response to grid requirements is limited.

- Type C (Type 3). This configuration is known as doubly-fed induction generator (DFIG). The current in the electric generators rotor is controlled by a power converter. Thus, electrical losses are lower and the response to grid requirements is enhanced. Since the power converter is only connected to the rotor of the generator, the rated power of the converter is around 30% of the rated power of the wind turbine.

- Type D (Type 4). A full power converter enables the decoupling of the generator from the grid frequency so that the frequency on the generator side can be fully controlled and the use of a gearbox can be avoided. Additionally, the full power converter provides enhanced grid services. A synchronous electrical generator (which can be either an electrically excited synchronous generator (EESG) or

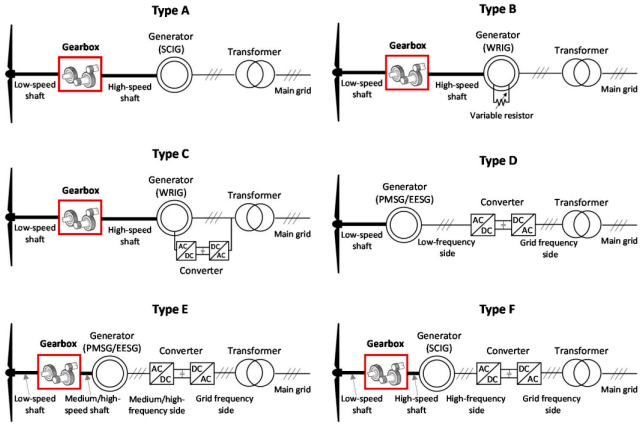


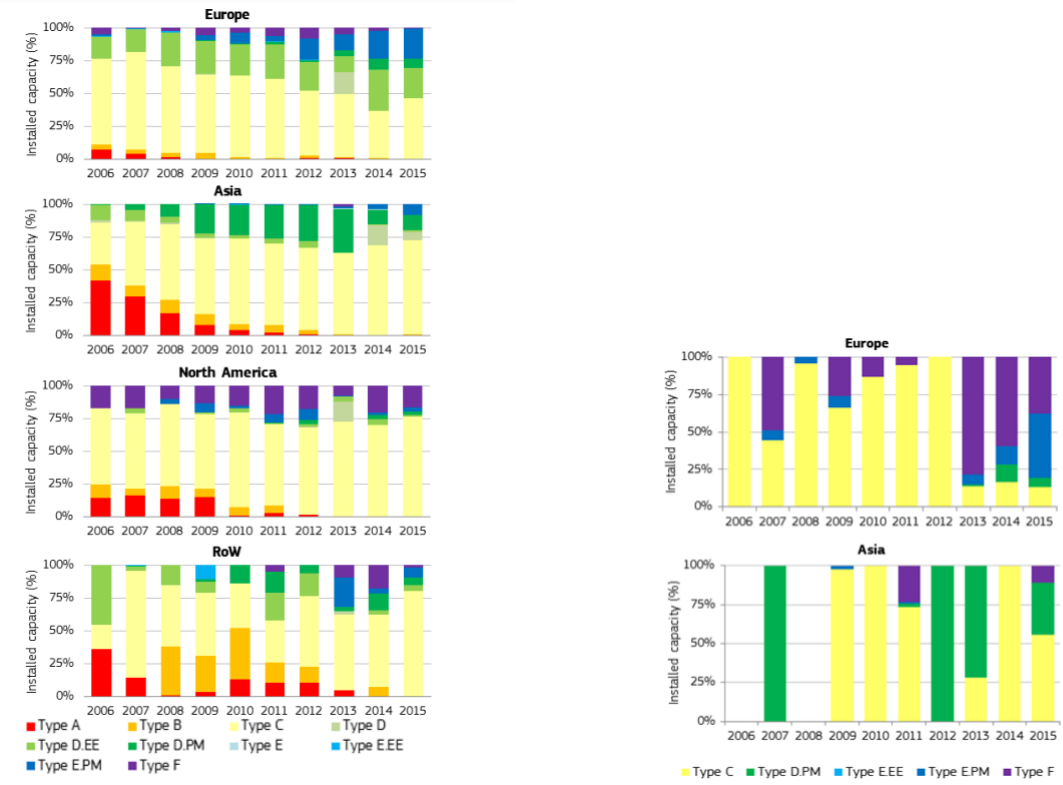
Figure 2.1: Wind turbine drivetrain configurations [4]

a permanent magnet synchronous generator (PMSG)) is directly coupled to the main shaft of the rotor (which operates at a rotational speed around 540 rpm, depending on the wind turbine size).

- Type E (Type 5). Gearbox-equipped wind turbine with a full power converter and medium/high-speed synchronous generator, which can be EESG or PMSG. This configuration enables reducing the size of the generator (compared to Type D), while maintaining the advantage of using a total power converter (with greater grid capabilities). In this arrangement, it is possible to choose between a relatively small gearbox (with moderate gear ratios) at the expense of using a large medium-speed (about 500 rpm) synchronous generator. On the other hand, it is possible to assemble a gearbox with a higher gear ratio in order to reduce the size of the generator, being a high-speed configuration with synchronous generator.

- Type F (Type 6). Gearbox-equipped wind turbine with a full power converter and high-speed asynchronous generator. As the full power converter enables the speed to be controlled by modifying the operating frequency, a squirrel cage induction generator (SCIG) is generally employed in this configuration.

In summary, types A, B and C correspond to geared high-speed wind turbines, type D is direct drive configuration and types E and F represent hybrid arrangements.



(a) Onshore

(b) Offshore

Figure 2.2: Evolution of the share of installed capacity by drive train configuration in onshore and offshore wind turbines by geographical zone Source: JRC Wind Energy Database [5]

The evolution of both onshore and offshore markets in terms of drivetrain types are shown in Figures 2.2a, 2.2b.

The onshore wind market is mainly dominated by type C configuration and to a lesser extent type D, especially in Europe and Asia . Hybrid arrangements have progressively gained ground in the last years although in a different way among geographical zones. We can observe more type E configurations in Europe and type F configurations in North America. Configurations of types A and B have steadily decreased and they currently represent a marginal market share. Most of type D and type E configurations in the Asian market use PMSGs while EESGs are more common in Europe.

The offshore wind market has evolved from a dominant type C configuration (geared high-speed DFIG) towards both direct drive (type D) and hybrid arrangements (types

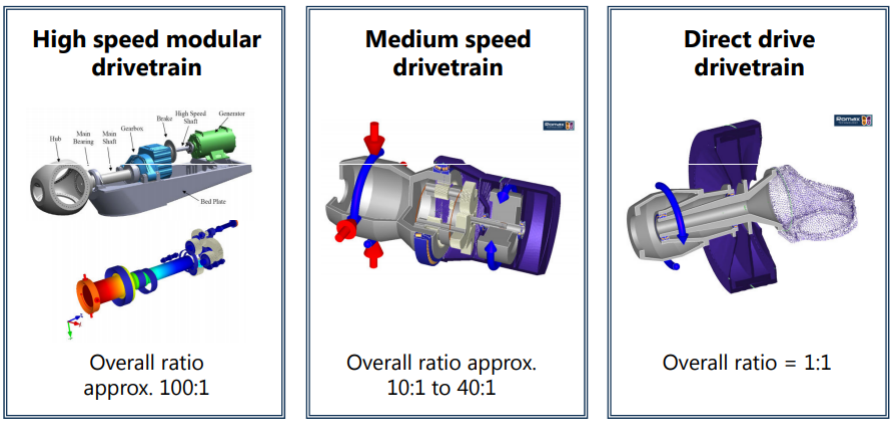


Figure 2.3: Wind turbine drivetrain configurations [1]

E and F)

In terms of ratio, a simple drivetrain classification found in wind industry is shown in Figure 2.3. Some modern multi-MW wind turbines, especially offshore, have eliminated the gearbox to reduce gearbox downtime and increase reliability. However, this has come with the expense of higher electrical sub-assembly failures. This type of drivetrain configuration hasn't been operational for as many years as the geared one, therefore industry experience is still immature.

A gearbox consists of several (usually three) planetary or helical stages. A wind turbine gearbox typically contains one low-speed planetary stage with two parallel stages or two planetary stages with one highs-peed parallel stage. In general, a planetary stage is designed for gear ratio up to seven and a parallel stage usually up to five [27].

Gearboxes are seen as the least reliable part of high-speed wind turbine configurations, although most often it is the third stage (the high speed one) that is problematic. Several studies, including the European project Reliawind, concluded that the electrical systems (including the power converter) and the pitch system cause more failures. It was also found that electrical system failures are not necessarily the highest cost, nor do they cause more downtime that any other turbine sub-assembly [28]. In addition, research shows that gearbox failures are most often due to unexpected loads originating somewhere else, e.g. in the turbine rotor or in its control system as a consequence of forcing the generator to maintain grid frequency [29]. More detailed data, models and

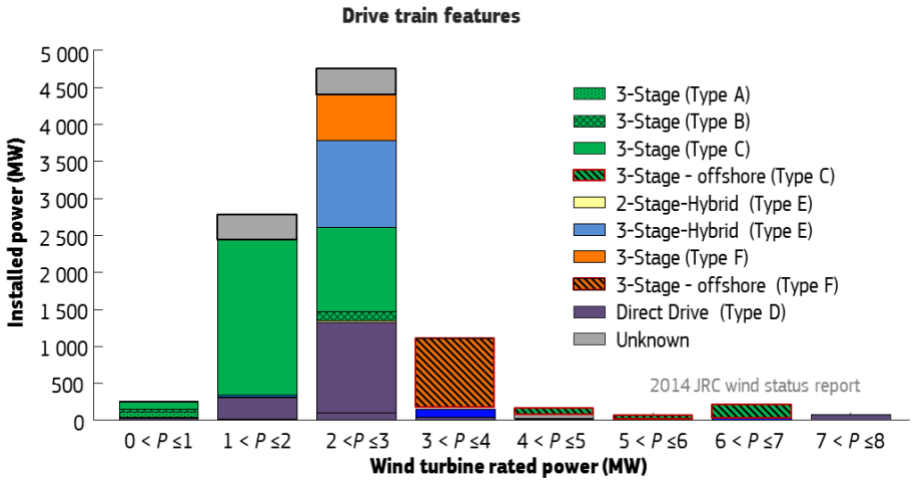


Figure 2.4: Drive train features classified according to wind turbine rated power. Data corresponding to installations in EU MS during 2012. Source: JRC Database [5]

experience are needed to close the loop between design and operation reliability.

Figure 2.4 shows the type of drive train employed in wind turbines installed in European Union Member States during 2012, classified according to wind turbine rated power. Onshore installations were mainly dominated by wind turbines in the range 1-3 MW, whereas turbines with higher rated power (in the range 3-7 MW) are employed for offshore installations as well. As regards drive train configuration, a 3-stage gearbox coupled with an asynchronous generator (Type C) is the most common arrangement for wind turbines under 3 MW. However, Type D (either direct drive or hybrid configuration) is the prevailing arrangement for wind turbines specifically in the 2-3 MW range. Looking at the offshore market, most wind turbines installed during 2012 were Type F with a three-stage gearbox and a full converter, the Siemens machines. Nevertheless, this scenario is expected to change as most wind turbines addressed to the offshore market introduced in recent years are based on permanent magnets either direct drive (Type D) or hybrid drive train (Type E). Incidentally, Figure 2.4 shows a clear segmentation in terms of rated power of wind turbines aimed at the onshore and offshore market.

As discussed in this section, a common gearbox choice in wind turbines is a planetary gearbox, due to its large power transmission capacity and compact design. Owing to

harsh operating conditions of wind turbines, gearboxes are prone to damage. Also, due to the complexity of the gearbox, fault identification and isolation is a complex task and the information provided by the condition monitoring systems is challenging to analyse. It is therefore vital to develop models for effective planetary gearbox fault diagnosis and this is going to be addressed throughout the thesis.

2.2 Reliability of Wind Turbines

Reliability has always been a vital aspect in the assessment of industrial products. One can see reliability theory as a tool for analyzing and improving the availability of the system. In the study of reliability, Weibull distribution and Bathtub-Curve are well known methods which are concisely describing failure rates. The failure rate of a component is often high in the initial phase of its lifetime. This can be explained by the fact that there may be undiscovered defects in the components. When the component has survived the initial period, the failure rate stabilizes at a level where it remains for a certain time until it starts to increase again as the component begin to wear out. There are other types of curves as well but the bathtub curve is a good choice for mechanical components such as gearboxes [30]. According to the bathtub curve, monitoring of machinery can be important in all stages of its lifetime [31]. Reliability studies can be aided by information on mean time between failures available from surveys, including [32]. Information about the life of electrical machines has also been published from experience in the wind industry [33].

In terms of wind energy, reliability describes the ability of a wind turbine to function as it should over its design life. A turbine is reliable if it is available to produce energy whenever the wind speed is higher than the wind turbines cut in speed throughout its design life. If a wind turbine experiences a failure within its design life, the turbine reliability is reduced. Reliability studies can focus on wind farm level or wind turbine level. A wind turbine is composed of many subsystems that include electrical, mechanical, structural and software. The procedure to determine the wind turbine reliability design can be performed on both the overall system level as well as on sub-system levels. Some primary studies or reliability analysis of wind turbines can be found in [34], [35]. Most

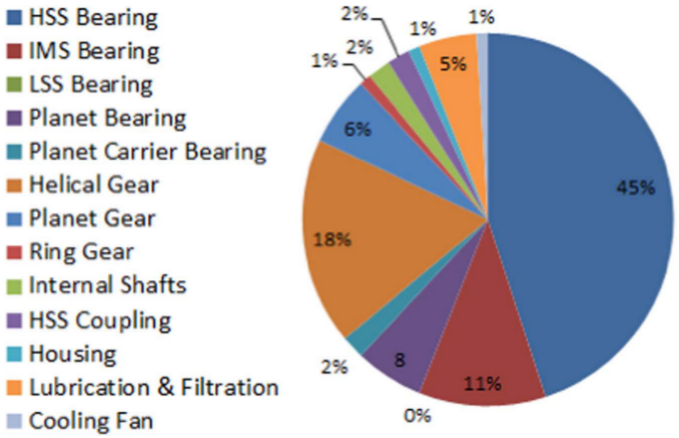


Figure 2.5: Gearbox failures based on 257 damage records released in 2014 [6]

wind turbine reliability models are based on either the homogeneous Poisson process or power law process. A recent review of reliability analysis methods in wind energy is given in [36] and a focus on risk analysis is presented in [37].

According to the latest reliability reviews, the gearbox is a major contributor to wind turbine downtime [2]. There exist a few gearbox failure databases which indicate that different gearboxes can fail in different ways, but bearings seem to be the largest contributor [6], [38] (Figure 2.5). Many factors, such as friction, lead to increased temperature. Further, mechanical stresses cause shaft cracks, tooth breakage, shattering, and in the worst-case scenario, damage to the tower (collateral damage). Major gearbox failures require a crane for handling, replacement, and greasing. Therefore, the cost is increased because of costs associated with crane rental and labor, as well as economic losses that are incurred. The gearbox has to work in random loading conditions and the reliability of its components is compromised compared to lab tests, were loading conditions are more controlled. Gearbox reliability has a direct effect on future liability for the Original Equipment Manufacturer (OEM) (warranty costs) and for the operator (O&M costs).

There has been some reliability literature focusing on individual components [39], [40], [41]. Reference [39] focuses on gearboxes and the failure rates of 3 stage, 2 stage and individual components of gearboxes. These failure rates are based on an expert failure mode and effects analysis (FMEA) rather than empirical data. The results

of this FMEA provide a failure rate for different gearbox types in a per turbine per year format. Empirical data from operating turbines show gearbox failure rates are low in comparison to other wind turbine subsystem failure rates, however this lower failure rate doesnt transfer to higher availability because of the high downtime related to gearboxes.

Reliability of any design is therefore an important feature and this can be established by gathering information from past faults that occurred in the design and then formulating novel strategies and techniques to minimize these shortcomings. By doing so, the availability and robustness of that design can be enhanced. One method of achieving this level is to implement efficient maintenance strategies. These will be analysed in the following sections.

2.3 Condition Monitoring

2.3.1 Condition Monitoring in Machinery

The main two categories of maintenance are either corrective or preventive maintenance. Corrective maintenance is a run-to-failure approach and is only performed once a component fails completely, whereas preventive maintenance is performed before the occurrence of a potential failure [42]. Preventive maintenance can be further classified into scheduled maintenance and condition-based maintenance. Scheduled maintenance refers to maintenance that happens at a fixed frequency whereas condition-based maintenance involves continuous health monitoring of wind turbine components. This type of maintenance utilises an estimation of the current and future condition of a component in order to provide an optimised maintenance scheduling that prevents failures without resorting to over-maintenance [43]. This optimised maintenance scheduling offers extended machine lifetime, as well as reduced maintenance costs and downtime.

According to [44], "CBM is a form of preventive maintenance which includes assessment of physical conditions, analysis and the possible ensuing maintenance actions. CBM strategies allow for failure prevention by understanding the physics of failure and, subsequently, the corresponding initiation of targeted maintenance activities. CBM is

a maintenance program that recommends maintenance actions based on the information collected through condition monitoring. It attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviours of a physical asset. A CBM program, if properly established and effectively implemented, can significantly reduce maintenance cost by reducing the number of unnecessary scheduled preventive maintenance operations [45].

Diagnostics and prognostics are two key aspects in a CBM program. Diagnostics deals with fault detection, isolation, and identification. Fault detection is a task to indicate whether the observed behaviour of a system is different than the expected one; fault isolation is a task to locate the system component that is faulty; and fault identification is a task to limit the scope of fault and determine the root cause of failure. Prognostics deals with condition assessment which includes predicting and determining the performance life remaining, by modeling fault progression. The current thesis focuses on diagnostics in the sense of fault prediction. Fault prediction is a task to determine an impeding fault and estimate the probability of its occurrence.

A CBM program can be used to do both diagnostics and prognostics. CMSs are applied to various engineering fields. In civil, railways, aerospace, and mechanical structures, CMSs has been implemented for decades successfully. A comprehensive works that presents projects implementing CMSs in different engineering fields and in several regions of the world can be found in [46].

It is worthwhile pointing out that if a failure sequence is rapid, effective condition monitoring is impossible. This is for example the case for electrical faults detected by protection, where the period of action may be only seconds. However, if the failure sequence is days or even months, then condition monitoring has the potential to provide early warning of impending failure [47].

The health condition prediction methods can be divided into model-based methods and data-driven methods [48]. The model-based methods, also known as the physics-of failure methods, perform reliability estimation using equipment physical models and damage propagation models. Model-based methods have been reported for analyzing component reliability such as bearings [49] and gearboxes [50], [51].

However, physics-of-failure models pose some limitations in building models and calculating dynamic responses in complex systems, such as wind turbines. On the other hand, data-driven methods directly utilize the collected condition monitoring data for health condition prediction, and do not require physics-of-failure models. It is not necessary to understand the mechanics and propagation of a damage, which is a limitation in physics-based methods. Examples of the data-driven methods include the proportional hazards model [10], the Bayesian prognostics methods [52], and the Artificial Neural Network (ANN) based prognostics methods [53], [54].

2.3.2 Condition Monitoring in Wind Industry

Current practice in the wind turbine industry involves preventive maintenance and its value in terms of cost savings is discussed in [55]. To monitor general operation and performance of wind turbines, SCADA systems were initially installed. As turbines increased in size and complexity and owing to recent developments in the field of sensing and signal processing, modern wind turbines started being equipped with CMS systems for the active remote monitoring of their components [56]. CM systems provide signals (e.g. vibration,electrical) whose information is used to take suitable maintenance actions and improve protection of costly assets.

The question to be posed when installing those systems is whether their cost is justified. The answer is not straightforward and can be system specific, but the information obtained through them can provide useful insight into the machine working condition and help handling future faults. The installation of CMS can be paid back by the reduction in O&M costs. This can be achieved through decreased downtime, decreased mean time between failure and thus increased production rates. Moreover, upcoming faults can be foreseen through condition monitoring systems, so proper planning, maintenance actions and shipping of components. A study analysing the profitability of condition monitoring both in onshore and offshore wind through life cycle cost analysis is presented in [57]. CMS can offer cost saving potential even for onshore wind turbines, but commercial benefits strongly depend on the ability of these systems to reliably predict a developing fault [38], [58]. A case study on the value of CMS systems

with an application on wind turbine gearboxes is presented in [38]. A reliability-based fault path investigation framework using both quantitative and qualitative methods is given in [37] and it can give valuable information to CMS systems. A thorough failure mode effects and criticality analysis for offshore wind turbine systems with the aim of improving condition based maintenance strategies is given in [59]. In this study, systems for which condition monitoring would generate highest value in wind are analysed and parameters that need to be monitored by a specific system from failure cause to failure mode are understood. The prioritized failure modes of the gearbox include both raceway wear and gear cracks.

UK Supergen Wind Technologies Consortium carried out a survey which shows that 14 out of 20 commercially available wind turbine CMSs include gearbox vibration monitoring. This has led to successful incipient gearbox fault detection and prevention of complete damage [17].

From the CM perspective, a wind turbine gearbox consists of three major components: gears, bearings, and lubricant. CM aims to detect failures at such an early stage that no disassembly of the gearbox is needed. The main condition monitoring techniques in the wind industry are presented in [11], [12], [13]. Monitoring techniques include vibration analysis, oil analysis, thermography, accoustic monitoring, ultrasonic testing, electrical effects, process parameters. Among these different techniques, vibration analysis and oil monitoring are the most predominantly used for wind turbine gearbox applications. The possible reason for this might be their established successes in other industries. A review of wind turbine gearbox condition monitoring methods is given in [60].

Monitoring can either be done on-line or off-line. On-line observation is usually preferred because it can provide a deeper insight into the system performance under loads, as well as both long and short term trends. Both vibration and oil analysis can be applied on-line and off-line [61]. On-line monitoring is also often incorporated into SCADA systems, so that alarms can be triggered in case of an upcoming failure in the system, something particularly useful for inaccessible locations. On-line systems can also potentially reduce the labour cost for off-line inspection.

This PhD focuses on data driven methods for condition monitoring of wind turbine gearboxes using SCADA and vibration data, since they are the most established methods in the industry.

2.4 Data Driven Methods for Condition Monitoring

Conventional reliability analysis mostly provides population-based assessment. On the contrary, individualized failure prediction results take into account signals that can show degradation of the system. This section provides an overview of failure prediction methods using data-driven approaches.

In fault diagnosis, three important aspects are important: time of fault appearance, location of fault and severity of fault. Thus, anomaly detection is first performed, to identify deviation when deviation from normal operation occurs. Then fault localization, narrows the problem source to the specific component or subsystem. Finally, fault classification, discriminates known and unknown faults and identifies the type of the fault if it is previously known [62].

A typical Prognostics and Health Management (PHM) workflow can be conceptually illustrated, as shown in Figure 2.6. This includes data collection, data processing, feature selection from processed data and application of some kind of statistical modeling, which often involves a training phase to establish baseline condition and data driven methods to detect anomalies (these will be discussed in the following paragraphs). The next step is at the actual fault detection and if a degradation is observed, then diagnostics, prognostics and maintenance are the next actions depending on the decision support.

A decision support system aims to analyse various amounts of data, compile comprehensive information and use this information in problem solving and decision making. Given the large amounts of condition monitoring data, statistical methods have been utilised for health management decision support systems and fault diagnostics. Automation is an important part of this process. Various topics in fault diagnosis emphasising on model-based and data-driven approaches were covered in [63]. A recent review on the PHM field is given in [48]. Recent advancements of PHM methodologies,

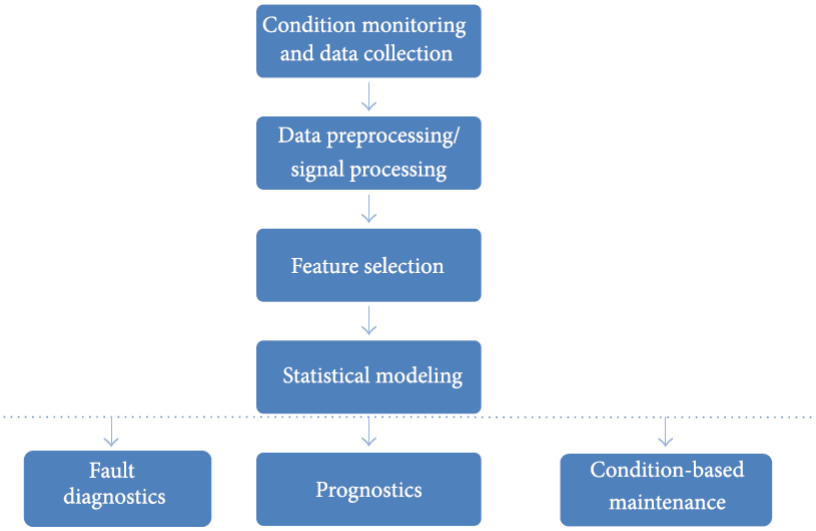


Figure 2.6: Typical flow of PHM systems

with a focus on data-driven approaches, are analysed in [64]. AI and machine learning techniques are increasingly being applied in machine diagnostics. An extensive review of AI approaches for fault diagnostics are given in [65].

AI and machine learning are two terms that are often used interchangeably. AI is the science of training machines to perform human tasks. Machine learning is a subset of AI that trains the machine on how to learn. Machine learning is the process of building an inductive model that learns from a limited amount of data without specialist intervention. This learning implies finding an underlying set of structures that are useful to understand relationships in data that might not be exactly similar to that on which learning occurred. In the machine learning model taxonomy, supervised learning predicts an output variable using labeled input data, while unsupervised learning draws inferences from data without labeled inputs (clustering). For supervised learning there is a distinction between models that predict a numeric variable (regression) or a categorical variable (classifiers) [66].

Many popular methods from machine learning and artificial intelligence are applied in the PHM context, such as Support Vector Machines (SVMs) [67] [68], k-nearest neighbors [69] [70] [71], and decision trees [72]. ANNs have attracted a lot of research interest and are applied widely in many engineering applications for fault detection [73].

Deep learning approaches have recently began to be applied in the field of fault diagnosis [74]. Deep belief networks have also become quite popular in the machine learning field and a multi-sensor health diagnosis method using them is elaborated in [75]. It is worth mentioning that the sensitivity to a given fault is often a function of operating conditions and the nature of the anomaly, therefore environmental parameters need to be taken into account. For example, self-organizing maps are used to categorise and recognise the respective operating conditions [62]. The interested reader can find more about the fundamentals of these methods in [76], although some will be elaborated further in the following sections.

On wind turbines, various data driven methods using machine learning have been applied [66]. There are many dataset types coming from wind turbines and most monitoring models discussed in the literature are based on operational and event datasets. Most models in the literature use SCADA, simulated or experimental vibration data; Both classification methods and regression have been utilised with ANNs, SVMs and decision trees being most popular.

2.4.1 Supervised Models

Supervised learning assigns a label to a vector of input data according to some certain similarity assessment criteria. During the supervised model training, 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 etc. Supervised learning is further categorised into regression and classification.

Classification models learn from the data input and labels and then uses this supervised learning process to classify new observations. This data set may simply be bi-class or multi-class. Some examples of classification problems are: speech recognition, handwriting recognition, bio metric identification, document classification etc.

Regression models use independent variables as inputs in order to predict the numeric output of dependent variables. In the context of reliability and condition monitoring, the component modeled is assumed to be performing at its optimum state.

This optimum (healthy) state is often a function of time, with failure rates having a low steady phase in the second part of the bathtub curve, as explained in Chapter 3. Ideally, normal behaviour data about the component should be recorded during the period when likelihood of failures is low. Having this optimum state, a baseline can be established. At every new time step, the output is predicted based on the derived model. The difference between the measured and predicted output value then constitute the time series of residuals which can be analyzed to determine the presence of a fault.

Decision Trees

A decision tree is one of the most easily interpreted supervised learning methods. It's a hierarchically organized structure, with each node splitting the data space into pieces based on value of a feature. As shown in Figure 2.7a [77], the sample population is divided into sub-regions, through nodes (circles). Leaf nodes (squares) do not split any further and each leaf node usually maps every point in its input region to the same output.

A very deep tree can lead to overfitting and this can be avoided by controlling the depth of the tree. The three parameters that affect the depth of the tree are the maximum number of splits, the minimum leaf size and minimum parent size.

Random Forests

Bootstrap-aggregated decision trees have proven to be a robust classifier [78]. They combine the results of many decision trees, thus reducing the effects of overfitting and improving generalization. Similarly to the decision trees, the hyperparameters often need to be tuned are the maximum number of trees, the depth of the trees and the maximum number of features considered for splitting a node.

kNN

Given a positive integer k, an unseen observation x_{obs} and a similarity metric (distance) d, kNN classifier performs the following steps:

1. The distance d between x_{obs} and each training observation is calculated for the whole dataset. The k points closest to x_{obs} are the subset A_k. An odd number of k is commonly used to prevent ties.

2. The conditional probability of each class is computed, which is the fraction of A_k points with a given class label.

3. x_{obs} is finally assigned to the class with the largest probability.

The choice of k is really important. A small value for k provides the most flexible fit, which will have low bias but high variance. Larger values of k will have smoother decision boundaries which means lower variance but increased bias. The value of k can be tuned for example using grid search in the log-scaled range of $[1 \quad \max(2, \text{NObservations}/2)]$, where NObservations is the number of observations. The similarity distance metric is euclidean.

It should be noted that kNN is one of the simplest and non-parametric machine learning algorithms but it requires the entire training data set to be stored, leading to expensive computation if the data set is large.

SVMs

In classification problems, SVMs aim at finding the decision boundary to separate different classes and maximize the margin. SVMs construct an optimal separating hyperplane between classes in a high dimensional feature space. Hyperplane is an (n minus 1)-dimensional subspace for an n-dimensional space. In a 2-dimension space, its hyperplane will be 1-dimension, which is just a line. This hyperplane creates the biggest margin between the training points of the different classes. For linearly separable classes, margin is the maximal width of the slab parallel to the hyperplane that has no interior data points (Figure 2.7c) [79]. A larger margin leads to lower generalisation error. The support vectors are the data points that are closest to the separating hyperplane and are located on the boundary of the slab.

The mathematical formulation follows [79]. Assuming x_j is a multi-dimensional set of training data points and y_j are the categories (only 2 categories are introduced first). For a dimension d, $x_j \in \mathbb{R}^{\mathrm{d}}$ and $y_j \pm 1$. The hyperplane equation is given in Eq 2.1.

$$f(x) = x'\beta + b = 0 \tag{2.1}$$

where $\beta \in \mathbb{R}^{\mathrm{d}}$ and b is a real number.

All the data points for the class 1 are on one side, and all the data points for class 1 on the other.

The problem of finding the best hyperplane is defined as the problem of finding β and b that minimize $||\beta||$ such that for all data points $y_j f(x_j) \geq 1$. The support vectors are the points x_j on the boundary for which $y_j f(x_j) = 1$.

Minimising $||\beta||$ is a quadratic programming problem and the optimal solution enables the classification of a vector z based on the sign of the classification score $f\hat{(}z)$, which represents the distance from the decision boundary.

In the linearly separable case, SVM is trying to find the hyperplane that maximizes the margin, with the condition that both classes are classified correctly. But in reality, datasets are probably never linearly separable, so the condition of 100% correctly classified by a hyperplane will never be met. SVMs can perform non-linear classification using soft margins and kernels. Soft margins are hyperplanes that separates many, but not all data points. Kernels operate as a similarity function. The kernel function offers a map of the originally inseparable data to a higher dimensional space where they can be linearly separated.

The problem tries to minimise 2.2, such that $y_j f(x_j) \leq 1 - \xi_j$ and $\xi_j > 0$. Slack variables are defined in the context of optimisation problems, to transform an inequality expression into an equality expression (with an added slack variable). ξ_j are slack variables and C is a penalty parameter.

$$(\frac{1}{2}\beta'\beta + C\sum_j \xi_j) \tag{2.2}$$

The penalty parameter represents the tolerance (soft) when finding the decision boundary and is an important hyperparameter for the SVM. The bigger the C, the more penalty SVM gets when it makes misclassification. Therefore, the narrower the margin is the fewer support vectors the decision boundary will depend on.

The problem of finding the optimal hyper plane is an optimization problem and can be solved by optimization techniques. Lagrange multipliers can be used to solve it analytically). C is sometimes called a box constraint and keeps the allowable values of

the Lagrange multipliers in a "box" bounded region.

The Kernel trick utilizes existing features, applies some transformations, and creates new features. Those new features are the key for SVM to find the nonlinear decision boundary. A Gaussian kernel is often chosen, because it depends on the relative distance between elements and it has good approximation capabilities [80]. The Radial Basis Function Gaussian kernel is a transformer/processor that generates new features by measuring the distance between points to specific centers.

$$\phi(x, center) = exp(-\gamma \left\| x - center \right\|^2) \tag{2.3}$$

γ controls the influence of new features $\phi(x, center)$ on the decision boundary. The higher the gamma, the more influence the features will have on the decision boundary.

Binary classification was first introduced, but multiclass classification can be achieved by reducing the single multiclass problem into multiple binary classification problems. Common methods for such reduction include one-vs-all or one-vs-one. Classification of new instances for the one-versus-all case is done by a winner-takes-all strategy, in which the classifier with the highest output function assigns the class (it is important that the output functions be calibrated to produce comparable scores). For the one-versus-one approach, classification is done by a max-wins voting strategy, in which every classifier assigns the instance to one of the two classes, then the vote for the assigned class is increased by one vote, and finally the class with the most votes determines the instance classification. The interested reader can find out more in [81], [82].

For regression problems, let x_n is a multivariate set of N observations with observed response values y_n. The linear function is given in Eq (2.4) and it should deviate from y_n by a value no greater than ϵ for each training point x and at the same time be as flat as possible.

$$f(x) = x'\beta + b \tag{2.4}$$

This is formulated as a convex optimization problem to minimize Eq (2.5) [83]. The slack variables ξ_n, ξ_n^* are introduced for each point to deal with otherwise infeasible

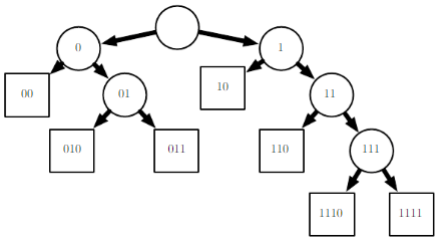
constraints. This approach is similar to the soft margin in SVM classification, because the slack variables allow regression errors to exist up to the value of ξ_n, ξ_n^*, yet still satisfy the required conditions.

$$\begin{aligned} J(\beta) &= \frac{1}{2}\beta'\beta + C\sum_{n=1}^{N}(\xi_n + \xi_n^*) \\ \forall n :& y_n - (x_n'\beta + b) \leq \epsilon + \xi_n \\ \forall n :& (x_n'\beta + b) - y_n \leq \epsilon + \xi_n^* \\ \forall n :& \xi_n \geq 0 \\ \forall n :& \xi_n^* \geq 0 \end{aligned} \tag{2.5}$$

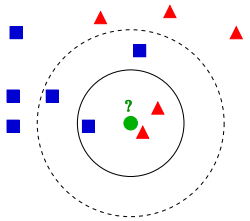
The constant C was introduced in the classification as the box constraint. In regression it's a positive numeric value that controls the penalty imposed on observations that lie outside the margin ϵ and performs regularization, therefore helping to prevent overfitting. The loss function is the distance between observed value y and the ϵ boundary (Eq 2.6).

$$L = \begin{cases} 0 & \text{if } |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon & \text{otherwise} \end{cases} \tag{2.6}$$

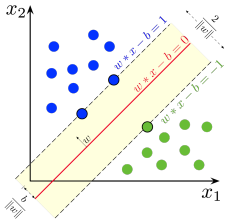
For non-linear problems, kernel functions can be applied similarly to classification.



(a) Diagrams of a decision tree. The tee nodes get the inputs and send them to the child node on the left (0) or the child node on the right (1).



(b) kNN classification example. The test sample (green circle) should be classified either to the class of blue squares or to the class of red triangles. If k = 3 (solid line circle) it is assigned to the class of triangles but if k=5 it is assigned to the class of squares.



(c) SVM classification example for two classes. The maximum-margin hyperplane and margins are shown.

Figure 2.7: Classification Algorithms

Artificial Neural Networks

An ANN has three distinctive part: an input layer consisting of nodes, one or more hidden layers of neurons and an output layer. They perform computations through a process of learning which is achieved through the interconnection of neurons. The inputs and their connectivity with the neurons depends on weight parameters. Each neuron has a nonlinear transfer function that determines the internal activity of the neuron and an activation function that provides the threshold for producing the output. The choice of activation function is determined by the nature of data and assumed distribution of target variables.

ANNs can be used both for regression and classification problems. Regression tries to map a set of continuous inputs to another set of continuous outputs, so neural networks for regression have multiple input neurons and one output neuron. A representation is shown in Figure 2.8.

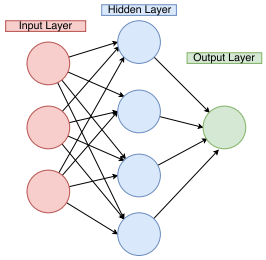


Figure 2.8: Neural Network Representation for Regression

Self-organizing Maps

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. SOM is suitable for both supervised and unsupervised learning schemes. This algorithm will be elaborated further in the next section.

Statistical Pattern Recognition

Statistical pattern cognition techniques can be used to detect faults when there is lack of domain/expert knowledge, but changes in stochastic parameters are indicating system abnormality. Using historical data and a given confidence interval, the control limit for monitored parameters can be set according to statistical significance.

Labeling

During the training phase, each input vector together with the corresponding system state indicated by a label is fed into the machine learning model. The input vector can consist of features extracted from pre-processed time series of signals relevant to the modeled component. For gearbox CM there might be labels such as healthy, high speed gear failure, planet bearing failure, etc.

The labels specifying the category to which training instances belong are usually assigned by human/expert judgement. This is time consuming, error prone and likely to result in a set of labeled vectors with an unbalanced number of classes [84]. The problem of unbalanced classes can be addressed using under-sampling (remove instances belonging to the majority class), over-sampling (sample more instances from minority

class), SMOTE, Tomek-links (which removes points in the majority class that are considered borderline, noise or redundant) [85], [86], [87].

Semi-Supervised models

Semi-supervised models include one-class classifiers. These can be anomaly or outlier detectors. This group uses a semi-supervised mode of training wherein a class/pattern/set of normal condition operation is learned. This is useful in condition monitoring when only a large set of normal operating data is available but not enough historic failure examples with labeled data have been acquired. This is often the case in real word industrial applications. The premise of one-class classifiers is that any deviation from the trained pattern is considered an anomaly, which can indicate a fault event.

One class classifiers can be statistical, neural network based [88], [89] or use support vector data description (SVDD) [90]. SVDDs are inspired by support vector classifiers but instead of using a separating hyperplane to distinguish two classes, they use a spherical boundary (enclosing normal condition data samples) defined by outermost data samples, the support vectors. They have been used in pump anomaly detection in various operating modes [91] as well as bearing fault detection [92] [93].

As it is introduced by, one class SVMs basically separate all the data points from the origin and maximize the distance from this hyperplane to the origin. The objective function to be minimised is given in Eq (2.7).

$$\frac{1}{2}\left\|\beta\right\|^2 \frac{1}{\nu n}\sum_{j=1}^{n}\xi_j - \rho \tag{2.7}$$

with respect to β, ξ, ρ subject to

$$\beta\phi(x_j) \geq \rho - \xi_j$$

$$\xi_j \geq 0$$

In this formulation, parameter ν is very important; it sets an upper bound on the fraction of outliers (training examples regarded out-of-class) and it is a lower bound on the number of training examples used as Support Vector.

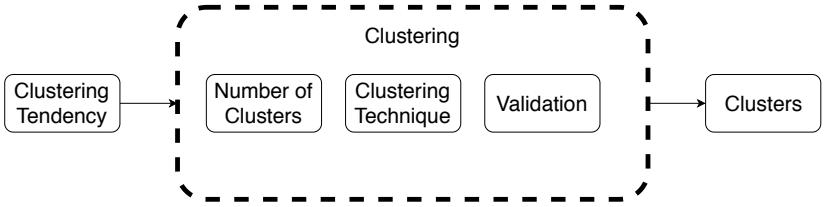


Figure 2.9: Overview of clustering

A small value of ν leads to fewer support vectors and, therefore, a smooth, crude decision boundary. A large value of ν leads to more support vectors and, therefore, a curvy, flexible decision boundary. The optimal value of ν should be large enough to capture the data complexity and small enough to avoid overfitting. Also, $0 < \leq 1$.

2.4.2 Unsupervised models

Unsupervised learning involves the grouping of data points and according to this approach, a set of data points that have similar properties should be clustered in a specific group.

The fundamental principle of clustering techniques is to calculate a certain distance Δ. Let x and y be two units clustered in group C_i and d be any distance measurement technique such as Euclidean distance, Minkowski, Manhattan [94].

$$\begin{aligned} \text{Maximum Distance}, \Delta_{max} &= _{x,y \in C_i}\ max(d(x,y)) \\ \text{Distance from Centroid}, \Delta_{cen} &= \frac{\sum_{x \in C_i} d(x, \bar{x_i})}{|C_i|} \end{aligned} \tag{2.8}$$

Where $\bar{x_i}$ is the cluster centroid of group C_i.

