Data-driven Operations & Maintenance for Offshore Wind Farms: Tools and Methodologies

Alexios Koltsidopoulos Papatzimos

Thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Engineering
as part of the Industrial Doctorate of Offshore Renewable Energy, jointly awarded by
the University of Edinburgh, the University of Exeter and the University of Strathclyde.
The work presented has been conducted under the industrial supervision of
EDF Energy R&D UK Centre Limited
2019

Ἓν οἶδα ὅτι οὐδὲν οἶδα

I know one thing; that I know nothing

-Socrates

This phrase is not one that Socrates himself is ever recorded saying. It is believed to be paraphrased in Latin and later on back-translated to ancient Greek. It is interesting to understand the context in which this phrase occurs; Socrates, having gone to a wise man, and having discussed with him, withdraws and thinks that "I seem at least in this one small thing to be wiser than this man [who believed himself to be wise], that the things I do not know, I do not imagine that I know"[1]. Socrates, since he denied any kind of knowledge, then tried to find someone wiser than himself among politicians, poets, and craftsmen. It appeared that politicians claimed wisdom without knowledge, poets could touch people with their words, but did not know their meaning and craftsmen could claim knowledge only in specific and narrow fields [2]. The interpretation of this phrase might be Socrates' awareness of his own ignorance.

Socrates' journey to knowledge and wisdom really strikes me as it reminds me that I should always work towards justifying my findings with solid proof and deep understanding, that I need to be down to earth and modest, and that I have to develop a holistic personality. To me it also means that by submitting this doctoral thesis I am not calling myself an expert in anything; I would be an ignorant to do so. Of course I have learnt a lot of things on the way, but this is just the beginning of my journey...

[1] Plato, Apology 21d.
[2] Plato; Morris Kaplan (2009). The Socratic Dialogues. Kaplan Publishing.

Abstract

Offshore wind assets have reached multi-GW scale and additional capacity is being installed and developed. To achieve demanding cost of energy targets, awarded by competitive auctions, the operations and maintenance (O&M) of these assets have to become increasingly efficient, whilst ensuring compliance and effectiveness. Existing offshore wind farm assets generate a significant amount of inhomogeneous operational data. These data contain rich information about the condition of the assets, which are rarely fully utilized by the operators and service providers. This thesis provides useful methodologies and tools that can help wind farm owners, operators and service providers to reduce their O&M costs by better managing their data, integrating processes and providing data-driven decision making.

The developed methodologies and tools are being presented through several case studies, showing the effectiveness of the solutions and their potential cost reduction opportunities. These are split into the following four themes:

i. **Data management** techniques, methodologies and case studies, aiming to improve data collection and data integration strategies for a data informed decision making.

ii. **Processes** and best practices for workflow improvements and automated data collection and standardization.

iii. **Data analytics** including reliability, diagnostic and prognostic methodologies and case studies.

iv. **Maintenance planning** including enhanced planning strategies, decision support frameworks and optimized maintenance operations.

All of the above frameworks, methodologies and case studies are linked together as they provide insights for data-driven decision making, which results in better informed and thus less costly maintenance strategies.

The methodologies and case studies presented will assist in creating data-driven O&M processes and allowing the full utilization of the produced offshore wind farm data.

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree or professional qualification. Except where otherwise acknowledged, the work presented is entirely my own. Parts of the work outlined in this thesis have been published or are under review for publication:

- Section 3.1 is based on Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2017) Towards automated and integrated data collection-standardising workflow processes for the offshore wind industry. In *Proc. of Offshore Wind Energy 2017*, 6-8 June, London, UK.

- Parts of Sections 3.2, 3.3, 5.1 and 5.3 are based on Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2017) An Integrated Data Management Approach for Offshore Wind Turbine Failure Root Cause Analysis. In *Proc. of 36th International Conference on Ocean, Offshore and Arctic Engineering (OMAE)*, OMAE2017-62279, 25–30 June, Trondheim, Norway.

- Sections 3.3, 5.2 and 6.1 are based on:

  – Koltsidopoulos Papatzimos, A., Thies, P.R., Dawood, T. (2018) Data-Driven Operational Reliability Assessment for Offshore Wind Farms. *Reliability Engineering & System Safety*, (Under Review).

  – Koltsidopoulos Papatzimos, A., Thies, P.R., Dawood, T. (2018) Offshore Wind Turbine Fault Alarm Prediction. *Wind Energy*, (accepted).

- Sections 3.4 and 5.4 are based on Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2018). Data Insights from an Offshore Wind Turbine Gearbox Replacement. *Journal of Physics: Conference Series*, **1104**(1).

- Section 4.1 and parts of Section 5.4 are based on Koltsidopoulos Papatzimos, A., Dawood, T., Thies, PR. (2018) Operational Data to Maintenance Optimization: Closing the Loop in Offshore Wind O&M. In *Proc. of the ASME 2018 1st*

International Offshore Wind Technical Conference (IOWTC), 4-7 Nov, San Francisco, USA.

- Sections 4.2 and 6.3 are based on:
  - Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2017) On Risk-based Inspections for Offshore Wind Farms: A Case Study. In *Proc. of 56th Annual British Conf. of Non-Destructive Testing (NDT 2017)*, The British Institute of Non-Destructive Testing (BINDT), 4-7 Sept, Telford, UK.
  - Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2018) Cost effective, risk-based inspection planning for offshore wind farms. *Insight - Non-Destructive Testing and Condition Monitoring*, **60**(6), 299-305.
- Sections 4.3 and 6.2 are based on Koltsidopoulos Papatzimos, A., Thies, P.R., Lonchampt, J., Joly, A., Dawood, T., (2019) Data Informed Lifetime Reliability Prediction for Offshore Wind Farms. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, San Francisco, CA, 2018.

Alexios Koltsidopoulos Papatzimos
July 2019

Acknowledgements

First and foremost I would like to express my gratitude to my industrial and academic supervisors Dr. Tariq Dawood and Dr. Philipp R. Thies. Without their advice and support over those three years, this work would not have been possible. Tariq helped me explore my limits and encouraged me to take initiatives, making this thesis and my industrial experience well rounded. Philipp was always there with useful insights and continuous support, helping me to formulate the research questions and add the academic elements to my work.

I would also like to thank EDF Energy R&D UK Centre for hosting me for three years and special thanks to my line managers; Erwan Galenne, Dr. Ellen-Marie Pavageau and Dr. Marie Berthelot for never saying no to any of my requests and for their continuous support.

Special thanks also go to my colleagues at EDF Energy R&D UK Centre for the supportive and friendly environment and especially to Florian for helping me at the beginning of my project, EDF R&D in France for their expertise and collaboration, EDF Renouvelables and EDF Renewables UK for helping me to understand the industrial context, challenges and the operations and maintenance applications and all the people I have met and worked with from Universities, SMEs and industrial partners that have shared their knowledge along the way.

I am also grateful to my fellow IDCOREans across the different cohorts and especially to Esteve for exchanging ideas, support and advice along this journey.

Last but not least, I would like to thank my family and friends for their love, mental support and encouragement during this period.

Funding from the Energy Technologies Institute (ETI) and the RCUK Energy Programme for the Industrial Doctoral Centre for Offshore Renewable Energy (Grant number EP/J500847/1) is gratefully acknowledged.

Contents

List of Figures

List of Tables

Nomenclature

Acronyms

ANN	Artificial Neural Networks
AR	Autoregressive
AWP	Approved Work Procedure
BS	British Standards
C-RPN	Cost-Risk Priority number
CA	Criticality Analysis
CBM	Condition-based Maintenance
CMMS	Computerized Maintenance Management System
CMS	Condition Monitoring System
CM	Condition Monitoring
CTV	Crew Transfer Vessel
DB	Database
DI	Data Integration
DM	Data Management
DSF	Decision Support Framework
ETL	Extract, Transform, Load
EU	European Union
FFT	Fast Fourier Transform
FMEA	Failure Mode and Effects Analysis
FMECA	Failure Mode, Effect and Criticality Analysis
FRCA	Failure Root Cause Analysis
FN	False Negative

FP	False Positive
HS-IS	High Speed Intermediate Shaft
HSS	High Speed Shaft
HV	Heavy Voltage
HW	High Wind
I&M	Inspection and Maintenance
IEA	International Energy Agency
IEC	International Electrotechnical Commission
IMS	Intermediate Shaft
IoT	Internet of Things
ISO	International Organization for Standardization
kNN	K-Nearest Neighbours
KPI	Key Performance Indicator
LS-IS	Low Speed Intermediate Shaft
LSS	Low Speed Shaft
LV	Low Voltage
LW	Low Wind
MCS	Monte Carlo Simulation
ML	Machine Learning
MS-AR	Markov Switching Autoregressive Model
MS-HR	Markov Switching Autoregressive
MTBF	Mean Time Between Failures
MTBV	Mean Time Between Visits
NDT	Non-Destructive Testing
O&M	Operation and Maintenance
OEM	Original Equipment Manufacturer
OPEX	Operational Expenditure
OREDA	Offshore and Onshore Reliability Data
P-F	Potential-to-functional Failure

PBA	Production-Based Availability
PoF	Probability of Failure
PT&I	Predictive Testing & Inspection
RAF	Approved Form
RBI	Risk-based Inspection
RBM	Risk-based Maintenance
RCA	Root Cause Analysis
RCM	Reliability-centred Maintenance
RDBS	Relational Database System
RDB	Relational Database
RDS-PP	Reference Designation System for Power Plants
RMS	Remote Monitoring System
rms	Root mean square
RM	Risk Matrix
ROP	Routine Operating Procedure
RPN	Risk Priority Number
RUL	Remaining Useful Life
SCADA	Supervisory Control and Data Acquisition
SHM	Structural Health Monitoring
SME	Small and Medium-sized Enterprises
SOV	Service Operation Vessel
SQL	Structured Query Language
SVM	Support Vector Machine
TBA	Time-Based Availability
TI	Turbulence Intensity
TN	True Negative
TP	Transition Piece (Chapters 2,4,6 and 7)
TP	True Positive (Chapters 3 and 5)
WT	Wind Turbine

Greek Symbols

β_i	Weibull distribution shape factor
ΔT	Temperature difference
λ_i	Weibull distribution scale factor
σ_i	Stress
σ_U	Standard deviation of wind speed
τ	Aggregated service time

Roman Symbols

a_t	Weight of trees
ACC	Accuracy
C	Accumulated damage
C	Slack variable
C_a	Cost without corrective action
C_D	Cost of detection
c_i^p	Cost of activity carried out at any moment against the same cost
c_i^r	Failure repair cost
C_n	Cost after corrective action
C_t	Set of distinct classes in a decision tree
CR	Corrosion rate
D	Detection method
d_st	Distance between neighbours for KNN algorithms
$e(t)$	White noise disturbance value
F_1	F_1 Score
h_i	Penalty function
I	Number of intervals of which data are collected
I_G	Gini's index
K	Number of subassemblies
m	Number of observations
N_i	Average number of cycles to failure (Section 6.3)

