Applications of Machine Learning in Diagnostics and Prognostics of Wind Turbine High Speed Generator Failure

Alan Turnbull

Wind & Marine Energy Systems CDT

Electronic & Electrical Engineering

University of Strathclyde, Glasgow

September 8, 2021

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

The cost of wind energy has decreased over the last decade as technology has matured and the industry has benefited greatly from economies of scale. That being said, operations and maintenance still make up a significant proportion of the overall costs and needs to be reduced over the coming years as sites, particularly offshore, get larger and more remote. One of the key tools to achieve this is through enhancements of both SCADA and condition monitoring system analytics, leading to more informed and optimised operational decisions.

Specifically examining the wind turbine generator and highspeed assembly, this thesis aims to showcase how machine learning techniques can be utilised to enhance vibration spectral analysis and SCADA analysis for early and more automated fault detection. First this will be performed separately based on features extracted from the vibration spectra and performance data in isolation before a framework will be presented to combine data sources to create a single anomaly detection model for early fault diagnosis. Additionally by further utilising vibration based analysis, machine learning techniques and a synchronised database of failures, remaining useful life prediction will also be explored for generator bearing faults, a key component when it comes to increasing wind turbine generator reliability. It will be shown that through early diagnosis and accurate prognosis, component replacements can be planned and optimised before catastrophic failures and large downtimes occur. Moreover, results also indicate that this can have a significant impact on the costs of operation and maintenance over the lifetime of an offshore development.

Contents

Abstract **ii**

List of Figures **viii**

List of Tables **xii**

Acknowledgements **xv**

1 Introduction **2**

1.1 Thesis Background 2

1.2 Research Question 4

1.3 Novelty of Research 6

1.4 Approach to Research 6

1.5 Important Definitions 7

1.6 Research Output 8

2 Literature Review **10**

2.1 Wind Turbine Generator Technology 10

2.2 Wind Turbine Reliability 13

2.3 Wind Turbine Condition Monitoring 17

2.3.1 Classes of Monitoring Systems 17

2.3.2 Condition Monitoring and Diagnostic Tools 18

2.4 Applying Machine Learning Techniques to Wind Turbine Condition Monitoring 19

2.4.1 Supervised Learning 21

2.4.2 Unsupervised Learning 28
2.4.3 Techniques Used Throughout this Thesis 31
2.4.4 Existing Research in Applying Data Driven Methods to SCADA and Vibration Analysis 32
2.5 Cost of Operations & Maintenance 35

3 Wind Turbine Fault Diagnostics using machine learning to enhance vibration and SCADA based analysis 36
3.1 Chapter Contribution 36
3.2 Bearing Diagnosis through Vibration Analysis 38
3.2.1 Bearing Fault Types 38
3.2.2 Time Domain Analysis 40
3.2.3 Frequency Domain Analysis - Fourier Transform 40
3.2.4 Order Tracking 40
3.2.5 Envelope Analysis 41
3.3 Generator Diagnostics through Vibration Analysis 42
3.3.1 Misalignment 42
3.3.2 Rotor Imbalance 43
3.3.3 Analysis Techniques 44
3.4 Feature Extraction 44
3.5 Feature Classification 46
3.6 Case Study 46
3.6.1 Dataset Description 47
3.6.2 Overall Methodology Framework 48
3.6.3 Model Setup 49
3.6.4 Validation 49
3.7 Case Study Results 51
3.7.1 Initial Results 51
3.7.2 Analysis of Classifier Results 53
3.7.3 Results including Turbine Baseline 55
3.7.4 Fault Diagnostics Indicator Dependencies 56

3.8 Diagnosing Problems with SCADA Data 59
3.8.1 Data Pre-processing 59
3.8.2 A Note on Trending 60
3.9 Normal Behaviour Model 61
3.9.1 Model Types 62
3.9.2 Modelling Process 62
3.9.3 Feature Selection 63
3.9.4 Training and Validation 64
3.10 Case Study - Generator Bearing Fault Detection 65
3.10.1 Data Description, Visualisation and Feature Analysis 65
3.10.2 Model Setup Detail 68
3.10.3 Model Testing & Validation 69
3.10.4 Results 72
3.11 Conclusions & Discussion 72
3.12 Future Work 75
3.12.1 Vibration Diagnostics 75
3.12.2 SCADA Temperature Diagnostics 75

4 Combining SCADA and Vibration Data into a Single Anomaly Detection Model 77

4.1 Chapter Contribution 77
4.2 Understanding the Problem 78
4.2.1 Up-sampling vs Down-sampling 80
4.2.2 Expected Variation 81
4.3 Modelling Framework 82
4.3.1 Normal Behaviour Model 83
4.3.2 Combined Anomaly Detection Classifier 83
4.4 Assessing Error through Classification 84
4.4.1 Case Study Description 84
4.4.2 Methodology 85
4.4.3 Summary of Data 86

4.4.4 NBM Specification 87
4.4.5 NBM Training and Validation 88
4.4.6 Single Class SVM Model 89
4.4.7 Anomaly Detection Results 91
4.4.8 Comparison of Results with Other Methods 92
4.5 Combining SCADA and Vibration for Fault Detection 95
4.5.1 Case Study Description 95
4.5.2 Normal Behaviour Model 96
4.5.3 Single Class SVM Model 99
4.5.4 Anomaly Detection Results 101
4.5.5 Comparison with Single Models 101
4.6 Conclusions 103
4.7 Future Work 104

5 Failure Prognosis using Classification of Vibration Features under Varying Operating Conditions. 105

5.1 Chapter Contribution 105
5.2 Background Information 106
5.2.1 Prognosis vs Diagnosis 106
5.2.2 Vibration Based Indicators 107
5.3 Introduction to Case Study 108
5.3.1 Failure Case 108
5.3.2 Dataset 109
5.3.3 Variation in Operating Conditions 110
5.4 Modelling Framework 111
5.5 Feature Extraction 113
5.5.1 Time Domain Analysis 113
5.5.2 Fourier & Order Analysis 114
5.5.3 Operational Characteristics 114
5.6 Classification Models and Prediction Algorithms 115
5.6.1 Baseline Classification Model 115

5.6.2 Two-stage Classification Model 116
5.6.3 Prediction Algorithms 116
5.6.4 Validation of Results 117
5.7 Results & Discussion 118
5.7.1 Baseline Classification Model 118
5.7.2 Two-stage Classification Model 118
5.8 Conclusion 124
5.9 Future Work 126

6 Analysis of the Cost Impact of Maintenance Strategies Enabled by Modern CMS for Offshore Wind Farms 127

6.1 Chapter Contribution 127
6.2 Maintenance Strategies 128
6.2.1 Types of Maintenance Strategies 128
6.2.2 O&M Decision Making 130
6.3 Methodology and Model Overview 131
6.3.1 Methodology 132
6.3.2 O&M Cost Model Description 133
6.3.3 Theoretical Site Characteristics 134
6.3.4 Analysis Cases 135
6.3.5 Baseline Failure Rates 138
6.4 Results 139
6.4.1 Baseline Cost 139
6.4.2 Effects of Predictive and Condition-based Maintenance Strategies 139
6.4.3 Analysis of Wind Farm Size and Location 140
6.5 Discussion 143
6.5.1 The Cost Impact of Advanced Maintenance Strategies 143
6.5.2 Result Limitations 144
6.6 Conclusion 145
6.7 Future Work 145

7 Overall Conclusions & Future Work **147**

7.1 Conclusions 147

7.1.1 Chapter 2 Summary 148

7.1.2 Chapter 3 Summary 148

7.1.3 Chapter 4 Summary 149

7.1.4 Chapter 5 Summary 149

7.1.5 Chapter 6 Summary 150

7.2 Discussion 150

7.3 Future Work 152

Bibliography **153**

List of Figures

1.1 Approach to overall research chapters. 7

2.1 Drive train arrangements usually employed in commercial wind turbines [1] 11
2.2 Generator topology classification with respect to configurations types. Adapted from [2] 12
2.3 Evolution of drive train configuration in onshore wind turbines by geographical zone (Source: JRC Wind Energy Database. "NA" represents Not Available records in the Joint Research Centre (JRC) Wind Energy) [3] 13
2.4 Wind turbine sub-assembly failure rates and downtimes [4] 15
2.5 Typical positions of accelerometers used for a condition monitoring based on vibration analysis [8]. 18
2.6 Typical condition monitoring work flow. 20
2.7 Difference between underfitting (bias) and overfitting (variance). 22
2.8 Example of kNN and SVM algorithm. 23
2.9 Example of Decision Tree and Random Forest algorithms. 26
2.10 ANN algorithm example diagram. 27
2.11 Clustering basic principle. 29

3.1 Location of bearings on the highspeed assembly where: 1) Upwind HS bearing, 2) Downwind rotor side HS bearing, 3) Downwind gen side HS bearing, 4) Drive end generator bearing and 5) Non-drive end generator bearing. 39
3.2 Speed signal and tachometer pulses over example vibration sample . . . 41

List of Figures

3.3 Misalignment of coupling between highspeed shaft and generator. 43
3.4 Static imbalance in generator. 44
3.5 Classification: two-class vs three-class. 47
3.6 Methodology for Component Health Classification. 49
3.7 Confusion Matrix - Results Set 1. 52
3.8 Diagnostic Health Indicator Variation. 53
3.9 Net Diagnostic Health Indicator Variation. 55
3.10 Confusion Matrix - Results Set 2. 57
3.11 Prediction results for each diagnostic for test cases 58
3.12 Data cleaning 61
3.13 NBM anomaly detection process. 63
3.14 Model based monitoring based on NBM. Adapted from [5], [6] 64
3.15 Breakdown of dataset for model testing and validation. 65
3.16 Visualisation of SCADA relationships to target variable. 68
3.17 Correlation heatmap (left) and Feature Importance (right) of selected SCADA variables. 69
3.18 Cross validation error. 70
3.19 Model validation. 71
3.20 Monthly alarms. 73

4.1 Example data acquisition rate for different data sources. 79
4.2 Example data showing variation within a series of 10 minute periods. . . 81
4.3 Framework for combining data streams. 82
4.4 NBM diagram 84
4.5 Methodology for Case Study 1 85
4.6 Schematic of two-layer feed forward neural network. 88
4.7 Distribution of error from 12 month training period 89
4.8 Trained SVM model with only 2 features; 1) RMSE and 2) Max error . 91
4.9 Heatmap of anomalies leading up to failure 92
4.10 Comparison of SVM classification approach to RMSE 94
4.11 Damage to Generator bearing and shaft at time of replacement. 95

4.12 Methodology showing 2 data sources combined into single anomaly detection model; SCADA and Vibration. 96
4.13 Schematic of Random Forest Regressive model - input features and output dependant on model. 98
4.14 Trained SVM model with only 2 features; 1) SCADA RMSE and 2) Vibration Absolute error 101
4.15 Heatmap of anomalies leading up to failure 101
4.16 Plot of weekly errors leading up to failure for individual NBM's 102

5.1 Key differences between Diagnosis and Prognosis used in this Thesis. . . 107
5.2 Diagram showing estimated positions of accelerometers used to measure generator bearing vibration. 110
5.3 Vibration sample acquisition conditions 110
5.4 Overall framework for predicting failure 112
5.5 Confusion matrix of trained algorithms for all data (100 class 0 and 100 class 1 samples). 119
5.6 Plot of clusters based on three chosen variables, normalised to ensure confidentiality. 'X' represents each cluster centroid position. 121
5.7 Plot of accuracy for each cluster ordered by extremity of operating conditions at time of vibration sample. 122
5.8 Confusion matrix of trained algorithms for group 3 (left) and group 4 (right). 123

6.1 Cost of repair in relation to maintenance strategy 130
6.2 Flow diagram of engineering decision process. 132
6.3 Overall approach to chapter analysis. 133
6.4 Simplified cost model structural overview, adapted from [7] 135
6.5 Adjusted failure rates. 138
6.6 Percentage cost reduction of predictive maintenance strategies. 140
6.7 Baseline direct O&M costs and lost production, Scenarios 2-4. 141

6.8 Percentage cost reduction of predictive maintenance strategies - comparison of sites (Direct O&M Costs). 142
6.9 Percentage cost reduction of predictive maintenance strategies - comparison of sites (Lost Production Costs). 143

List of Tables

1.1 Important definitions 8

2.1 Generator failure rates study comparison 16

3.1 Diagnostic Method Summary 44
3.2 Parameters to Consider for Feature Extraction 45
3.3 Data Description 48
3.4 Diagnostic Health Indicator Description 50
3.5 Training Data Structure 50
3.6 Classification Results - Model Accuracy Comparison 51
3.7 Classification Results - Model Accuracy with Baseline 56
3.8 Test setup demonstrating real time application 58
3.9 Data pre-processing summary 60
3.10 Analysis cases 66
3.11 Available SCADA channels 67
3.12 SCADA features used in model 69

4.1 Case study 1 - Summary of data 86
4.2 Case study 1 - model features 87
4.3 SVM error model feature 90
4.4 Anomaly rate using single class SVM apprach. 92
4.5 Anomaly rate using standard RMSE approach 93
4.6 Summary of SCADA data 97
4.7 Case study 2 - Summary of Vibration data 97

4.8 Case study 2 - SCADA model features 99
4.9 Case study 2 - vibration model features 100
4.10 SVM error combined model features 100

5.1 Pool of data available for analysis 111
5.2 Time-domain features 113
5.3 Order-domain features 115
5.4 Operational characteristics 115
5.5 Cluster descriptions 120
5.6 Results for 5 clusters 120
5.7 Results for 3 clusters 124
5.8 Results for 4 clusters 124
5.9 Results by power binning 125

6.1 Maintenance strategy definitions 129
6.2 Intervention Categories 136
6.3 Analysis cases 137
6.4 Baseline failure rates [8] 139
6.5 Baseline costs overview 139

Acknowledgements

I would like to begin by gratefully acknowledging the EPSRC grant EP/L016680/1 which funded this work. I would like to thank the Wind and Marine Energy Systems CDT management and administrative staff, in particular Professor Bill Leithead, Professor Alasdair McDonald and Drew Smith who gave me the opportunity to be part of this wonderful research group.

A special thanks must go to my supervisors Dr James Carroll and Professor Alasdair McDonald for all the patience, support and guidance given to me throughout my PhD. I am extremely thankful to have had the opportunity to learn from them both over the past 4 years. I would also like to thank industry partners for sharing data and making this research possible. Unfortunately, for confidentiality reasons, I cannot name the companies, but I am sincerely grateful. To all other academics, researchers and industry partners involved throughout this research, I am thankful for the kindness, wisdom and experience shared.

I am also very grateful to my friends from the 2016 CDT cohort and the team spirit that was formed to help me through the stressful moments. Finally, and most importantly I would like to thank my family. In particular my wife Cate, who has shown unwavering support throughout this process always encouraging me to be the best version of myself. My parents Joy and Ronnie and my brother Jimmy for everything they have done for me that has allowed me to pursue a career in engineering.

Chapter 0. Acknowledgements

Chapter 1

Introduction

1.1 Thesis Background

In order for wind energy to continue to compete with traditional methods of generating electricity such as fossil fuels, the levelised cost of energy (LCOE) must be reduced in the coming years to meet ambitious targets set, for example, in the latest round of UK offshore leasing auctions. Costs associated with the operation and maintenance (O&M) of a wind farm makes up a significant proportion of total lifetime costs. In fact, up to 30% of the total energy cost can be spent on O&M for some large offshore developments [9]. With wind farms moving into harsher environments further offshore, this value is only expected to increase in the future. As more money is spent on O&M, innovations surrounding asset management have the potential to greatly influence the overall LCOE. According to a study found in [10], innovations associated with operations, maintenance and service are anticipated to reduce the LCOE by approximately 2% between 2014 and 2025. Generator faults can contribute significantly to the overall downtime experienced by a wind farm due to component failure, with around 1 failure per year in state of the art offshore wind turbines (WT) [4, 8, 11–14].

One of the areas in which significant improvement can be made is through the introduction of turbine condition-based maintenance [10]. All large utility scale WTs have supervisory control and data acquisition (SCADA) systems as standard, which are primarily used for performance monitoring [6], however developers are now opting

for more sophisticated condition monitoring (CM) systems to gain better insight into WT condition. The most widely developed and adopted CM systems are based on vibration analysis, with sensors placed throughout the nacelle to gain insight into the dynamic performance of WT sub-systems, and in turn identify any potential issues or faults. However, with SCADA systems currently fitted to every turbine installed in the last 10-15 years, it is also imperative to extract as much information from this data as possible, especially in cases in which CM systems are not readily available.

Wind turbine technology is continuously evolving, with operators now having the potential to own diverse sets of generation fleets located globally, all of which could be at very different stages of their design life. This creates a difficult situation where assets need to be monitored closely to ensure maximum availability across the entire portfolio. With today's rapid advances in automation and big data analytics, the industry is well placed to realise the potential operational expenditure (OPEX) savings associated with more complicated asset management strategies through the use of intelligent data acquisition, storage, processing and analysis. With increased wind farm capacity and asset portfolios growing in size there is a growing need for automated fault detection, which can utilise proven techniques and flag issues in real time across an entire fleet. With so much data now being gathered across multiple systems and platforms, understanding how to best use the data both in isolation and together is becoming increasingly important, having the potential to reduce O&M costs significantly if leveraged correctly.

The current chapter will first of all introduce the research question, the motivation behind this thesis and novelty of output. It will also give a high level approach and provide all important definitions required in the context of this research.

Chapter 2 will provide an extensive literature review of wind turbine drivetrains, associated reliability and cost of operations and maintenance. An overview will also be provided of state of the art condition monitoring systems, machine learning (ML) theory and applications of data driven diagnostics and prognostics used in wind energy research.

Chapter 3 will discuss how a range of machine learning techniques can be leveraged

to augment wind turbine diagnostics through normal behaviour modelling and fault classification. This chapter will first of all provide insight into utilising vibration based condition monitoring to detect a range of faults with high levels of accuracy. Secondly the use of SCADA temperature modelling will be explored to detect generator bearing faults.

Chapter 4 will present an enhanced methodology based on normal behaviour modelling to detect generator bearing faults, first of all by analysing SCADA data only, with a particular emphasis on how multiple error metrics can be used to detect faults earlier and more consistently than existing methods. Building on this same methodology it will then consider an example in which both SCADA and vibration data are available, presenting an approach to combine both analysis methods into a single fault detection model.

Chapter 5 will discuss how additional WT's and examples of similar failures can be used to further increase confidence in the diagnosis, while providing some insight into remaining useful life. A two-stage methodology will be presented to first of all cluster wind turbines by multiple operating conditions building on traditional binning techniques. Features from vibration data are then extracted in order to classify the turbine as 'healthy' or 'unhealthy', predicting failure within 1-2 months before occurrence.

Chapter 6 will then move on to looking at the costs of wind farm operation and maintenance in relation to different available condition monitoring and maintenance strategies. In particular this chapter will focus on modelling offshore wind farms, providing insight into how costs can be saved through early intervention and repair to reduce large component failures and major replacements.

Finally, Chapter 7 will present any conclusions that can be drawn from the research, discussing the findings, key advantages of the methodologies presented and the limitations surrounding the data sets used and any future work.

1.2 Research Question

The aim of the research is to provide insight into the following research question:

"How can machine learning techniques be leveraged to improve wind turbine generator diagnostics and prognostics, and what impact can using such approaches have on the overall wind farm O&M cost?"

To answer this primary research question several areas of wind turbine condition monitoring must be first explored individually. Using several case studies, the research will focus on applying techniques on known faults and failures associated with wind turbine doubly fed induction generators (DFIGs), with the overall aim of providing a broad understanding of real world application, practicalities and benefits. With that being said, this thesis will be broken down into 4 key research questions:

1. *How best can SCADA and vibration based condition monitoring systems be used in isolation to diagnose a range of common DFIG faults, and which machine learning techniques and decision metrics provide the most favourable results?*

2. *What are the benefits of having both SCADA and vibration data, and can they effectively be combined into a single anomaly detection model for fault diagnostics?*

3. *To what extent can previous experience of failure be used to predict remaining useful life in similar machines based on purely data-driven approaches, and how do different site operational conditions affect prediction accuracy?*

4. *How can implementing advanced monitoring and predictive maintenance strategies affect O&M costs throughout the life of a wind farm?*

These secondary research questions are answered throughout each chapter of this thesis and map onto Chapters 3, 4, 5 and 6 respectively. The beginning of each chapter will set out the research question to be answered, present the research and draw its own unique conclusions. These will then be brought together in the context of the primary research question in the final chapter.

1.3 Novelty of Research

By carrying out the literature review, a novel area of research was determined with few publications in the following areas:

1. Comparison of different machine learning techniques to detect generator faults with the use of real world SCADA and vibration data, with a focus on how different data sources, techniques and decision metrics can affect fault detection ability and times.

2. Insight into how SCADA and vibration analysis can be used together through synchronised data sets to increase confidence in a particular diagnosis, comparing this to existing methods of completing each analysis in isolation.

3. Understanding which features can be identified and extracted in the latter stages leading up to failure. The ability to compare failure prediction times and thresholds to provide a platform on which to build a robust approach to predicting failure and remaining useful life across similar machines at different stages of their design life.

4. Analysis of the extent that different maintenance strategies can affect the O&M costs of wind farms over the entire operational life.

1.4 Approach to Research

There are several steps in predicting failure using SCADA and vibration analysis, changing with the techniques and data set being used. In depth methodologies will be described throughout Chapters 3, 4 and 5, in which a variety of analysis and ML techniques will be applied as necessary. Figure 1.1 however shows a high level approach to this research, which forms the grounds for the layout of chapters throughout this thesis. When considering fault diagnostics (highlighted in green in Figure 1.1), at the

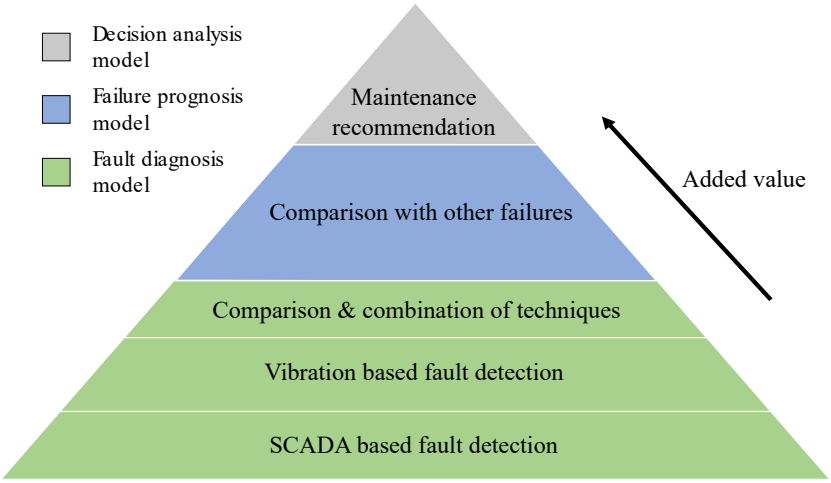


Figure 1.1: Approach to overall research chapters.

most basic level of analysis both SCADA and vibration analysis can be treated completely in isolation, making use of any available data streams. If leveraged correctly, these data sources can also be analysed, results compared and combined together to build confidence in any particular diagnosis, as will be described in Chapter 4. To develop these techniques for fault prognosis (highlighted in blue in Figure 1.1) and gain insight into remaining useful life, information is also required for similar faults in other identical machines in order to compare and classify accordingly, as will be discussed in Chapter 5. Finally, these methods can then be adapted further to provide insight into maintenance recommendations through a decision support or analysis model (highlighted in grey in Figure 1.1), which will be discussed in Chapter 6 as part of a cost analysis of different maintenance strategies. The more information is learned about a particular fault, moving up the pyramid, more value can be added and extracted to assist with asset management and O&M strategies.

1.5 Important Definitions

To avoid any confusion there are several key definitions which must first be introduced and understood before explaining any techniques, case studies, methodologies and results. These definitions are specific to this thesis and may differ from other definitions observed in literature. These definitions can be found in Table 1.1.

Table 1.1: Important definitions

Phrase	Definition
Fault	Fault or anomaly within a system which has the potential to develop into a failure
Failure	Failure of a component due to a single or multiple faults which causes the wind turbine to cease operation
Failure mode	Specific point and type of failure which causes the wind turbine to cease operation
Root cause	Underlying reason which leads to a particular failure mode with either single or multiple root causes for a single failure mode.
Diagnostics	Ability to determine if a fault is present in a system with either known or unknown fault location, severity or remaining useful life
Prognostics	Ability to determine fault severity and predict failure mode and remaining useful life
Feature	Used synonymously with predictor/input
Target	Used synonymously with response/output

1.6 Research Output

Published peer reviewed journal and conference articles:

1. *"Investigation of the relationship between main-bearing loads and wind field characteristics"*, Journal of Physics, 2017 [15]

2. *"Prediction of wind turbine generator bearing failure through analysis of high-frequency vibration data and the application of support vector machine algorithms"*, Journal of Physics, 2018 [16]

3. *"Prediction of wind turbine generator failure using 2-stage cluster-classification methodology"*, Wiley Wind Energy, 2019 [17]

4. *"Wind turbine main-bearing loading and wind field characteristics"*, Wiley Wind Energy, 2019 [18]

5. *"Combining SCADA and vibration to predict wind turbine component failure"*,

Wiley Wind Energy, 2020 [19]

6. *"Effect of time history on normal behaviour modelling to predict wind turbine failure"*, Energies, 2020 [20]

7. *"Cost benefit of implementing advanced monitoring and predictive maintenance strategies for offshore wind farms"*, Energies, 2021 [21]

Presentations:

1. *"Prediction of wind turbine generator bearing failure through analysis of high-frequency vibration data and the application of support vector machine algorithms"*, IET RPG, 2017

2. *"Wind turbine main-bearing loading and wind field characteristics"*, Deepwind, 2018

3. *"Comparison of anomaly detection techniques to predict generator bearing failure using SCADA data"*, WESC, 2019

Chapter 2

Literature Review

2.1 Wind Turbine Generator Technology

Wind turbine generator and drivetrain technology has developed rapidly over the last decade as utility-scale wind turbines have increased in size and contributed to a greater share of the electricity market [3, 22, 23]. This fundamental shift in the energy mix requires wind turbines to cope with greater flexibility in generation, with wind farms now operating more like traditional power plants to meet changing electricity grid conditions. The generator and wider drivetrain configuration has adapted with both grid requirements and the constant need to increase reliability and decrease downtime due to unplanned maintenance [24].

When describing a wind turbine drivetrain configuration it is typically expressed as a series of assemblies and components required to convert the kinetic energy in the rotor to electrical energy needed for a stable grid connection. In modern utility-scale wind turbines there are 4 major categories as described in [25] and [24]. Figure 2.1 shows these generator types and configurations in relation to key drivetrain components. The following list provides details of each configuration in relation to generator type. Note type A and B have been excluded from this review due to a focus on current utility-scale technology that meets modern grid requirements [26, 27].

- Configuration 1 (Type C). Doubly-fed induction generator (DFIG). Partial power converter is used to control the electrical current in the generator's rotor. Since

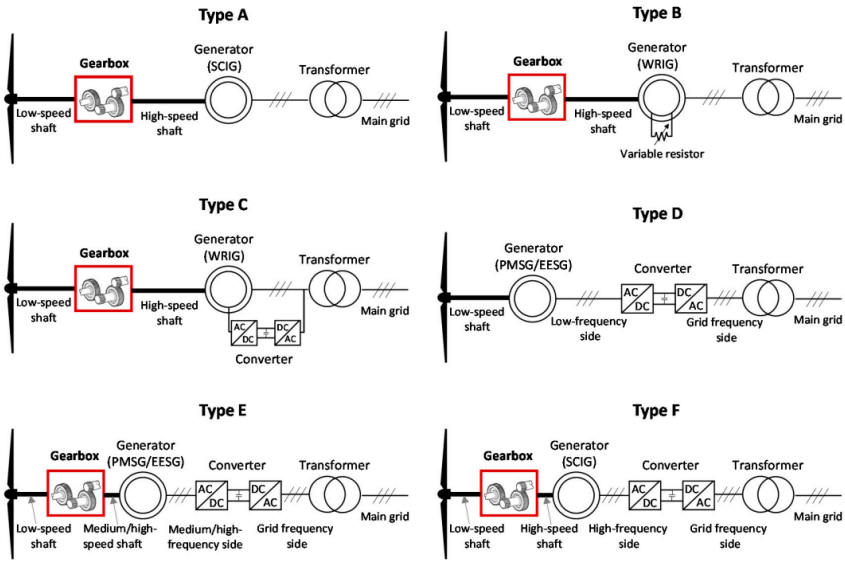


Figure 2.1: Drive train arrangements usually employed in commercial wind turbines [1]

the power converter is connected only to the rotor of the generator, the rated power of the converter is much lower (around 30% of the rated power of the WT), keeping costs of electronics down while enhancing response to grid requirements.

- Configuration 2 (Type D). Full power converter enables the decoupling of the generator and the grid frequency. This means that the frequency on the generator side can be fully controlled allowing for enhanced grid services and the use of a gearbox can be avoided. A synchronous electrical generator (which can be either an electrically excited synchronous generator (EESG) or a permanent magnet synchronous generator (PMSG)) is directly coupled to the main shaft of the rotor which operates at a rotational speed around 5-40 rpm, depending on the wind turbine size.

- Configuration 3 (Type E). Gearbox-equipped wind turbine with a full power converter and medium/high-speed synchronous generator, which can be EESG or PMSG. 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

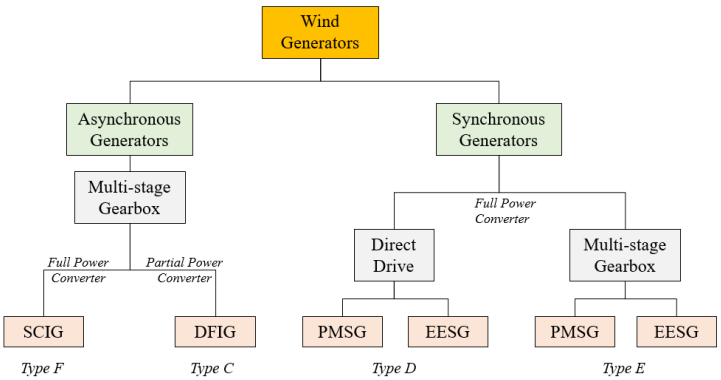


Figure 2.2: Generator topology classification with respect to configurations types. Adapted from [2]

generator (high-speed configuration with synchronous generator).

- Configuration 4 (Type F). 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.

Figure 2.2 shows more clearly how each of these drivetrain configurations relate to the generator types mainly observed in industry. More broadly, Type C corresponds to geared multistage highspeed wind turbine, Type D is a direct drive machine, while type E and F are hybrid models. In terms of market share, geared turbines continue to dominate the global market, as shown in Figure 2.3, with the vast proportion of these turbines onshore made up of a DFIG arrangement below 3MW rated power output. This is particularly true across Europe, Asia and North America. Further analysis presented in [3] shows the evolution of configuration types with geographical location, with Type F more prevalent in North America and Types D and E having more market share in Europe and Asia. If we look offshore across Europe, DFIG models dominated the early market predominately close to shore. This has vastly changed over the last 5-7 years with direct drive and hybrid models now making up a significant proportion. PMSGs have seen an explosion in the Asian market in particular, while EESGs are typically more common in European waters. Having said this, PMSGs have been

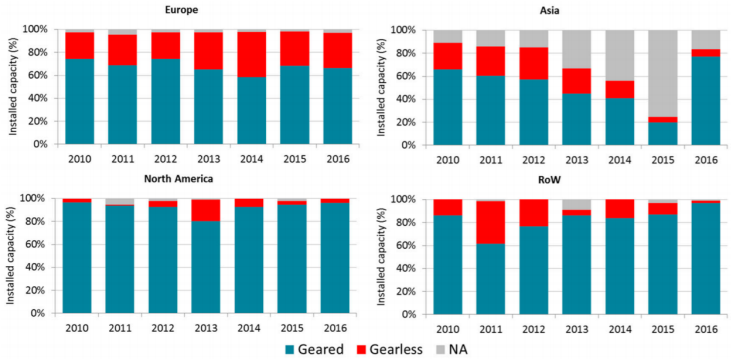


Figure 2.3: Evolution of drive train configuration in onshore wind turbines by geographical zone (Source: JRC Wind Energy Database. "NA" represents Not Available records in the Joint Research Centre (JRC) Wind Energy) [3]

gaining more traction in Europe as turbines increase beyond 5-6MW [22,25,28–31]. The technological shift towards PMSGs over other types of generator is predominately due to the increased reliability that can be achieved with less components and importantly no gearbox. The next section will discuss reliability of the systems described above.

2.2 Wind Turbine Reliability

Understanding component reliability is a fundamental part of engineering design, particularly when considering the design lifetime of critical products and infrastructure when costs of failure are at a premium. Due to wind farms being located in increasingly more remote locations better reliability is one of the most critical drivers for decreasing the O&M costs for wind energy. Reliability is often described as the probability that a product or a system will perform its intended functions satisfactorily. This means operating without failure and within specified performance limits, for a specified length of time, when operating under specified environmental and usage conditions.

One of the most common ways in which to present failure rates is the number of failures per turbine per year.

$$\lambda = \frac{\sum_{i=1}^{I} \sum_{k=1}^{K} \frac{n_{i,k}}{N_i}}{\sum_{i=1}^{I} \frac{T_i}{8760}} \tag{2.1}$$

where λ is failure rate per turbine per year, I is the number of intervals for which the

data is collected, K is the number of sub-assemblies, $n_{i,k}$ is the number of failures, N_i is the number of turbines and T_i is the total time period in hours. Other such terms used to describe reliability are mean time to failure (MTTF), mean time to repair (MTTR) and mean time between failures (MTBF), all having simple relationships as follows:

$$MTTF = \frac{1}{\lambda} \tag{2.2}$$

$$MTTR = \frac{1}{\mu}; \tag{2.3}$$

$$MTBF = MTTF + MTTR \tag{2.4}$$

where λ is failure rate per turbine per year as stated previously and μ is the repair rate, which is simply the transition rate from the failed to operational state [32, 33].

There have been several studies to date investigating the reliability of wind turbines and their sub assemblies, and it remains a crucial aspect of research to continuously update records for new WT technology. One such study in 2007, which shows failure rates and associated downtime are obtained from analysing 1,500 German turbines over a 15 year period [4]. It should be noted that while this is a large population the majority of turbines are rated under 1 MW, which is relatively small when compared to the 10-12 MW machines being installed offshore today. With regards to failure rate, Figure 2.4 shows that although the generator may not fail the most regularly in comparison to other sub-assemblies, it actually has the largest average down time per failure at over 7 days. It is the product of failure rate and downtime per failure which will effect total wind turbine availability and is what makes generators a valuable component to consider.

Another study worth mentioning by Spinato in 2009, [13], consists of a dataset of more than 6000 wind turbines in Denmark and Germany over an 11 year period. It calculates a generator failure rate of between 0.05 and 0.135 failures per turbine per year, with the range reflecting different ages and sizes of turbines studied. In general lower failure rates are associated with smaller turbines, while the higher end of the scale is associated with larger turbines, including direct drive machines, which are known to have higher rates of generator failure [11].

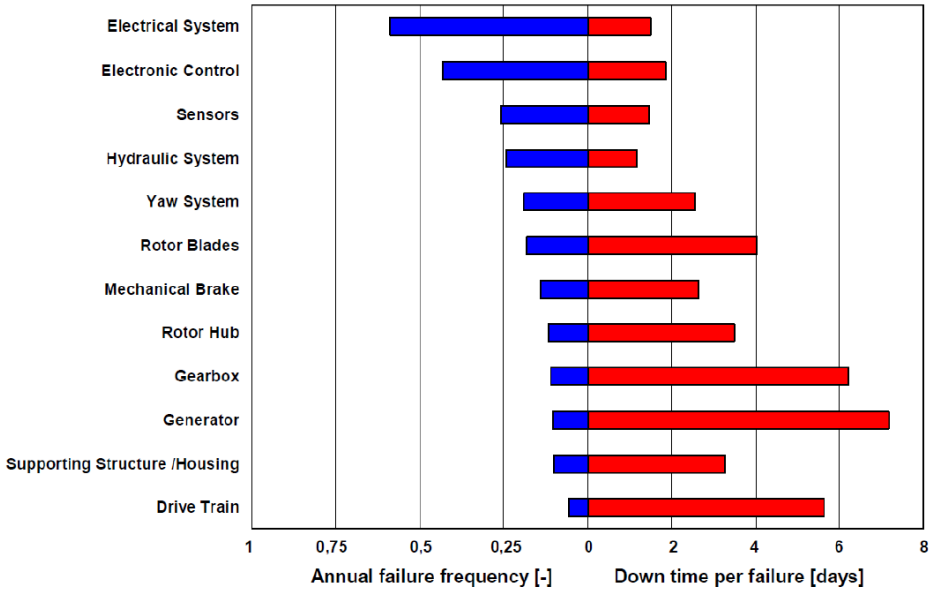


Figure 2.4: Wind turbine sub-assembly failure rates and downtimes [4]

In 2010 a similar study from Reliawind, [14], calculates a failure rate of 0.14. Considering the fact that this investigation includes larger, more modern wind turbines to those used in Spinato's study, the failure rate is remarkably consistent with the top end of the scale found in [13].

In 2015 Carroll, [11], considers a dataset of 2222 wind turbines, each of similar power rating between 1.5 and 2.5 MW. They differ only by their generator and drivetrain configuration and have been categorised into two groups for the study. The first configuration in the analysis is a doubly fed induction generator (DFIG), with the second being a permanent magnet generator (PMG) with fully rated converter (FRC). All turbine generators and converters are in their first 5 years of operation and located in wind farms throughout Europe. This provides a unique insight into failure rates with a large population of very similar machines. The DFIG configuration showed a failure rate of 0.123 while the PMG population was calculated at 0.076. Table 1 provides a comparison between each study, which suggests that the rate of generator failure increases with turbine size, leading to lower reliability in larger machines. However, it should be noted that no such reliability study currently exists which considers turbines beyond approximately 3 MW, which can verify if past trends still hold for such large

Table 2.1: Generator failure rates study comparison

Study	**failure rate**
Spinato [13]	0.135
Reliawind [14]	0.14
Carroll, DFIG [11]	0.123
Carroll, PMG [11]	0.076

and increasingly complex wind turbines.