The general goal of clustering is to minimize the dispersion within a group as well as maximize the isolation of groups among each other.

A general overview of the clustering procedure followed in the thesis is given in Figure 2.9. Each of the steps in explained in more detail in the following paragraphs.

The first step is clustering tendency, which basically examines if there is any grouping structure to the data- otherwise clustering is meaningless. The grouping structure can be determined by using some tests for randomness, uniformity or homogeneity. Those tests are often hypothesis tests where the null hypothesis assumes homogeneous

data and a statistic is computed to determine the validity of this hypothesis. Some tests include the Cox-Lewis test [95], the Holgate [96] and T2 [97]. A commonly used one is the Hopkins statistic H [98] [99].

When the actual clustering procedure starts, an important parameter to be determined is the number of clusters. This is used as input in many clustering algorithms. Methods and taxonomy to determine the number of clusters are presented in [100]. These methods are primarily based on the quality of the partition- such as within-cluster cohesion and between-cluster separation . Some commonly used methods are Aikaike Information Criterion, Bayes Information Criterion, Minimum Description Length.

There are numerous types of clustering algorithms which mainly differ in how the samples are grouped. The basis for grouping depends on the algorithms fundamental notion of a cluster and how it efficiently finds the clusters. The interested reader can find surveys of clustering methods in [101]. The main categories are hierarchical clustering [102], centroid-based clustering (like k-means), distribution based clustering (like Gaussian Mixture Models) and spectral clustering. Some of them will be further discussed in this section.

k-means

This algorithm groups data into clusters by measuring the distance between data points [103]. The most popular measure is Euclidean distance. The k-means algorithm in a few steps can be described as following and are shown in Figure 2.10:

1. k points are selected as the initial centroids
2. All points are assigned to the closest centroid
3. k clusters are formed
4. The centroid of each cluster is recomputed
5. The data points are reassigned to the centroid that is closest to them now
6. Steps 2-5 are repeated until cluster doesn't change

Self-Organising Maps

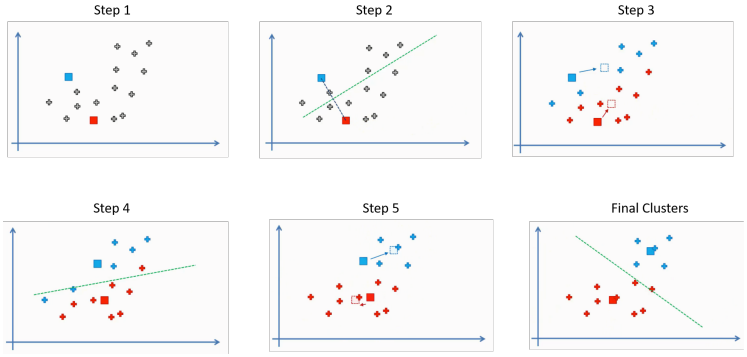


Figure 2.10: Steps of k-means clustering algorithm

A Self-Organising Map (SOM) is a type of ANN that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples. This low-dimensional space is the map and therefore SOM can be used for dimensionality reduction. Unlike traditional ANNs which use error correction learning, SOMs use competitive learning for training. 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.

The algorithm is described by the following steps:

- The node weights are initialized.
- A random vector is chosen from the training dataset.
- Every node is examined to calculate which ones weights are most like the input vector. Therefore, the distance from each weight to the sample vector is calculated. The weight with the shortest distance is the winner. A common distance metric used is the Euclidean distance. The winning node is commonly known as the Best Matching Unit (BMU).
- Then the neighbourhood of the BMU is calculated. The amount of neighbours decreases over time.
- The winning weight is rewarded with becoming more like the sample vector. The neighbours also become more like the sample vector. The closer a node is to the

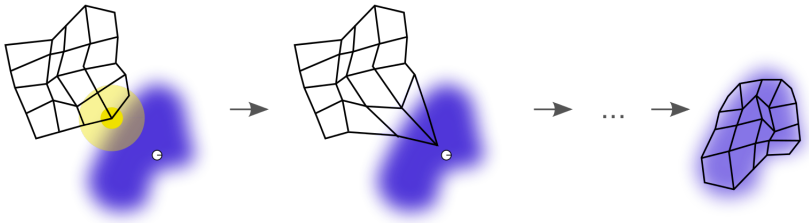


Figure 2.11: Self Organising Map Training [7]

BMU, the more its weights get altered and the farther away the neighbor is from the BMU, the less it learns.

The competitive update procedure is shown in Eq. 2.9:

$$\begin{aligned} w_j &= [w_{j1}, w_{j2}, w_{j3}...w_{jn}], j = 1, 2, ...m \\ w_j(t+1) &= w_j(t) + a(t)h_{j,w_c}(t)(x - w_j(t)) \end{aligned} \tag{2.9}$$

where w are the weights, h_{j,w_c} is the Gaussian kernel function around the BMU, and $a(t)$ denotes step size, which monotonically decreases with step iterations.

This process is illustrated simply in Figure 2.11. The blue area shows the training data distribution and the small white disc is the current training sample drawn from that distribution. At first (left) the SOM nodes are arbitrarily positioned in the data space. The node nearest to the training node (highlighted in yellow) is selected, and is moved towards the training datum, as (to a lesser extent) are its neighbours on the grid. After many iterations the grid tends to approximate the data distribution (right).

Principal Component Analysis

Principal component analysis (PCA) is a widely used statistical technique for unsupervised dimension reduction.

Dimension reduction is useful because in general, the number of available training samples (N) should exceed the number of features (L) as the complexity of a model cannot exceed the complexity of the training dataset. Depending on the application, the features nature and specific assumptions, different N/L ratio recommendations are provided in literature [104].

PCA provides the orthogonal transformation of possibly correlated variables into a

set of linearly uncorrelated ones (principal components). To perform PCA on a given dataset, the following steps are required [105]:

1. The mean of each data dimension is calculated and subtracted from the original dataset in order to obtain the adjusted dataset.

2. The datasets covariance matrix is calculated.

3. Eigenvalues and unit eigenvectors of the covariance matrix are calculated.

4. Eigenvectors are sorted by eigenvalue, highest to lowest. A number n of features is selected based on the explained variance complexity trade-off and a feature vector is obtained by combining the first n eigenvectors.

5. The post-PCA dataset matrix is obtained by multiplying the transpose of the feature vector by the transpose of the adjusted dataset matrix.

By only retaining a number of features, the dimensionality of the modelling is reduced providing a basis for better results given a limited amount of samples. Nevertheless, at the same time there exists the inherent trade- off of some variance from the dataset being lost. As such, an optimal number of principal components should be selected so that most of the datasets variance remains explained while the total number of features is reduced.

PCA can be used both as a dimensionality reduction technique and as a clustering technique.

2.4.3 Hyperparameter Tuning

A hyperparameter is a parameter whose value is set before the learning process begins. Hyperparameters are not updated during the learning and are used to configure either the model (e.g. number of decisions trees and their depth, number of layers of a deep neural network) or the algorithm used to lower the cost function (learning rate for gradient descent algorithm). Hyperparameters should not be confused with model parameters and they can't be directly trained from the data. Model parameters are learned during training when a loss function is optimised using e.g. gradient descent.

The model parameters specify how to transform the input data into the desired output whereas the hyperparameters define how the model is actually structured. Model parameters are the properties of the training data that are learnt during training by the machine learning model. An example of model parameter is the weights and biases of an ANN, which are learnt during the training process. Model hyperparameters, are common for similar models and can't be learnt during training but are set beforehand. Hyperparameters can also be optimised when designing models and that's the reason why the distinction between parameters and hyperparameters is nuanced.

Hyperparameter tuning is choosing a set of optimal hyperparameters for a learning algorithm. Tuning strategies include grid search and random search. Grid search works by searching exhaustively through a specified subset of hyperparameters. This means it can often be time consuming and computationally expensive but very effective in finding the optimal parameters. Random search differs from grid search mainly in that it searches the specified subset of hyperparameters randomly instead of exhaustively. The major benefit being decreased processing time.

2.4.4 Training,Validation and Testing

For model evaluation and testing purposes, the dataset is split into 3 sets: training, validation and testing. The training dataset is the actual dataset that we use to train the model. The model learns and is fitted from this data.

Validation dataset is the sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. Finally, the test dataset is the sample of data used to provide an unbiased evaluation of a final model fit on the training set.

A standard approach in machine learning is to use cross validation techniques accompanied by hold-out sets. For model validation, the original dataset is split into training (typically 70% of the data) and testing (30%). Using the training data, k-fold cross validation (typically k=10) is performed with different hyperparameters for each model tested noting the results in terms of validation measures.

Cross validation is often called out-of-sample testing is a model validation technique

used for assessing generalisation of results of statistical analysis to independent datasets. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset, and validating the analysis on the other subset. To reduce variability, in most methods multiple rounds of cross-validation are performed using different partitions, and the validation results are combined (often averaged) over the rounds to give an estimate of the model's predictive performance. In typical cross-validation, results of multiple runs of model-testing are averaged together, but the holdout method involves a single run, making it the simplest form of cross-validation. In k-fold cross-validation, the original sample is randomly partitioned into k equal sized sub-samples. Of the k sub-samples, a single sub-sample is retained as the validation data for testing the model, and the remaining k 1 sub-samples are used as training data. The cross-validation process is then repeated k times, with each of the k sub-samples used exactly once as the validation data. The k results can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling (see below) is that all observations are used for both training and validation, and each observation is used for validation exactly once. Stratified cross validation is often chosen in classification of imbalanced classes, to ensure that the train and test sets have approximately the same percentage of samples of each target class as the complete set. In the holdout method, points are randomly assigned to two sets (a training and test set). More details on these methods can be found in [106].

There are various metrics to evaluate model performance. These are given in Eq. 2.10.

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\ Recall &= \frac{TP}{TP+FN} \\ Precision &= \frac{TP}{TP+FP} \\ Specificity &= \frac{TN}{TN+FP} \\ F1 &= 2\frac{Precision\dot{R}ecall}{Precision+Recall} \end{aligned} \tag{2.10}$$

Where TP is the number of True Positives (when the component was predicted as healthy and it was in fact healthy), FN is the False Negatives (when the model predicted unhealthy and it was in fact healthy), TN is True Negative (predicted as unhealthy and actually unhealthy) and FP is False Positive (FP, when the model predicted healthy and the instance was unhealthy).

The validity of the regression-based models can be expressed through various metrics. Some of them are the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination R^2 which are shown in Eq 2.11.

$$\begin{aligned} \text{MAE} &= \frac{1}{N}\sum_{i=1}^{N}|y_i - x_i| \\ \text{RMSE} &= \sqrt{\frac{1}{N}(y_i - x_i)^2} \\ R^2 \quad &= 1 - \frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \end{aligned} \tag{2.11}$$

Where N is the number of data points, y is the actual output and x is the predicted output.

In clustering analysis validation provides a metric to the data partitioning quality. A comprehensive survey can be found in [107]. A few of them are given in Eq 2.12.

$$\begin{aligned} \text{Silhouette} &= \frac{B_i - A_i}{max(B_i, A_i)} \\ \text{Calinski-Harabasz} &= \frac{SS_B}{SS_w}\frac{N-k}{k-1} \end{aligned} \tag{2.12}$$

The silhouette value for each point is a measure of how similar that point is to points in its own cluster, when compared to points in other clusters,where A_i is the average distance from the ith point to the other points in the same cluster as i, and B_i is the minimum average distance from the ith point to points in a different cluster, minimized over clusters.

The silhouette value ranges from 1 to 1. A high silhouette value indicates that i is well matched to its own cluster, and poorly matched to other clusters. If most points have a high silhouette value, then the clustering solution is appropriate. If many points have a low or negative silhouette value, then the clustering solution might have

too many or too few clusters. Silhouette values can be used as clustering evaluation criterion with any distance metric.

The Calinski-Harabasz criterion is sometimes called the variance ratio criterion. where SS_B is the overall between-cluster variance, SS_W is the overall within-cluster variance, k is the number of clusters, and N is the number of observations. Well-defined clusters have a large between-cluster variance and a small within-cluster variance. The larger the ratio, the better the data partition.

2.4.5 Models Used in the Thesis

The previous section presented some of commonly used available methods for training machine learning models. Most of those are utilised throughout the thesis for both supervised and unsupervised learning. Algorithms used include SVMs, Random Forests, Neural Networks and Self-Organising Maps.

2.4.6 Analysing SCADA Data

SCADA systems have been widely used in wind turbine condition monitoring applications, because they are already installed in most machines and therefore offers a cost effective diagnostic solution. Due to this, using SCADA data for failure detection has been extensively researched in the literature using a variety of approaches. A comprehensive, recent review on the use of SCADA data in condition monitoring of wind turbines is given by [108]. The review splits the approaches using SCADA data for fault detection into three main categories; trending, clustering, and normal behaviour modelling.

Trending using principal component analysis with an auto-associative neural network can be used to train normal operation data and the principal components produced are evaluated through statistic model variations [109]. A technique using correlations among relevant SCADA data is investigated in [14]. The relationship between gearbox temperature, power output and efficiency are derived using the first law of thermodynamics in [110], [111]. It is shown that temperature trend rises, while the efficiency drops months before a planetary gear failure. Clustering on SCADA data has been

performed by [112] and [109] using ANN self organising maps, because of their ability to represent and visualise large datasets. The first proposed a method that trains only normal operational data and calculates the distance between new input data and the best matching neuron for abnormality detection, whereas the second uses different failure categories as clusters. Both trending and clustering have limitations in online monitoring because of the difficulty in interpreting the results and changes in data.

Normal behaviour models have been used in various condition monitoring applications. Anomalies have been detected from normal operation in gas turbines [113] and transformers [114]. The main principle of this method is to model a measured parameter using operational examples. Then the residual of the measured minus the modeled signal may indicate a fault, if the deviation between the modeled and the measured signal increases. Normal behaviour models have been introduced in SCADA data analysis using either linear and polynomial approaches or ANNs. ANNs are suitable for SCADA data because of their ability to capture non-linear relationships between observations, as shown in [115] and [15] for gearbox bearing and cooling models. The advantages of ANNss over linear models are demonstrated in [116] through bearing damages in offshore wind turbines. A model of the main bearing temperature using ANN is developed in [117], and it is shown that significant indications of fault are apparent three months before the turbine stopped due to overheating. [118] tested ANN configurations with different number of neurons and activation functions in order to detect a bearing fault using high frequency SCADA data. A self evolving ANN has been proposed to automatically update monitoring and prognostic models in [119], [120].

Other research includes a methodology classifying and predicting turbine faults using real wind turbine SCADA data in [121].Time-sequence and probability-based methods are proposed for SCADA alarm analysis in [122]. A non-linear state estimation technique model to model a healthy wind turbine gearbox using historical data is presented in [123]. This methodology uses a memory matrix and a weigting vector to model each operational state and a hypothesis test for anomaly detection. SCADA data used in industry consist of mainly 10 minute average values, however the benefits of high frequency data for performance monitoring are discussed and presented in [124].

2.4.7 Analysing Vibration Data

Vibration analysis in the context of the present work requires the installation of acceleration transducers on the gearbox surface, which offers sensitivity in fault discovery.

In order to analyse vibration signals, their nature and fault signature needs to be understood. A local gear defect, like a crack, can affect the gear mesh stiffness and the dynamic properties of the whole gearbox (a pitting fault propagation model can be found in [125]). This changes the vibration signal, which is reflected in amplitude and frequency modulation as shown in [126]. A model for fault diagnosis of planetary gearboxes considering the modulation effects and the vibration transfer path is given in [127]. Sideband patterns in planet gears are explained in [128] and results are evaluated in a dynamometer test bed. A general model for faulty rolling element bearing vibration signals can be found in [129]. A dynamic model of the wind turbine planetary drivetrain is used to determine the vibration signatures of the planet bearing inner and outer race defects in [130].

Diagnosis of gearboxes through vibration signals can be performed using time, frequency or time-frequency methods [45].

Time domain analysis is based on the time waveform. Time domain analysis methods include mainly calculation on descriptive statistics broadband analysis methods, so the monitored parameters are peak, root mean square (RMS), crest factor , peak-to-peak interval, mean, standard deviation, skewness, and kurtosis. Successful results for the RMS indicator are presented in [131], but the signals were sampled at constant position intervals in a lab experiment. Simply monitoring those parameters in the time domain has not shown that successful results in real wind turbine variable speed signals that contain noise, that's why another popular time domain method for rotating machinery called Time Synchronous Averaging (TSA) has been used [132].

Frequency domain analysis is based on the transformed signal in the frequency domain, which makes it easier for specific frequency components of interest to be identified. The most common practice is spectrum analysis using Fast Fourier Transform. Spectrum analysis either looks at the whole spectrum or at certain frequency components and extracts features. The power spectrum is broadly used in spectrum analysis.

Traditional frequency domain analysis through Fast Fourier Transform can sometimes be insufficient for the analysis of bearing faults. One of the primary studies on bearing diagnostics found that bearing fault signals are found in the high frequency region of resonances excited by internal impacts [133]. That is the reason why the analysis of envelope signals reveals more diagnostic information than the analysis of raw signals, since the signal is bandpass filtered in a high frequency band where the fault impulses are amplified by structural resonances [134]. A comprehensive explanation of envelope analysis for bearings is given in [135]. Furthermore, other spectra have been developed like the cepstrum [136], which is capable of detecting harmonics and sideband patterns in the power spectrum.

Time-frequency domain methods have been proposed in order to handle non-stationary signals, which are often common in machinery faults. In time-frequency analysis the signal power or energy is represented in two dimensions of both time and frequency. The most popular ones are Short-Time Fourier Transform [137] and Wigner-Viller distribution. Another widely used method, which instead of a time frequency representation is based on time-scale representation of signal, is wavelet transform [138], [139].

Different vibration analysis methods regarding wind turbine gearboxes are evaluated and presented in [140]. To understand the dynamic responses of wind turbine gearboxes under different loading conditions, two identical gearboxes are tested. One was tested on a dynamometer and the other was first tested in the dynamometer, and then field tested in a turbine in a nearby wind plant. In the field, the test gearbox experienced two oil loss events that resulted in damage to its internal bearings and gears. 12 accelerometers are installed close to the various stages of 2 gearboxes. A single vibration measurement (collection of signal only at 1 timestamp for 10 minutes at 40kHz) from only 1 operating condition for each sensor of the healthy gearbox is collected. For the damaged gearbox, single vibration measurements of 3 operating conditions of faulty samples is presented. The faulty samples include damaged gears and bearings throughout the gearbox. The project involved the analysis of the collected vibration data by several independent research partners and conclusions were drawn from the comparison of their analysis results. However, the single measurements of

the signals can be a limitation to generalise those conclusions. Condition indicators for vibrations are summarised in [141]. TSA is commonly used in gearbox diagnostics and applications in wind industry can be found in [142], [143]. TSA algorithm was evaluated on gearbox experiments in [142] but did not show very promising results when applied to a real wind turbine planetary gear in [143]. Sideband energy ratio is a simple method that can detect amplitude and frequency modulation in the spectrum, indicating the gear's health condition and has been successfully demonstrated [144]. A similar technique is sideband power factor presented in [145]. Cepstrum analysis on gearbox fault diagnosis, indicating families of sidebands that are caused by modulation when a fault occurs [146].

Spectral Kurtosis (SK) [147] is a valuable tool for extracting transients buried in noise and locating the frequency bands with a high amount of impulsiveness, while at the same time it filters the signal to maximise that impulsiveness. Applications of the spectral kurtosis can be found in [148], [149]. A ring gear fault on a wind turbine planetary gearbox is detected using SK in [143] and the results seem to be much more promising compared to TSA. A comprehensive tutorial on rolling element bearing diagnostics using SK is given in [150]. An alternative time frequency method is multi scale enveloping spectrogram using various scales for wavelet transforms, and an application of this on roller bearings is given in [151]. The underlying periodicities in vibration signals can sometimes better be described by cyclostationarity which can be analysed through spectral correlation [152]. The spectral correlation algorithm has been made more computationally efficient in the work presented by Antoni [153]. An application of cyclostationary process in early gear teeth spalling on a test bench can be found in [154].

SK has been optimised for rolling element bearings using autoregressive (AR) models [155], complex Morlet wavelets [155] and the Kurtogram [147]. SK in high speed machines can be enhanced through minimum entropy deconvolution, which effectively deconvolves the effect of the transmission path and clarifies the impulses, even where they are not separated in the original signal, as presented in [156]. Signal to noise ratio is improved through linear prediction and self adaptive noise cancellation, while spec-

tral kurtosis is used to identify the frequency band with the high level of impulsiveness in [157].

A local mean decomposition method is used to decompose the multi-component amplitude-modulated and frequency-modulated signal into a number of product functions and extracts the fault information from these functions through multi-scale permutation entropy. Fault diagnosis is then conducted using improved support vector machines based on binary trees in [158]. Empirical Mode Decomposition (EMD) is a valuable time-frequency tool, especially for the non-stationary vibration signals of wind turbines. An application of EMD on a wind turbine gearbox gear pitting fault is presented in [159]. EMD combined with energy separation can provide a framework for detecting planetary gearbox faults, as presented in [160]. Ensemble EMD and singular value decomposition are both used and presented in [161]. Time and frequency features are combined and a high-dimensional hybrid-domain feature vector is created, so the state of the bearing is assessed through relative compensation distance.

An initial application of wavelets to gearbox diagnostics is described in [162]. The use of discrete wavelet transform for noise cancellation and continuous wavelet transform for condition monitoring and fault diagnosis in wind turbines is demonstrated in [163]. A continuous Morlet wavelet transform is used to filter out noise in vibration signals and Wiegner-Ville distribution is used as a time-frequency analysis method in [164] and its efficiency is applied and tested on an experimental wind turbine gear. Adaptive optimal kernel time-frequency analysis is demonstrated on wind turbine lab experimental non-stationary signals from both local and a distributed faults on gears in [165].

Multi fault detection using offline analysis of vibration signals coming from wind turbine gearboxes is investigated in [166]. For this purpose, conventional methods such as Hilbert transform, cepstrum analysis and multi-scale enveloping spectrogram are used to reveal faults in gears and bearings.

2.5 Closing the literature gap

There has been plenty of research interest in wind turbine gearbox condition monitoring. The present thesis will try to close the gap of the lack of an integrated wind turbine gearbox fault detection and diagnostic framework using artificial intelligence and a large population of real wind turbine operational data.

This thesis initially performs reliability analysis of offshore wind turbine drivetrains, giving some valuable information on failure rates of wind turbine drivetrain components, shown in Chapter 3. Then, in Chapter 4 a planet bearing gearbox fault is analysed using SCADA data and various temperature monitoring methods that were discussed in this Chapter, such as normal behaviour modelling. Chapter 5 analyses vibration signals and compares various signal processing methods for gear and bearing fault detection. A feature extraction approach is proposed and those features as used as inputs in pattern recognition models that perform anomaly detection, fault isolation and diagnosis, These models are applied in wind turbine units and the whole wind turbine fleet in 6. Machine learning models that were discussed in this chapter are used, such as SVMs and SOMs.

The proposed methodology will aid in gaining a better insight into various sources of data coming from wind turbine gearboxes, explore and compare various analysis methods in terms of signal processing and machine learning. The results can be used in understanding gearbox failures and how incipient faults can be detected.

Chapter 3

Reliability of Wind Turbine Drivetrain Components

3.1 Chapter Contribution

The reliability of wind turbines is one of the most critical drivers of the O&M costs for wind energy, therefore it is important to gain a better understanding of turbine reliability. A review on reliability and its impact on cost of energy for wind is given in [167]. Reliability is defined in [168] as the probability that a product or a system will perform its intended functions satisfactorily (i.e., without failure and within specified performance limits) for a specified length of time, when operating under specified environmental and usage conditions. As described in Chapter 2, in the context of wind energy, reliability describes the wind turbine's ability to produce energy whenever the wind speed is higher than the wind turbine's cut-in speed and lower the wind turbine's cut-out speed throughout its design life.

It is discussed in almost all the literature reviewed in Section 2.2 that there is not enough data in the public domain for conducting reliability analysis of wind turbines. Most existing reliability studies have focused on onshore, especially for older turbines with low power ratings. As multi MW wind turbines are increasingly being installed offshore, reliability and failure rate analysis is becoming even more crucial due to lack of accessibility. When it comes to critical sub-assemblies , such as the gearbox, that

consist of multiple components it is important to understand failure rates and failure modes for different turbine configurations and drivetrain types. It is discussed in [167] that the wind turbine sub-assemblies that are more critical in terms of downtime are gearbox and generator, which can reach up to more than 300 hours per year.

There is thus an opportunity to carry out novel research through the analysis of a large offshore population of multi MW turbines with different wind turbine types in terms of drivetrain failure rates. This chapter aims to answer the following research question:

"What are the most critical failure rates and failure modes of modern multi MW wind turbine gearbox components?"

The contributions of the chapter are as follows:

1. Perform failure rate analysis of offshore wind turbine drivetrain sub-assemblies
2. Present failure rates from multi MW (more than 5MW) wind turbines
3. Present both geared and direct drive machine failure rates
4. Further analyse failure modes from top failure rates

3.2 Failure Rates and Failure Rate Categories

Failure is defined in different ways in the literature. [169] and [34] have stated that there is no standard format for defining failure rates. The definition is often influenced by the extent of data available to the analyst. For example, some papers can include the amount of downtime in their failure definitions and some cannot, depending on data availability. In [18] a failure is defined as the termination of the ability of an item to perform a required function. However the definition provided in [21] includes downtime in its definition and a failure is defined as the stoppage of a turbine for one or more hours that requires at least a manual restart to return it to operation.

In the analysis carried out in the following sections a failure is defined as failure of an equipment unit that compromises one or several functions, according to ISO 14224 [170]. A failure can be critical or degraded. Critical failures require immediate

action towards cessation of performing the function, even though actual operation can continue for a short period of time. A critical failure results in an unscheduled repair. Degraded failures can be gradual, partial or both. The function can be compromised by reduced or erratic outputs. An immediate repair can normally be delayed but, in time, such failures can develop into a critical failure if corrective actions are not taken.

The present study provides failure rates in a per turbine per year format, as it has been already been widely used in the literature [34], [171]. The formula used is given in Eq. (3.1).

$$\lambda = \frac{\sum_{i=1}^{I} \sum_{k=1}^{K} \frac{n_{i,k}}{N_i}}{\sum_{i=1}^{I} \frac{T_i}{8760}} \tag{3.1}$$

Where λ is failure rate per turbine per year, I is the number of intervals for which data are collected, K is the number of sub-assemblies, $n_{i,l}$ is the number of failures N_i is the number of turbines and T_i is the total time period in hours.

Reference [172] provides an overview of other terms used in wind turbine reliability papers. It states that $1/\lambda$ is the mean time to failure (MTTF). The repair rate is the transition rate from the failed to operational state and is represented by μ. $1/\mu$ is the mean time to repair (MTTR). The mean time between failures (MTBF) is the sum of the MTTR and MTTF.

The Weibull distribution is widely used in reliability and life data analysis due to its versatility. Depending on the values of the parameters, the Weibull distribution can be used to model a variety of life behaviours. The shape parameter β is quite important in this. Different values of the shape parameter can have effects on the behaviour of the distribution, by causing the distribution equations to reduce to those of other distributions. For example, when $\beta = 1$, the pdf of the three-parameter Weibull reduces to that of the two-parameter exponential distribution.

$$f(t) = \frac{\beta}{\theta} \frac{t}{\theta}^{\beta-1} e^{-\frac{t}{\theta}^{\beta}} \tag{3.2}$$

Where β is the shape parameter, θ is the scale parameter that has the dimensions of time and t is time.

A wind turbine is considered a repairable system, which means it can usually be

returned to operation after a failure by some repair process other than complete system replacement. Minimal repair is often assumed with wind turbines, returning the wind turbine back to the condition it was in before failure [173].

Most models for repairable systems have the forms of non-homogenous Poisson processes, like the famous Crow-AMSAA model [174] which uses power law function, and bounded intensity process model [175] which uses bounded exponential function. The functional form of Crow-AMSAA model is given in Eq. (3.3).

$$\lambda(t) = \frac{\beta}{\theta}\frac{t}{\theta}^{\beta-1} \tag{3.3}$$

Where λ is the failure rate, θ is the scale parameter, β is the shape parameter and t is time.

Many mathematical concepts apply to reliability engineering, particularly from the areas of probability and statistics. The Weibull distribution is considered one of the most applicable ones for reliability engineering when evaluating the reliability and failure characteristics of a machine. A mixed Weibull distribution with one sub-population with $\beta < 1$, one sub-population with $\beta = 1$ and one sub-population with $\beta > 1$ would have a failure rate plot that is identical to the bathtub curve and describes the reliability of a repairable system. Figure 3.1 shows the effect of shape parameters in failure rate distributions. Any shape parameter below 1 demonstrates a reliability improvement with time, above 1 shows a decline in reliability with time and a shape parameter of 1 shows a steady failure rate. These shape parameters are evident in the bathtub curve. The bathtub curve is a form of hazard function and is a way to model data distribution in survival analysis, also called instantaneous failure rate.

The first part of the curve is a decreasing failure rate, known as early failures, the second part is a constant failure rate, known as intrinsic failures and third part is an increasing failure rate, known as deterioration.

As mentioned, it is β that determines which stage of the bathtub curve the failure trend follows. The early failures and deterioration section are a special case non-homogenous Poisson process and can be represented by the power law process. When $\beta < 1$, the failure intensity decreases with time and hence the model describes a reli-

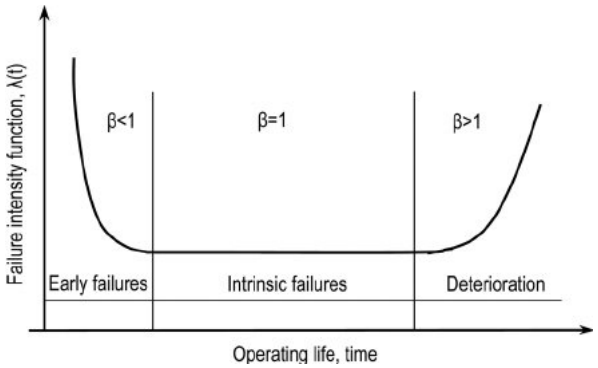


Figure 3.1: Bathtub curve and failure rates [8]

ability improvement situation. When $\beta > 1$, the failure intensity increases with time and so the model can be used to the situations in which a reliability deterioration is observed. When $\beta = 1$ in this equation the process becomes a homogenous poisson process meaning that the failures are random and can be represented with an average failure rate.

The two failure rate categories in this analysis are based on whether a component or a sub-assembly was replaced or not. In case of a replacement, the whole gearbox, a stage or just a gear/bearing may have been exchanged for a new one. If a replacement did not occur it is still considered as a failure if the equipment unit's functions are still compromised, as defined previously in this section.

3.3 Population

In the database analysed, there are alarms from the condition monitoring system of offshore wind turbine drivetrains. The population analysed in this study consists of ˜ 1200 offshore wind turbines from over 20 offshore wind farms. There are 8 different turbine models analysed. The number of failures analysed in total is 1044. For confidentiality reasons the exact nominal power, blade size or drive train configuration of the turbine type used in this analysis cannot be provided. However it can be stated that they are all modern multi MW scale turbine, ranging from 2 to 10MW. The drive train types also vary and include high speed machines, medium speed machines and direct drive machines (type C, E and D as described in 2.1).

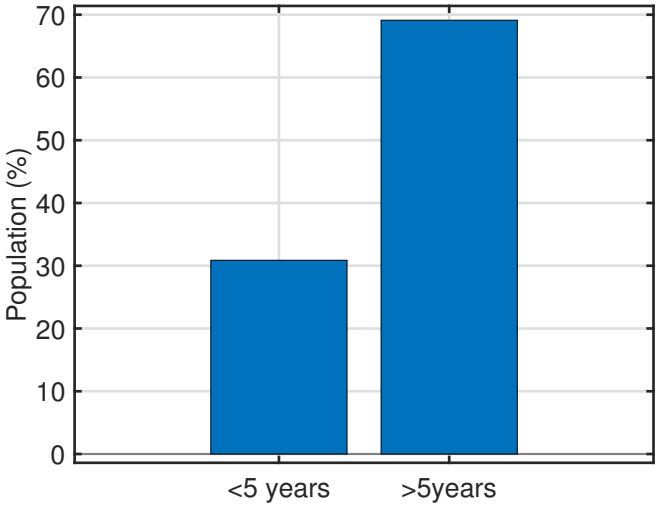


Figure 3.2: Operating years of analysed population

The years of installation of the analysed population are shown in Figure 3.2. It can be observed that almost 70% of the turbines is installed for more than 5 years. It should be noted that all wind farms are operating for at least a year.

Since the population includes different turbine models and sizes, a breakdown of these is presented in Figure 3.3. Around 25% of the population is rated at more than 5MW.

Finally, the different drivetrain types are shown in Figure 3.4. Almost 80% consists of geared machines, which is representative of the current state of the market since manufacturers are just starting to introduce direct drive and low speed generators offshore. The geared machines have up to 3 stages, which can be either 1 planetary and 2 parallel stages or 2 planetary and 1 parallel stage. The generator types include both asynchronous and synchronous permanent magnet configurations.

3.4 Component Failure Rates

The main drivetrain parts are the main bearing, the gearbox and the generator. The various drivetrain components along with their associated drivetrain parts are shown in Table 3.1.

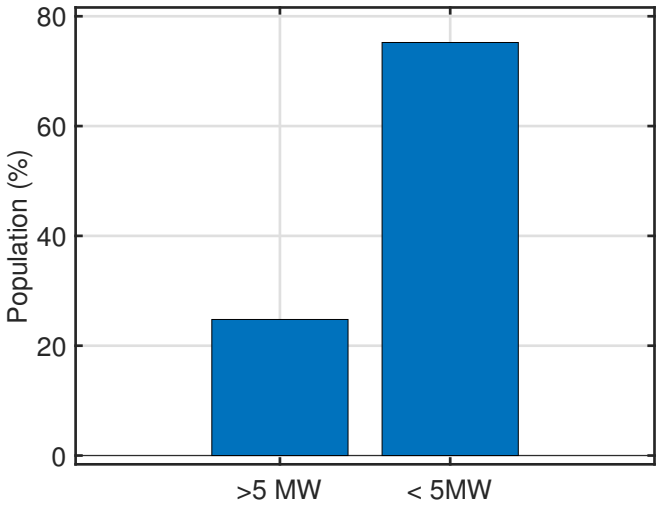


Figure 3.3: Power rating of analysed population

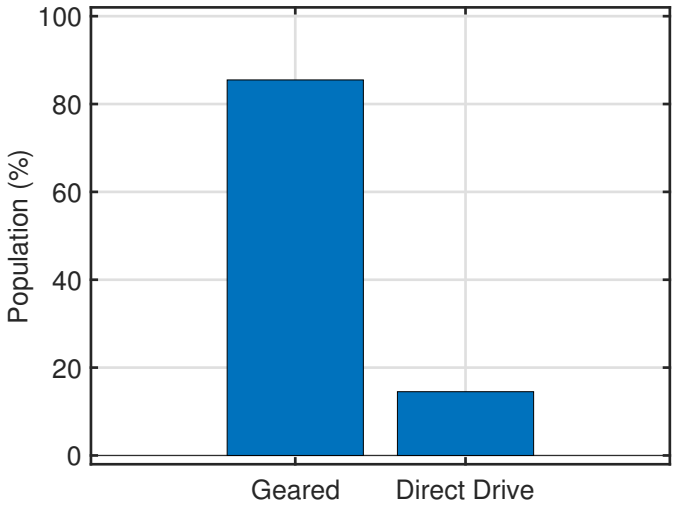


Figure 3.4: Drivetrain configuration of analysed population

Drivetrain Component	Drivetrain Location
HS bearings	Gearbox
HS pinion-wheel	
HS stage/shaft	
IMS bearings	
IMS pinion-wheel	
IMS stage/shaft	
Planet bearings	
Ring gear	
Sun pinion-wheel	
Planet stage/shaft	
Planet wheel	
Idle pump	
Generator bearings	Generator
Main bearings	Main bearing

Table 3.1: List of drievtrain components and corresponding locations

The failure rates of the various drivetrain components for the whole population of turbines is shown in Figure 3.5. It can be observed that the top fault locations for the population analysed are gearbox high speed bearings, generator bearings and main bearings.

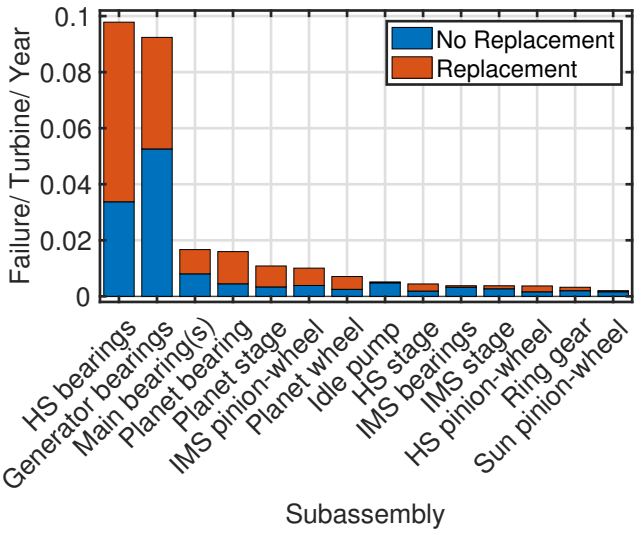


Figure 3.5: Failure rates for drivetrain fault locations

The total failure rate of the gearbox is calculated to be around 0.21 failures per turbine per year. This is close to the mean gearbox failure rate, within the range given

in [167], where all currently existing wind turbine reliability datasets are compared. Nevertheless, the weighted average failure rate for offshore turbines is more than 0.5 in [167]. This indicates that there can be some discrepancy between different reliability studies, depending on location, wind turbine type and power rating. Wind turbine gearboxes do not have as high failure rates as other sub-assemblies, such as pitch system which can reach to more than 1 failure per turbine per year. However, they are very critical in terms of downtime.

3.5 Failure Rate by Operational Year

The component failure rates with time are examined in [173]. The deterioration phase dominates the results for the gearbox, which is justified by the fact that the gearbox is a sub-assembly of mature technology for which no real development can be implemented once installed. The Power Law Process model could not be implemented on the gearbox reliability data in [173] due to the rejection of the goodness-of-fit test for any aggregation of the failure data and the lack of a sufficient number of failures. On the contrary, in the case of the generator the intensity of failures is either improving or constant. The difference in trend between gearbox and generator is explained due to the difference in maturity of technology. Designers and operators have made evident efforts to improve the reliability of the wind turbine generators compared to gearboxes.

Figure 3.6 shows the drivetrain failure rate per operational year of the turbine population. A high number of failure rates occurs around the 4th year of operation, which agrees with what is presented in another offshore wind turbine study [3]. That spike could be attributed to end fo warranty reasons. However, it should be noted that that the population of turbines analysed with over 15 years is smaller than the overall population with turbines only coming from 2 wind farms. By that point the assets have not reached deterioration state.

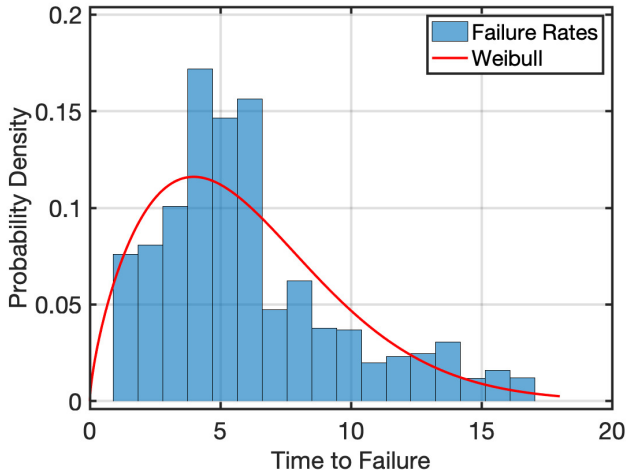


Figure 3.7: Weibull distribution of failures over time

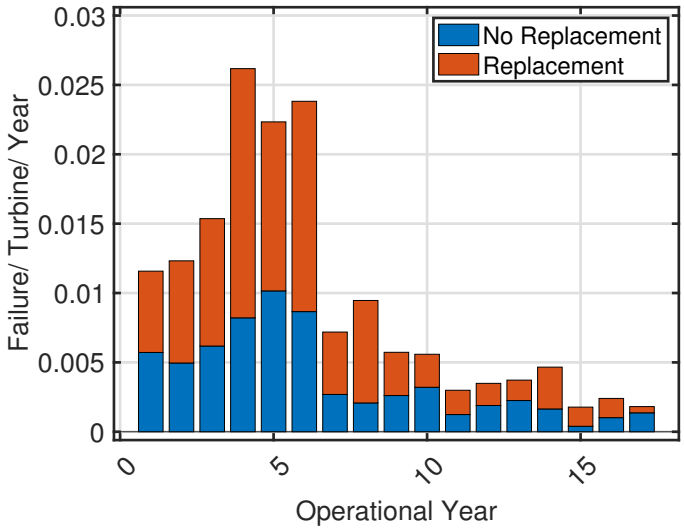


Figure 3.6: Failure rates by operational year

Based on the analysed data, a Weibull distribution is fit, which is shown in Figure 3.7. The scale parameter is 6.73 and the shape parameter is 1.69.