N_i	Number of turbines (Chapter 2)
n_i	Number of accumulated cycles
$n_{i,k}$	Number of failures
O	Occurrence probability
q	AR delay operator
R	Restriction
$R(t)$	Weibull distribution
R^2	Coefficient of determination
RL	Remaining life
S	Severity of effect
S_c	Setup cost
S_l	Linear space
S_t	Set of indices of selected trees
T	Training data
t	Time
t_d	Decision tree
t_0	Original designed structure thickness
t_{actual}	Current structure thickness
$t_{required}$	Minimum allowable structure thickness
U	Average wind speed
x_i	Optimal component replacement age
y	Year
y_{bag}	Tree weighted average

Part I

Formulating the Research Question

Chapter 1

Introduction

1.1 Background and Research Context

The drive for reduction of greenhouse gas emissions has been building over the last decades in response to the global temperature rise. Renewable energy sources will have the fastest growth in the electricity sector, providing almost 30% of power demand in 2023, up from 24% in 2017 [1]. They are expected to play a significant role in the decarbonisation of the future electricity and energy mix, which will aid in the reduction of greenhouse gasses. The installed capacity of renewable energy is continuously increasing around the globe, due to favourable policies and economies of scale that have allowed renewable energy technologies to compete against more traditional energy generation, such as nuclear, coal and gas plants. Although renewable energy sources generate variable loads, compared to the more constant profile provided by traditional energy sources that are used as a baseload, when combined with battery storage and with effective grid interconnections, they can provide a competitive case against traditional energy generation sources.

Both the need for low carbon generation and policy were the main reasons for growth in renewables up until now. In the recent years, economies of scale have reduced significantly the cost of renewable energy sources, making them cost competitive to traditional energy sources. Growth markets include Europe, USA and Asia; mainly China, India, Japan and Taiwan. The European Union (EU) Renewable Energy Directive and the Paris agreement have set clear goals for the increase of the renewable energy sources in Europe and around the globe. The EU aims for at least 32% [2] of the energy consumption to be generated by renewable energy sources by 2030 and the Paris agreement has indicated a 65% of the world's primary energy supply to be provided by renewable energy sources by 2050, up from around 15% in 2017 [3]. Wind energy has one of the highest growths of all the renewable energy sources, due to the highly available wind resource and its competitive price that is driven by policy regulations, high subsidies and by the deregulation of the electricity markets.

Development of onshore wind turbines began in the 1970s and proceeded towards commercialisation in the 1980s. Offshore wind farm developments, Figure 1.1, have been driven by higher and more consistent wind resource offshore, possibility for the development of larger turbines, which supports the economies of scale, and higher public acceptability due to less visual and noise impact. The first commercial wind project installed offshore, began in Denmark in 1991 at the Vindeby wind farm, building from knowledge transferred from the oil & gas industry. Offshore wind farm developments continued through the decade and since the beginning of 2000, the developments have been growing exponentially, as shown in Figure 1.2, reaching 23.3 GW globally by the end of 2018 and they are expected to reach 120 GW of installed capacity by 2030 [5]. The majority of developments have been mainly in Europe, especially in the United Kingdom and Germany, which together at the end of 2017 accounted for over two thirds of the total offshore wind power installed worldwide. These developments have been favoured by government subsidies, shallow waters and large wind resource in the area. Outside Europe, developments have begun in China, Taiwan, Japan and in the

Figure 1.1: Teesside offshore wind farm [4].

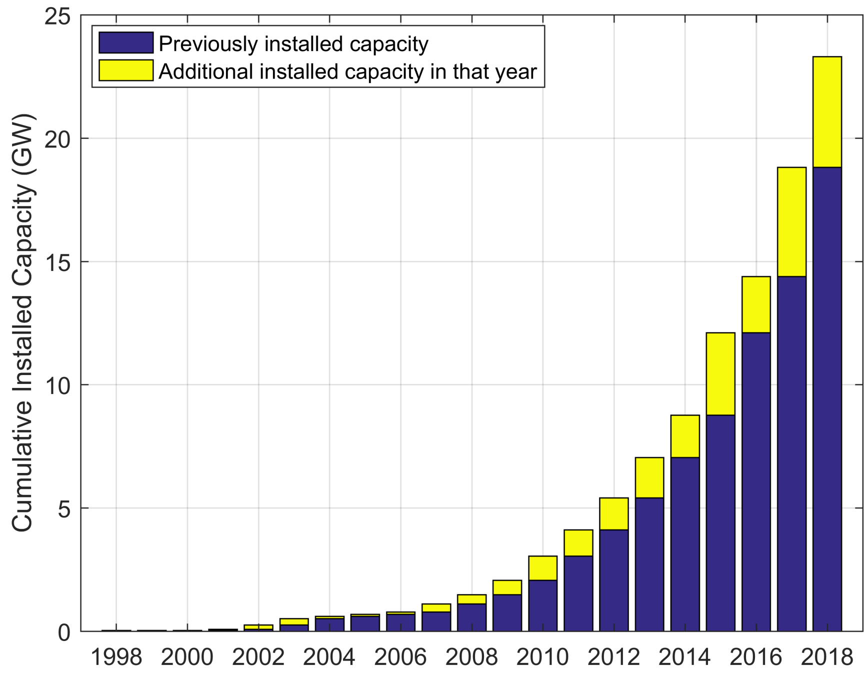


Figure 1.2: Cumulative worldwide offshore wind installations 1998- 2018 [5, 6, 7].

USA. The driver for the recent growth of offshore wind farms around the globe, is the continuously falling costs of new developments, political support and the need for a more diverse energy mix. Traditionally, the development of renewable energy sources have been supported by government subsidies. Recent competitive auctions have reduced the subsidy levels to record low values of £57.5 per MWh in the UK in 2018 and to subsidy-free levels in Germany and the Netherlands, excluding grid connection costs, allowing offshore wind to compete with conventional energy sources.

Figure 1.3 shows the cost breakdown of a typical offshore wind farm. This study was the most recent one when this thesis was written, however the overall cost is significantly higher to the one from the recent auctions that was discussed above, the split of costs is expected to remain the same for newer projects. Operations and maintenance (O&M) costs are an important cost driver, especially since these costs are reoccurring during the full lifetime of the project. Thus, reducing O&M related costs will have a clear influence on the total cost of energy. Moreover, inefficient O&M could lead to higher costs in the future. To achieve demanding cost of energy targets,

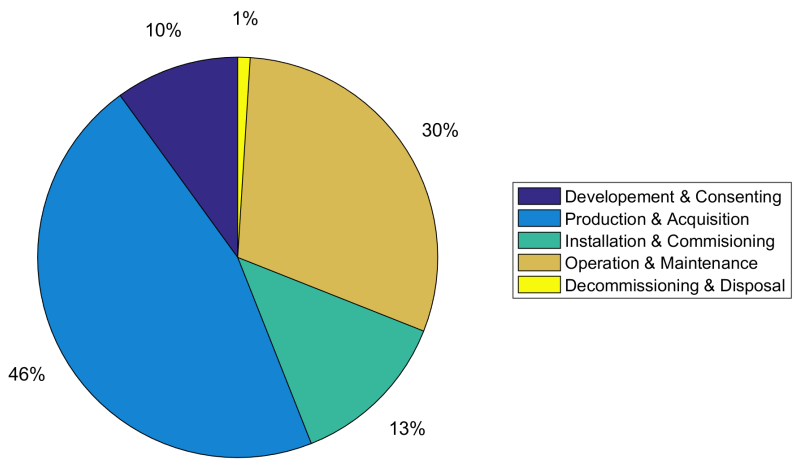


Figure 1.3: Lifecycle cost breakdown of an offshore wind turbine project [10], based on £140/MWh, corresponding to the administrative strike price for 2018/19.

awarded by competitive auctions, the O&M of these assets has to become increasingly efficient, whilst ensuring compliance and effectiveness. Furthermore, effective O&M can increase the assets' availability and offer the possibility of life extension. Typical O&M activities, are expected to reach 30% of the life-time cost of the offshore wind farm. These include logistical, inspection and maintenance activities. Any possible design defects could result in additional costs during the operational life of the assets [8, 9]. The whole O&M supply chain is summarized in Figure 1.4, showing the onshore and offshore logistics, including vessels, helicopters, turbine/ foundation/ array cable maintenance, offshore substations and onshore logistic along with office activities. As the offshore wind farm installations are constantly increasing, over the next few years the offshore wind O&M will become a significant industrial sector on its own. According to the UK Government projections for 2025, the offshore wind industry O&M could be worth around £2bn per annum. The main customers for O&M services are the project owners, the original equipment manufacturers (OEM) of the turbines, the transmission owners and any other third party providers of vessels, components and labour. Due to this complex structure, there can often be a lack of communication between the OEM and the wind farm operator especially on the data provided from the former to the latter [12]. O&M challenges are different for each stakeholder, which could result in a conflict of interest when prioritizing and executing maintenance activities in an asset.

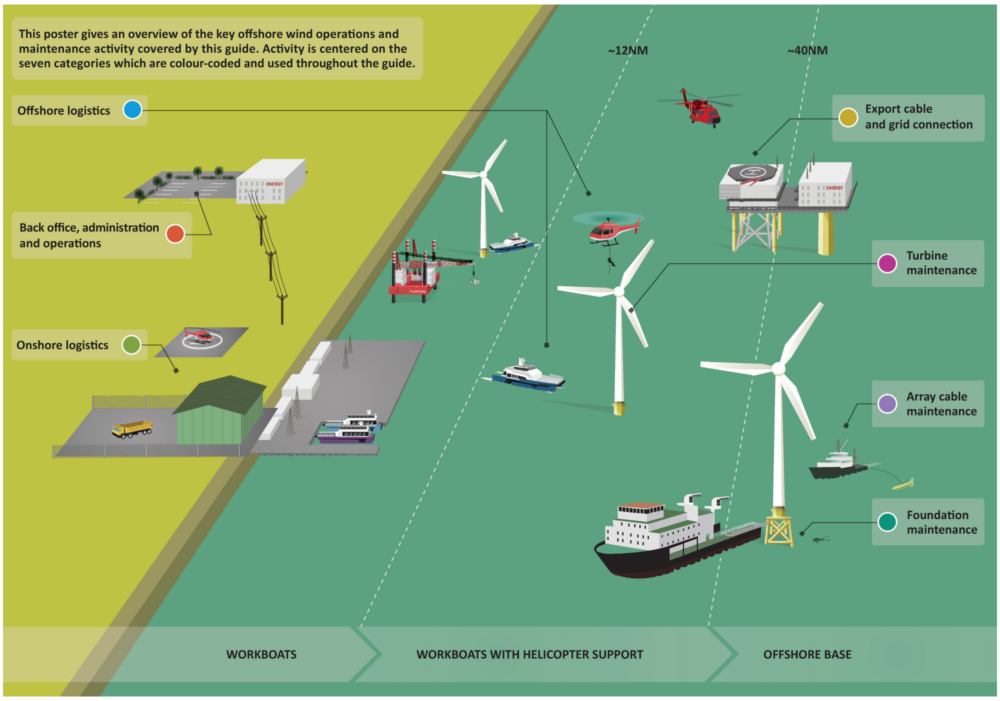


Figure 1.4: Overview of offshore wind O&M activities, as presented from the Scottish Enterprise and the Crown Estate [11].