When comparing DFIG and PMG configurations, the PMG has a failure rate 40% lower to that of the DFIG, which actually could be lowered further if minor failures related to its cooling and lubrication system were not included in [11]. Spinato's study [13], also suggests that direct drive machines are not necessarily more reliable when compared to their geared counterpart. In fact if you take into consideration aggregated failure intensity of the generator and converter in the DD machine, it is greater than that of the gearbox, generator and converter in geared drive wind turbines.

Finally in a more recent study in 2018 [34], Artigao identified thirteen reliability studies in scientific literature to review. This work made an effort to unify the various studies to obtain comparable results. Failure rates are normalised and it is shown that the control system, gearbox, electric system, generator and hub & blades are the most critical assemblies with regards to wind turbine condition monitoring. Taking into account the top-three contributors to failure rates for each study, the most recurrent ones across all studies are the electric, control and yaw systems, and hub and blades categories; for downtime these are the gearbox, generator, and braking and electric system. Mechanical and electrical/control components show similar failure rates. However, mechanical components cause higher amount of downtime when compared to electrical/control ones, reaching more than 75% of the total downtime.

With failure rates of this magnitude leading to high levels of downtime it is clear that improvements must be made during the design stage to increase the reliability of wind turbine generators. However, with the current rate of wind farm development both onshore and offshore, it is expected that these numbers will only increase in the near future. It is therefore imperative that assets are managed intelligently through

the use of condition monitoring systems to allow operators to effectively and efficiently plan maintenance work which minimises turbine downtime, or alternatively provides reliable information in the case that power must be limited or turbine shut down to avoid serious or dangerous failure.

2.3 Wind Turbine Condition Monitoring

2.3.1 Classes of Monitoring Systems

There are several systems and approaches which exist to determine how a particular wind turbine is operating and perhaps more importantly performing.

The first class of system is the Supervisory Control and Data Acquisition (SCADA) system, which is used to determine and record basic functionality and health. This typically involves storing 10-minute averaged signals along with the associated standard deviation of active power output, measured wind speed, gearbox bearing and lubrication oil temperatures, generator winding and bearing temperatures, as well as phase currents and power factor. Every wind turbine currently installed has a SCADA system installed as standard.

The second class of system is the structural health monitoring (SHM) system, which is typically interested in (but not limited to) the structural integrity of the tower, support structure and foundations. As these faults are typically driven by blade-passing frequency low frequency sampling of below 5Hz can be used.

The final class of system, which is most applicable to generator failure, is condition monitoring (CM) and diagnostic systems. These systems can be used to provide reliable information on the performance and operational health of a wind turbine power train to detect early signs of any anomalies that may exist within subassemblies. Once an anomaly is detected, these systems have the potential to diagnose specific faults, severities and locations depending on the particular issue. This information can then be used to schedule repair work and calculate any losses or required downtime.

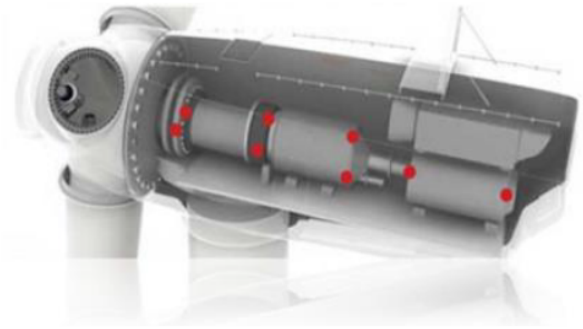


Figure 2.5: Typical positions of accelerometers used for a condition monitoring based on vibration analysis [8].

2.3.2 Condition Monitoring and Diagnostic Tools

Most condition monitoring systems used in wind turbines have been developed from traditional methods applied to rotating machinery, which are primarily based on drive-train vibration analysis. This typically involves placing accelerometers at key locations on the main bearing(s), gearbox and generator bearings, as shown in Figure 2.5. The sampling frequency used to measure the vibration signal is usually variable, but typically has a maximum of between 40 and 50kHz (although a higher sampling frequency of up to 96kHz is available in some products) [35]. Analysis of these signals is typically done both in the time domain and frequency domain.

More sophisticated systems now look beyond traditional methods of analysing mechanical vibration by also considering electrical signals. This may involve high frequency current signature analysis to gain further insight into how the generator is performing and operating, as performance is ultimately linked to power output [36,37]. Due to the high volume of data generated with such high sampling frequency, CM systems tend to work alongside SCADA systems logging data periodically. It can then be used for investigative purposes if a fault is suspected or detected.

One other technique worth mentioning is oil debris analysis, which can be done periodically by taking oil samples during routine inspections or maintenance work, or by having an inline system fitted to the gearbox oil inlet. These types of systems can count oil particles (both ferrous and non-ferrous) of different sizes and provide live

remote access to the data through the CMS system network [38].

This thesis focuses on data driven methods for condition monitoring using both SCADA and vibration data, the two most widely used techniques in industry. Oil particle counting or debris analysis, although useful for overall drivetrain condition monitoring, is more applicable to gearbox faults and will not be considered as part of this thesis. Although applicable, unfortunately no high frequency current data could be obtained.

2.4 Applying Machine Learning Techniques to Wind Turbine Condition Monitoring

A range of data driven techniques have been published in relation to wind turbine condition monitoring. This section provides an overview of techniques and approaches observed in literature with regards to SCADA and vibration monitoring.

It is worth explicitly mentioning that fault prognostics considers individual turbine component health, as opposed to a population-based assessment when considering more conventional reliability analysis. Depending on the techniques and data source it is important to consider the capabilities of modelling approaches in relation to what the model can achieve. Broadly speaking there are two key aspects to data-driven modelling, the first of which is fault detection (diagnostics) and the second of which is failure prediction (prognostics). With regards to fault diagnosis there are several levels of capabilities, which are often used together but can sometimes cause confusion. The first level, usually performed first, is anomaly detection, which aims to identify any deviation from normal operating conditions. These models will typically not have capabilities to explicitly locate fault position within a specific assembly, sub-system or component. To do this we need to have some level of fault classification, which can be used to narrow down or identify the source of the problem to component/sub-component level, or even down to the failure mode itself depending on the model inputs and capabilities. Once a fault has been detected and classified, we move through to the failure prognostics stage, where fault severity must be established and degradation

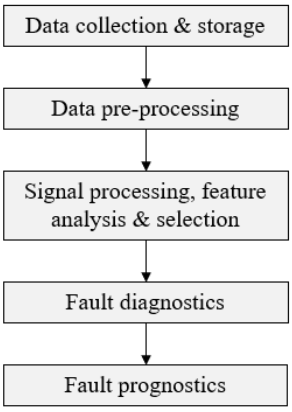


Figure 2.6: Typical condition monitoring work flow.

measured. This stage is the most difficult which, for purely data-driven models, is achieved by utilising known examples of faults and/or failures. A typical condition monitoring workflow is shown in Figure 2.6 highlighting where the key steps described above fit into the overall process.

The aim of any condition monitoring system is to aid decision makers in their operations and maintenance strategy. For this reason a decision support system must have the ability to create insight from various amounts of data across different data sources. On a fundamental level machine learning can be thought of as a process of building a model automatically from data. How the model learns from the data depends on the algorithm, but in all cases it is the underlying structure and relationships in the data that are of importance and hence is what the algorithm is trying to understand. Machine learning can be categorised as either supervised learning, which predicts an output variable from labeled inputs, and unsupervised learning, which deduces a desired output from data inputs without labeling. For supervised learning there is a distinction between models that predict a numeric variable (regression model) or a categorical variable (classification model).

Machine learning techniques are increasingly being applied to wind turbine diagnostics and prognostics, with multiple review papers now having been published in the area over the last few years which can be found in [6, 39, 40]. According to [41] almost two-thirds of methods published in research used classification algorithms with the rest

relying on regressive based approaches.

Machine learning algorithms can be categorised further into parametric and non-parametric models. Parametric models estimate parameters to learn a function which relates the training inputs and outputs. Once an algorithm is trained predictions can be made without the need for the original training dataset. In contrast, non-parametric models cannot be categorised by a fixed set of parameters. The interested reader can find more about the fundamentals of these methods in [42] and how to apply them in [43], although some techniques will be explained in further detail in the following sections.

2.4.1 Supervised Learning

Supervised learning algorithms require input data to be labeled in in order to learn the structure and relationships between input and output variables or classes. Depending on the required output supervised learning techniques are categorised into regression or classification problems.

Classification models learn to classify data into discrete classes, which can be binary (2 classes) or multi-class (3+ classes). To do this they learn from input data which is labeled into categories. This process of labelling data is done prior to training, and is typically achieved by assigning a category to training samples by human or expert judgement. It is important to try and produce a balanced set of labeled data with a similar number of instances in each class, however the problem of unbalanced classes can be addressed by under-sampling (remove instances belonging to the majority class for training) or over-sampling (sample more instances from minority class for training) [44]. Later on in this thesis we will see examples of under-sampling to balance training sets for classification.

Unlike classification models, regression models predict an independent numerical output using a set of independent input variables. A simple example of this could be to predict power output based upon wind speed and generator speed.

The terms ‘bias’ and ‘variance’ are often used in literature (and will be throughout this thesis) in relation to supervised model performance. In general high variance is

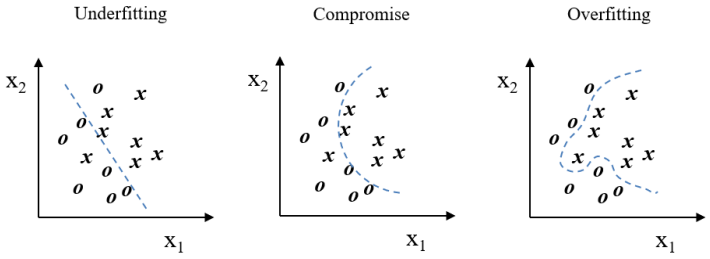
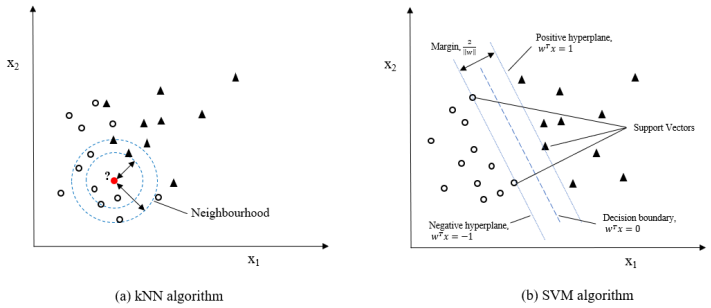


Figure 2.7: Difference between underfitting (bias) and overfitting (variance).

proportional to overfitting, while high bias is proportional to underfitting, which can be explained through Figure 2.7. The first example on the left shows a model that is underfitting, therefore does does correctly split the data into each discrete class producing a high number of misclassifications. On the right is an example of a model which is overfitting, meaning that although the data is split into the correct classes, any new data that that is marginally outside of the current trained behaviour risks being misclassified. In practice a compromise between the two is usually desirable. In the context of training machine learning models, variance is a measure of the variability (or consistency) of the model prediction for classifying a particular example if the model was retrained multiple times. If a model has high variance we can say that it is sensitive to randomness (noise) in the training data. Conversely, bias is a measure of the difference between model predictions and correct values if the model is rebuilt multiple times on different training datasets. If a model has bias we can treat this as the systematic error that is not due to randomness or noise. Understanding model bias and variance is an important part of model validation and can be achieved through training and evaluating subsets using cross-validation, however this will be explained in greater detail in later chapters. Some common algorithms will now be briefly described.

K-nearest Neighbours (kNN)

The kNN algorithm belongs to a subcategory of non-parametric models described as instance-based learning. This is different to all other models described later as they work by memorizing the entire training dataset. The algorithm is relatively straight-



(a) kNN algorithm (b) SVM algorithm

Figure 2.8: Example of kNN and SVM algorithm.

forward and can be summarised as follows:

1. Define the number of k points and a suitable distance metric, where k is the number of neighbours the algorithm will use to calculate an output.

2. Calculate the distance d between the unseen observation x_{obs} and each observation used for training.

3. Determine the k number of closest observations and create a subset to analyse.

4. Probability of x_{obs} belonging to each class is calculated and assigned to the class of the highest probability.

The most important hyperparameter here is k and can have a large impact on the model performance. A small value for k provides the most flexible fit, which typically means low bias but high variance. Larger values of k will have smoother decision boundaries meaning lower variance but increased bias. The kNN algorithm can be applied to classification or regression problems and is one of the simplest models to apply and visualise, as shown in Figure 2.8(a). As stated previously this is a non-parametric model meaning it requires the entire dataset to be stored, which in practice can lead to storage and computational efficiency problems. That being said, one of the key advantages of using a memory-based approach is that the classifier can immediately adapt when new data is collected.

Support Vector Machines (SVM)

SVM's are more widely used in classification problems, and work by defining a decision boundary that can separate classes in multi-dimensional space (ie can deal with a large number of features). SVM's goal is to not only find a decision boundary but to optimise the hyperplane between classes. For a n-dimensional space the hyperplane will be an (n-1)-dimensional subspace which creates the largest margin between training points of the different classes. If we first consider the simplest case of two linearly separable classes, the margin is simply the maximum distance perpendicular to the hyperplane that has zero interior data points, as demonstrated in Figure 2.8(b). For the simple linear example as shown in Figure 2.8(b) shows the linear decision boundary, $w^T x = 0$, with the margin separating the positive and negative hyperplane represented as $w^T x = 1$ and $w^T x = -1$ respectively, where w is the SVM weight parameter, or coefficient. Support vectors are defined as the data points that are closest to the hyperplane supporting the decision boundary. In practice minimising $\frac{1}{2}\|w\|^2$ is used to solve the problem of maximising the margin with the use of quadratic programming. The full mathematical formulation behind the SVM algorithm is relatively simple and for the interested reader further details can be found in several references including [45], [46] and [43]. By maximising the margin the classes are separated by a greater distance leading to better algorithm performance by having lower generalisation error and less problems with overfitting.

There are two other scenarios worth mentioning with regards to SVM classification, each requiring a more complex solution than previously described. The first is used when dealing with nonlinearly separable cases for which the slack variable, ζ, was introduced by Vladimir Vapnik in 1995. This was initially introduced because linear constraints were not suitable for nonlinearly separable data and needed to be relaxed to allow for optimisation in the presence of misclassifications. To do this of course required appropriate cost penalties to be put in place for any misclassification. For these cases the new objective function to be minimised is as follows, where ζ is the slack variable and C is the penalty cost parameter:

$$\frac{1}{2}\|w\|^2 + C\left(\zeta^{(i)}\right) \tag{2.5}$$

The second scenario is regarding the use of a kernel SVM to solve nonlinear problems. The fundamental approach behind kernel methods is to create nonlinear combinations of the original features to project them onto a higher-dimensional space via a mapping function, ϕ, to then train a linear SVM to classify the data in this new feature space. The same mapping function, ϕ, can then be used to transform new data to be classified. In order to save expensive computational time mapping and training through quadratic programming a so-called *kernel trick* is implemented in practice. This is where a kernel function is defined, the most common of which is the radial basis function (RBF), or Gaussian kernel as it is otherwise known. Again, the interested reader can find more details surrounding the mathematics behind this in [45].

Decision Trees

Decision tree classifiers provide an approach that can be very easily interpreted and visualised regardless of model dimensionality (number of features). Starting with a root node, a decision tree is a hierarchical structure with a series of learned questions that splits the data based on the feature that provides the largest information gain. This splitting process is repeated for each child node until all leaves are considered pure, meaning that all training examples at each node belong to the same class.

The downside of a decision tree is that it can easily lead to the model overfitting if a tree is too deep. Tree depth can be limited by setting a maximum number of splits, a minimum leaf size and a minimum parent size; an act known as pruning. In order to split nodes efficiently at the most informative inputs an objective function we want to optimise must be defined. For example if we want to maximise the information gained, IG, at each split we can define the objective function as follows:

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^{m} \frac{N_j}{N_p} I(D_j) \tag{2.6}$$

where, f is the feature to perform the split, D_p and D_j are the datasets of the parent

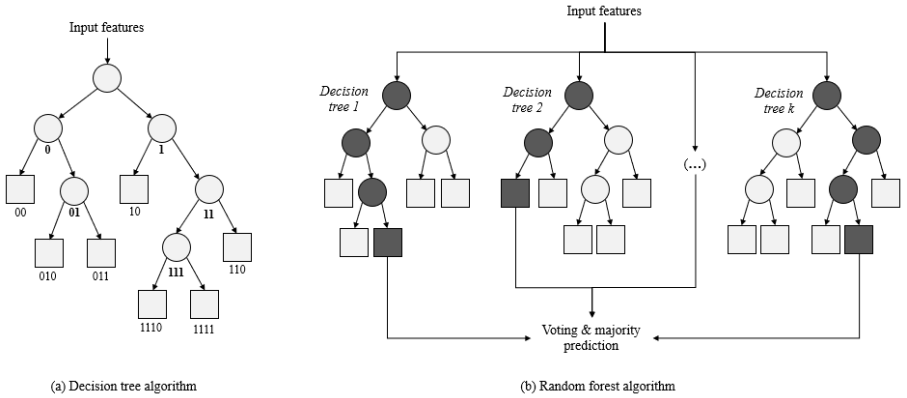


Figure 2.9: Example of Decision Tree and Random Forest algorithms.

and jth child node, I is the chosen impurity measure, N_p is the total number of training examples of the parent node, while N_j is the total number of samples of the jth child node. There are three common impurity measures used in simple decisions trees; Gini impurity, entropy and classification error. As an objective function the Gini impurity aims to minimise the probability of misclassification while entropy aims to maximise the mutual information in the tree. In practice both of these impurity measures yield similar results when growing a tree [43]. Classification error is more suited to assessing the tree for pruning rather than initially designing the tree. A simple representation of a decision tree splitting process is shown in Figure 2.9(a).

Random Forests

Ensemble methods in machine learning utilise multiple algorithms to gain more accurate and stable predictions than could be achieved with a single algorithm. Random forests can be considered an ensemble of decision trees, which aims to overcome some of the limitations of a single decision tree of overfitting and high variance. The random forest algorithm is summarised with the following steps:

1. Take a random bootstrap sample of size n, which is randomly chosen.

2. Grow a decision tree from the data sample taken, splitting the nodes in accordance with the objective function as described in the above section on decision tree algorithms.

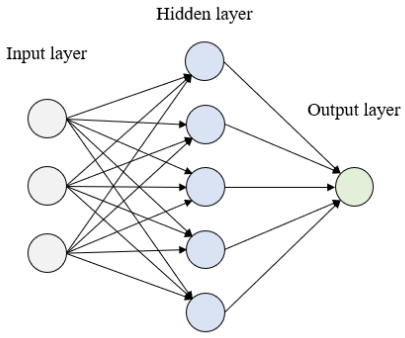


Figure 2.10: ANN algorithm example diagram.

3. Repeat steps 1-2 k times, with k being the number of decision trees in the forest.

4. Assign a class by majority vote considering each tree prediction.

In theory this process provides a more robust, generalised model which is less susceptible to overfitting. Like decision trees, the random forest algorithm can be adapted to both classification and regression problems. Again, a simple representation of a random forest prediction process is shown in Figure 2.9(b).

Artificial Neural Networks (ANN)

The ANN algorithm has become incredibly popular in research with regards to wind turbine condition monitoring, likely due to ease of application and adaptability to both regression and classification problems. An ANN has three distinct layers, as depicted in Figure 2.10 - an input layer, hidden layer(s) and an output layer. The input layer is made up of a number of nodes, which are connected to a series of neurons in the hidden layer. The inputs and the connectivity with the internal neurons are computed and learned in a iterative process to determine weight parameters. Each neuron has a nonlinear transfer function to determine its own unique inputs and outputs with an activation function to provide an output limit. Further reading on ANN theory and practice can be found in [47] and [48].

Semi-Supervised Models

The most notable of the semi-supervised category of models is the single class classifier. Unlike supervised models these algorithms learn patterns in a data belonging to a single class, with the goal of detecting outliers that do not match the original learned behaviour. With regards to wind turbine condition monitoring, this means that any deviation from the trained behaviour will indicate an anomaly and flag as not belonging to the class describing normal operation. This could then indicate a possible system fault. Several types of single class classifiers exist and can be based on several of the techniques already discussed including decision trees, neural networks and SVMs. Taking SVM as an example, instead of finding the maximum margin of the hyperplane between two or more classes of features, single class classifiers determine the boundary which encloses the outermost support vectors. Single class SVMs seem to be the most used semi-supervised model in literature, likely due to ease of model visualisation, with a wide range of applications including wind turbine bearing fault detection [49, 50].

2.4.2 Unsupervised Learning

The key difference when it comes to unsupervised learning is that models do not require labelled data in order to find underlying patterns. As labelling is often a time consuming and manual task, removing this step has obvious benefits, however these approaches come with limitations which should also be understood. The main applications of unsupervised learning is to segment groups by some shared attributes, detect anomalies that do not fit to any group, or to simplify datasets by aggregating variables with similar attributes.

The two types of unsupervised learned that will be discussed in this section are; grouping data through clustering techniques and reducing dimensionality through principle component analysis (PCA).

Clustering

Clustering techniques utilise similar properties of data points to cluster data into specific groups. Some of the most common clustering algorithms are k-means clustering,

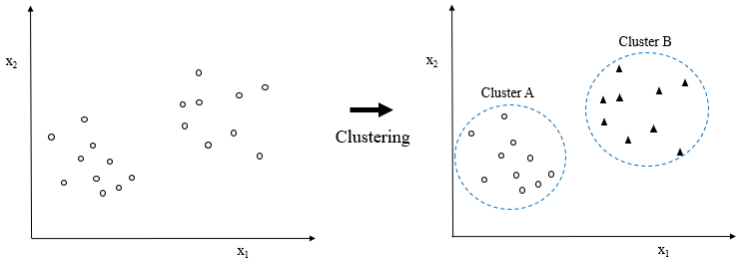


Figure 2.11: Clustering basic principle.

hierarchichal clustering and Gaussian clustering models (also described as distribution based clustering). Figure 2.11 shows the basic idea of clustering to define set clusters or groups, and the interested reader can find more details of a range of clustering techniques in [51] and [52]. The simplest algorithm (and most widely used to due ease of application and low computational effort) is k-means clustering, which determines the distance between datapoints, with closer data points more likely to belong to the same cluster. The most common distance to use in k-means clustering is the squared Euclidean distance, which can be described as the distance between two points x and y in m-dimensional space as follows [53]:

$$d(x,y)^2 = \sum_{j=1}^{m}(x_j - y_j)^2 = \|x - y\|_2^2 \tag{2.7}$$

where j is the jth dimension or feature in the context of machine learning. The k-means algorithm works by minimising the cluster inertia factor, which is simply the sum of the squared errors within the cluster calculated as follows:

$$CI = \sum_{i=1}^{n}\sum_{j=1}^{k} w^{(i,j)}\|x^{(i)} - \mu^{(j)}\|_2^2 \tag{2.8}$$

where $\mu(j)$ is the centroid for cluster j and $w(i,j)$ is the cluster number (ie. if sample $x(i)$ is in cluster j, $w = 1$). To cluster the data the algorithm performs the following steps:

1. Number of clusters, k, is chosen in which to split the data

2. k centroid locations are chosen at random

3. Each data point is assigned to the nearest centroid using the euclidean distance
4. Cluster inertia is calculated from each data point and centroid location
5. New centroid locations are determined by calculating the minimum quadratic error of the data points to the centre of each cluster, moving the centroid towards that point
6. Repeat step 3, 4 and 5.

An important consideration when implementing the k-means algorithm is the number of clusters and hence centroids to generate. This value can be achieved through domain knowledge, however if an analytical solution is required the elbow method is often implemented to make the decision, which works by plotting the ascending values of k versus the total error obtained for that k value. Other hyperparameters to consider are the maximum number of iterations to run the algorithm (steps 3-5 above), and the number of times the full algorithm will run with different random initial centroid positions (or seeds). It should be noted that the output of k-means clustering will not be the same each time the algorithm is run even with a fixed dataset and hyperparameters due to the random initiation of centroid positions.

Another clustering method worth discussing in a little more detail is hierarchical clustering, which in contrast to k-means, allows the number of optimum clusters to be determined by the algorithm. The ability to visualise the cluster relationships by dendogram plots is also an attractive reason for choosing hierarchical clustering. Finally Gaussian Mixture Models (GMM), belonging to a group of soft clustering techniques allow for a probabilistic approach to clustering by assigning a likelihood of a data point belonging to any particular cluster based on its closeness. For those interested, more detail on clustering techniques and the mathematics behind the algorithms can be found in [51].

Principal Component Analysis

Principal component analysis (PCA) is a unsupervised technique that can be used in order to reduce the dimension of a dataset. When considering model dimensionality

the number of training samples (N) should not exceed the number of features (L), with different N/L ratio required based on the specific application, features available and underlying assumptions [54]. PCA can be performed with the following steps, which provides the orthogonal transformation of possibly correlated variables into a set of linearly uncorrelated ones (principal components) [55]:

1. The mean of each data feature (dimension) is calculated and subtracted from the original dataset
2. Covariance matrix is calculated
3. Eigenvalues and unit eigenvectors of the covariance matrix are calculated
4. Eigenvectors are sorted from highest to lowest eigenvalue.
5. N number of features is chosen based on explained variance complexity trade-off and a feature vector is obtained by combining the first n eigenvectors
6. The post-PCA dataset matrix is obtained by multiplying the transpose of the feature vector by the transpose of the the adjusted dataset matrix

In terms of dimension reduction, the optimum number of principal components is a trade off between gaining the maximum reduction in dimensionality while maintaining as much explained variance as possible. This will limit the amount of information lost from the original dataset while reducing the number of features.

2.4.3 Techniques Used Throughout this Thesis

Most of the data driven models and techniques described in the previous sections will be utilised at some point during this thesis. Where possible models will be compared, however different models can be chosen to perform the same task and it is not possible to compare all eventualities. Models may be chosen at times in reflection of the best perceived model based on both literature and experience at the time of the specific research or analysis. When implementing machine learning models both MATLAB's *Machine Learning Toolbox* and Python based *Scikit-learn* and *Tensorflow* packages have been utilised at different points in the thesis.

2.4.4 Existing Research in Applying Data Driven Methods to SCADA and Vibration Analysis

In an effort to optimise CMS activities, machine learning techniques are increasingly being used to enhance wind turbine diagnostics and prognostics, with extensive publications and multiple review papers now having been published over the last few years which can be found in [6, 39, 40]. In a review by Stecto et.al in 2018 [41] models were classified by typical ML steps, including data sources, feature selection and extraction, model selection (classification, regression), validation and decision-making. Findings showed that most models use SCADA or simulated data, with almost two-thirds of methods using classification and the rest relying on regression.

Focusing on SCADA data only, one comprehensive review [6], splits up the general approaches to data driven fault detection into three main categories; trending, clustering, and normal behaviour modelling. Initial research into trending involved using regressive techniques to develop simple relationships between 2 variables, for example power against component temperature, or increasing this to three variables by including ambient temperature [56]. A technique using correlations among relevant SCADA data is investigated in [57], which used power binning to track changes across operating conditions through time. Temperature trending as a function of power output over different time scales was investigated in [58]. Principle Component Analysis can also be utilised for trending purposes using an auto-regressive neural network as shown in [59]. Building upon these techniques research began to focus on clustering algorithms, which could more easily be applied to online systems due to the ability to automatically set thresholds and classify observations. Initial attempts used k-means clustering to measure anomalies based on the Euclidean distance from cluster centroid positions [60]. ANN self organising maps were later used to visualise large datasets in [59] and [61]. These techniques were then later applied in [62] where examples were presented to detect gearbox failure by comparing the quantisation error.

Research involving normal behaviour models has been widely published in recent years across various condition monitoring applications. The full process of developing NBMs will be discussed throughout Chapters 3 and 4, however the main principle is

very similar to previous methods whereby normal operation is used as a basis for which to base anomaly detection. The key difference for NBMs is that they rely on empirically modelling a measured parameter during a training phase to set the boundaries of what is considered normal behaviour. The residual of the measured parameter minus modelled parameter is used as a fault indicator, with normal behaviour having residual values of approximately zero (± expected noise). [6] differentiated NBMs into two distinct categories; Full Signal ReConstruction (FSRC) and AutoRegressive with eXogenous input modelling (ARX). The former uses only signals other than the target variable to predict the target variable, while the latter also makes use of the historical values of the target variable.

Models range from simple linear and polynomial fits such as [5], which used a linear ARX model to detect generator bearing failure by modelling bearing temperature and [62], which developed higher order polynomial FSRC models of drive train temperatures. Similar approaches have also been shown to detect faults of other wind turbine components such as transformers in [63]. To improve upon these modelling techniques ANNs were introduced to capture non-linear relationships between observations and increase model dimensionality, as demonstrated in [64] through detection of bearing damage in offshore wind turbines. In [65], it was demonstrated how ANNs could be used to detect main bearing damage three months before the turbine was stopped due to overheating by modelling main bearing temperature. ANNs adaptability to different data was also shown in [66], where high frequency SCADA data was successfully used to detect bearing faults. Two back propagated Neural Networks (BPNNs) were used in [67], one to select relevant features, and the other to detect anomalies based on the RMSE between the measured and modelled target parameter. Other presented techniques include [68], which used a kNN algorithm to detect incipient failure in two turbines up to 6 months before failure. Nonlinear auto-regressive neural networks with exogenous inputs (NARX) models have recently been used in [69–71] to detect a range of gearbox component faults. Other research into fault detection using SCADA data include [72], which used a non-linear state estimation technique to model gearbox behaviour, while probabilistic based methods were presented in [73].

Research involving data driven methods and high frequency vibration data has also been published extensively over the last decade in relation to wind turbine diagnostics. Unlike SCADA analysis, vibration signal analysis can make use of component kinematics and fault signatures to make a specific diagnosis. Much of the early work focused on wind turbine gearboxes due to reliability issues and contribution to turbine downtime, however this has since extended to other drivetrain components such as generators, main bearings and blades. It was shown in [74] that diagnosis of gearboxes can be performed using time, frequency or time-frequency methods to analyse vibration signals.

Methods to analyse the signal in the time domain have been proposed and are often based on statistical analysis methods to describe the time waveform such as peak value, RMS, kurtosis, mean, standard deviation and skewness. In lab conditions this has proven to be a successful approach for both gearboxes [75] and generator bearings [76]. To apply these simple techniques in real world applications however does not prove useful due to variable loads, non-stationary signals and additional noise.

There are many techniques using the frequency domain that have been proposed across different components and assemblies depending on specific fault signatures. The simplest of these is using Fast Fourier Transform, which looks at either the whole spectrum or specific frequency bands to extract features as demonstrated in [77]. When it comes to bearing diagnosis, Fast Fourier Transform is typically inadequate and requires more advanced signal processing techniques such as envelope analysis as shown in [78]. Regardless of the method used, once features have been extracted from the raw vibration signal, machine learning methods have proven to be a useful method of automatically classifying and detecting faults as shown in [16], which employed an SVM classifier to successfully detect generator bearing faults. Gaussian Mixture Models (GMMs) were used in [79] to detect low speed bearing faults using both frequency and time domain features. A comparison of wind turbine gearbox vibration analysis algorithms based on feature extraction and classification can be found in [80, 81].

2.5 Cost of Operations & Maintenance

The costs of wind energy can be broken down into Capital Expenditure (CAPEX) and Operational Expenditure (OPEX). The CAPEX includes all one-time expenditure associated with wind farm development, deployment and commissioning. This typically includes all turbine hardware and installation costs, electrical infrastructure costs, civil works, SCADA and monitoring systems as well as all required permits and licensing.

OPEX includes all ongoing expenditure to operate and maintain the wind farm. The most important OPEX costs include the costs of running the site, including land and sub-station rental, insurance and taxes, management and administration, scheduled O&M activity costs as well as some allowances for unplanned maintenance to repair any unexpected failures. According to several studies including [9], technical reports by wind industry specialists and UK government experts [82, 83] the costs of OPEX can be up to 40% of total costs for some offshore developments.

Not much literature exists when it comes to quantifying the impact of predictive maintenance strategies on wind farm OPEX. A review on reliability and its impact on cost of energy for wind is given in [84]. One attempt to quantify the impact of both costs and revenue can be found in [85]. The influence of extending potential-to-functional failure intervals has on offshore wind turbine availability is explored in [86].

Chapter 3

Wind Turbine Fault Diagnostics using machine learning to enhance vibration and SCADA based analysis

3.1 Chapter Contribution

Vibration analysis is widely used within the wind industry for early detection of a range of mechanical and electrical faults associated with both the generator and gearbox highspeed (HS) assembly. As described throughout the literature review, the most common of these component faults are typically associated with the bearing assemblies, with faults located on either the inner race, outer race or/and rolling element itself. The fault type within each bearing assembly can be determined by the unique frequencies associated with each damage type, with a turbine typically consisting of several different bearing assembly configurations. To determine which bearing assembly a particular fault type is located within, the exact bearing design and kinematics must first be understood. Although detecting these faults through vibration analysis is well understood, the wind industry poses a difficult dilemma when it comes to large offshore wind farms, when inspection is not easily and cheaply available, requiring ev-

ery wind turbine to be remotely monitored and analysed individually with specialist engineering knowledge. Reducing engineering hours spent analysing these types of sites offers a huge challenge to operators, however by taking specialist engineering knowledge and allowing classification algorithms to help detect and locate faults offers a scalable approach with great potential of driving down costs.

Using SCADA systems to analyse WT performance has become standard industry practice over the last decade, with SCADA systems installed as standard in all modern machines currently in operation. With so much data readily available, operators are now looking at how SCADA data can be used to enhance and augment more traditional vibration based WT condition monitoring. To understand fully how SCADA data can be leveraged to detect generator and high speed faults, data must be made available leading up to catastrophic failure; something that is rarely available to researchers but would allow more models to be tested, validated and compared. In doing so could lead to standardisation of approaches and further understanding of what data is actually required for operators to reap optimal benefits. Based on the literature review in Section 2, the majority of research to date in this area has focused on gearbox faults, with comparatively less using WT generator case studies. As Section 2 also shows, the reliability of WT generators and HS assemblies is only marginally better than that of the gearboxes, therefore understanding how models can be used on such faults can still add great value to operators.

This chapter aims to answer the following research questions:

"How best can SCADA and vibration based condition monitoring systems be used in isolation to diagnose a range of common DFIG faults, and which machine learning techniques and decision metrics provide the most favourable results?"

The contributions of this chapter are as follows:

1. Provide an understanding of the most effective vibration analysis techniques to diagnose bearing faults and other issues associated with the generator and high-speed assembly.

2. Demonstrate how classification algorithms can be used to determine component health, fault type and location.

3. Convey how data selection, feature engineering and model selection can significantly influence results when using normal behaviour models for fault detection using SCADA data

4. Provide insight into the importance of correctly setting thresholds when evaluating the effectiveness of SCADA temperature-based normal behaviour models to detect faults

3.2 Bearing Diagnosis through Vibration Analysis

This chapter will primarily focus on bearing faults associated with the highspeed assembly and generator which, for the type of wind turbine used in this study gave a total of 5 different bearing locations. Three of these were located on the HS shaft of the gearbox, while the final two were located on either side of the generator. For the purposes of this chapter, these were; 1) Upwind HS bearing, 2) Downwind rotor side HS bearing, 3) Downwind gen side HS bearing, 4) Drive end generator bearing and 5) Non-drive end generator bearing. Figure 3.1 shows the HS bearing locations in relation to the generator and gearbox.

3.2.1 Bearing Fault Types

When a fault is present somewhere in the bearing assembly, frequencies will be introduced into the spectrum, which occur due to shock loading at the fault location. This loading will excite high frequency resonance in the surrounding structure that can be picked up by any transducers if positioned closely enough to the source. The characteristic fault frequency will ultimately depend on the fault location, and can be determined from the following equations:

$$BPFI = f\frac{N}{2}\left(1 + \frac{B}{P}cos\theta\right) \tag{3.1}$$

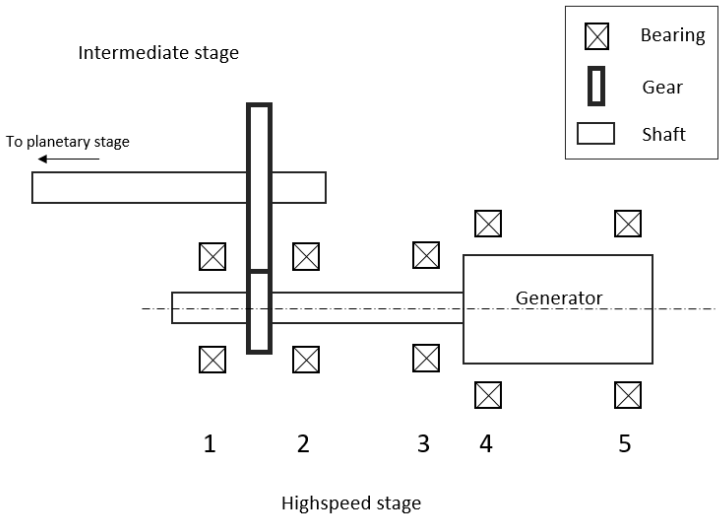


Figure 3.1: Location of bearings on the highspeed assembly where: 1) Upwind HS bearing, 2) Downwind rotor side HS bearing, 3) Downwind gen side HS bearing, 4) Drive end generator bearing and 5) Non-drive end generator bearing.

$$BPFO = f\left(1 - \frac{B}{P}cos\theta\right) \tag{3.2}$$

$$BSF = f\frac{P}{2B}\left(1 + \left(\frac{B}{P}cos\theta\right)^2\right) \tag{3.3}$$

$$FTF = \frac{f}{2}\left(1 + \frac{B}{P}cos\theta\right) \tag{3.4}$$

where BPFI and BPFO are ball passing frequencies at the inner and outer races respectively, BSF is the rolling element ball passing frequency and FTF is the fundamental train, or cage frequency. N is the number of bearings, P is the radius of rotation, B is the bearing diameter and θ represents the bearing contact angle. Depending on the fault type and severity, different combinations of these fault frequencies may be observed in the spectra, along with fundamental shaft frequencies and harmonics.

There are several techniques that can be used to give insight into component vibration by analysing acceleration measurements, mainly surrounding time-domain and frequency-domain analysis. Although more information will be provided in upcoming sections, an interested reader can find more detail on vibration based analysis for

rotating machinery in literature such as [87–90].

3.2.2 Time Domain Analysis

The vibration signal can be analysed in a number of ways, the simplest of which is in the time domain. Basic statistical analysis techniques can provide important information about the signal and although it is not sufficient to actually detect fault frequencies and diagnose faults, it is certainly a useful method in which to detect any obvious irregularities or signal interference.