3.6 Failure Modes

A failure mode is defined as the effect by which a failure is observed on the failed item [170]. The failure modes of the drive train components that fail the most, gearbox

high speed bearings and generator bearings, are shown in Figures 3.8, 3.9. Both for gearbox high speed bearings and for generator bearings, outer race defects is the failure mode that contributes most to faults. Inner race defects follow, whereas rolling element defects are not encountered that often. It should be noted that these results are related to geared wind turbines.

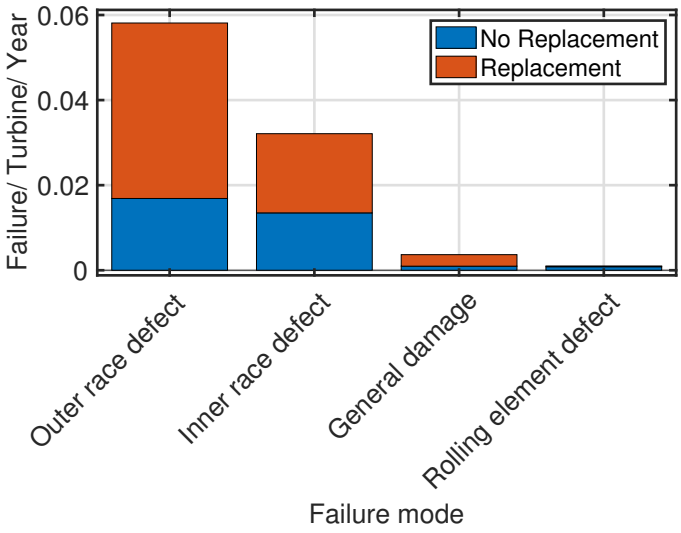


Figure 3.8: High speed bearings failure modes

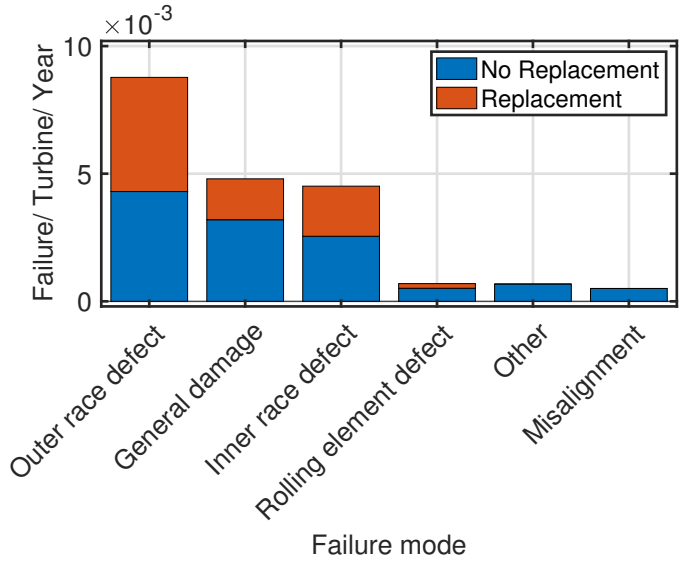


Figure 3.9: Generator bearings failure modes

3.7 Failure Rates by Turbine Size

The failure rates based on turbine size are shown in Figures 3.10, 3.11. For turbines with power rating >5MW generator bearings are the major failure contributor, followed by planet wheels and main bearings. Turbines rated lower than 5MW fail mainly due to gearbox high speed and generator bearings.

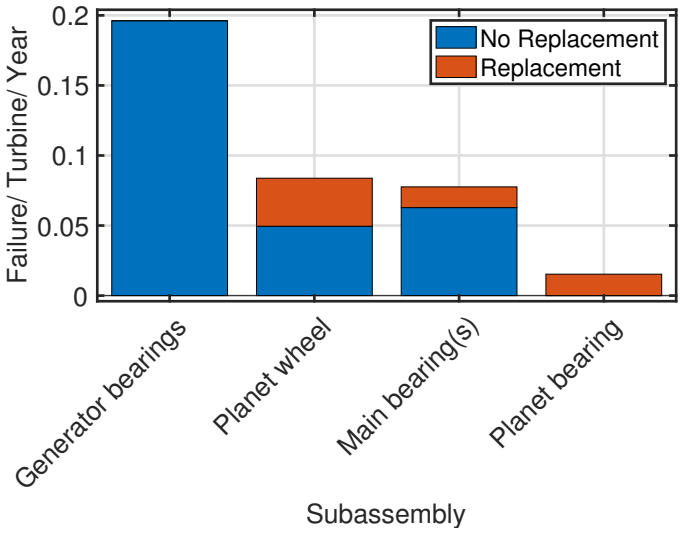


Figure 3.10: Failure rates for turbines rates >5MW

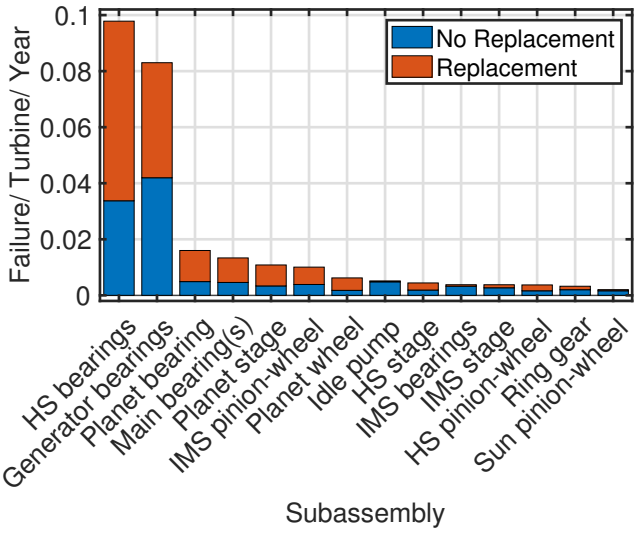


Figure 3.11: Failure rates for turbines rates <=5MW

3.8 Failure Rates by Turbine Configuration

Failure rates are further analysed in this section based on drivetrain configuration. Overall failure rates depending on drivetrain types are shown in Figure 3.12. Geared turbines have higher failure rates, however it should be highlighted that the analysed population consists mainly from geared turbines, whereas there are not a lot of operational data from direct drive turbines.

The geared turbines' failure rates are shown in Figure 3.13 and resemble the overall failure rates. It should be pointed out that the direct drive failure rates available consist only from mechanical ones, so only main bearing failure rates are recorded in this database. This is shown in Figure 3.14.

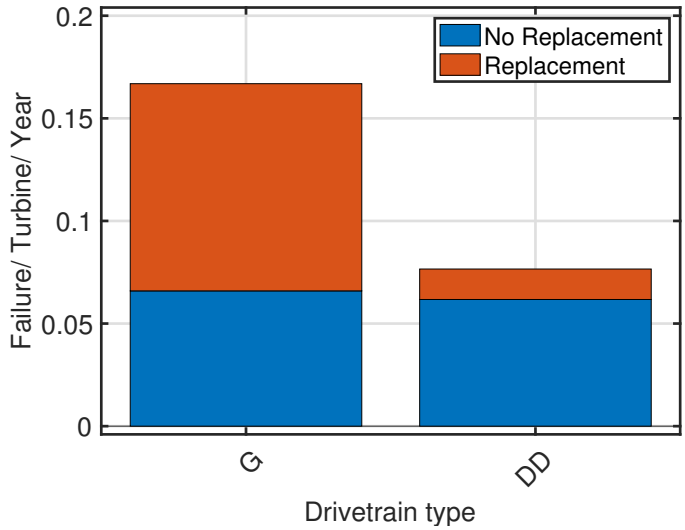


Figure 3.12: Failure rates based on drivetrain type

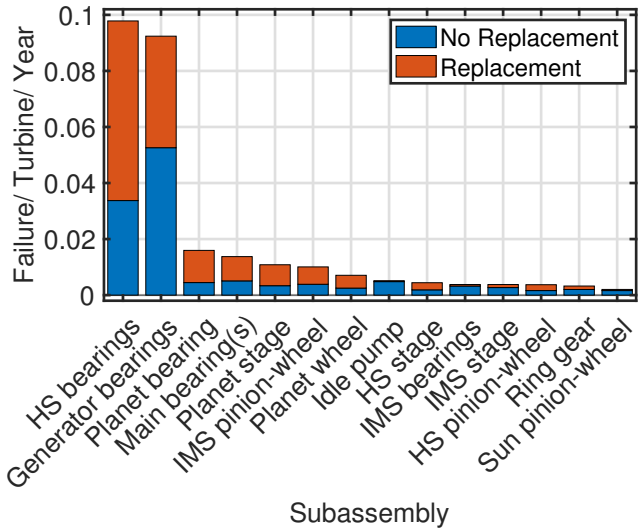


Figure 3.13: Geared turbines failure rates

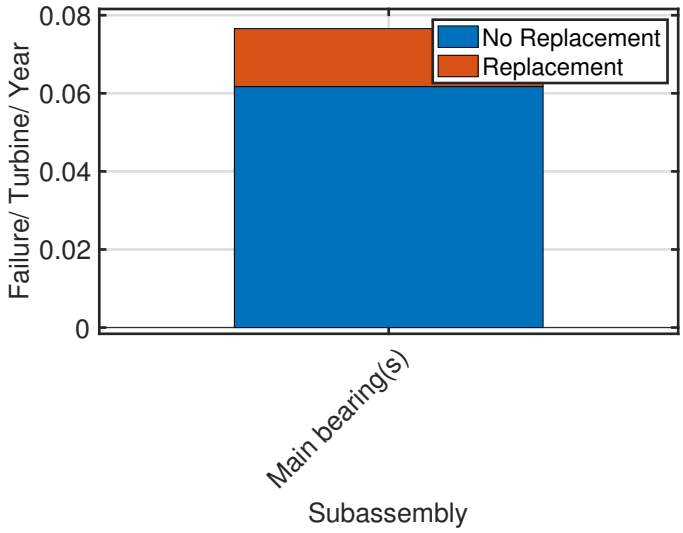


Figure 3.14: Direct drive turbines failure rates

3.9 Replacement categories

In this section, the various replacement categories are further analysed. The replacement categories are further broken down into full gearbox replacement or just a sub-assembly replacement. The third category is just an inspection or repair. It can be seen in Figure 3.15 that the majority of maintenance actions are sub-assembly replacements,

followed by repairs.

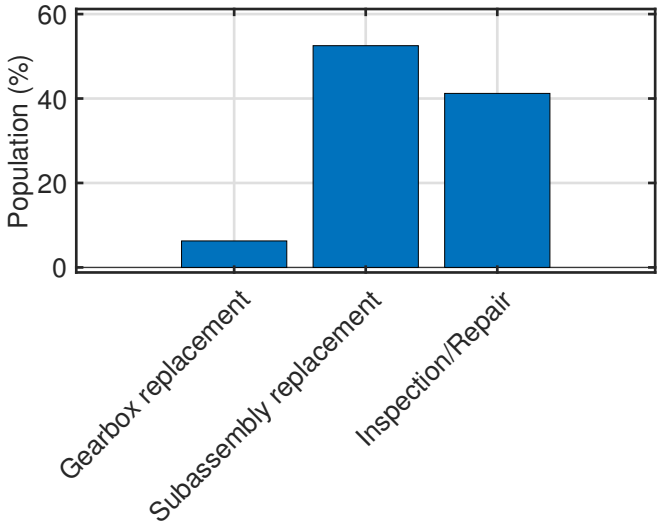


Figure 3.15: Gearbox replacements

Since full gearbox replacements are the most significant in terms of cost, a breakdown of the sub-assemblies that cause the most gearbox replacements are shown in Figure 3.16. Most gearbox replacements happen due to faults in the planetary stage, especially in the planet bearings. This is expected, since planet fault replacement require complete disassembly of the gearbox.

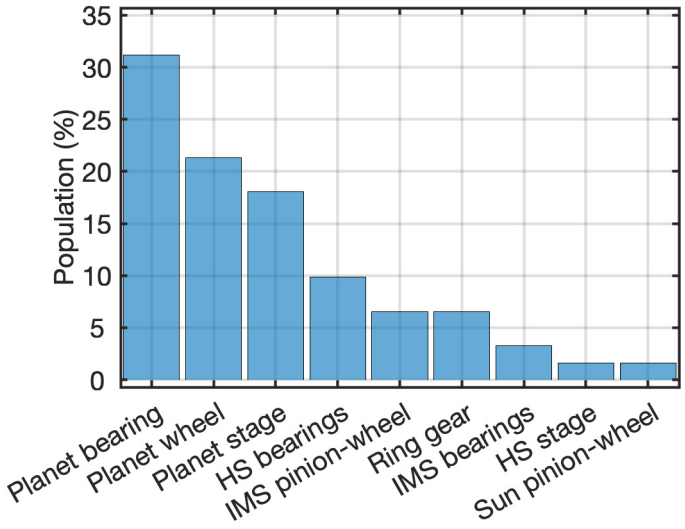


Figure 3.16: Fault locations for replaced gearboxes

3.10 Conclusions

In the current chapter, the reliability of wind turbine drivetrains is studied through a large population of offshore wind turbine failure data. The top fault locations in the populations analysed are high speed bearings and generator bearings. The top failure modes for bearings are outer race defects. Also, it is observed that the failure rate is high in the first operational years and with a higher rate around year 4, which does not strictly follow the bathtub curve.

In terms of turbine size most failures in turbines rated higher than 5MW occur in generator bearings. Geared and gearless machines also exhibit differences, as expected. The main source of mechanical failures for direct drive machines are main bearings.

High speed bearings seem to be a critical component in wind turbine drivetrains, which is in accordance with previous studies [6].

It should be emphasized, that as discussed in [167], failure rates between different datasets can have some discrepancy, depending on the volume of data, location, power rating and turbine configuration.

The results of this chapter give strong indication on how integrated CBM strategies should be designed. However, a systematic failure mode effect and criticality analysis and risk analysis should also be performed in combination in order to generate the highest value from CBM.

Chapter 4

On the Use of SCADA Data for Wind Turbine Gearbox Failure Prediction

4.1 Chapter Contribution

SCADA systems have proven to be an effective way for wind turbine condition monitoring. Based on the literature as analysed in Chapter 2, faults can be detected using SCADA. However, there is not so much information available on the temperature behaviour at various stages of the gearbox and how it can develop in different wind turbines leading up to a component catastrophic failure. Real failure example information is rarely available which makes it challenging to compare and validate models. This chapter aims to answer the following research question:

"Can SCADA systems be used for wind turbine gearbox fault detection and isolation?"

The contributions of the chapter are as follows:

1. Give an insight into the temperatures inside the gearbox, both for normal operation and operation close to a catastrophic gearbox failure.

2. Demonstrate various data analysis methods for wind turbine gearbox failure pre-

diction using SCADA

3. Develop both an individual and a global normal behaviour model based on a large population of turbines.

4. Propose a framework for early incipient fault detection and validate it against both onshore and offshore wind turbines.

All models are tested and validated against real wind turbine data from operating wind turbines with respect to a planet bearing failure. Planet bearings may not have the highest failure rate out of the whole drivetrain, but they are accounted for the most gearbox replacements, as shown in Figure 3.16.

4.2 Data Pre-processing

In order to remove outliers, values that are more than three scaled median absolute deviations away from the median of each sensor are discarded. Furthermore, full curtailment events are not included in the modelling process, so data points that have wind speed more than cut-in and power equal or less than zero are removed. Curtailment is a reduction in the output of a generator from what it could otherwise produce given available resources, typically on an involuntary basis [176]. However, there is not any further curtailment/de-rating strategy information provided, which could affect the modeling results. Data points where maximum wind speed has reached more than the cut-out wind speed (around 25 m/s) are also filtered out because beyond this wind speed the turbine is stopped. These points will not fit any pattern, thus cannot be taught to the model. Finally, standardisation is perfomed using zscore.

The cluster filter is applied on the training data and aims to remove outliers depending on the operating conditions of the wind turbine. A simple threshold does not take into account the nonlinear operational characteristics of the system. A multivariate outlier detection approach based on Mahalanobis distance is used in [177]. A similar approach is extended and used in [178], by dividing the data based on operating power and temperatures ranges. This paper utilizes agglomerative hierarchical

clustering [179]. Essentially, the distance between every pair of objects in a data set is computed. This information is then used in the linkage function which determines how the objects in the data set should be grouped into clusters that form a binary hierarchical cluster tree. The distance in this paper is calculated in the Euclidean space and the inner squared distance is computed using Ward's algorithm.

The distance is calculated for each data vector in the training data set from its cluster centre. The Mahalanobis distance values can be estimated by a loglogistic distribution as elaborated in [178] and data below the probability threshold of 2.5% are filtered out. The distribution of the Mahalanobis distance values of the training data set for an operating wind turbine is shown in Figure 4.1. It can be observed that the values can be estimated by a loglogistic probability distribution function.

A probability threshold of 2.5% is chosen. The cluster filter is used only on the training data set, and therefore false alarms due to curtailment could occur in the implementation stage. It is suggested that the condition monitoring process is blocked during power curtailment, which is available in most modern SCADA system logs. This information was not available in this case study though.

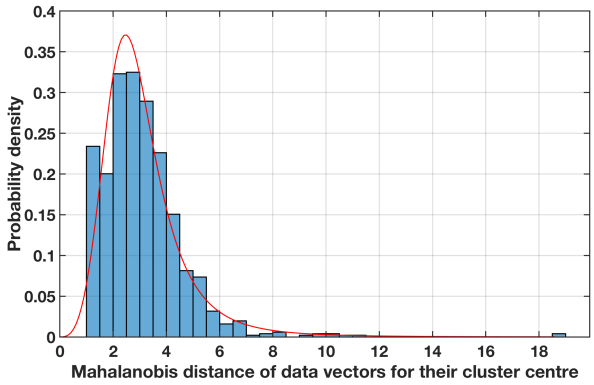


Figure 4.1: The histogram and probability density function fit for Mahalanobis distance values of data vectors from its cluster centre.

The power curve before and after pre-processing is shown in Figure 4.2.

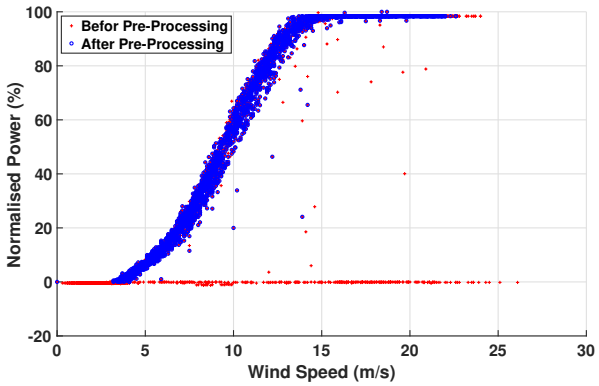


Figure 4.2: Power curve of wind turbine 1 before and after data pre-processing.

4.3 Trending

Wind turbines operate under variable conditions and trends are not always easily interpretable, since a change in the value of a SCADA parameter is accordingly not necessarily evidence for a fault. However, it's still a very straightforward way of monitoring the condition of assets, therefore quite a few studies have used trending to detect early signs of degradation. One simple way is to monitor the change of ratios of SCADA parameters and how they change over time.

The proposed condition monitoring strategy is based on [14]. The correlation of SCADA data variables is modeled using a polynomial equation. The coeffients of the model for the historic data and the tested data are calculated and a criterion is established based on the difference between the historic model and the calculated relationship in the new acquired data whose condition is being monitored.

Let x, y be two SCADA variables. Their relationship can be modeled based on Eq 4.1.

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + ... a_k x_i^k \tag{4.1}$$

Where i is the number of samples of SCADA variables and k is the order of the equation.

The coefficient matrix can be calculated based on least squares

$$A = (X^T X)^{-1} X^T Y \tag{4.2}$$

Where

$$X = \begin{bmatrix} 1 & x_1 & ... & x_1^k \\ 1 & x_2 & & x_2^k \\ ... & & & \\ 1 & x_n & & x_n^k \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}, A = \begin{bmatrix} a_0 \\ a_1 \\ ... \\ a_k \end{bmatrix}$$

The condition monitoring criterion can therefore be established by Eq 4.3.

$$c = \frac{\int_{x_{min}}^{x_{max}} \left| \sum_{j=0}^{k} (a_j - b_j) x^j dx \right|}{x_{max} - x_{min}} \tag{4.3}$$

4.4 Clustering

In order to automate the classification between normal and abnormal operations, clustering have been used for CM using SCADA.

This can be achieved using Principal Component Analysis (PCA) and manual interpretation of the reduced dimension of results. Semi-supervised methods like one-class SVMs can be applied on the data and the outliers are classified as anomalies as described in Chapter 2.

Another method commonly applied in the literature is Self Organising Maps, where a useful visualisation can be achieved and information can be represented. The condition of the healthy turbine can be represented on the map. When new data points are acquired, the distance of time series data points from the centers of the Self Organising Map can indicate the evolution of fault. For example, the Euclidean distance between the input vector and its cluster center can be used as the health indicator that represents the extent of similarity between present condition and baseline 4.4. This is called Minimum Quantisation Error (MQE).

$$MQE = \sqrt{e_1^2 + e_2^2 + ... + e_n^2} \tag{4.4}$$

Where n is the number of features and e the Euclidean distance between each feature

and the centre.

4.5 Normal Behaviour Modelling

It has been discussed in the literature that gearbox fault detection -assuming that a gearbox fault is reflected in the SCADA data somehow- is a multivariate problem that does not depend on clear temperature trends, but on trends with respect to other variables in the system. Consequently, normal behaviour modeling has been widely used for SCADA data analysis and failure prediction. This is achieved using regression based methods.

Models can either be parametric (e.g. polynomial curve) or non-parametric (e.g. ANNs). ANNs are one of the algorithms that have shown most promising results. ANNs have the ability to model complex relationships by finding patterns between inputs and outputs. That's what makes them so useful for models using SCADA parameters, since the relationships between the variables can be nonlinear.

4.5.1 Neural Network Model Training

The weights are firstly set using random initialisation. The input values are propagated through the network and the error between the calculated output and the desired output is fed back to the network. The weights and biases of the network are tuned to optimize network performance. The network performance function used is the mean squared error. The Levenberg-Marquardt back propagation training algorithm is chosen. This optimisation algorithm has the combined advantage of Newtons method, which converges rapidly and Gradient descent method, which is assured to converge. Generalisation of the network is improved using regularisation, therefore by adding a term that consists of the mean of the sum of squares of the network weights and biases. One of the problems that occur during neural network training is called overfitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large. Balancing bias and variance helps in avoiding overfitting [180], [181]. Using this regularisation performance function causes the

network to have smaller weights and biases, and this forces the network response to be smoother and less likely to overfit.

An ANN of 1 layer is used, which is the most common shallow structures. The hyperparameters chosen to be optimised are the number of neurons, and the learning rate. The ranges of the hyperparameters to be used as input in the Bayesian optimiser are chosen within ranges as suggested in [182]. The optimal regularisation parameters are found using the Bayesian framework as described in [183], [184]. Holdout cross validation of training and validation sets with a ratio 1/3 is used inside the objective function. This method is explained in Section 2.4.4.

The models are trained using only the data furthest in time before the component failure, which are labeled as "healthy" (for the case study presented periods are given in Figure 4.4).

The "healthy" data is split randomly into 80% and 20%, with only the first subset being used in the training and optimisation process as described in the previous paragraph, and the rest is used to assess the model performance on unseen normal data. Once the training process has been completed, the model is used to predict the rest of the test subsets of the different periods before the bearing failure. The error between the modeled and the actual values is used as a metric of abnormality of the gearbox operation.

4.5.2 Monitoring Framework

An anomaly detection condition monitoring framework that can detect incipient gearbox failures is proposed in this section. The framework is based on the assumption that in case of abnormal behaviour, the normal behaviour model prediction output will differ quite significantly from the measured signal. Therefore, the errors will increase. An extensive review of anomaly detection methods is given in [185]. There exist different techniques that can be grouped into classification-based, nearest neighbor-based, clustering-based and statistical. This study focuses on statistical techniques as suggested in [186], because of the nature of the errors. Assuming that the error is defined as the difference between the measured variable from the SCADA system and the pre-

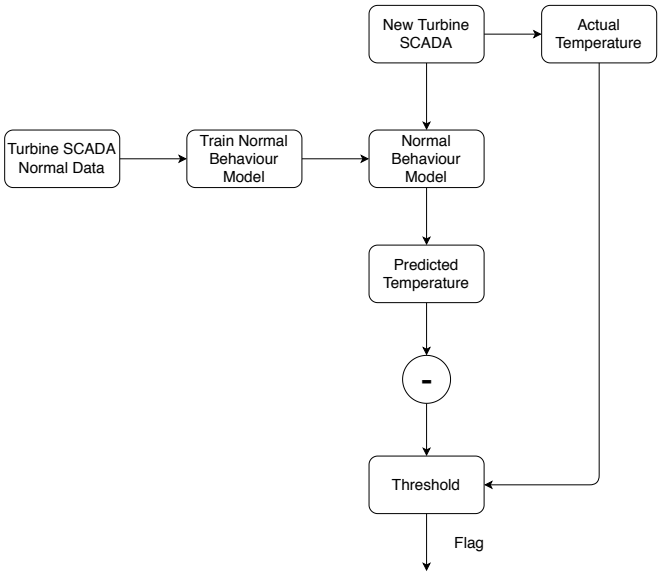


Figure 4.3: Normal Behaviour Modelling Framework

dicted ANN model, then in a successfully trained model the errors should be normally distributed around zero.

The prediction error from the training phase is used to estimate the parameters of the underlying statistical process, using a maximum likelihood estimator. These parameters are used to set the thresholds for anomaly detection. The thresholds form the upper and lower bounds of the normal range of the SCADA data. Instances that have a low probability of being generated from the trained model, based on the applied test statistic, are classified as anomalies. The probabilistic assumption made is that prediction errors which have a probability of being generated by the trained model smaller than 0.01% are considered an anomaly, as suggested in [186]. The choice of the probability influences the sensitivity of the system and can be adjusted if required, depending on whether false alarms or misdiagnosis is more important to the operator. In order to reduce variance, the normality thresholds are calculated based on 1 day averages of the 10 minute resolution data. Figure 4.20b shows an example of the boundary between normal and abnormal behavior, in terms of the probability density function.

The process described is shown in Figure 4.3.

4.6 Case Study

The aim of this case study is to find patterns at different time periods prior to a gearbox failure based on SCADA channels. There are available data from 199 turbines in 24 different wind farms, both onshore and offshore. They are the same turbine model that failed from the same gearbox failure mode, which will be analysed in the following paragraphs.

For 10 turbines there were continuous data from 18 months before failure until within 1 month of failure occurrence.

For the rest of the 189 turbine the data provided is quite sparse; for every turbine there are 10 min SCADA for 1 week for specific time periods with respect to a component failure. Those specific time periods are:

- 18 months
- 1 year
- 6 months
- 3 months
- 2 months
- 1 month

Only about 1/4 of the turbines have data from 2 and 3 months before failure. Also, for around 1/7 of the turbines there are no available data for more than a year before failure. All turbines have available data for 1 year, 6 months and 1 month before failure. The data acquired 18 months before failure will be labeled as "healthy", since this is the data that is furthest away in time before the failure and based on some vibration signals for a population of turbines, the gearbox seems to be in a normal operating state. This subset of "healthy" data will be used in training the normal models in most of the algorithms presented.

A structure of the data and the time periods used for each trained model are shown in Figure 4.4.

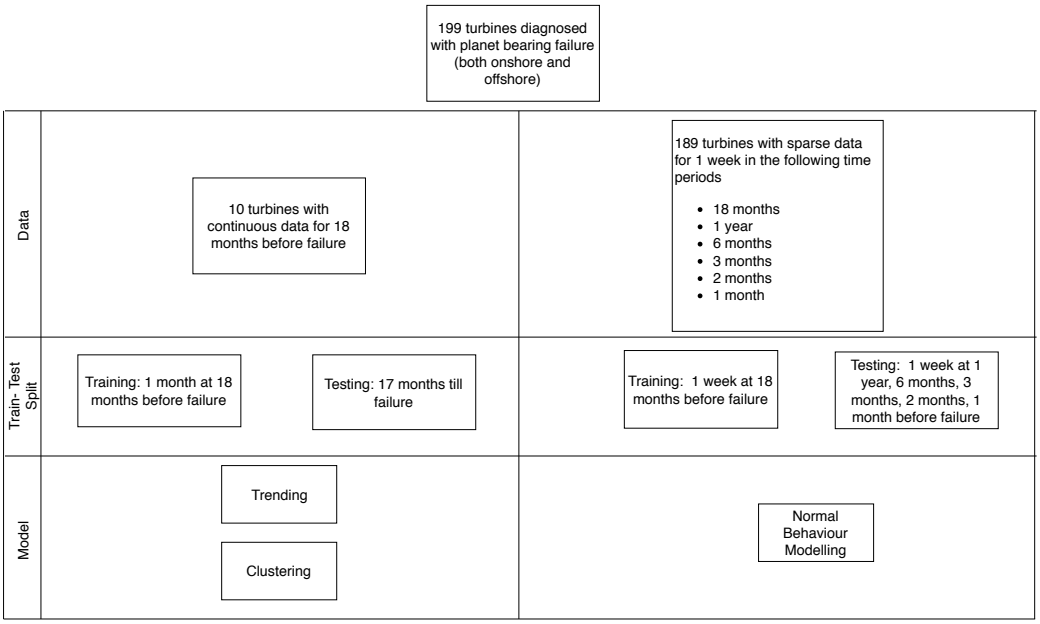


Figure 4.4: Data description for SCADA case studies

The gearbox examined has a structure commonly found in offshore wind turbines, where high step-up ratios and compactness are required. It consists of two planetary stages and one parallel stage. The main shaft (hollow shaft) is connected to the planet carrier of the second planetary stage and the high speed stage of the gearbox is coupled to the generator. The ring gears of the planetary stages are fixed. The failure mode studied occurs on a first planetary stage planet bearing. It starts on the inner race way with debris eventually effecting the outer raceway. Experience has shown that this type of failure mode can have quite a long period of initiation and propagation until catastrophic failure- in the order of magnitude of months- which depends of course on the operating and loading conditions of each turbine.

An indicative gearbox drawing is given in Figure 4.5.

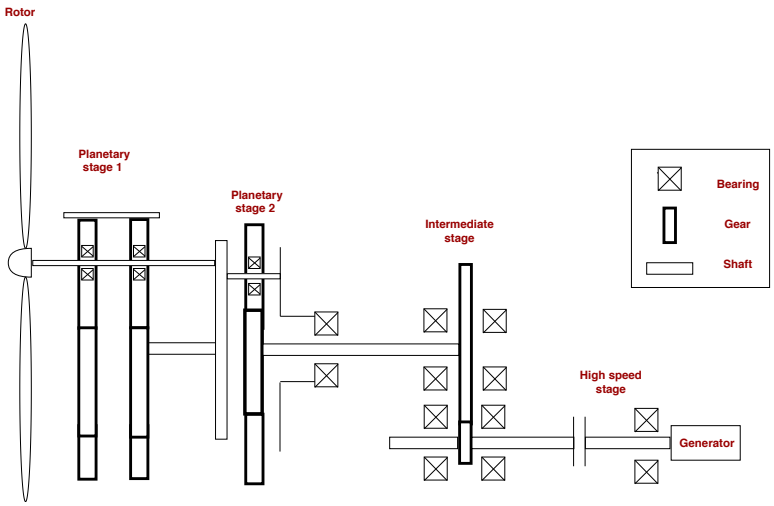


Figure 4.5: Gearbox high level drawing with bearing positions.

The gearbox sensors installed throughout the gearbox are given in Table 4.1. The exact locations of the sensors has not been validated by the authors, but the sensor names give an indication of their position in the gearbox. It seems that the temperature sensors are mostly either measuring the high speed stage, the hollow shaft or the oil. No sensor is installed in the planetary stage.

Sensor Name	**Variable Name**
Gearbox Oil Temperature Bottom Level	T_{obl}
Gearbox Oil Temperature Higher Level	T_{ohl}
High Speed Shaft Temperature Rotor End	T_{hir}
High Speed Shaft Temperature Generator End	T_{hig}
High Speed Shaft Temperature Middle	T_{him}
Hollow Shaft Temperature Rotor End	T_{hor}
Hollow Shaft Temperature Generator End	T_{hog}
Nacelle Temperature	T_{nac}
Ambient Temperature	T_{amb}
Generator Speed	N_{gen}
Rotor Speed	N_{rot}
Wind Speed	V_{win}
Electrical Power	P_{el}

Table 4.1: List of gearbox SCADA sensors.

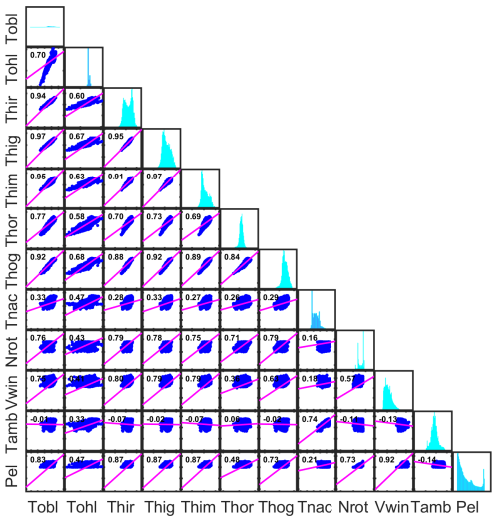
4.6.1 Data Visualisation

In order to understand the different variables of the system and build more informed models, an insight to the system is given initially through visualisation. This will be presented in this section.

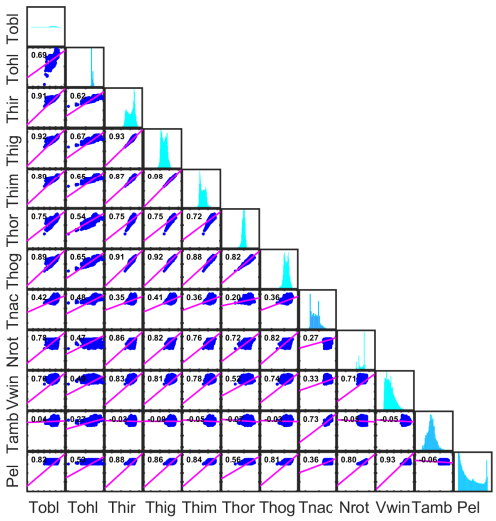
Correlations between SCADA variables

In order to model the various variables, their relationships need to be understood. Figures 4.6a-4.7d shows the correlations for the onshore and offshore turbines for the different time periods before failure. The temperatures from the bearings on the same shaft are highly correlated. The rotor speed has also a higher correlation with the oil temperature closer to failure, probably due to the correlation of the flow rate with heat extraction. It is interesting to notice that the ambient temperature is not correlated with most of the temperatures inside the gearbox, probably due to the presence of a cooling system. The rotor speed is expected to have a positive correlation with the gearbox temperature, but that correlation seems to generally increase close to the presence of a fault.

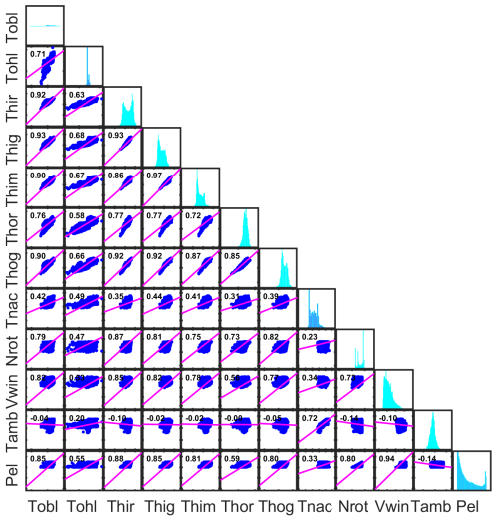
Most of the variables have uniform bell shaped or bimodal distributions. It's worthwhile pointing out that some temperatures seem to change distribution shape towards failure, especially onshore. These are mainly the oil temperature at the bottom level and the bearing temperatures closer to the generator end of the shafts.



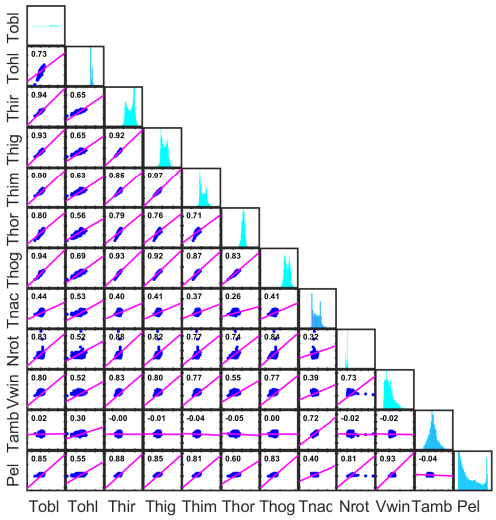
(a) Correlation plot (18 months before failure).



(b) Correlation plot (1 year before failure).

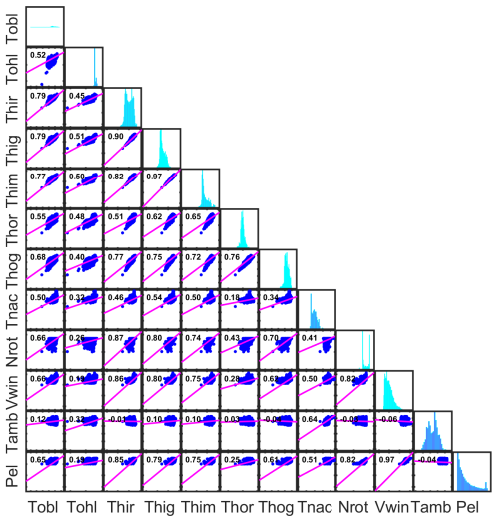


(c) Correlation plot (6 months before failure).

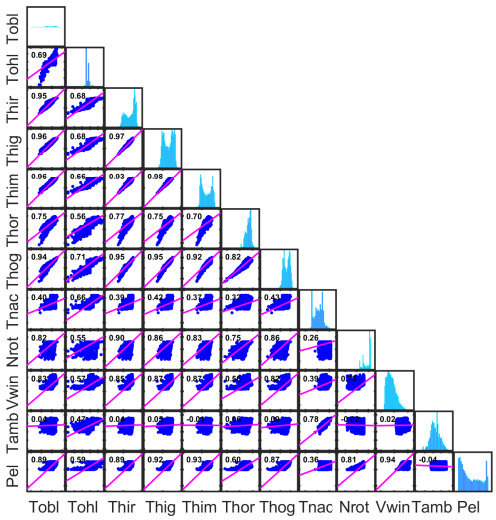


(d) Correlation plot (1 month before failure).

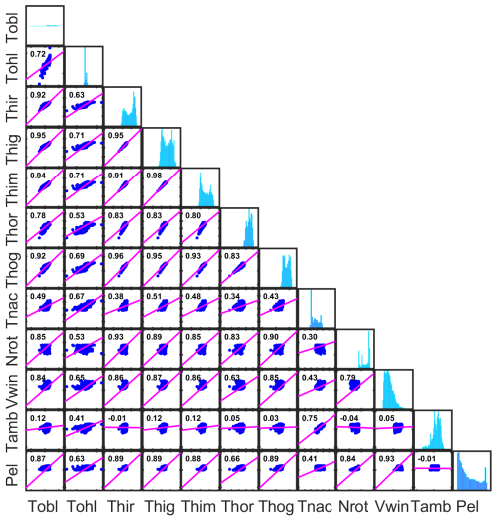
Figure 4.6: Correlation plots for onshore wind turbines



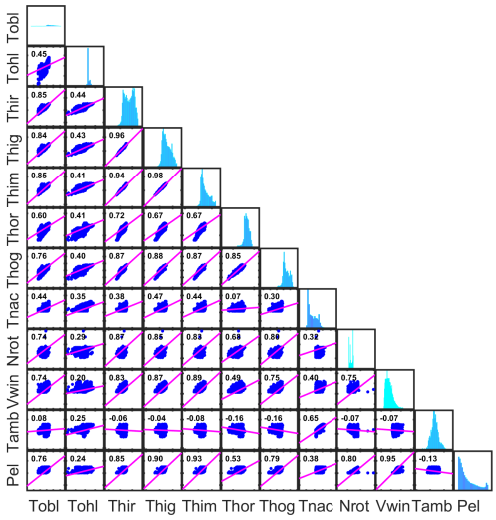
(a) Correlation plot (18 months before failure).



(b) Correlation plot (1 year before failure).



(c) Correlation plot (6 months before failure).



(d) Correlation plot (1 month before failure).

Figure 4.7: Correlation plots for offshore wind turbines

Temperature Difference at Various Stages of the Gearbox

The temperature difference between the channels at various stages across the gearbox are examined for the different time periods before failure. Figures 4.8 and 4.9 show the gearbox temperatures and their evolution with respect to time before failure for onshore and offshore wind turbines respectively. The hollow shaft temperatures seem to be higher but with a smaller variation than the high speed shaft temperatures. The oil temperatures are lower than the bearing temperatures for both onshore and offshore turbines.