The most vulnerable stakeholders are the operators as they do not always have access to sufficient data about their assets and can experience inefficient communication with the OEMs that could result in duplicate work. This problem, in conjunction with the increasing amount of operational data produced from different sources with different temporal and format characteristics, creates a need to better organize, analyse and relate information easily and efficiently. This could ease the operational and strategic decision making.

At the same time, collaborative work is needed from the industry for data sharing and data standardization. At the moment, there is no standardised process for data collection and taxonomy in the wind industry, making it difficult for wind farm owners to choose between the Reference Designation System for Power Plants (RDS-PP) and the International Electrotechnical Commission's (IEC) 61400 terminologies for categorizing failures and linking them to the components and the sensors. This would allow a uniform terminology across the industry which will aid data sharing and allow more effective and data-driven operational decision making.

1.2 Work Context

The work has been carried out while placed at EDF Energy R&D UK Centre and it has been tailored towards the company's operational and research needs. A lot of the insights provided in the thesis have been derived from work carried out with subject matter experts from different parts of the business. This EngD thesis mainly covers the asset owner and operator's view, challenges and addressed questions, but the methodologies and results presented can be useful to all the offshore wind O&M stakeholders.

1.3 Aim and Objectives

The aim of this EngD project is to create tools and methodologies with a focus on data-driven decision making that will support O&M cost reduction efforts for offshore wind farms, with a focus on the owner and operator's perspective.

The specific project objectives are as follows:

- Identify the current state of the art on managing and utilizing operation and maintenance data of offshore wind farms and the associated challenges of the wider sector.
- Create tools and methodologies that will improve the operational work flow and aid in better use of the generated operational and maintenance data.
- Create case studies and useful results that will inform offshore wind farm owners and operators to better manage their assets. Aiming to increase the lifetime of the offshore wind farms and to reduce the overall O&M costs, while at the same time improving the assets' availability, through data-driven decision making processes.

1.4 Contribution to Academic and Industrial Knowledge

This thesis seeks to make contributions to both academic and industrial knowledge in the field of offshore wind O&M. The academic contribution is accomplished by presenting data, industrial practices and applied analysis, including:

- Methods for managing "big data".
- Data analytics techniques validated for offshore wind turbines.
- Management processes for offshore wind O&M decision making.

The industrial contribution of this thesis is achieved by investigating new techniques that could be used by offshore wind farm operators such as EDF to follow data-driven methodologies and practices for the asset management of their current and future offshore wind farm developments, including:

- Recommendations for improved industrial processes for data collection, architecture and data management.
- A framework and application of integrating inhomogeneous operational data sources.
- Applications of reliability analysis and correlation of environmental conditions with wind turbine component failures.
- Application of predictive analytics for a wind turbine gearbox and for wind turbine fault alarms.
- A framework and application for risk-based operations.
- Application of failure root cause analysis of a wind turbine yaw system failure.
- Strategic maintenance planning improvement, with a focus on reliability.

1.5 Thesis Outline

The thesis is divided into four parts, which are outlined along with the individual chapters in Figure 1.5.

Part I discusses the formulation of the research question. Chapter 1 has introduced the thesis background and problem statement. Chapter 2 presents a literature review that identifies the state of the art in O&M for offshore wind farms, with a focus on data management, reliability analysis, data analytics and operational planning. Key results from significant bodies of work are explored and previous studies are compared and critically assessed. The literature review concludes with the current research gaps in the data management and data-driven O&M context, as the focus of this thesis. At the end of this chapter, the overarching approach to research is presented, along with a description of the Teesside offshore wind farm.

Part II introduces the different tools and methodologies explored and developed as part of this research. Chapter 3 examines the different methods followed for operational workflow improvements, data management, reliability analysis and machine learning techniques. Chapter 4 presents the methods used to utilize the above techniques during the planning of the maintenance operations through a decision support framework for the machine learning models, risk-based operations and through the development of an O&M strategy planning tool.

Part III presents the results for the operational analysis and the maintenance optimization case studies. Chapter 5 includes examples of asset reliability analytics, failure root cause analysis and diagnostic & predictive analytics. Chapter 6 includes alarm prediction, improved O&M planning and a risk-based inspection case study.

Part IV presents the discussion and concluding remarks. Chapter 7 discusses the results of the analysis, relating them back to the objective of developing a data-driven decision making processes for offshore wind and asset management. It provides the main conclusions in the context of offshore wind O&M data-driven decision making. Further work in the research area is also discussed.

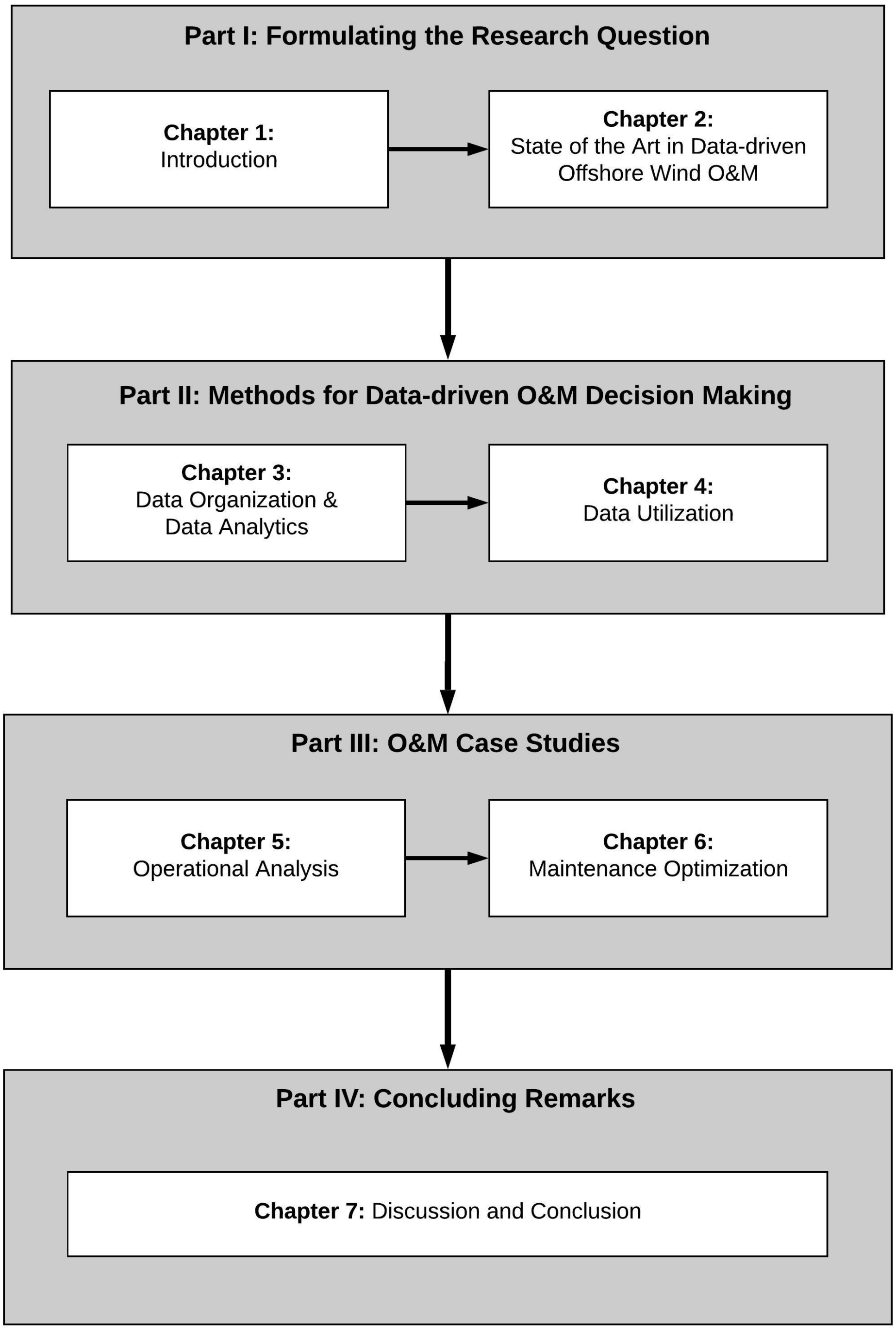


Figure 1.5: Thesis structure

Chapter 2

State of the Art in Data-driven Offshore Wind O&M

This chapter presents a state-of-the-art review for data-driven O&M in offshore wind. These include O&M strategies, reliability analysis, data management, data analytics and operational workflow management for the offshore wind sector. It provides a critical assessment and identifies the research gaps that have been used as the building blocks of this thesis that have informed the research question for this work. It also outlines the overarching approach to research that includes the overall methodology and test site for all the case studies presented in this thesis.

Typical offshore wind operations are shown in Figure 2.1, where different parts of an offshore wind farm are presented including the export and inter-array cables, the substation, the wind turbine foundations (fixed and floating), the tower and the blades. Operations include crew transfer to the turbine, inspection and cleaning of the foundations, cable survey and repairs, heavy-lift operations, helicopter transfer, bathymetric survey and remote surveillance.