3.2.3 Frequency Domain Analysis - Fourier Transform

The frequency domain is one of the most commonly used methods to analyse vibration in rotating equipment. Fourier analysis is a technique widely used to convert an input signal in the time domain to an output in the frequency domain using a fast Fourier transform (FFT) algorithm. The FFT algorithm samples a signal over a specific time period and divides it into its frequency components, with each sinusoidal component having a unique frequency with its own amplitude and phase. The two most important factors for consideration when performing any frequency based analysis are the time signal sample rate and sample length, which will influence the range of frequencies that can be analysed in the frequency domain.

It is important to focus on the range of frequencies which are associated with the mechanical rotation of the generator shaft, which will allow any indicators of a fault to be detected. In general, this frequency range will be the mean generator shaft rotational frequency and associated harmonics, along with fundamental bearing fault frequencies described previously.

3.2.4 Order Tracking

During each vibration sample, the generator shaft speed can vary, often significantly, meaning that the signal is not stationary. This can produce a smearing effect on the FFT spectrum which is somewhat proportional to the range of shaft speeds experienced over the sample. The Fourier transform can therefore be adapted for a sliding

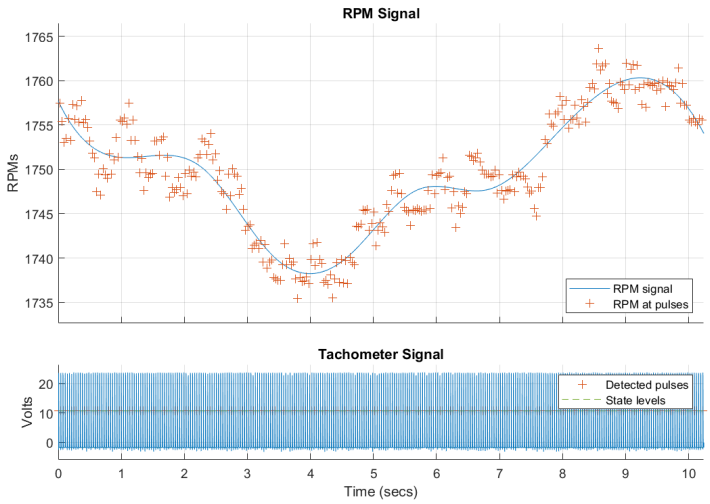


Figure 3.2: Speed signal and tachometer pulses over example vibration sample

time window by using the short-time Fourier transform, where a spectro-temporal representation of the signal is obtained. This is used for order analysis and allows the spectral values to be tracked in time [91,92]. Order analysis can therefore be described as a re-sampling technique which is effective when analysing non-stationary signals. When a signal is re-sampled using this technique it moves from a variable shaft speed and constant sampling rate, to constant shaft speed and variable sample rate. When computing the frequency-RPM map it is important that a sensible resolution bandwidth is chosen to capture all the desired features of the signal therefore losing as little information as possible. Order tracking relies on tachometer/encoder pulses which are synchronised with the accelerometer measurements and can therefore track the shaft speed over the vibration sample. Figure 3.2 shows an example of the variation in shaft speed over the vibration sample along with the associated tachometer pulses.

3.2.5 Envelope Analysis

When a defect is present in the bearing assembly (regardless of fault type) there are two fundamental ways in which to detect the change in vibration. The first is in detecting the impacts directly, as described in Section 3.2.3 with frequency domain analysis looking for specific fault frequencies. The second is through modulation of the fault fre-

quencies at high frequency ranges corresponding to the structure's resonance frequency which acts as a carrier signal. This modulation can be determined through envelope analysis, which removes the high-frequency components and focuses on lower-frequency repetitive behaviour, extracting this information from the signal and representing it on an envelope spectrum. Envelope signals reveals more diagnostic information than the analysis of raw signals because the signal is bandpass filtered in a high frequency band where the fault impulses are amplified by structural resonances [93].

Envelope analysis offers the opportunity to detect bearing faults earlier than otherwise possible using a standard Fourier transform. This is because high noise levels can conceal the harmonics associated with bearing fault frequencies in the standard spectrum. As well as this, heightened random noise can be a sign of bearing damage deterioration, therefore it can sometimes be hard to distinguish fault frequencies. Due to the nature of rotor loading this is certainly the case for wind turbine applications meaning envelope analysis is an extremely powerful tool.

It should be noted that while envelope analysis may be the best technique in diagnosing high speed bearing faults, it cannot be used as the only tool if gauging vibration levels or if an assessment of component condition is required. As bearing damage progresses and the number of defects increases, the envelope signal will not necessarily correlate accordingly. For high levels of bearing damage the standard FFT spectrum still needs to be used to capture changes in amplitude of the fundamental frequencies of vibration.

3.3 Generator Diagnostics through Vibration Analysis

Other common issues considered in this chapter are misalignment between the high-speed shaft and generator, and rotor imbalance within the generator assembly.

3.3.1 Misalignment

Misalignment occurs between the highspeed and generator shaft when the centre-lines of the coupled shafts do not coincide with each other. This can occur in two fundamental

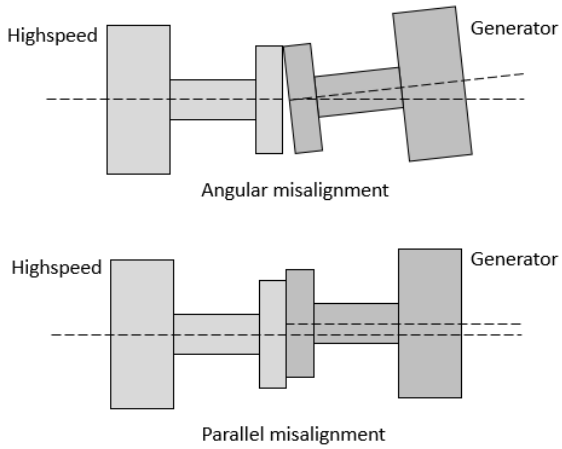


Figure 3.3: Misalignment of coupling between highspeed shaft and generator.

ways causing either angular or parallel misalignment.

Angular misalignment occurs when the shafts meet but are not parallel, as demonstrated in Figure 3.3. From a mechanical loading perspective, this sort of misalignment will produce a bending moment on the shaft, which in turn causes increased vibration at shaft frequency. This can be detected in both the radial and axial direction depending on the angular misalignment.

Parallel misalignment occurs when the shafts are parallel but not coincident, as Figure 3.3 again demonstrates. This form of misalignment causes increased vibration in the radial direction at both 1 and 2 times shaft frequency.

When detecting real life cases of misalignment there is usually a combination of both angular and parallel misalignment, therefore any indicators can be made up of 1 and 2 times shaft speed in both the radial and axial direction.

3.3.2 Rotor Imbalance

Rotor imbalance fundamentally occurs when the centre of mass and axis of rotation do not align. This can occur for a variety of reasons including manufacturing tolerances, mechanical damage, electromagnetic imbalances in the generator etc, however all have the same dominant fault frequency when analysing and detecting a fault of this nature. For clarity Figure 3.4 shows a simple representation of static rotor imbalance, with the

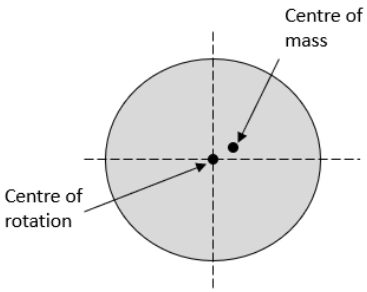


Figure 3.4: Static imbalance in generator.

Table 3.1: Diagnostic Method Summary

Component	Fault Type	Method
Bearing	Outer Race damage	Envelope Analysis
Bearing	Inner Race damage	Envelope Analysis
Bearing	Rolling Element damage	Envelope Analysis
HS Coupling	Misalignment	FFT with Order Tracking
Generator	Rotor Imbalance	FFT with Order Tracking

centre of mass offset from the centre of rotation, which is the simplest form of imbalance found on the generator rotor assembly. Vibration will increase at shaft frequency if rotor imbalance is introduced to the assembly, which can typically be picked up on the radial sensor.

3.3.3 Analysis Techniques

Both misalignment and imbalance are detected using FFT with order analysis as described in Sections 3.2.3 and 3.2.4. Table 3.1 summarises the diagnostic indicators used for the faults detected in this Section.

3.4 Feature Extraction

There are several steps to extract features from vibration data in order to diagnose a fault. First of all the time series signal must be converted to the desired spectrum using the range of techniques in Sections 3.2 and 3.3. The spectrum chosen should

Table 3.2: Parameters to Consider for Feature Extraction

Parameter	**Description**
Sensor channel	Vibration sensor channel to take vibration data - determines the location and direction of installed vibration sensor
Frequency range	Range of frequencies in the computed frequency spectrum
Resolution	Resolution of spectrum determined by frequency range and number of lines on spectrum. Sensor will have limitations on maximum resolution.
Fmin	Minimum required frequency in spectrum or envelope
Fmax	Maximum required frequency in spectrum or envelope
No. of Harmonics	Number of harmonics desired to capture in spectrum
Search Range	Percentage around fault frequency to determine vibration peak or energy

reflect the range of frequencies that will change if a fault is present in the system. Once the spectrum has been computed there are also other parameters which should be considered that can influence the component health indicator calculated for feature classification. Table 3.2 lists key parameters that must be chosen for feature extraction using vibration data in order to calculate a component health diagnostic indicator.

It is worth noting that many of these parameters were already set prior to this research study during installation and commissioning of the CMS hardware and could not be altered. For clarity, the sensor channel could be chosen from several installed sensors (the closest sensor to the known fault location was chosen in each case), while frequency range, resolution, Fmin and Fmax were all pre-determined based on established industry practices on data-acquisition and sensor capabilities. The number of harmonics and search range could be chosen exclusively as long as the value was within the limits set by the other parameters described above. The number of harmonics was chosen in each case to cover all expected changes in vibration, while accounting for other component frequencies that may coincide or interfere with the harmonic ranges. This ultimately depended on the type of turbine, drivetrain and associated kinematics.

3.5 Feature Classification

Once health indicators for each fault type have been calculated through the feature extraction process, data needed to be labelled for classification. For this application, classification can be binary, for example when the component is healthy or unhealthy, or multi class. If a multi class system is chosen, a measure of damage is required in order to differentiate between different stages of fault progression. This could potentially be achieved through endoscope inspections or analysis of different data depending on the fault type (oil debris for example for gearbox health), however it is difficult to distinguish between stages of fault progression and will always ultimately rely on some element of expert opinion or subjectivity. Figure 3.5 highlights the differences between binary and multi-class systems using a typical progression of a bearing health indicator over time. Throughout this chapter, diagnostics are used to indicate component health, however no attempt is made to estimate any remaining useful life or classify damage severity. To do this additional information is required to determine failure or replacement dates, which will be discussed in much more detail throughout Chapter 5.

Classification algorithm types were discussed in depth throughout the literature review, with multiple tested and compared in this chapter. The classifiers chosen for this chapter were Simple Decision Trees, K-nearest Neighbour and Support Vector Machines.

3.6 Case Study

The case study used for this chapter was chosen to best showcase how classification algorithms can be used to detect the range of faults discussed in Section 3.3 with a high degree of accuracy. No information was known about the level of damage at the time of each vibration sample, nor were the turbine faults run to failure. As the components were proactively inspected and exchanged, no information could be inferred in terms of remaining useful life. Once exchanged no inspection report was available to classify the damage at the time of replacement. As this is the case a binary classification system was chosen, as described in Figure 3.5, which ultimately aims to simply determine if

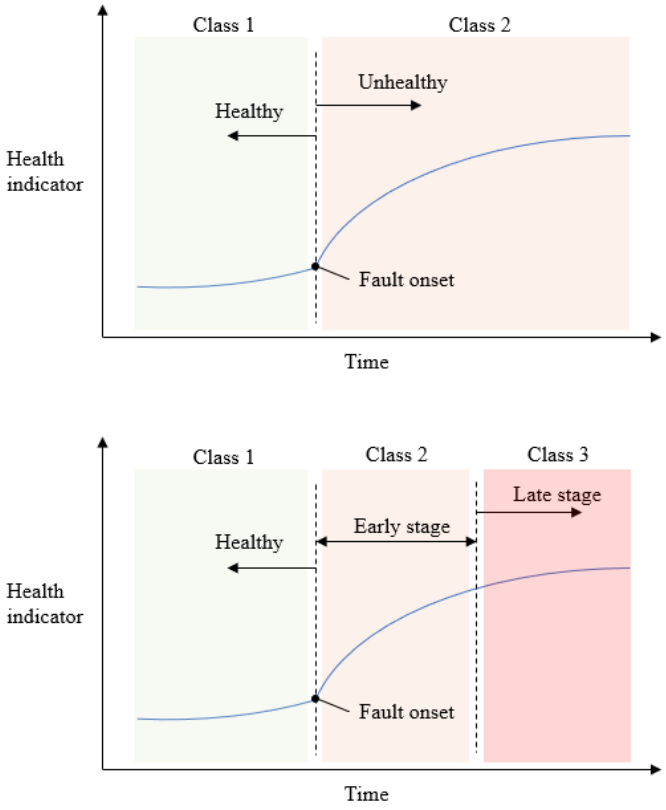


Figure 3.5: Classification: two-class vs three-class.

a component is healthy or unhealthy (specific fault type) with the highest degree of accuracy possible.

3.6.1 Dataset Description

Data was gathered across two different classes, with class A being defined as a healthy component with no fault present operating as expected under normal loading conditions. Class B was defined as the component having a known fault of a particular fault type. Multiple examples of each fault type were found across different turbines on the same wind farm. Vibration samples for each class were taken for each fault type, with features calculated as described in Section 3.4.

Requirements were put in place for data acquisition which limited the operating range at which the vibration samples were taken to a single power bin of greater than

Table 3.3: Data Description

Parameter	**Fault Type 1**	**Fault Type 2**	**Fault Type 3**	**Fault Type 4**
Fault Type	HS Bearing	Gen Bearing	Imbalance	HS Coupling
No. Turbines	3	4	2	3
Power Range	>90% P_r	>90% P_r	>90% P_r	>90% P_r
No. Data Samples:				
Class A	100	130	70	90
Class B	100	130	70	90

90% of rated power. This was done to try and limit the effects of operating conditions on drive train loading which will in turn affect vibration amplitude. Studies have shown that this can have a large impact on classification accuracy, and will be discussed in more depth throughout Chapters 4 and 5 in cases for which using data from a single bin was not possible.

The turbine used for this study was between 1 and 2 MW Rated Power, asynchronous generator with fully rated converter and a 3 stage gearbox. For confidentiality reasons the exact type and model cannot be stated explicitly. Table 3.3 summarises the data used throughout this chapter.

3.6.2 Overall Methodology Framework

The overall methodology consisted of building and training four separate classification models based on fault type. Using the feature extraction process detailed in Section 3.4, diagnostic health indicators were first of all calculated, producing a dataset consisting of a single row of features describing each vibration data sample. Full description of the features used is described in Section 3.6.3. Once the individual datasets for each fault type were calculated, they were then labelled based on the binary classes 'Healthy' or 'Unhealthy' from known fault health states. Classifier models were then trained and validated using cross-fold validation techniques, as described in Section 3.6.4, to assess model accuracy. Once trained, the model could then be used on previously unseen data to make predictions on component health. Figure 3.6 shows a block diagram

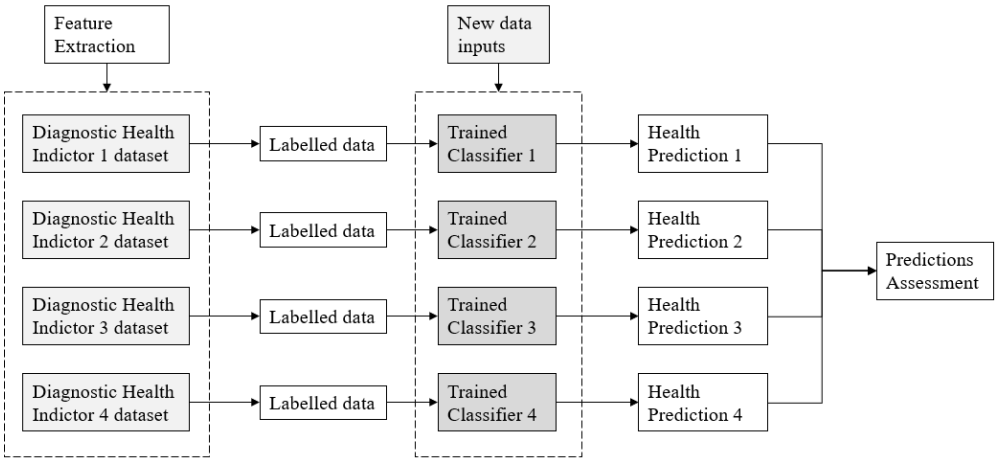


Figure 3.6: Methodology for Component Health Classification.

of methodology used to go from feature extraction on raw training data through to assessing component health predictions.

3.6.3 Model Setup

For this analysis fault types were grouped by component, however this analysis did not distinguish between fault mode. For example, Fault type 1 was considered to be any bearing fault on the HS Gen side bearing detected on the HS generator side sensor. Health indicators were developed to pick up any fault mode within this fault group, and did not distinguish between an inner race fault, outer race fault or fault to the rolling element itself. Damage could be located at 1 or more of these locations with the bearing component still classed as 'Unhealthy', or Class 1 for training purposes, with 'Healthy' being Class 0. Power Output at the time the vibration sample was also taken and used as an indicator, although due to the restrictions in data acquisition this was not a significant factor. Table 3.4 describes the diagnostic health indicators used for model inputs, while Table 3.5 shows an example of the training data structure.

3.6.4 Validation

Validation of each classifier was achieved separately during the training and validation phase using 5-fold cross validation. This was done in order to determine the overall

Table 3.4: Diagnostic Health Indicator Description

Fault Type	Sensor	Indicator	No. Harmonics
1	HS Gen side	BPFO, BPFI, BSF	3
2	Gen DE	BPFO, BPFI, BSF	3
3	Gen DE	HSS Frequency	2
4	HS Gen side	HSS Frequency	0

Table 3.5: Training Data Structure

Sample No.	Diagnostic Indicator	Power Output[a]	Class
1	0.0124	0.96	0
2	0.026	0.91	0
...	...	...	...
100	0.045	0.92	1

[a] Power Normalised based on Rated Power = 1

accuracy of each algorithm. As discussed previously, this method involves partitioning the data into subsets of a predetermined ratio, one of which is then omitted from training and used to test the algorithm. For this example using 5-fold, 20% of the data was used for cross validation purposes. The process is then repeated using different sup-populations and an average accuracy calculated to use as a performance indicator. The prediction process can then be evaluated using a confusion matrix giving correct/incorrect classification and the likelihood of false positives/ negatives. Four common metrics that are used to evaluate model performance are shown in equations 3.5 - 3.8, where TP is the number of true positives and TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.6}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.7}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.8}$$

Table 3.6: Classification Results - Model Accuracy Comparison

Fault Type	**Classifier Type**	**Model Accuracy (%)**
1	Decision Tree	89.5
1	SVM	91.5
1	kNN	91.0
2	Decision Tree	92.7
2	SVM	94.6
2	kNN	94.2
3	Decision Tree	93.6
3	SVM	90.7
3	kNN	92.9
4	Decision Tree	88.3
4	SVM	87.2
4	kNN	86.7

3.7 Case Study Results

First of all results will be presented using absolute values of the diagnostic health indicator using a Confusion matrix to showcase algorithm accuracy and likelihood of false positives and negatives. These results will then be discussed and refined to increase accuracy by identifying key areas of improvement and sources of training confusion. Finally, new data will introduced to the model to generate predictions, in which some of the limitations to fault identification and automatic classification will be presented and discussed.

3.7.1 Initial Results

Several classification algorithms were tested and compared for each fault type, with varied results. Out of the 3 algorithms presented in Table 3.6, accuracy results were quite similar indicating that algorithm type was not a significant factor. In fact there was less than 3% between algorithms across all fault types, averaging just 2.1% difference. For fault types 1 and 2 Support Vector Machines proved to be the most accurate, while for faults 3 and 4 Decision Trees were marginally better. The kNN algorithm was

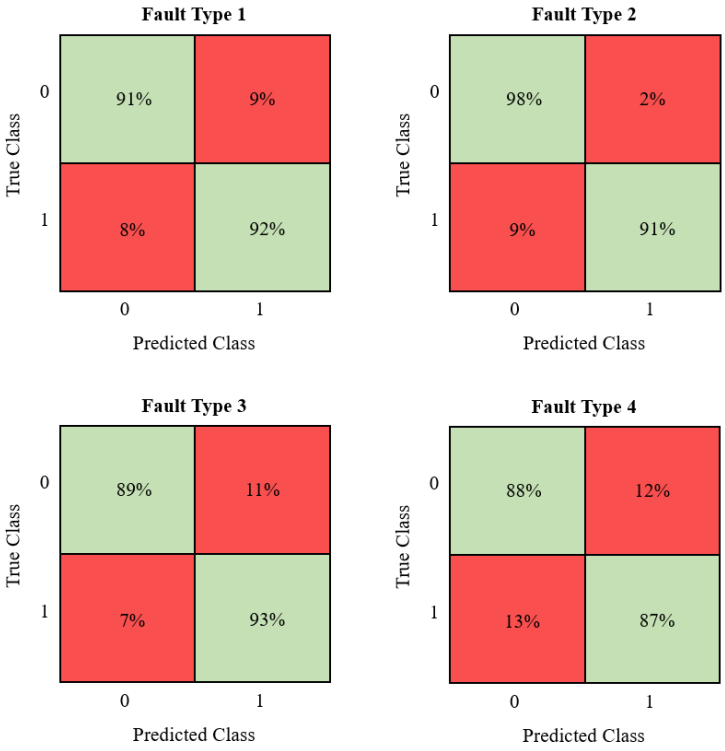


Figure 3.7: Confusion Matrix - Results Set 1.

consistently 2nd place, proving its flexibility for simple applications such as this.

Using the cross validation process described in Section 3.6.4, a confusion matrix was produced for each trained classification model. Figure 3.7 shows the validation and testing results for each of the 4 fault types described in Table 3.3 in the form of a confusion matrix. For each individual Fault Type these matrices show the true class on the y-axis and predicted class on the x-axis, with the average percentage of entries in each class over the cross fold validation sub data sets. Each row of the confusion matrix sums to 100%. When true class is equal to predicted class, this is known as the true positive rate of a particular class, while the false negative rate is described as when the true class does not equal the predicted class. Fault 1 for example could be described as having a 91% true positive rate of predicting when a highspeed bearing is healthy, with a 9% false positive rate. It also has a 92% accuracy when detecting an unhealthy highspeed bearing with an 8% false positive rate. This logic can be used to describe each of the 4 fault types, and will be the chosen metric throughout this chapter when

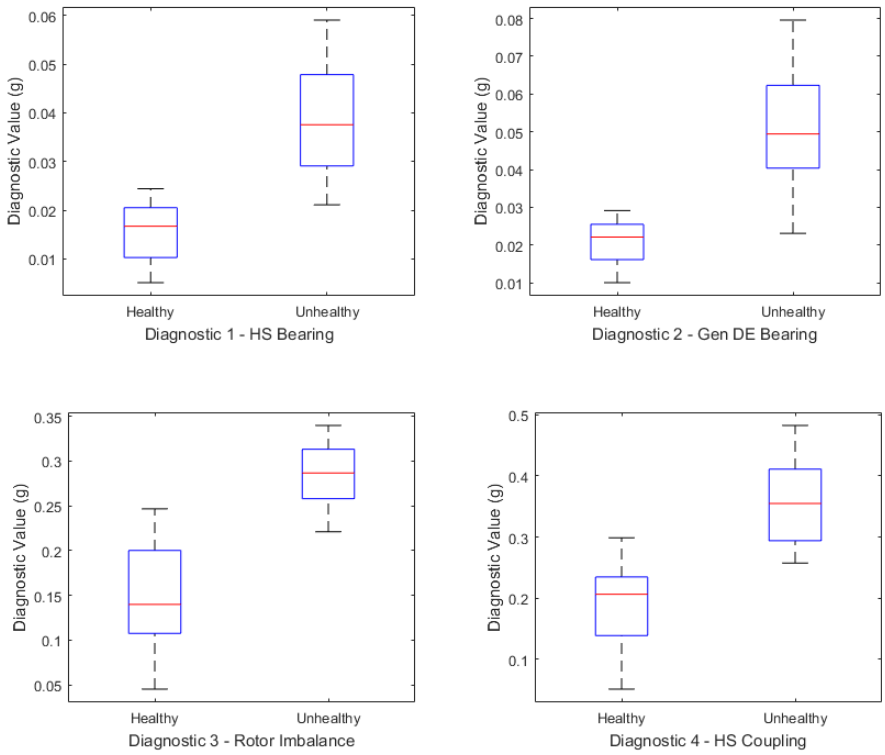


Figure 3.8: Diagnostic Health Indicator Variation.

evaluating classification accuracy.

3.7.2 Analysis of Classifier Results

To understand the the observed spread in algorithm accuracy the health diagnostic indicators can be analysed in further detail. Figure 3.8 shows a boxplot of diagnostic values used as classifier inputs for each fault type, separated into binary health states 'Healthy' and 'Unhealthy'. For each boxplot, the central mark indicates the median, and the top and bottom edges of the box indicate the 75th and 25th percentiles respectively. The whiskers extend to the most extreme data points. What is immediately obvious is the significant overlap between health states in all fault types. Of course, some variation is expected due to the nature of turbine operation conditions, however there are also other factors which could contribute to the disparity. One reason could

be intermittent problems on site during data acquisition that may negatively effect signal quality. Other key reasons could include varying operating conditions at the time of vibration signal (Power Output and Rotor Speed) although this has been accounted for to some extent by limiting data acquisition and using instantaneous Power Output values as a model input. Different wind speeds and loading experienced throughout the drivetrain at Rated Power will also contribute, however without further information on rotor speed and wind speed to better model the operating conditions, the effects of this cannot be considered or analysed. The final key contributing factor is likely to be related to differences present in each wind turbines used to train, test and validate the classifier. To study the effects of individual turbines a baseline was calculated for each diagnostic turbine pair. The relative net amplitude was then determined by:

$$A_{baseline} = \sum_{n=1}^{max} A_n / n_{max} \tag{3.9}$$

$$A_{net} = A_n / A_{baseline} \tag{3.10}$$

where $A_{baseline}$ is the turbine baseline diagnostic amplitude, A_n is the data sample diagnostic amplitude, n_{max} is the number of samples for that turbine fault type pair and A_{net} is the relative net amplitude for each individual data sample.

A plot of this new net amplitude, as shown previously with the absolute value is presented in Figure 3.9. If we first of all take a look at the healthy data across all diagnostics, it can be seen that it is now centred around zero, with amplitudes both positive and negative as the turbine mean is subtracted in each sample. The relative position of the unhealthy samples in comparison to the healthy ones also appears to have increased. Of course there is still some overlap due to aforementioned reasons, however it provides a major step forward in accuracy when transferring knowledge and experience from one turbine to another for diagnostics.

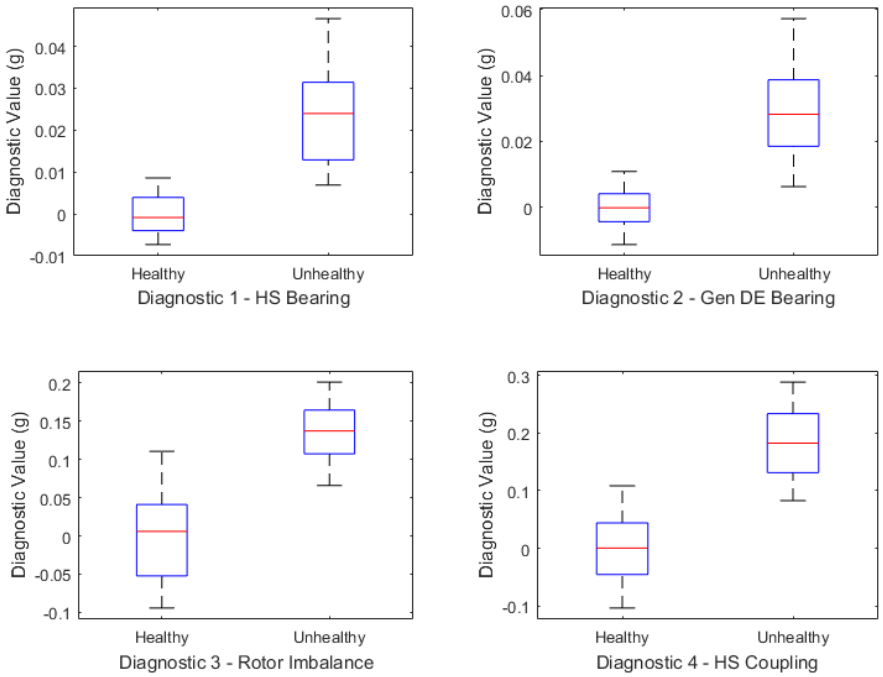


Figure 3.9: Net Diagnostic Health Indicator Variation.

3.7.3 Results including Turbine Baseline

The net diagnostic amplitude for each sample was calculated and used with the normalised power measurement to train the same classifiers for comparison. Results show that by taking into consideration a vibration baseline for each individual turbine accuracy can be increased. Table 3.7 shows the percentage increase for each algorithm across all trained models, while Figure 3.10 shows the new confusion matrices associated with each fault type. All three algorithms had a similar performance when comparing model accuracy, with just 3.6% difference across all classifiers and fault types averaging 95.0%. The kNN algorithm actually performed marginally better than both decision tree and support vector models when considering maximum accuracy, however did prove to be the most inconsistent. Decision tree models were consistently high with a range of only 1.3% while maintaining an average of 94.75%.

Table 3.7: Classification Results - Model Accuracy with Baseline

Fault Type	**Classifier Type**	**Model Accuracy (%)**	**Delta (%)**
1	Decision Tree	94.5	+5.0
1	SVM	95.0	+3.5
1	kNN	96.0	+5.0
2	Decision Tree	94.6	+1.9
2	SVM	96.5	+1.9
2	kNN	96.5	+2.3
3	Decision Tree	94.3	+0.7
3	SVM	93.6	+2.9
3	kNN	92.9	+0.0
4	Decision Tree	95.6	+7.3
4	SVM	95.6	+8.4
4	kNN	95.6	+8.9

3.7.4 Fault Diagnostics Indicator Dependencies

So far throughout this chapter all results have considered classifier performance of each of the four fault types individually. Real time applications of such systems however, would involve assessing all health indicators in parallel as data is recorded through the CMS platform. To replicate this one month of data was taken from both the highspeed shaft sensor and generator drive-end sensor (either side of the generator and highspeed shaft coupling) for one example of each fault type. In each case all four diagnostic health indicators were calculated as previously and passed through the decision tree trained algorithms. A total of 5 test cases were used, one for each fault type plus one test case with no known faults, as described in Table 3.8, where as previously 0 is a healthy component state and 1 is an unhealthy component state.

Ideally results would show a high percentage of correct predictions across all indicators. For example when considering fault type 1, diagnostic health indicator 1 would predict an unhealthy fault class, while the remaining 3 indicators would predict a healthy class. This would mean that each indicator is insensitive to other faults on the drivetrain, and will not produce a high number false alarms.

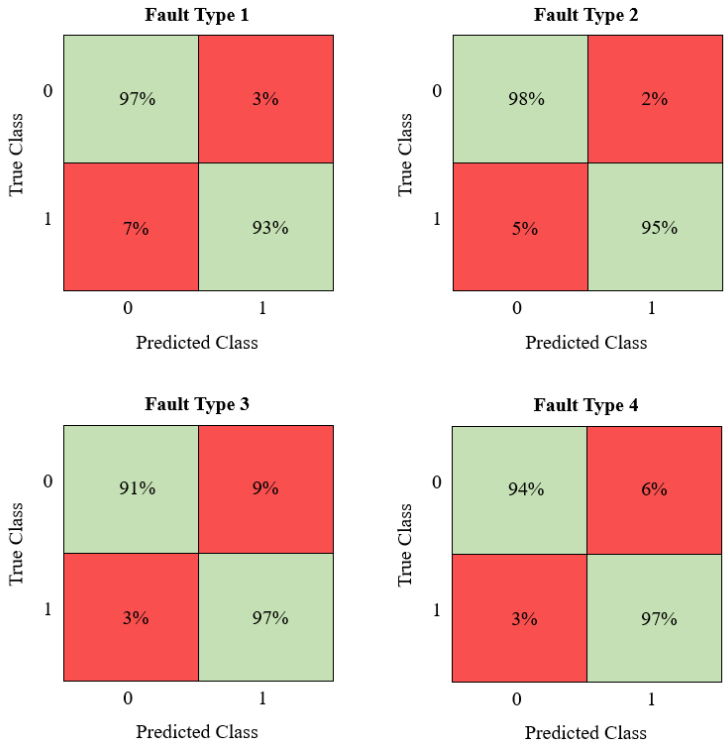


Figure 3.10: Confusion Matrix - Results Set 2.

Figure 3.11 shows results for the 5 test cases described in Table 3.8. These are shown through a stacked bar plot, with one bar per diagnostic for each test case. Healthy predictions are shown in yellow, while unhealthy predictions are shown in purple. Test case 1 had no known faults, with all 4 trained classifiers having an overall accuracy of 95.8%, and only 4 out of the 96 samples producing a false positive results. Similar levels of accuracy were achieved in the second (highspeed bearing) and third (generator bearing) test cases, with an overall accuracy of 95.5% and 97.15% respectively.

Fault types 3 and 4 produced varying results, which delivers key insight into how issues with similar fault frequencies and load paths can manifest into false alarms when looking across the drivetrain in real time. Considering only fault types 3 and 4 now, which were imbalance and rotor coupling damage respectively, we can see that in either case produces a high level of false positive results. This could be down to two main reasons, the first is that the fault has been classified as a single fault type in the operations and maintenance log, however there are several real issues being detected.

Table 3.8: Test setup demonstrating real time application

Test Case No.	**Diagnostic True State** $d_1 - d_2 - d_3 - d_4$	**No. Samples per Diagnostic**
1	0-0-0-0	24
2	1-0-0-0	28
3	0-1-0-0	26
4	0-0-1-0	23
5	0-0-0-1	26

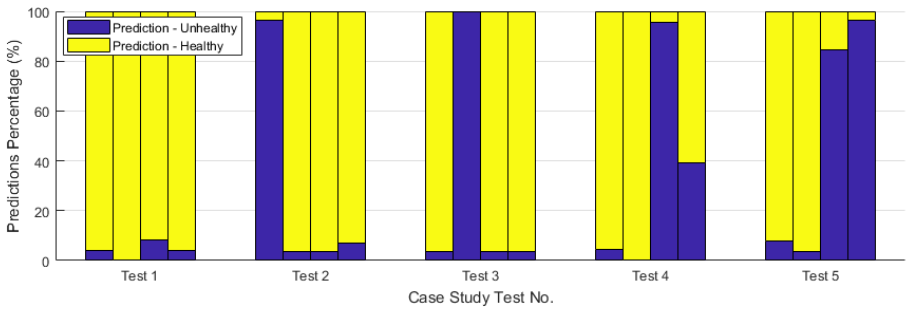


Figure 3.11: Prediction results for each diagnostic for test cases

The second explanation is due to the crossover of fault frequencies and proximity of sensors. Imbalance and rotor coupling damage show up in the fundamental shaft speed and for the former, some harmonics of the fundamental shaft speed. The highspeed and generator shafts, connected through the coupling, are rotating at the same speed. Diagnostics were calculated from the closest sensor in each case, however in reality due to the load path of the faults, either sensor may pick up changes in vibration due to either fault. This is the most likely explanation for the high levels of false positives observed in either case. Taking this into consideration, algorithm performance for these cases were still high in line with previous results for detecting some issue is present. In order to distinguish between these similar fault types conclusively turbine inspection may be required.

3.8 Diagnosing Problems with SCADA Data

The use of SCADA data for anomaly detection in wind turbines has been widely researched and published, with literature covering a range of machine learning and statistical based techniques to detect faults and under-performance. Similar to vibration analysis, most of the SCADA based fault detection has focused on gearbox oil and temperature modelling. This section will apply some of these established techniques to wind turbine generator faults, to both showcase their effectiveness in comparison to vibration analysis and to lay foundations for building on these techniques in later chapters.

3.8.1 Data Pre-processing

It is important to adopt a consistent approach to data pre-processing in order to fairly and objectively compare anomaly detection model results. The ultimate goal of pre-processsing is to be left only with data which describes normal operating conditions of the WT, filtering out any data in which the WT is operating outside normal operating parameters and hence does not fit the expected relationships. First of all, data points using basic logic were removed based on the known operating strategy and control system of the WT. More specifically, this involved removing all data from when the wind turbine had been shut-down during any periods of planned maintenance, when the wind speed was below the required cut-in wind speed of that particular turbine model (typically around 4m/s), or when the wind speed was above cut-out wind speed, meaning that the wind turbine was actively shut down or idling.

The most challenging operating condition to account for is if a WT undergoes periods of curtailment, which is when the generator power output is actively decreased from what it could produce given the available incoming wind speed. This could be done for a variety of reasons including noise reduction, reducing loads due to known faults or lower total wind farm power output due to grid restrictions. If blade pitch angle is known in relation to wind speed and power output this can be made a relatively simple exercise in data filtering, however in cases such as this when blade pitch angle is

Table 3.9: Data pre-processing summary

Criteria	Description	Method
1	WT shut-down	$P_{out} < 0$
2	Wind speed below cut-in speed	$U_{wind} < U_{cut-in}$
3	Wind speed above cut-out speed	$U_{wind} < U_{cut-out}$
4	Wind speed curtailment	Euclidean distance

not available it becomes a more complicated affair. Many methods have been proposed to get around this based on multivariate outlier detection. Since all of the examples used in this study have very little issues with curtailed periods, the analysis simply uses the Euclidean distance from each data vector mean to establish outliers. It should be noted that this part of the pre-processing was only done on the training dataset, therefore outliers due to curtailment could still be detected as false flags while testing the anomaly detection model with new data.

Finally, due to confidentially agreements between data owners and researchers, all power and torque measurements throughout this report have been normalised. A summary of the pre-processing criteria and methods used can be found in Table 3.9, while an example of a power curve before and after pre-processing can be found in Figure 3.12.