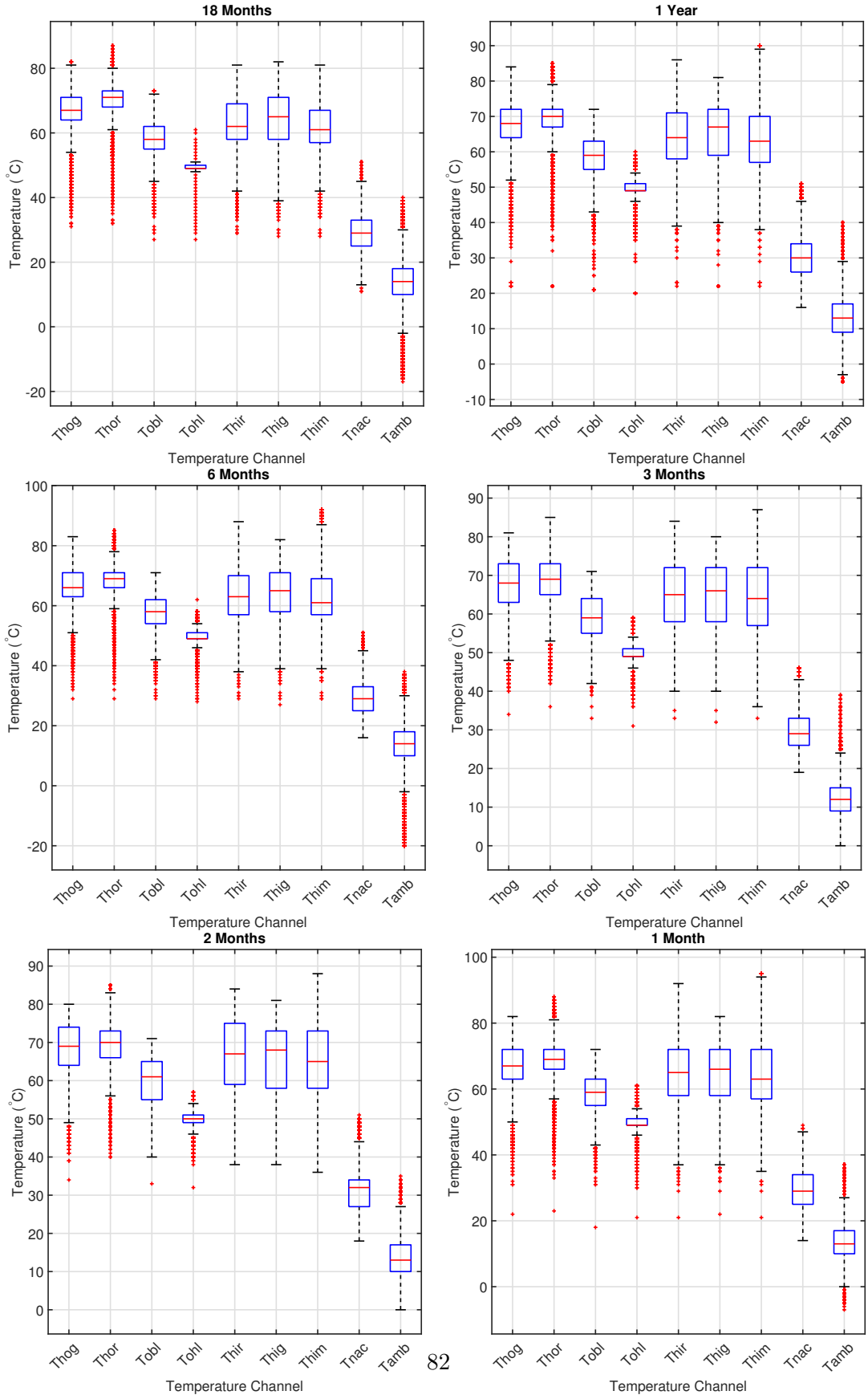
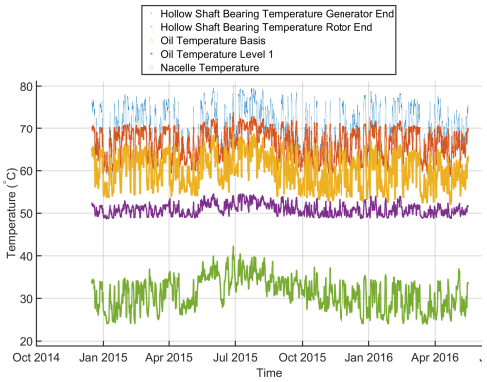
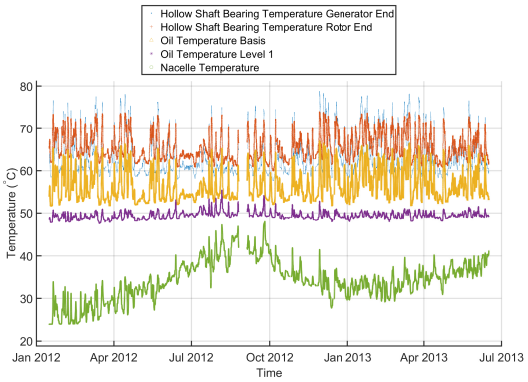


Figure 4.8: Boxplots of temperature variables, for different time periods before failure (onshore turbines).

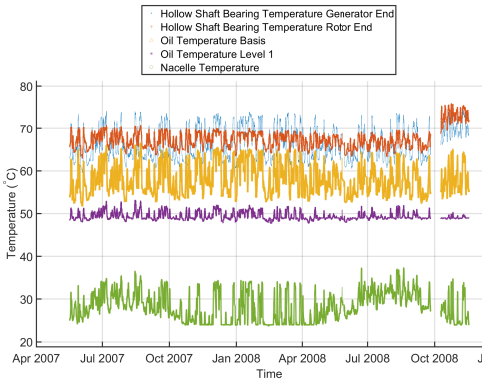
Figures 4.10a- 4.11d show the evolution of gearbox temperatures over time for 2 onshore and 2 offshore turbines. A moving average filter is applied to make the parameter ranges more interpretable. The evolution is different for each turbine which means that it is hard to assume a common fault propagation behaviour for all of them. The high speed shaft middle bearing temperature seems to be affected progressively before failure in a few turbines.



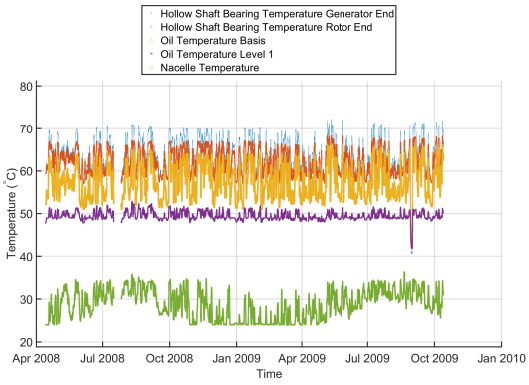
(a) Nacelle temperatures, oil temperatures and temperatures in the hollow shaft bearings (Turbine 1).



(b) Nacelle temperatures, oil temperatures and temperatures in the hollow shaft bearings (Turbine 2).



(c) Nacelle temperatures, oil temperatures and temperatures in the hollow shaft bearings (Turbine 3).



(d) Nacelle temperatures, oil temperatures and temperatures in the hollow shaft bearings (Turbine 4).

Figure 4.10: Gearbox temperature evolution leading up to a bearing failure for 4 different turbines

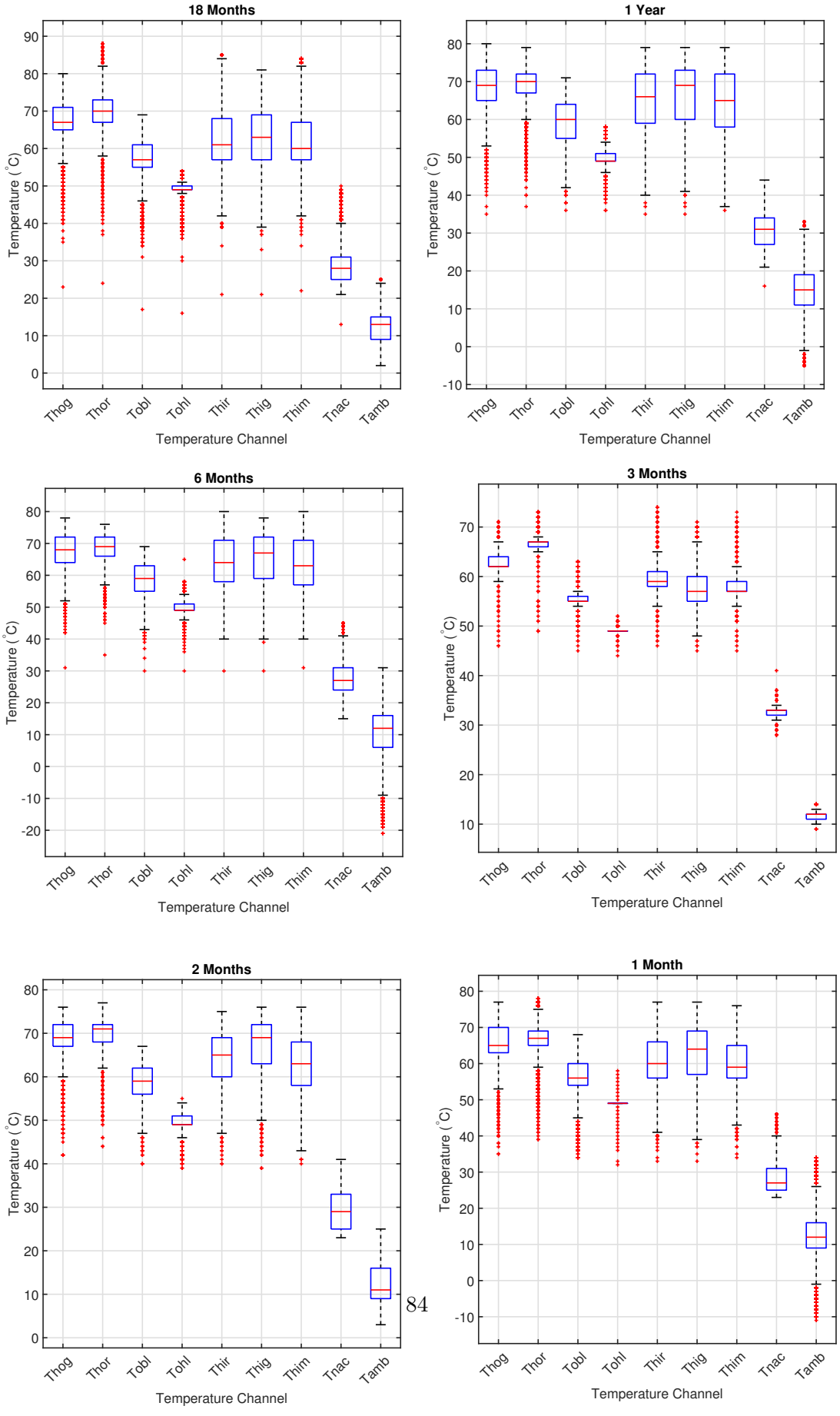
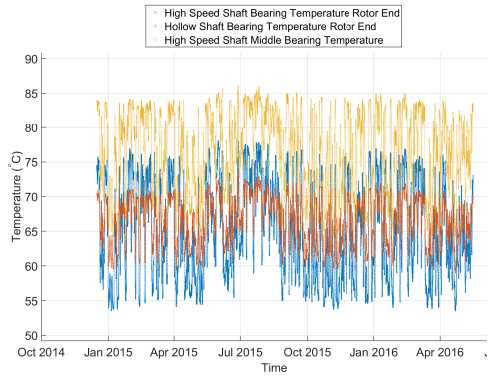
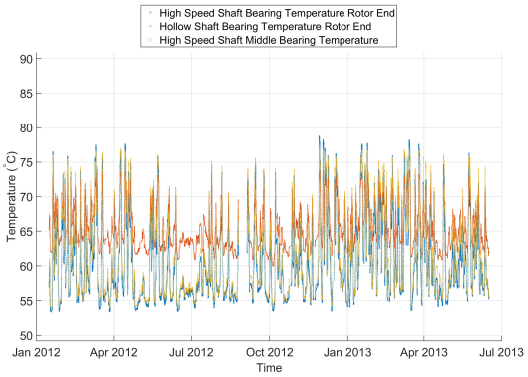


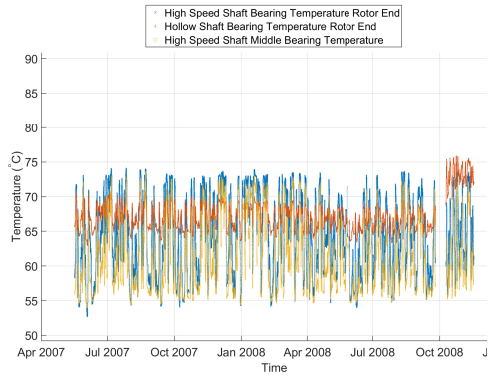
Figure 4.9: Boxplots of temperature variables, for different time periods before failure (offshore turbines).



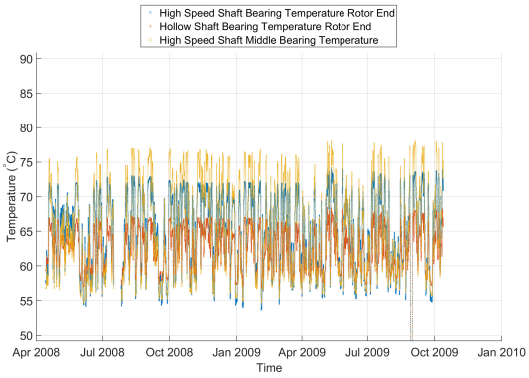
(a) High speed shaft bearing temperatures compared to hollow shaft (Turbine 1).



(b) High speed shaft bearing temperatures compared to hollow shaft (Turbine 2).



(c) High speed shaft bearing temperatures compared to hollow shaft (Turbine 3).



(d) High speed shaft bearing temperatures compared to hollow shaft (Turbine 4).

Figure 4.11: Temperature evolution of high speed shaft and hollow shaft for 4 different turbines leading up to a bearing failure.

4.6.2 Trending Results

Some indicative results for trending analysis are shown in Figures 4.12, 4.13 for an onshore (left) and an offshore (right) wind turbine. The condition monitoring criterion parameter c is also shown. According to literature, trending and monitoring c can often not give consistent results [108], which was also the case in the present case study. The hollow shaft bearing temperatures seem to exhibit some trend in Turbine 3, which could be expected since these are the measurements close to the fault.

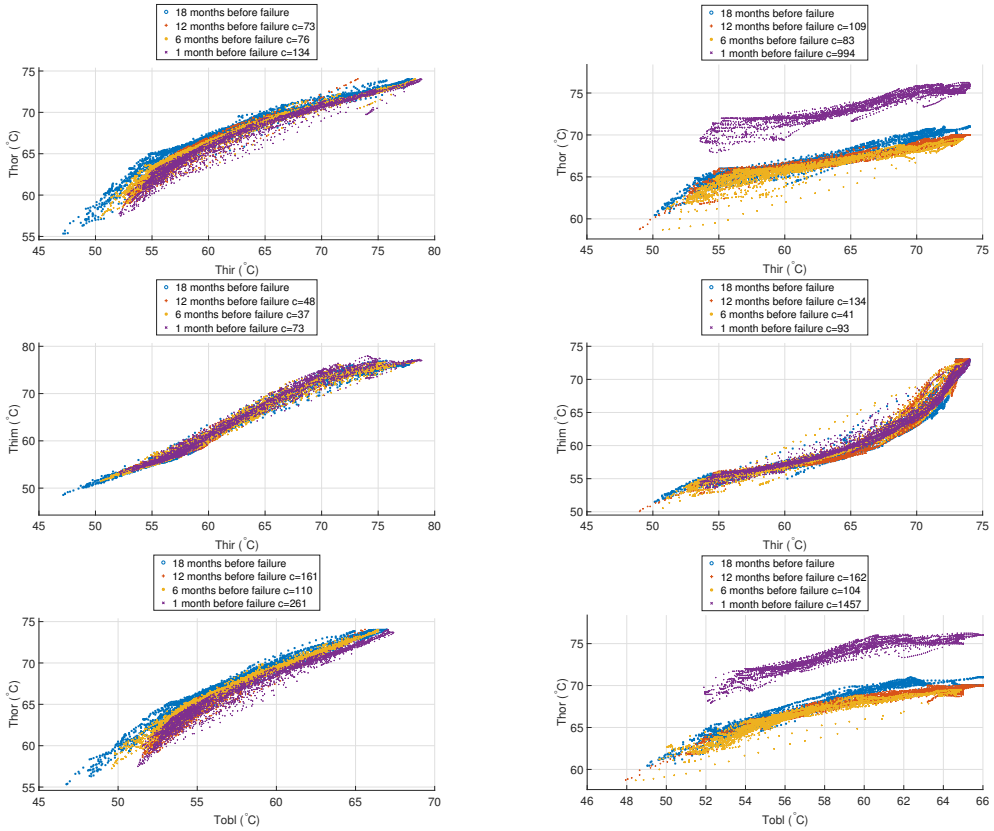


Figure 4.12: Trending results between various temperatures across the gearbox. (left column=onshore- Turbine 1 right column=offshore- Turbine 3)

4.6.3 Clustering Results

Results after applying dimensionality reduction are shown in Figure 4.14. Turbine 1 (Figure 4.14a) shows a different clustering formation between the different stages towards failure. The data collected from the period 1 month before failure, is clustered in a different part of the 2 dimensional feature space compared to the data that is collected 18 months before. The progression of the fault is gradually showing in the 2 other periods plotted (12 months and 6 months). In most cases no clear apparent structure in the data was observed (see Figure 4.14b).

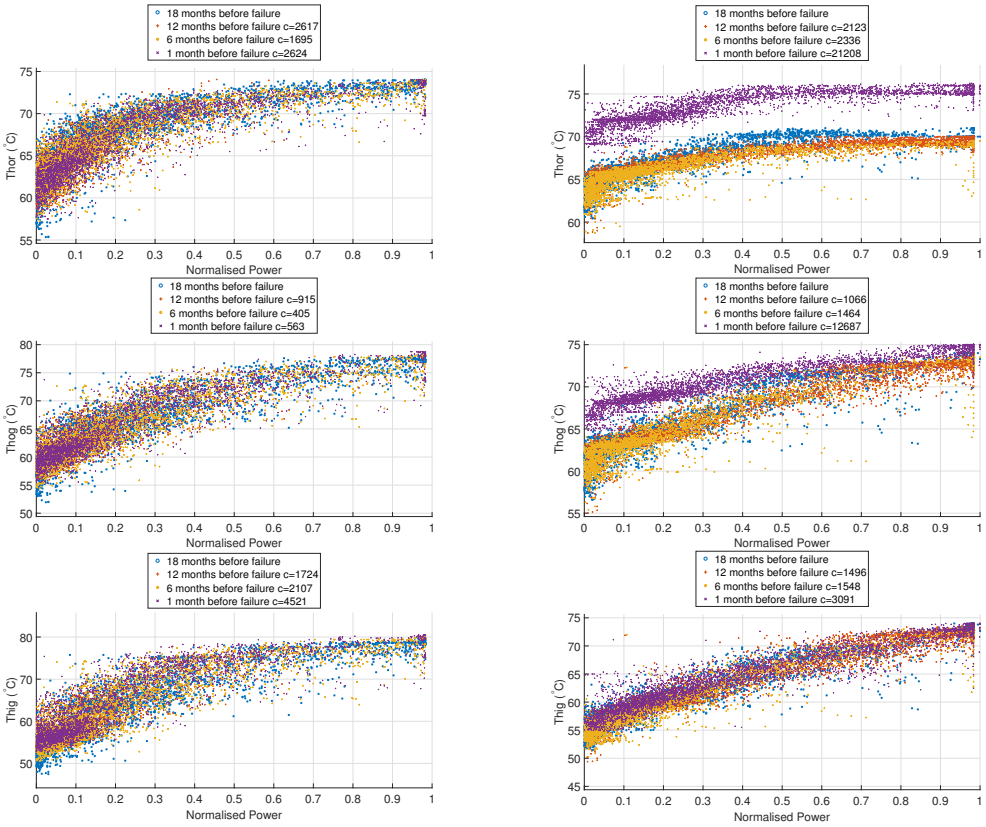
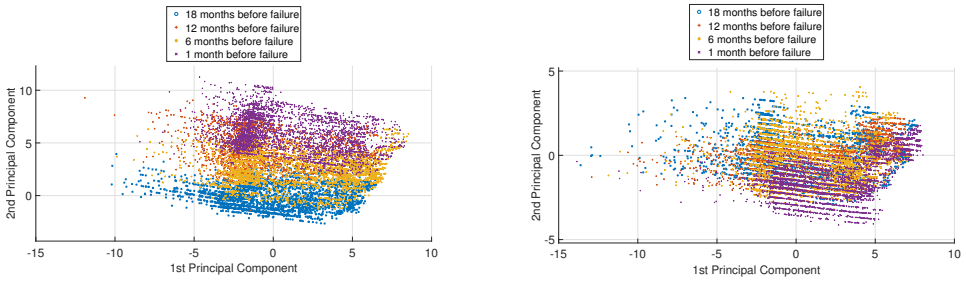


Figure 4.13: Trending results between power level and various temperatures across the gearbox. (left column=onshore- Turbine 1, right column=offshore- Turbine 3)



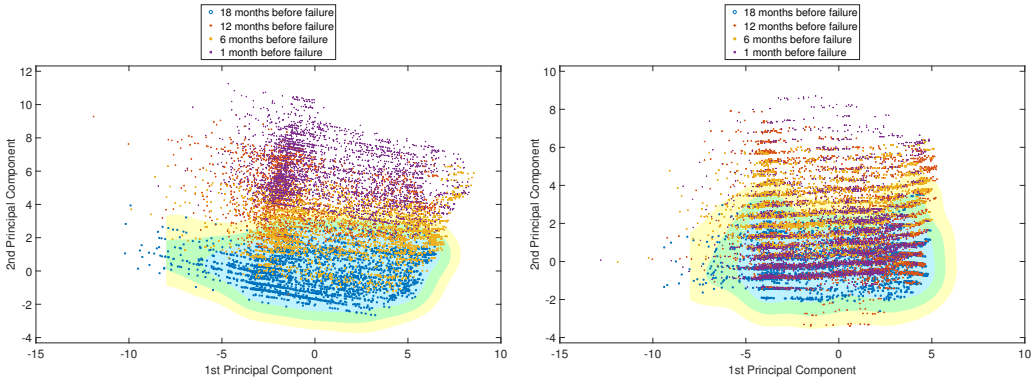
(a) Principal component analysis for Turbine 1 (onshore).

(b) Principal component analysis for Turbine 3 (offshore).

Figure 4.14: Principal component analysis results

PCA has a few limitations, relying on linear assumptions being one of them. PCA is focused on finding orthogonal projections of the dataset that contains the highest variance possible in order to find linear correlations in the dataset, but if data is not linear correlated PCA can nor represent them adequately. Also, mean and covariance do not describe some distributions, being mainly important for Gaussians.

Some results of one-class classification using SVMs and the data after the application of PCA are shown in Figure 4.15. A detailed formulation of this algorithm is given in Chapter 2. The training was performed for the data collected between 18 and 17 months before failure. The thresholds to detect 5%, 1% and 0.1% outliers are calculated. These are easily visualised in a 2 dimensional projection of the dataset after the application of PCA, as shown in Figure 4.15. The blue region is the normal-training data, the green is 5% outliers, the yellow is 1% and the white is 0.1%. Anomalies can potentially be detected based on the thresholds set, with the outliers representing abnormal operation.



(a) One-class classification regions for Turbine 1 (onshore).

(b) One-class classification regions for Turbine 3 (offshore).

Figure 4.15: One-class classification results

Self organising maps (SOM) have proven to be a powerful clustering tool as described in Chapter 2. This method is applied to the analysed wind turbine datasets. The datasets analysed belong to the turbines that have 18 months of continuous data acquired, so that a the behaviour of the distance parameter can be thoroughly and dynamically examined. The dimensions of the map are chosen based on a rule proposed in [187]. The rule suggests that $M = 5\sqrt{N}$ where M is the number of neurons, which is an integer close to the result of the right hand side of the equation, and N is the number of observations.

An example of a self organising map is shown in Figure 4.16 for an offshore turbine.The unified distance matrix is depicted, which is a projection of the shape of the 13- dimensional dataset. The colour represents the magnitude of the Euclidean distance between connected map units. This facilitates a qualitative understanding of how the SOM has conformed itself to the shape of the input data [112].

The training period was again 18 months before the component failure for 1 month and the evolution of MQE is shown in Figure 4.17. It is clear that the distance from the map during the training period has a constant low variance and non monotonic value which continues in a similar manner during the beginning of the test period and then increases significantly leading up to failure. Most of the other turbines did not show such promising results.

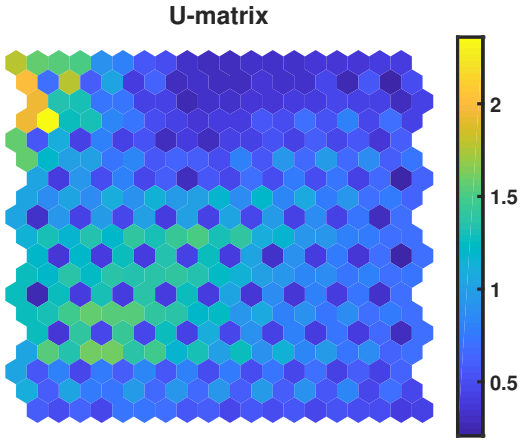


Figure 4.16: Unified distance matrix for an offshore turbine

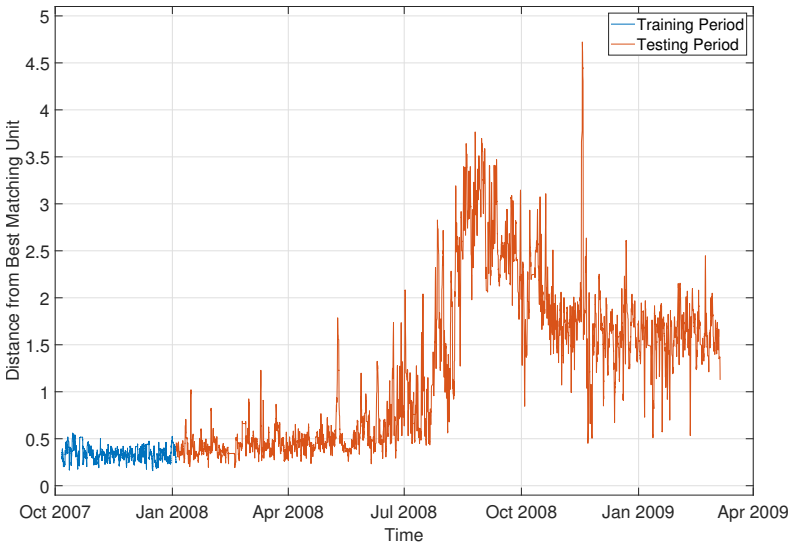


Figure 4.17: Distance from Self Organising MAP

	Input	Output
Model 1	N_{rot}	T_{hog}
	P_{el}	
	T_{obl}	
	T_{hor}	
	T_{hig}	
Model 2	V_{win}	T_{hig}
	P_{el}	
	T_{obl}	
	T_{hog}	
	T_{him}	

Table 4.2: Normal Behaviour Model Inputs and Outputs.

4.6.4 Normal Behaviour Modeling Results

The purpose of the models built is to see if the temperature sensors can give an indication of the fault present and if the stage of the gearbox with the fault can be isolated. That's why one sensor close to the fault location and one on a stage further are chosen to investigate this. There is no direct sensor on the planetary stage which makes this process more challenging. The temperature in the hollow shaft and the high speed shaft are modeled, to examine in which stages of the gearbox this fault can be predicted. The inputs and outputs of the normal models are chosen based on the correlations presented in the previous section and are given in Table 4.2.

Since there is a big population of turbines with a historic failure example, there are two types of normal behaviour models trained, for each of the temperature prediction models shown in Table 4.2. The two types of normal behaviour models are global and individual. The global model uses data from all the population of turbines for training, whereas the individual models are trained on each wind turbine unit respectively. The training period for both models is during the period when the turbines are considered to be in normal operation,

Initially, a global model that incorporates data for all turbines is trained. Since the model includes both onshore and offshore turbines, this is added as a categorical variable. For that reason, one-hot encoding is used which creates new binary variables, indicating the presence of each possible category from the original data. One-hot en-

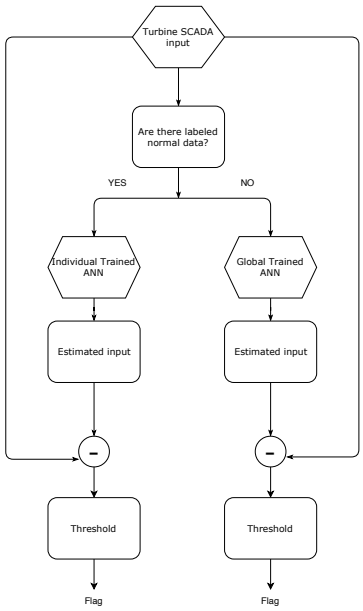


Figure 4.18: Framework flowchart

coding is a process by which categorical variables are converted into a form that could be provided as input to regression algorithms, by using binary variables. Then, normal behaviour models are trained and optimised for each turbine individually. The two models that predict the hollow shaft temperature and the temperature in the bearing of the high speed shaft respectively are applied. The models are trained and tested on all turbines. The ANN models are used to estimate the temperature and the error is calculated over the course of 1 day with a moving window of two hours. If the error value has a probability of occurrence of less than 0.01%, then it is considered to be an anomaly in the gearbox. The flowchart of the applied methodology is given in Figure 4.18.

The regression results for the global temperature prediction Model 1 are given in Figures 4.19a, 4.19b. The training and validation that was not included in the training seem to have a good fit, whereas the prediction of the temperature 1 month before failure is not that successful. The R^2 values for the different time periods before failure are given in Table 4.3. Both the training and the validation have an R^2 close to 1, which means that the model has captured fairly accurately the relationships between the parameters and has not overfitted. The normal behaviour model seems to predict less

	Healthy (Training)	Healthy (Validation)	1 Year	6 Months	1 Month
Model 1	0.96	0.96	0.81	0.79	0.69
Model 2	0.95	0.94	0.82	0.78	0.68

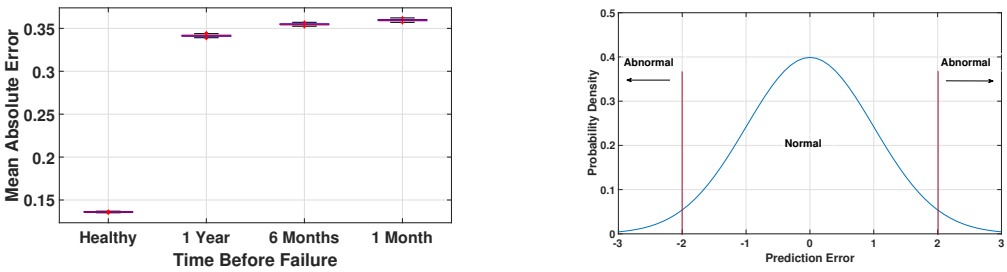
Table 4.3: R^2 results for time periods before failure

accurately the temperature one month before the bearing failure. The mean absolute error is calculated using bootstrapping [188] for the different time periods before failure and is shown in Figure 4.20a. As expected, the prediction error seems to increase close to failure which indicates that the temperature can be used to predict incipient planet bearing inner race defects.

Both the hollow shaft bearing temperature model and the high speed shaft bearing temperature model give an indication of the incipient planet bearing fault, which would make fault isolation quite challenging, even in stage level.

The results of the fault detection framework applied in each turbine are shown in Table 4.4. A false alarm means that the model predicted a failure during the period that the turbine is considered to be in normal operation and it's the period that is included in the training process but a subset of data is kept for testing purposes.

The nature of errors illustrated in Figures 4.19c, 4.19d, where the error of the normal behaviour is distributed around zero but close to the component failure the mean is shifted and the standard deviation increases.

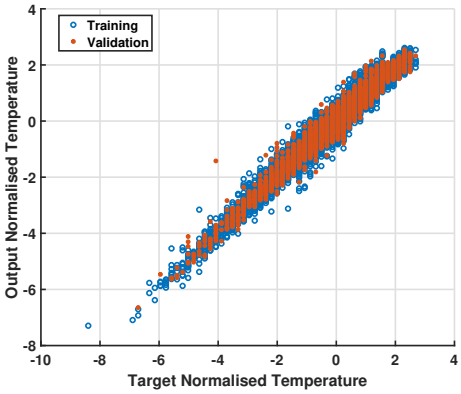


(a) Boxplot of mean absolute errors at different times before failure.

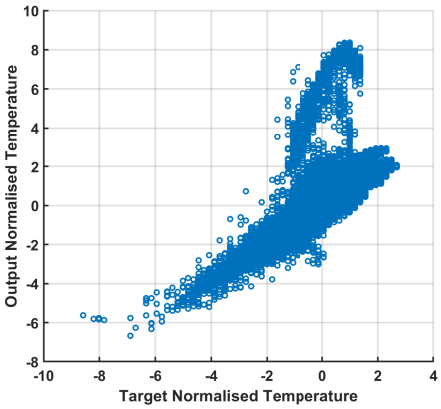
(b) Probability density distribution of prediction error.

Figure 4.20: Boxplot of errors and indicative probability density function

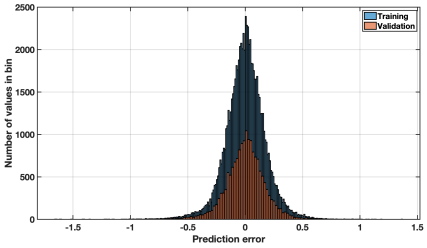
For the turbines that did not have labeled "healthy" data, a global normal behaviour model is trained for the turbines that had "healthy" data and the rest of the turbine



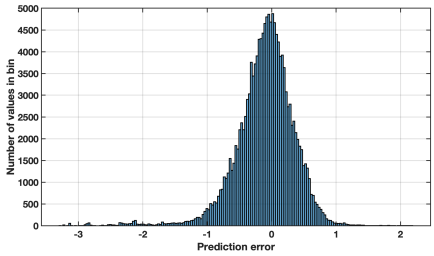
(a) Healthy regression results



(b) 1 month before failure regression results



(c) Healthy data histogram of errors for both training and validation subsets



(d) 1 month before failure data error histogram

Figure 4.19: Regression results and errors for healthy and close to failure data. The errors increase close to the occurance of a planet bearing failure.

			1 Year	6 Months	2 Months	3 Months	1 Month	Not Detected	False Alarm
Turbines with normal data	Model 1	Individual	138	21	4	1	6	0	0
		Global	71	4	3	0	0	0	92
	Model 2	Individual	131	22	5	3	9	0	0
		Global	69	0	0	0	0	0	101
Turbines without normal data	Model 1	Individual	-	-	-	-	-	-	-
		Global	12	2	0	0	0	0	15
	Model 2	Individual	-	-	-	-	-	-	-
		Global	14	1	0	0	0	0	14

Table 4.4: Model results showing when fault was detected for the population of turbines

temperatures are predicted. That way it can be explored if models can be made for turbines and used for other turbines of the same make that do not have normal data. This could aid the operators in detecting abnormalities in turbines that do not have enough historical labeled data. One way to validate if this assumption holds is to test the model on turbines that have labeled "healthy" data, by excluding them in the training process of the global model and predicting the temperatures in all the time periods.

The models trained individually for each turbine seem to outperform the models trained globally for a fleet of turbines and then tested on a new one. The global models seem to give plenty of false alarms, which basically indicates that global models do not capture the normal operation accurately or that the period considered as normal operation is not normal for a population turbines.

It is important to note that even though all turbines failed due to the same failure mode, each failure developed in a different way and in a different pace due to operating conditions. Furthermore, even though all turbines have the same model, there could be slight differences in the gearbox components. Also, it is only confirmed when the catastrophic failure happened in each turbine and not when exactly it started developing, so it is hard to validate whether the prediction of a failure actually corresponds to an incipient failure.

4.7 Conclusions

This chapter investigates the use of SCADA data for temperature monitoring inside the wind turbine gearbox. First, data visualisation for greater understanding is performed on a population of turbines of the same make that failed from a planet bearing inner race defect. The temperature values and the behaviour towards failure differs from turbine to turbine. It is also interesting to notice that the ambient temperature does not have a significant effect on the gearbox bearing temperatures. No clear trends can be deducted from observing the temperature behaviour.

Trending and clustering methods and how they can be utilised for SCADA data and gearbox health monitoring are demonstrated. The behaviour of various monitored

parameters -with a focus on gearbox temperatures- progressively up to a gearbox failure is studied. Methods of dimensionality reduction and unsupervised learning are applied to gain a better understanding of the multi-dimensional dataset. Trends and fault development can be more clear in some cases than others. According to the presented results and past literature work, trending and clustering algorithms have not exhibited consistently successful results and often make the interpretation of results difficult.

The concept of evaluating a residual of measured minus modelled signal provides a failure indicator which is easy to interpret, and for that reason normal behaviour models are widely applied in the context of SCADA data health monitoring. Therefore, normal behaviour models are built based on data that are assumed to be in normal operation. Models are built and optimised using neural networks for two different temperatures across the gearbox. It seems that both the hollow shaft bearing temperature and the high speed shaft bearing temperature can give an indication of a planet bearing fault, which shows that diagnosis could be challenging in a blind test case since heat propagates throughout the gearbox.

The models are trained for each turbine individually and for a population of turbines collectively. The individual turbine models have successfully demonstrated the prediction of planet bearing incipient failures months before the gearbox breakdown. Global models made for a population of wind turbines used to predict failures in other turbines of the same make that have not been included in the training process are also explored. It should be noted that even though the same failure mode is examined in all turbines, each turbine will end up failing in a different way due to operating conditions and failure development. Generalisation of models from one turbine to another have proven to be challenging.

It should be highlighted that most models presented rely heavily on the selection of training data, training algorithm, training process and threshold setting. These dependencies can result in either undetected faults or frequent false alarms.

Furthermore, fault initiation and propagation can not be easily known from historic field failure examples due to lack of accessibility, accurate physical modeling and limited amount of monitored parameters. These introduce an extra element of uncertainty in

the validation of the discussed models.

It is concluded based on the findings of this chapter that SCADA can be used for gearbox fault detection. However, isolation and diagnosis could be quite challenging, since heat propagates and there is normally not a dedicated sensor installed in every component. Also, generalisation from one turbine to another (even among same models) is not straightforward. Fault causes and propagation might need to be better understood through the aid of physical modeling.

Chapter 5

Feature Extraction using High Frequency Vibration Data

5.1 Chapter Contribution

From the CM perspective, a wind turbine gearbox consists of three major components: gears, bearings, and lubricant. CM aims to detect failures at such an early stage that maintenance actions can be planned in advance. This thesis focuses on the vibration analysis approach for gears and bearings.

Various vibration analysis methods have been proposed in the literature regarding wind turbine gearboxes and some of them are evaluated and presented in [140]. The methods presented in the literature usually compare one or two signal processing algorithms. Also, the lack of real failure examples makes it challenging to benchmark the effectiveness of diagnostic algorithms. Either only scarce real examples or examples from simulations and lab experiments are used to validate vibration analysis techniques. This chapter aims to answer the following research question:

"What are the most effective vibration analysis algorithms for wind turbine gearbox diagnostics?"

The question is answered by comparing various signal processing algorithms using samples from real wind turbine run-to-failure historic data. Features are extracted from from processed signals that act as health indicators. These features are used in a later

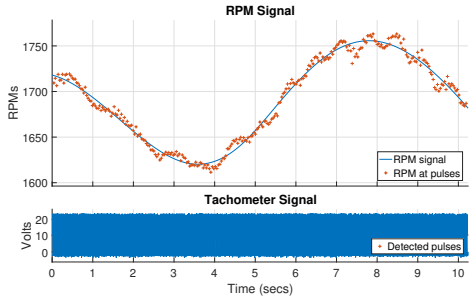


Figure 5.1: Tachometer Pulse and RPM Signal

chapter as inputs in machine learning models.

5.2 Gear Diagnostics

This section presents an overview of the most important existing vibration diagnostic algorithms for gears and perform a comparative study on signals from 4 failure examples. Most gear diagnostic algorithms are based on the fact that a local gear defect, like a crack, can affect the stiffness of the neighbouring teeth. This changes the vibration signal, which is reflected in amplitude and frequency modulation as shown in [126]. A model for fault diagnosis of planetary gearboxes considering the modulation effects and the vibration transfer path is given in [127].

5.2.1 Order Tracking

The rotational speed of the wind turbine is not constant because it depends on the wind speed and the torque controller. Speed variability is reflected in the vibration signals, which are not stationary. This causes spectral leakage in the frequency domain. This problem can be solved through computed order tracking [189]. The vibration frequencies are proportional to the rotational speed and the constants of proportionality are the orders. The non-stationary vibration signals are resampled in the angular domain, creating cyclo-stationary signals.