The focus of this thesis is on data-driven decision making for the reduction of the offshore wind O&M costs. The different components of the data-driven decision framework used for this thesis are shown in the Venn diagram, Figure 2.2. These are the generated data, the analysis of the data, the processes followed and the planning strategies for executing the maintenance strategies. There are four overlapping areas

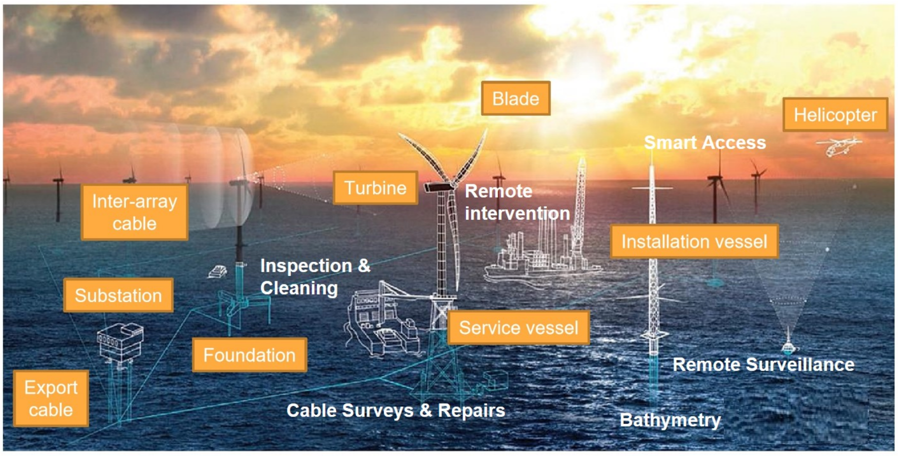


Figure 2.1: Offshore wind O&M schematic [13].

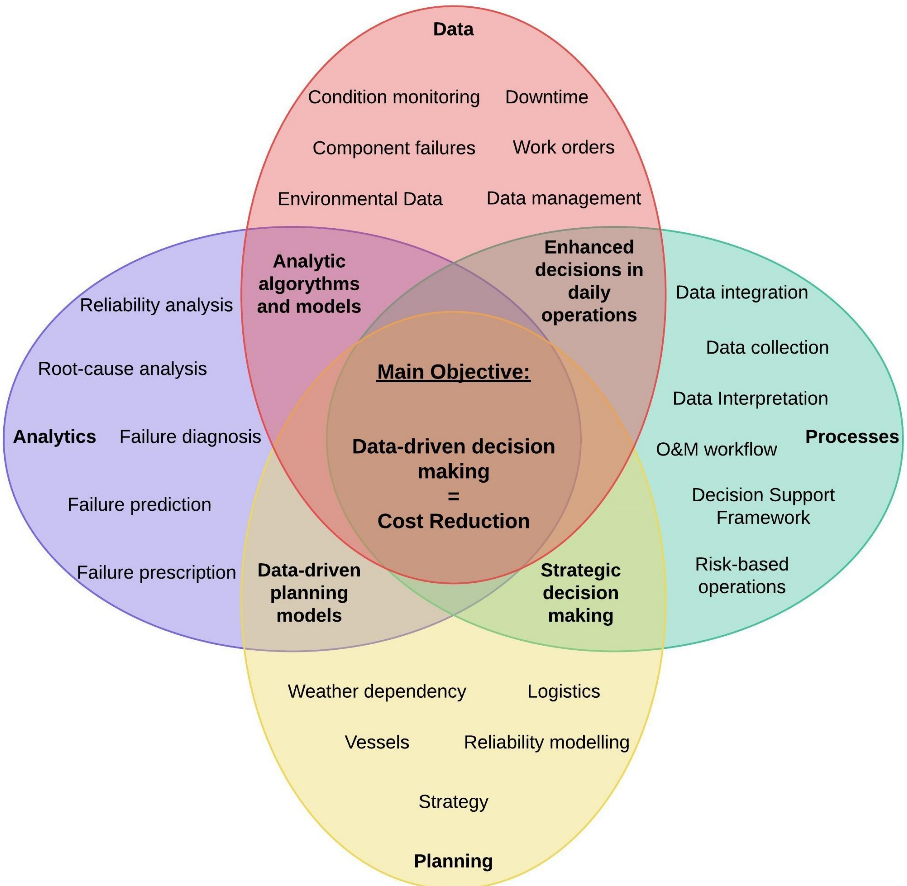


Figure 2.2: Venn diagram showing the four pillars (data, analytics, processes, planning) of offshore wind O&M data-driven decision making.

in the Venn diagram. The top red circle shows the data sources and the different systems that are being used in order to capture the state of the turbines and the environmental parameters and how these data sources can be better managed. The purple circle on the left indicates the analysis that has been done in this thesis, including creating reliability profiles for turbine components, investigating the root cause analysis of failures, predicting and diagnosing failures. The green circle shows the improvements in the processes, including the data collection, data integration and improvements in the operational workflow through risk-based operations and decision support frameworks. Finally, the bottom circle shows the planning, which includes logistics, vessels and reliability modelling for strategic long-term decision making. The data and analytics are combined to generate analytical algorithms and models, processes in combination with the generated data can provide enhanced decisions for daily operational decision making, analytics and strategic planning can lead to robust data-driven planning models and processes and future planning can lead to a better strategic decision making, by transferring the lessons learnt from the operations. A successful combination of all these four components will enhance data-driven decision making in offshore wind O&M and help reduce the operational expenditure (OPEX).

2.1 Maintenance Strategies and Planning

O&M of an offshore wind farm refers to the phase after the commissioning of the farm, when the turbines are running and maintenance operations are taking place. This is normally for a duration of 20-25 years. Recent studies have shown that O&M costs of offshore wind farms could reach up to 30% of the total lifetime costs of the farm [10], making it an integral part for cost reduction. This cost reduction can be achieved by implementing appropriate maintenance strategies and applying robust and tested remote monitoring systems (RMS).

The ideal maintenance strategy should offer high availability while providing the lowest possible operational costs. The main maintenance strategies adopted can be categorized as preventive and corrective.

2.1.1 Strategies

For offshore wind farms, the maintenance strategies are governed by the weather conditions, as unfavourable weather can delay previously scheduled maintenance activities. Thus, the different maintenance decisions need to take this into consideration, as the date of the maintenance task might need to be shifted according to the weather conditions. Maintenance strategies could be either preventive or corrective. Preventive maintenance could also include calendar-based, condition-based and weather-based maintenance. This is the type of maintenance that is planned, usually in fixed intervals and could include regular inspections, replacements and retrofits. Preventive maintenance strategies could include planned component replacements, retrofits or annual maintenance and inspection campaigns on the turbine's critical components and access points, such as lifts, ladders and pathways. Corrective maintenance includes planned and unplanned maintenance, which are determined by the condition of the asset and driven by inspection or RMS's inputs.

The ability to gain good access to the farm is pivotal to performing the operations needed. This question of access is becoming more important as wind farms are moving further offshore. Daily maintenance operations are mainly performed using crew transfer vessels (CTVs), which are usually catamarans able to carry 10-15 technicians. For urgent operations, faster transfers or when weather conditions are not favourable for a CTV, helicopters can be used to carry technicians to the turbines. For turbines located far offshore and for longer operations, service operation vessels (SOVs) are used, able to accommodate tens of technicians and the required equipment and spare parts and stay offshore for numerous weeks. For heavy-lift operations jack-up vessels are used, able to transport, lift and replace large turbine components such as gearboxes, blades and main bearings.

Typically for at least the first five years of operation of the farm the maintenance of the turbines is performed by the OEM, as part of the turbines' warranty. A key benefit of outsourcing maintenance is that it helps the wind farm operators to gain access to trained technicians and experts. On the contrary, there is less availability of operational data, visibility of operations and influence in decision making. It can however free up internal manpower, leading to a reduction in overhead costs. Moreover it can reduce the risk to the wind farm owner, as it gets shifted onto the maintenance service provider. There is generally a shift to longer term contracts by OEMs, up to 15 years, which is due to low operational costs as the banking sector demands longer term service contracts for the financing of new projects and additionally due to the higher uncertainty of newer and larger wind turbine models. However, in certain cases, according to Orsted, in-house O&M can be cheaper [14]. One of the key reasons for this is the sense of ownership the team now has over the farm. However, there is a right balance that needs to be carefully examined when deciding to outsource or build in-house capabilities. The larger the uncertainty with in-house maintenance is, the more capital expenditure that is needed to build and train the team and constantly ensure that they are up to date with new upgrades and turbine models. However, it could just be a matter of portfolio size, as a study of Accenture concluded [15]. It was shown that large operators with global portfolios of more than 2GW have relied heavily on sophisticated control rooms to pool investment and optimize cost and performance across their portfolios. Some players with up to 1GW within a single country have largely relied on taking the O&M value chain in-house for greater control, enabling optimization across sites in-country. Meanwhile, medium-size independent power producers have focused more on optimizing their energy management capability, choosing to outsource most of their O&M activity to OEMs. There are definitely several parameters that need to be taken under consideration and this decision will be very site or operator specific.

2.1.2 Reliability-centered Maintenance

Another important type of maintenance that has been widely used by the aerospace industry is reliability-centred maintenance (RCM). It integrates preventive maintenance, predictive testing and inspection (PT&I), repair (also called reactive maintenance), and proactive maintenance to increase the probability that a machine or component will function in the required manner over its design life-cycle with a minimum amount of maintenance and downtime, as shown in Figure 2.3 [16]. The International Energy Agency (IEA) Wind Energy Task 11 has included RCM as part of preventive maintenance strategies for wind turbines. RCM will find the optimum time for maintenance through analysis of past data.

As Figure 2.4 indicates, both condition-based maintenance (CBM) and RCM could have the same results in monitored components. As an effect, it can be assumed that those two maintenance strategies would offer the best solution if implemented consecutively. CBM can be followed for components that are being remotely monitored and RCM for non monitored components. This can keep the spare parts up-to-date and allow more time for planning. The limitation of RCM is that it requires a lot of years of reliability information.

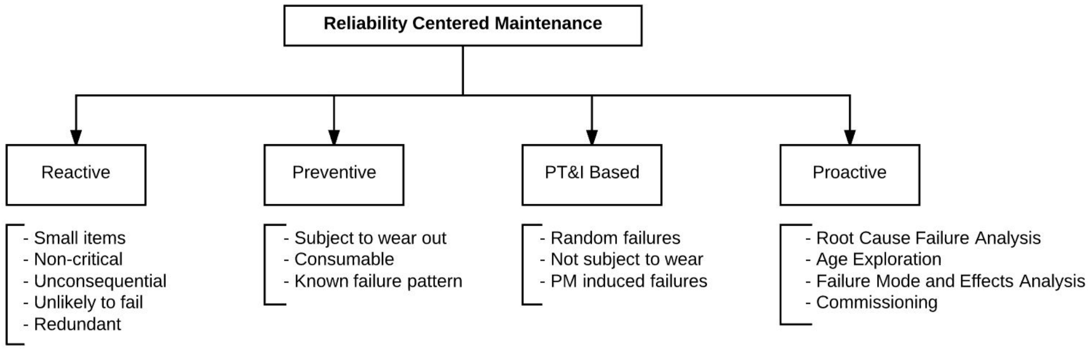


Figure 2.3: Reliability-centred maintenance applied by NASA [16].