3.8.2 A Note on Trending

Trending offers a simple approach to monitor SCADA parameters and their relationships over time. WT's by design operate under constantly changing conditions, reacting to short-term stochastic load cycles caused by varying wind speed. As discussed throughout the literature review, this means that many relationships monitored through the SCADA system can be considered highly non linear, changing with operating conditions and control strategies. Looking at longer time frames WT's also react to diurnal and seasonal temperature cycles, which is especially important when considering temperature distributions throughout the nacelle. Developing simple trends under such variable conditions is therefore extremely difficult, and just because there is a long or short term change in a single parameter or relationship does not necessarily mean a

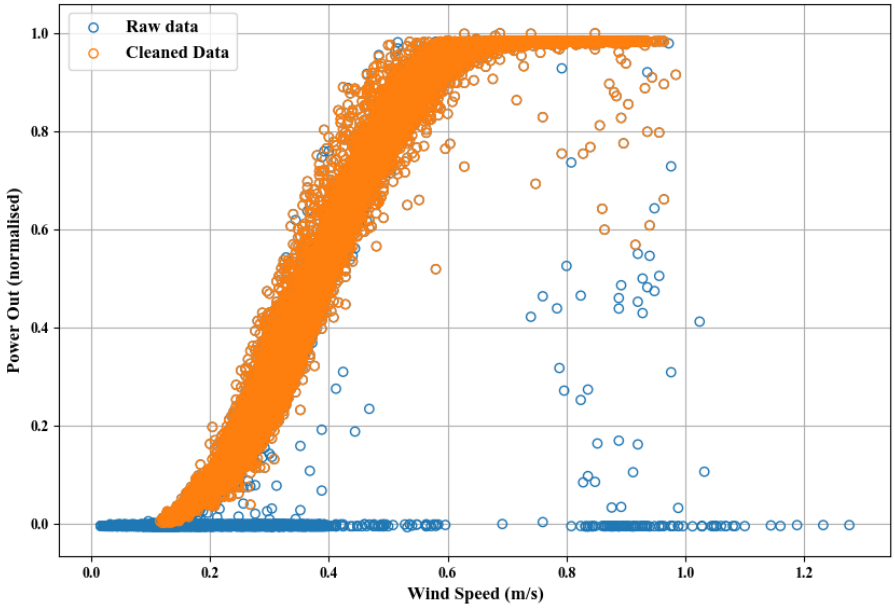


Figure 3.12: Data cleaning

fault exists within the system. Trending methods have however been proven to detect long term degradation of drive-train components and also allow for a baseline to be established in which to compare other more complex methodologies. Trending for condition monitoring purposes involves developing a relationship between parameters over a training period which can describe the normal operating of a particular component, for example generator bearing temperature and power output. This relationship can then be monitored and changes in the relationship evaluated through time. Thresholds can then be set to establish how far the trend can deviate from the original before it is considered anomalous. Although a useful starting point, trending results will not be presented in this chapter due to the amount of literature already existing in this area, with a comprehensive review found in [6].

3.9 Normal Behaviour Model

As discussed in the literature review, NBM's have become a focal point in both academic research and industry with regards to SCADA based wind turbine fault detection. As alluded to in previous sub section, simple trending is typically not suitable to de-

tect faults within the generator and highspeed assembly as any relationships between parameters are multivariate and highly non-linear. To create higher dimensional relationships requires more complex modelling techniques, which can be achieved through a range of regressive based machine learning models.

3.9.1 Model Types

As highlighted in the literature review, regressive based NBM's can fundamentally be either parametric or non-parametric based. The type of NBM considered throughout this section is the Random Forest model, a non-parametric model geared for handling large datasets of high dimensional nonlinear relationships.

3.9.2 Modelling Process

Although different machine learning algorithms could be utilised, the overall model framework remained the same, as shown in Figure 3.13. Historical SCADA data which can adequately describe normal operating conditions without any faults is first required to train the NBM. The underlying principle of this approach is that once a fault is introduced in to the system, the temperature diagnostic which describes a particular fault or component health, will significantly deviate from expected behaviour. This deviation will manifest when analysing the model error.

The NBM works by empirically modelling the diagnostic based on a set of input parameters. The process is summarised in more detail in Figure 3.14, where $u(t)$ are the input variables at timestep t, $\hat{G}(t)$ represents the data-driven NBM to predict target variable, $\hat{y}(t)$ while $G(t)$ constitutes the process of obtaining the measured target variable $y(t)$ through the required in field sensor. Finally, $e(t)$ represents the error between the predicted and measured value.

The threshold set to distinguish normal behaviour from an anomaly is determined from the training phase, in which statistical analysis can be used to calculate upper and lower boundaries.

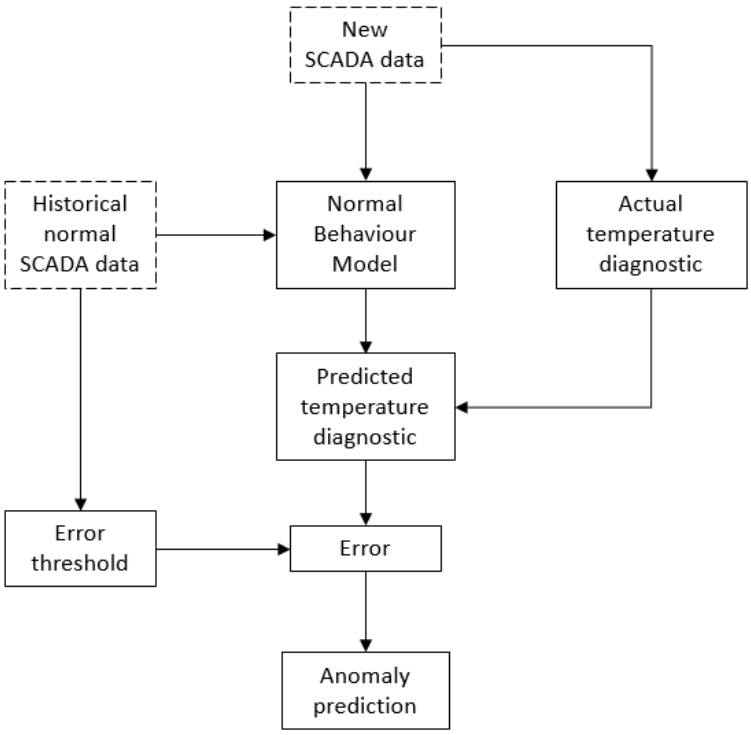


Figure 3.13: NBM anomaly detection process.

3.9.3 Feature Selection

Feature selection is an important aspect to consider when building an anomaly detection or normal behaviour model, especially when working with high dimensional datasets such as SCADA. A number of techniques were used to assess feature importance throughout this thesis beyond expert system operational knowledge, which in itself is a very useful approach. Other techniques included univariate analysis, for which we examine each feature individually to determine the strength of the relationship with the response variable. This can be useful to disregard irrelevant features which are noisy or uncorrelated, however the metrics used are typically quite restricting and not applicable for ranking high dimensional non-linear relationships. For example, one common metric to understand the relationship is the Pearson correlation coefficient, which measures the linear correlation between two variables. This could produce a low score based on linear correlation but is actually very correlated though a higher order non-linear relationship.

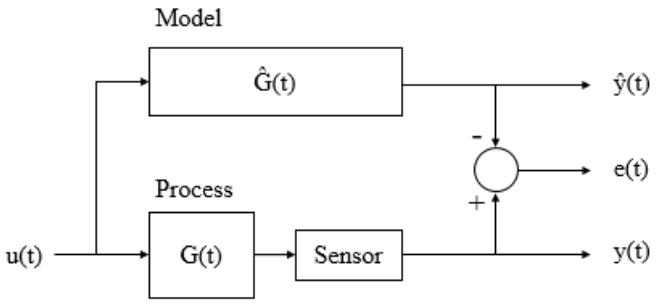


Figure 3.14: Model based monitoring based on NBM. Adapted from [5], [6]

To navigate this issue another useful way to rank features is through a model based approach. For ease of application a tree based model was most often utilised throughout this thesis and involves training a random forest model and assessing feature importance based on the average decrease in impurity across all decision trees in the forest. As this is automatically computed during the training phase, it offers a convenient approach to assessing feature importance for multiple reasons. Crucially, as these models are logic based, they do not require input normalisation and can deal with non-linear relationships. One drawback can be overfitting therefore the depth of the tree should be limited and cross-validation should be applied to ensure generalisation.

3.9.4 Training and Validation

As previously discussed, validation of all models was achieved through cross-validation. The number of folds used was dependant on the data used for training. Figure 3.15 shows the breakdown of training, validation and testing with regards to the data leading up to component replacement. The data is initially divided into a training and testing dataset. The initial testing dataset is used for testing how the trained algorithm behaves detecting alarms leading up to component replacement. The initial training dataset is used to validate and test the algorithm inputs and tune hyperparameters. As clarified in Figure 3.15, this training dataset is further reduced down into training and test subsets for performing cross-validation. In this example 10-fold cross-validation is utilised with 75% of initial training data.

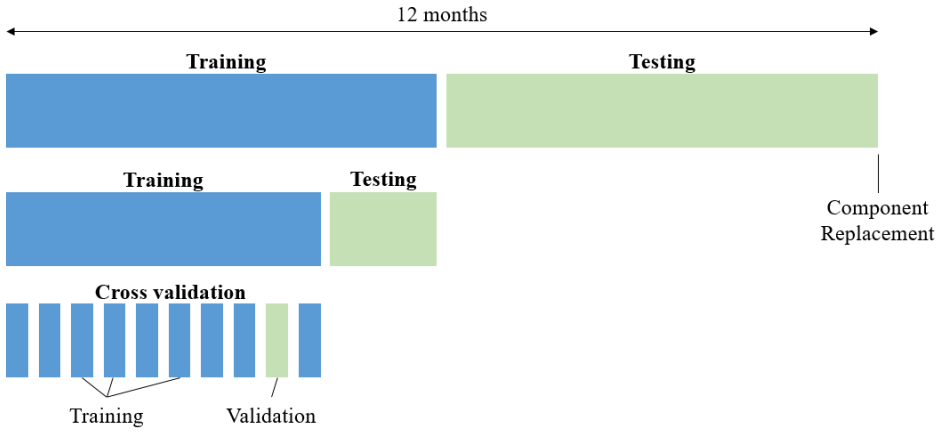


Figure 3.15: Breakdown of dataset for model testing and validation.

3.10 Case Study - Generator Bearing Fault Detection

This case study will present results from 4 generator bearing replacements, each utilising the same temperature NBM to detect anomalous behaviour and generate alarms. Each case study will present results utilising two different ratios to initially split the data into training to testing sets. Data utilised for each turbine can be found in Table 3.10, which describes the relationship and breakdown of raw data to data used for testing and training. Each raw data sample represents all SCADA channels at an individual timestamp, while cleaned data represented data available after pre-processing as described in Section 3.8.1.

3.10.1 Data Description, Visualisation and Feature Analysis

One year of SCADA data was gathered for each turbine prior to generator bearing replacement. A full description of the available channels can be found in Table 3.11, which included 10 minute measurements of power output, rotor and generator speeds, wind speed, ambient temperature and in the nacelle, as well as temperatures of key generator components including the bearing, slip ring and windings for phases 1, 2 and 3.

In order to make informed decisions surrounding feature selection SCADA variables can first be visualised and correlations established. Figure 3.16 shows these relation-

Table 3.10: Analysis cases

Case Study	Wind Turbine	Train:Test ratio	Raw data	Cleaned data	Data (Train:Test)
1	1	1:1	52704	44609	22304:22304
2	1	1:2	52704	44609	14870:29738
3	2	1:1	40096	23824	11192:11192
4	2	1:2	40096	23824	7941:15882
5	3	1:1	52233	36248	18123:18124
6	3	1:2	52233	36248	12083:24164
7	4	1:1	52630	42550	21274:21274
8	4	1:2	52630	42550	14183:28366

ships between variables. Figure 3.16 (a) shows the generator bearing temperature in relation to operational parameters, which includes the power output, wind speed, rotor speed and generator speed. Figure 3.16 (b) shows the temperature distributions associated with the generator and surroundings, which includes bearings, slip ring, phase windings (A, B and C) components, along with average nacelle and ambient temperature. Note that the data shown in both of the figures described above represents the raw data before cleaning and pre-processing. A scatter plot is shown for each pair of SCADA variables along with a histogram describing the distribution divided into 10 equal bins. These plots highlight the variability of generator bearing temperature over each of the operational parameters. Looking more closely at the correlation, as previously stated we can calculate the linear dependence between each pair of features using the Pearson correlation coefficient, r:

$$r = \frac{\sum_{i=1}^{n}[(x^i - \mu_x)(x^i - \mu_x)]}{\sqrt{\sum_{i=1}^{n}(x^i - \mu_x)^2}\sqrt{\sum_{i=1}^{n}(y^i - \mu_y)^2}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{3.11}$$

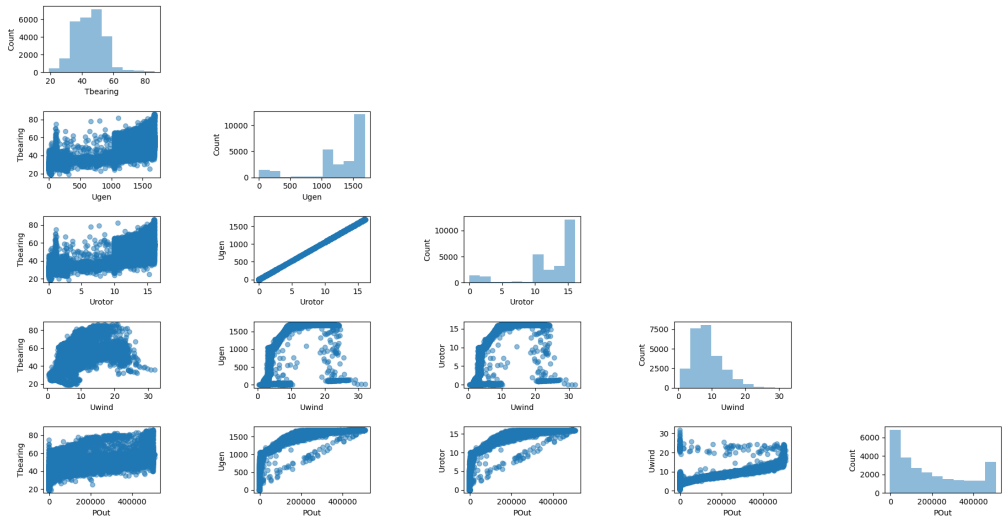
where μ denotes the mean of the corresponding variable, σ_x and σ_y are the standard deviations of feature x and y, while σ_{xy} is the covariance between features. The Pearson correlation can be calculated as the covariance between two features divided by the product of their standard deviation as shown above. Figure 3.17 shows the Pearson

Table 3.11: Available SCADA channels

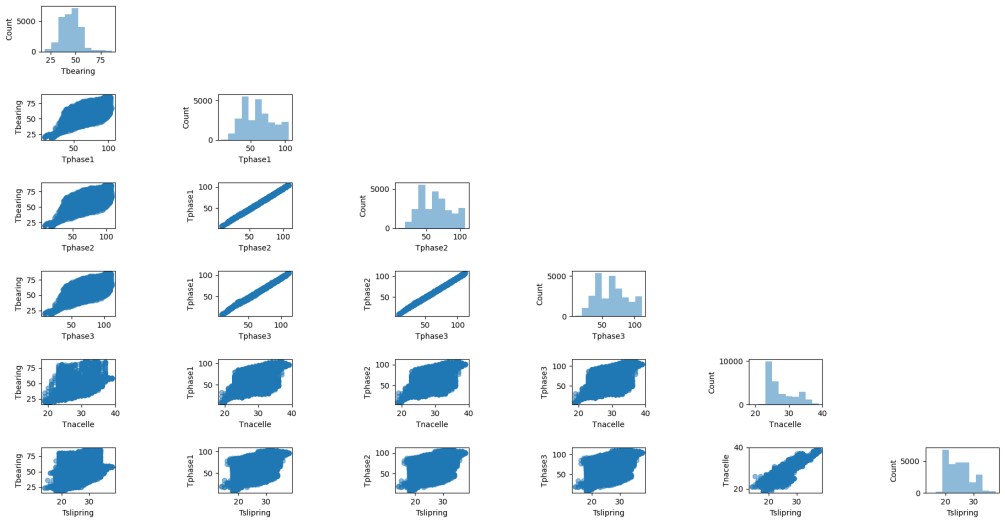
Channel No.	Feature	Description
1	P_{out}	Average power output
2	U_{rotor}	Average rotor speed
3	U_{gen}	Average generator speed
4	U_{wind}	Average wind speed
5	$T_{ambient}$	Average ambient temperature
6	$T_{nacelle}$	Average nacelle temperature
7	$T_{slipring}$	Average generator slip ring temperature
8	T_{phase1}	Average generator phase 1 temperature
9	T_{phase2}	Average generator phase 2 temperature
10	T_{phase3}	Average generator phase 3 temperature
11	$T_{bearing}$	Average bearing temperature over

correlation coefficient for each pair of features described in Table 3.11. The r value has a range of between -1 and 1, with two feature showing total positive correlation if r=1 and true negative correlation when r=-1. Two features will be completely uncorrelated if r=0.

Next to the correlation heatmap in Figure 3.17 is a barplot highlighting feature importance as calculated by the random forest regressive model, as described in section 3.9.3. Pairs of features for which r=1/-1 (such as between rotor speed and generator speed) were removed from this analysis as only one of the pair is required to describe the relationship, with the other being redundant. Analysis shows that generator speed is the most important feature to predict generator bearing temperature, which is followed by wind speed and power output. Other temperatures such as generator phase temperature and slip ring temperature have less impact on the model prediction. These two contrasting plots highlight the difficulties when assessing the correlation between SCADA variables. Feature analysis requires careful consideration to capture complex relationships between variables that cannot be represented by r values alone.



(a) Operational parameter relationships & distribution



(b) Temperature relationships & distribution

Figure 3.16: Visualisation of SCADA relationships to target variable.

3.10.2 Model Setup Detail

The model used for this case study was a random forest regressive model, with the model predictors and response features highlighted in Table 3.12 for clarity. Features were chosen based on the analysis described in the previous section. Model hyperparameters were determined through model testing and validation processes. Key hyperparameter values included; max tree depth of 4, number of tree estimators of 10, minimum samples

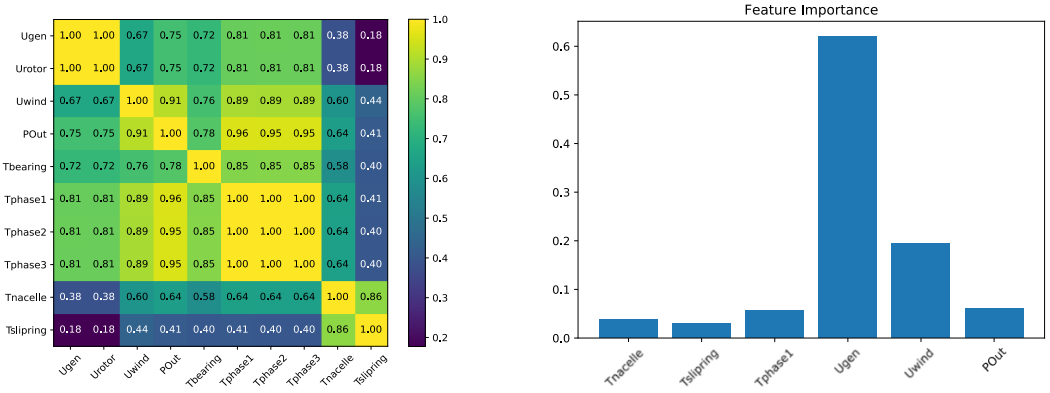


Figure 3.17: Correlation heatmap (left) and Feature Importance (right) of selected SCADA variables.

Table 3.12: SCADA features used in model

Feature No.	**Feature**	**Model Training**
1	P_{out}	Predictor
2	U_{gen}	Predictor
3	U_{wind}	Predictor
4	$T_{nacelle}$	Predictor
5	$T_{slipring}$	Predictor
6	T_{phase1}	Predictor
7	$T_{bearing}$	Response

per split of 2 with the quality of split measured by the mean squared error (MSE).

3.10.3 Model Testing & Validation

Cross validation scores across all 8 case studies (described in Table 3.10) can be found in Figure 3.18, for which the same pre-processing and features engineering pipeline was applied. In each case the model was setup as detailed in the previous section, with 6 features to predict generator bearing temperature. For each case study the average coefficient of determination (R-squared score) of the 10-fold cross validation is plotted along with the maximum and minimum score. The R-squared score is closely related to the variance and MSE and can be described as the the proportion of the variance in

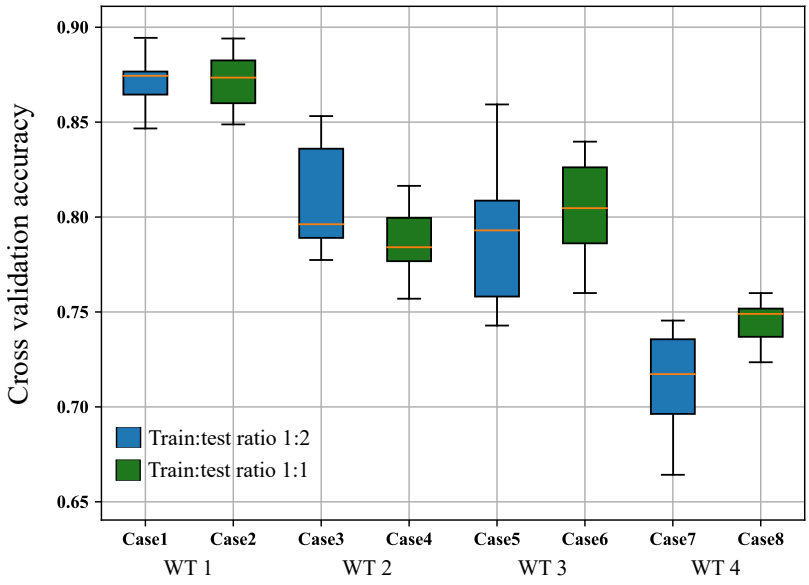


Figure 3.18: Cross validation error.

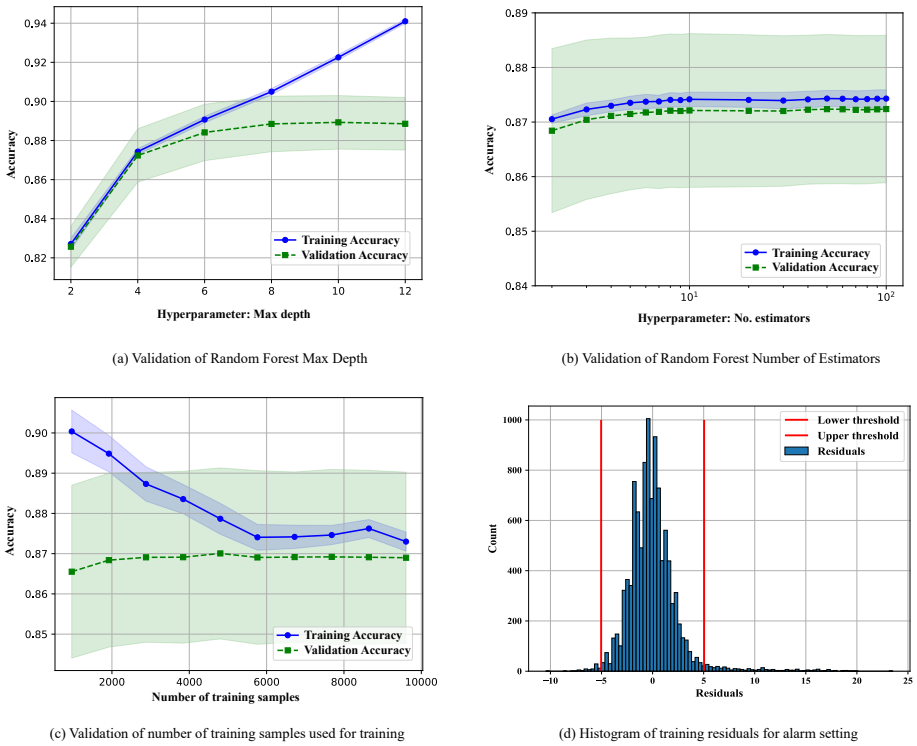
the dependent variable that is predictable from the independent variable(s):

$$MSE = \frac{1}{N}(y_i - x_i)^2 \tag{3.12}$$

$$R^2 = \frac{SS_{reg}}{SS_{tot}} \tag{3.13}$$

where SS_{tot} refers to the total sum of squares (proportional to the variance) and SS_{reg} refers to the sum of squares of residuals. An ideal fit would give an r-squared score of 1. There is a clear disparity in accuracy between wind turbine fault cases, ranging from close to 0.88 for wind turbine 1 to 0.72 for wind turbine 4. This showcases the variability of predictability of bearing temperature between turbines highlighting the difficulty in producing one technique or model that is applicable to SCADA data across a site.

Now considering different train:test ratios we can analyse the effect of the number of training samples has on accuracy across each turbine. The key observation is that the mean r-squared score does not change significantly in each case, however the variation



(a) Validation of Random Forest Max Depth

(b) Validation of Random Forest Number of Estimators

(c) Validation of number of training samples used for training

(d) Histogram of training residuals for alarm setting

Figure 3.19: Model validation.

is consistently larger when a smaller training set is used. This can be explored further through a training curve, shown in Figure 3.19 (c) for case study 1, which highlights the problem with overfitting when the number of training samples is reduced. This pattern was consistent when looking at all case studies. Figure 3.19 plot (a) and (b) show validation curves of key model hyperparameters used for training. These curves formed the bases of tuning and choosing the values described in the previous section, and show the importance, particularly of tree depth, of correctly validating and optimising the chosen model.

Finally, Figure 3.19 plot (d) shows the error residuals of the training period, which are not only used to assess model accuracy, but to set alarm limits that can detect anomalous behaviour. Here the upper and lower limits are set to the mean +/- two times the standard deviation.

3.10.4 Results

Results will be presented for each case study considering the number of alarms raised each month based on a consistent set of rules. The upper alarm threshold was determined by the standard deviation of the training residuals as described above. As we move through time, to limit false positives we don't want an alarm to be raised at every point above the threshold due to expected variability in the data. It is consistent anomalous behaviour that is required to be detected. To achieve this there are two main approaches that will be used throughout this work when dealing with SCADA data; rolling average and hysteresis. The rolling average simply takes the rolling average of the data, which can be weighted towards most recent data, based on a chosen set number of points. Alternatively, a hysteresis will require a set number of points to be continuously above a threshold. For example, if the applied hysteresis is 5 and 3 points in row cross the threshold followed by one below, the hysteresis value would reset to zero.

The results presented in Figure 3.20 have a hysteresis of 12 (two hours) applied, with an alarm threshold calculated as the mean plus two standard deviations. Firstly, results show a clear increase in the number of alarms for each case study, which is particularly true for the final 2-3 months before replacement. Results also show a similar pattern of alarms for each of the train:test ratios in all cases, suggesting that 3 months of SCADA data, for these test fault cases is seemingly enough. This reiterates what is observed by the learning curve shown in Figure 3.19 (c), which shows a levelling off of accuracy after approximately 6000 data samples. This gives confidence that the model is picking up true anomalies rather than a problem associated with overfitting.

3.11 Conclusions & Discussion

This chapter has shown real world applications of how classification algorithms can be used to detect and classify fault types across the generator and highspeed assembly. To achieve this, the most effective vibration analysis techniques have been demonstrated to create effective diagnostic health indicators for a particular fault by extracting am-

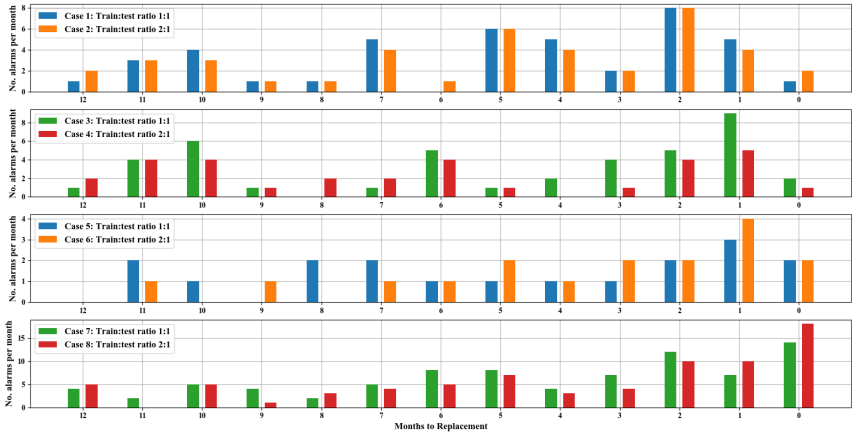


Figure 3.20: Monthly alarms.

plitudes at key frequencies. These diagnostics were then fed through a classification algorithm, of which several were compared and tested, to detect and classify the fault.

Results show that by using a binary classification of 'healthy' and 'unhealthy' components, several different faults could be classified with an average accuracy of over 95%. This level of accuracy is achieved by using data from the closest sensor when considering the load path of a particular fault type in isolation. When comparing performance between the three chosen algorithms (Decision Tree, kNN and SVM), decision trees were the most consistent across all fault types, with difference in accuracy of only 1.3% while maintaining an average of 94.75%. When considering multiple health indicators across the drivetrain simultaneously, this chapter has also shown that interdependencies between fault frequencies and sensors do exist. When this occurs it becomes more difficult to determine the fault type and location conclusively, however still provides evidence of a fault within a particular assembly, which can then be further investigated through inspection or additional analysis. This chapter has provided evidence of interdependencies across generator and coupling issues when making a diagnosis, however, bearing faults across the highspeed and generator could be diagnosed independently. This is explained due to different bearing types having distinct fault frequencies producing accurate health indicators. For cases in which bearing types or fault frequencies

were a closer match, it would be harder to isolate the fault location producing higher levels of dependency between diagnostics. Finally with regards to vibration diagnostics, this chapter has helped shine some light on some of the difficulties when training an algorithm for fault detection across similar turbines of different loading and vibration levels. A simple turbine baselining method has been described to help negate these effects, which has been shown to increase algorithm accuracy by up to 8.9%.

In terms of SCADA temperature analysis of generator bearings, it has been shown that normal behaviour modelling can be a powerful technique when it comes to detecting changes in operating behaviour, which has already been widely shown in literature across other components of the drivetrain. Due to the complex nature of the relationship between operating conditions, loading and temperatures, more advanced machine learning techniques that can handle these non-linear high dimensional relationships are more suitable. Results show that Random Forest models can model these relationships, with generator rotor speed being by far the most important variable to predict and model generator bearing temperature. All four fault cases showed a significant increase in alarms 2-3 months prior to component replacement, at which point the turbine was shut down. It is not known if the turbine in question was shut down for scheduled repair, breached an upper warning limit initiating an unplanned shut down or due to component failure.

In a recent paper published by McKinnon, Turnbull et.al, [20], similar case studies were used to study the effects of time history on model accuracy in greater detail. These findings are worth discussing in the context of the results presented in this chapter. In [20] the effects of the training time period on model performance are explored by considering models trained with both 12 and 6 months of data, with 12 months of training providing the best performance over the two case studies presented. These findings differ from results presented in this chapter, where a sensitivity study on training samples showed limited accuracy model increase beyond 6000 training samples, which equates to approximately 3 months of SCADA data. This may indicate a limiting factor on autoregressive models with exogenous inputs (ARX), for which it was proven provide more accurate modelling approach when compared to full signal reconstruction

(FSRC) models, however this may be at the cost of requiring more training data. It also suggests a variability in required training periods for turbines in different geographical conditions where seasonality effects may be stronger. Further work is required to understand the true underlying reason for this observed variation before any attempt to standardise an approach can be made. For now each optimised trained model remains case specific.

3.12 Future Work

3.12.1 Vibration Diagnostics

Future work could be completed to further investigate dependencies of diagnostics across multiple drivetrain components and assemblies, an area which could help decrease uncertainty of diagnostics. This would reduce time and effort required to inspect and confirm any diagnosis before potential replacements, particularly useful in an offshore environment when technician time up a turbine is at a premium.

Another key area of improvement would be related to turbine environmental loading and scaling. This chapter has shown that, even at similar turbine operating conditions, vibration differs between identical turbine types. In order to train accurate classifiers to detect faults across different turbines, or indeed different operating conditions over different sites, more information may be required to assist with classification, increasing model dimensionality, which in turn requires more thoughtful feature engineering and model selection.

3.12.2 SCADA Temperature Diagnostics

Normal behaviour modelling for specific components is now well understood and has been shown to detect a range of faults by modelling key operational attributes for each turbine. Optimising turbine level models is a difficult task and requires understanding of individual turbine behaviour, loading, expected noise and effects surrounding seasonality. There is still a lack of understanding to what extent models can be transferred between components and turbines of the same type on the same wind farm, as well

as how well models and thresholds scale with turbine size. Although this chapter has helped shine some light on these issues for a particular turbine model and fault mode, access to better quality data across different models and sites (with varying conditions) is required to make breakthroughs with regards to a sustainable and automated approach.

To what extent any alarms produced by NBM's can be used to actually diagnose an unknown issue for a turbine in operation is also up for debate. SCADA analysis is more likely to be useful in practice in an attempt to lower false alarm rates and give more diagnostic certainty in conjunction with vibration analysis, which will be discussed throughout the next chapter.

Chapter 4

Combining SCADA and Vibration Data into a Single Anomaly Detection Model

4.1 Chapter Contribution

As alluded to in previous chapters, operators can now rely on multiple sources of data to help effectively monitor wind turbine condition, to make informed operational decisions which can minimise downtime, increasing availability and profitability of any given site. Two such approaches are SCADA temperature and vibration monitoring, which have both been discussed in detail throughout Chapter 3, and are typically performed in isolation and compared over time for both fault diagnostics and in some cases prognostics. This requires analysis and interpretation of two individual models, perhaps even by two separate specialist engineering teams, looking at SCADA data and vibration data respectively. Data sources may not be stored in the same location, or indeed analysed by the same engineer, team or even company for a particular site depending on the asset management and O&M contracts in place.

That being said, as the wind industry continues to make strides towards improved digitalisation, new sites are being installed with greater importance placed on good data management to provide a platform for analysis. This allows different data sources to

be more easily accessed at an improved cost. Combining data sources where applicable would allow for one single governing component health status to be established which can detect deviations in both vibration and temperature.

This chapter aims to answer the following research question:

"What are the benefits of having both SCADA and vibration data, and can they effectively be combined into a single anomaly detection model for fault diagnostics?"

The contributions of this chapter are as follows:

1. Highlight the key difficulties, challenges and opportunities of combining data sources with respect to WT condition monitoring
2. Provide a framework in which to combine different data sources into a single anomaly detection model
3. Demonstrate an automated approach to effectively establish a single decision boundary to detect an anomaly based on multiple normal behaviour models
4. Give insight into the benefits and drawbacks of combining data sources and analysis techniques.

4.2 Understanding the Problem

Before describing the modelling framework it is first important to understand the key problems and motivations associated with combining multiple models and data sources into a single anomaly detection model. While this chapter will focus on vibration and temperature models, it is equally applicable to other condition monitoring techniques involving oil particle counting (OPC) or current signature analysis. The key issue is a very simple one on the surface, and surrounds the fact that data resolution and sample rates of such systems are inherently different.

Considering temperature and vibration again, temperature data will likely come from the WT SCADA system. This typically consists of 10 minute averaged data,

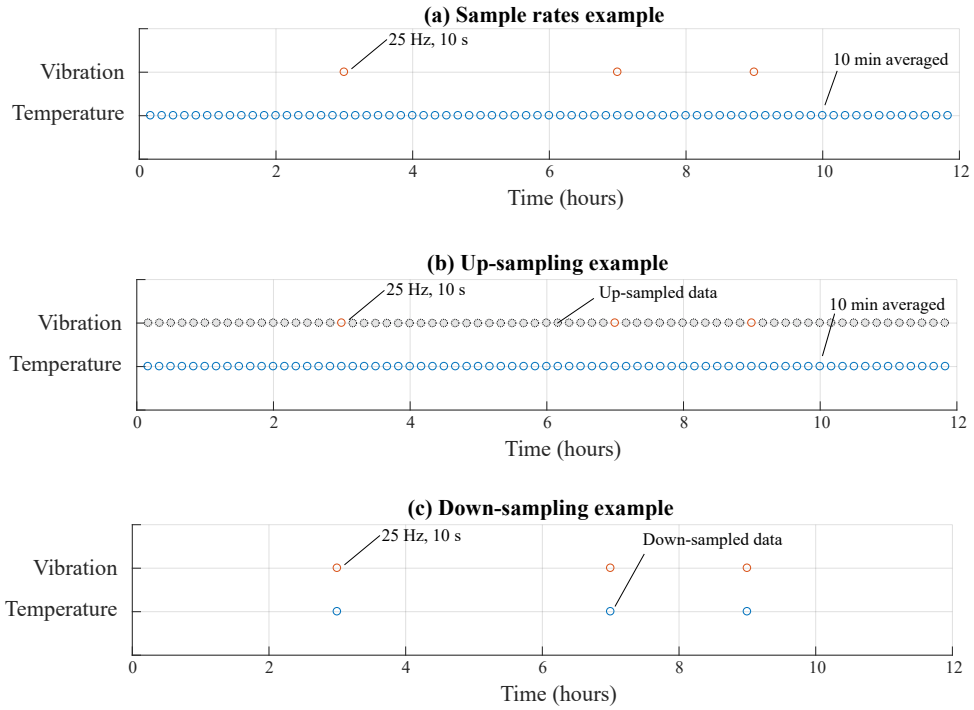


Figure 4.1: Example data acquisition rate for different data sources.

with potentially some limited information about the variation of temperature within that 10 minute period. The frequency of acquisition will not change unless there is a problem with the SCADA system preventing data being recorded. The vibration data on the other hand will be acquired at a much higher frequency of up to approximately 25kHz for a sample time of over 25s depending on the component being monitored. Slower rotating components will likely have longer sample times in order to capture the full shaft rotation for frequency domain analysis. Sample rates may also change depending on turbine operating state in order to maximise efficiency of data storage. Furthermore, companies that offer vibration data acquisition tend to implement decay rates to reduce the volume of older data being stored in the system. This means that if looking at examples of failure in the past, unless careful with data management there may be less frequent vibration samples as you go further back in time. A typical example of acquisition rates of both data streams is demonstrated using example (a) of Figure 4.1.

4.2.1 Up-sampling vs Down-sampling

The obvious solution would be to either up-sample or down-sample the raw data in order to align each data sample rate with the other. There is a variety of issues associated with either case which is worth explaining in detail.