The shaft speed variation for a vibration sample along with the tachometer pulse is shown in Figure 5.1. As depicted, the speed has a 5% variation during the 10s acquisition period.

5.2.2 Sideband Analysis

Sideband analysis is an simple diagnostic method of vibration signals. Order tracking is first applied and as a consequence frequency components become less smeared. In gear faults the modulation is reflected as sidebands around the gear mesh frequency and its harmonics, as explained in the previous section. These sidebands are separated by the shaft frequency of the gear [126]. Sideband analysis extracts health indicators based on this modulation. These are the following:

- Sideband Energy Ratio (SER)
- Sideband Average Power (SAP)
- Sideband Modulation Lifting Factor (SMLF)
- Sideband Power Factor (SBPF)

SER is shown in Eq. (5.1). The sum of the amplitudes of the first six sideband peaks on each side of the center mesh frequency $\sum_{i=-6}^{6} A_{SB,i}$ is divided by the amplitude of the center mesh frequency A_F [144].

$$\text{SER} = \frac{\sum_{i=-6}^{6} A_{SB,i}}{A_F} \tag{5.1}$$

For a healthy gear mesh the sidebands have a small amplitude compared to the center mesh frequency. As damage develops on a gear tooth, the sidebands rise in amplitude which results in a larger SER value.

SAP is introduced by the authors [190] calculates the power of the signal around the centre mesh frequency. The amplitude A around the mesh frequency and with a window up to six sidebands is used. The power is normalised with the length N of the data segment.

$$\text{SAP} = \frac{\sum_{i=1}^{N} |A(i)|^2}{N} \tag{5.2}$$

Sideband modulation lifting factor (SMLF) or sideband level factor (SLF) [191] is defined as the sum of the first order sideband about the fundamental gear mesh

frequency divided by the standard deviation of the signal of interest.

$$\text{SMLF} = \frac{\sum_{i=-1}^{1} A_{SB,i}}{\text{std(x)}} \tag{5.3}$$

The SBPF algorithm is introduced by [145] and it sums the Power Spectrum amplitudes of the meshing frequency second harmonic and its first 5 sideband peaks on each side. Zappala's formula is given as $PSA(2xf_{mesh,HS}) + \sum_{i=-5}^{5}(PSA(SB_i)$ but the authors have modified the definition in Eq. 5.4 to be consistent with the rest of the formulas given in this Section.

$$\text{SBPF} = (2A_F)^2 + \sum_{i=-5}^{5} A_{SB,i}^2 \tag{5.4}$$

Other condition indicators for gear fault diagnosis can be found in [191].

5.2.3 Cepstrum Analysis

Cepstrum is a nonlinear signal processing technique used to identify and separate harmonic families. One of its earliest applications was in speech and echo analysis, with the aim of detecting periodic structures [192], [193].

The cepstrum can be defined as a spectrum of a logarithmic spectrum, which means logarithmic amplitude and linear frequency scale. It can therefore be a really useful tool for detecting periodicity in a spectrum. That periodicity can be families of harmonics. In gearbox vibration spectra, families of sidebands with uniform spacing rise as an indication of various fault types due to modulation. The spacing of these sidebands reveals valuable information about the source of vibration.

Given a real signal y and its Fourier Transform Y, different cepstrum forms are given in Eq. (5.5):

$$\text{C}_{\text{cplx}}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log[Y(e^{j\omega})]e^{j\omega}d\omega \tag{5.5}$$

Depending on whether the phase information is kept or not as an input in the inverse Fourier Transform, the cepstrum can be complex or real. The cepstrum is complex if the phase information of the original time waveform is kept.

The independent variable of the cepstrum τ has time dimensions and it is called "quefrency". High quefrency means rapid fluctuations in the spectrum (small frequency spacings) ans low quefrency means slow changes in the spectrum (big frequency spacings).

One of the advantages of the cepstrum over the classic spectrum analysis is that the cepstrum is less sensitive to secondary effects and it gives a measure of the sideband activity over the whole spectrum. This whole sideband activity is represented basically by one line in the cepstrum. One example of secondary factors that affect the spectrum can be that in practice, both amplitude and frequency modulation can cause asymmetry in the sidebands because they either reinforce or cancel each other out. Transmission paths can also affect this, and that would be reflected in the spectrum.

One of the disadvantages of the cepstrum is that it requires a baseline, since the harmonics could also be related to shaft problems. Also, if the faults are distributed the cepstrum would not capture the modulation [140].

It is often recommended to perform a zoom analysis on the spectrum before performing cepstrum analysis, so as to obtain sufficient resolution [146].

5.2.4 Time Synchronous Averaging

TSA resamples the vibration data synchronously with a shaft. The main concept behind TSA is to average together a series of signal segments, each corresponding to one period of a synchronising signal. In order to perform TSA, signal phased-locked with the angular position of a shaft within the system is needed. This can be provided by a tachometer signal. The results is that vibration components associated with a specific shaft or gear are isolated and all other components are averaged out. This makes it very suitable for gearbox vibration analysis.

The TSA signal is defined in Eq. (5.6)

$$y_{\mathrm{TSA}}(t) = \frac{1}{N} \sum_{i=0}^{N-1} y(t + iLT_s) \tag{5.6}$$

Where $y(t)$ is the measured and resampled time signal, N is the number of gear revo-

lutions, L is the number of samples in one revolution and T_s is the resampling period.

An extensive review of TSA algorithm can be found in [142]. Statistical indicators can be extracted from the TSA (such as root mean square, crest factor) and can give information about the state of the component.

5.2.5 Amplitude Modulation

Amplitude Modulation (AM) analysis is the absolute value of the Hilbert transform of the narrowband signal, since primary gear meshing characteristics extracted from narrowband analysis is the subject of interest [140]. Narrowband analysis operates the TSA signal by filtering out all the tones except that of the gear mesh and with a given bandwidth, which can be achieved through band-pass filtering.

Modulation is a non-linear effect in which several signals interact with one another to produce new signals with frequencies not present in the original signals. Amplitude modulation is defined as the multiplication of one time-domain signal by another time-domain signal. Gear defects or faults can increase the kurtosis of the signal significantly. AM is sensitive to eccentric gears and broken or soft tooth faults. The derivative of the amplitude modulation can reveal also useful information regarding a fault presence.

5.2.6 Empirical Mode Decomposition

EMD was proposed to adaptively decompose a nonstationary signal into a collection of intrinsic mode functions (IMFs). There are two conditions that each IMF must satisfy:

1. The number of extrema (maxima and minima) must be equal to the number of zero crossing points or differ at most by one.

2. At any point, the mean value of the envelope defined by the local maxima and the one defined by the local minima is zero.

The decomposition starts with the research of the maxima and the minima along the signal. Subsequently all the maxima and minima are interpolated by a spline. Therefore, this maxima and minima are basically the upper and lower envelope. The mean envelope is subtracted from the signal to obtain the IMF. If the IMF cannot satisfy

the aforementioned conditions, it will be treated as a raw signal and the research for maxima and minima will be repeated. When the two conditions are met, the IMF is subtracted from the original signal and the resulting residual signal is used to follow the same process and calculate the rest of the IMFs. The EMD algorithm, applied to the original signal, stops when the residual signal is a constant or monotonic function, after the extraction of the last IMF. The algorithm is explained in more detail in [194].

Although EMD is well known and widely used, it suffers from a major drawback of mode mixing. Either an IMF represents a mixture of components of widely different signals, or a single signal spreads over several IMFs. Ensemble empirical mode decomposition is used in order to overcome this problem [195], which is a noise assisted method.

Fourier transform is applicable only for stationary and periodic signals. For non-stationary signals, in which frequency value changes at any moment, the definition of the instantaneous frequency is more appropriate. The instantaneous frequency is the frequency of a sine or a cosine function that locally fits the signal [194]. By applying the Hilbert transform to each IMF the amplitude function can be calculated through the envelope and the instantaneous frequency can be calculated through the phase.

5.2.7 Spectral Kurtosis

SK is a very powerful tool of time frequency analysis in signal processing and it essentially extends the concept of kurtosis and it indicates how the impulsiveness of a signal varies in frequency. Its principle is analogous to the power spectral density which decomposes the power of the signal in frequency, except the fourth-order statistics instead of second-order are used. Thus, SK is capable of detecting transients in a signal even when they are buried in noise, by indicating in which frequency bands they take place.

In a case of a gear fault, for example a tooth crack, the excitation affects structural responses. SK aims to detect those non-Gaussian excitation components along with the frequency [143]. Based on the SK, the concept of the kurtogram was proposed as a tool for blind identification of detection filters for diagnostics [196].

The method of gear diagnostics based on the spectral kurtosis is as follows:

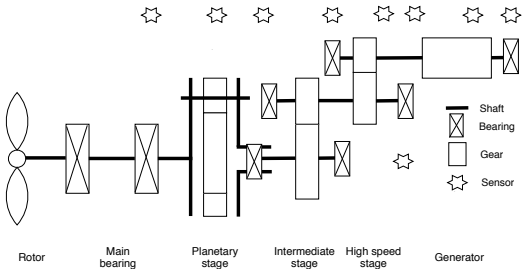


Figure 5.2: Gearbox Internal Nomenclature

1. Calculate the fast kurtogram of the vibration signal

2. Design the band pass filter based on the results of the kurtogram

3. Analyze filtered signal

5.2.8 Gear Case Study

Planetary gearboxes are widely used in wind turbines due to their large transmission ratio. The case study presented concerns a wind turbine gearbox consisting of three stages: one low speed planetary stage (PS) and two parallel stages, namely a high speed (HS) stage and an intermediate speed (IS) stage. The main shaft is connected to the planet carrier (PLC) and the HS pinion of the gearbox is coupled to the generator. Figure 5.2 shows the internal gearbox structure. The calculation of the gear frequencies is shown in Table 5.1.

Table 5.1: Fundamental Gear Frequencies

Gear Element	**# Teeth**	**Speed**	**GMF**
PS Planet Carrier		f_a	
PS Ring Gear	Z_r	Fixed	
PS Planet Gear	Z_p	$f_a \frac{Z_r}{Z_p}$	$f_a Z_r$
PS Sun Pinion	Z_s	$f_s = f_a(1 - \frac{Z_r}{Z_s})$	
LS Gear	Z_l		$Z_{ip} f_i$
IS Pinion	Z_{ip}	$f_i = \frac{Z_l}{Z_{ip}} f_s$	
IS Gear Wheel	Z_{ig}		$Z_h f_h$
HS Pinion	Z_h	$f_h = \frac{Z_{ig}}{Z_h} f_i$	

The wind turbine rating and the signal acquisition characteristics are shown in

Table 5.2: Wind Turbine and Signal Acquisition Characteristics

Turbine Rating	**Signal Duration**	**Sampling Rate**
1.5-3 MW	10-11 s	25-26 kHz

Table 5.2 [1]. Accelerometers are installed in the different stages of the gearbox and a tachometer is installed on the high speed shaft. Operational variables are also recorded, like the generator speed, the wind speed and the power produced.

The failure presented in this case study occurs on the intermediate stage pinion tooth. The failure mode is gear tooth tip and flank fracture, as shown in Figure 5.3. Root cause analysis has not been completed by the manufacturer and is out with the scope of this paper.

Figure 5.3: Broken pinion on intermediate shaft

Failure examples from 4 wind turbines are collected. There exist vibration datasets at different times progressively prior to failure. The oldest dataset is collected 2.5 years before the incident and according to the maintenance reports the gearbox at this time is in a healthy state. In total there are 72 vibration data samples from different loading conditions.

The signals are collected in 3 groups with respect to the time that the failure was recorded; healthy, 5-6 months before failure, 1-2 months before failure.

The signals -after being order tracked and for similar loading conditions- are shown in Figure 5.4. Vibrations signatures at different times prior to failure are as expected according to the theory described in Section 5.2.2, with sidebands rising around the

[1] Ranges are provided for the rated power, sampling period and sampling frequency for confidentiality reasons.

centre frequency (order 2), which corresponds to the second harmonic of the gear mesh frequency.

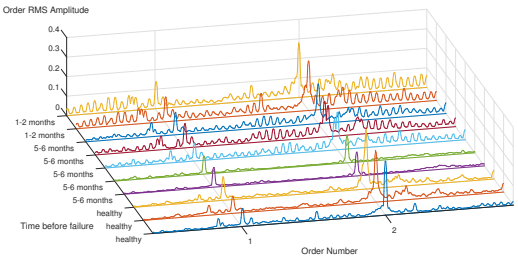


Figure 5.4: Order Spectra Timeline Towards Failure

Cepstral analysis was applied on the case study dataset. The analysis was based on the amplitude order spectrum. The results for similar loading conditions are shown in Figures 5.5. Sidebands are clear in the period of 1-2 months before failure and sometimes are also apparent in the period of 5-6 months before failure. This modulation is reflected in the cepstra, when a spike is noticed at around 0.17 seconds which corresponds to the intermediate speed shaft speed (5.8 Hz).

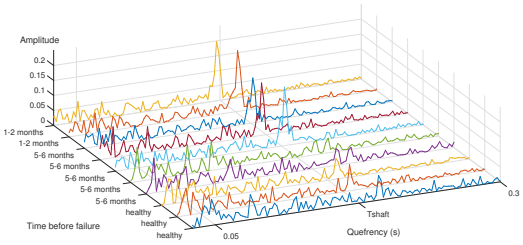


Figure 5.5: Cepstra Timeline Towards Failure

TSA was also applied according to section 5.2.4. As shown from the plots towards failure, there is not a significant visible difference between the healthy vibration signals and the faulty ones. It is discussed in [143] that TSA can under-perform in diagnostics of gears because of the elastic deflection of the generator shaft, resonant amplification effects and possible loss of tooth contact occurring in a wide load range, which can introduce chaotic effects in signals. Also, only around 10s of the signal are available, which means limited number of revolutions to be averaged.

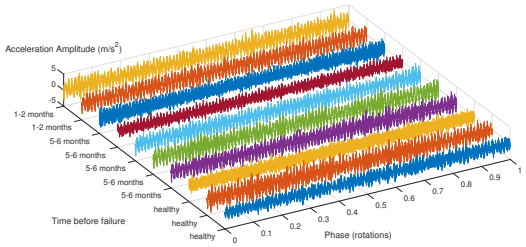


Figure 5.6: TSA Plots Towards Failure

The AM results are given in Figure 5.7. Large values and variations of modulation are noticed close to the component failure. AM can be effective in diagnosing gear faults as expected from previous studies in the literature [140].

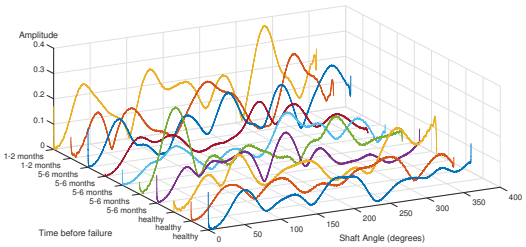


Figure 5.7: AM Plots Towards Failure

The envelope spectra of the selected IMFs are shown in Figure 5.8. Spikes in the multiples of the shaft frequency only appear very close to the component failure. The ineffectiveness of EMD can be attributed to the fact that its theoretical basis is purely empirical and one of its main drawbacks is mode mixing, as discussed in [197].

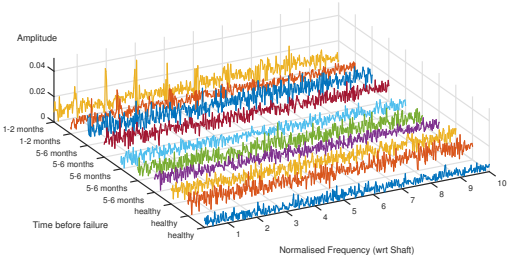


Figure 5.8: EMD Plots Towards Failure

The SK plots are shown in Figure 5.9. In some cases spectral kurtosis was really

effective in revealing the fault pattern, as shown on the last plot in the figure, where the spikes are separated by the intermediate shaft period. The method however was not consistently effective in all the samples. This can be attributed to the filter limitations in the Finite Impulse Response, which is used in the SK algorithm.

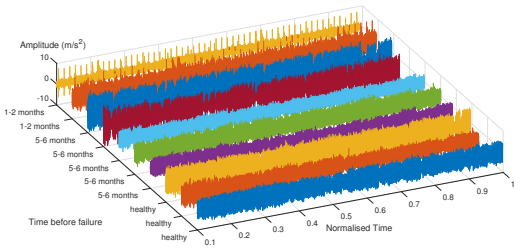


Figure 5.9: Spectral Kurtosis Plots Towards Failure

5.3 Bearing Diagnostics

When a rolling element strikes a local fault in the inner or outer race (or when a faulty rolling element strikes the inner or the outer race), a shock is introduced. That shock excites high frequency resonances of the structure between the bearing and the transducer. Those resonances depend on the strength of the load borne by the rolling elements and are further amplitude modulated by this effect. Also, the transmission path from the fault to the transducers should be taken into account and in case of a moving fault, the transfer function of the transmission path also varies. This section aims to present techniques for analysing bearing faults, which are usually more challenging than gears.

The characteristic bearing fault frequencies are determined according to Eq. (5.7a)-

(5.7d).

$$\text{BPFI} = f\frac{N}{2}\left(1 + \frac{B}{P}\cos(\theta)\right) \tag{5.7a}$$

$$\text{BPFO} = f\left(1 - \frac{B}{P}\cos(\theta)\right) \tag{5.7b}$$

$$\text{FTF} = \frac{f}{2}\left(1 - \frac{B}{P}\cos(\theta)\right) \tag{5.7c}$$

$$\text{BSF} = f\frac{P}{2B}\left(1 - \left(\frac{B}{P}\cos(\theta)\right)^2\right) \tag{5.7d}$$

Where BPFI and BPFO are the ball passing frequencies inner and outer respectively, FTF is the fundamental train frequency (cage speed) and BSF is the ball spin frequency (the frequency with which the fault strikes the same race- inner or outer). Those frequencies depend on the shaft frequency f and some parameters regarding the bearing dimensions. Those are N the number of balls, B the ball diameter, P the pitch diameter and θ the contact angle.

It should be taken into account that the above mentioned frequencies in reality have some slip, due to the variation of the angle θ with the position of each rolling element in the bearing, because the ratio of local radial to axial load changes. That slip is considered to be around 1-2% [150].

5.3.1 Separating Gear from Bearing Signals

Bearing signals in gearboxes are often masked by discrete frequency components coming from gears, even if there are no gear faults present. Therefore, it is useful to separate gear from bearing components.

The easiest way to achieve this is through using the residual from linear prediction [198]. Another more effective method is based on the self-adaptive noise cancellation technique. When two components are to be separated and the one is deterministic whereas the other one is random, then the reference signal can be made a delayed version of the primary signal. If the delay is longer than the correlation length of the random signal, then the adaptive filter will only find the transfer function between the

deterministic part of the signal and the delayed version of itself. The adaptive filter is a recursive filter that uses the least mean squares algorithm [199]. Further information on the filter can be seen in [200].

The filter coefficients calculation speed can be improved if the adaptation is performed on blocks of data, by using the overlap-save technique for fast convolution by means of Fast Fourier Transforms.

The filter essentially depends on the choice of three parameters: the time delay, the filter length and the forgetting factor. The delay should be large enough so that the delayed version becomes uncorrelated with the noise in the signal but not so large that the periodic components become uncorrelated [2]. The forgetting factor should be set sufficiently small so that the filter equations converge. The choice of the filter length is also a compromise between sufficient selectivity of the frequency response and convergence.

5.3.2 Spectral Kurtosis

SK was analysed in Section 5.2.7 and is a useful tool for detecting transients in a signal.

Faults associated with rolling element bearings give rise to modulated impulses. Thus, the SK will be large in frequency windows where the fault signal is dominant and small where the spectrum is dominated by stationary signals. The algorithms is explained in [147].

In order to calculate the SK, the short time Fourier transform of the signal is computed by moving a time window along the signal. The window must be shorter than the spacing between pulses, but longer than individual pulses so as to get a maximum value of SK.

5.3.3 Envelope Analysis

Envelope analysis of bearing vibration signals is based on the demodulation of the high frequency resonance associated with bearing impacts [135]. As already explained, the

[2] In theory the correlation time between periodic signals is infinite but in practice the components of the real signal have small distribution in periodicity, so ultimately the autocorrelation function will tend to zero

impacts caused from rolling element bearings modulate the signal at the associated bearing fault frequencies. Therefore, sidebands appear in the spectrum around the resonant frequency, which is hard to know a priori. It is advisable that a resonant frequency that is higher than the shaft and gear harmonics is chosen. Also, it is suggested in the literature to pick a frequency window close to the accelerometer resonance. In the bearing envelope analysis the signal is multiplied with a high frequency, complex signal centered at a hypothesized resonant frequency. This is then low pass filtered to remove the high frequency image, decimated, and the spectral power density is estimated.

5.3.4 Bearing Case Study

To validate the suggested methodology, a case study on vibration data from offshore wind turbines is presented. The wind turbines belong to the same manufacturer, have the same rated power and gearbox type and belong to 4 different offshore wind farms.

In the case study considered the gearbox has a structure commonly found offshore. It consists of two planetary stages (PS1, PS2) and one parallel stage (HS). The main shaft is connected to the PLC of PS1 and the HS pinion of the gearbox is coupled to the generator. The ring gears of the planetary stages are fixed.

The gearbox internal structure is shown in Figure 5.10 and the component speeds are calculated according to Table 5.3.

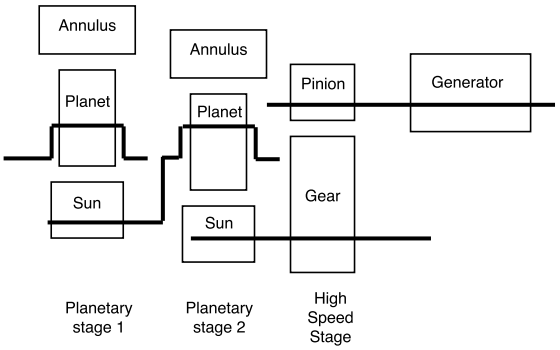


Figure 5.10: Gearbox Internal Nomenclature

The turbines are rated at between 2.5 and 3.5 MW. The vibration data acquisition system consists of 9 accelerometers and a tachometer on the high speed shaft. The generator speed, the wind speed and the power produced by the turbine are also recorded.

Table 5.3: Speed Calculation for Double Planetary Stage Gearbox

Gear Element	**# Teeth**	**Speed**
PS1 Planet Carrier		f_a
PS1 Ring Gear	Z_{r1}	Fixed
PS1 Planet	Z_{p1}	$f_{p1} = f_a \frac{Z_{r1}}{Z_{p1}}$
PS1 Sun Pinion	Z_{s1}	$f_{s1} = f_a(1 + \frac{Z_{r1}}{Z_{s1}})$
PS2 Ring Gear	Z_{r2}	Fixed
PS2 Planet	Z_{p2}	$f_{p2} = f_{s1} \frac{Z_{r2}}{Z_{p2}}$
PS2 Sun Pinion	Z_{s2}	$f_{s2} = f_{s1}(1 + \frac{Z_{r2}}{Z_{s2}})$
IMS Gear Wheel	Z_i	$f_i = f_{s2}$
HSS Output Pinion	Z_h	$f_h = f_i \frac{Z_i}{Z_h}$

The acquisition time of the signal is between 10 and 11s and the sampling frequency is over 25kHz. [3]

The failure mode examined is the same in all wind turbines. This failure occurs on a first planetary stage planet bearing. The failure starts on the inner race way with debris eventually effecting the outer raceway. Figure 5.11 shows the bearing in a faulty state.

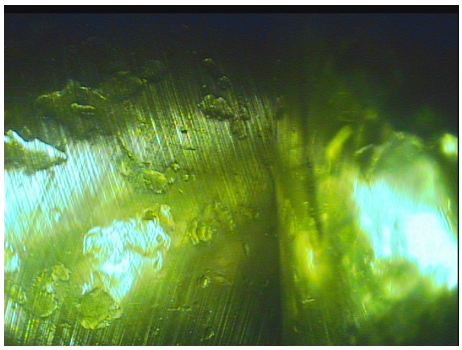


Figure 5.11: Faulty bearing.

Data is collected for the wind turbines at various time steps prior to failure. The oldest dataset dates back to 2.5 years prior to failure and according to the maintenance reports the gearbox at this time is in a healthy state.

The vibration signal of the channel that is closest to the fault is used. Only the first 2 seconds of the data is used on which the signal is assumed to be relatively stationary

[3] Ranges are provided for the rated power, sampling period and sampling frequency for confidentiality reasons.

as the wind conditions will not have changed significantly. Initially the RMS of the raw vibration signals is calculated, since it could easily reveal some present faults. The 95 vibration samples from four different wind farms are grouped according to their acquisition time with respect to the catastrophic failure. The results are shown in Figure 5.12 as a function of the reference torque. It seems that no matter whether the signals are collected from healthy a bearing or a bearing close to failure, the RMS does not reveal any noticable change in condition. This indicates that further signal processing should be investigated and applied in order to reveal fault signatures.

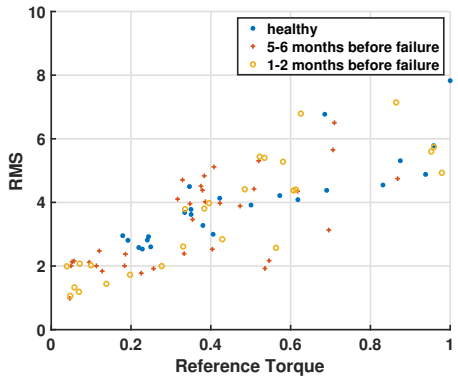


Figure 5.12: RMS of raw vibration signals

Some representative results in the frequency domain after processing the vibration signals are shown in Figures 5.13, 5.14. Figure 5.13 shows the envelope spectra progressively at various times prior to failure with a constant demodulation band. The demodulation band is chosen close to the resonance frequency: 11-12.5kHz. The three spectra are selected to be at similar loading conditions (80% of full load). The same signals are analysed as explained in the methodology and the optimum demodulation band is chosen based on spectral kurtosis. The results are shown in Figure 5.14. The bearing defect frequencies and the harmonics are much more prominent in the second case. The BPFI in this example is around 9 Hz. The sample that's up to 2 months before failure also has some spikes at the BPFO and its harmonics (around 7Hz).

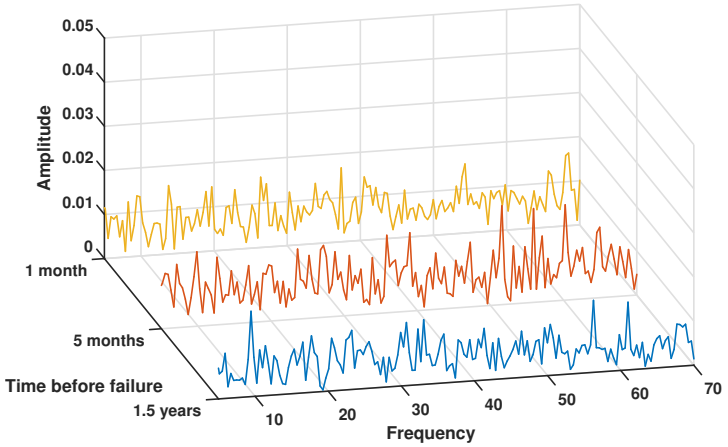


Figure 5.13: Envelope Spectra (constant demodulation band)

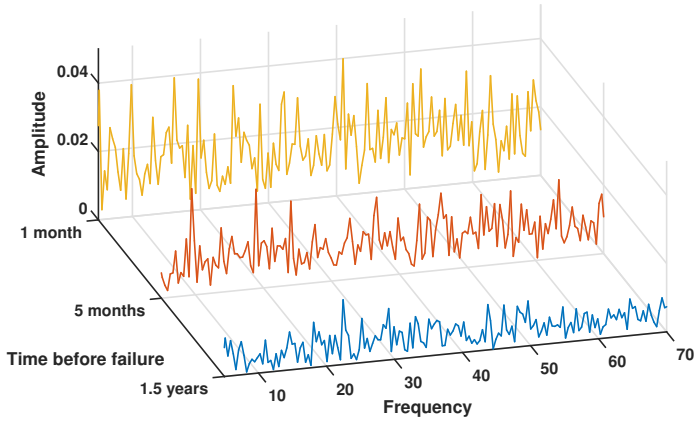


Figure 5.14: Envelope Spectra (based on spectral kurtosis)

5.4 Effect of Operational Parameters on Vibration Signals

Wind turbines operate under variable conditions of randomly changing weather, temperature, wind shear, wind speed, and load. This affects the wind turbine operating conditions and this is reflected in the acceleration patters and amplitudes. Therefore, these changing conditions are apparent in the vibration signals. This will be examined in the present section, in which new failure examples are examined.

5.4.1 Correlation among parameters

The effect of operational parameters on vibration signals is examined using vibration data from wind turbines in an offshore wind farm, located in the UK. The two turbines examined are the same model and make and there are two gearbox fault incidents confirmed. The gearbox is a double planetary stage gearbox with one parallel stage. The first fault examined is a planet bearing fault on the first planetary stage. The failure mode is inner race defect. The second fault examined is a fault on the high speed pinion. The failure mode is confirmed to have started as tooth wear and progressed to general damage, which led to the replacement of both the pinion and the bearings on the stage. The design characteristics of the wind turbines examined shown in Table 5.4[4].

Table 5.4: Turbine Characteristics of Operational Effects Case Study

Type	**Rated Power**	**Rated Speed**	**Gearbox Type**
Offshore	3-4MW	12m/s	Double Planetary

The correlation among the various operational parameters and the vibration signals is initially investigated. The wind turbine operates under variable speed conditions until rated wind speed is reached. Afterwards, pitch regulation allows the wind turbine to actively change the angle of attack of the air on the blades, thus controlling the power output. Another aspect of horizontal axis wind turbines is yawing, which ensures that the wind turbine is producing the maximal amount of electric energy at all times. The yaw drive is used to keep the rotor facing into the wind as the wind direction changes. Pitch and yaw are two variables that can also give indication on the operating condition of the machine.

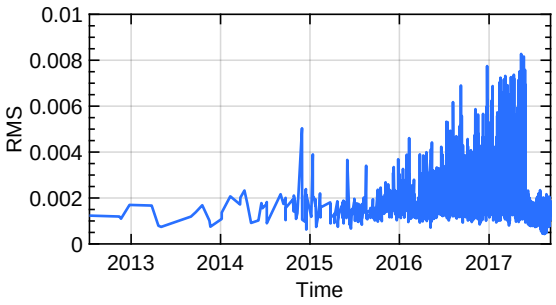
Accelerometer sensors are placed on the two ends of the high speed stage and on the planetary stage. Vibration signals are sampled in the time domain and tracked using a tachometer mounted on the high speed shaft. Then the signals are transformed and analysed in the frequency domain. The condition indicators extracted depend on the component being monitored. A comprehensive modeling of gearbox faults and their signatures on vibration signals can be found in [201]. For bearing faults, the

[4]Ranges are provided for confidentiality reasons.

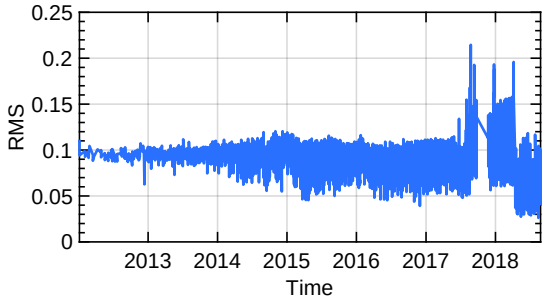
Fast Fourier Transform is being applied and the narrowband RMS around the fifth harmonic of the ball passing frequency is examined. This is chosen based on domain knowledge and is gearbox specific. For gear faults, the cepstrum is being applied and the peak of the modulating shaft frequency is examined. The choice of the condition indicators can vary among vibration analysts, but the aforementioned have proven to have successfully detected historical faults and are used by the industry.

The fault condition indicators are given as a function of time in Figure 5.15 for the two gearbox faults, a planetary bearing inner race defect and a high speed gear damage. The alarm for the faulty bearing was activated on September 2016 and the bearing was replaced on May 2017. The alarm for the gear was activated on September 2017 was replaced on May 2018. There is a large variation in the evolution of the RMS. In terms of diagnostics, it is important to reduce the scatter in the condition indicator and be able to observe its trending. By plotting the RMS at different power levels, the trend towards failure is more clearly monotonic with smaller variations, as can be seen in Figure 5.16. The RMS seems to increase with power as expected.

The bearing fault has a slower progression, since it is a spall that developed over time. The gear fault is a tooth crack that happened more abruptly. This can be seen in Figure 5.15.

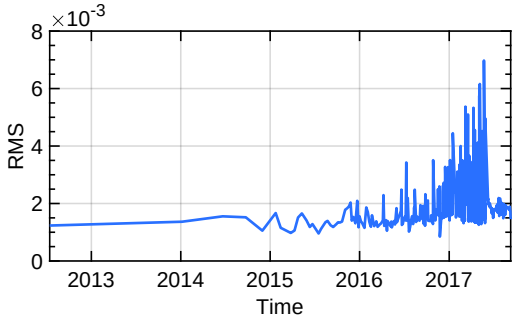


(a) Bearing failure

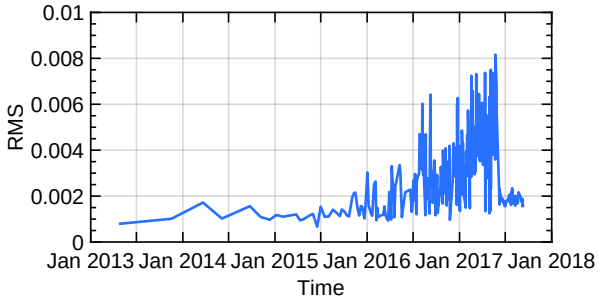


(b) Gear failure

Figure 5.15: RMS leading up to failures



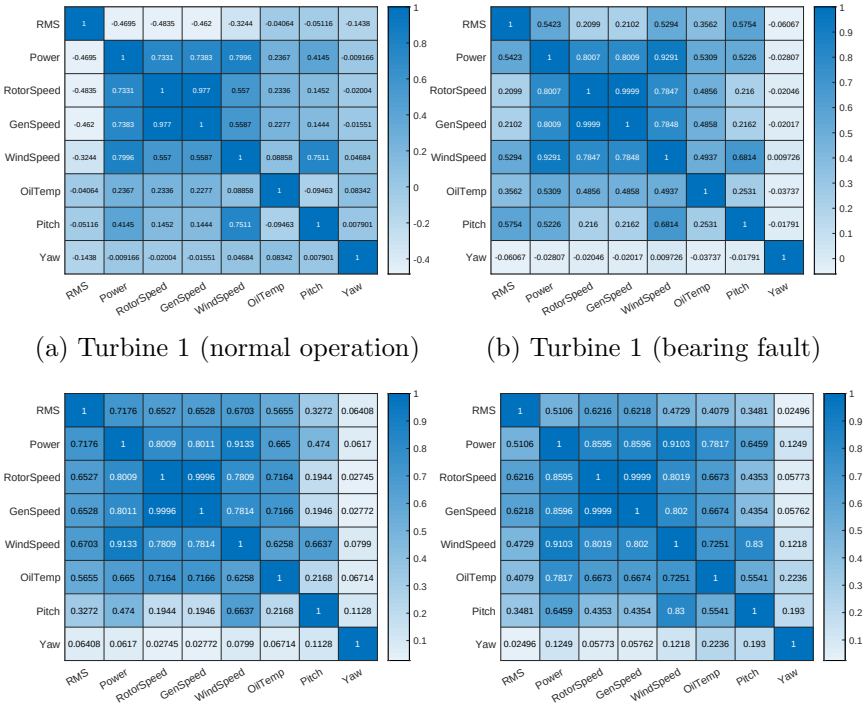
(a) RMS evolution at 70% of rated power



(b) RMS at rated power

Figure 5.16: RMS at different power levels for bearing fault

The correlation matrix of the RMS condition indicator and the various operational parameters is shown in Figures 5.17 for the two wind turbines of the case study. Two time periods are demonstrated for each turbine, one for normal operation (without known faults) and one for operation after the condition monitoring alarm was activated and before the component was replaced. As expected, the rotor and generator rpm are highly correlated, since the one is proportional to the other by a factor of the gearbox ratio. The power, wind speed, rotor speed and generator speed are also correlated, as described in the previous paragraph. The yaw is barely correlated with any of the variables, which is also expected because it depends more on the direction of the wind. The oil temperature seems to be more highly correlated to the rest of the variables during abnormal operation.



(a) Turbine 1 (normal operation)

(b) Turbine 1 (bearing fault)

(c) Turbine 2 (normal operation)

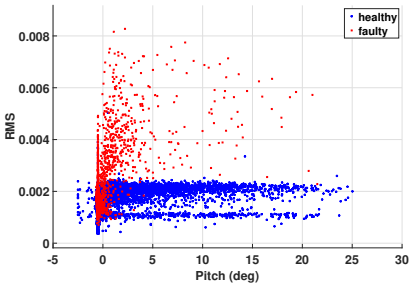
(d) Turbine 2 (gear fault)

Figure 5.17: Correlation plots of wind turbines at different health states

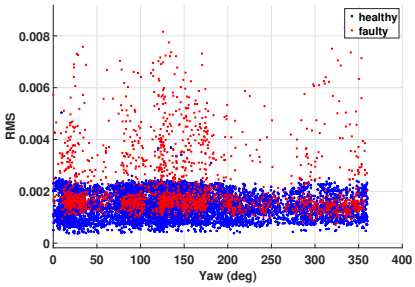
For more clarity, the change of the relationship of the condition indicators and the operational parameters during the 2 health states is shown in Figures 5.18, 5.19. It is interesting to notice that in the bearing fault case study, the condition indicator has a significant increase in higher power levels and rotor speeds, after the fault has started developing. This indicates that diagnosing faults in power bins close to rated power could improve results. In the gear tooth crack case study, the relationship is less consistent. There also seems to be a spike in the condition indicator of the faulty gear for wind speed around 7m/s, potentially due to some transient conditions that affect the operation of the turbine during the gear fault.

5.4.2 Removal of operational effects

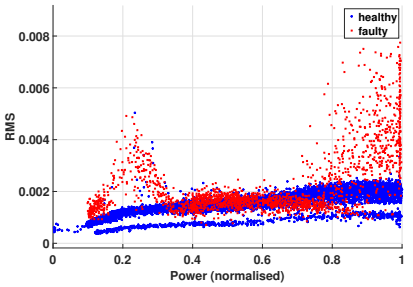
As described in the previous section, the relationship between the RMS and the Torque changes due to fault propagation. This relationship needs to be modelled dynamically. A simple linear model is proposed in [142], where condition indicators are modelled as



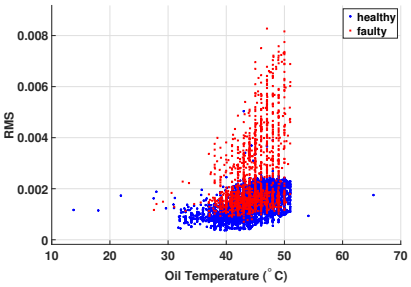
(a) RMS towards failure for different pitch angles



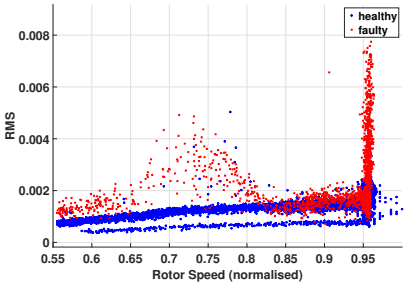
(b) RMS towards failure for different yaw angles



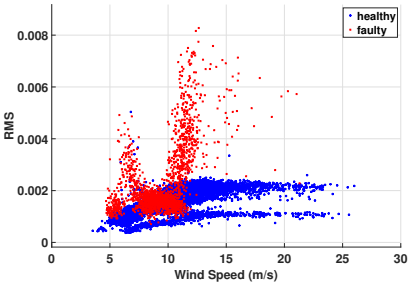
(c) RMS towards failure for different power levels



(d) RMS towards failure for different oil temperatures

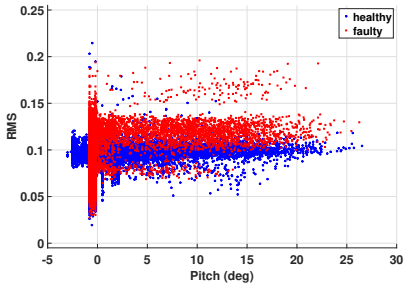


(e) RMS towards failure for different rotor speeds

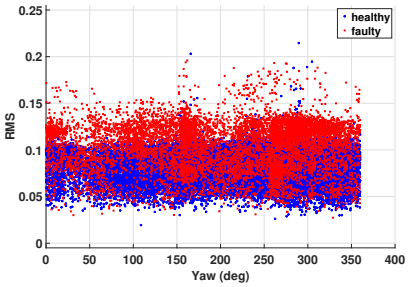


(f) RMS towards failure for different wind speeds

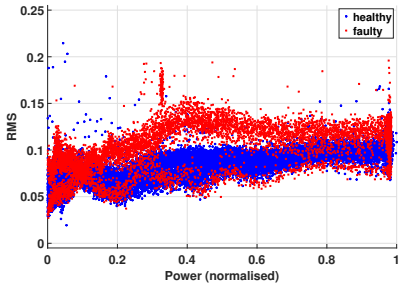
Figure 5.18: Effect of operational parameters on gear fault condition indicator for normal and abnormal operation (Turbine 1)



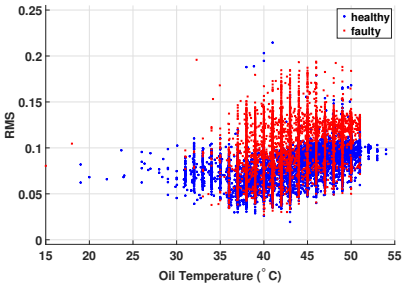
(a) RMS towards failure for different pitch angles



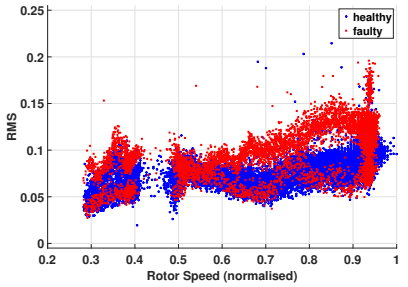
(b) RMS towards failure for different yaw angles



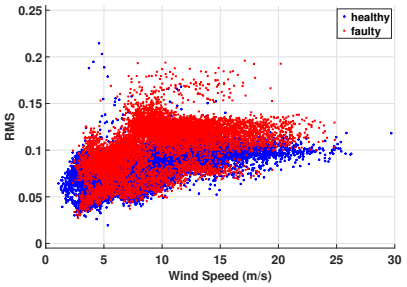
(c) RMS towards failure for different power levels



(d) RMS towards failure for different oil temperatures



(e) RMS towards failure for different rotor speeds



(f) RMS towards failure for different wind speeds

Figure 5.19: Effect of operational parameters on gear fault condition indicator for normal and abnormal operation (Turbine 2)

a least square fit regression. Other non linear models could be used in a similar way. However, as discussed in [202], the evolution of the degradation process is not captured properly in that way, because it is performed in a batch manner. Therefore, a state observer can be utilised.