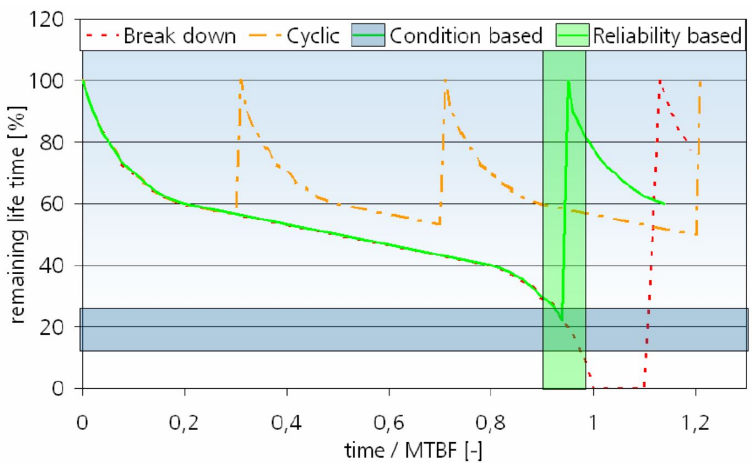


Figure 2.4: Comparison of Cyclic (preventive), CBM and RCM for the remaining lifetime of a component against the mean time between failures (MTBF) [17].

Another type of maintenance implemented in a lot of industries is the risk-based maintenance (RBM). RBM can be considered as a more advanced step from RCM, as it combines the following:

- The probability of failure of a component or an assembly.
- The consequences of the failure, usually identified by a risk impact matrix.

2.1.3 Risk-based Operations

A common approach is to perform annual inspections to all or a sample of the turbine's components. This usually includes all the internal and external access equipment, along with any mechanical, welded and electrical connections. However, this is a time consuming and costly process, taking into account that most of the inspected structural components have expected lifetime that exceeds the 20 year designed lifetime of the wind farm [18]. Moreover, most of the drivetrain components are fitted with condition monitoring systems (CMS) and supervisory control and data acquisition (SCADA) systems, as discussed in more detail in Section 2.2. At the same time, several of these inspections have the opportunity to be combined with other planned operations, to share resources and reduce costs.

Consequently, the concepts of RCM and risk-based inspection (RBI) and RBM have been developed and successfully implemented in several industries, as discussed in Section 2.1.2, including the aerospace [16], chemical [19], oil & gas [20] and nuclear [21]. Frameworks and studies have been computationally applied in the offshore wind industry [22], but the offshore wind farm operators are still sceptical in the implementation of such a framework, mainly due to the lack of field data. The main difference between RCM and RBI/RBM lies in the level of analysis undertaken to define maintenance and inspection intervals. Where cost-risk optimisation is incorporated to help tailor maintenance regimes, it is considered as an RBI technique [23]. Both methods show an attempt to utilize the existing knowledge of the asset in order to perform future maintenance actions effectively and to reduce the associated overall costs. As operational cost reduction is a critical aspect for future offshore wind farm projects, this thesis is investigating risk-based operation techniques.

The definition of risk depends on the context in which it is applied. A recent review article has identified and grouped the different definitions of risk concepts in the following [24]: (i) Expected consequences, (ii) Probability and scenarios/ consequences/ severity of consequences, (iii) Event or consequence, (iv) Consequences/ damage/

severity of these and uncertainty and (v) The effect of uncertainty on objectives. The International Organization for Standardization's (ISO) standards have defined risk as a measure of the likelihood of failure and its potential impact [25]. This could refer to the system as a whole, or to its individual components, as well as to people, the environment and economic losses [26]. It is usually assessed by a risk matrix that includes the likelihood (y-axis) and consequences (x-axis) of failure. Several authors have challenged the effectiveness of a risk matrix and its outputs [27, 28]. As an effect, recent improvements of the above definition have also added an uncertainty interval to the equation [29]. Regardless of the different weaknesses of this method, it is still considered a robust, pragmatic method, being applied by several industries today.

Despite the evident cost reduction opportunities of risk-based operations [30] and the successful implementation in other industries, research and implementation in the offshore wind field is limited.

2.1.4 O&M Strategy Modelling and Simulation Tools

In order to better decide on the different O&M strategies, several tools have been developed over the years, trying to compare different maintenance scenarios and estimate their overall OPEX cost. A list of these tools that are still under development or commercially available when this literature study was conducted, are listed below:

A. ECUME O&M is an internal tool developed by EDF R&D [31], which uses a Markov-chain weather simulation model to generare different weather scenarios. The tool is running a Monte Carlo Simulation (MCS) to randomize the different input parameters, which include all the typical O&M parameters and different probabilistic failure rates inputs. Outputs include different cost and availability calculations and breakdowns.

B. StrathOW-OM is a tool developed by the University of Strathclyde for an offshore wind operator [32]. The tool uses a failure event generation based on a Weibull distribution and uses a probabilistic weather model. It generates the different scenarios using a Monte Carlo simulation and it produces cost and availability outputs. It uses a decision support model based on Bayesian networks.

C. NOWIcob is a tool developed during the LEANWIND project, which was funded by the FP7 European Commission research and innovation activities. The tool is a complete lifecycle software made up of different modules, that has all the typical inputs and allows the input of failure rates or Weibull distributions [33]. It uses a discrete time-based model and its unique features include the consideration of wake and electrical system losses.

D. MAINTSYS/ Uis Sim Model is a tool initially developed at the University of Stavanger and then accelerated into a startup. The tool uses agent-based and discrete event paradigms [34] for simulating the different scenarios. The weather data can be either historical or synthetic with a Markov-chain process and the failures are time-dependent using a non-homogeneous Poisson process for failure generation.

E. ECN O&M is one of the oldest tools in the market. It uses a MCS to generate the different scenarios and it is deterministic, but can incorporate a stochastic module of failure rates, which outputs cumulative distribution function of costs [35]. The tool has been validated against a period of 4 years of an operating offshore wind farm.

F. O2M is a consulting tool build by DNV-GL. It uses reactive maintenance strategy only and it has been validated against an existing offshore wind farm.

G. Mermaid is an installation tool developed by James Fischer (ex. Mojo Maritime), that is also adding an O&M module to it [36].

H. O&M Expert is being developed by MAREi at the University of Cork. It uses MCS and it is built in modular blocks. It includes different maintenance strategies and it can also perform layout optimization.

More detail and a comparison of the different tools is provided in Table 2.1. Overall the tools have unique capabilities and different limitations. It is important for the tools to be modular and scalable in order to incorporate new features and capabilities in the future, as the industry is transforming. The main limitation of all the tools is that they lack a way to run multiple cases at the same time and run complex optimization scenarios. Some tools such as NOWIcob and O&M expert have been designed with a lot of features and complexities that might also require detailed inputs, which might not be available and could add more uncertainties to the final results. On the contrary ECN O&M has a more simplistic approach, assuming that most of the complexities in the decision making would average out over the years, allowing a faster running time. MAINSYS has focused on the visualization functionalities, allowing a better understanding of the different scenarios implemented.

Table 2.1: List of O&M tools and their features.

Tool	Weather	Reliability	Logistics	Advantages	Limitations
A	Synthetic/ historical	Failure rates and probabilistic values	MCS	Easy to use and run	Difficult to implement changes in the code
B	Historical	Probabilistic	Bayesian network	Modular design	Lack of user interface
C	Stochastic	Weibull	Discrete time-based model	Wake and electrical system losses, life-cycle costs	Detailed inputs needed might not always be available
D	Synthetic/ historical	Time-dependent Poisson process	Agent-based	Cloud solution, large database of turbines and vessels	Lack of optimization capabilities.
E	Synthetic/ historical	Failure rates and Probabilistic	MCS	4 year validation against a commercial wind farm	Mainly deterministic inputs.
F	Historical	Stochastic	MCS	Validated against commercial wind farm's serial failures, large pool of failure rates.	Only able to simulate reactive maintenance.
G	Historical	Probabilistic	MCS	Built on an existing installation tool	Lack of operational implementation
H	Synthetic/ historical	Failure rates & probabilistic	MCS	Layout optimization, modular blocks and cloud solution	Lack of operational implementation and optimization capabilities.

Some of the above tools have been benchmarked against each other for some simple scenarios in two different papers [37, 38]. The simulation results showed major differences of up to 60% between the different scenarios and tools for the availability and OPEX cost calculations. This is due to the different modelling approaches and assumptions that the tools have. In more detail, the models have different approaches in the following aspects:

- Different failure generation engines.
- Maintenance operation modelling differences.
- Different approach on modelling vessels' chartering.

This raises questions on how realistic and trustworthy those models can be. It is also interesting to mention that none of the tools has been tested against full lifetime OPEX costs of an offshore wind farm.

2.2 O&M Data

As offshore wind farms are unmanned and located far out in the ocean, they need many sensors to monitor their condition and performance. For example, a single turbine at Teesside offshore wind farm consists of more than 40 sensors, creating around 1500 individual data streams of information every day. In addition, information about daily maintenance operations, wind and wave information from meteorological masts and wave buoys, reports, pictures and videos from inspections performed are also logged. An example of the different data sources for Teesside offshore wind farm is shown in Figure 2.5.

2.2.1 Remote Monitoring

Condition monitoring is defined as the continuous evaluation of the health of a plant and equipment throughout its serviceable life. Condition monitoring of machines has increased in importance as engineering processes are being automated and human interaction is being reduced. Consequently, for wind turbines in offshore environments condition monitoring equipment is crucial, as intervention for inspection might not be possible due to weather conditions and it can also be costly.

The monitoring of an offshore wind farm typically includes three types of RMS's, with different sensor types and sampling rates. These include SCADA and vibration CMS, which are described below in more detail.