First of all let us consider up-sampling vibration data using example (b) shown in Figure 4.1. This poses a particular challenge due to the variable nature of wind turbine loading. Relationships between operating conditions, loading throughout the drivetrain and hence, vibration are also highly nonlinear. SCADA data is captured every 10 minutes regardless of operating condition, or in other words, over the full range of the power curve. Taking a broad view of up-sampling, relationships would have to be understood between the vibration and temperature diagnostic for all values you would want to predict over. This is not a trivial task in itself, however it is further complicated by fluctuations within the 10 minute period you are up-sampling to. For example, an instantaneous value of power at the time of vibration measurement may be known, which can be mapped to a 10 minute mean temperature measurement encompassing the same time period as when the vibration sample was taken. Wind speed can change significantly within the 10 minute time period leading to differences between the SCADA based mean power and vibration based instantaneous power. Additionally, there will always be some level of uncertainty with regards to timestamps of different systems not being synced accurately. In other words, we cannot say with absolute certainty we are comparing like for like when up-sampling data, which can lead to model uncertainty and prediction inaccuracy.

Down-sampling data offers similar challenges with regards to the differences between operating states within the 10 minute averaged SCADA sample and 10 second vibration sample taken within this larger time window. It also has the additional drawback of losing information for the diagnostic model as you remove data in line with example (c) shown in Figure 4.1. That being said, it does put a stop to any additional uncertainty being introduced to modelling by up-sampling. The choice will ultimately depend on the dataset in question, and may involve a hybrid of the two approaches described above.

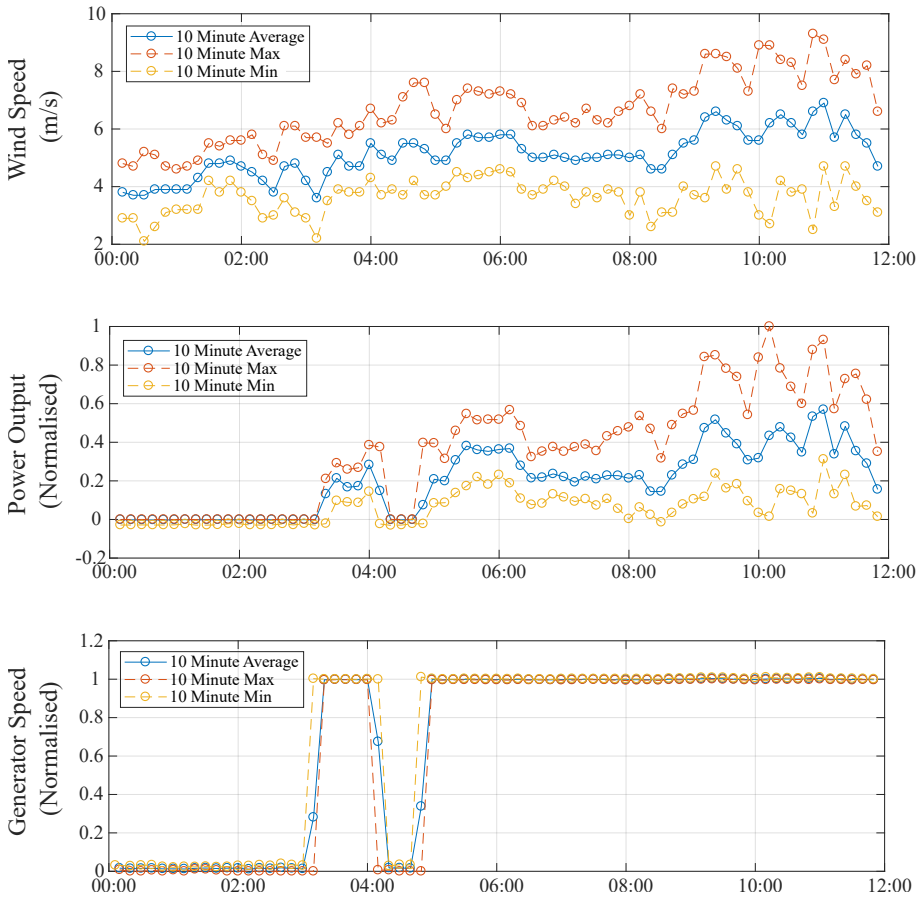


Figure 4.2: Example data showing variation within a series of 10 minute periods.

4.2.2 Expected Variation

To provide insight into the extent of the problem it is important to first understand the expected variation over time. Figure 4.2 shows variations in wind speed, power output and generator rotor speed over a typical 12 hour period over which we would want to build an anomaly detection model. The plot shows the 10 minute average value for each operational parameter along with the maximum and minimum recorded value during each 10 minute period. Data has been normalised to the maximum value of each parameter (excluding wind speed) experienced over the 12 hour window.

This plot shows evidence of two factors described above: first of all it shows substantial variation within the 10 minute averaged data, with variation of up to approximately 90% when considering changes in both wind speed and associated power output.

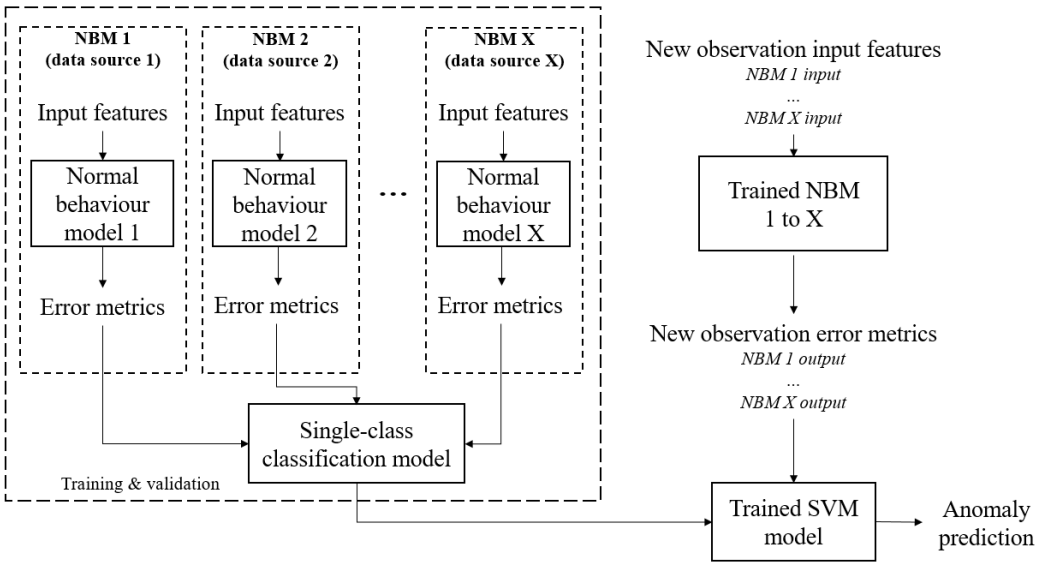


Figure 4.3: Framework for combining data streams.

It also highlights the problems which could occur if timestamps of acquisition systems are not synced, with an offset of just 1 hour producing over 20% difference in operating parameter, a value which could likely increase over a different 12 hour period.

4.3 Modelling Framework

The previous section discussed some of the limitations surrounding up-sampling and down-sampling raw data, highlighting why it is important to have a robust approach to fuse data sets. With that in mind, to protect as much information as possible in the combined model, individual normal behaviour models were first created and later combined when assessing error. Figure 4.3 provides a step by step diagram of the framework used for creating a single anomaly detection model from multiple sources. The overall aim of this is to create a single anomaly detection model for a particular component, which encompasses more than one data source. To make this possible, each data source needs to add value to the model when detecting damage within the component through anomalous behaviour. For this reason each data source must be able to sustain its own diagnostic health indicator in isolation.

A normal behaviour model is first developed for each diagnostic indicator to model

expected behaviour using that particular data source. An example of this can be seen in Figure 4.3 as NBM 1 to NBM X, where X is the maximum practical number of models for a particular component depending on the monitoring systems. The specific inputs and error used for making and assessing predictions will be discussed in detail when describing each case study. The important point to note at this stage is that each individual NBM has its own unique inputs, predictions and associated error with different data sample rates. These trained models are saved and used to make predictions on new data using the same input features.

Error metrics are then calculated over a set time resolution to consider all trained NBM's. This new dataset will now contain error metrics at the same timestamp regardless of original data source to use in the anomaly detection model. This anomaly detection element is achieved by applying a single class Support Vector Machine (SVM) classifier.

4.3.1 Normal Behaviour Model

As discussed in the literature review and previous chapter, the NBM works by empirically modelling the diagnostic health indicator value based on a range of input parameters unique to each data source. The process is summarised in Figure 4.4, where $u(t_i)$ are the input variables at timestep t, $\hat{G}(t_i)$ represents the data-driven NBM to predict target variable, $\hat{y}(t_i)$ while $G(t_i)$ constitutes the process of obtaining the measured target variable $y(t_i)$ through the required in field sensor. Finally, $e(t_i)$ represents the error between the predicted and measured value. Error metrics, $e(t_c)$ can then be calculated at the chosen timestep for the combined model.

There are multiple regressive base machine learning algorithms which could be chosen to create the NBM's, all of which are discussed extensively in the literature review. Throughout this chapter both Random Forest and Neural Networks are utilised.

4.3.2 Combined Anomaly Detection Classifier

Anomaly detection is achieved using a single class SVM classifier to assess error distribution by creating a complex decision boundary based on multiple error statistics over

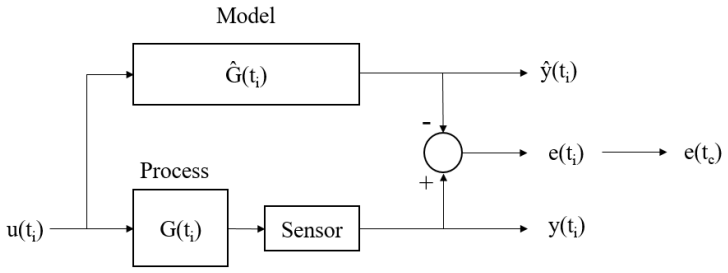


Figure 4.4: NBM diagram

a given time period. This means that RMSE is no longer used in isolation, in theory allowing any observed change in error distribution to be detected which lies outside the decision boundary. It also provides a useful metric when assessing the anomaly as to how far outside the boundary a point lies, which can be use to infer how anomalous the data sample is.

4.4 Assessing Error through Classification

To assess the robustness of the outlined approach the methodology was first of all tested on SCADA data only. This allows a direct comparison to be made with simpler techniques to showcase and prove the broader concept prior to combining data streams.

4.4.1 Case Study Description

The first failure mode used for this study is associated with the high speed shaft (HSS) of the gearbox. The initial point of failure was the HSS, however on inspection noticeable damage was also observed to the gears, with the gearbox also needing to be replaced. Root cause analysis suggested that the failure occurred due to misalignment causing out of plane loads both on the shaft and through the gearbox. For confidentiality reasons the exact type and model cannot be explicitly stated, however the WT in question has a doubly fed induction generator (DFIG) and is of between 500 kW and 1 MW at rated power. It has a pitch regulated variable speed control strategy, and had been operated onshore for over 15 years prior to failure.

Once the failure date was determined from the O&M log, 10 minute averaged

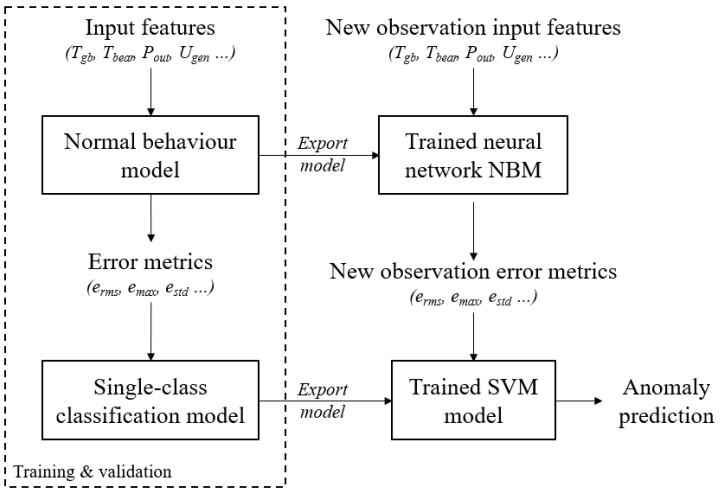


Figure 4.5: Methodology for Case Study 1

SCADA data for 18 months leading up to failure was collected. The WT in question had no CM systems installed therefore the fault type could not be analysed or confirmed through spectral analysis of vibration.

4.4.2 Methodology

The adapted methodology for this case study is described in Figure 4.5. The first step in the process is to develop the NBM using a neural network (see Section 4.4.4 for model details), which takes a variety of input features to predict gearbox oil temperature. Input features were chosen through feature analysis (as described in earlier sections) which could adequately describe WT performance, as well as relevant temperatures throughout the nacelle and components related to the gearbox and generator. Once the error had been minimised through the training and validation process (as described in Section 4.4.5), the error between model prediction and measured temperature is then calculated for each time step.

The second stage of the process is to evaluate the error output and determine a threshold which can adequately distinguish between normal and anomalous behaviour. To achieve this, error metrics were first calculated to describe the error distribution over a chosen time period, which for this case study daily metrics were calculated. For the

Table 4.1: Case study 1 - Summary of data

Model Phase	**Raw data (10 min)**	**Coverage (%)**	**Averaged data**	**Cleaned data**
Model development	48096	92	8016	6974
Training (70%)	-	-	-	*4882*
Validation (15%)	-	-	-	*1046*
Testing (15%)	-	-	-	*1046*
Model implementation	26634	94	4439	3695

same training dataset used for the NBM, a single class SVM classifier was trained using the error metrics to develop a decision boundary, as described in Section 4.3. Whilst other classifiers could have been chosen, an SVM model was deemed well suited to this specific application, mainly due to the models inherit ability to establish a continuous, complex decision boundary in which a spatial representation can easily be visualised.

Once both models have been trained and validated they could then be used on new data points to detect anomalies. This is done by using the NBM to first predict temperature based on the same inputs, assessing the error between predicted and measured value, before finally feeding those error metrics into the classifier to determine if the new data point was normal or anomalous. The purpose of presenting the first case study is to prove the methodology with a single data source, showcasing how it can outperform simpler approaches to evaluate NBM error such as using RMSE as a single metric.

4.4.3 Summary of Data

A summary of the data used in this case study can be found in Table 4.1, which distinguishes the amount of raw data available in comparison to the data available for each development phase after cleaning. Data coverage is also stated, which specifies the percentage of data available for modelling compared to how much data is theoretically possible if the acquisition system recorded 100% of data without issue.

Table 4.2: Case study 1 - model features

Feature No.	Feature	Description	Layer
1	P_{out}	Power output	Input layer
2	U_{wind}	Wind speed	Input layer
3	U_{hss}	High speed shaft speed	Input layer
4	T_{nac}	Nacelle temperature	Input layer
5	T_{gen}	Generator phase temperature	Input layer
6	T_{bear}	Bearing temperature	Input layer
7	T_{gb}	Gearbox oil temperature	Output layer

4.4.4 NBM Specification

A two-layer feed-forward neural network was utilised for the NBM which had an input layer consisting of 6 WT operational parameters to predict a single output (Feature No. 7), as described in Table 4.2. Model input features include parameters to describe the operating conditions (Features No. 1-3) and temperature distribution throughout the nacelle (Features No. 4-6). As previously stated, the model output parameter used for anomaly detection was the gearbox oil temperature. Additional model parameters for the neural network were chosen to reflect the number of input and output features required [94] which in this case utilised a single hidden layer with 7 neurons. This provided a balance between the accuracy of prediction and computational time for training, and although this wasn't a restricting factor due to the relatively small dataset, there was no need to go beyond 7 neurons for this application due to the accuracy achieved. Figure 4.6 shows a diagram of a two-layer feed-forward neural network used for this case study.

Of the 18 months of SCADA data that was gathered prior to failure, the initial 12 months (which we know had no serious faults through analysing the O&M logs) was used for training and validating the NBM, with the final 6 months used to test and track error leading up to failure. From the initial data set consisting of continuous 10 minute mean values the data was cleaned to remove any periods of curtailment or downtime due to scheduled maintenance. The process behind this has been discussed throughout

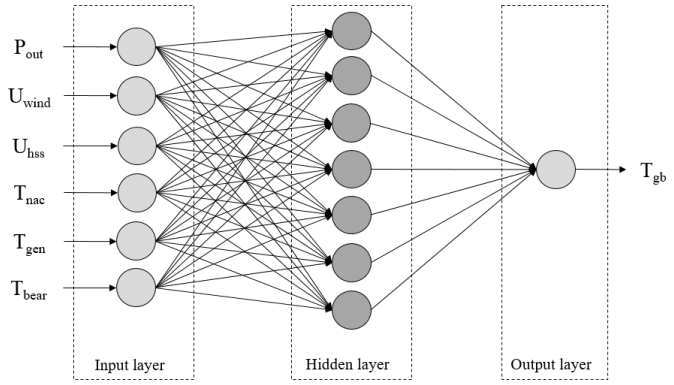


Figure 4.6: Schematic of two-layer feed forward neural network.

the previous chapter. The hourly averages were then calculated to remove some of the higher resolution fluctuations which can decrease model accuracy, allowing the model to be more tailored towards longer term behavioural trends. Once this process had been completed, this left a total of 6974 samples for training and validating the model, with an additional 3695 samples left for testing and tracking the error prior to failure.

4.4.5 NBM Training and Validation

Out of the 3695 samples in the 12 months training data set, 70% was used for training and 15% for validation, with the remaining 15% used to test the model independently as described in Table 4.1. From experience this provides a good balance between the number of samples required for training, validating and testing the model based on the volume of data used in this study. During the training phase, the training and validation samples are chosen at random and then fed into the neural network, which is adjusted in line with the error between predicted and known values of the target variable. It is then validated with the validation samples and mean square error is calculated for the new data points. This process is repeated until the mean square error no longer increases for the validation data set, indicating that the neural network generalizes well and is no longer over-fitting to the training set. Once this has been achieved the model is then independently tested with the testing dataset (the 15% of data leftover after randomly selecting training and validation data) to ensure generalisation.

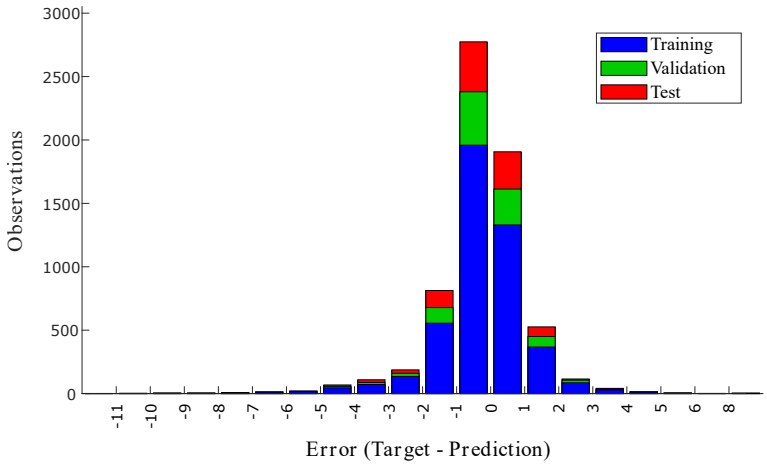


Figure 4.7: Distribution of error from 12 month training period

To get a sense of the error distribution for this particular model, a histogram can be observed in Figure 4.7, showing the model error over the 12 month training period, broken down into the phases described above. The mean squared error over the training, validation and testing phases were 2.364, 2.254 and 3.075 respectively, with correlation coefficient (r) values of 0.859, 0.859 and 0.832. With r-values deviating approximately 3.2% between data sets, this model generalises relatively well, which sets a good foundation when trying to detect anomalies leading up to failure.

4.4.6 Single Class SVM Model

Once a baseline expected error has been established from the training period there are several ways in which to then compare residuals moving forward. Typically this is achieved by comparing the daily or weekly RMSE with the RMSE of the training period to give an indication of whether the temperature (or other chosen parameter) is acting as predicted. This approach however does come with limitations, stemming mainly from the fact that only one parameter is used to describe an error which, over any particular time period is multi-faceted and has a unique error distribution associated with it. More than one parameter can be looked at in isolation, or indeed the entire distribution could be described and tracked in time, however this introduces a different problem; how to robustly set thresholds which will indicate the fault. Using

Table 4.3: SVM error model feature

Feature No.	Feature	Description
1	e_{rms}	RMSE
2	e_{min}	Min error
3	e_{max}	Max error
4	e_{std}	Standard deviation of error distribution
5	$e_{kurtosis}$	Kurtosis of error distribution

a single class support vector machine aims to address these limitations, first of all by considering multiple parameters which can effectively describe the distribution of error over a chosen time period, and secondly to set more complex boundaries which can more precisely describe the threshold to indicate a fault.

A single class SVM model was developed to evaluate the error distribution each day based on the NBM output. For each daily error distribution the parameters stated in Table 4.3 were calculated and used as inputs to the SVM model.

To begin with let us consider a single class SVM model with only 2 features as an example. Using RMSE and maximum error, the model can be trained with the cleaned 12 month data set, giving a total of 290 samples. Figure 4.8 shows each observation in the trained model, with support vectors (which influence the decision boundaries) circled in red. On the right hand side of Figure 4.8 there is a scale which corresponds to the contour line boundaries indicating an anomaly. In the model shown, any score below zero would constitute an outlier or anomaly, which lies outside the decision boundary. The more negative a score the further away from the central cluster the observation is. This process was then repeated with all parameters which describe the daily error distribution as shown in Table 4.3. Although these models could not be visualised in the same manner, the scoring system remained the same, and could be used to detect outliers in new observations leading up to failure. The model was trained to recognise 1% of data as anomalies in the training period, therefore a similar percentile would be expected moving forward if no fault was present in the system.

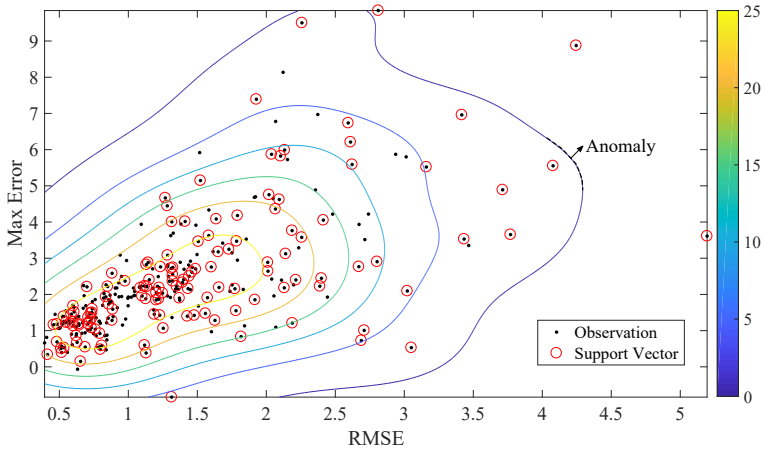


Figure 4.8: Trained SVM model with only 2 features; 1) RMSE and 2) Max error

4.4.7 Anomaly Detection Results

Using the models described in Sections 4.4.5 and 4.4.6 the error over the 6 months leading up to failure was evaluated. For every 1-hour time step the error output was determined and the daily statistics shown in Table 4.3 calculated. If features fed into the single class SVM model were considered part of the same class (of normal behaviour), it resulted in a positive score and was assigned an anomaly score of zero. If the model output score was below zero, this was considered an outlier. In order to create a heatmap, this score was then inverted to give a positive anomaly score which provides an indication of fault severity, or distance from the determined frontier which defined the class.

Figure 4.9 shows a heatmap of the daily anomaly score leading up to failure. A total of 154 days are shown, which accounts for all 24 hour periods in each month while the turbine was operating under normal operating conditions. If the WT was curtailed, shut down due to planned maintenance activities or the wind was below cut in speed or above cut out speed for an entire 24 hours, data was removed during the cleaning process described in the previous sections. If during the 24 hour period the WT partially experienced normal operating conditions, this data was still used to calculate the daily error metrics.

Table 4.4 gives a breakdown of anomalies detected per month along with associated

Table 4.4: Anomaly rate using single class SVM apprach.

Month	**Number of anomalies**	**Percentage of anomalies**	**Turbine active days per month (Days of month in data)**
May	0	0%	22 (22)
June	2	8%	25 (30)
July	1	3%	30 (31)
August	4	13%	31 (31)
September	7	24%	29 (30)
October	6	25%	30 (31)
November	1	14%	7 (7)

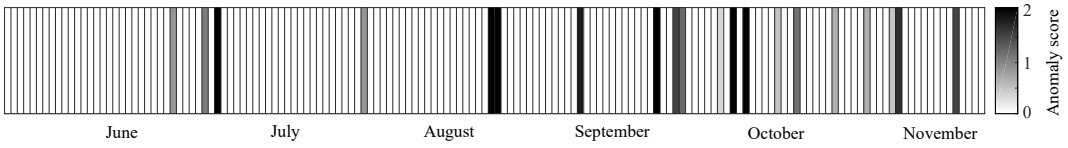


Figure 4.9: Heatmap of anomalies leading up to failure

percentages taking into consideration the number of active days. In general, results show that anomalies increase towards failure, with a maximum of 8% per month in the 4-6 months before failure, increasing to 25% in the 2 months directly before failure. In terms of detection time, the first meaningful and consistent anomalies are observed in August, giving approximately 3 months lead time. A slight decrease in anomalies are reported in November just before failure, however this could be down to the significantly less active days (1 week) in the month to assess and detect anomalous behaviour before failure occurred. There could also be a physical explanation to fluctuations in anomalies, with anomaly rate correlating with time of damage progressing within the highspeed assembly. There is no way of knowing with confidence without corresponding inspections on site.

4.4.8 Comparison of Results with Other Methods

It is important to compare results obtained using this methodology with standard approaches as found in literature such as [6, 95, 96], as discussed in depth throughout

Table 4.5: Anomaly rate using standard RMSE approach

Date	**Number of anomalies** *(threshold 1, 2, 3)*	**Percentage of anomalies** *(threshold 1, 2, 3)*
May	8 , 1 , 0	36% , 4.5% , 0%
June	11, 1, 0	44% , 4% , 0%
July	9, 0, 0	30% , 0% , 0%
August	11, 2, 0	35.5% , 6.5% , 0%
September	14, 6, 2	48% , 20.7% , 6.9%
October	23, 2, 0	76.6% , 6% , 0%
November	4, 1, 0	57% , 14.3% , 0%

the literature review. To do this the RMSE and standard deviation were first calculated over the entire training period, which could then be compared to the daily RMSE for the 6 months prior to failure. Three different thresholds were used to compare the daily RMSE with the training period; the first threshold was simply the RMSE, the second was 1 standard deviation above the RMSE, while the third threshold was 2 standard deviations above the RMSE. This gave error thresholds of 1.57, 3.15 and 4.72 degrees Celsius respectively, whereby any new observation with a daily RMSE of above these limits was deemed an anomaly. Results using all 3 thresholds individually are shown in Table 4.5.

The first threshold shows a clear increase in anomalies leading up to failure, however with such large proportion of errors across the entire 6 months, in practice this would lead to a large amount of false alarms. The second threshold actually performs relatively well, with an average of less than 5% of anomalies detected over the first 2 months. Moving closer to failure also results in a clear increase in anomaly rate using this threshold, however it does not perform as well as the SVM model presented earlier in the chapter. This is best observed through Figure 4.10, which shows a direct comparison of results. By plotting a simple linear regression of each set of results it becomes obvious that the SVM model not only detects on average more anomalies per month, but is also more consistent over the final two months. It is also important to note that both models have a very low anomaly detection rate in May, with the SVM model actually

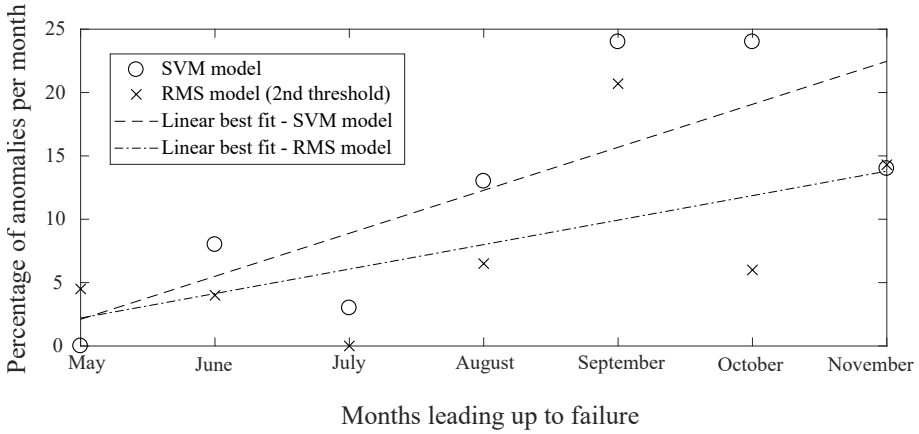


Figure 4.10: Comparison of SVM classification approach to RMSE

detecting less anomalies than the threshold based on the RMSE and standard deviation, giving confidence that increase in anomaly rate actually relates to the fault. Using the third threshold, anomalies were only detected for a short spell 2 months prior to failure, indicating that the limit was set too high to track progression, however does offer an insight into the largest changes in behaviour, albeit for a short period of time.

Results here have demonstrated that single class SVM models are a successful method to evaluate errors output by a NBM, which in this case used a two-layer feed forward neural network. Using a SVM model acts as an improvement to existing techniques by first of all allowing multiple parameters to be used which can effectively describe the error distribution over a chosen time period. Secondly it allows for a more complex decision boundary to be formed which can detect and distinguish anomalous behaviour from normal behaviour. For this first case study results showed anomalies in gearbox oil temperature can be detected up to 3 months before the HSS failure. It also shows a larger and more consistent increase when compared to simple thresholds based on RMSE alone. As stated previously, the decrease in anomaly rate observed in November could be attributed to the fact there was only 7 days, therefore less data in which to make the calculation, or perhaps related to some physical phenomena regarding damage progression.

Figure 4.11: Damage to Generator bearing and shaft at time of replacement.

4.5 Combining SCADA and Vibration for Fault Detection

4.5.1 Case Study Description

The second case study focuses on generator bearing failure, which in this case stems from raised bearing temperatures leading to bearing inner ring growth resulting in the bearing inner ring spinning on the generator rotor shaft at the drive end, with Figure 4.11 showing damage that can occur over time from a fault of this nature. In order to ensure confidentiality again the exact power output and wind turbine and bearing type used is not provided, however it can be stated that it was a doubly-fed induction generator (DFIG) with a rated power of between 2 and 4 MW, again utilising a variable speed, pitch regulated control strategy.

Once the failure was identified in the wind turbine OEM O&M event log, both SCADA data and vibration data were retrieved for the year leading up to failure. To ensure dates collected were correct SCADA data was checked directly after the failure date to ensure wind turbine downtime occurred as stated in the events log. SCADA data consisted of 10-minute mean values taken continuously over the entire year, while vibration data samples were taken 1 week apart at both the drive end and non-drive end of the generator. Each vibration sample consisted of approximately 10 seconds of data taken with a sampling frequency of approximately 25 kHz. A summary of SCADA and vibration data available for the second case study can be found in Table 4.6 and 4.7 respectively.

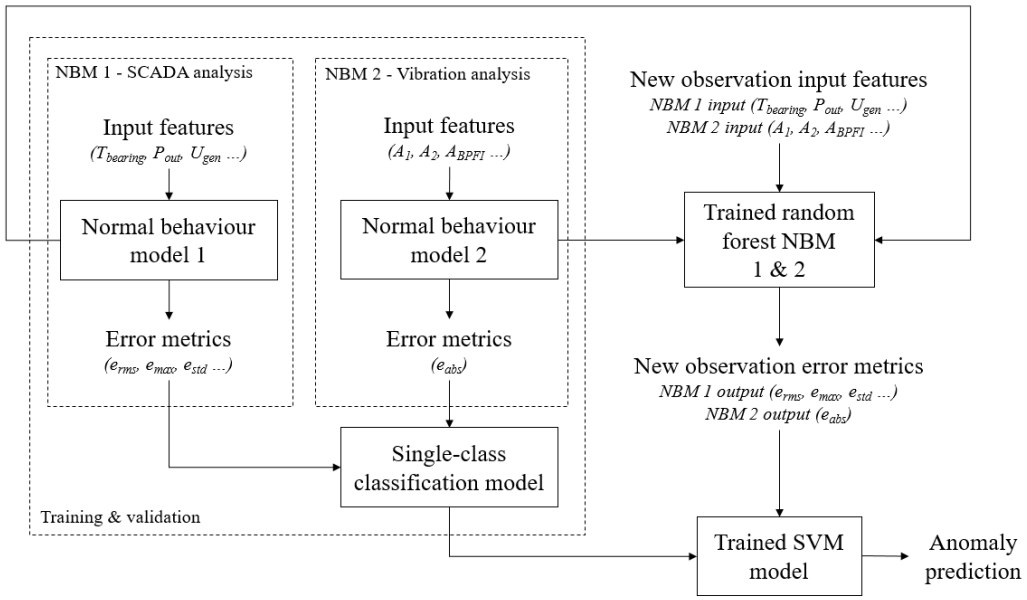


Figure 4.12: Methodology showing 2 data sources combined into single anomaly detection model; SCADA and Vibration.

The methodology for this case study, which can be found in Figure 4.12 (adapted from Figure 4.3), builds upon the first case study described in Section 4.4 by encompassing two data sources. NBM 1, which is based on SCADA analysis, predicts generator bearing temperature relying on multiple operational parameters and temperatures as inputs using a random forest regressive model (see Section 4.5.2 for details), and compares these predictions with observed temperature measurements. NBM 2, which is based on vibration analysis, predicts a summation of fault frequency amplitudes and relies on a variety of spectral features and operating conditions as inputs. Error is evaluated for both models and metrics describing the error distribution are calculated for the same weekly time resolution in each case. These error metrics are then used as inputs features to train a single class SVM model for anomaly detection.

4.5.2 Normal Behaviour Model

For the second case study both SCADA and vibration are used, however for the NBM independent analysis was performed for each data source following the same methodology as used previously. Again, Figure 4.4 illustrates the process of using these types of

Table 4.6: Summary of SCADA data

Model Phase	**Raw data (10 min)**	**Coverage (%)**	**Averaged data**	**Cleaned data**
Total	52104	99	8785	5275
Model development	26052	-	4391	2634
Model implementation	26052	-	4394	2641

Table 4.7: Case study 2 - Summary of Vibration data

Sensor location	**No. Samples**	**Frequency**	**Sample time**
Generator drive end	56 (1 week apart)	Approx. 25kHz	Approx. 10 sec
Generator non-drive end	56 (1 week apart)	Approx. 25kHz	Approx. 10 sec

models to detect faults, with the residual error, $e(t)$, this time between either measured temperature or vibration, $y(t)$, and model prediction $\hat{y}(t)$ used as an indicator for a potential fault.

Although a neural network was used in the first case study, in reality there are many supervised regressive machine learning models that could be chosen to represent normal behaviour of the target variable. To highlight this fact, a Random Forest algorithm was selected for the second case study. More detail can be found in the literature review, but as a short recap random forests are ensemble learning models that build multiple decision trees and merge them together to get a more accurate and stable prediction. Using ensemble methods allows for overall better performance and reduces the risk of over-fitting, which can be an issue when utilising solitary decision trees. Figure 4.13 shows a diagram of this type of Random Forest Regressive model in relation to model input features, prediction output and how these relate to the number of decision trees used.

SCADA NBM Description

The SCADA channels available for this analysis are stated in Table 4.8. Like the previous study, data describes the operating conditions and power output at each timestamp, as well as a range of temperatures throughout the nacelle at selected locations,

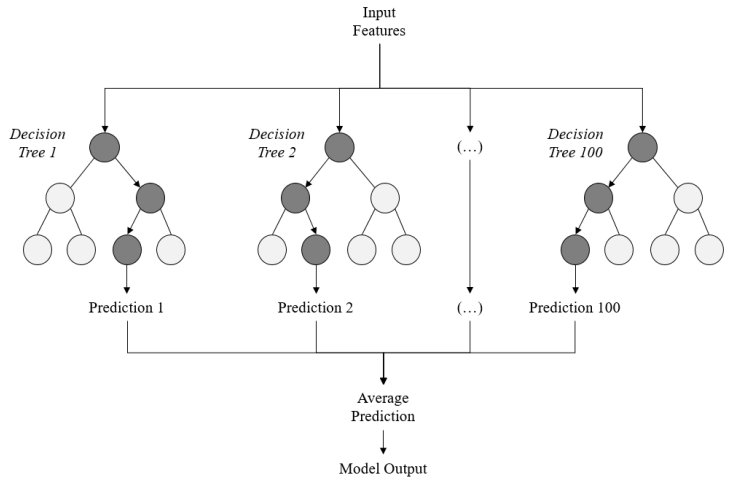


Figure 4.13: Schematic of Random Forest Regressive model - input features and output dependant on model.

this time to give a representation of overall generator performance and condition at each instance. Feature importance was determined through a combination of a model based approach and expert knowledge, with this having been discussed in more detail throughout Chapter 3.

As with the first case study, data cleaning is a crucial component in producing any normal behaviour model which can detect anomalies effectively. For this case study data was cleaned to remove times in which the wind turbine was operating in conditions out with the normal torque speed operating strategy. Primarily this was any sustained periods of shutdown or power de-rating, which could occur for a number of reasons including maintenance work or grid constraints.

Vibration NBM Description

Unlike the SCADA data, in which features can be taken directly, features must first of all be identified and extracted from the vibration signal. The features which are extracted depend on the failure mode which is trying to be detected, with different failures having different markers, or fault identifiers, within the signal. As discussed in detail in Chapter 3, the raw time domain signal is not sufficient to extract these identifiers, therefore frequency domain analysis is required, where the time domain

Table 4.8: Case study 2 - SCADA model features

Feature No.	Feature	Model	Description
1	P_{out}	Predictor	Average power output
2	U_{rotor}	Predictor	Average rotor speed
3	U_{gen}	Predictor	Average generator speed
4	U_{wind}	Predictor	Average wind speed
5	$T_{ambient}$	Predictor	Average ambient temperature
6	$T_{nacelle}$	Predictor	Average nacelle temperature
7	$T_{slipring}$	Predictor	Average generator slip ring temperature
8	T_{phase1}	Predictor	Average generator phase 1 temperature
9	T_{phase2}	Predictor	Average generator phase 2 temperature
10	T_{phase3}	Predictor	Average generator phase 3 temperature
11	$T_{bearing}$	Response	Average bearing temperature over

signal is converted into the frequency domain using a fast Fourier transform (FFT) while also implementing order analysis techniques. Table 4.9 highlights the input and output parameters used for the vibration based NBM. Input predictors consisted of the power output at the time the vibration sample was taken, along with vibration amplitudes at shaft frequency and two harmonics. The output response, or diagnostic, was the sum of vibration amplitudes at known fault frequencies, which in this case was the ball passing frequency of the inner race (BPFI) together with two harmonics and associated side bands at shaft frequency. All amplitudes were identified and values extracted in the order domain.