Given the measurement output y and input of a system u, the state observer estimates the state variable x [203]. The plant equation as shown in Eq. (5.8).

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{5.8}$$

Where $\dot{x}$ is the rate of change of the state variable.

The observer is modelled like the plant but it also includes an additional closed loop term that takes into account for inaccuracies the A and B matrices (Eq. (5.9)).

$$\dot{\hat{x}} = A\hat{x} + Bu + K(y - C\hat{x}) \tag{5.9}$$

Where $y - C\hat{x}$ is the output prediction error of the estimated state $\hat{x}$ and K is the observer gain. A Kalman filter can be used to optimally set the gain matrix. Figure 5.20 represents the plant and state observer.

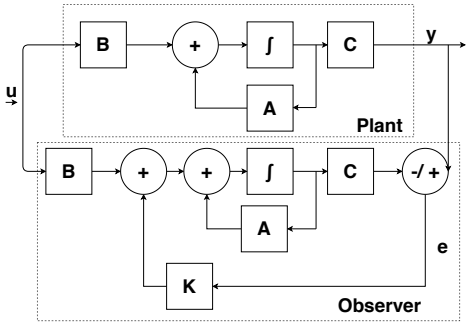


Figure 5.20: Plant and state observer

Kalman filters are commonly applied in continuously changing systems and are well suited for real time problems. Their advantages include light memory and speed- since only the previous state is used. The Kalman filter is an recursive continuous state

observer using knowledge of the systems model, process noise and measurement noise.

The Kalman filter process is divided into two steps:

1. The prediction step, which uses a previously estimated state and the linear model to predict the value of the next state as well as the state estimate covariance:

$$\begin{aligned} \hat{x}_{k|k-1} &= Ax_{k|k-1} + Bu_k \\ P_{k|k-1} &= AP_{k-1|k-1}A^T + Q \end{aligned} \tag{5.10}$$

2. The update step, which uses the current measurement of the output together with the statistical properties of the model, to correct the state estimate. The values calculated is the innovation covariance, the Kalman gain resulting in the updated state estimate and state estimate covariance:

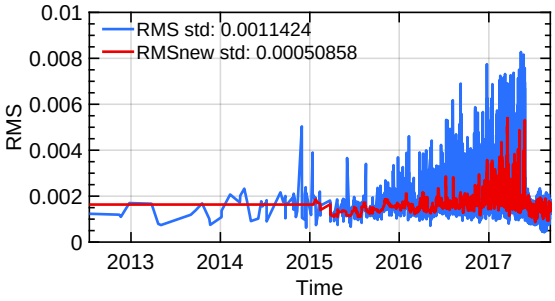
$$\begin{aligned} S_k &= CP_{k|k-1}C^T + R \\ K_k &= P_{k|k-1}C^T S_k^{-1} \\ \hat{x}_{k|k} &= Ax_{k|k-1} + K_k(Z_k - C\hat{x}_{k|k-1}) \\ P_{k|k-1} &= (I - K_k C)P_{k|k-1} \end{aligned} \tag{5.11}$$

The measurement matrix K is given by Eq. 5.12 and the transition matrix A is given by Eq. 5.13.

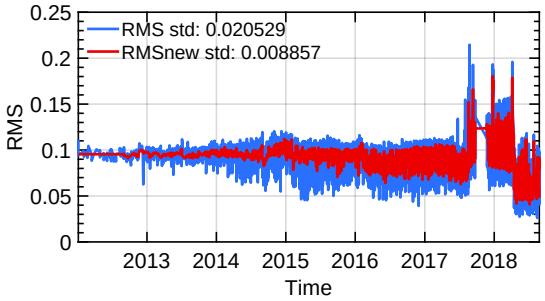
$$K = \begin{bmatrix} 1 & 0 & \text{Power} & 0 \end{bmatrix} \tag{5.12}$$

$$A = \begin{bmatrix} 1 & t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.13}$$

The two steps is repeated for every sample: $k = 1, 2, \ldots, K$. The plant noise is



(a) Filter for condition indicator (bearing fault)



(b) Filter for condition indicator (gear fault)

Figure 5.21: Recursive filter for condition indicators of gearbox faults. Standard deviation decreases.

Table 5.5: Standard deviation of RMS before and after filtering

	Std before filter	**Std after filter**	**Change**
Bearing Fault	0.001142	0.00050858	-55%
Gear Fault	0.20529	0.008857	-56%

modeled as white noise process.

The recursive filter is applied on the condition indicators presented in the previous section. The results are presented in Figure 5.21. The standard deviation is decreased after the filter application, removing the scatter due to load variations. The results of the reduction in the standard deviation (std) are given in Table 5.5. A reduction in variance of more than 50% is achieved in both cases.

5.5 Feature extraction

Various signal processing algorithms are presented in Sections 5.2, 5.3. Due to variability in vibration samples, health conditions can be captured more accurately in some cases by one signal processing method and more accurately in other cases by other. Also, different types of components (e.g gears or bearings) and fault patterns require different signal processing methods. After all, in real time fault cases, the location of the fault will not be known. Therefore, various signal processing methods should be applied in parallel to raw vibration signals from wind turbines, in order to capture all possible fault locations and failure modes. Once signals are processed, features are extracted that act as health indicators. This helps in the diagnostic and decision making process, since only a few indicators have to be analysed and tracked instead of whole vibration signals. Additionally, these features can be used as inputs in machine learning models, as analysed in Chapter 6. This section presents a signal processing pipeline that is suggested to be used on wind turbine vibration signals and a feature extraction methodology.

5.5.1 Signal Processing Pipeline

Depending on the fault location and failure mode examined, different signal processing techniques can be more effective. The paper proposes a set of parallel signal processing pipelines to reveal potential fault signatures. These are as follows:

- Order Spectrum
- Cepstrum
- Separation + Spectral Kurtosis + Envelope Spectrum

The order spectrum and cepstrum are particularly useful for gear faults, whereas the separation combined with spectral kurtosis and envelope is useful for bearing faults.

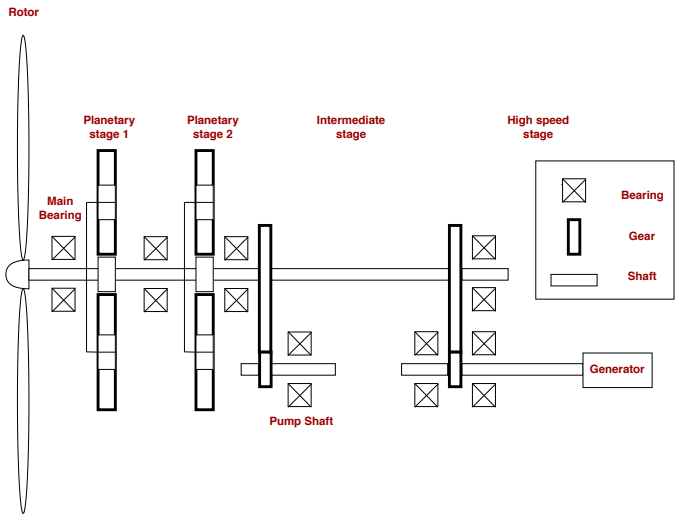


Figure 5.22: Common Wind Turbine Gearbox Configuration

5.5.2 Features

Feature extraction is used to compress high-dimensional time series (such as sensor signals) by keeping their main characteristics intact while discarding noise and removing correlations. This should speed up model training and produce better outcomes than when applied to the original, raw data.

After signals processing the next step is feature extraction from these signals. These features are health condition indicators. A summary of vibration condition indicators can be found in [141].

Most existing operating wind turbines (up to 6MW) have a conventional high speed geared drivetrain with an overall ratio of around 100:1. The gearboxes in these configurations have a one or two planetary stages and one or two parallel stages. A typical wind turbine gearbox configuration is shown in Figure 5.22. The framework presented will be demonstrated on this type of gearbox, however the proposed methodology can be applied to any other type.

Fault characteristic frequency is a widely useful fault characterizing feature in rotating machine fault diagnosis because the fault induced features repeat themselves periodical as the rotational frequency is a constant value. As for the gearbox, the meshing frequency and its harmonic, together with the sidebands around them are of

Table 5.6: Fault characteristic frequencies of different components of the planetary gearbox and the spectral lines distribution around the meshing frequency or the resonance frequency

Component	**Fault Frequency**	**Sideband**
PLC 1	f_{plc1}	-
Sun 1	$f_{s1} = (1 - \frac{Z_{r1}}{Z_{s1}}) f_{plc1} - f_{plc1}$	$f_{s1} Z_{s1} \pm k f_{s1}$
Planet 1	$f_{p1} = \frac{Z_{s1}}{Z_{p1}} f_{s1}$	$f_{s1} Z_{s1} \pm k f_{p1} \pm l f_{plc1}$
Ring 1	$f_{r1} = \frac{Z_{s1}}{Z_{r1}} f_{s1}$	$f_{s1} Z_{s1} \pm k f_{r1}$
PLC 1 Bearing	$f_{plc1,o} = \frac{P}{B}\left(1 - \frac{B}{P}\cos(\theta)\right)$	$f_n \pm k f_{plc1,o}$
	$f_{plc1,i} = \frac{n_{plc1}}{2}\left(1 + \frac{B_{plc1}}{P_{plc1}}\cos(\theta_{plc1})\right)$	$f_n \pm k f_{plc1,i}$
	$f_{plc1,b} = \frac{n_{plc1}}{2}\left(1 + (\frac{B_{plc1}}{P_{plc1}}\cos(\theta_{plc1}))^2\right)$	$f_n \pm k f_{plc1,b}$
Planet 1 Bearing	$f_{p1,o} = \frac{n_{p1}}{2}\left(1 - \frac{B_{p1}}{P_{p1}}\cos(\theta_{p1})\right)$	$f_n \pm k f_{p1,o}$
	$f_{p1,i} = \frac{n_{p1}}{2}\left(1 + \frac{B_{p1}}{P_{p1}}\cos(\theta_{p1})\right)$	$f_n \pm k f_{p1,i}$
	$f_{p1,b} = \frac{P_{p1}}{B_{p1}}\left(1 + (\frac{B_{p1}}{P_{p1}}\cos(\theta_{p1}))^2\right)$	$f_n \pm k f_{p1,b}$
PLC 2	$f_{plc2} = f_{s1}$	-
Sun 2	$f_{s2} = (1 - \frac{Z_{r2}}{Z_{s2}}) f_{plc2} - f_{plc2}$	$f_{s2} Z_{s2} \pm k f_{s2}$
Planet 2	$f_{p2} = \frac{Z_{s2}}{Z_{p2}} f_{s2}$	$f_{s2} Z_{s2} \pm k f_{p2} \pm l f_{plc2}$
Ring 2	$f_{r2} = \frac{Z_{s2}}{Z_{r2}} f_{s2}$	$f_{s2} Z_{s2} \pm k f_{r2}$
PLC 2 Bearing	$f_{plc2,o} = \frac{P}{B}\left(1 - \frac{B}{P}\cos(\theta)\right)$	$f_n \pm k f_{plc2,o}$
	$f_{plc2,i} = \frac{n_{plc2}}{2}\left(1 + \frac{B_{plc2}}{P_{plc2}}\cos(\theta_{plc2})\right)$	$f_n \pm k f_{plc2,i}$
	$f_{plc2,b} = \frac{n_{plc2}}{2}\left(1 + (\frac{B_{plc2}}{P_{plc2}}\cos(\theta_{plc2}))^2\right)$	$f_n \pm k f_{plc2,b}$
Planet 2 Bearing	$f_{p2,o} = \frac{n_{p1}}{2}\left(1 - \frac{B_{plc1}}{P_{plc1}}\cos(\theta_{p2})\right)$	$f_n \pm k f_{p2,o}$
	$f_{p2,i} = \frac{n_{p1}}{2}\left(1 + \frac{B_{p1}}{P_{p1}}\cos(\theta_{p2})\right)$	$f_n \pm k f_{p2,i}$
	$f_{p2,b} = \frac{P_{p1}}{B_{p1}}\left(1 + (\frac{B_{p1}}{P_{p1}}\cos(\theta_{p2}))^2\right)$	$f_n \pm k f_{p2,b}$
HS Gear	$f_{hsg} = f_{s2}$	$f_{hsg} Z_{hsg} \pm k f_{hsg}$
HS Pinion	$f_{hsp} = f_{hsg}\frac{Z_{hsg}}{Z_{hsp}}$	$f_{hsg} Z_{hsg} \pm k f_{hsp}$
HS Bearing 1	$f_{hsb1,o} = \frac{n_{p1}}{2}\left(1 - \frac{B_{hsb1}}{P_{hsb1}}\cos(\theta_{hsb1})\right)$	$f_n \pm k f_{hsb1,o}$
	$f_{hsb1,i} = \frac{n_{hsb1}}{2}\left(1 + \frac{B_{hsb1}}{P_{hsb1}}\cos(\theta_{hsb1})\right)$	$f_n \pm k f_{hsb1,i}$
	$f_{hsb1,b} = \frac{P_{hsb1}}{B_{hsb1}}\left(1 + (\frac{B_{hsb1}}{P_{hsb1}}\cos(\theta_{hsb1}))^2\right)$	$f_n \pm k f_{hsb1,b}$
HS Bearing 2	$f_{hsb2,o} = \frac{n_{p1}}{2}\left(1 - \frac{B_{hsb2}}{P_{hsb2}}\cos(\theta_{hsb2})\right)$	$f_n \pm k f_{hsb2,o}$
	$f_{hsb2,i} = \frac{n_{hsb2}}{2}\left(1 + \frac{B_{hsb2}}{P_{hsb2}}\cos(\theta_{hsb2})\right)$	$f_n \pm k f_{hsb2,i}$
	$f_{hsb2,b} = \frac{P_{hsb2}}{B_{hsb2}}\left(1 + (\frac{B_{hsb2}}{P_{hsb2}}\cos(\theta_{hsb2}))^2\right)$	$f_n \pm k f_{hsb2,b}$

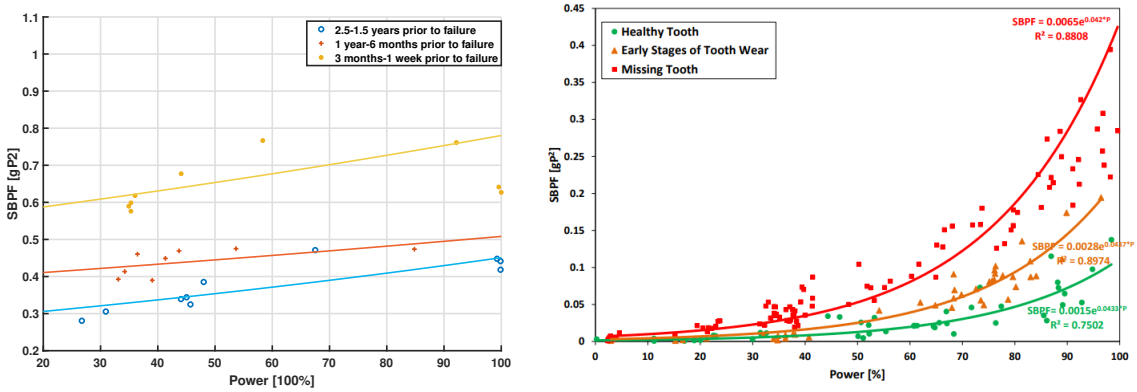
great importance in the corresponding fault diagnosis [204]. Table 5.6 summarizes the equations of fault characteristic frequencies of different localized faults and the distribution rules of these fault characteristic frequencies in the corresponding spectrum (as in [127]).

The features calculated are descriptive statistics (e.g. RMS) and are shown in Table 5.7. These features are calculated based on the fault frequencies and the narrowband sidebands. As far as the FFT is considered, for gears the first 2 harmonics of the gear mesh frequency are considered, with a narrowband of up to 6 sidebands on either side. For bearings, the first 5 harmonics of the bearing fault frequency with a 5% window are considered. Regarding the cepstrum, a 5% around the respective quefrencies is taken and regarding the envelope spectrum, the broadband statistical features are calculated.

Table 5.7: Calculated features from processed signals and operational parameters used as inputs in the machine learning model.

Feature	**Calculation**
RMS	$\sqrt{\frac{1}{N}\Sigma_{i=1}^{N}A(i)^2}$
Kurtosis	$\frac{\sum_{i=-6}^{6}A_{SB,i}}{A_F}$
Crest Factor	$\frac{max\|A(i)\|}{\sqrt{\frac{1}{N}\Sigma_{i=1}^{N}A(i)^2}}$
Peak-to-Peak	$Max(A(i)) - Min(A(i))$
Root Sum of Squares	$\sqrt{\Sigma_{i=1}^{N}\|A(i)\|^2}$
Power	Provided channel
Speed	Provided channel

Some features regarding gear sideband analysis can be calculated from the order spectrum (as introduced in Section 5.2.2 and these are summarised in Table 5.8. The SBPF feature is shown in Figure 5.23 for both field (Figure 5.23a- own elaboration) and seeded-fault in test bed (Figure 5.23b [145]). In both cases the features show an exponential behaviour with respect to load and there is a progressive trend towards failure. Therefore, this indicates that methods developed using lab tests can be applied in field data for diagnosis and prognosis, but a more statistically significant dataset from operating wind turbines should be used to validate this.



(a) Field data

(b) Seeded-fault tests [145]

Figure 5.23: Influence of the fault severity and the variable load operating conditions on the SBPF values

Table 5.8: Calculated gear features from processed signals and used as inputs in the machine learning model.

Feature	Calculation
SER	$\frac{\sum_{i=-6}^{6} A_{SB,i}}{A_F}$
SAP	$\frac{\sum_{i=1}^{N} \lvert A(i)\rvert^2}{N}$
SMLF	$\frac{\sum_{i=-1}^{1} A_{SB,i}}{\mathrm{std(x)}}$
SBPF	$2A_F^2 + \sum_{i=-5}^{5} A_{SB,i}^2$

5.6 Conclusions

This chapter studied various vibration analysis algorithms for wind turbine gearbox diagnostics and explored their effectiveness.

The signal processing algorithms were compared using samples from real wind turbine run-to-failure historic data. Both gear and bearing faults are examined, and due to the nature of kinematics and fault progression, different signal processing algorithms have proven to be more effective for each gearbox component. For gears, sideband analysis and cepstrum seem to have promising results. For bearings, analysis can prove to be more challenging since their fault signatures are not deterministic and often masked by more dominant components (such as gears), especially in planetary stages. In that

case, signal separation, filtering techniques and envelope analysis can be utilised.

Vibration signals are affected by environmental and operating parameters, which in wind turbines can exhibit large variation. This can make diagnosis even more challenging, since it's hard to create a baseline model. The operating speed and torque level have a large effect -as expected- and this is further demonstrated through a case study of operating wind turbine data. The use of a recursive filter is proposed, which can reduce noise and variance.

Vibration data contain a lot of condensed information and in wind turbine drivetrains they often come as time series (of a few seconds or minutes) with very high frequency sampling rate to make sure that all possible fault signatures can be captured. To optimally make use of vibration data, instead of using them raw, features can be extracted from the signals that can act as health indicators.

This chapter suggests a signal processing pipeline that takes as inputs raw vibration signals and performs various processing techniques in parallel, so that fault signatures for different types of faults and components inside the gearbox are captured. Afterwards, statistical features are extracted based on the kinematic characteristics of the gearbox. This can be used as inputs in statistical fault detection models, as will be elaborated further in the next Chapter. The methodology presented is demonstrated on specific gearbox types, but without loss of generalisation it can be applied in any gearbox type.

Chapter 6

AI Diagnostic Models for Wind Turbine Gearboxes

6.1 Chapter Contribution

Recent developments in sensors and signal processing systems, big data management, machine learning and improvements in computational capabilities have opened-up opportunities for integrated and in-depth CM analytics, where different types of data can facilitate informed, reliable, cost-effective and robust decision making in CM [66].

Traditional CM approaches in wind industry involve rule-based systems. This involves a lot of manual analysis and can become impractical as data volume becomes larger.

This chapter aims to answer the following research question:

"How can AI techniques be utilised in wind turbine gearbox component diagnostics?"

The key contributions of the chapter are as follows:

1. An automated systematic framework for anomaly detection, fault isolation and diagnosis based on AI techniques.

2. Present both supervised methods based on historic data and unsupervised methods based on fleet data.

3. Give insight into machine learning model building and output interpretation through hyperparameter tuning and probabilistic outputs.

4. Test AI model generalisation.

6.2 Rule-Based Condition Monitoring

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 severity levels.

The motivation behind any rule-based systems is to capture expertise in a particular application domain, represent in a modular fashion, for easy transfer to enduser. In such systems, an expert's knowledge is encoded in the form of inference rules of the form given in Eq 6.1 [205].

$$\text{If} \quad E_1, E_2, E_3, ..., E_n \quad \text{then} \quad H_i \tag{6.1}$$

where E_i's are pieces of evidence or symptoms and H refers to the hypothesis or conclusions. The set of rules constitute the knowledge base of the rule-based expert system and the inference engine uses the given knowledge base in the solution of a problem.

Diagnosis and repair in maintenance engineering often require experience-based tasks, demanding skilled engineers and their problem solving abilities [206]. Domain knowledge gives a unique understanding of how systems work in the context of their surroundings, an understanding of how systems and subsystems behave when they fail along with their symptoms.

However, as the installed wind capacity grows, so does the level of available data to be analysed, which makes it impractical. It requires tremendous expert knowledge to set appropriate threshold for each monitoring parameter. These thresholds need 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 (as shown in Chapter 5), it is hardly possible to set a constant threshold without considering their correlations.

6.3 Data-driven condition monitoring

When robust rule-based models are difficult to establish for a system primarily due to its complexity, data-driven techniques can be used to learn and develop behavioral and functional models from collected data. Data-driven fault detection has become very popular because of its minimal requirement on a priori knowledge on the system being monitored, as well as the proliferation of numerous data-driven algorithms and approaches.

The core assumption of data-driven methods is that the data has characteristics, artifacts or features that can indicate or is correlated with a fault event.

The process of data-driven methods can be seen in Figure 2.6. The first step is data collection from the wind turbine. Data can include condition (e.g vibration) and operation data (such as power) giving context to condition. Afterwards, signal pre-processing is applied to remove noise and reveal fault signatures (e.g. band-pass filtering, envelope). When a large feature space needs to be analysed, feature selection or dimensionality reduction is applied to retain relevant health indicators that will be used to develop the data-driven fault detection model. Then, the training phase of data-driven models can include supervised, unsupervised or semi-supervised methods. The training phase often involves the establishment of a baseline model. For fault detection, various metrics are computed (such as similarity values from distance to baseline model). Finally, when a fault is indicated, a diagnosis step can follow.

The present Chapter discusses methods for data-driven wind turbine condition monitoring. Methods are applied on wind turbine individual units and also on fleet-wide scale.

6.4 Unit Specific Monitoring

6.4.1 Overview

Unit specific monitoring refers to techniques using historical data of a single turbine for model training. The training process aims to map the distribution of features under healthy status. This procedure often requires comprehensive historical data covering all possible working regimes of wind turbines and in known conditions/statuses. The monitoring process is performed through comparing the testing feature vector and baseline model. Supervised learning process is feasible when data of different failure modes is available. For rule-based monitoring, the optimal threshold for fault alarm can be set by experience.

The limitations of unit specific monitoring are the following [207]:

- Lack of historical data: Unit-specific data-driven fault detection model are effective when sufficient historical data and the context information of their health status are available.
- Dynamic operating conditions: The CMSs acquisition is at sparse times (daily, weekly) and not linked to operating regime. Therefore, it will take a long time to accumulate historical data of one signal of one turbine that covers all its operation regimes.
- Availability of health status context: 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.
- Threshold setting: impractical to do for each feature and failure mode, especially as data volume grows.

6.4.2 Framework

The framework proposed in this thesis is depicted in Figure 6.1. This novel framework is proposed by the authors. The steps include signal collection, signal processing, feature

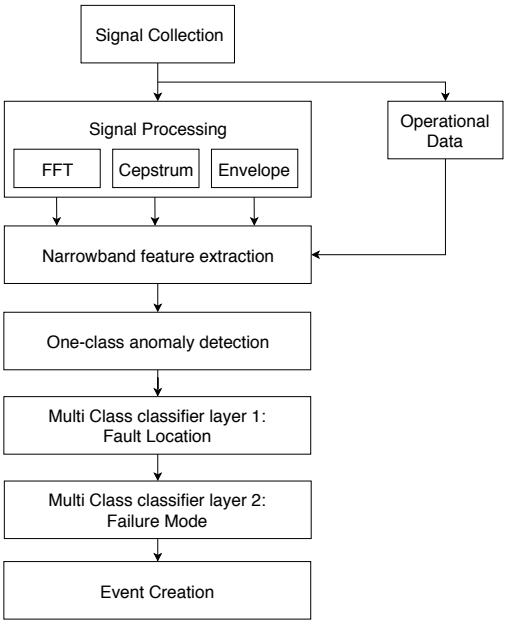


Figure 6.1: Wind Turbine Gearbox Diagnostic Framework

extraction and use of a serious of layers of pattern recognition models for anomaly detection, fault isolation and failure mode analysis.

Feature Extraction

The step of feature extraction is elaborated further in Chapter 5. Vibration signals are collected in the time domain and further processed in the frequency or time-frequency domain where condition indicators (features) are extracted. The features and their calculation can be found in Section 5.5.

Fault Detection

The first layer of the pattern recognition model is fault detection. When not enough faulty examples are available, anomaly detection models are trained based on normal operating conditions. These models can be one-class Support Vector Machines (SVMs). When a fault occurs, the outliers are outside the normal boundaries and this is flagged as an anomaly.

Table 6.1: Possible gearbox fault locations and failure modes.

Fault Location	**Failure Mode**
Sun 1	Tooth Crack
Planet 1	Tooth Wear
Ring 1	
Ring 2	
Planet 2	
Sun 2	
High Speed Gear	
High Speed Pinion	
Planet Carrier 1 Bearing	Outer Race Defect
Planet 1 Bearing	Inner Race Defect
Planet Carrier 2	Rolling Element Defect
Planet Carrier 2 Bearing	
Planet 2 Bearing	
High Speed Gear Bearing	
High Speed Pinion Bearing	
High Speed Shaft	Misalignment

Fault Isolation and Diagnosis

If a fault is present based on the anomaly detection classifier and when historic fault examples are available, models can be trained using labeled faulty data. These models are trained based on historical data, with the aim of performing gearbox incipient fault isolation and diagnosis. The model is essentially a set of two multi class classifiers. The first classifier gives the fault location and the second classifier gives the failure mode.

The potential fault locations and failure modes (and therefore the classes of the classifiers) are shown in Table 6.1.

It should be noted that the labeling of historical data can compromise the automated process of building the framework.

6.5 Fleet-Wide Monitoring

6.5.1 Overview

In fleet-wide monitoring, instead of tracking an individual machine over time, the wind turbine is compared to the rest of the fleet. This can be useful when historical data

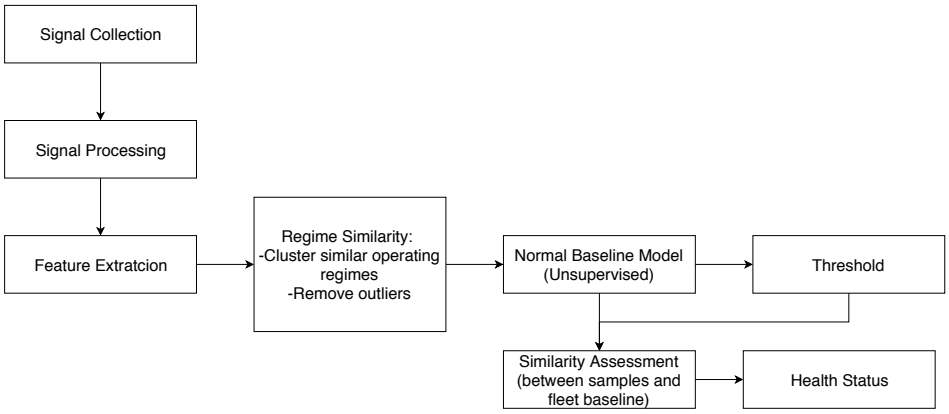


Figure 6.2: Fleet monitoring framework

from each turbine are not available for training and can often give the context in which a turbine is operating.

Given a network of similar, not necessary exactly identical, machines such as a group of wind turbines in wind farms, fleet clustering aims to aggregate units into groups. A primary motivation for performing fleet clustering is that it seeks to create local fault detection models when global models cannot be established [208]. This means it is not possible to establish a fault detection model for all wind turbines in a flee due to different operating characteristics, then turbines with similar characteristics can be aggregated into groups.

The fleet-wide monitoring methodology has two basic assumptions/limitations:

- The majority of the wind turbines in the fleet are in healthy condition.
- The wind turbines of the fleet exhibit statistically similar feature set under healthy condition.

6.5.2 Framework

The framework for unsupervised fleet-wide monitoring is shown in Figure 6.2. Each of the steps are elaborated further in the following sections.

Feature Extraction

Features are extracted systematically using the procedure described in Section 5.5.2. Vibration signals are processed from all the turbines in the fleet and condition indicators are calculated.

Regime Similarity

The first step for fleet-based data-driven health assessment is to identify outliers that are the failure samples and normal/healthy samples in the fleet data. This regime similarity process can be performed using clustering (e.g. DBSCAN) or regression (e.g. support vector regression).

Baseline Model

A clustering tendency method is used to determine whether unit aggregation is actually needed. Where there is no apparent structure in the fleet, then all the units are considered in a single group. The actual clustering process is initiated when a group structure is determined. 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 systems, rotating speed and power are the most significant operation regime parameters that can affect acceleration signals and these can be used in establishing the baseline model.

Similarity Assessment

After the baseline model is trained, new features can be used as input. The similarity of the new data with the baseline model is calculated based on some distance metric. This may be used as a health indicator.

For example, the Euclidean distance between the input vector and its cluster center can be used as the health indicator that represents the extent of similarity between

present condition and baseline , which was introduced in 4.4 in Chapter 4 and is repeated here.

$$MQE = \sqrt{e_1^2 + e_2^2 + ... + e_n^2}$$

Where n is the number of features and e the Euclidean distance between each feature and the centre. Therefore, the larger the MQE, the more critical the condition is. The contribution of each feature can be calculated based on 6.2, which can aid in diagnosis.

$$C_{feat,i} = \frac{e_i^2}{MQE} \tag{6.2}$$

Threshold Setting

The threshold for fault detection can be determined at 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 control limit should be generally optimal for the fleet [207].

Fore example, if the distance distribution for training data is assumed to be Gaussian, a threshold can be determined by putting 95% confidence level of variance for the MQE of training data.

6.6 Case Studies

6.6.1 Single Unit Monitoring

The first case study will focus on a single unit (one wind turbine) and provide a detailed demonstration of how a developed fault can be shown on the monitored parameters and how a machine learning model can be used for fault detection.

The wind turbine examined is an offshore wind turbine, rated between 2-4MW [1]. The gearbox is a planetary gearbox with 1 planetary stage and 2 parallel stages. The configuration is shown in Figure 6.3. A CMS alarm was activated on October 2015,

[1] Ranges Provided for Confidentiality Reasons

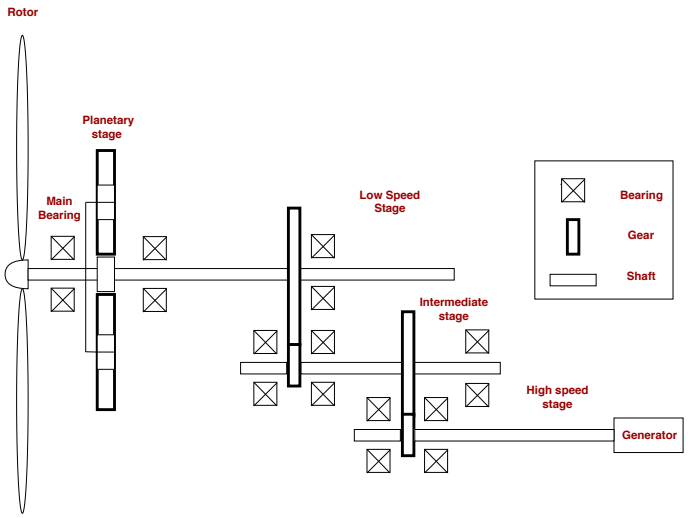


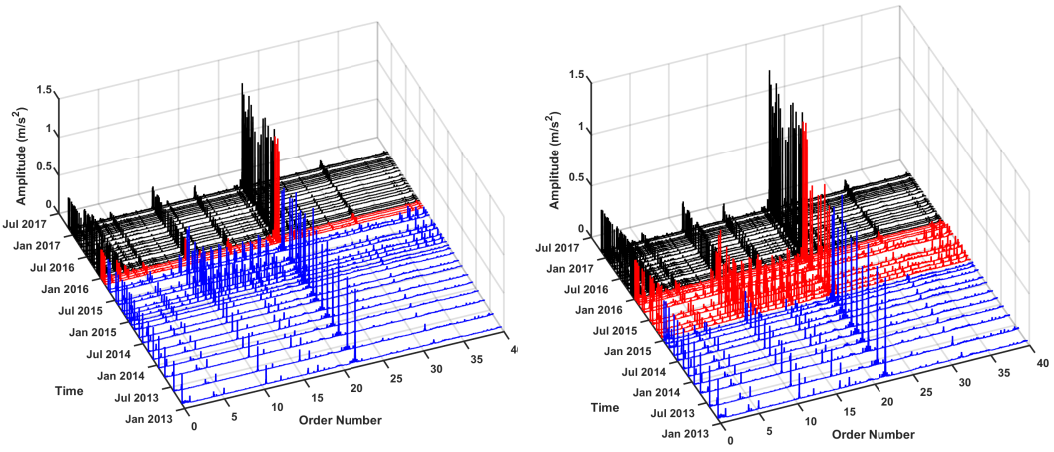
Figure 6.3: Gearbox configuration for 1 planetary stage and 2 parallel stages

indicating a high speed pinion tooth crack fault. The order spectrum waterfall of the vibration signal is shown in Figure 6.4. Maintenance took place in December 2015.

Figures 6.4a and 6.4b have colour coded the differences between rule-based and machine learning based condition monitoring. In both cases, blue represents the period that the system was considered healthy; Figure 6.4a shows in red the alarm activation until replacement. Figure 6.4b shows in red the period that the samples are labeled as faulty using expert judgement. It is clear from the spectrum that there are sideband fault indicators months before the alarm activation.

Features are extracted according to the proposed methodology and two of them are shown in Figure 6.5. The features shown are standardised Kurtosis and RMS around the gear mesh frequency, as described in 5.5. The authors decided to present a case study that shows clear separation of feature in order to better demonstrate the methodology. However, in many real world application scenarios this separation might not be that straightforward, which emphasizes the value for using machine learning techniques.

The machine learning algorithm chosen to train this pattern recognition model is support vector machines, which is powerful in terms of having a regularisation parameter (therefore avoiding overfitting) and a kernel function. 10-fold cross validation is used to assess the performance of the fitted model. Both a one-class and a two-class



(a) Order spectra leading up to failure for rule based system.

(b) Order spectra leading up to failure for machine learning based system.

Figure 6.4: Order spectra leading up to failure. Color coding shows difference between threshold of rule based system and labels on machine learning based. Blue is before alarm activation (healthy period), red is after alarm activation and black is after replacement.

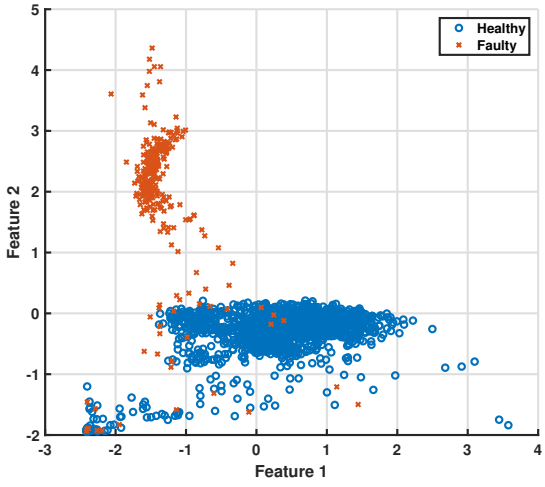


Figure 6.5: Feature separation between healthy and faulty turbine states

Table 6.2: Performance metrics for 1-class classification using SVMs

Outlier Rate	5%
Recall	94%

SVM are demonstrated in this case study.

Initially, the one-class application is presented. Only normal (healthy) data are used to train the model. The results are shown in Figure 6.6, where 2 features are depicted for visualisation purposes. The boundary separating the healthy data from the rest of the data occurs where the contour value is 0. The contours basically show the distance from the hyperplane, which is calculated for each data point thorugh the SVM model. The colour of the contour depicts that distance as shown on the figure bar. Some metrics are shown in Table 6.2. The outlier rate is the training examples that are considered out of class. The data with scores less than 0 are the outliers which are considered as faulty. Knowing the actual labels, recall is a measure that tells us what proportion of the data that actually was faulty was diagnosed by the algorithm as faulty.

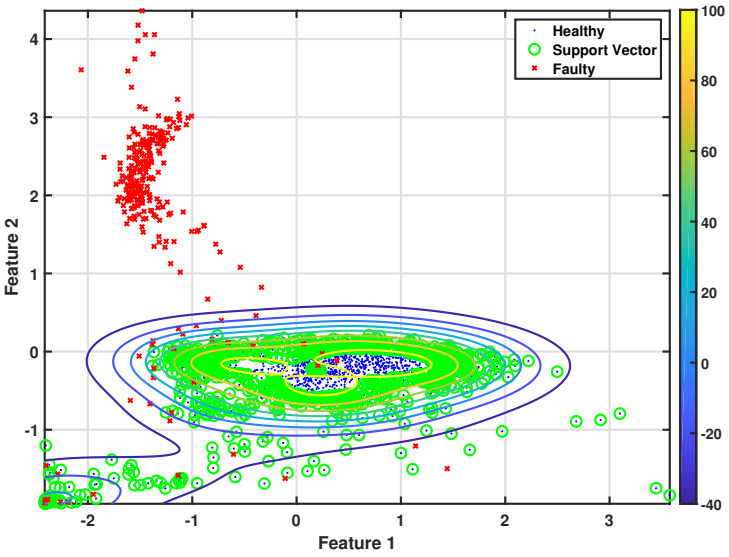


Figure 6.6: Outlier detection using one-class SVMs

Afterwards, a two-class SVM is used. Hyperparameters are optimised using Bayesian

optimisation [209] [210] [211]. A good fit is one with low cross-validation loss. The objective function is shown in Figure 6.7. The hyperparameters optimised are the regularisation parameter (C) and the radial basis function scale (γ), which were discussed in Chapter 2. The objective function computes the cross-validation loss at these parameters. The optimisation stops when the maximum set number of iterations is reached. A cost for missclassification is also specified. Double the penalty is applied to missclassification of faulty data than healthy. This is because prioritisation want to be given on minimising missed detections and also because the samples from faulty data are much less than the healthy, which is often the case in the field. The cost given will obviously yield more false alarms.