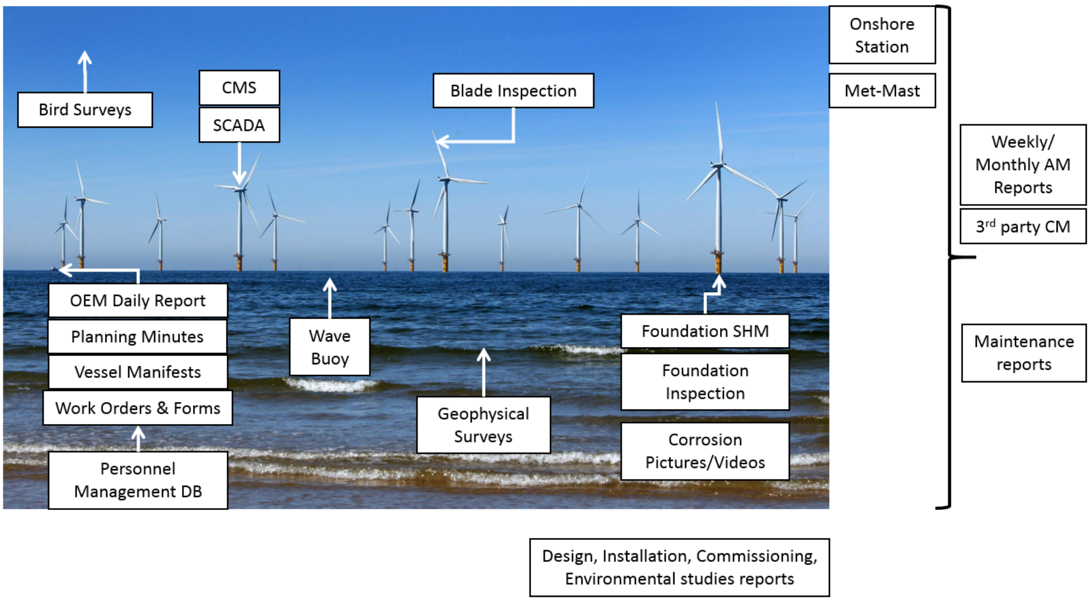


Figure 2.5: Teesside Data

2.2.1.1 Supervisory Control and Data Acquisition

SCADA systems have been originally used by the oil & gas industry, where temperature, pressure and other types of sensors are measured remotely, where it was not possible to monitor the equipment otherwise. Such systems have also been popular for wind turbines, since it is hard to continuously monitor unmanned structures, while located at a large distance. SCADA systems provide a range of sampling frequencies ranging from 1 second sampling to 10 minute average data from different sensors employed on the turbine.

The benefit of the SCADA systems is that it can provide information from different parts of the turbine, including wind speed, rotational speed of components, temperatures, energy production and alarms in a relatively easy way to analyse and interpret the information. On the other hand, its low data rate does not allow the depth of analysis required for precise failure diagnosis. Overall, it is considered by OEMs and operators as a good value for money, although it has been argued that SCADA systems do not accurately inform the operator about the condition of the wind turbines [39]. This is because the systems do not allow an in-depth analysis of the root causes and the location of the fault, sometimes leading to false alarms. As sensors are monitoring larger components with multiple subcomponents, it can be challenging, just from the data, to identify the exact location of the failure and whether it is an actual component or sensor fault. Therefore, an element of human intervention might still be needed.

2.2.1.2 Condition Monitoring System

CMS first appeared in the wind industry after pressure from insurance companies, resulting from large number of gearbox failures [40]. The technology was adapted from previous condition monitoring (CM) knowledge of rotating machines, which later became part of the wind turbine certification process. To ensure a minimum functionality of the CMS, a certification procedure was worked out, resulting in a set of guidelines [41], which has evolved into a DNV GL standard [42] required to be followed by any CMS.

Different CMS techniques include vibration, oil debris and analysis, as well as strain and electrical gauge signals. Turbine OEMs and operators use a variety of different combinations of these techniques depending on the turbine type.

In general, CMS include a bigger variety of sensors, at higher sampling frequencies compared to SCADA, which is more appropriate for early fault diagnosis and prognosis [43]. CMS's typically generate a large number of data points every second, which cannot always be saved due to storage limitations. For analysis purposes these data are usually processed and saved into other formats. Examples of different analysis methods used for CMS data, include:

- Fast Fourier Transform (FFT), which is an algorithm that samples a signal over a period of time and divides it into its frequency components.
- Cepstrum, which is the inverse Fourier transform of the logarithm of the signal's spectrum.
- Envelope, which is generated by passing the time domain signal through a ban-pass filter and then through an enveloper, to extract the repetition rate of the spiky bursts of energy.
- Root mean square (rms), which is the arithmetic mean of the squares of a set of numbers.

CMS's are usually deployed at the drivetrain of the wind turbine, on the gearbox, generator, main bearing, as well as on the tower and the foundation. The later ones are know as structural health monitoring systems (SHM).

Drivetrain CMS A typical system architecture is shown in Figure 2.6 [44]. The data are generated from the sensor, usually daily or in specified intervals, and they are processed into different forms. Then, they are categorized in the different power bands that have been captured for easier analysis. The final files are then archived and any previous ones are deleted from the servers. Operators would typically only have access to the final set of archived files.

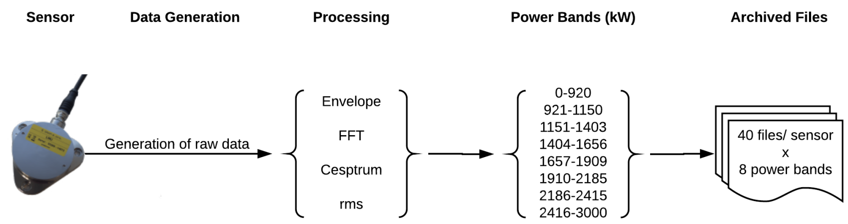


Figure 2.6: Gram & Juhl CMS drivetrain data processing.

Structural Health Monitoring SHM systems use acceleration and displacement sensors to measure the vibrations of the turbine structure on the foundation [45]. Most of the wind turbine foundations installed have been designed with high safety factors, since the recommended standards have been translated from the oil & gas industry [46]. Thus, SHM systems on the turbine foundations are not always used. Due to the high cost of these systems, operators are instrumenting a representative sample of around 10% of the turbines in the farm. These could be the ones that are exposed to harsher environmental conditions or turbine foundations with known issues.

Oil Particle Counter Due to the importance of the gearbox failures, in terms of downtime, oil particle counters are fitted in modern wind turbine gearboxes. Gears and bearings are wearing components and produce ferrous and non-ferrous debris from their normal operation. These systems can distinguish between the different sizes and types of debris and provide a count to the CMS, which can then be translated into a slow or fast component degradation.

In general, CMS data can provide a very detailed overview into the different turbine components at high frequency. This can ease the fault detection and diagnosis capabilities of the operators. On the contrary, due to the large amount of data generated it can be difficult and costly to analyze them and will also require experienced personnel.

2.2.2 Operational Data

Operational data include any other sources of information generated by the wind turbine technicians, engineers and asset managers. These include work orders, which are capturing the work carried out by technicians during maintenance activities, operational and inspection reports, images and videos.

2.2.3 Data Management

The intelligent use and intepretation of the turbines' monitoring system data described above on a real time basis could potentially provide ways that will allow wind farm owners and operators to reduce the O&M costs.

There is a limited amount of literature in the area of O&M databases (DBs) in the wind industry. A framework for data integration by following the IEC standards was suggested by [47], as well as its proposed ontology at [48]. The framework includes three patterns; data source handling (information from various sources with different formats), semantic model (concepts with semantics) and information provisioning (efficient ways to access reliable information). Although these studies have been proven useful for understanding the proposed framework, the concepts and the ontologies, they have not been applied by the authors on wind farm data. Moreover, the monitoring systems for the proposed model are not specified, making it hard to translate theory into practice.

Big data challenges have begun to appear in the offshore wind industry, as the data generated from multi-MW wind turbines, are being produced at high frequencies, large volumes, different formats and they need to be trustworthy in order to be utilized [49].

A review of different DBs and big data techniques [50] has concluded that big data can be treated in two ways, as shown in Figure 2.7:

- For structured data, which in this case are the met mast, wave buoy, SCADA, CMS and SHM, it is suggested to manipulate them in a relational database (RDB) in order to meet the schema-on-write constraints, once they have been pre-processed. RDBs are usually using a Structured Query Language (SQL).

- In the case of unstructured data, which include reports, images, etc., it is recommended to initially store them into a distributed DB that will then be retrieved after meeting the schema-on-read constraints and can be processed using a No-SQL database.

This approach differs from the one followed by [51], where the use of a No-SQL DB approach is suggested, even for structured CMS and SCADA data. The justification provided is that SQL DBs are limited in scalability. This is not always the case, as with a well defined relational schema, the RDB can be scalable, as it has been shown in [52, 53].

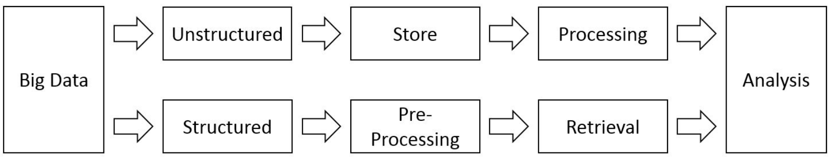


Figure 2.7: Transforming big data for analysis [50].

2.2.4 Data Architecture

It is important to also consider the data architecture framework where the data will be stored and further processed. There are two main processes; the centralized and the distributed. Both of them could be stored and managed on servers or on the cloud. Regardless the different data architecture process, it is important to have an integration framework in place, such as the one shown in Figure 2.8 to make sure that the different parts of the business have access to the right data.

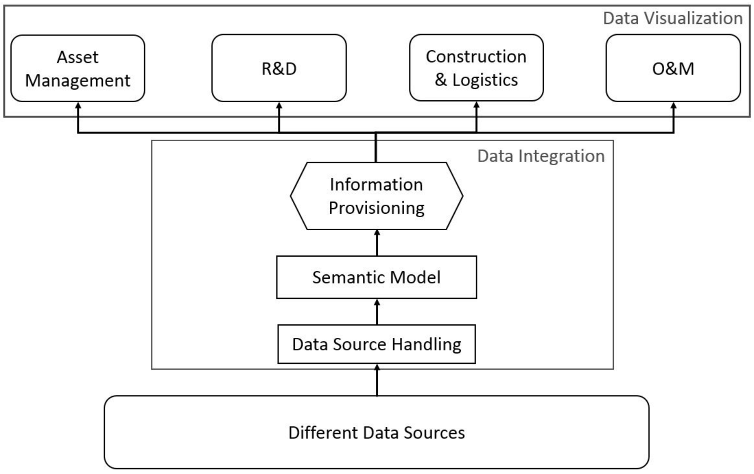


Figure 2.8: Data integration framework by [47].

2.2.4.1 Centralized Process

All data are collected to a single centralized storage area and processed upon completion by a single computer system with large memory, processor and storage capabilities. This process can be useful for cases where minimum manpower is desirable, as well as when the collection and consumption of the data occurs in the same location.

2.2.4.2 Distributed Process

The data can be distributed across different locations and data centres and processed into a centralized storage. There are several distributed architectures, but they are all sharing the same principles. The advantages of such a system are that it is scalable, flexible and allows parallel processes. As a downside, there is a lack of data and process redundancy, it is resource intensive and requires a large storage capacity.