4.5.3 Single Class SVM Model

Features describing the error from both the SCADA and vibration NBM's were used for the classification model. The features used for the second case study are described in Table 4.10 alongside those used in the first. The key difference in case study 2 being the fact that weekly stats were used instead of the daily features used in the first case study. This is in line with the lowest data resolution of the vibration samples, therefore only a single feature could be used to train the classifier from the vibration NBM. For

Table 4.9: Case study 2 - vibration model features

Feature No.	**Feature**	**Model**	**Description**
1	A_1	Predictor	Amplitude of peak at order number 1
2	A_2	Predictor	Amplitude of peak at order number 2
3	A_3	Predictor	Amplitude of peak at order number 3
4	P_{out}	Predictor	Generator power output at time of vibration sample
5	$A_{diagnostic}$	Response	Sum of amplitude peaks at fault frequencies

Table 4.10: SVM error combined model features

Feature No.	Feature	Description	Input
1	e_{rms}	RMSE	SCADA
2	e_{min}	Min error	SCADA
3	e_{max}	Max error	SCADA
4	e_{std}	Standard deviation of error distribution	SCADA
5	$e_{kurtosis}$	Kurtosis of error distribution	SCADA
6	e_{abs}	Absolute error	Vibration

this particular case study sample size was deemed more important than the number of features per sample, which could be increased for example, if monthly metrics were used instead.

Figure 4.14 shows an example of the trained SVM classifier for the 6 months of data using the weekly SCADA RMSE and vibration absolute error. This was again extended to include all features described in Table 4.10 in order to get a more comprehensive understanding of weekly error distribution. It should be noted that although the scale on Figure 4.14 ranges from 2 to -1.5, zero remained the learned frontier, with 1 data point from the training data in this example outside this threshold.

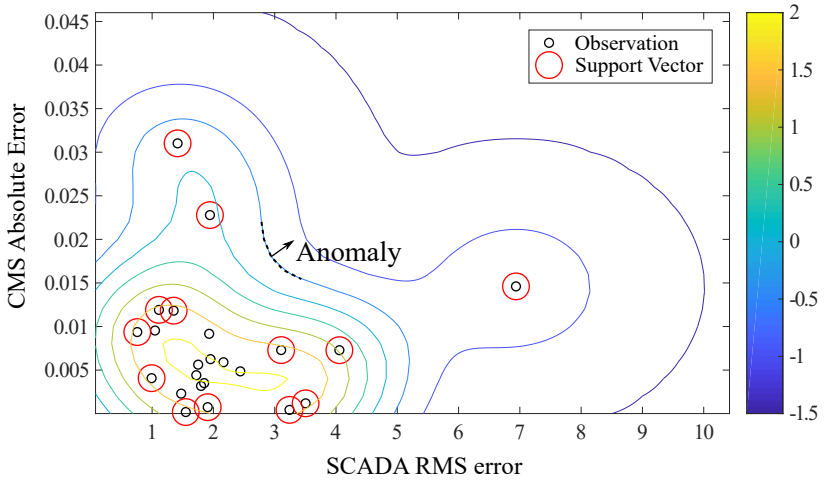


Figure 4.14: Trained SVM model with only 2 features; 1) SCADA RMSE and 2) Vibration Absolute error

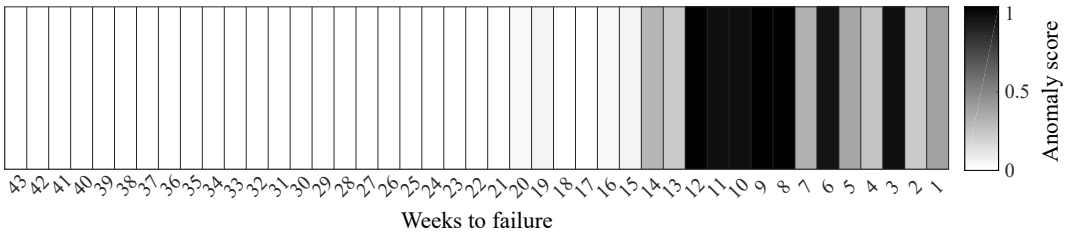


Figure 4.15: Heatmap of anomalies leading up to failure

4.5.4 Anomaly Detection Results

Results show that anomalies were detected consistently 16 weeks prior to failure, with errors associated with the highest anomaly score appearing between 10 and 12 weeks before failure. As with the previous case study, if the new observation was deemed normal behaviour it was simply given a score of zero, with anomalies given a higher score the further from the decision boundary they stray. Figure 4.15 shows the heatmap of anomalies for approximately 10 months leading up to failure.

4.5.5 Comparison with Single Models

For completeness the results from Section 4.5.4 have been compared with each NBM individually using a single threshold based on the RMSE and standard deviation of the training period (see threshold 2 in Section 4.4.8 for details). Figure 4.16 shows

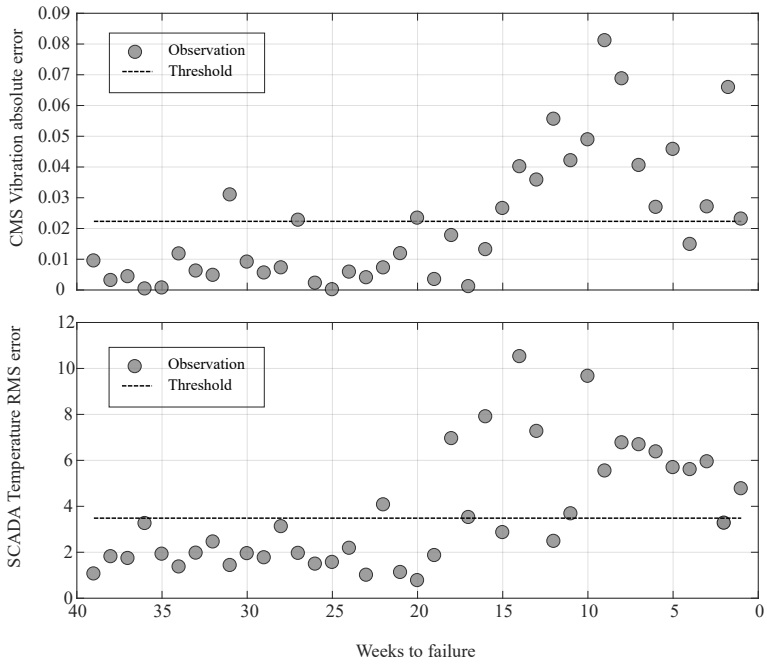


Figure 4.16: Plot of weekly errors leading up to failure for individual NBM's

the weekly error for each model along with the calculated threshold. Results shows an initial step change in vibration levels at approximately 14 weeks to failure, with an increase in temperature actually occurring in the 3 weeks prior to this. The heatmap shown in Figure 4.15 demonstrates that the SVM model captures changes in both temperature and vibration. This rise is then followed by a slight decrease in average vibration and temperature in the 7 weeks directly before failure, which is again in line with Figure 4.16.

Results show that an SVM classifier is an effective method in evaluating the error from two different NBM's that can describe bearing failure, each with a different data source and analysis techniques. Figure 4.14 shows that the decision boundary can be modelled to account for each error, and allows for anomalies to be detected for new observations in which the temperature error, vibration error, or indeed both, deviate from expected levels.

The decrease in vibration and temperature levels which occur approximately 7 weeks directly before failure could be attributed to the type of failure used in this case study. Bearing faults can see increases in temperature or vibration due to pitting (in this case

on the inner race) which can smooth over time due to further wear. This can often then be followed by additional step changes as more pitting or damage to the inner race and bearing assembly occurs, or in this case, failure. As this analysis focuses on detecting the bearing fault at the end of life the step changes in vibration and temperature are quite apparent. To see any initial changes in vibration due to early fault onset more vibration samples would have to be taken going further back in time as it likely occurred prior to this data set. These samples were not made available for use in this study.

As highlighted previously the main benefit for using the SVM classifier is to combine models into a single threshold, which in this case produces consistent anomalies 14 weeks prior to failure. If looking at temperature and vibration in isolation this is not the case, with each model dropping below the anomaly threshold at some stage in the final weeks. In terms of early fault detection or failure prediction, this study does not show substantial improvements to actual lead time from first anomaly detection to failure.

4.6 Conclusions

This chapter has highlighted the key issues surrounding fusion of multiple raw data sources to detect WT faults. A framework has been presented which aims to navigate around some of these potential sources of uncertainty by combining data streams when assessing NBM error. This is useful as it allows for individual NBM's to be first set up and trained without any loss of information. This has been demonstrated by 2 case studies, the first of which allows a direct comparison to be made of assessing error using a single class SVM classifier to simpler threshold techniques observed in literature. The second study then combines temperature and vibration models to detect a generator bearing fault. As shown, this fault had progressed quite substantially and could therefore be detected in both temperature and vibration. By using a combined classifier, a single anomaly boundary can be established to use as a single health indicator capable of detecting changes in either parameter. It should be highlighted that these conclusions have been drawn based on the small amount of case studies presented

in this chapter. Without applying these techniques on more data across different case studies it is difficult to say conclusively if the SVM classifier offers a better alternative to simpler thresholds.

4.7 Future Work

Future work is required to validate this approach on further generator faults through temperature and vibration monitoring. Additionally, by introducing oil debris analysis this methodology could be adapted for gearbox related issues through monitoring of oil, temperature and vibration. This would allow for a single holistic health indicator to be established for any component throughout the drivetrain, reducing duplicate alarms being generated from multiple sources. This may ultimately reduce time spent by condition monitoring engineers clearing alarms, allowing more focus to be applied to real faults or issues.

Chapter 5

Failure Prognosis using Classification of Vibration Features under Varying Operating Conditions.

5.1 Chapter Contribution

Up until this point Chapters 3 and 4 have utilised different machine learning, data analysis and signal processing techniques to detect, and some cases locate, a range of faults across the generator and high speed assembly. In order for prognostic models to be developed, wind turbines must be analysed during conditions of catastrophic failure. Only once these examples are gathered can any sort of data-driven remaining useful life be determined.

This chapter aims to answer the following research question:

"To what extent can previous experience of failure be used to predict remaining useful life in similar machines based on purely data-driven approaches, and how do different site operational conditions effect prediction accuracy?"

The contributions of this chapter are as follows:

1. Demonstrate how classification algorithms can be used to determine remaining useful life of generator bearings using vibration based condition monitoring systems and historical examples of failure.

2. Provide insight into how operating conditions at the time of vibration measurement affect diagnostic features and classification results which estimate remaining useful life.

3. Present a framework which relies on clustering methods to group historical data based on operating conditions to more accurately estimate remaining useful life.

5.2 Background Information

5.2.1 Prognosis vs Diagnosis

This chapter will focus on wind turbine prognostics using vibration based analysis supplemented by machine learning techniques. As discussed in depth throughout Chapter 3, there are a number of well established and proven techniques using vibration analysis to detect faults in such systems, traditionally based around Fourier analysis, envelope analysis and order analysis techniques, which can all be used to identify and extract features from the signal based on rotational frequencies [77].

The key differentiation between a diagnosis and prognosis is as follows; prognosis aims to estimate remaining useful life before failure of a component with an existing fault whereas diagnosis only aims to diagnose the fault itself. A diagnosis may also be very specific to an exact component and fault type. When dealing with prognostics of failure, the initial fault type may or may not be the eventual point of failure. There could be a single concentrated area of damage or it may have spread to multiple areas across the component or assembly at the point of failure. The following analysis throughout this chapter is based on known generator bearing failures, however it is not known with absolute certainty if the points of failure and specific failure mode in question were identical.

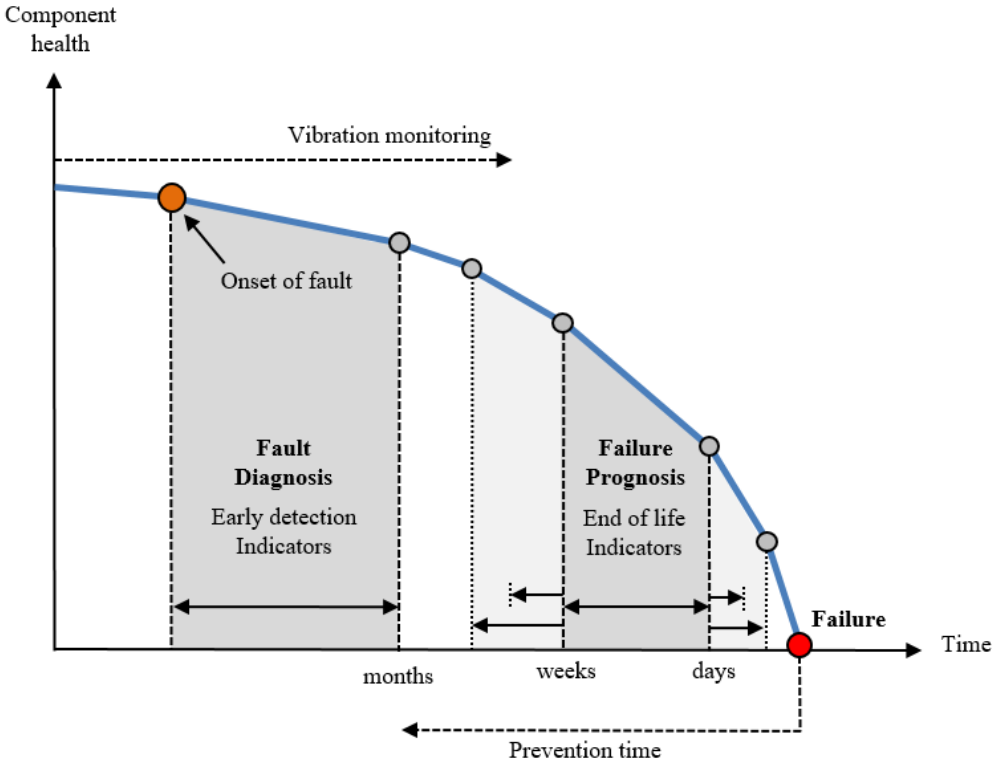


Figure 5.1: Key differences between Diagnosis and Prognosis used in this Thesis.

5.2.2 Vibration Based Indicators

Now that the difference between diagnostics and prognostics has been defined it is equally important to distinguish the difference between required features. When making a diagnosis, the ultimate goal is maximising prevention time by detecting a fault as early as possible, which can then be monitoring more closely and appropriate proactive maintenance carried out that minimises lost production. This is expressed visually in Figure 5.1, where early detection indicators are used to detect the fault as close to fault onset as possible. These indicators were described in detail throughout Chapter 3 when performing fault diagnosis.

As a fault progresses the same indicators used for diagnosis may not be suitable to successfully predict failure. Take generator bearings as an example (for which the case study in this chapter will later center around), where envelope analysis techniques are widely used to detect fault frequencies modulated at carrier frequency. This method

works well for early detection, however as damage worsens, modulation at these higher carrier frequencies becomes less distinct and harder to track. This may be simply due to signal processing of a now noisier signal, or it could be due to physical effects of component wear as initial sharp edges become more rounded, or more likely a combination of the two. What this means is that increases in amplitude in the envelope spectrum cannot be correlated to fault progression once significant damage has occurred. This phenomenon requires additional indicators that are capable of detecting changes in frequencies which occur before component failure, as described in Figure 5.1 as end of life indicators. Features used for this will be described in detail later as the case study is presented.

5.3 Introduction to Case Study

5.3.1 Failure Case

This chapter will focus on generator bearing failure. From discussions with industry partners, it is believed that this failure case stems from raised bearing temperatures leading to bearing inner ring growth resulting in the bearing inner ring spinning on the generator rotor shaft at the drive end. It should be noted that this was determined by assessing only some of the failure examples used throughout this chapter, therefore it cannot be known if every failure was a result of the same issue described above. Root causes sometimes stem from design and manufacturing issues such as imperfections in material grade, out of tolerances, poor shaft alignment and improper installation methods. Other causes include operational and maintenance issues such as high loading, unbalanced electromagnetic forces, damage while in transit or exacerbated by inadequate cooling and inspection strategies [97]. Again, exact root cause can not be stated conclusively, or indeed if each example of failure used in the study had the same root cause.

5.3.2 Dataset

In order to track and compare component condition, vibration data was retrieved at different points in time leading up to a generator bearing failure. To achieve this, events associated with generator bearing failure from a wind turbine OEM were analysed until multiple examples of the same failure mode were identified. This was then cross checked with SCADA data to ensure dates were correlated by checking failure dates with down time experienced by the wind turbine. To ensure best results possible all examples were from an identical generator and drivetrain configuration. In order to guarantee confidentiality the exact power output, wind turbine specification and bearing type used is not provided, however it can be stated that it was a doubly-fed induction generator (DFIG) with a rated power of between 2 and 4 MW. Each turbine utilised a variable speed, pitch regulated control strategy. Generator rotor speed at rated power was partly determined by grid frequency, where examples were found for both 50 and 60 Hz.

A total of 15 different wind turbines from eight wind farms were used in the study, in each of which between 4 and 10 vibration samples were gathered a week apart 1 year, 5-6 months and 1-2 months before failure. In most cases data was gathered from each of the three groups described above. In some cases however not every group could be provided, in which case data was used where available. This is clarified in Table 5.1. Each sample consisted of approximately 10 seconds of data taken with a sampling frequency of approximately 25 kHz at both the drive end and non-drive end generator bearing (see Figure 5.2 for clarification).

Once the population of failure examples had been identified, data was gathered prior to occurrence and labeled based on the time to failure. A binary classification system was used, with samples labeled 'satisfactory' taken at more than 2 months prior to failure and labeled 'unsatisfactory' if the sample was taken between failure and 2 months prior to failure. This gave a total of 306 vibration samples, consisting of 204 in class 0 (satisfactory) and 102 in class 1 (unsatisfactory) for both the drive end and non-drive end sensor. In relation to component prognostics, unsatisfactory conditions would likely lead to immediate action being required on site to deal with the issue.

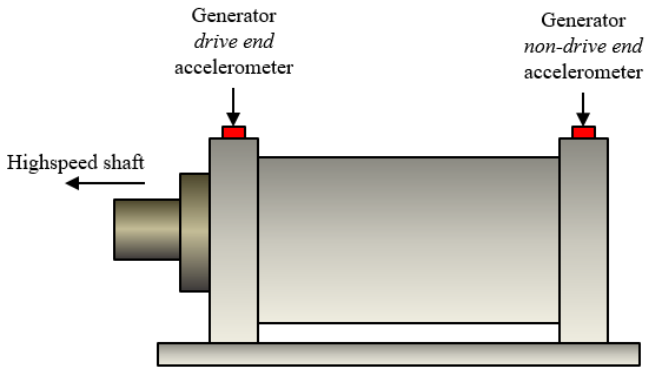


Figure 5.2: Diagram showing estimated positions of accelerometers used to measure generator bearing vibration.

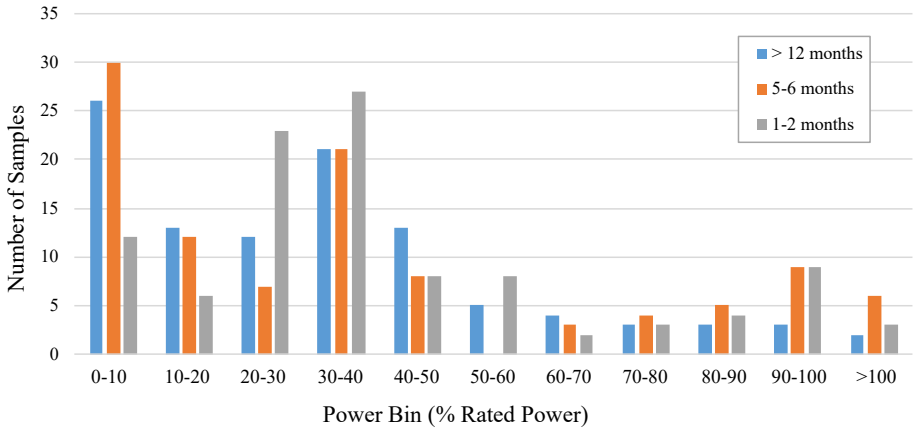


Figure 5.3: Vibration sample acquisition conditions

5.3.3 Variation in Operating Conditions

Vibration samples were stored and gathered 1 week apart across all WT operating conditions. Figure 5.3 shows the spread and frequency of power output at the time each vibration sample was taken, grouped into the three times outlined in Table 5.1. This shows the majority of vibration samples were collected while the WT was operating below 50% of rated power, with a small proportion actually at rated power. With the exception of one power and time bin combination (1-2 months, 50-60% Rated Power), all other combinations have at least one corresponding sample, allowing for a full spread of operational conditions at various times to failure to be analysed.

Table 5.1: Pool of data available for analysis

Wind Turbine No.	**At least 1 year before failure**	**5-6 months before failure**	**1-2 months before failure**
1	10 samples	9 samples	7 samples
2	None	None	4 samples
3	9 samples	7 samples	8 samples
4	8 samples	7 samples	8 samples
5	10 samples	10 samples	7 samples
6	9 samples	8 samples	8 samples
7	8 samples	7 samples	7 samples
8	8 samples	8 samples	8 samples
9	9 samples	9 samples	6 samples
10	None	9 samples	7 samples
11	None	9 samples	7 samples
12	9 samples	9 samples	5 samples
13	8 samples	8 samples	6 samples
14	None	None	7 samples
15	7 samples	9 samples	7 samples
Total	**95 samples**	**109 samples**	**102 samples**

5.4 Modelling Framework

As alluded to earlier, the research presented in this chapter draws upon synchronised databases of generator bearing vibration time series and failure events from a wind turbine OEM. This allows multiple vibration signal examples of the same failure mode in different turbines at a number of time intervals leading up to failure to be analysed and compared.

Previous studies have shown that operating parameters at the time of vibration measurement can have a substantial effect on vibration spectra, and hence the features which are extracted to detect the fault [98, 99]. To help mitigate these issues vibration samples can be binned based on power output at the time of measurement, a technique widely adopted by developers of condition monitoring software today. Expanding on this idea of grouping vibration data, this paper presents a two-stage methodology to predict generator bearing failure 1-2 months before occurrence. The first stage will use

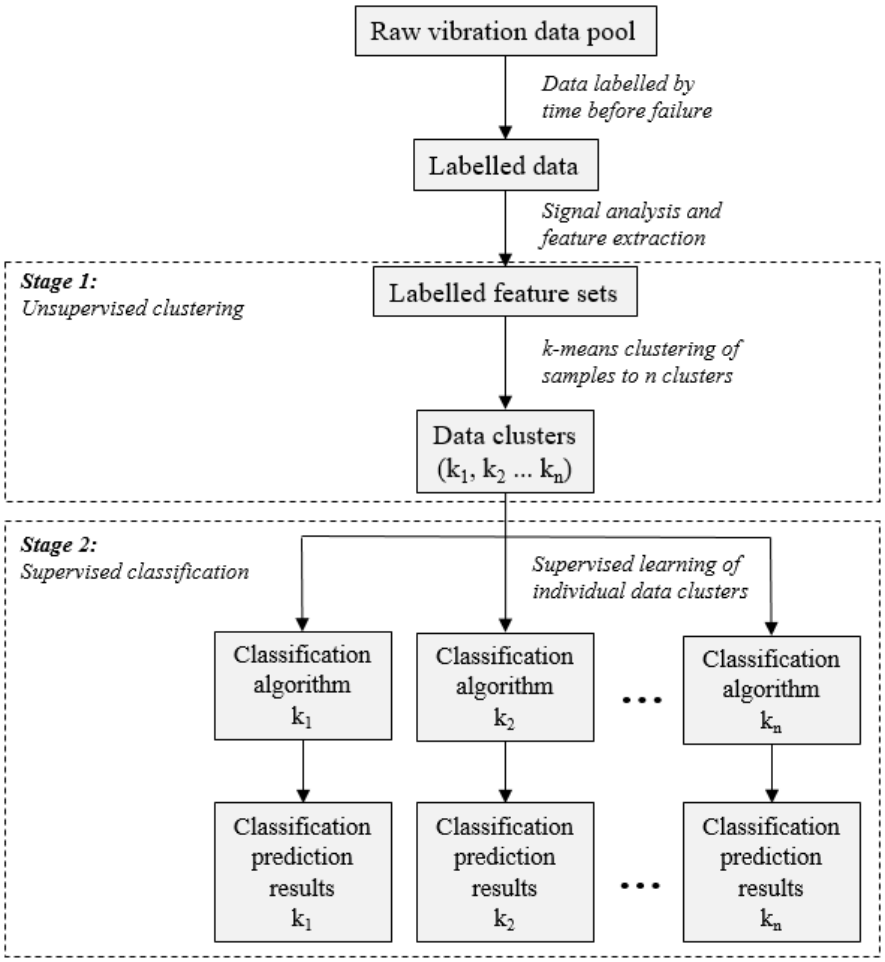


Figure 5.4: Overall framework for predicting failure

k-means clustering to group data by operating conditions, which will act as an advancement to existing power binning techniques by considering more parameters to define measurement sample bins. The second stage uses decision trees and support vector machines for feature classification. K-means clustering separates data based on Euclidean distance from a set number of cluster centroids, therefore does not discriminate clusters based on sample size. Both Decision Trees and SVMs are well established techniques, with the former scaling well to larger groups and the latter known to perform better with smaller groups [41, 100].

Table 5.2: Time-domain features

Feature No.	Feature	Formula
1.1	Maximum	$x_{max} = max\{x(t)\}$
1.2	Minimum	$x_{min} = min\{x(t)\}$
1.3	Mean	$u_x = \frac{1}{T}\int_0^T x(t)dt$
1.4	RMS	$x_{rms} = \left[\frac{1}{T}\int_0^T x^2(t)dt\right]^{1/2}$
1.5	Standard deviation	$\sigma_x = \sqrt{\frac{1}{T}\int_0^T [x(t) - u_x]^2 dt}$
1.6	Kurtosis	$\beta = \frac{1}{T}\int_0^T [x(t) - u_x]^4 dt$

5.5 Feature Extraction

As with diagnostics, most of the experience to draw upon to analyse vibration in generators and other rotating machinery for prognostics purposes comes from industries which utilise large, fixed speed synchronous machines. Modern wind turbines employ variable speed control strategies and, along with the stochastic nature of the wind, produce load patterns that are far more varied than traditional generators. The analysis of such vibration signals is therefore more challenging and as such, makes diagnosing faults and making a prognosis in wind turbine generators more difficult. Although spoken about previously, it is worth mentioning some of the techniques in the context of prognostics.

5.5.1 Time Domain Analysis

Time domain analysis offers a simple method of leveraging basic statistical analysis techniques to provide important information about the signal. When it comes to detecting high levels of abnormal vibration which may be the case in cases of severe damage it can be particularly useful. Although not suitable to diagnose any particular issue, it can certainly be useful information when used alongside other features of vibration. Table 5.2 shows the features that were extracted and used to analyse the signal, where $x(t)$ is the acceleration at time t, x_{max} and x_{min} are the maximum and minimum measured acceleration, u_x, x_{rms}, σ_x and β are the mean acceleration, RMS, standard deviation and kurtosis over signal length T.

5.5.2 Fourier & Order Analysis

Again the fastest and most widely used technique for analysing vibration signals is through the Fourier Transform, which represents the signal in the frequency domain, in which spectral peaks and signal energy can be analysed as approaching failure [101].

As with previous diagnostics, the vibration spectra gathered for this analysis had non-stationary raw time-domain signals, meaning the generator shaft speed varied during acquisition. To mitigate the spectral smearing that this produces order tracking was again utilised, which was discussed in Section 3.2.4. Table 5.3 shows the features that were extracted and used to analyse the signal using this technique. The spectral peak frequencies were chosen to reflect amplitude gains expected for a bearing fault such as this when severe damage is present causing relative movement within the internal assembly of the generator [97, 99, 102, 103].

Although all WTs in the dataset were the same size and topology, they did not all use the same bearing manufacturer and model. This meant that the ball passing frequency for the inner and outer races (BPFI and BPFO), as well as the rolling element deterioration frequency (BSF) could not be calculated consistently for each WT. For this reason only frequencies which were common to all WTs were used in the analysis, which included synchronous vibration at rotor speed fundamental frequency (Order No. 1) along with the 1st (Order No. 2) and 2nd (Order No. 3) harmonics. Non-synchronous vibration indicators were also extracted at order numbers 0.5, 1.5 , 2.5 and 3.5. While a simpler approach could have been taken such as calculating the RMS of a broadband frequency range, using separate amplitudes allowed individual frequencies to be isolated and used as individual features.

5.5.3 Operational Characteristics

Operational characteristics are used as a method of understanding the loading conditions at the time of vibration measurement. The features that were extracted to represent this are shown in Table 5.4 and were used in two instances; first of all in the baseline model to assist with training the classification algorithms, and secondly in two-stage model used independently from classification during the clustering stage. A

Table 5.3: Order-domain features

Feature No.	Feature	Description
2.1	$A_{0.5}$	Amplitude of peak at order number 0.5
2.2	A_1	Amplitude of peak at order number 1
2.3	$A_{1.5}$	Amplitude of peak at order number 1.5
2.4	A_2	Amplitude of peak at order number 2
2.5	$A_{2.5}$	Amplitude of peak at order number 2.5
2.6	A_3	Amplitude of peak at order number 3
2.7	$A_{3.5}$	Amplitude of peak at order number 3.5

Table 5.4: Operational characteristics

Feature No.	Feature	Description
3.1	P_{out}	Average electrical power out during 10s signal
3.2	Ω_{gen}	Average shaft speed during 10s signal
3.3	τ_{gen}	Average electromagnetic torque during 10s signal, $\tau_{gen} = P_{out}/\Omega_{gen}$
3.4	U_{wind}	Average wind speed during 10s signal from wind turbine anemometer

wind turbine control strategy is defined on a torque-speed curve while the operating performance is indicated by the power output in relation to wind speed. By considering these 4 features a representation of both relationships can be achieved.

5.6 Classification Models and Prediction Algorithms

5.6.1 Baseline Classification Model

The baseline classification model was trained using classification algorithms with all available data regardless of turbine and operating conditions at the time of vibration measurement. Both support vector machine (SVM) and decision tree classifiers were tested, with decision trees proving to be best suited to the baseline model, likely due to their ability to scale easily to large data populations. Once algorithms were trained they could be validated using 5-fold cross validation (see Section 3.6.4 for further details from previous study). This provided a baseline to which the two-stage methodology

could be fairly compared to evaluate the effectiveness of clustering data by operating conditions to produce sub-populations of data.

5.6.2 Two-stage Classification Model

For the two-stage model, as shown in Figure 5.4, features were clustered using k-means clustering to create groups containing vibration samples taken at similar operating conditions. The same two classification algorithms as described above were then trained on features within each cluster, with the best algorithm chosen based on performance. The final algorithm used for each cluster therefore changed with the number and spread of samples in that particular group. Once algorithms were trained they could be validated using 5-fold cross validation, before the chosen algorithm was exported for each cluster. Once exported the chosen algorithm could be used for any unseen data belonging to any particular cluster. Testing SVM's and decision trees provided the opportunity for high levels of accuracy across clusters containing both small and large quantities of data, which will be discussed in more detail when presenting results in Section 5.7.

5.6.3 Prediction Algorithms

The classification algorithms described in this Section are used across both the baseline and two-stage model, while clustering is only required in the latter. All algorithms are widely known techniques and can be readily applied, as discussed throughout the literature review. There are many types of common clustering algorithms which can find hidden patterns in large data sets including hierarchical clustering, k-means clustering, hidden Markov models and gaussian mixture models, however for this chapter k-means clustering was chosen due to its ease of application, fast computational speed and ability to manually choose the number of clusters [94].

Once features were successfully grouped, classification algorithms were then trained and tested based on the validation process outlined in Section 4.6.4, with the best chosen and applied to the set of features specific to that cluster. There are a variety of classifiers available that use supervised learning processes to classify data. For this research decision trees and support vector machines with polynomial kernel were chosen.

Decision trees scale well making them suitable for datasets with a large number of samples, while SVM's typically are more suitable for datasets with a smaller number of samples. In such cases however, they provide an opportunity for more complex decision boundaries to be established [94]. It is worth quickly reiterating the importance of model validation and correctly setting hyperparamters to limit overfitting, especially when it comes to decision trees. This has been well documented in both the literature review and throughout Chapter 3, therefore will not be expanded upon in this chapter, however for the interested reader more information found about their application in [41, 90].

5.6.4 Validation of Results

Similar to previous chapters, final model inputs were chosen by examining and refining the available features in different ways; first of all by expert knowledge, secondly by a model based approach (such as using random forest impurity score), and thirdly by an iterative method of using different permutations of features to discover which could best be used to produce the greatest overall accuracy. For each labeled vibration sample leading up to failure the features described in Tables 5.2 to 5.4 were extracted and used. These were then clustered with data in each cluster according to Table 5.5. To ensure the algorithm is trained in a balanced manner, the same number of data points were chosen from each class, with random samples chosen for the class with more samples in any particular group. For example, if one cluster contained 40 healthy samples and 32 unhealthy samples, 32 samples for each health class were used to train the model. Cross-validation was used to determine the overall accuracy of the algorithm. This method involves partitioning the data into subsets of a predetermined ratio, one of which is then omitted from training and used to test the algorithm. For this example 20% of the data was used for validation purposes. The process is then repeated using different sup-populations and an average accuracy calculated to use as a performance indicator. The prediction process can then be evaluated using a confusion matrix giving correct/incorrect classification and the likelihood of false positives/ negatives.

5.7 Results & Discussion

5.7.1 Baseline Classification Model

The baseline model using all available data had a maximum classification accuracy of 67% using a decision tree classifier. The confusion matrix is presented in Figure 5.5, which shows the likelihood of false positives or negatives when classifying the bearings as being of satisfactory condition or requiring action by having an expected RUL of 1-2 months before failure. The data population used for both training and validation was limited only by the number of available class 1 examples. The results shown here used 100 random vibration samples from each health class and regardless of the random population of data used, the accuracy never increased beyond 67%, with the mean accuracy just over 64% for this particular classification algorithm. The fact that accuracy changes with training data highlights the limitations of this approach and suggests that a more robust methodology is required to choose data more carefully for supervised training.

The results above used all order domain features described in Table 5.3, along with the time series RMS, standard deviation and kurtosis (features 1.4-1.6) in Table 5.2. For the baseline model, operational characteristics outlined in Table 5.4 were also utilised for classification.

5.7.2 Two-stage Classification Model

Clustering Results

K-means clustering partitions data using an iterative process initiated with a random guess of centroid positions. This means that although the number of clusters can remain constant, different data clusters are obtained every time the algorithm is run with a different seed. A sensitivity study performed using 10 random seeds with 5 clusters showed that 96% of the data points were consistently grouped into a particular cluster. Results will focus on 5 clusters, chosen to provide a balance between separating the vibration samples based on operating conditions while still having enough data in each cluster to perform classification and validation. Using 5 clusters also manages to cover

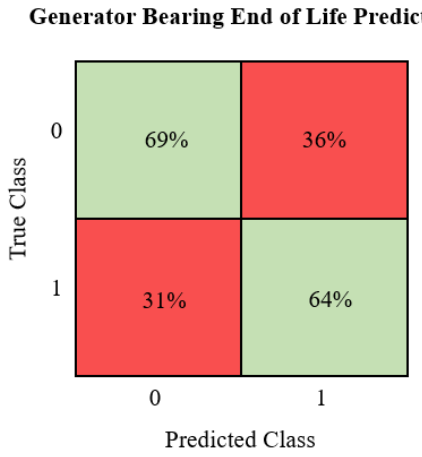


Figure 5.5: Confusion matrix of trained algorithms for all data (100 class 0 and 100 class 1 samples).

all parts of the torque-speed operating curve including important transition points. The effect of number of clusters on classification accuracy will be explored in greater detail throughout the results section.

Samples were clustered based on three properties which describe the operating conditions at the time the signal was taken; average electromagnetic torque (feature 3.3), wind speed (feature 3.4) and generator speed (feature 3.2). The average shaft speed and torque gives a good indication of loading conditions on the generator shaft at the time of measurement, while using wind speed allows for an understanding of any de-rating or adjustments from the normal torque speed curve operating points. Only 3 out of the 4 parameters described in Table 4 were required due to the direct relationship between power and torque. For this analysis using electromagnetic torque provided more consistency in the clustering process. The clusters are described qualitatively in Table 5.5, while Figure 5.6 shows the data clustered into the 5 groups to assist with visualisation. Each point in Figure 5.6 represents the coordinates of the chosen operation characteristics used for clustering for a single vibration sample. Note that these characteristics are common to both the drive-end and non drive-end sample. The centroid position is also given in Table 5.5, which describes each cluster quantitatively with respect to the normalised wind speed, electromagnetic torque and generator speed.

Table 5.5: Cluster descriptions

Cluster No.	**Data in cluster (%)**	**Centroid position** $(\tau_{gen}, U_{wind}, U_{gen})$	**Cluster description**
1	15.0	(0.093, 0.19, 0.61)	low torque, low wind speed, low gen speed
2	4.2	(0.23, 0.25, 0.72)	low-med torque, low-med wind speed, low-med gen speed
3	20.5	(0.27, 0.31, 0.87)	medium torque, medium wind speed, medium gen speed
4	41.0	(0.38, 0.38, 0.99)	medium torque, rated wind speed, high gen speed
5	19.3	(0.83, 0.63, 0.99)	rated torque, rated wind speed, high gen speed

Table 5.6: Results for 5 clusters

Cluster No.	**Algorithm**	**Overall accuracy (%)**	**Difference from base model (%)**
1	Decision Tree	61	-6
2	SVM	70	+3
3	Decision Tree	81.6	+14.6
4	Decision Tree	78.6	+11.6
5	Decision Tree	65.2	-1.8

Classification Results

The results presented first of all for the two-stage classification model use the 5 clusters of data shown in the previous section. Each cluster goes through a training and validation process for each of the classification algorithms as described in Section 5.6.3. Table 5.6 presents the best algorithm and accuracy for each of the clusters, along with the percentage accuracy difference from the baseline model described in Section 5.7.1, for which all data was taken and classified without any clustering.