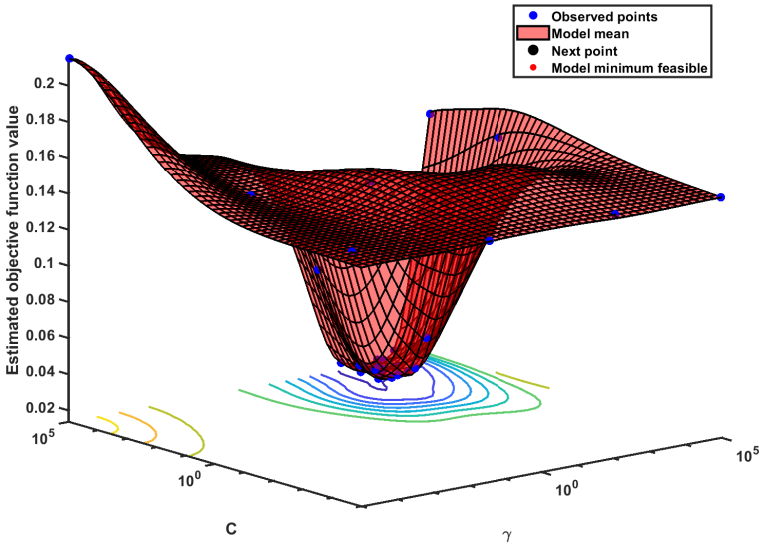


Figure 6.7: Objective function model for SVM

The confusion matrix of the trained model is shown in Table 6.3. Due to the penalty applied, all the faulty data are classified correctly, but more false alarms happen. The metrics assessing the classifier are summarised in Table 6.4.

The deterministic class assignments which are the output of the classifier model can be turned into probabilistic ones. The SVM classification score for classifying observation is the signed distance to the decision boundary ranging from $-\infty$ to $+\infty$.

Table 6.3: Confusion matrix for 2-class classification using SVMs. Horizontal= Actual, Vertical = Predicted.

faulty	245 13%	0 0%	100% 0%
healthy	32 1.7%	1610 85.3%	98.1% 1.9%
	88.4% 11.6%	100% 0%	98.3% 1.7%
	faulty	**healthy**	

Table 6.4: Performance metrics for 2-class classification using SVMs

Precision	0.99
Recall	0.94
F1 Score	0.96

A positive score for a class indicates that the observation is predicted to be in that class. A negative score indicates otherwise. The posterior probability is the probability that an observation belongs in a particular class, given the data. For SVM, the posterior probability is a function of the score that an observation belongs in class. For inseperable classes, this posterior probability can be modeled by a sigmoid function.

This can be shown in Figure 6.8, where the posterior probability contours are shown. 1 represents the probability of belonging to the healthy class and -1 represents the probability of belonging to the faulty class.

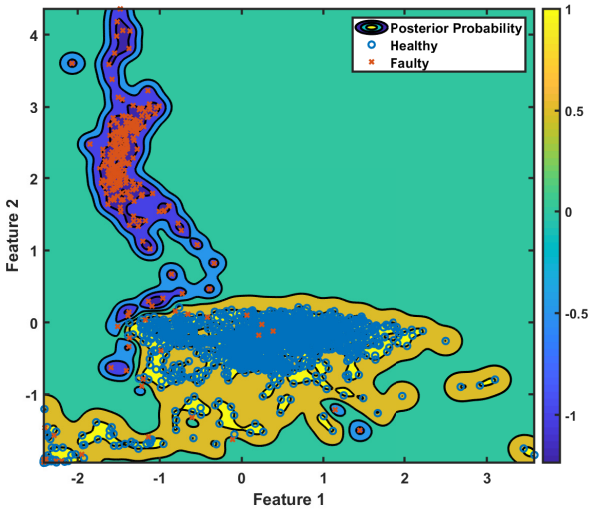


Figure 6.8: Posterior probabilities of SVM scores for fault classification

6.6.2 Population of Single Units Monitoring

This Section provides a fault detection case study for a population of wind turbine units. Models are trained and tested for a group of wind turbines located in various wind farms.

Data Description

43 wind turbines with confirmed gearbox faults are considered in this case study. The wind turbines are offshore, located in 7 different wind farms, same model and are rated between 2 and 4 MW. The gearbox examined is the same type as shown in Figure 6.9.

The confirmed gearbox fault locations from the 43 turbines presented in this case study are located on the high speed stage and are as follows:

- Generator bearing
  - General damage
- High speed bearing type 1, generator side
  - Inner race defect

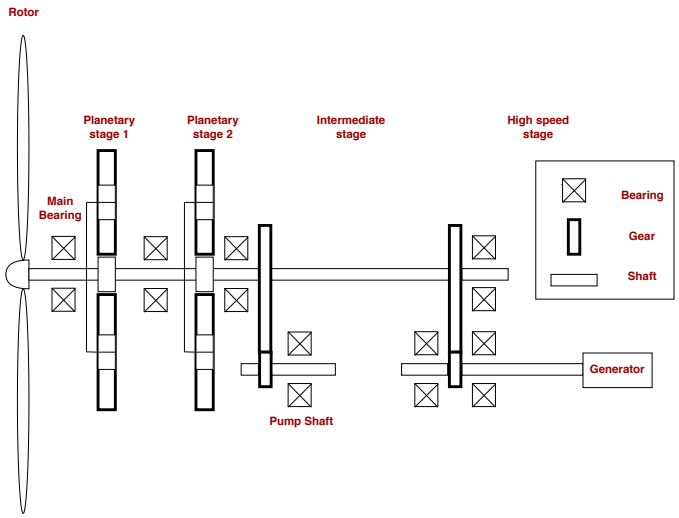


Figure 6.9: Double Planetary Stage Wind Turbine Gearbox Configuration

  - Outer race defect
  - Rolling element defect

- High speed bearing type 2, generator side
  - Inner race defect
  - Outer race defect
  - Rolling element defect
- High speed bearing, rotor side
  - Inner race defect
- High speed pinion
  - Tooth wear
  - Tooth crack
- Intermediate speed pinion
  - Tooth crack

The signals are collected from an accelerometer positioned on the top of the planetary stage in the gearbox.

Methodology Implementation

In mechanical components, like gears and bearings, the degradation can be gradual until a catastrophic failure occurs. In this case study, the historic alarms of the events are known. The labeling is done using expert knowledge, by inspecting the vibration signals and looking at when fault frequencies start to occur based on each failure example. The samples that are outside the window of the degradation period are labeled as healthy. The samples that are inside the window of the degradation period are labeled as faulty for the one class classifier. For the two layers of the multi class diagnostic classifiers, the fault location and failure mode are added as classes respectively. Often when the alarm is activated, the turbine stops. That means there are often not as many faulty examples as healthy and that in general the dataset is unbalanced for most faults. Stratified cross validation is used therefore.

An example of two vibration signals, one healthy and one with a high speed pinion fault is shown in Figure 6.10.

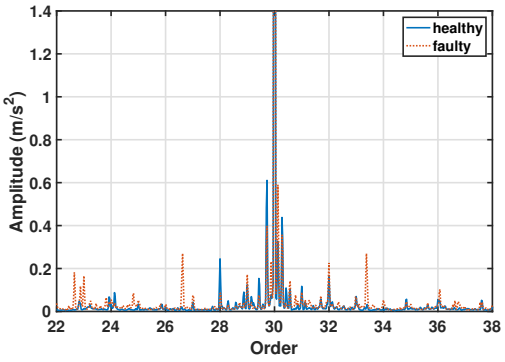


Figure 6.10: Order spectra of healthy and faulty signals

For the one-class classification, A one-class SVM classifier [212] is trained. SVMs offer superior generalisation capabilities [67]. It is a semi-supervised algorithm that learns a decision boundary using only normal data points and then testing the likelihood of a test instance being within the boundary of the learnt model. The results of the one class classification are shown in Table 6.5. The precision and recall scores indicate an adequate anomaly detection capability.

The pattern recognition model used for multi class classification is RUSboosted deci-

Table 6.5: One-class classification outlier detection

Precision	0.87
Recall	0.91

sion trees because they have proven to be a robust classifier that alleviates the problem of imbalanced classes [213]. In the present case study there are various types of fault examples used, however the fault classes have unequal number of samples since some failures were encountered in more wind turbines than others. RUSBoost is an algorithm that combines boosting and data sampling. It applies RUS, which is a technique that randomly removes examples from the majority class. The hyperparameters tuned are the maximum number of splits, the number of learning cycles and the learning rate. These hyperparameters are tuned using random grid search and cross validation.

Results

The results of the stratified cross validation are shown in normalised confusion matrices in Tables 6.6-6.8. Based on Table 6.6, the fraction of signals classified correctly based on their fault locations is satisfactory, and in most cases more than 80%. The failure examples that have scarce samples (like the intermediate speed pinion) are still more prone to missclassifications (40% for intermediate speed pinion).

The second layer of classifiers has as output the failure mode and results are shown in Tables 6.7, 6.8. The results indicate that the inner from the outer race bearing defect is more easily distinguishable than the tooth crack and wear incidents on gears. This can be justified because the symptoms for tooth cracks and wears have very similar effect on vibration signals, whereas bearing fault frequencies between inner and outer race are distinctively different. It is discussed in Chapter 2 that according to [59], the prioritized failure modes of the gearbox include both raceway wear and gear cracks.

To introduce a certain robustness cumulative counting of the classified spectra has been introduced. This illustrated for a high speed bearing and a high speed gear fault that were not included in the training set and is shown in Figures 6.11,6.12. Each time a the classification model outputs a fault, the counter adds up to this specific fault category. If another fault (or no fault) is predicted in the next sample, the counter

Table 6.6: Normalised confusion matrix with fault location.Horizontal= Actual, Vertical = Predicted.

Gen bearing	297 65%	70 15%	3 1%	77 17%	10 2%	0 0%
HS bearing type 1, gen side	8 <1%	32259 90%	65 <1%	3171 9%	440 <1%	1 1%
HS bearing type 2, gen side	0	312 11%	2144 79%	234 8%	54 2%	0
HS bearing, rotor side	7 <1%	2733 6%	102 <1%	43666 93%	642 1%	0
HS pinion	0	851 7%	27 <1%	1042 8%	11016 85%	0
IMS pinion	0	8 12%	1 1%	9 14%	8 12%	33 60%
	Gen bearing	**HS bearing type 1, gen side**	**HS bearing type 2, gen side**	**HS bearing, rotor side**	**HS pinion**	**IMS pinion**

Table 6.7: Normalised confusion matrix with failure mode for high speed bearing. Horizontal= Actual, Vertical = Predicted.

Inner Race	39809 92%	2989 7%	618 1%
Outer Race	88 3%	3152 97%	7 <1%
Rolling Element	10 2%	3 1%	474 97%
	Inner Race	**Outer Race**	**Rolling Element**

Table 6.8: Normalised confusion matrix with failure mode for high speed gear. Horizontal= Actual, Vertical = Predicted.

Tooth Crack	100 83%	21 17%
Tooth Wear	102 4%	2521 96%
	Tooth Crack	**Tooth Wear**

reduces. In Figure 6.11, the counter starts increasing before the actual start of the alarm. After the bearing is replaced and the system is assumed to return to normal operating condition, the counter starts decreasing. In Figure 6.12, the classifier does

not detect the fault earlier than the actual alarm which could be attribute to either the scarcity of training data for this fault location or to the fact that the fault had a sudden development.

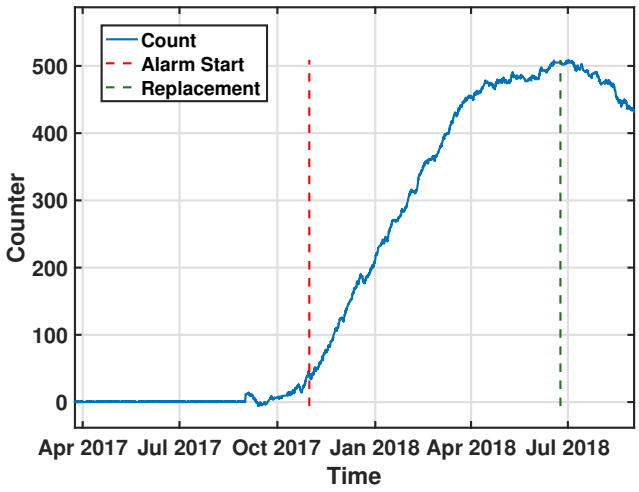


Figure 6.11: Counter for high speed bearing fault

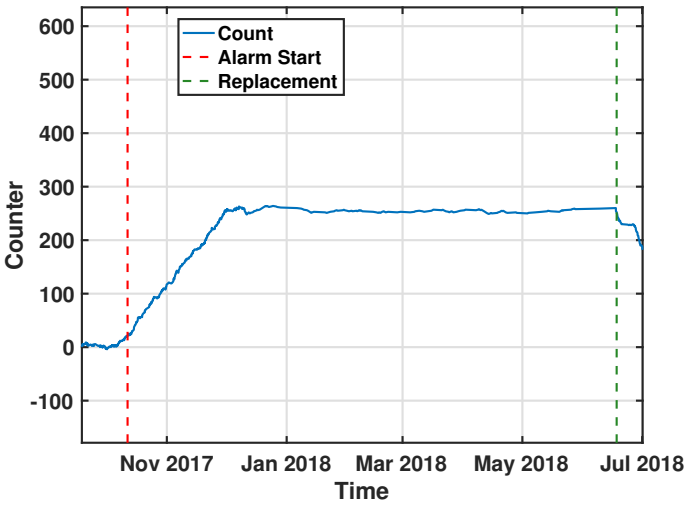


Figure 6.12: Counter for high speed gear fault

6.6.3 Fleet Based Monitoring

In this section a fleet-wide based monitoring case study is presented, using data from an operating wind farm.

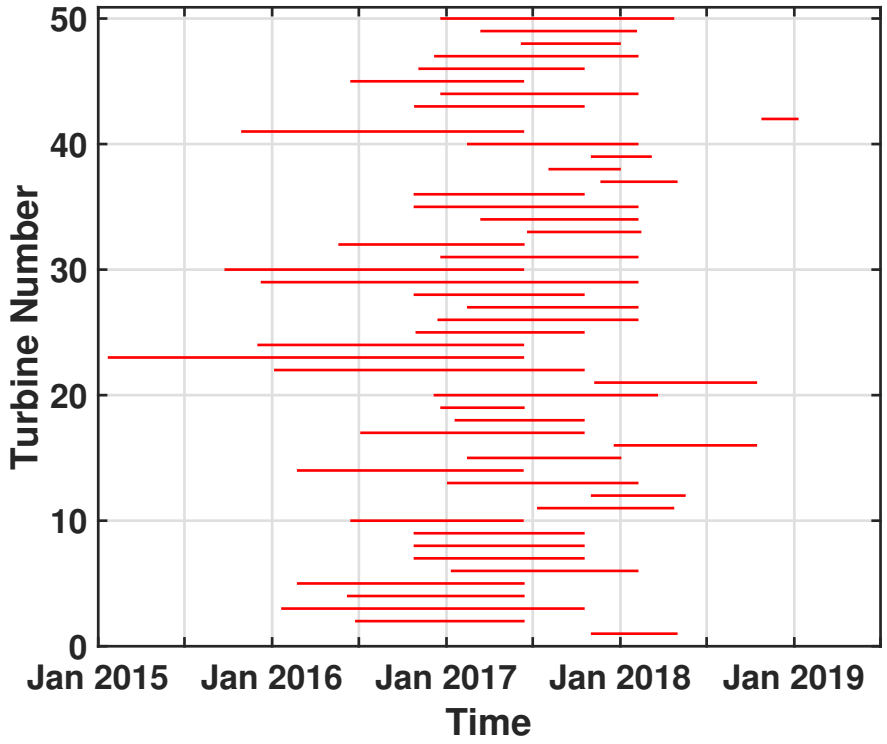


Figure 6.13: Wind farm faults

Data description

The case study involves an offshore wind farm which has more than 50 turbines. The turbines are rated between 2 and 4MW and the gearbox configuration is 2 planetary stage 1 parallel stage. For data limitation reasons, the case study will focus on the sensor of the the high speed stage of the gearbox. In this particular wind farm, there were recorded incidents of gearbox high speed bearing faults in the majority of the population of the wind farm (70% of turbines).

The recorded incidents for a sub-population of the wind farm are shown in Figure 6.13. It can be observed that there is a period when the majority of wind turbines had a high speed bearing fault. The red lines illustrate the period of the alarm being active. When the red line disappears a maintenance action has taken place.

Methodology implementation

In this case study, the obvious outliers 3 scaled median absolute deviations away from the median are removed. Each calculated feature is classified as in either normal operating condition or outlier. That way, a set of baseline data can be created and used for model training. This baseline dataset is 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 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.

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 fault detection model. In this case study, the chosen model algorithm is Self-Organizing Maps. When being used for health assessment under unsupervised learning situation, it can automatically recognize the spatial distribution of baseline data, and gives a representation of the cluster with m neurons.

The threshold for MQE can be determined from a statistical point of view. Since the kernel function 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 $C_{feat,i}$ of each feature can be further calculated by Eq. 6.2 to identify the source of variation.

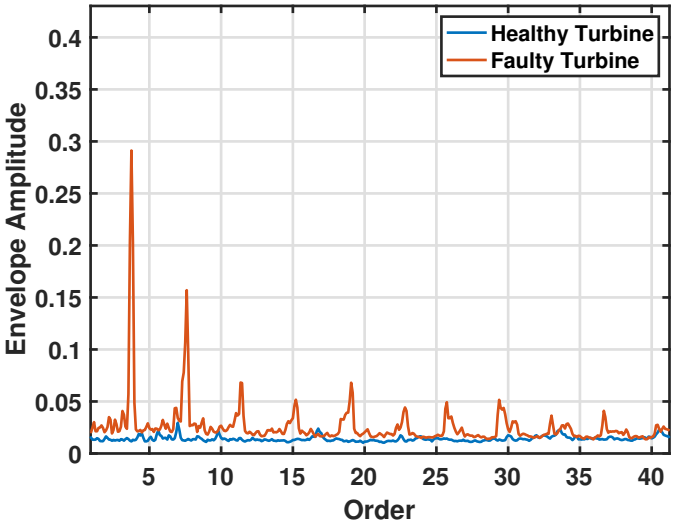


Figure 6.14: Envelope spectrum comparison for 2 different turbines of the same fleet

Results

The envelope spectrum is compared for 2 different turbines at the same time for similar operating regimes. Harmonics at the ball passing frequency are clear on the turbine with bearing fault. This is shown in Figure 6.14;

Afterwards, the RMS is extracted from the envelope spectrum and its evolution in time for rated power level is examined. As shown in Figure 6.15, there is a clear trend leading up to failure. The alarm was activated October 2018 and the bearing was replaced in end of December 2018. It is interesting to notice that the vibration level changes after replacement.

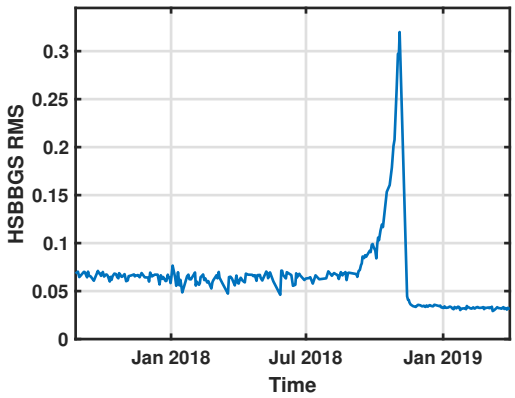


Figure 6.15: RMS timeline for turbine 36

Initially a simple observation of the data can give some some insight into the population to be analysed. A snapshot of the condition indicator (RMS) for a day in a sub-set of turbines in the wind farm is shown in Figure 6.16. This box plot shows the variation of the condition indicator over the whole fleet for various power levels. The plot shows that there is both within-machine variance and also within-fleet variance. The RMS for the whole wind farm over the same day as a function of the power level is shown in Figure 6.17.

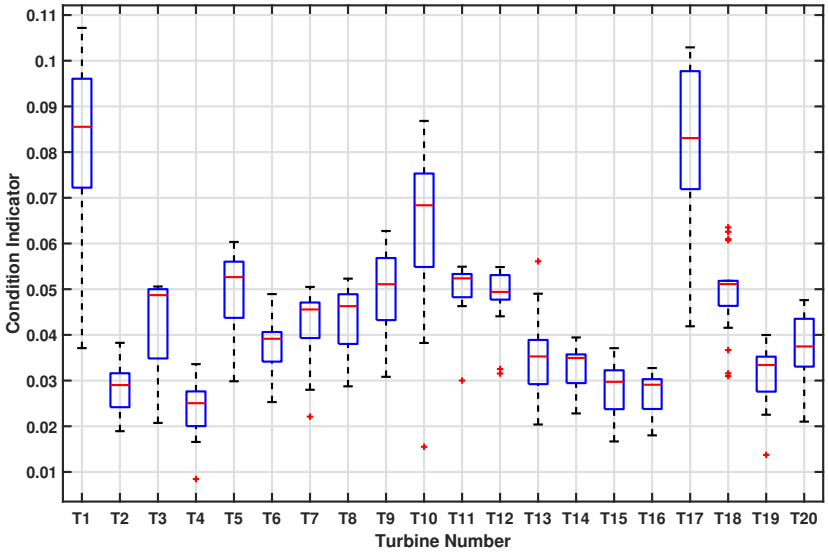


Figure 6.16: Boxplot of condition indicator for a sub-set of turbines

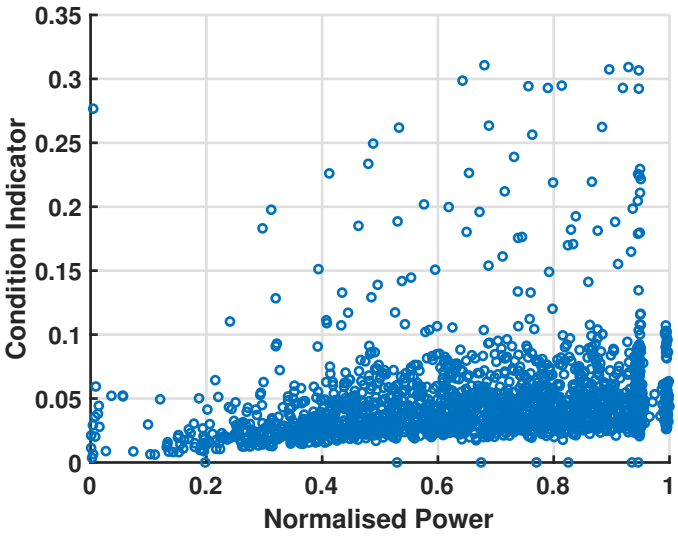


Figure 6.17: Snapshot of condition indicator

Assuming a Gaussian distribution for the condition indicator, a comparison of the condition indicator of the faulty turbine and the rest of the fleet is shown in Figure 6.18. The deviation of the turbine mean from the fleet is clearly noticeable from September 2018, a month before the alarm. As depicted in Figure 6.19 turbine mean is already almost 2 standard deviations away from the fleet at the time of the alarm. For comparison purposes, the standard deviation of a faulty turbine during 2017 away from the fleet mean is depicted in Figure 6.20. As shown in Figure 6.13, most of the turbines have an alarm activated during this period.

The results of the support vector regression are shown in Figure 6.21. The 95% upper confidence bound is also shown and the outliers are removed. The rest of the data consists the set that is used to train the baseline model.

After the baseline model is trained, health assessment is performed on the wind farm, by testing each turbine and calculating the MQE. The results of a subpopulation of the farm are shown in the colourmap in Figure 6.22. It is clear that turbine 36 exceeds the threshold, and therefore is diagnosed as faulty. The contribution of features is shown in Figure 6.23. HSBBGS stands for features related to high speed ball bearing generator side, HSRBGS stands for features related to high speed roller bearing generator side and HSRBRS stands for features related to high speed roller bearing generator side.

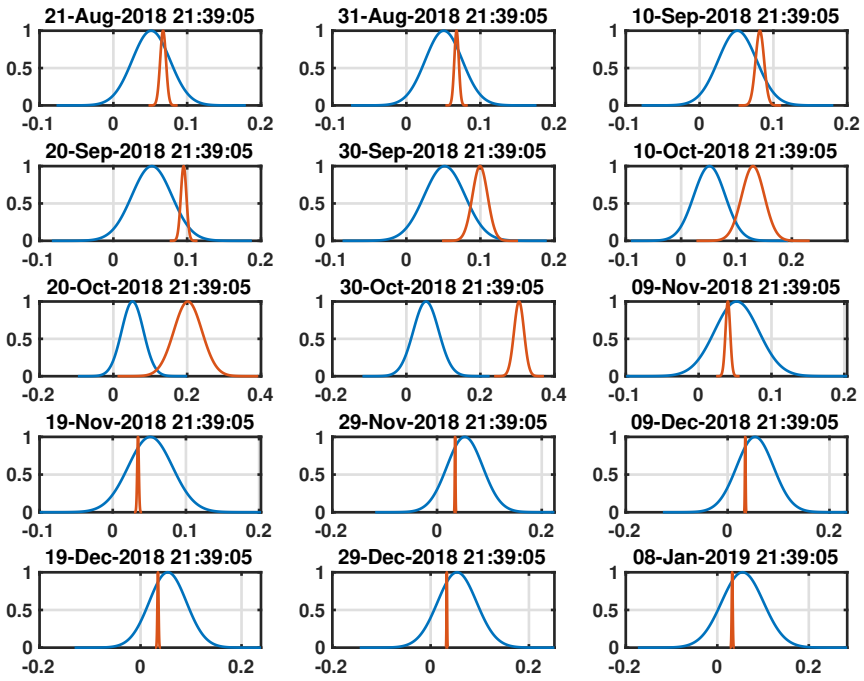


Figure 6.18: Distribution of farm and faulty turbine condition indicator (blue for farm, orange for faulty turbine). Replacement occurred in November 2018.

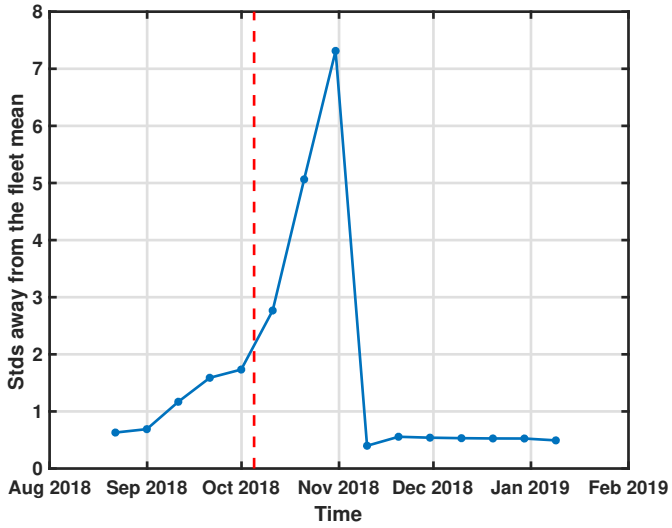


Figure 6.19: Standard deviations away from fleet mean. Majority of turbines are healthy. Dashed line shows date of CMS alarm activation.

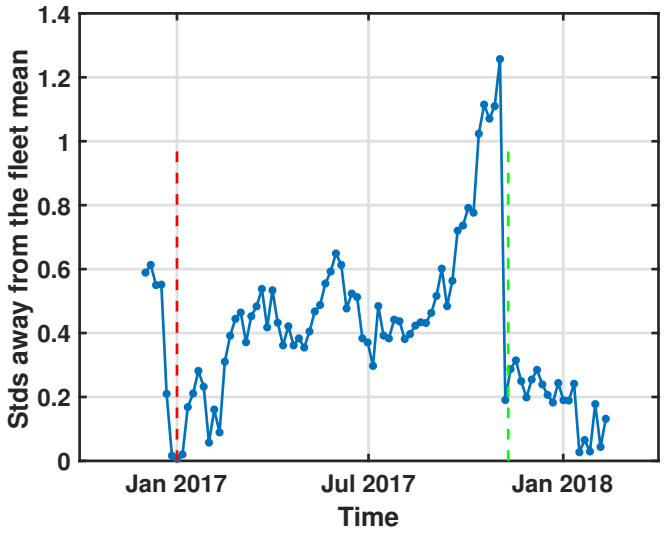


Figure 6.20: Standard deviations away from fleet mean. Majority of turbines have a fault. Red dashed line shows date of CMS alarm activation and green dashed line shows date of replacement.

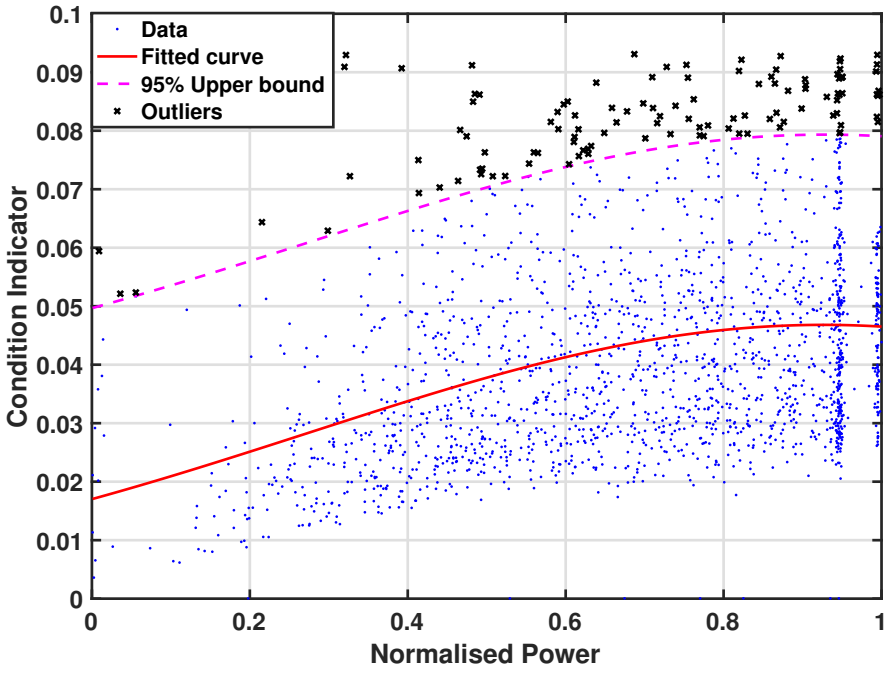


Figure 6.21: Regression results of condition indicator

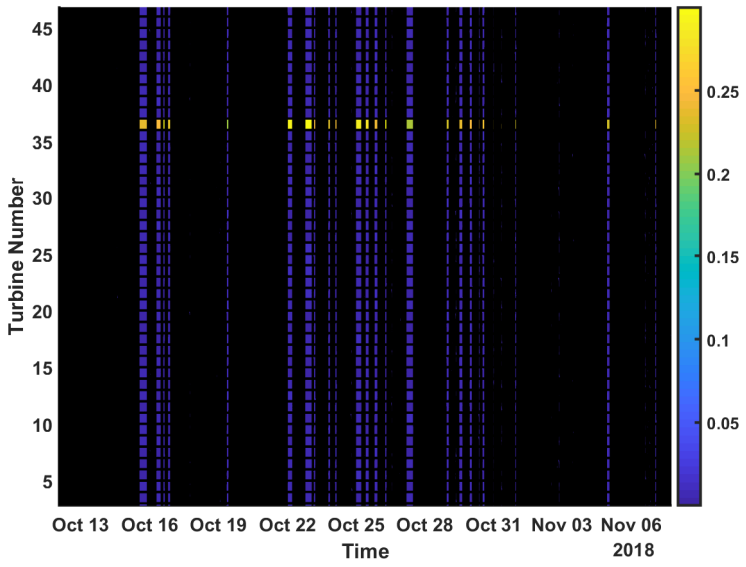


Figure 6.22: Health assessment for a sub-population of the fleet of turbines

It is confirmed that turbine 36 had a high speed bearing fault during that period and the fleet-based system detected the fault 3 weeks before the actual system alarm.

6.7 Conclusions

This chapter presented a number of possible contributions of AI in the diagnostic process of wind turbine gearboxes. An automated framework based on signal processing, feature extraction and training of machine learning models is presented. The models can either be fleet based or unit specific based.

The capabilities of using AI models are demonstrated through the various case studies. Supervised, unsupervised and semi-supervised models can be used to predict failures. Weights on false negatives or false positives can be assigned depending on the desired design of the model. Deterministic class assignments can be assigned to probabilities, for further interpretation of results.

In unit specific based monitoring, historical data are used to train machine learning models. Initially, normal (healthy) data can be used to train one-class classification models. These models learn the behaviour of normal data and in the case of SVMs,

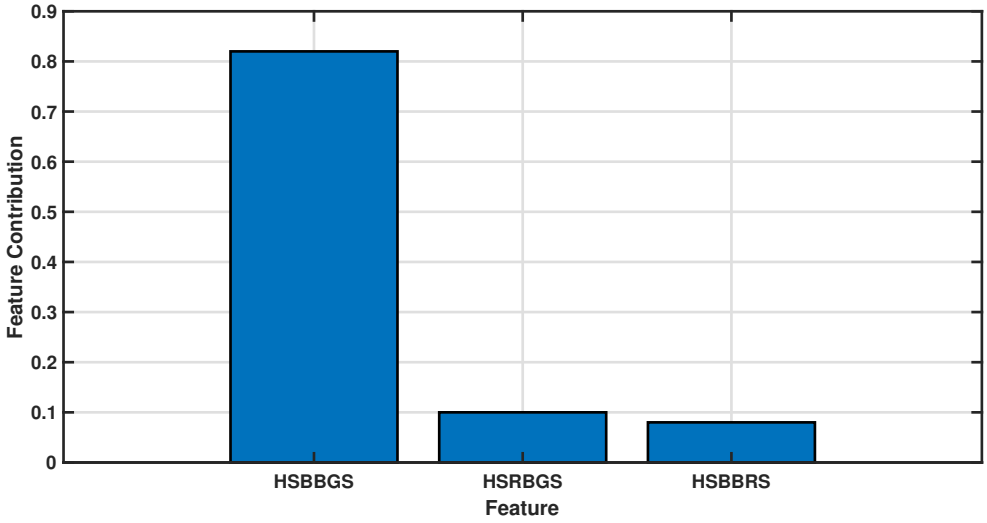


Figure 6.23: Feature contribution

create a boundary around them. If training is performed successfully, then when new faulty data are used as input, they are flagged as anomalies. If historical examples of faults are available, these can be used to train supervised classification models. The supervised classification models can perform fault isolation and diagnosis, giving as output the fault location of the gearbox and the failure mode. The main limitations of unit specific models is the amount of training data required, the historic failure example availability and the generalisation capabilities.

In fleet based monitoring, no historical data are required for model development. Instead, the turbines are compared to a fleet of similar machines and fault detection is performed based on the principle that if a turbine deviates from the rest of the fleet, then there is a potential fault. The main limitations of this methodology is that it relies on the assumption that the majority of the fleet is healthy. It is demonstrated in the case study that the fleet based monitoring framework can have successful monitoring capabilities. However, there was also a case when the majority of the fleet suffered from a fault and thus the limitation of the methodology are further proven.

It's hard to debate whether unit-specific based or fleet based monitoring is more useful. Both methods have their advantages and limitations, as highlighted throughout the chapter. Fleet based monitoring can give the context in which a wind turbine is

operating, but unit-specific monitoring could potentially contain more information on fault development. Fleet based monitoring can potentially implemented more easily but relies on having enough similar operating machines without a fault. Unit-specific monitoring might require more failure examples. Both models can be used by operators for decision making.

All in all, the traditional methods of rule-based monitoring for wind turbine gearboxes is -up to now- the industry standard. These methods are quite established and easily interpreted, because an alarm is activated when a threshold of some specific measured system parameters is exceeded. It is therefore easy for the diagnosis engineers to analyse the potential fault source and recommend maintenance activities. However, the volume of data for analysis for wind farms is expected to grow significantly within the following years. This is due to the fact that the installed wind capacity grows and also data transmission methods and storage capabilities are improving because technological advances. Therefore, rule-based methods can become more impractical.

AI techniques have received a large interest over the past decades, both in theoretical aspects and also in some various applications of different fields (such as image recognition or market segmentation). There have also been some applications in the field of condition monitoring, but in wind energy this is not yet the industry practice. There have been some case studies in the literature proving the significance of AI, but most of them rely on synthetic data or laboratory data. Even in the few studies that have used operational data from real wind turbines, it's difficult to generalise the effectiveness of the results because most of them rely on small datasets from specific operational conditions. Also, interpreting and tuning machine learning models can be difficult, therefore they are often used as black box. There are indications that AI can be effective in wind turbine gearbox condition monitoring. The present thesis has demonstrated some promising results using the largest dataset of operational wind turbine vibration data. There are various things that can influence the effectiveness and accuracy of machine learning models- including volume and quality of training data, availability of labeled data, process of labeling, feature extraction, choice of machine learning algorithm and training process of machine learning model (e.g. hyperparameter tuning). It is recom-

mended that these models are built by experts who have both domain knowledge of wind turbine gearbox systems and solid background in data analysis techniques.

Chapter 7

Conclusions and Future Work

Based on the problem statement that the repair cost and downtime of the wind turbine gearbox is too high, the objective of this thesis was to answer the following research question:

"How can incipient wind turbine gearbox faults be predicted based on a combination of condition monitoring data and the use of AI before catastrophic failure occurs?"

To answer this primary research question a number of other smaller secondary research questions were first answered. These secondary research questions were answered throughout each chapter of this thesis. This chapter provides a summary of the major concluding points from each of the previous chapters in this thesis. The main research question from each chapter and the overall research question shown above will be addressed. More detailed conclusions, discussions or answers to secondary research questions can be found at the end of each chapter. The conclusions shown in this chapter will be based on certain assumptions and will have certain levels of uncertainty around them. Details of these assumptions and uncertainty levels can be found in the chapters themselves. This chapter will conclude by providing an overview on how the work in this thesis will contribute to the wind industry and by outlining future work.

7.1 Conclusions

The conclusions from each chapter are summarised in this section and some overall discussion points are highlighted.

7.1.1 Conclusions from Chapter 2

Wind turbine gearbox reliability has been a subject of interest in the literature. It might not be the component with the highest failure rate, but it has one of the highest impacts on downtime and repair costs. Efficient monitoring of gearboxes provides therefore a large opportunity for operation and maintenance cost reduction. Current industry practice mainly involves threshold setting, which can be sub-optimal in terms of data volume and non-linearity of parameters. There have been many studies on vibration condition monitoring and SCADA data analysis. However, there have not been enough studies on providing a framework for data processing, feature extraction and the use of machine learning techniques for gearbox fault detection, isolation and diagnosis. This is addressed in the current thesis.

7.1.2 Conclusions from Chapter 3

"What are the most critical failure rates and failure modes of modern multi MW wind turbine gearbox components?"

Wind turbine drivetrain failures need to be better understood in order to improve the manufacturing process, invest in more efficient condition monitoring systems and perform more accurate reliability analysis. It seems that high speed stage bearing failures are critical for geared wind turbines. However, most full gearbox replacements happen due to faults in the planetary stage. This thesis addressed both faults.

7.1.3 Conclusions from Chapter 4

"Can SCADA systems be used for wind turbine gearbox fault detection and isolation?"

SCADA data are a cost-efficient way to monitor the condition of wind turbines. Operational anomalies are contextual in wind turbines, which means that a change in

a parameter is relative to other parameters in the system. There have been various mathematical models applied in the literature, like trending clustering and normal behaviour modeling. Some of these methods have been applied and demonstrated in this thesis on wind turbine gearboxes. Anomaly detection of an incipient fault can be successfully achieved and a monitoring framework is proposed.

The challenge when it comes to wind turbine gearboxes, where multiple sub-assemblies that are dependent on each other, diagnosis can be more challenging. A temperature deviation in the system does not make fault isolation easy, since heat propagates. This can be especially challenging when there are limited number of sensors installed. Also, training models and choosing a normal period can significantly affect the accuracy of the results. There has to be a balance between a sufficient amount of training data that cover all operational regimes and environmental conditions but also not including degradation data in the healthy period of the dataset.

In field applications, it hard to know when a fault started developing, even when historical examples are available. It was demonstrated in the case study that even when there are examples from the same failure mode from different turbines, there is a vast variation in the data structure and fault development.

7.1.4 Conclusions from Chapter 5

"What are the most effective vibration analysis algorithms for wind turbine gearbox diagnostics?"

Vibration data are the most efficient and commonly used source of wind turbine drivetrain condition monitoring. Vibration signals are collected in high sampling frequency and small time intervals, giving information about drivetrain dynamics. Fault signatures can often be revealed through vibration data, however this is often dependent on the signal processing algorithm utilised. The chapter explained and demonstrated on real wind turbine data a few of the most commonly used signal processing algorithms for wind turbine drivetrain condition monitoring. The various methods for detecting gear and bearing faults are discussed. Bearing faults often pose more challenges in terms of diagnosis, because their signatures are non deterministic and often masked by

gear components. Order, cepstrum and envelope analysis have proven to be effective in analysing vibration signals.