2.2.5 Reliability Analysis

Reliability metrics are used in the wind industry as well as in numerous other industries in order to describe quantitatively the performance characteristics of a system or its components. Reliability is defined by the British Standards (BS) as the ability of an item to perform a required function under stated conditions for a stated period of time [54]. Wind turbine reliability databases are mainly using aggregated failure rate data for different wind turbine components and systems, allowing wind turbine operators to benchmark against each other or to measure performance against predefined targets. This section compares the reliability databases used by the industry, as well as different types of reliability assessment.

2.2.5.1 Reliability and Failure Rates

Reliability Although the wind industry has tried to standardize several terms, and different authors have highlighted the importance of this, there are still different definitions used for the taxonomy of components, (sub)assemblies and (sub)systems [55].

Another important aspect to consider is the definition of failure in the offshore wind sector, since different definitions have been used in the literature. The IEC 60050-191 standards [56] define failure as the termination of the ability of an item to perform a required function. A similar definition is given by the oil & gas industry in the Offshore and Onshore Reliability Data (OREDA) Handbook, stating that a failure is the termination or the degradation of the ability of an item to perform its required functions [57]. A more detailed breakdown of the failures is described by OREDA into the following: (i) Complete failure of the item (meaning that the component will no longer be in use), (ii) Failure of part of the item that causes unavailability of the item for corrective action, (iii) Failure discovered during inspection or preventive maintenance that requires repair, (iv) Failure on safety devices or control/monitoring devices that necessitate shutdown or reduction of the capabilities of the item below specified limits.

At the same time, it is also stated that unavailability due to preventive or planned maintenance or due to shutdown because of external conditions (unless there is a maintenance activity) is considered as outage and not as failures.

Failure Rates In the offshore wind industry, there has not been a robust uniform failure definition. RELIAWIND has defined the failure as any event that includes the stoppage of a turbine for one or more hours that requires a manual reset as a minimum in order to return back to operation [58]. The definition has been used by [59]. Another definition that appears in [60, 61, 62] suggests that a failure is characterized as any visit to the turbine outside of a scheduled operation, in which material is consumed [60]. The latter definition is closer to the one defined by OREDA and it is considered more accurate for the purposes this thesis, as it is based on an industry that has been operating for numerous years under harsh environmental conditions. It also allows easier comparison with the results of these studies.

A failure rate λ is defined by the OREDA Handbook [57] as the observed number of failures n over the aggregated service time τ, Eq. 2.1. This is considered for a homogeneous sample that includes identical items that have been operating under the same operational and environmental conditions.

$$\lambda = \frac{n}{\tau} \tag{2.1}$$

Eq. 2.1 was then translated by [63] for the offshore wind industry in Eq. 2.2 and used by [55, 61, 60, 12, 63]. The numerator is the number of failures in all periods ($n_{i,k}$), per turbine (N_i), with I being the number of intervals for which the data are collected and K the number of subassemblies. The denominator is the sum of the length in hours of the interval (T_i), divided by the total number of hours in the year.

$$\lambda = \frac{\sum_{i=1}^{I} \sum_{k=1}^{K} \frac{n_{i,k}}{N_i}}{\sum_{i=1}^{I} \frac{T_i}{8760}} \tag{2.2}$$

2.2.5.2 Reliability Databases

Several reliability databases have been generated for onshore and offshore wind farms. A list of these, along with the dates, sample and country representing can be found in Table 2.2. They all include samples from different turbine manufacturers and wind farm locations. Such databases allow wind farm developers and analysts to create different reliability statistics for wind farm components and predict their remaining lifetime. At the same time, they can compare different operational locations and turbine manufacturers against each other, which can inform future wind farm developers on their choices. The issue with all these databases is that they are all using different standards, processes, definitions and terminology. This makes the comparison and the utilization of the provided information very challenging. Moreover, when wind farm operators participate in more than one of these projects, they need to duplicate work, since different inputs or data manipulation techniques are required.

Table 2.2: List of wind turbine reliability databases.

Database	**Date**	**Sample**	**Type**	**Country**
WMEP [64]	1989- 2006	1,500	Onshore	Germany
LWK [65]	1993- 2006	>650	Onshore	Germany
Windstats [66]	1994- 2004	up to 7,000	Onshore	Germany & Denmark
VTT [67]	1996- 2008	78	Onshore	Finland
Elforsk (Vindstat) [68]	1997-2010	936	Onshore	Sweden
Offshore-WMEP [69]	2007- today	297	Offshore	EU (mainly Germany)
RELIAWIND [70]	2008- 2011	350	Onshore	EU
CREW [71]	2011- today	>800	Onshore	USA
SPARTA [72]	2013- today	1,045	Offshore	UK
Strathclyde [73]	2007- 2014 (est.)	350	Offshore	UK

For example, by comparing the failure rate outputs of (i) RELIAWIND, (ii) SPARTA and (iii) Strathclyde studies (Table 2.2), the reported failure rate is (i) 16-20, (ii) 15.8 and (iii) 6.8 per turbine/year respectively. This means that 2 different offshore wind turbine studies present average failure rates with 2.3 times difference between them and an onshore wind, which has higher failure rates compared to the offshore ones. This difference could be due to several different factors, including different manufacturers and sites, location in the farm or different maintenance strategies. Some of the most important reasons could be the diverse terminology of a failure used in the industry, the age of the assets, the turbine make and the location of the wind farm.

Figure 2.9 maps the different databases with regards to their competences. The databases are classified under three categories; benchmarking (databases that include basic information on number of failures, downtime, etc.), reliability analysis (databases that additionally include failure rates per time and component/ (sub)assembly, etc.) and the ability to facilitate O&M decision making (databases that additionally include information on repair cost, detailed information on repair actions, failure mode and effects analysis (FMEA), etc.). As it can be seen, only the WMEP and Offshore-WMEP databases provide an approach that does not only benchmark and/or create reliability metrics, but is also envisioning the generation of a database of faults that will facilitate faster fault diagnosis and offer O&M decision support.

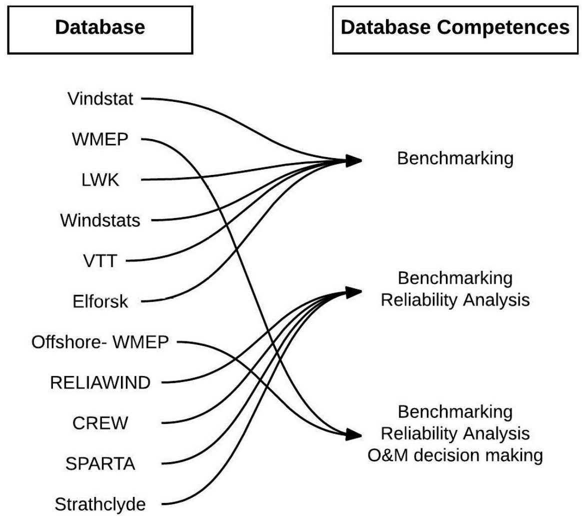
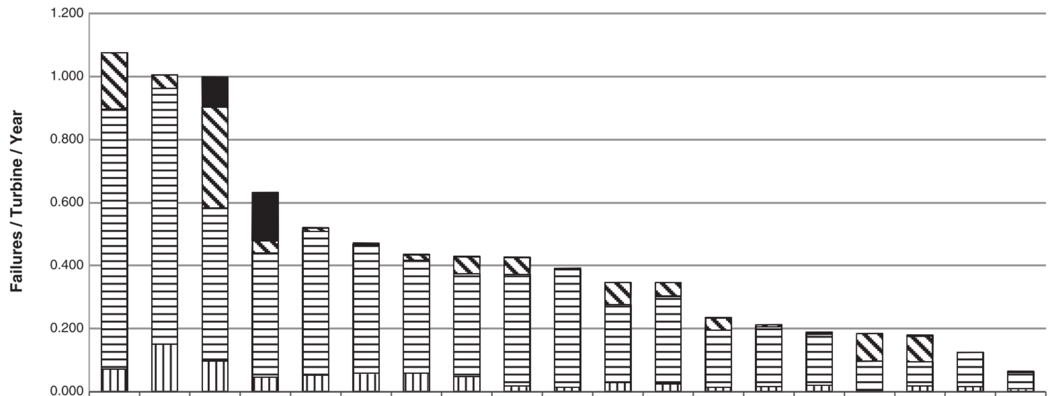


Figure 2.9: Reliability database competences.

Offshore Wind The most recent offshore wind reliability analysis has been presented in two studies from the same authors [61, 62] and presented in Figures 2.10 and 2.11. They provide a detailed analysis of a large population of ∼350 offshore wind turbines of up to 9 years of operation, which gives useful inputs for O&M cost models. The methodology that these studies used, does not fully utilize all the capabilities of the provided data from the operators, as it does not foresee how these data could be further aggregated. Another offshore wind study, assessing the performance of a number of Round 1 projects for their first few years of operations, does not present any reliability statistics for them [74]. One of its main conclusions in order to increase the farm's availability is to encourage more proactive O&M; this will raise O&M costs and at the same time increase energy yield, which will mitigate high capital costs by improving annual energy production. Another study is based on 36 turbines and analyses their stoppages rather than the failure rates, which includes stoppages for any type of operation, even scheduled maintenance and inspection, which makes it impossible to calculate the actual failure rates from it [75].



	Pitch / Hyd	Other Components	Generator	Gearbox	Blades	Grease / Oil / Cooling Liq.	Electrical Components	Contactor / Circuit Breaker / Relay	Controls	Safety	Sensors	Pumps/ Motors	Hub	Heaters / Coolers	Yaw System	Tower / Foundation	Power Supply / Converter	Service Items	Transformer
Major Replacement	0.001	0.001	0.095	0.154	0.001	0.000	0.002	0.002	0.001	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.005	0.000	0.001
Major Repair	0.179	0.042	0.321	0.038	0.010	0.006	0.016	0.054	0.054	0.004	0.070	0.043	0.038	0.007	0.006	0.089	0.081	0.001	0.003
Minor Repair	0.824	0.812	0.485	0.395	0.456	0.407	0.358	0.326	0.355	0.373	0.247	0.278	0.182	0.190	0.162	0.092	0.076	0.108	0.052
No Cost Data	0.072	0.150	0.098	0.046	0.053	0.058	0.059	0.048	0.018	0.015	0.029	0.025	0.014	0.016	0.020	0.004	0.018	0.016	0.009

Figure 2.10: Offshore wind failure rate for subassemblies and cost category [61].