Two of the groups showed a significant increase in accuracy with a maximum of up to 81.6% for group 3. These two groups (3 and 4) accounted for 61.5% of all data,

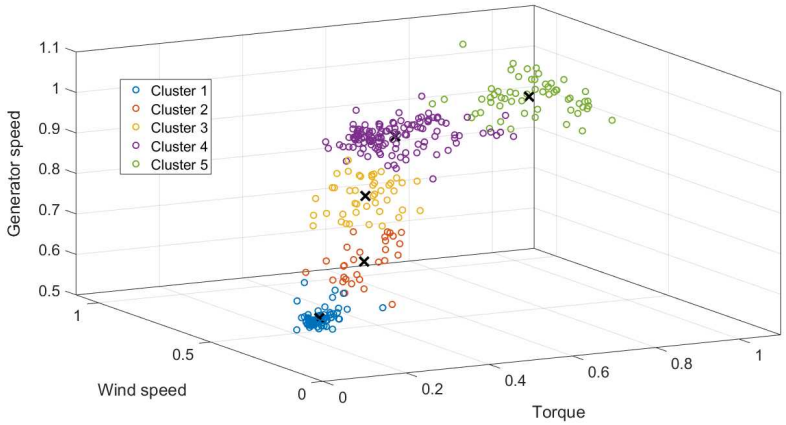


Figure 5.6: Plot of clusters based on three chosen variables, normalised to ensure confidentiality. 'X' represents each cluster centroid position.

while group 2, which saw a comparatively modest rise of 3% in accuracy accounted for a further 4.2%. Interestingly, the two groups (1 and 5) which underperformed in comparison to the baseline model had the 2 extreme datasets of either very high operating conditions or very low. This suggests that if the operating conditions are too low at the time the vibration sample is taken it becomes more difficult to distinguish high levels of damage as the bearing fault progresses. Alternatively, if the WT is operating at rated power with rated torque, shaft speed and wind speed, this also holds true. This indicated that optimal operating conditions may exist to identify faults and monitor progression. The algorithms perform much better using features from mid-range operating conditions, with the highest accuracy when in and around rated wind speed. Figure 5.7 shows the overall accuracy of the algorithms for each group in comparison to the baseline, with the groups ordered based on the extremity of operating conditions at the time the vibrations sample was taken.

In terms of features analysis, the dominant frequency (the point at which the most increase in amplitude was observed to indicate the fault) changed between turbines, therefore all order domain features were used in every case. The RMS, standard deviation and kurtosis proved to be the most significant of the time domain features, with accuracy generally increasing if at least one of these three features was used.

To compare absolute accuracy of both approaches, an overall weighted accuracy for

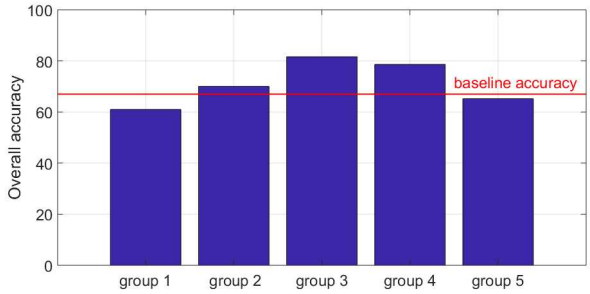


Figure 5.7: Plot of accuracy for each cluster ordered by extremity of operating conditions at time of vibration sample.

the two-stage model can be calculated by;

$$\gamma_{overall} = \sum_{i}^{n} \gamma_i \delta_i \tag{5.1}$$

where $\gamma_{overall}$ is the overall weighted accuracy, γ_i is the cluster accuracy and δ_i is the percentage of data in cluster i for n number of clusters. For the 5 clusters in this analysis this gives an overall weighted accuracy of 73% for the two-stage methodology considering all data. If groups 1 and 5 are excluded, the weighted accuracy of the remaining groups increases to 79%, and represents all data in mid-range operating conditions.

Looking at the best performing algorithms in Table 5.6, decision trees consistently outperformed SVMs in the 4 groups with the highest number of data samples. As expected, SVMs performed better in group 2, which had significantly less data. Having the option of both algorithms proved to be a useful method to ensure the highest accuracy was achieved for each group of varying size.

Considering now the top two performing groups in more detail, Figure 5.8 shows the confusion matrix for each training and validation phase. Results indicate a balance between predicting false-positives and false-negatives, with each group having slightly less misclassifications for class 1 samples to that of class 0. Fault propagation times will differ for every wind turbine used in the study, therefore the features extracted as fault indicators will also change in time leading up to failure. Classifying all data 1-2 months before failure into a single class is one of the limitations surrounding this

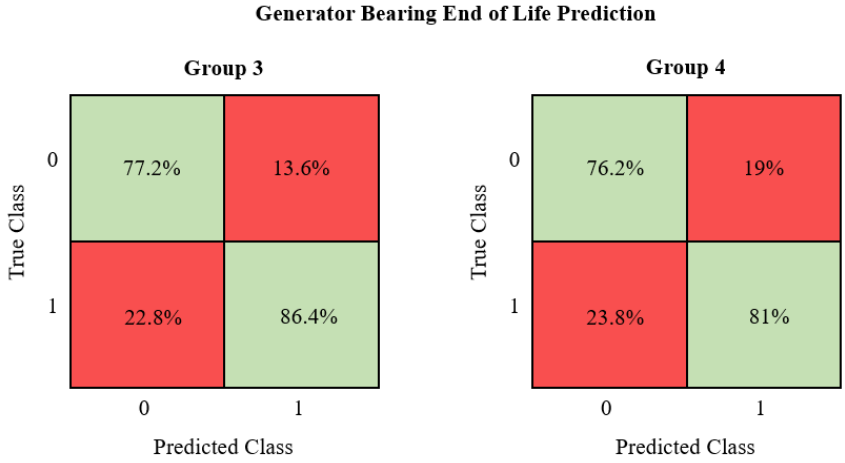


Figure 5.8: Confusion matrix of trained algorithms for group 3 (left) and group 4 (right).

type of classification process, which considers multiple examples without making an allowance for the fact each bearing has its own unique propagation time and failure threshold.

Effect of Number of Clusters on Classification Accuracy

Earlier results showed the two-stage classification model with 5 clusters as this provided a balance between the number of clusters and number of data within each cluster. For completeness, results for other cluster numbers will also be provided to showcase the variation and importance of correctly choosing this value. If 3 or 4 were chosen an increase in accuracy at mid range operating conditions still occurred, however the same level of improvements were not observed. Tables 5.7 and 5.8 show the breakdown of clusters and accuracy along with the associated cluster centroid positions and algorithms used. Considering the weighted accuracy described in equation 5.1 this can be calculated for 3 clusters as 69.8% and 69.1% for 4 clusters. If 6 or more clusters were chosen, groups started to get too sparse to perform classification on every cluster, (as even with 5 clusters group 2 consisted of only 4.2% of the data) and accuracy generally decreased. If more data was available, it would be worth revisiting the number of clusters to determine the optimal range.

Table 5.7: Results for 3 clusters

Cluster No.	Data (%)	Centroid	Algorithm	Overall accuracy (%)
1	31.1	(0.13, 0.21, 0.64)	Decision Tree	62
2	50.1	(0.34, 0.36, 0.96)	Decision Tree	78.2
3	18.8	(0.81, 0.62, 0.99)	SVM	60.4

Table 5.8: Results for 4 clusters

Cluster No.	Data (%)	Centroid	Algorithm	Overall accuracy (%)
1	30.3	(0.13, 0.20, 0.63)	Decision Tree	62.5
2	43.6	(0.31, 0.35, 0.95)	Decision Tree	79.7
3	14.5	(0.60, 0.48, 0.99)	SVM	63.0
4	11.6	(0.90, 0.68, 0.99)	SVM	53.6

Comparison of Clustering and Power Binning

Up to this point clustering has been compared only to the baseline model, which considers all available data. Clustering offers the opportunity to group vibration data using a variety of operating parameters, and results have shown improved accuracy in doing so, however, it is also important to compare this approach to standard power binning practices. A similar methodology was used as previously describe in Figure 5.4, but instead of clustering in stage 1 data was divided into 5 even bins based on power output only. Each bin was then classified individually, with results showing again an improvement to the baseline model, but not reaching the level of accuracy achieved through clustering. Results are presented in Table 5.9, where mid-range operating conditions again provide the highest prediction accuracy.

5.8 Conclusion

Predictive maintenance strategies that use previous failures to learn and predict failure and remaining useful life of components in different wind turbines have the potential to make substantial savings to costs associated with O&M. This research indicates

Table 5.9: Results by power binning

Bin No.	**Power range (%)**	**Data in bin (%)**	**Algorithm**	**Overall accuracy (%)**
1	0-20	18.3	Decision Tree	63.6
2	20-40	45.8	Decision Tree	68.2
3	40-60	15.8	Decision Tree	71.1
4	60-80	5.9	SVM	67.1
5	80-100	14.2	SVM	55.2

that machine learning classification algorithms can be applied to specific features to successfully predict generator bearing failure 1-2 months before occurrence with an accuracy of up 81.6%. To achieve this level of accuracy operating conditions at the time each vibration sample was taken must be considered independently of classification, as accuracy falls significantly to 67% without it, as shown in the baseline model. This finding suggests that an optimal condition monitoring zone may exist in which to more accurately track faults for prognostic purposes, with the same logic also applying to diagnostics. A simple approach to improve baseline accuracy is manual or rule-based removal of data at extremes of the power curve, however it is difficult to do this robustly and fairly. Binning vibration samples by power output to create groups which can be fairly compared is standard practice, however this paper has demonstrated that k-means clustering is a more successful method by considering more operating parameters prior to classification.

This work shows that the operating conditions at the time of vibration measurement can greatly influence the ability to detect a fault, with mid-range operating conditions leading to highest accuracy when predicting remaining useful life based on vibration analysis for this particular wind turbine and failure mode. The examples shown find a balance between false-positives and false-negatives, meaning that equal weight is given to either case however from a commercial perspective it may be worth refining the model to either minimise or maximise false-positives, which could be further explored through cost-benefit analysis.

The two-stage methodology described in this paper provides a scalable prediction

model which could be applied to a range of faults across the drivetrain. By considering different features of vibration that can effectively describe high levels of damage of a particular fault, which could also rely on different vibration measurements and analysis techniques to those used in this paper, the model could be adapted accordingly.

Furthermore, this chapter only uses vibration features common to all bearing specifications. Improvement in accuracy could be made by bringing in additional fault frequencies specific to each bearing type if all bearing kinematic details are known. Conversely, in practice keeping kinematic details up to date is time consuming for engineers and often details are not known as wind turbine components are replaced and wind turbines move between owners and asset managers. Providing accurate failure prediction without requiring this specific information is hugely valuable as a site gets older.

5.9 Future Work

There are several areas that future work could be done to enhance the research presented in this chapter. First of all, more specific bearing fault frequencies could be used in conjunction with rotational frequencies to assess damage and predict failure. It would be interesting to understand to what extent this increases accuracy or allows for earlier prediction as more specific indicators of the fault progression are modelled. Getting more continuous timestamps of vibration data leading up to failure would also allow for a more robust prognosis and model validation. This would allow for a much more granular look at how vibration features develop through to failure.

Furthermore, if SCADA data is also gathered over the same time period leading up to failure, identifying and comparing other metrics to classify or predict failure would be an interesting study. For example cumulative energy yield or time at rated power before failure could be used instead of a basic measure of time. This may act as a better metric as it takes into consideration operational condition as well as time, which is more closely related to experienced vibration and fatigue.

Chapter 6

Analysis of the Cost Impact of Maintenance Strategies Enabled by Modern CMS for Offshore Wind Farms

6.1 Chapter Contribution

Modern condition monitoring systems and data analytics enable wind farm owners and operators the opportunity to gain further insight into wind turbine operation and component health. Emerging predictive and condition-based maintenance strategies could potentially reduce the LCOE by reducing money spent on wind farm O&M. That being said, there is actually very little research that attempts to quantify such strategies over the lifetime of a wind farm. This chapter will focus on offshore wind farms, where it is believed that the largest gains can be made by employing improved maintenance strategies. More specifically, this chapter will explore the cost associated with different replacement strategies once a fault has been detected. Analysis will take into consideration potential consequential damage if an asset is kept running longer while a fault is present, as well as the costs associated with early repair.

This chapter aims to answer the following research question:

"How can implementing advanced monitoring and predictive maintenance strategies affect O&M costs throughout the life of a wind farm?"

The contributions of this chapter are as follows:

1. Provide insight into the cost associated with minor repairs and major replacements of the generator and gearbox for large offshore wind farms.
2. Determine the costs associated with implementing both predictive and condition-based maintenance strategies in comparison to a more reactive maintenance strategy.
3. Weigh up the cost benefit between reaching maximum achievable design life and replacing a component too early when implementing such strategies.

6.2 Maintenance Strategies

Maintenance strategies have evolved over the last 2 decades in line with increased turbine size and reduction in technology costs. With sites now reaching for higher average wind speeds in more remote geographical locations (particularly offshore), it is imperative that O&M costs are as low as possible to keep the LCOE down. One aspect of achieving this is employing more optimal maintenance strategies, made possible with increased monitoring capabilities, improved digitalisation and more in depth analysis of data. This allows engineers and operators to assess asset performance, understand reliability and make informed maintenance decisions that can drive down costs over the lifetime of the site.

6.2.1 Types of Maintenance Strategies

There are several terms observed in literature to describe the different maintenance strategies, which are not always exactly the same. For clarity Table 6.1 defines each strategy, as used in this chapter with respect to component replacement. Four strategies

Table 6.1: Maintenance strategy definitions

Maintenance Strategy	**Description**
Preventative	Routine maintenance to minimise the risk of faults developing.
Reactive	Maintenance performed retrospectively after component failure.
Predictive	Prognosis is performed after fault is detected with replacement scheduled accordingly based on availability of resources and site conditions.
Condition-based	Ongoing assessment performed once a fault has been detected and maintenance is performed when condition has worsened to a set threshold.

are described including preventative, reactive, predictive and condition-based, however it will be the latter two that will form much of the analysis.

Preventative maintenance is the act of performing routine maintenance that will minimise the risk of faults developing in the first place. This type of strategy is inefficient, does not utilise a components full design life and is wasteful of both materials and technician time performing the work. It allows for a less hands on approach to condition monitoring, meaning costs can be saved by not requiring advanced monitoring hardware and associated ongoing monitoring.

Reactive maintenance allows components to run to failure, making use of there entire design life, but risking other components health at failure. This type of strategy reduces cost of routine O&M while providing a hands off approach to condition monitoring, however, means more expensive replacements and downtime while waiting for spare parts and access delays.

Predictive maintenance and condition-based maintenance are similar in nature. They both require condition monitoring to detect faults and engineers to assess component condition and make a prognosis to some extent. Predictive maintenance involves acting on that prognosis, and perhaps reassessing if the condition does worsen. The goal of this is to perform maintenance activity at as low a cost as possible within a suitable time frame. Some design life will be compromised for minimising risk of extended periods of downtime.

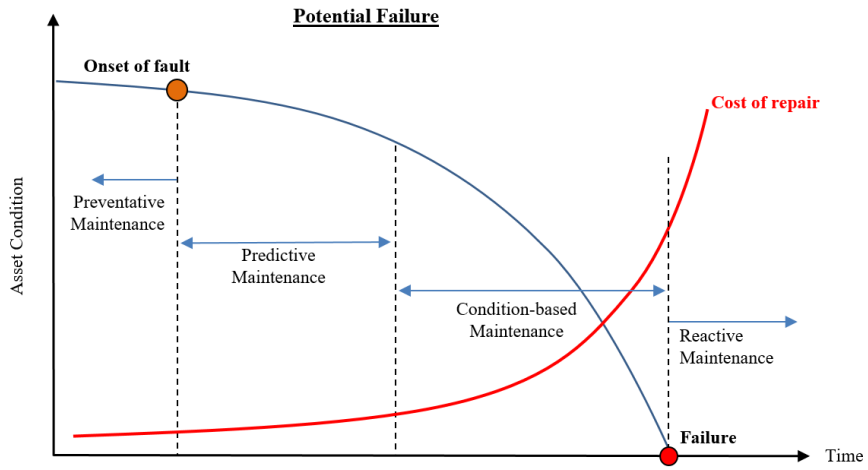


Figure 6.1: Cost of repair in relation to maintenance strategy

Condition-based maintenance aims to get extra life out of the component by continuously monitoring component health and setting hard upper limits for maintenance. Fault and hence damage progression is not always predictable, therefore there is added risk here of further damage to additional components or even assemblies within the drivetrain. This can increase the cost of maintenance activities further down the line. Figure 6.1 shows the relationship between asset condition and cost of repair for the four strategies outlined above.

6.2.2 O&M Decision Making

Decision makers have multiple factors to consider when assessing maintenance which include, but are not limited to: wind turbine availability contracts, weather forecasts, component lead times, technician availability, other work on site, turbine scheduled maintenance, O&M contracts, fault severity and prognosis recommendation.

Figure 6.2 shows a typical decision flow for making a diagnosis and prognosis of a major component. Note depending on asset management, O&M and condition monitoring contracts in place the decision maker may be different. This flow chart provides a general high level overview. From the initial alarm (vibration or SCADA), assuming it is not a false alarm, a diagnosis is made of fault type. Depending on the fault type and what conclusions can be drawn from the data, an inspection may be required

for clarification of damage level and location. If an inspection is required it will be planned and performed over a suitable timeline and inspection report used to validate observations in the data. From here prognosis will be performed with or without the inspection report depending on if it was deemed necessary. As discussed in previous chapters, prognosis can be made based on previous examples of faults and failures in a number of ways looking at vibration and temperature of the current fault in relation to similar historical cases. However, another factor to consider is the cost of different approaches, something that is often overlooked in industry as decisions are made without insight into how this may be affected over the lifetime of the asset. This analysis aims to address and provide insight into the costs associated with deciding whether an intervention (or replacement) should be planned and performed at this stage, or whether owners would be better served waiting and getting more life out of a component at the risk of further damage or running the turbine at lower loads.

For the purposes of this analysis, predictive maintenance would aim to plan for replacement at this stage, while a condition-based approach would seek to continue monitoring through higher set alarms. In order to fully appreciate these decisions cost must be analysed, taking into consideration all factors discussed at the beginning of this section. This chapter however will attempt to make a start at this by considering the effects of running a fault for longer through a condition-based maintenance scheme over earlier replacement when employing predictive maintenance.

6.3 Methodology and Model Overview

To ensure a robust analysis of availability and O&M costs for different maintenance strategies a number of hypothetical offshore sites were analysed with varying distances from shore and size (number of turbines). This allows changes in downtime, vessel costs, total repair costs and times to be analysed and compared. Data used for both wind and sea state was assumed to be the same for all hypothetical sites.

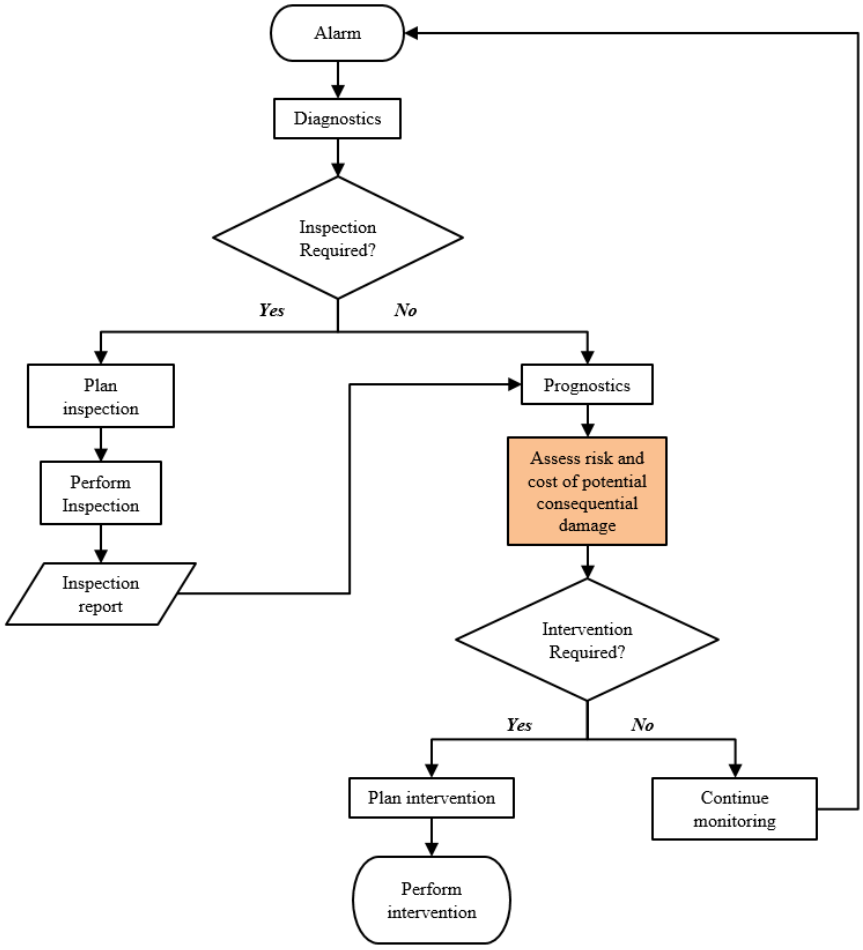


Figure 6.2: Flow diagram of engineering decision process.

6.3.1 Methodology

Figure 6.3 shows the step by step approach taken in this chapter to perform the analysis. As the quality of results obtained was highly dependent on setting the correct inputs for each case, time was first taken to become familiar with the model and adjustable parameters (the cost model utilised is described in more detail in Section 6.3.2). From here the baseline model parameters were chosen for each hypothetical site, as described in Section 6.3.3. Test cases were then defined that could synthesise each of the maintenance strategies. These are detailed in Section 6.3.4. Finally the models could be run, outputs analysed for each maintenance strategy across different sites and conclusions formed.

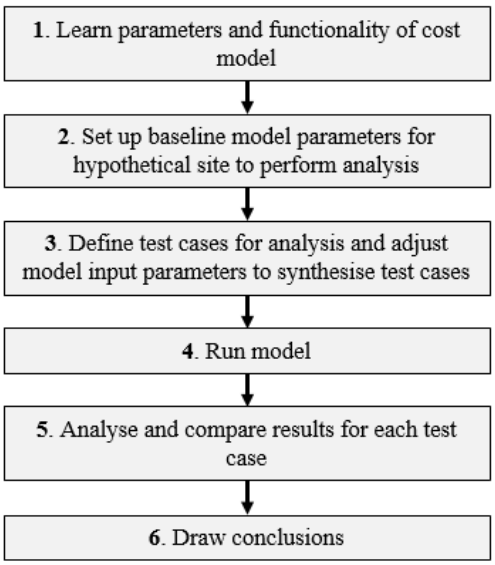


Figure 6.3: Overall approach to chapter analysis.

6.3.2 O&M Cost Model Description

The benchmarked O&M cost model used throughout this analysis was developed by the University of Strathclyde and is a time based simulation of operations of an offshore wind farm over its lifetime. To provide an overview, a Monte Carlo Markov Chain is used to model failure behaviour, with maintenance operations simulated based on both site conditions and available resources. Model input parameters are used to determine these constraints used over the lifetime of the asset operations. The model has the capability to calculate availability, turbine downtime, power production and maintenance resource allocation of the simulated wind farm, along with associated costs. Lost revenue is determined from the power produced at each simulated wind speed time step. Losses associated with electrical transmission and wind farm arrays are represented by efficiency coefficients, with the value of power produced determined by a combination of the market price of electricity and the value of UK support mechanisms. A constant price of electricity was therefore used throughout the simulation of 90 £/MWh. It should be noted that these costs may not reflect current prices, however are still valid to give a good indication of changes in price with differing strategies. The lost revenue cost due to maintenance is calculated using availability of the wind farm, which

is defined as the number of operational turbines divided by the total number of turbines. Wind distribution is characterised by a two parameter Weibull curve, however climate data more broadly consisted of wind speed, wave height and wave period, each of which had its own seasonal variability which influenced both generation capacity and accessibility over the year. The expected repair cost of each intervention category was provided by the industrial partner for the three-stage PMG FRC wind turbine configuration. This chapter will focus on analysing costs for both operations and maintenance of theoretical sites, however will discuss results in relation to overall availability, downtime and power production. A diagram describing model inputs and outputs, as well as the key features of the simulation loop can be found in Figure 6.4. Full details of all model input parameters and interdependencies can be found in [7, 104].

6.3.3 Theoretical Site Characteristics

Four different offshore wind farms were simulated of different sizes and distances from shore. Three distances of 25km, 50km and 100km were chosen to represent "near", "mid" and "far" offshore sites. To then understand the effect of wind farm size on costs two different sizes of 50 and 100 turbines were also chosen to represent "medium" and "large" sites. A three-stage PMG FRC turbine was utilised throughout the analysis, for which representative failure rates were gathered and used based on [105]. Although this technology is not representative of some of the larger turbines currently being installed offshore, it does offer an opportunity for more trusted reliability rates to be used calculated over a larger population of operational turbines. For all sites modelled it was assumed that each had the same weather and climate characteristics as per FINO climate data from an offshore research platform. The site chosen to represent the climate of all cases was located 45km offshore in the North Sea, and corresponds to both existing and future wind farms currently being developed. Although different sea state data may improve analysis authenticity by avoiding this assumption, having the same sea states allows for a direct comparison between distances and sizes to be made. The same wind speed is also used across each site, which as found in [105], is a valid assumption based on the observed mean wind speeds across 60 offshore wind farms in

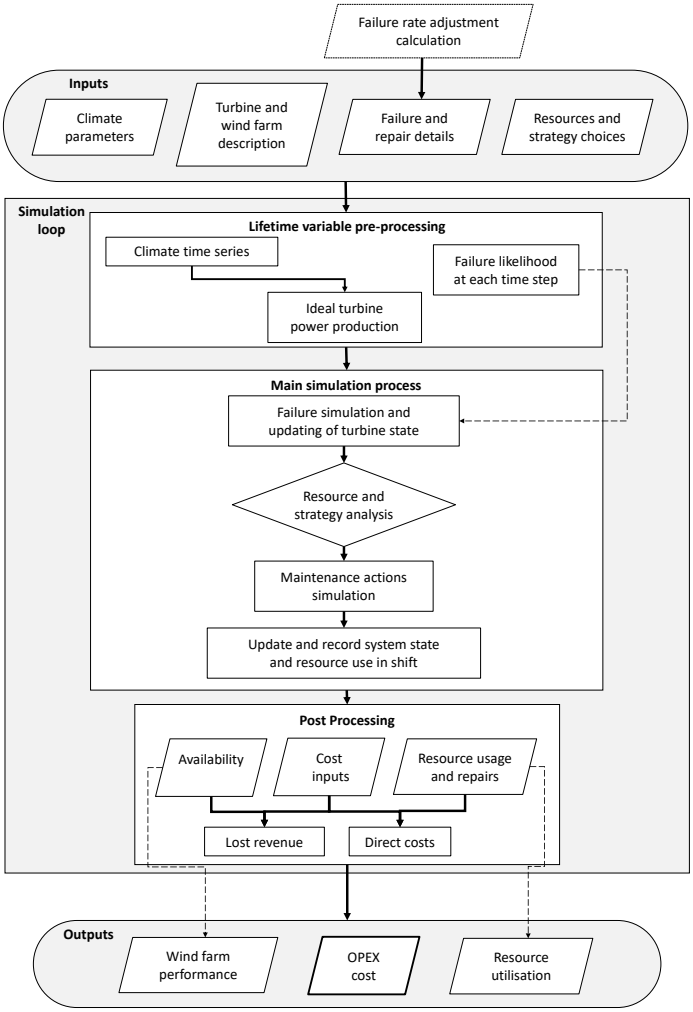


Figure 6.4: Simplified cost model structural overview, adapted from [7]

operation ranging from 1km to just over 200km from shore, which showed less than 2% deviation. Each wind farm simulated utilised a modern multi MW offshore wind turbine, however the exact power rating cannot be provided for confidentiality reasons.

6.3.4 Analysis Cases

Analysis cases have been designed in an attempt to synthesis scenarios to represent the maintenance strategies described above. Before describing how this is achieved, Table 6.2 details how intervention types are categorised for a generator and gearbox

Table 6.2: Intervention Categories

System	Intervention Category	Intervention Example	Jack-up Required?
Generator	Major Replacement	Full generator replacement	HLV
Generator	Major Repair	Slip-ring replacement	CTV
Generator	Minor Repair	Re-alignment	CTV
Gearbox	Major Replacement	Full gearbox replacement	HLV
Gearbox	Major Repair	Highspeed assembly replacement	CTV
Gearbox	Minor Repair	Oil system flush	CTV

within the cost model. Each intervention type has distinctive constraints and resource requirements which must be met when looking for suitable windows to perform the work. It should be noted that a major replacement requires a heavy lift vessel (HLV) in either case, while both major and minor repairs only require a crew transfer vessel (CTV). This is key, as by intervening earlier HLV costs can be avoided by performing a major repair rather than a major replacement.

The premise of this analysis is the underlying assumption that, as certain faults are left unaddressed, through time damage has the potential to worsen and spread throughout a component or assembly. This leads to more expensive repairs or replacements along with additional inspections over the life of the fault progression. There may also be higher refurbishment costs if the component is to be re-used. Conversely, leaving the component running fundamentally extends the useful life of the component, reduces the risk of premature replacement or repair with little to no observed damage or change after intervention. This will ultimately increase the mean time between failures for a turbine and site. In other words, each strategy has its own unique opportunity cost that must be considered. It should be noted that this analysis does not attempt to take into consideration O&M and asset management contracts (such as warranty periods) that may affect decision making.

First of all a baseline is determined for each of the scenarios described in Table 6.3 by using assembly level failure rates established in [105]. The assemblies used were ‘Generator’, ‘Gearbox’, ‘Converter’ and ‘Rest of Turbine’, each broken down into intervention categories ‘Major Replacement’, ‘Major Repair’ and ‘Minor Repair’. The

Table 6.3: Analysis cases

Scenario No.	**No. Turbines**	**Distance Offshore**	**Percentage Failures**	**Percentage RUL**
1	100	25km	10-40%	30-90%
2	50	25km	10-40%	30-90%
3	200	25km	10-40%	30-90%
4	100	50km	10-40%	30-90%
5	100	100km	10-40%	30-90%

failure rates per turbine per year for both the generator and gearbox were adjusted from the baseline as per the following equations:

$$MajorReplacement : \lambda_{adjusted} = (1 - P_f)\lambda_{baseline} \tag{6.1}$$

$$MajorRepair : \lambda_{adjusted} = (1 + P_f)\frac{\lambda_{baseline}}{P_{RUL}} \tag{6.2}$$

where $\lambda_{baseline}$ is the baseline failure rate representing a reactive maintenance strategy, $\lambda_{adjusted}$ is the adjusted failure rate representing the predictive or condition-based strategy, P_f is the percentage of failures per turbine per year that are diagnosed and repairable in line with the explanation of repair categories above, and P_{RUL} is the percentage of component remaining useful life that is utilised prior to repair (but before failure when major replacement is required). It is difficult to set a definitive number for P_f as it is impossible to understand the root cause of every major replacement and determine if the failure could have been avoided with better condition monitoring practices and earlier intervention. A range of P_f values between 10% and 40% are therefore used to account for this uncertainty. With regards to P_{RUL}, the baseline assumes a value of 100% as the component was run to failure for all major replacements as per the reactive maintenance strategy. For a predictive maintenance strategy, the lower the percentage the earlier the repair took place, ultimately leading to a higher rate of major repairs. Values of 30%, 50%, 70% and 90% were used for each scenario to understand the effects of earlier repair on the cost of maintenance activities, as described for each analysis case in Table 6.3. This gave a total of 16 different wind farms to be simulated

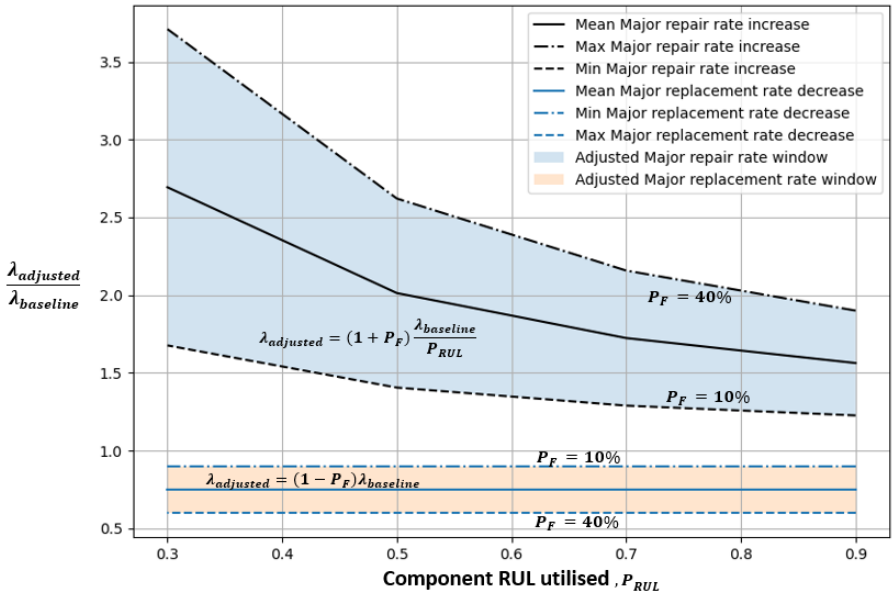


Figure 6.5: Adjusted failure rates.

per scenario along with the scenario baseline simulation.

Figure 6.5 shows the adjusted failure rate in relation to the baseline failure rate and P_{RUL} (for which $\lambda_{adjusted}/\lambda_{baseline} = 1$). For the range of P_f values described above, the major replacement rate is reduced as per equation 6.2, and is not dependent on P_{RUL}. Conversely, the major repair rate increases as per equation 6.1, with lower P_{RUL} values causing a larger increase in repair rate.

6.3.5 Baseline Failure Rates

The baseline failure rates used for the model can be found in Table 6.4, taken from [8], which provides the rates as the number of expected failures per turbine per year for each category. Although these were the categories used when setting up the model parameters, only gearbox and generator failures were considered and adjusted for the purposes of this analysis. This was done for both major replacements and major repairs as described above, with minor repairs remaining constant for every component/assembly category.

Table 6.4: Baseline failure rates [8]

Failure Category	Gearbox	Generator	Converter	Rest of Turbine
Major Replacement	0.154	0.095	0.005	0.11
Major Repair	0.038	0.321	0.081	0.622
Minor Repair	0395	0.485	0.076	5.222

Table 6.5: Baseline costs overview

Costs Category	£/MWh
Lost Revenue	13.14
Transport Costs	13.93
Staff & Repair Costs	3.80
Total O&M Costs	**30.87 ± 0.15**
Total Direct O&M Costs	**17.73 ± 0.098**

6.4 Results

6.4.1 Baseline Cost

First of all the baseline costs are considered, which for this analysis will form the reactive maintenance strategy described previously in Table 6.1 to compare against. Presented in Table 6.5 are costs, in £/MWh, of the total O&M costs broken down into lost revenue, transport, staff and repair costs. Total direct O&M costs are also given, which is the combination of transport, staff and repair costs. The uncertainty provided in the table is determined through running 5 Monte Carlo simulations with the same failure rates inputs, each made up of 25 random simulations and converging with approximately 0.15% accuracy.

6.4.2 Effects of Predictive and Condition-based Maintenance Strategies

The direct O&M costs with the failure rate adjustments can be found in Figure 6.6, which shows the costs in relation to the percentage RUL utilised as a percentage change from the baseline costs. Disregarding percentage RUL utilised for a moment, overall it can be seen that by taking a proactive approach to maintenance between 2% and

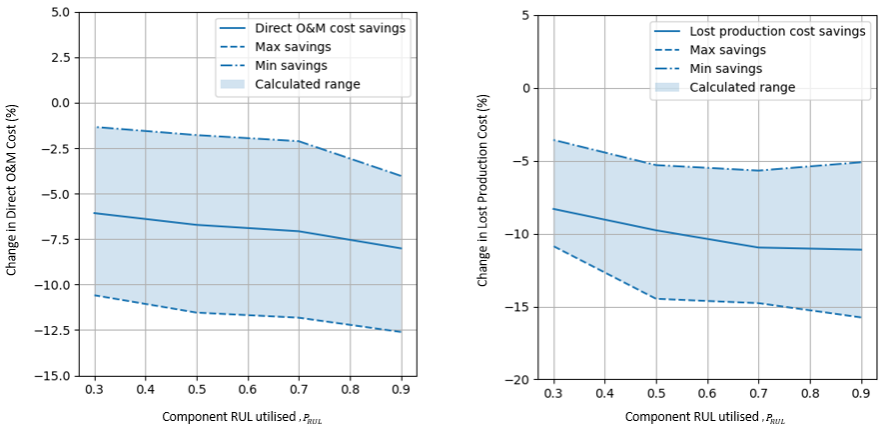


Figure 6.6: Percentage cost reduction of predictive maintenance strategies.

12.5% can be saved by repairing a component early before major replacement is required depending on the percentage of faults detected and acted upon. For the baseline case used this equates to a saving of approximately £0.355 - 2.128 per MWh over the lifetime of the wind farm. With regards to changes in lost production, also found in Figure 6.6, between 3.6% and 15.75% less energy is lost across the range of simulations.

If we now think about strategies in more detail, results suggest that by utilising as much component life as possible before repair an extra 1-2% can be saved on average in direct O&M costs over the lifetime of the wind farm. Over the lifetime of the wind farm this equates to approximately £0.177 - 0.355 per MWh when using the baseline rates presented in Table 6.5. There also appears to be a closer correlation between lost production and utilised life of the component, suggesting that a higher energy yield can be obtained over the site life by running components for longer before replacement (assuming a HLV is not required).