As the volume of vibration data grows, advanced data analytic techniques should be employed. Feature engineering is a very important step in data analytics. In the context of fault diagnosis and prognosis, features contain useful and often easily interpreted information on fault propagation. A systematic and automated feature extraction process is proposed, based on physics of failure knowledge and drivetrain kinematic data.

Vibration signals can exhibit a lot of variation and sensitivity to external parameters, such as operational conditions. Load (exhibited through rotational speed and power output) is the most important of them, since it significantly affects gearbox operations.

7.1.5 Conclusions from Chapter 6

The large amount of data streams coming through wind turbines and the non-linearities resulting from load variations render traditional rule based methods less efficient. AI techniques can be utilised to improve vibration condition monitoring and maintenance decisions on wind farms. The models can be supervised or unsupervised and also unit-based or fleet based. The suggested frameworks using AI models are applied on a large amount of operational wind turbine historic data with known gearbox faults.

For supervised models, the amount of historical failure example data required could be a barrier for large scale implementation. For unsupervised and semi-supervised models, the main barrier is assumption of normality state of the data- either for the majority of the fleet in the case of fleet-wide monitoring or for the trained period in the case of unit-specific semi-supervised monitoring.

The effectiveness, sensitivity and generalisation of AI models in wind turbine condition monitoring needs to be better understood through more case studies.

7.1.6 Overall discussion

Wind turbines operate under varying speed and load, which results in non-linearity for the monitored parameters. That's one of the reasons that setting thresholds can be sub-optimal. Also, a model trained on an individual turbine might not be applicable to the other turbine, even if it is the same machine model.

Regarding the machine learning models, the most important part in machine learning process is understanding the data before building the model, the physics behind it and its relation to the desired output. Each algorithm is different in terms of what kind of data and what problem setting it works best for. There is no algorithm that is universally the best for all datasets, therefore the choice must be made according to the problem.

The end goal for fault detection is finding a model that is the digital twin of the monitored equipment/process. This model will be able to predict the response towards new operating modes, novel fault cases, external perturbations and even maintenance events (such as repair and setup change/adjustments).

Furthermore, the models often rely on the assumption that data adequacy and quality is sufficient. The data collected from the available sensors might no be suitable in detecting a specific fault. More often, due to limitations with storage and capacity, data is sampled sparsely which effectively limits the bandwidth of information that can be extracted. Even when the right machine parameter is being measured by the appropriate sensor, data quality can be improved by using hardware and/or digital filters that isolate the important aspect of the data while attenuating unnecessary or unwanted data artifacts. The introduction of new wind turbine models and functionalities will bring a new challenge in monitoring processes. The current sensing system might not be sensitive enough to detect underlying characteristics of new events.

Noise is also an important aspect that needs to be taken into account in wind turbine monitoring. There also can be cases where ambient conditions are severe beyond what the equipment has been designed for and may exhibit features that are outside its normal range of operating conditions, which may lead to false alarms. Weather seasonality can also affect a system's long term behaviour. In order to address this, fault

detection methods can employ multiple models to account for different perturbations, but as the number of perturbations to consider grows, the complexity and number of models may not be practical anymore.

Maintenance events (repair/replacement/conditioning) can also affect the fault detection system. It's a common assumption that the right corrective maintenance repair activity was performed, but there always lies the element of human error. Maintenance events can change the mechanical properties (e.g.stiffness) of the system and can cause deviations in the measured features. Data-driven methods typically employ a re-training of the model after every maintenance event.

7.2 Future Work

A number of areas have been mentioned throughout this thesis in which improvements could be made with further work. The following paragraphs will outline what further work could be carried out to improve and build on this thesis.

Data is collected from various sensors across the wind turbine and usually evaluated independently. Future work could include the combination of various data sources. For example vibrations, SCADA, and oil samples can be combined in an integrated decision support system which may increase the maintenance action confidence.

At the same time, the wind turbine drivetrain is composed of many components. Multi-component systems dependencies can be taken into account such as structural dependencies, economic dependencies and stochastic dependencies. Structural dependence concerns on the physical structure of the system, whereby inspecting or replacing one component may require intervention of other components, for example their dismantlement. Economic dependence deals with the cost effectiveness of certain maintenance policies, whereby combining the maintenance of multiple components can affect costs. Stochastic dependence means that the degradation of one component can affect the degradation process of other components, usually accelerating the other [214].

The thesis focused on methods for providing diagnostic solutions for wind turbine drivetrains using data driven methods. Estimating the remaining useful life is the next step towards robust asset management strategies. This can be achieved by using

hybrid methods (both physics-based and data-based) to more accurately model damage accumulation and future conditions of the system. The above can prove particularly useful as new, larger, more expensive wind turbines with new drivetrain technologies are introduced in the growing offshore wind sector.

Finally, a substantial amount of operating wind turbines will reach the end of their planned service life in the near future. Lifetime extension scenarios for wind turbine drivetrains should be explored and modeled.

Bibliography

[1] J. Coultate, "Wind turbine drivetrain technology and cost drivers," *Romax Technology*, 2011.

[2] E. Artigao, S. Martín-Martínez, A. Honrubia-Escribano, and E. Gómez-Lázaro, "Wind turbine reliability: A comprehensive review towards effective condition monitoring development," *Applied energy*, vol. 228, pp. 1569–1583, 2018.

[3] J. Carroll, A. McDonald, and D. McMillan, "Failure rate, repair time and unscheduled o&m cost analysis of offshore wind turbines," *Wind Energy*, 2015.

[4] C. Vázquez-Hernández, J. Serrano-González, and G. Centeno, "A market-based analysis on the main characteristics of gearboxes used in onshore wind turbines," *Energies*, vol. 10, no. 11, p. 1686, 2017.

[5] C. V. Hernández, T. Telsnig, and A. V. Pradas, "Jrc wind energy status report 2016 edition," *Market, Technology and Regulatory Aspects of Wind Energy*, 2017.

[6] S. Sheng, "Gearbox reliability database: Yesterday, today, and tomorrow (presentation)," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2014.

[7] "Attribution: Mcld [cc by-sa 3.0]." `https://creativecommons.org/licenses/by-sa/3.0`.

[8] F. Spinato, P. J. Tavner, G. Van Bussel, and E. Koutoulakos, "Reliability of wind turbine subassemblies," *IET Renewable Power Generation*, vol. 3, no. 4, pp. 387–401, 2009.

[9] "Breaking: Uk offshore wind strike prices slide down to gbp 39.65/mwh." https://www.offshorewind.biz/2019/09/20/uk-offshore-wind-strike-prices-slide-down-to-gbp-39-65-mwh/. Accessed: 2019-10-09.

[10] Z. Tian and H. Liao, "Condition based maintenance optimization for multi-component systems using proportional hazards model," *Reliability Engineering & System Safety*, vol. 96, no. 5, pp. 581–589, 2011.

[11] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renewable and Sustainable energy reviews*, vol. 13, no. 1, pp. 1–39, 2009.

[12] Y. Amirat, M. E. H. Benbouzid, E. Al-Ahmar, B. Bensaker, and S. Turri, "A brief status on condition monitoring and fault diagnosis in wind energy conversion systems," *Renewable and sustainable energy reviews*, vol. 13, no. 9, pp. 2629–2636, 2009.

[13] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez, and M. Papaelias, "Condition monitoring of wind turbines: Techniques and methods," *Renewable Energy*, vol. 46, pp. 169–178, 2012.

[14] W. Yang, R. Court, and J. Jiang, "Wind turbine condition monitoring by the approach of scada data analysis," *Renewable Energy*, vol. 53, pp. 365–376, 2013.

[15] A. Zaher, S. McArthur, D. Infield, and Y. Patel, "Online wind turbine fault detection through automated scada data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, 2009.

[16] C. S. Gray and S. J. Watson, "Physics of failure approach to wind turbine condition based maintenance," *Wind Energy*, vol. 13, no. 5, pp. 395–405, 2010.

[17] C. J. Crabtree, D. Zappalá, and P. J. Tavner, "Survey of commercially available condition monitoring systems for wind turbines.," 2014.

[18] J. Ribrant and L. Bertling, "Survey of failures in wind power systems with focus on swedish wind power plants during 1997-2005," in *Power Engineering Society General Meeting, 2007. IEEE*, pp. 1–8, IEEE, 2007.

[19] B. Hahn, M. Durstewitz, and K. Rohrig, "Reliability of wind turbines," *Wind energy*, pp. 329–332, 2007.

[20] W. Eggersglüß, "Wind energie ix-xiv, praxis-ergebnisse 1995-2004," *Landwirtschaftskammer Schleswig-Holstein, Osterrönfeld, Germany*, vol. 2004, 1995.

[21] M. Wilkinson, B. Hendriks, F. Spinato, K. Harman, E. Gomez, H. Bulacio, J. Roca, P. Tavner, Y. Feng, and H. Long, "Methodology and results of the reliawind reliability field study," in *European Wind Energy Conference and Exhibition 2010, EWEC 2010*, vol. 3, pp. 1984–2004, Sheffield, 2010.

[22] D. McMillan and G. W. Ault, "Quantification of condition monitoring benefit for offshore wind turbines," *Wind Engineering*, vol. 31, no. 4, pp. 267–285, 2007.

[23] C. J. Crabtree, "Operational and reliability analysis of offshore wind farms," in *Proceedings of the scientific track of the European Wind Energy Association conference*, 2012.

[24] J. Puigcorbe and A. De-Beaumont, "Wind turbine gearbox reliability: the impact of rotor support," *Renew. Energy World Mag*, vol. 14, no. 3, pp. 67–73, 2010.

[25] A. D. Hansen, F. Iov, F. Blaabjerg, and L. H. Hansen, "Review of contemporary wind turbine concepts and their market penetration," *Wind Engineering*, vol. 28, no. 3, pp. 247–263, 2004.

[26] J. Serrano-González and R. Lacal-Arántegui, "Technological evolution of onshore wind turbinesa market-based analysis," *Wind Energy*, vol. 19, no. 12, pp. 2171–2187, 2016.

[27] Z. Chen, "Wind turbine drive train systems," in *Wind Energy Systems*, pp. 208–246, Elsevier, 2011.

[28] M. Wilkinson, K. Harman, B. Hendriks, F. Spinato, T. van Delft, G. Garrad, and U. Thomas, "Measuring wind turbine reliability-results of the reliawind project," in *EWEA Conference*, pp. 1–8, 2011.

[29] R. L. Arántegui and J. S. Gonzáles, "Jrc wind status report: Technology, market and economic aspects of wind energy in europe. 2015," 2014.

[30] J. Ribrant, "Reliability performance and maintenancea survey of failures in wind power systems," *KTH School of Electrical Engineering*, pp. 59–72, 2006.

[31] D. J. Wilkins, "The bathtub curve and product failure behavior," *Reliability HotWire*, vol. 21, no. NOV, 2002.

[32] M. R. W. Group *et al.*, "Report of large motor reliability survey of industrial and commercial installations, part i," *IEEE Trans. Ind. Appl*, vol. 21, no. 4, pp. 853–864, 1985.

[33] P. Tavner, G. Van Bussel, and F. Spinato, "Machine and converter reliabilities in wind turbines," 2006.

[34] P. Tavner, J. Xiang, and F. Spinato, "Reliability analysis for wind turbines," *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, vol. 10, no. 1, pp. 1–18, 2007.

[35] S. S. Smater and A. D. Dominguez-Garcia, "A framework for reliability and performance assessment of wind energy conversion systems," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2235–2245, 2011.

[36] L. Alhmoud and B. Wang, "A review of the state-of-the-art in wind-energy reliability analysis," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1643–1651, 2018.

[37] M. Leimeister and A. Kolios, "A review of reliability-based methods for risk analysis and their application in the offshore wind industry," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 1065–1076, 2018.

[38] A. Van Horenbeek, J. Van Ostaeyen, J. R. Duflou, and L. Pintelon, "Quantifying the added value of an imperfectly performing condition monitoring systemapplication to a wind turbine gearbox," *Reliability Engineering & System Safety*, vol. 111, pp. 45–57, 2013.

[39] K. Smolders, H. Long, Y. Feng, and P. Tavner, "Reliability analysis and prediction of wind turbine gearboxes," in *European Wind Energy Conference and Exhibition 2010, EWEC 2010*, vol. 4, pp. 2660–2682, Sheffield, 2010.

[40] K. Alewine, "Wind turbine generator failure modes," *Shermco Industries*, 2011.

[41] E. Echavarria, B. Hahn, G. Van Bussel, and T. Tomiyama, "Reliability of wind turbine technology through time," *Journal of Solar Energy Engineering*, vol. 130, no. 3, p. 031005, 2008.

[42] J. J. Nielsen and J. D. Sørensen, "On risk-based operation and maintenance of offshore wind turbine components," *Reliability Engineering & System Safety*, vol. 96, no. 1, pp. 218–229, 2011.

[43] K. Verbert, B. De Schutter, and R. Babuška, "Timely condition-based maintenance planning for multi-component systems," *Reliability Engineering & System Safety*, vol. 159, pp. 310–321, 2017.

[44] E. Standard, "En 13306: Maintenance terminology," tech. rep., 2017.

[45] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[46] H. Van der Auweraer and B. Peeters, "International research projects on structural health monitoring: an overview," *Structural Health Monitoring*, vol. 2, no. 4, pp. 341–358, 2003.

[47] P. Tavner, *Condition monitoring of rotating electrical machines*, vol. 56. IET, 2008.

[48] T. Xia, Y. Dong, L. Xiao, S. Du, E. Pan, and L. Xi, "Recent advances in prognostics and health management for advanced manufacturing paradigms," *Reliability Engineering & System Safety*, 2018.

[49] S. Marble and B. P. Morton, "Predicting the remaining life of propulsion system bearings," in *Aerospace Conference, 2006 IEEE*, pp. 8–pp, IEEE, 2006.

[50] G. J. Kacprzynski, M. J. Roemer, G. Modgil, A. Palladino, and K. Maynard, "Enhancement of physics-of-failure prognostic models with system level features," in *Aerospace Conference Proceedings, 2002. IEEE*, vol. 6, pp. 6–6, IEEE, 2002.

[51] C. J. Li and H. Lee, "Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics," *Mechanical systems and signal processing*, vol. 19, no. 4, pp. 836–846, 2005.

[52] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distributions from component degradation signals: A bayesian approach," *IiE Transactions*, vol. 37, no. 6, pp. 543–557, 2005.

[53] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," *Journal of Intelligent Manufacturing*, vol. 23, no. 2, pp. 227–237, 2012.

[54] Z. Tian, L. Wong, and N. Safaei, "A neural network approach for remaining useful life prediction utilizing both failure and suspension histories," *Mechanical Systems and Signal Processing*, vol. 24, no. 5, pp. 1542–1555, 2010.

[55] J. J. Nielsen and J. D. Sørensen, "Risk-based operation and maintenance of offshore wind turbines," 2013.

[56] E. Byon, E. Pérez, Y. Ding, and L. Ntaimo, "Simulation of wind farm operations and maintenance using discrete event system specification," *Simulation*, vol. 87, no. 12, pp. 1093–1117, 2011.

[57] J. Nilsson and L. Bertling, "Maintenance management of wind power systems using condition monitoring systemslife cycle cost analysis for two case studies," *IEEE Transactions on energy conversion*, vol. 22, no. 1, pp. 223–229, 2007.

[58] J. Nielsen and J. Sørensen, "Methods for risk-based planning of o&m of wind turbines," *Energies*, vol. 7, no. 10, pp. 6645–6664, 2014.

[59] M. N. Scheu, L. Tremps, U. Smolka, A. Kolios, and F. Brennan, "A systematic failure mode effects and criticality analysis for offshore wind turbine systems towards integrated condition based maintenance strategies," *Ocean Engineering*, vol. 176, pp. 118–133, 2019.

[60] J. P. Salameh, S. Cauet, E. Etien, A. Sakout, and L. Rambault, "Gearbox condition monitoring in wind turbines: A review," *Mechanical Systems and Signal Processing*, vol. 111, pp. 251–264, 2018.

[61] S. Sheng and P. S. Veers, *Wind Turbine Drivetrain Condition Monitoring, an Overview*. National Renewable Energy Laboratory Golden, Colo, USA, 2011.

[62] J. Liu, D. Djurdjanovic, K. A. Marko, and J. Ni, "A divide and conquer approach to anomaly detection, localization and diagnosis," *Mechanical Systems and Signal Processing*, vol. 23, no. 8, pp. 2488–2499, 2009.

[63] J. Korbicz, J. M. Koscielny, Z. Kowalczuk, and W. Cholewa, *Fault diagnosis: models, artificial intelligence, applications*. Springer Science & Business Media, 2012.

[64] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: A review on data driven approaches," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[65] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[66] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, and G. Nenadic, "Machine learning methods for wind turbine condition monitoring: A review," *Renewable energy*, 2018.

[67] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical systems and signal processing*, vol. 21, no. 6, pp. 2560–2574, 2007.

[68] R. Hao, Z. Peng, Z. Feng, and F. Chu, "Application of support vector machine based on pattern spectrum entropy in fault diagnostics of rolling element bearings," *Measurement Science and Technology*, vol. 22, no. 4, p. 045708, 2011.

[69] Y. Lei and M. J. Zuo, "Gear crack level identification based on weighted k nearest neighbor classification algorithm," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1535–1547, 2009.

[70] R. Casimir, E. Boutleux, G. Clerc, and A. Yahoui, "The use of features selection and nearest neighbors rule for faults diagnostic in induction motors," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 2, pp. 169–177, 2006.

[71] D. Wang, "K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited," *Mechanical Systems and Signal Processing*, vol. 70, pp. 201–208, 2016.

[72] B.-S. Yang, M.-S. Oh, A. C. C. Tan, *et al.*, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1840–1849, 2009.

[73] S. Agarwal, A. Swetapadma, and C. Panigrahi, "An improved method using artificial neural network for fault detection and fault pole identification in voltage source converter-based high-voltage direct current transmission lines," *Arabian Journal for Science and Engineering*, vol. 43, no. 8, pp. 4005–4018, 2018.

[74] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.

[75] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering & System Safety*, vol. 115, pp. 124–135, 2013.

[76] M. B. Christopher, *PATTERN RECOGNITION AND MACHINE LEARNING.* Springer-Verlag New York, 2016.

[77] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning.* MIT Press, 2016.

[78] L. Breiman, "Random forests machine learning. 45: 5–32," *View Article PubMed/NCBI Google Scholar*, 2001.

[79] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer series in statistics New York, NY, USA:, 2001.

[80] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *et al.*, "A practical guide to support vector classification," 2003.

[81] K.-B. Duan and S. S. Keerthi, "Which is the best multiclass svm method? an empirical study," in *International workshop on multiple classifier systems*, pp. 278–285, Springer, 2005.

[82] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[83] V. Vapnik, *The nature of statistical learning theory.* Springer science & business media, 2013.

[84] T. Elhassan and M. Aljurf, "Classification of imbalance data using tomek link (t-link) combined with random under-sampling (rus) as a data reduction method.",," 2016.

[85] S. Kotsiantis, D. Kanellopoulos, P. Pintelas, *et al.*, "Handling imbalanced datasets: A review," *GESTS International Transactions on Computer Science and Engineering*, vol. 30, no. 1, pp. 25–36, 2006.

[86] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[87] I. Tomek, "An experiment with the edited nearest-neighbor rule," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 6, no. 6, pp. 448–452, 1976.

[88] M. Markou and S. Singh, "Novelty detection: a reviewpart 1: statistical approaches," *Signal processing*, vol. 83, no. 12, pp. 2481–2497, 2003.

[89] M. Markou and S. Singh, "Novelty detection: a reviewpart 2:: neural network based approaches," *Signal processing*, vol. 83, no. 12, pp. 2499–2521, 2003.

[90] D. M. Tax and R. P. Duin, "Support vector data description," *Machine learning*, vol. 54, no. 1, pp. 45–66, 2004.

[91] D. M. Tax and R. P. Duin, "Support vector domain description," *Pattern recognition letters*, vol. 20, no. 11-13, pp. 1191–1199, 1999.

[92] Y. Pan, J. Chen, and L. Guo, "Robust bearing performance degradation assessment method based on improved wavelet packet–support vector data description," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 669–681, 2009.

[93] H. Wang and J. Chen, "Performance degradation assessment of rolling bearing based on bispectrum and support vector data description," *Journal of Vibration and Control*, vol. 20, no. 13, pp. 2032–2041, 2014.

[94] M.-M. Deza and E. Deza, *Dictionary of distances*. Elsevier, 2006.

[95] E. Panayirci and R. C. Dubes, "A test for multidimensional clustering tendency," *Pattern Recognition*, vol. 16, no. 4, pp. 433–444, 1983.

[96] P. Holgate, "Some new tests of randomness," *The Journal of Ecology*, pp. 261–266, 1965.

[97] P. J. Diggle, J. Besag, and J. T. Gleaves, "Statistical analysis of spatial point patterns by means of distance methods," *Biometrics*, pp. 659–667, 1976.

[98] A. Banerjee and R. N. Dave, "Validating clusters using the hopkins statistic," in *2004 IEEE International conference on fuzzy systems (IEEE Cat. No. 04CH37542)*, vol. 1, pp. 149–153, IEEE, 2004.

[99] J. F. Pierna and D. Massart, "Improved algorithm for clustering tendency," *Analytica Chimica Acta*, vol. 408, no. 1-2, pp. 13–20, 2000.

[100] B. Mirkin, "Choosing the number of clusters," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 252–260, 2011.

[101] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping multidimensional data*, pp. 25–71, Springer, 2006.

[102] Y. Zhao, G. Karypis, and U. Fayyad, "Hierarchical clustering algorithms for document datasets," *Data mining and knowledge discovery*, vol. 10, no. 2, pp. 141–168, 2005.

[103] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA, 1967.

[104] J. Hua, Z. Xiong, J. Lowey, E. Suh, and E. R. Dougherty, "Optimal number of features as a function of sample size for various classification rules," *Bioinformatics*, vol. 21, no. 8, pp. 1509–1515, 2004.

[105] L. Rokach and O. Maimon, "The data mining and knowledge discovery handbook: A complete guide for researchers and practitioners," 2005.

[106] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, pp. 1137–1145, Montreal, Canada, 1995.

[107] K. Wang, B. Wang, and L. Peng, "Cvap: validation for cluster analyses," *Data Science Journal*, pp. 0904220071–0904220071, 2009.

[108] J. Tautz-Weinert and S. J. Watson, "Using scada data for wind turbine condition monitoring–a review," *IET Renewable Power Generation*, vol. 11, no. 4, pp. 382–394, 2016.

[109] K. Kim, G. Parthasarathy, O. Uluyol, W. Foslien, S. Sheng, and P. Fleming, "Use of scada data for failure detection in wind turbines," in *ASME 2011 5th International conference on energy sustainability*, pp. 2071–2079, American Society of Mechanical Engineers, 2011.

[110] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, "Monitoring wind turbine gearboxes," *Wind Energy*, vol. 16, no. 5, pp. 728–740, 2013.

[111] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, "Use of scada and cms signals for failure detection and diagnosis of a wind turbine gearbox," in *European Wind Energy Conference and Exhibition 2011, EWEC 2011*, pp. 17–19, Sheffield, 2011.

[112] S. Catmull, "Self-organising map based condition monitoring of wind turbines," in *EWEA Annual Conf*, vol. 2011, 2011.

[113] L. Tarassenko, A. Nairac, N. Townsend, I. Buxton, and P. Cowley, "Novelty detection for the identification of abnormalities," *International Journal of Systems Science*, vol. 31, no. 11, pp. 1427–1439, 2000.

[114] S. McArthur, V. Catterson, and J. McDonald, "A multi-agent condition monitoring architecture to support transmission and distribution asset management," 2005.

[115] M. C. Garcia, M. A. Sanz-Bobi, and J. del Pico, "Simap: Intelligent system for predictive maintenance: Application to the health condition monitoring of a windturbine gearbox," *Computers in Industry*, vol. 57, no. 6, pp. 552–568, 2006.

[116] M. Schlechtingen and I. F. Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mechanical systems and signal processing*, vol. 25, no. 5, pp. 1849–1875, 2011.

[117] Z.-Y. Zhang and K.-S. Wang, "Wind turbine fault detection based on scada data analysis using ann," *Advances in Manufacturing*, vol. 2, no. 1, pp. 70–78, 2014.

[118] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: A data-mining approach," *Renewable Energy*, vol. 48, pp. 110–116, 2012.

[119] P. Bangalore and L. B. Tjernberg, "An approach for self evolving neural network based algorithm for fault prognosis in wind turbine," in *PowerTech (POWERTECH), 2013 IEEE Grenoble*, pp. 1–6, IEEE, 2013.

[120] P. Bangalore and L. B. Tjernberg, "Self evolving neural network based algorithm for fault prognosis in wind turbines: A case study," in *Probabilistic Methods Applied to Power Systems (PMAPS), 2014 International Conference on*, pp. 1–6, IEEE, 2014.

[121] K. Leahy, R. L. Hu, I. C. Konstantakopoulos, C. J. Spanos, and A. M. Agogino, "Diagnosing wind turbine faults using machine learning techniques applied to operational data," in *Prognostics and Health Management (ICPHM), 2016 IEEE International Conference on*, pp. 1–8, IEEE, 2016.

[122] Y. Qiu, Y. Feng, P. Tavner, P. Richardson, G. Erdos, and B. Chen, "Wind turbine scada alarm analysis for improving reliability," *Wind Energy*, vol. 15, no. 8, pp. 951–966, 2012.

[123] Y. Wang and D. Infield, "Supervisory control and data acquisition data-based non-linear state estimation technique for wind turbine gearbox condition monitoring," *IET Renewable Power Generation*, vol. 7, no. 4, pp. 350–358, 2013.

[124] E. Gonzalez, B. Stephen, D. Infield, and J. J. Melero, "Using high-frequency scada data for wind turbine performance monitoring: A sensitivity study," *Renewable energy*, vol. 131, pp. 841–853, 2019.

[125] X.-H. Liang, Z.-L. Liu, J. Pan, and M. J. Zuo, "Spur gear tooth pitting propagation assessment using model-based analysis," *Chinese Journal of Mechanical Engineering*, vol. 30, no. 6, pp. 1369–1382, 2017.

[126] P. McFadden, "Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration," *Journal of vibration, acoustics, stress, and reliability in design*, vol. 108, no. 2, pp. 165–170, 1986.

[127] Z. Feng and M. J. Zuo, "Vibration signal models for fault diagnosis of planetary gearboxes," *Journal of Sound and Vibration*, vol. 331, no. 22, pp. 4919–4939, 2012.

[128] L. Hong, J. S. Dhupia, and S. Sheng, "An explanation of frequency features enabling detection of faults in equally spaced planetary gearbox," *Mechanism and Machine Theory*, vol. 73, pp. 169–183, 2014.

[129] Y.-F. Wang and P. J. Kootsookos, "Modeling of low shaft speed bearing faults for condition monitoring," *Mechanical Systems and Signal Processing*, vol. 12, no. 3, pp. 415–426, 1998.

[130] S. Jain, P. Whiteley, and H. Hunt, "Detection of planet bearing faults in wind turbine gearboxes," in *Proceedings of International Conference on Noise and Vibration Engineering (ISMA12)/International Conference on Uncertainty in Structural Dynamics (USD12)*, pp. 4351–4362, 2012.

[131] S. A. Abouel-seoud, "Fault detection enhancement in wind turbine planetary gearbox via stationary vibration waveform data," *Journal of Low Frequency Noise, Vibration and Active Control*, vol. 37, no. 3, pp. 477–494, 2018.

[132] D. Hochmann and M. Sadok, "Theory of synchronous averaging/sup/spl omega," in *Aerospace Conference, 2004. Proceedings. 2004 IEEE*, vol. 6, pp. 3636–3653, IEEE, 2004.

[133] H. Balderston, "Incipient failure detection: Incipient failure detection in ball bearings," tech. rep., BOEING CO SEATTLE WA AEROSPACE SYSTEMS DIV, 1969.

[134] M. S. Darlow, R. H. Badgley, and G. Hogg, "Application of high-frequency resonance techniques for bearing diagnostics in helicopter gearboxes," tech. rep., MECHANICAL TECHNOLOGY INC LATHAM NY, 1974.

[135] D. Hochmann and E. Bechhoefer, "Envelope bearing analysis: theory and practice," in *Aerospace Conference, 2005 IEEE*, pp. 3658–3666, IEEE, 2005.

[136] R. B. Randall, "A history of cepstrum analysis and its application to mechanical problems," *Mechanical Systems and Signal Processing*, 2016.

[137] W. Wang and P. McFadden, "Early detection of gear failure by vibration analysis i. calculation of the time-frequency distribution," *Mechanical Systems and Signal Processing*, vol. 7, no. 3, pp. 193–203, 1993.

[138] R. K. Young, *Wavelet theory and its applications*, vol. 189. Springer Science & Business Media, 2012.

[139] W. Staszewski and G. Tomlinson, "Application of the wavelet transform to fault detection in a spur gear," *Mechanical Systems and Signal Processing*, vol. 8, no. 3, pp. 289–307, 1994.

[140] S. Sheng, "Wind turbine gearbox condition monitoring round robin study–vibration analysis," *Contract*, vol. 303, pp. 275–3000, 2012.

[141] J. Zhu, T. Nostrand, C. Spiegel, and B. Morton, "Survey of condition indicators for condition monitoring systems," in *Annu. Conf. Progn. Heal. Manag. Soc*, vol. 5, pp. 1–13, 2014.

[142] E. Bechhoefer and M. Kingsley, "A review of time synchronous average algorithms," in *Annual Conference of the Prognostics and Health Management Society, San Diego, CA, Sept*, pp. 24–33, 2009.

[143] T. Barszcz and R. B. Randall, "Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352–1365, 2009.

[144] J. Hanna, C. Hatch, M. Kalb, A. Weiss, and H. Luo, "Detection of wind turbine gear tooth defects using sideband energy ratio," *China Wind Power 2011; October, 19-21, 2011, Beijing, China*, 2011.

[145] D. Zappalá, P. J. Tavner, C. J. Crabtree, and S. Sheng, "Side-band algorithm for automatic wind turbine gearbox fault detection and diagnosis," *IET Renewable Power Generation*, vol. 8, no. 4, pp. 380–389, 2014.

[146] R. Randall, "Cepstrum analysis and gearbox fault-diagnosis," *Maintenance Management International*, vol. 3, no. 3, pp. 183–208, 1982.

[147] J. Antoni and R. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 308–331, 2006.

[148] R. Randall, J. Antoni, and N. Sawalhi, "Application of spectral kurtosis to bearing fault detection in rolling element bearings," in *Eleventh International Congress on Sound and Vibration. St. Petersburg*, 2004.

[149] N. Sawalhi and R. B. Randall, "The application of spectral kurtosis to bearing diagnostics," in *Proceedings of ACOUSTICS*, pp. 3–5, 2004.

[150] R. B. Randall and J. Antoni, "Rolling element bearing diagnosticsa tutorial," *Mechanical systems and signal processing*, vol. 25, no. 2, pp. 485–520, 2011.

[151] R. Yan and R. X. Gao, "Multi-scale enveloping spectrogram for vibration analysis in bearing defect diagnosis," *Tribology International*, vol. 42, no. 2, pp. 293–302, 2009.

[152] W. Gardner, "Measurement of spectral correlation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1111–1123, 1986.

[153] J. Antoni, G. Xin, and N. Hamzaoui, "Fast computation of the spectral correlation," *Mechanical Systems and Signal Processing*, vol. 92, pp. 248 – 277, 2017.

[154] C. Capdessus, M. Sidahmed, and J. Lacoume, "Cyclostationary processes: application in gear faults early diagnosis," *Mechanical systems and signal processing*, vol. 14, no. 3, pp. 371–385, 2000.

[155] N. Sawalhi and R. Randall, "Spectral kurtosis enhancement using autoregressive models," in *ACAM Conference, Melbourne, Australia*, 2005.

[156] N. Sawalhi, R. Randall, and H. Endo, "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2616–2633, 2007.

[157] C. Ruiz-Cárcel, E. Hernani-Ros, P. Chandra, Y. Cao, and D. Mba, "Application of linear prediction, self-adaptive noise cancellation, and spectral kurtosis in identifying natural damage of rolling element bearing in a gearbox," in *Proceedings of the 7th World Congress on Engineering Asset Management (WCEAM 2012)*, pp. 505–513, Springer, 2015.

[158] Y. Li, M. Xu, Y. Wei, and W. Huang, "A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree," *Measurement*, vol. 77, pp. 80–94, 2016.

[159] W. Teng, F. Wang, K. Zhang, Y. Liu, and X. Ding, "Pitting fault detection of a wind turbine gearbox using empirical mode decomposition," *Strojniški vestnik-Journal of Mechanical Engineering*, vol. 60, no. 1, pp. 12–20, 2014.

[160] Z. Feng, M. Liang, Y. Zhang, and S. Hou, "Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode

decomposition and energy separation," *Renewable Energy*, vol. 47, pp. 112–126, 2012.

[161] S. Kang, D. Ma, Y. Wang, C. Lan, Q. Chen, and V. Mikulovich, "Method of assessing the state of a rolling bearing based on the relative compensation distance of multiple-domain features and locally linear embedding," *Mechanical Systems and Signal Processing*, vol. 86, pp. 40–57, 2017.

[162] W. Wang and P. McFadden, "Application of wavelets to gearbox vibration signals for fault detection," *Journal of sound and vibration*, vol. 192, no. 5, pp. 927–939, 1996.

[163] W. Yang, P. Tavner, and M. Wilkinson, "Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train," *IET Renewable Power Generation*, vol. 3, no. 1, pp. 1–11, 2009.

[164] B. Tang, W. Liu, and T. Song, "Wind turbine fault diagnosis based on morlet wavelet transformation and wigner-ville distribution," *Renewable Energy*, vol. 35, no. 12, pp. 2862–2866, 2010.

[165] Z. Feng and M. Liang, "Fault diagnosis of wind turbine planetary gearbox under nonstationary conditions via adaptive optimal kernel time–frequency analysis," *Renewable Energy*, vol. 66, pp. 468–477, 2014.

[166] W. Teng, X. Ding, X. Zhang, Y. Liu, and Z. Ma, "Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform," *Renewable Energy*, vol. 93, pp. 591–598, 2016.

[167] C. Dao, B. Kazemtabrizi, and C. Crabtree, "Wind turbine reliability data review and impacts on levelised cost of energy," *Wind Energy*, 2019.

[168] L. Rademakers, B. Blok, B. Van den Horn, J. Jehee, A. Seebregts, and R. Van Otterlo, "Reliability analysis methods for wind turbines, task 1 of the project: Probabilistic safety assessment for wind turbines," *Netherlands energy research foundation, ECN Memorandum*, 1992.

[169] A. Graves, K. Harman, M. Wilkinson, and R. Walker, "Understanding availability trends of operating wind farms," in *AWEA Wind Power Conference, AWEA*, 2008.

[170] E. ISO, "14224: 2006," *Petroleum, petrochemical and natural gas industries-Collection and exchange of reliability and maintenance data for equipment.*

[171] J. Carroll, A. McDonald, and D. McMillan, "Reliability comparison of wind turbines with dfig and pmg drive trains," *IEEE Transactions on Energy Conversion*, vol. 30, no. 2, pp. 663–670, 2014.

[172] E. Byon, L. Ntaimo, C. Singh, and Y. Ding, "Wind energy facility reliability and maintenance," in *Handbook of Wind Power Systems*, pp. 639–672, Springer, 2013.

[173] F. Spinato, *The reliability of wind turbines.* PhD thesis, Durham University, 2008.

[174] L. H. Crow, "Reliability analysis for complex, repairable systems," tech. rep., ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITYABERDEEN PROVING GROUND MD, 1975.

[175] G. Pulcini, "A bounded intensity process for the reliability of repairable equipment," *Journal of Quality Technology*, vol. 33, no. 4, pp. 480–492, 2001.

[176] L. Bird, J. Cochran, and X. Wang, "Wind and solar energy curtailment: Experience and practices in the united states," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2014.

[177] A. Kusiak and A. Verma, "Monitoring wind farms with performance curves," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 192–199, 2013.

[178] P. Bangalore, S. Letzgus, D. Karlsson, and M. Patriksson, "An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox," *Wind Energy*, vol. 20, no. 8, pp. 1421–1438, 2017.

[179] L. Rokach and O. Maimon, "Clustering methods," in *Data mining and knowledge discovery handbook*, pp. 321–352, Springer, 2005.

[180] A. N. Tikhonov and V. I. Arsenin, *Solutions of ill-posed problems*, vol. 14. Winston, Washington, DC, 1977.

[181] G. C. Cawley and N. L. Talbot, "Preventing over-fitting during model selection via bayesian regularisation of the hyper-parameters," *Journal of Machine Learning Research*, vol. 8, no. Apr, pp. 841–861, 2007.

[182] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural networks: Tricks of the trade*, pp. 437–478, Springer, 2012.

[183] D. J. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.

[184] F. D. Foresee and M. T. Hagan, "Gauss-newton approximation to bayesian learning," in *Proceedings of the 1997 international joint conference on neural networks*, vol. 3, pp. 1930–1935, Piscataway: IEEE, 1997.

[185] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[186] M. Schlechtingen, I. F. Santos, and S. Achiche, "Wind turbine condition monitoring based on scada data using normal behavior models. part 1: System description," *Applied Soft Computing*, vol. 13, no. 1, pp. 259–270, 2013.

[187] J. Tian, M. H. Azarian, and M. Pecht, "Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm," in *Proceedings of the European Conference of the Prognostics and Health Management Society*, Citeseer, 2014.

[188] C. Z. Mooney, R. D. Duval, and R. Duvall, *Bootstrapping: A nonparametric approach to statistical inference*. No. 94-95, Sage, 1993.

[189] K. Fyfe and E. Munck, "Analysis of computed order tracking," *Mechanical Systems and Signal Processing*, vol. 11, no. 2, pp. 187–205, 1997.

[190] S. Koukoura, J. Carroll, and A. McDonald, "Wind turbine intelligent gear fault identification," in *Annual Conference of the PHM Society*, 2017.

[191] V. Sharma and A. Parey, "A review of gear fault diagnosis using various condition indicators," *Procedia Engineering*, vol. 144, pp. 253–263, 2016.

[192] H. Vold, M. Mains, and J. Blough, "Theoretical foundations for high performance order tracking with the vold-kalman tracking filter," tech. rep., SAE Technical Paper, 1997.

[193] B. P. Bogert, "The quefrency alanysis of time series for echoes: Cepstrum pseudo-autocovariance, cross-cepstrum, and saphe cracking," *Time series analysis*, pp. 209–243, 1963.

[194] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, vol. 454, pp. 903–995, The Royal Society, 1998.

[195] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[196] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 108–124, 2007.

[197] N. E. Huang and Z. Wu, "A review on hilbert-huang transform: Method and its applications to geophysical studies," *Reviews of geophysics*, vol. 46, no. 2, 2008.

[198] S. M. Kay and S. L. Marple, "Spectrum analysisa modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[199] B. Widrow and S. D. Stearns, *Adaptive signal processing*, vol. 15. Prentice-hall Englewood Cliffs, NJ, 1985.

[200] J. Antoni and R. Randall, "Unsupervised noise cancellation for vibration signals: part ievaluation of adaptive algorithms," *Mechanical Systems and Signal Processing*, vol. 18, no. 1, pp. 89–101, 2004.

[201] E. Hiroaki and S. Nader, "Gearbox simulation models with gear and bearing faults," in *Mechanical engineering*, InTech, 2012.

[202] E. Bechhoefer, P. Menon, and D. He, "A control theory approach to machinery health prognostics," in *66th Forum of the American Helicopter Society:" Rising to New Heights in Vertical Lift Technology", AHS Forum 66*, 2010.

[203] K. Ogata and Y. Yang, *Modern control engineering*, vol. 4. Prentice hall India, 2002.

[204] T. Wang, Q. Han, F. Chu, and Z. Feng, "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review," *Mechanical Systems and Signal Processing*, vol. 126, pp. 662–685, 2019.

[205] M. Nadakatti, A. Ramachandra, and A. Santosh Kumar, "Artificial intelligence-based condition monitoring for plant maintenance," *Assembly Automation*, vol. 28, no. 2, pp. 143–150, 2008.

[206] M. Grigoriu and C. Phillip, "Maintenance engineering using expert systems," *Expert Systems, Professional Engineering, May*, pp. 72–6, 1988.

[207] Y. Zhao, D. Li, A. Dong, J. Lin, D. Kang, and L. Shang, "Fault prognosis of wind turbine generator using scada data," in *North American Power Symposium (NAPS), 2016*, pp. 1–6, IEEE, 2016.

[208] E. R. Lapira, *Fault detection in a network of similar machines using clustering approach*. PhD thesis, University of Cincinnati, 2012.

[209] M. A. Gelbart, J. Snoek, and R. P. Adams, "Bayesian optimization with unknown constraints," *arXiv preprint arXiv:1403.5607*, 2014.

[210] A. D. Bull, "Convergence rates of efficient global optimization algorithms," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2879–2904, 2011.

[211] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, pp. 2951–2959, 2012.

[212] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[213] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Rusboost: A hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 1, pp. 185–197, 2009.

[214] R. Assaf *et al.*, *Prognostics and health management for multi-component systems.* PhD thesis, University of Salford, 2018.