6.4.3 Analysis of Wind Farm Size and Location

To consider the effect wind farm size and distance from shore has on O&M cost reduction Scenarios 2-4 (see Table 6.3) were simulated. For each scenario a baseline cost was again calculated by running 4 Monte Carlo simulations with identical input parameters. As with scenario 1, baseline failure rates were used with each simulated site made up

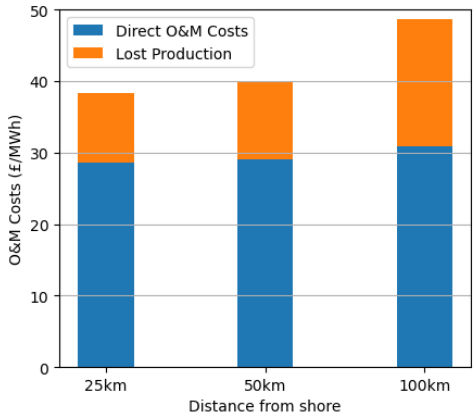


Figure 6.7: Baseline direct O&M costs and lost production, Scenarios 2-4.

of 25 random simulations. Figure 6.7 shows the baseline direct O&M costs for wind farms located 25km, 50km and 100km (Scenarios 2, 3 and 4 respectively), along with the expected lost revenue. Each Monte Carlo simulation had a convergence of below 0.1%. For each baseline we see a modest rise in direct O&M costs from approximately 28.6 at 25km to 31.1 £/MWh at 100km. The lost production however increases more substantially as the wind farm is located further offshore increasing on average from 10.7 to 17.7 £/MWh when moved from 50km to 100km offshore. This phenomenon was also found in [8], and can be explained through a restriction of resources applied in the cost model. This constraint meant that as the site is moved further offshore not enough technicians and CTVs are available to complete all repairs increasing overall downtime. The number of resources can be optimised to reduce this, however a relative comparison of strategies is then not possible due to higher base O&M costs.

Concentrating on direct O&M costs for each scenario the cost reduction in relation to the percentage RUL utilised was plotted in Figure 6.8, with (a), (b) and (c) representing scenarios 2, 3 and 4 respectively. As a whole, the calculated reduction in direct O&M costs are remarkably consistent across the different sites, suggesting that distance from shore has little impact on direct O&M costs savings for a medium sized wind farm. As a percentage of the baseline, results also suggest that similar savings can be made per MWh when compared to larger sites.

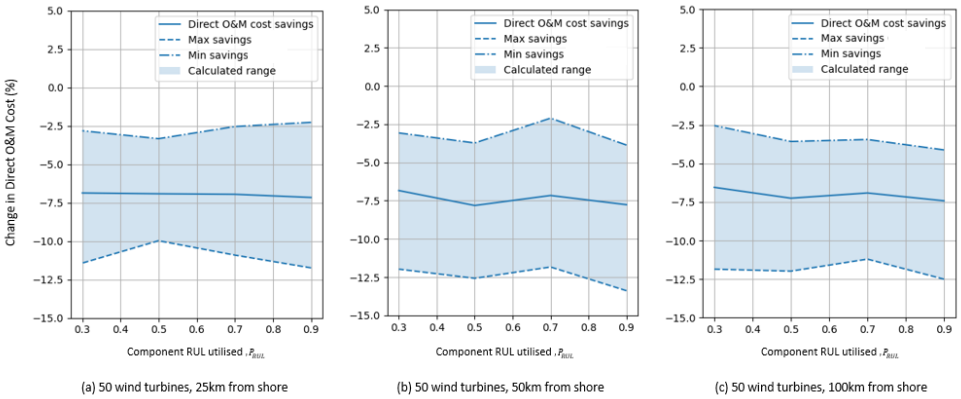


Figure 6.8: Percentage cost reduction of predictive maintenance strategies - comparison of sites (Direct O&M Costs).

Looking at lost production now in Figure 6.9 (a), (b) and (c), results show more deviation in potential savings with respect to distance from shore when compared to direct O&M costs, with values ranging from approximately 1.5% to 12% depending on P_{RUL} and P_f. One interesting observation is that results indicate cost reductions are not monotonic as a function of P_{RUL}, however, the notable increase at $P_{RUL} = 0.7$ can likely be attributed to stochastic variability in the simulation.

The notable decrease in saved lost production at 100km from shore, presented in Figure 6.9 (c), is due to the substantial increase in absolute lost production (see Figure 6.7), meaning that the percentage decrease relative to the absolute value will be lower. There is also stronger correlation between P_{RUL} and lost production in comparison to direct O&M costs, with greater values of P_{RUL} on average producing less lost production over the wind farm's 20 years operation. This trend is more obvious with lost production due to the direct link between turbine downtime and increased major repair rate. For direct O&M costs any additional CTV time due to increased major repair rate may not show as strongly due to the dominant effect of offsetting the cost of a HLV with a CTV.

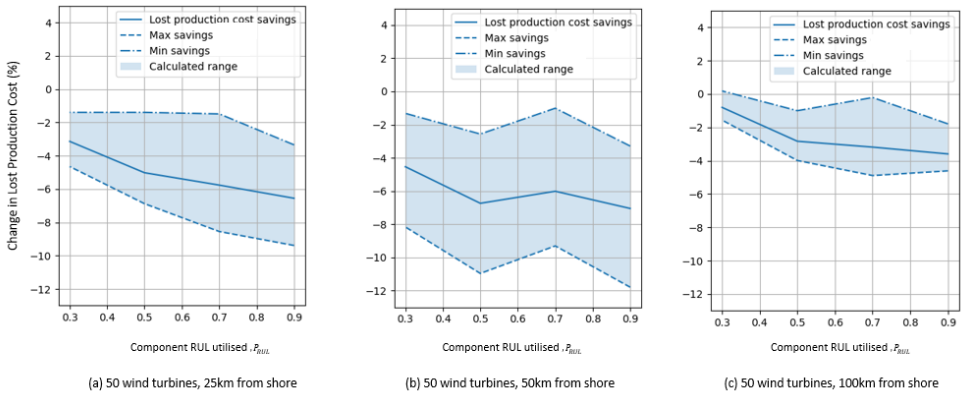


Figure 6.9: Percentage cost reduction of predictive maintenance strategies - comparison of sites (Lost Production Costs).

6.5 Discussion

The choice facing operators and owners with regards to O&M strategies is complex, and is often a reflection of the technology, monitoring capabilities and contracts in place at a particular wind farm. Regardless, this analysis aims to provide insight into how OPEX can be reduced over a wind farm life with alternative maintenance strategies assuming that advanced monitoring is possible to gain the required insight.

6.5.1 The Cost Impact of Advanced Maintenance Strategies

Two key parameters were analysed for a variety of sites; P_f to simulate a range of additional failures per turbine per year that are diagnosed early and are repairable without using a HLV, and P_{RUL} to simulate how early the component was repaired in relation to the expected remaining useful life before failure, at which time a HLV would be required. Considering direct O&M costs (transport, staff and repair costs), results indicate that identifying more faults and repairing them earlier is more important than utilising more of the component life, for which comparatively lower gains can be made. This highlights the importance of making informed maintenance decisions, as analysis shows that the risk of failure and requiring a HLV is much greater than repairing a component too early and compromising some design life. Due to the much higher cost of HLVs over CTVs, this is true even if statistically it means the same component

needs to be repaired several times over the wind farm life. In relation to maintenance strategies, this suggests that predictive maintenance may be the optimal solution for offshore wind farms under the conditions that an optimal weather window can be chosen to limit lost revenue and a HLV can be avoided by the early repair. Having said this, results also show that lost production can be further reduced over the wind farm operation lifetime by increasing P_{RUL} through a more condition-based approach, however gains are limited, therefore a trade-off between further cost savings and the risk of failure will have to be considered carefully. This decision will ultimately be fault case specific and require prudent diagnostics and prognostics.

6.5.2 Result Limitations

As with all results presented, it is important to reflect on limitations and the major causes of uncertainties within the analysis. To reduce uncertainty, failure and repair rates for the baseline reactive maintenance strategy have been taken from [8], for which rates were obtained with a population of offshore wind turbines over a number of years. Other studies do however exist with varying observed rates across different wind turbine models which would affect calculated costs. Vessel, staff and component repair costs were taken from the original cost model, and reflect costs at the time the model was created in 2016, which may not reflect current rates. Another key source of uncertainty when trying to understand costs is P_f, which as stated previously was set to values of between 10% and 40%. This range has been chosen to reflect the uncertainty, as it will ultimately be specific to each site and turbine with various impact factors such as differing loading patterns, ambient conditions, manufacturing and installation tolerances and levels of routine maintenance. Future work should attempt to better define this number in relation to each component and fault type. Results also do not take into consideration the reliability and extra cost associated with the CMS system and network infrastructure, which may have some impact on failure rates and overall costs of implementing predictive and condition based strategies.

6.6 Conclusion

In conclusion, this chapter has presented analysis in an attempt to bring much needed clarity on the impact advanced monitoring strategies can have on lifetime O&M costs. The key findings of this research are as follows:

- With the input parameters described, results showed a potential cost reduction of up to 12.5% in direct O&M costs (transport, staff and repair costs) and up to 15% reduction in lost production by utilising advanced monitoring strategies.
- Results showed that the major driving factor of realising these savings is through early intervention to avoid failure and major component replacement, and hence avert the need to use a HLV and instead use a CTV for a simpler repair.
- If weighing up the risk of component failure and replacing a component too early, results suggest that it is more cost effective to intervene earlier if HLVs can be avoided, even if that means more major repairs over the lifetime of the site.
- Using a more condition based approach and pushing component design life closer to the end of its RUL can further reduce costs, however results suggest that this could be due to increased availability through reduced lost revenue, rather than decreasing direct O&M costs.
- As a percentage of total costs potential savings are consistent across wind farm size and distance from shore up to 50km. Beyond 50km we see a percentage drop due to increased lost production, however the reasons for this phenomenon are due to modelling constraints and has been discussed previously.

6.7 Future Work

Future work in this area would be to further validate the results with more sites and resource allocation. In line with new proposed offshore sites it would also be useful to increase wind farm size to beyond 200 wind turbines to see if results presented for below 100 hold true for very large sites. This was not completed during this study due to

CPU limitations. With regards to this methodology specifically, future work could also involve more accurately choosing values for P_f through analysis of failure records and root cause analysis (RCA) reports to understand how many major component failures could on average be prevented. Finally, condition based approaches may involve the requirements of additional inspections to ensure components do not fail as damage worsens. The cost of these additional inspections could be explored further through the introduction of an additional intervention category with specific resource constraints and requirements.

Chapter 7

Overall Conclusions & Future Work

The aim of the research is to provide insight into the following research question:

"How can machine learning techniques be leveraged to improve wind turbine generator diagnostics and prognostics, and what impact can using such approaches have on the overall wind farm O&M cost?"

Each chapter has presented research to provide insight into a series of smaller research questions, however it is now important to discuss these findings in the context of the original primary research question above. The conclusions below will attempt to bring out the pertinent findings of each chapter, however full detail along with all uncertainties, assumptions and discussion can be found in each individual chapter. This chapter will conclude by outlining future work, and provide an overview of how the research and conclusions presented will contribute to the wind industry.

7.1 Conclusions

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

7.1.1 Chapter 2 Summary

The literature review first of all provided an overview of wind turbine generator reliability rates, showing that while the generator may not have the highest failure rate, it does have one of the highest impacts on overall downtime and repair costs of any drivetrain assembly. This highlights the importance of advanced condition monitoring and the opportunity to reduce O&M costs through effective and automated methods. There has been lots of literature published to date on using both SCADA and vibration data to more accurately and automatically diagnose a range of wind turbine faults, with techniques and methodologies presented across data processing, feature extraction and the application of machine learning techniques. This research however has tended to focus on gearbox faults due to the slightly higher failure rates and cost of repair when compared to generators. Studies applying techniques to diagnose and isolate a range of generator faults are not as readily available. The literature review also found that, although attempts have been made, there are also not enough studies that estimate remaining useful life of generators and their components, or indeed make any effort to quantify the O&M cost savings associated with advanced monitoring strategies. This thesis aims to address these shortcomings.

7.1.2 Chapter 3 Summary

Chapter 3 dealt with wind turbine generator fault diagnostics and more specifically how using machine learning can enhance traditional vibration and SCADA based analysis. With regards to vibration analysis, classification algorithms were applied to determine whether a particular component had a fault by classifying known fault-specific vibration based features. Results were promising highlighting a real opportunity for practical applications that can reduce expert engineering hours spent manually classifying faults. Limitations do however exist due to differing vibration baselines between turbines, varying operating conditions between sites and similarities between vibration indicators of different fault types. This required thought and expertise when setting up models to avoid a large number of false positive predictions.

With regards to SCADA data, there is a large opportunity to monitor wind turbine

health in a cost effective manner, especially in those turbines that do not have any active vibration based condition monitoring system. It has been shown that normal behaviour modelling can be a powerful technique when it comes to detecting changes in operating behaviour, which has already been widely shown in literature across other components of the drivetrain. For practical in field applications however SCADA based NBM's should be approached with caution. It was demonstrated in this chapter that models can be case specific, meaning that a different amount of training data and features are required for each turbine or fault, which could lead to high costs associated with setting up models and storing data. It is also very difficult to locate the fault without further analysis or inspection. That being said, SCADA data may still be under utilised in industry and does offer a cost efficient way of gaining some insight into component health, if used correctly.

7.1.3 Chapter 4 Summary

Leading on from Chapter 3, Chapter 4 aimed to provide insight into how SCADA and vibration data could be used together to create a single anomaly detection model for fault diagnosis, as well as highlight some of the challenges in doing so. A framework was presented to combine different data streams by building individual NBMs and assessing the error through a SVM classifier. It was shown that by doing so faults can be detected more consistently than using an individual model only relying on one source of information on component health. Models such as this are very specific to each fault case however, and would be challenging to implement in practice. More work is required to understand the entire benefits of combining models like this across different fault types, components, wind turbines and monitoring data sources.

7.1.4 Chapter 5 Summary

Vibration analysis is the most commonly used method of assessing component life and making a diagnosis. Chapter 5 looked at how data driven methods can be used to take this one step further and estimate component remaining useful life. Looking at examples of similar bearing failure across multiple turbines and sites, results showed that

vibration signatures can vary significantly across sites and operating conditions. This can be detrimental to classification accuracy when making a prognosis on component RUL. To navigate this issue a clustering methodology was presented to group vibration samples by a range of operating conditions including shaft speed, instantaneous torque, power output and wind speed to give a clear understanding of where each turbine was on the operating curve when the vibration sample was measured. By doing so prognosis accuracy was increased. Results also suggests that optimum operating conditions may exist in which to detect faults and estimate RUL, with accuracy increasing in operating conditions just before reaching rated power. This could have implications in practice on setting the parameters for data acquisition for both better performance and optimisation of data storage. That being said, with regards to the overall approach and supervised models in general, the amount of historical failure example data required could be a major barrier for large scale implementation.

7.1.5 Chapter 6 Summary

One of the key areas with limited publications is quantifying potential cost savings with regards to predictive maintenance strategies. Chapter 6 aimed to understand this important element in more detail. Analysis in this chapter looked specifically at offshore wind farms, where major component failures have the largest impact on both direct O&M costs and lost production. Results suggest that significant savings can be made by implementing predictive maintenance strategies, as by intervening earlier costs associated with HLVs can be reduced. Analysis also indicates that if weighing up the risk of component failure and replacing a component too early, it is more cost effective to intervene earlier if HLVs can be avoided, even if that means more major repairs over the lifetime of the site.

7.2 Discussion

Machine learning can be an effective method to enhance vibration and SCADA analysis, and provide an automated solution for setting much needed turbine specific thresholds.

Wind turbines operate under varying speed and loads, which results in differences in vibration and performance parameters, making one size fits all approaches to alarm threshold setting sub-optimal by causing false alarms. Additionally, if thresholds are not set correctly it can cause underlying faults and issues to be missed, leading to component failure at a substantial cost to owners through extensive repair work and additional lost production.

Data driven techniques require careful application and rely on expertise in both data science and wind turbine operation, performance and condition monitoring. Wind farm data is starting to be much more readily available to operators through careful storage and digitalisation tools and platforms. The most challenging element in realising the potential savings of predictive maintenance is most likely down to the assumptions of what constitutes normal behaviour and if models can be developed with high enough dimensionality at scale to cope with the varying load conditions and modes of operation without increasing manual effort dealing with false alarms. Getting relevant examples of failure on which to base estimates of RUL is also not without its own challenges at scale, whether that be storing the right data efficiently or simply having confidence in exact fault types and failure modes after replacement. How transferable examples of failure are as turbines scale up and new drivetrain designs are installed is also difficult to know.

With so much variation possible throughout the training and validation processes, providing a suitable benchmark to assess model performance is also problematic. A true understanding of model performance in comparison to other techniques can only be achieved by assessing both the alarms generated by any particular model over time, and the efficiency of the training process. For the the former, evidence of component health is required to match any alarms generated with fault progression, a condition which requires a carefully planned dataset. With regards to the latter, training efficiency would have to consider multiple criteria such as training length, data quality, number of features, as well as the ability to handle key operating constraints such as curtailment. In practice, models that can successfully predict component health while being scaled across an entire fleet covering all operating zones may be preferable to highly accurate,

turbine specific models that require specific training regimes. This trade-off between model accuracy and scalability is currently not well documented in literature.

Another point worth discussing in relation to the larger research question is with regards to model retraining. As maintenance is performed operational characteristics change over time, whether that is due to general wear, component replacements, weather and loading patterns or new control strategies. In real time applications this means that data driven models require retraining in order to reflect current behaviour and minimise false alarms. What constitutes an allowable change in normal behaviour and what constitutes a fault is often a grey area. Again this requires careful consideration by experienced engineers until such time that data driven methods can rely further on reinforcement learning to make these decisions without risking failure.

7.3 Future Work

Future work has been discussed in detail for each of the four chapters representing secondary research questions. In terms of importance to industry needs, the focus of academic study should centre on areas that can aid decision makers and save engineering time. The first area that can make a difference is through combining data sources for prognostic purposes, as fault progression is rarely linear and engineers in practice rely on multiple system parameters to understand component condition and the extent of any damage before estimating RUL.

Another area is through hybrid-modelling (relying on both physics-based and data-based modelling) to assist with fault prognosis to more accurately model damage progression and degradation. This will be particularly useful when trying to use proven techniques on larger wind turbines with newer drivetrain technology. This could save time and effort in transferring knowledge with sites or turbines with no data or failure history.

Finally, the true cost of predictive maintenance needs to be further explored in not only the context of reliability, direct O&M costs and lost production, but also taking into consideration the cost associated with increased digitalisation. This could include such things as increased storage and other IT infrastructure, monitoring contracts,

software etc. This may be especially important for not only new sites, but for lifetime extension scenarios when advanced monitoring may be most crucially required.

Bibliography

[1] 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, 10 2017. [Online]. Available: http://www.mdpi.com/1996-1073/10/11/1686

[2] S. Chatterjee and S. Chatterjee, "Review on the techno-commercial aspects of wind energy conversion system," *IET Renewable Power Generation*, vol. 12, no. 14, pp. 1581–1608, 10 2018.

[3] C. V. Hernandez, T. Telsnig, and A. V. Pradas, "JRC Wind Energy Status Report: 2016 Edition — EU Science Hub," 2017. [Online]. Available: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/jrc-wind-energy-status-report-2016-edition

[4] B. Hahn, M. Durstewitz, and K. Rohrig, "Reliability of wind turbines - Experiences of 15 years with 1,500 WTs," *Wind Energy*, pp. 328–330, 2007.

[5] W. Garlick, R. Dixon, and S. Watson, "A model-based approach to wind turbine condition monitoring using SCADA data," 2009.

[6] J. Tautz-Weinert and S. Watson, "Using SCADA data for wind turbine condition monitoring – a review," *IET Renewable Power Generation*, vol. 11, no. 4, pp. 382–394, 2017.

[7] I. Dinwoodie, "Modelling the operation and maintenance of offshore wind farms," Ph.D. dissertation, University of Strathclyde, 2014.

[8] J. Carroll, A. McDonald, and D. McMillan, "Failure rate, repair time and unscheduled O and M cost analysis of offshore wind turbines," *Wind Energy*, vol. 19, no. 6, pp. 1107–1119, 2016.

[9] C. J. Crabtree, D. Zappala, and S. I. Hogg, "Wind energy: UK experiences and offshore operational challenges," *Proceedings of the Institution of Mechanical Engineers Part a-Journal of Power and Energy*, vol. 229, no. 7, pp. 727–746, 2015.

[10] B. Valpy and E. Philip, "Future renewable energy costs: offshore wind, How technology innovation is anticipated to reduce the cost of energy from European offshore wind farms," *KIC InnoEnergy*, 2014.

[11] 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, 2015.

[12] S. Pfaffel, S. Faulstich, and K. Rohrig, "Performance and Reliability of Wind Turbines: A Review," *Energies*, vol. 10, no. 11, p. 1904, 11 2017. [Online]. Available: http://www.mdpi.com/1996-1073/10/11/1904

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

[14] M. Wilkinson, B. Hendriks, F. Spinato, K. Harman, E. Gomez, H. Bulacio, J. Roca, P. J. Tavner, Y. Feng, and H. Long, "Methodology and results of the reliawind reliability field study," in *European Wind Energy Conference*, 1 2010.

[15] E. Hart, A. Turnbull, D. McMillan, J. Feuchtwang, E. Golysheva, and R. Elliott, "Investigation of the relationship between main-bearing loads and wind field characteristics," in *Journal of Physics: Conference Series*, vol. 926, no. 1. Institute of Physics Publishing, 11 2017.

[16] A. Turnbull, J. Carroll, S. Koukoura, and A. McDonald, "Prediction of wind turbine generator bearing failure through analysis of high-frequency vibration data and the application of support vector machine algorithms," *The Journal of Engineering*, vol. 2019, no. 18, pp. 4965–4969, 7 2019.

[17] A. Turnbull, J. Carroll, A. McDonald, and S. Koukoura, "Prediction of wind turbine generator failure using two-stage cluster-classification methodology," *Wind Energy*, vol. 22, no. 11, pp. 1593–1602, 11 2019.

[18] E. Hart, A. Turnbull, J. Feuchtwang, D. McMillan, E. Golysheva, and R. Elliott, "Wind turbine main-bearing loading and wind field characteristics," *WIND ENERGY*, vol. 22, no. 11, pp. 1534–1547, 2019.

[19] A. Turnbull, J. Carroll, and A. McDonald, "Combining SCADA and vibration data into a single anomaly detection model to predict wind turbine component failure," *Wind Energy*, vol. 24, no. 3, pp. 197–211, 3 2021.

[20] C. McKinnon, A. Turnbull, S. Koukoura, J. Carroll, and A. McDonald, "Effect of time history on normal behaviour modelling using SCADA data to predict wind turbine failures," *Energies*, vol. 13, no. 18, p. 4745, 9 2020.

[21] A. Turnbull and J. Carroll, "Cost Benefit of Implementing Advanced Monitoring and Predictive Maintenance Strategies for Offshore Wind Farms," *ENERGIES*, vol. 14, no. 16, 2021.

[22] T. Ackermann and L. Söder, "An overview of wind energy-status 2002," *Renewable and Sustainable Energy Reviews*, vol. 6, no. 1-2, pp. 67–127, 2002.

[23] A. D. Hansen and L. H. Hansen, "Wind turbine concept market penetration over 10 years (1995-2004)," *Wind Energy*, vol. 10, no. 1, pp. 81–97, 1 2007.

[24] 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, 5 2004. [Online]. Available: http://journals.sagepub.com/doi/10.1260/0309524041590099

[25] J. Serrano-González and R. Lacal-Arántegui, "Technological evolution of onshore wind turbines—a market-based analysis," *Wind Energy*, vol. 19, no. 12, pp. 2171–2187, 12 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/we.1974

[26] H. Li and Z. Chen, "Overview of different wind generator systems and their comparisons," *IET Renewable Power Generation*, vol. 2, no. 2, pp. 123–138, 2008.

[27] J. A. Baroudi, V. Dinavahi, and A. M. Knight, "A review of power converter topologies for wind generators," *Renewable Energy*, vol. 32, no. 14, pp. 2369–2385, 11 2007.

[28] Z. Chen, "Wind turbine drive train systems," in *Wind Energy Systems: Optimising Design and Construction for Safe and Reliable Operation.* Elsevier Ltd., 1 2010, pp. 208–246.

[29] R. Llorente Iglesias, R. Lacal Arantegui, and M. Aguado Alonso, "Power electronics evolution in wind turbines - A market-based analysis," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 9, pp. 4982–4993, 12 2011.

[30] H.-J. Wagner and J. Mathur, *Introduction to Wind Energy Systems*, ser. Green Energy and Technology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. [Online]. Available: http://link.springer.com/10.1007/978-3-642-32976-0

[31] G. M. Joselin Herbert, S. Iniyan, E. Sreevalsan, and S. Rajapandian, "A review of wind energy technologies," *Renewable and Sustainable Energy Reviews*, vol. 11, no. 6, pp. 1117–1145, 8 2007.

[32] Eunshin Byon, Lewis Ntaimo, Chanan Singh, and Yu Ding, "Handbook of Wind Power Systems," ser. Energy Systems, P. M. Pardalos, S. Rebennack, M. V. F. Pereira, N. A. Iliadis, and V. Pappu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, ch. Wind Energy, pp. 639–672. [Online]. Available: http://link.springer.com/10.1007/978-3-642-41080-2

[33] *Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design*. Springer London, 2009.

[34] 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, 10 2018.

[35] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, "Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges," *Energies*, vol. 7, no. 4, pp. 2595–2630, 2014.

[36] I. Attoui and A. Omeiri, "Fault Diagnosis of an Induction Generator in a Wind Energy Conversion System Using Signal Processing Techniques," *Electric Power Components and Systems*, vol. 43, no. 20, pp. 2262–2275, 12 2015.

[37] M. R. Shahriar, P. Borghesani, and A. C. C. Tan, "Electrical Signature Analysis-Based Detection of External Bearing Faults in Electromechanical Drivetrains," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5941–5950, 2018.

[38] J. D. Zhu, J. M. Yoon, D. He, and E. Bechhoefer, "Online particle-contaminated lubrication oil condition monitoring and remaining useful life prediction for wind turbines," *Wind Energy*, vol. 18, no. 6, pp. 1131–1149, 2015.

[39] 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.

[40] 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, 8 2018.

[41] 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*, vol. 133, pp. 620–635, 4 2019. [Online]. Available: http://creativecommons.org/licenses/by/4.0/

[42] Christopher Bishop, *Pattern Recognition and Machine Learning.* Springer, 2016. [Online]. Available: https://www.springer.com/gp/book/9780387310732

[43] S. Raschka and V. Mirjalili, *Python Machine Learning*, 3rd ed., 2019. [Online]. Available: https://www.oreilly.com/library/view/python-machine-learning/9781783555130/

[44] S. Kotsiantis, D. Kanellopoulos, P. E. Pintelas, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: A review Special Issue "Ensemble Algorithms and Their Applications" View project EVOLUTIONARY INTELLIGENCE (SPRINGER): Special Issue on "Intelligent and Fuzzy Systems in Data Science and Big Data" View project Handling imbalanced datasets: A review," Tech. Rep., 2006. [Online]. Available: https://www.researchgate.net/publication/228084509

[45] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, 2001. [Online]. Available: http://statweb.stanford.edu/~tibs/book/preface.ps

[46] V. Vapnik, "The Nature of Statistical Learning Theory," 2000. [Online]. Available: https://books.google.co.uk/books?hl=en&lr=&id=EqgACAAAQBAJ&oi=fnd&pg=PR7&dq=the+nature+of+statistical+learning&ots=g4FZht7Y65&sig=GAhg9aeGRHNfygJUsh-JrYqT8l8&redir_esc=y#v=onepage&q=the%20nature%20of%20statistical%20learning&f=false

[47] K. Gurney, *An Introduction to Neural Networks*, 1997. [Online]. Available: https://www.google.co.uk/books/edition/An_Introduction_to_Neural_Networks/e0pZDwAAQBAJ?hl=en&gbpv=1&dq=neural+network+theory&printsec=frontcover

[48] A. Galushkin, *Neural Networks Theory*, 2007. [Online]. Available: https://www.google.co.uk/books/edition/Neural_Networks_Theory/ULds8NuzLtkC?hl=en&gbpv=0

[49] H. Wang and J. Chen, "Performance degradation assessment of rolling bearing based on bispectrum and support vector data description," *JVC/Journal of Vibration and Control*, vol. 20, no. 13, pp. 2032–2041, 10 2014.

[50] 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, 4 2009.

[51] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping Multidimensional Data: Recent Advances in Clustering*. Springer Berlin Heidelberg, 2006, pp. 25–71. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-28349-8_2

[52] J. Kogan, C. Nicholas, and M. Teboulle, *Grouping multidimensional data: Recent advances in clustering*. Springer Berlin Heidelberg, 2006.

[53] M. M. Deza and E. Deza, *Encyclopedia of Distances*. Springer Berlin Heidelberg, 2016.

[54] 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, 4 2005. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/15572470/

[55] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. Springer US, 2010.

[56] E. Wiggelinkhuizen, T. Verbruggen, H. Braam, L. Rademakers, J. Xiang, and S. Watson, "Assessment of condition monitoring techniques for offshore wind farms," *Journal of Solar Energy Engineering, Transactions of the ASME*, vol. 130, no. 3, pp. 0 310 041–0 310 049, 8 2008.

[57] 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, 5 2013.

[58] D. Astolfi, F. Castellani, and L. Terzi, "Fault prevention and diagnosis through SCADA temperature data analysis of an onshore wind farm," *Diagnostyka*, vol. 15, pp. 71–78, 1 2014.

[59] 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 5th International conference on energy sustainability*. American Society of Mechanical Engineers, 2011, pp. 2071–2079. [Online]. Available: https://www-webofscience-com.proxy.lib.strath.ac.uk/wos/woscc/full-record/WOS:000321076701053

[60] Z. Zhang, "Analysis of wind turbine vibrations based on SCADA data," *Journal of Solar Energy Engineering-transactions of The Asme - J SOL ENERGY ENG*, vol. 132, 8 2010.

[61] S. Catmull, "Self-Organising Map Based Condition Monitoring Of Wind Turbines," *European Wind Energy Association*, 2011.

[62] M. Wilkinson, B. Darnell, T. Van Delft, and K. Harman, "Comparison of methods for wind turbine condition monitoring with SCADA data," *IET Renewable Power Generation*, vol. 8, no. 4, pp. 390–397, 2014.

[63] S. McArthur, V. Catterson, and J. McDonald, "A multi-agent condition monitoring architecture to support transmission and distribution asset management," *3rd IEE International Conference on Reliability of Transmission and Distribution Networks*, pp. 87–91, 2005.

[64] M. Schlechtingen and I. Ferreira 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, 7 2011.

[65] 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.

[66] A. Verma and A. Kusiak, "Fault Monitoring of Wind Turbine Generator Brushes: A Data-Mining Approach," *Journal of Solar Energy Engineering-Transactions of the Asme*, vol. 134, no. 2, p. 9, 2012.

[67] Y. Yan, J. Li, and D. W. Gao, "Condition parameter modeling for anomaly detection in wind turbines," *Energies*, vol. 7, no. 5, pp. 3104–3120, 2014.

[68] Y. Pei, Z. Qian, S. Tao, and H. Yu, "Wind Turbine Condition Monitoring Using SCADA Data and Data Mining Method," *2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings*, pp. 3760–3764, 1 2019.

[69] P. Bangalore and L. B. Tjernberg, "An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings," *IEEE Transactions on Smart Grid*, vol. 6, pp. 980 – 987, 2015.

[70] 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, 8 2017.

[71] Y. Cui, P. Bangalore, and L. B. Tjernberg, "An anomaly detection approach based on machine learning and scada data for condition monitoring of wind turbines," *2018 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2018 - Proceedings*, 8 2018.

[72] 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.

[73] 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, 11 2012.

[74] 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, 10 2006.

[75] 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, 8 2017. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/1461348417725950

[76] G. X. Wu, Y. B. Zuo, and Y. H. Shi, "Research on Vibration Signal Feature Extraction Method to the Wind Turbine Generator," *Advanced Materials Research*, vol. 902, pp. 370–377, 2014. [Online]. Available: https://www.scientific.net/AMR.902.370

[77] E. Mollasalehi, "Data-driven and Model-based Bearing Fault Analysis - Wind Turbine Application," Ph.D. dissertation, University of Calgary, Calgary, 2017. [Online]. Available: http://dx.doi.org/10.11575/PRISM/25517

[78] S. Kim, D. An, and J.-H. Choi, "Diagnostics 101: A Tutorial for Fault Diagnostics of Rolling Element Bearing Using Envelope Analysis in MATLAB," *Applied Sciences 2020, Vol. 10, Page 7302*, vol. 10, no. 20, p. 7302, 10 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/20/7302/htmhttps://www.mdpi.com/2076-3417/10/20/7302

[79] J. Ogata and M. Murakawa, "Vibration-Based Anomaly Detection Using FLAC Features for Wind Turbine Condition Monitoring," *8th European Workshop On Structural Health Monitoring*, 2016.

[80] S. Koukoura, J. Carroll, A. McDonald, and S. Weiss, "Comparison of wind turbine gearbox vibration analysis algorithms based on feature extraction and classification," *IET Renewable Power Generation*, vol. 13, no. 14, pp. 2549–2557, 10 2019.

[81] S. Koukoura, J. Carroll, and A. McDonald, "On the use of AI based vibration condition monitoring of wind turbine gearboxes," *Journal of Physics:*

Conference Series, vol. 1222, no. 1, p. 012045, 5 2019. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1222/1/012045https://iopscience.iop.org/article/10.1088/1742-6596/1222/1/012045/meta

[82] DECC, "DECC Electricity Generation Costs 2013 - GOV.UK," Tech. Rep., 2013. [Online]. Available: https://www.gov.uk/government/publications/decc-electricity-generation-costs-2013

[83] GL Garrad Hassan, "Offshore wind operations and maintenance opportunities in Scotland," Tech. Rep., 2013.

[84] C. Dao, B. Kazemtabrizi, and C. Crabtree, "Wind turbine reliability data review and impacts on levelised cost of energy," *Wind Energy*, vol. 22, no. 12, pp. 1848–1871, 12 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2404

[85] A. May, D. McMillan, and S. Thöns, "Economic analysis of condition monitoring systems for offshore wind turbine sub-systems," *IET Renewable Power Generation*, vol. 9, no. 8, pp. 900–907, 11 2015.

[86] S. Koukoura, M. N. Scheu, and A. Kolios, "Influence of extended potential-to-functional failure intervals through condition monitoring systems on offshore wind turbine availability," *Reliability Engineering and System Safety*, vol. 208, p. 107404, 4 2021.

[87] A. Khadersab and S. Shivakumar, "Vibration Analysis Techniques for Rotating Machinery and its effect on Bearing Faults," in *Procedia Manufacturing*, vol. 20. Elsevier B.V., 1 2018, pp. 247–252.

[88] A. Muszynska, "Vibrational Diagnostics of Rotating Machinery Malfunctions," *International Journal of Rotating Machinery*, vol. 1, no. 3-4, pp. 237–266, 1995.

[89] S. Singh and M. Vishwakarma, "A Review of Vibration Analysis Techniques for Rotating Machines," Tech. Rep., 2015. [Online]. Available: www.ijert.org

[90] D. Kateris, D. Moshou, X.-E. Pantazi, I. Gravalos, N. Sawalhi, and S. Loutridis, "A machine learning approach for the condition monitoring of rotating machinery," *Journal of Mechanical Science and Technology*, vol. 28, no. 1, pp. 61–71, 2014. [Online]. Available: https://doi.org/10.1007/s12206-013-1102-y

[91] M. J. Kabir, A. M. T. Oo, M. Rabbani, and Ieee, "A Brief Review on Offshore Wind Turbine Fault Detection and Recent Development in Condition Monitoring Based Maintenance System," in *Australasian Universities Power Engineering Conference (AUPEC)*, ser. Australasian Universities Power Engineering Conference. NEW YORK: Ieee, 2015.

[92] A. Romero, S. Soua, T. H. Gan, and B. Wang, "Condition monitoring of a wind turbine drive train based on its power dependant vibrations," *Renewable Energy*, vol. 123, pp. 817–827, 2018.

[93] D. Hochmann and E. Bechhoefer, "Envelope bearing analysis: Theory and practice," *IEEE Aerospace Conference Proceedings*, vol. 2005, 2005.

[94] A. C. Muller and S. Guido, *Introduction to Machine Learning with Python*, 3rd ed. O'Reilly Media, 2017.

[95] F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, A. Stetco, M. Barnes, J. Keane, and G. Nenadic, "Machine learning methods for wind turbine condition monitoring: A review," *Renewable Energy*, vol. 133, pp. 620–635, 2019.

[96] Y. S. Hong, Y. M. Cho, S. H. Ahn, C. K. Song, and Z. Hameed, "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.

[97] C. Scheffer, *Practical Machinery Vibration Analysis and Predictive Maintenance - 1st Edition*. [Online]. Available: https://www.elsevier.com/books/practical-machinery-vibration-analysis-and-predictive-maintenance/scheffer/978-0-7506-6275-8

[98] H. D. M. de Azevedo, A. M. Araujo, and N. Bouchonneau, "A review of wind turbine bearing condition monitoring: State of the art and challenges," *Renewable & Sustainable Energy Reviews*, vol. 56, pp. 368–379, 2016.

[99] M. L. Hossain, A. Abu-Siada, and S. M. Muyeen, "Methods for Advanced Wind Turbine Condition Monitoring and Early Diagnosis: A Literature Review," *Energies*, vol. 11, no. 5, p. 14, 2018.

[100] L. Saidi, J. Ben Ali, and F. Fnaiech, "Application of higher order spectral features and support vector machines for bearing faults classification," *ISA Transactions*, vol. 54, pp. 193–206, 1 2015.

[101] J. L. Chen, J. Pan, Z. P. Li, Y. Y. Zi, and X. F. Chen, "Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals," *Renewable Energy*, vol. 89, pp. 80–92, 2016.

[102] P. Borghesani, P. Pennacchi, S. Chatterton, and R. Ricci, "The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 44, no. 1, pp. 118 – 133, 2014.

[103] Y. Wang, P. W. Tse, B. P. Tang, Y. Qin, L. Deng, T. Huang, and G. H. Xu, "Order spectrogram visualization for rolling bearing fault detection under speed variation conditions," *Mechanical Systems and Signal Processing*, vol. 122, pp. 580–596, 2019.

[104] J. Carroll, A. McDonald, I. Dinwoodie, D. McMillan, M. Revie, and I. Lazakis, "Availability, operation and maintenance costs of offshore wind turbines with different drive train configurations," *Wind Energy*, vol. 20, no. 2, pp. 361–378, 2 2017.

[105] J. Carroll, "Cost of energy modelling and reduction opportunities for offshore wind turbines," Ph.D. dissertation, University of Strathclyde, 2016.