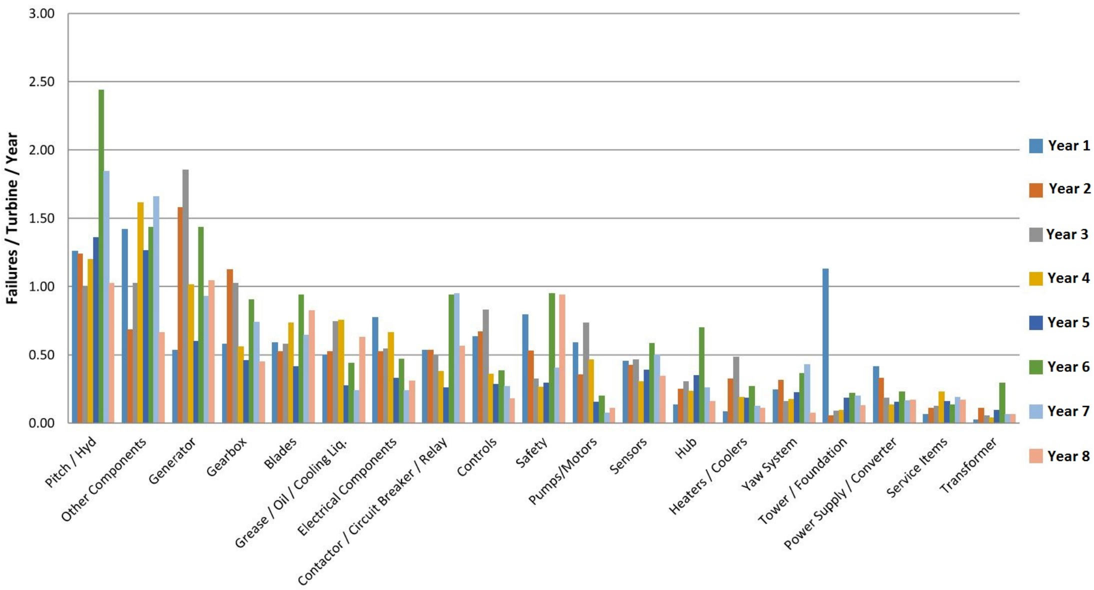


Figure 2.11: Offshore wind failure rates through time [62].

Onshore Wind Onshore wind turbines have a larger pool of operating experience and more published information is readily available. A summary of them in shown in Figure 2.12. Most of the studies are based on four databases. RELIAWIND is the most recent one and it includes SCADA data, work orders, alarms and service records from ∼350 turbines [70]. Windstats and LWK databases include ∼6000 turbines at 11 years of operation [63, 55]. Although these databases have smaller capacity turbines,

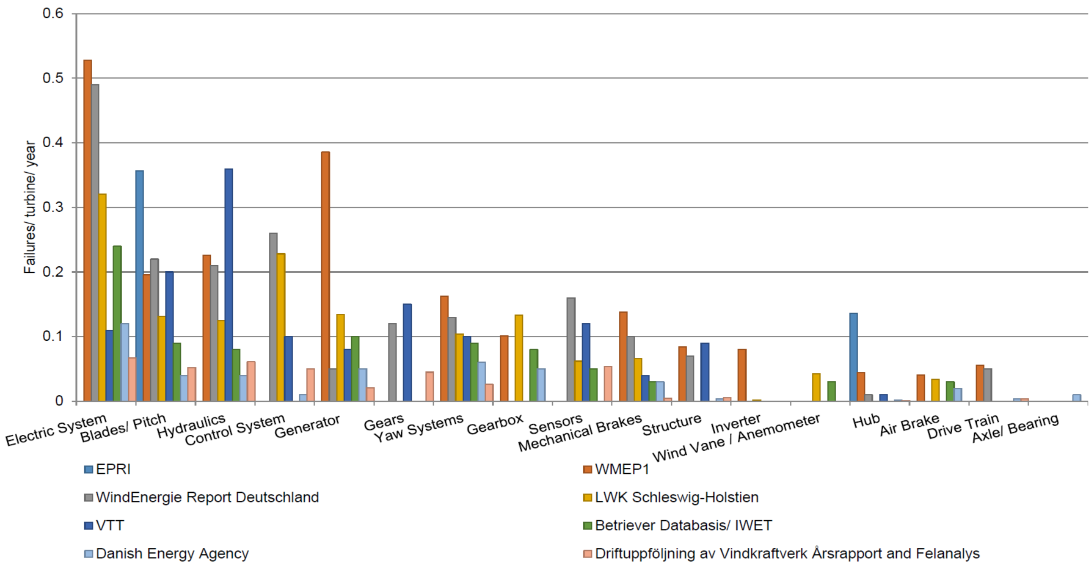


Figure 2.12: Onshore wind turbine failure rates from different databases [76].

most of them under 1 MW, they provide interesting reliability statistics that show the differences in failure rates for same capacity turbines between different manufacturers. For example, a difference with a factor up to 2, 3, and 4 has been shown for the blade, pitch and generator subassemblies respectively. Similar results can be seen in [77], where failure rates are grouped with the wind turbine's capacity and large differences between different capacities are noticed.

2.2.6 Reliability Assessment

In order to better understand the root cause of the failures in wind turbines, an attempt to correlate them with environmental conditions is investigated. Previous research has explored the failure rates of different offshore and onshore wind turbine subassemblies seeking to establish the correlation between failure rates and wind speeds at the turbine system level. One study has explored the offshore wind level failure rates [73] and one study has investigated the effect for onshore sites [12]. To the author's knowledge there are no publications that quantify the impact of offshore environmental conditions at a turbine component level, mainly due to the limited amount of available data. The effect of environmental conditions on wind turbine failures for onshore wind farms has been better understood and studied in previous research. One study has investigated the influence of wind speed for a large scale data set, using aggregated wind speed and failure rate data and tried to make this correlation [78]. Another study investigated three onshore wind farms over a year and attempted to correlate the wind speed and temperature against the number of failures observed by grouping all the turbine components into either mechanical or electrical [79]. An additional study investigated the effect of environmental conditions on converter systems by using output power, frequency and voltage inputs [80]. A more recent study has built a machine learning framework using failure data sets for onshore wind farms, along with weather and SCADA data to predict short-term failures for gearbox, generator, pitch system, yaw system and frequency converter [81].

Offshore wind failure rates are very important for modelling and estimating operation and maintenance costs. Previous research [82] has emphasized the significance and lack of failure statistics and the uncertainty around using fixed failure rate values. The study used old onshore wind failure rates, and hypothetical variations, which highlights the need for data-driven analysis. Additionally, the research need for operational analysis has been underlined at the 2016 Wind Energy Research Workshop, as a mean to further improve the development of the wind industry [83].

The existing literature focuses on data stemming solely from work orders, indicating the time that a replacement or a repair occurred. This data constraint makes it difficult to identify the exact time that the actual failure has happened. Turbine alarms can be a good indicator to detect the exact time when a failure has occurred and to provide valuable information that subsequently support the identification of potential failure locations and root causes [84].

2.3 O&M Processes

This section presents an industry review and challenges regarding typical O&M processes, as well as a common operational offshore wind workflow.

2.3.1 Industry Review

The previous section focused on the different reliability databases available. A review of the different standards, guidelines and data collection methods is presented here.

2.3.1.1 Standards and Guidelines

An inconsistency in standards and guidelines has previously been identified by the IEA Task 33, which made the following recommendations for reliability data sharing [85]:

1. Ensure the access to all relevant data.
2. Identify the use-case and be aware of the resulting data needs.
3. Map all wind turbine components to one taxonomy/ designation system.
4. Align operating states to IEC 61400-26.
5. Train staff understanding, what data collection is helpful for.
6. Support data quality by making use of computerized means.
7. Share reliability data to achieve a broad statistical basis.
8. Develop comprehensive wind-specific standard based on existing guidelines/ standards.
9. Develop component-/ material-specific definition of faults, location, and severity.

Although this is a very important step for the industry, there is still progress that needs to be made, since this task focusses only on the collection of specific type of information, that will allow industrial data sharing, to feed future maintenance models and allow operators to benchmark their assets and maintenance strategies. A holistic approach needs to be followed in order to improve the O&M workflow, without focusing only on opportunistic data collection. Guidance will need to be provided on the data management and data architecture frameworks that should be followed by the industry and the organizations.

Table 2.3 shows an overview of standards and guidelines that have been used by different studies. Moreover, additional standards and guidelines are presented, which could be modified and used by the offshore wind industry. They are classified in seven different categories that have been determined as critical for O&M data collection. There is neither a standard nor a guideline that can satisfy all the requirements. Additionally, very few of these solutions have been combined and used for a single case. One can see the separation between data management standards and reliability focused guidelines. Only one standard deals with the severity of the failure and the majority of the standards and guidelines focusing on taxonomy, maintenance, operation and failure definitions and categorization. Figure 2.13 shows how the different databases are using the existing standards and taxonomies and highlights the lack of a standardized methodology.

2.3.1.2 Data Collection

The different data collection techniques are also presented in Figure 2.13. In terms of data collection, if the RELIAWIND and Strathclyde studies are excluded, as they refer to historical data from the OEM databases, the rest of the databases require specific information that needs to be manually imported either during the operations via a log-book template or via a "fill-in the blanks" online or hand-filled template. Herewith, if an operator would like to contribute to more than one database, part of the manual labour work of data entry will be duplicated. At the same time, the results may be variable, since the results will be based on the inputs, which could include human errors.

2.3.2 O&M Process Challenges

Building an effective, efficient and robust data management system is a challenge faced by all industries. Evidence can be recognized in the examples shown below:

- Organizations waste 15-18% of their budgets dealing with data inaccuracies [98].
- The US economy loses $3.1 trillion a year because of poor data quality [98].
- The UK in 2013 could not deliver 1.4 million orders due to poor address data [98].

Table 2.3: Standards and Guidelines in terms of the areas that they cover for Severity, Taxonomy, Operation, Failure, Data Management (DM) and Data Integration (DI).

Name	Severity	Taxonomy	Maintenance	Operation	Failure	DM	DI
ISO 31000 [25]	X						
IEC 61400-25 [86]		X					
RDS-PP [87]		X					
ISO 14224 [88]			X		X		
ZEUS [89]		X	X	X	X		
Tavner [12]			X		X		
RELIAWIND [70]		X		X	X		
CREW [71]		X		X	X		
NERC GADS [90]		X		X			
IEC 61400-26 [91]				X			
DAMA [92]						X	
UNESCO IOC 73 [93]						X	
ISO/IEC 27001 [94]						X	
IEC 641850-7-3 [95]		X					X
ISO/IEC 10032 [96]						X	
ISO 15926-4 [97]							X
Ngyen [47, 48]						X	X

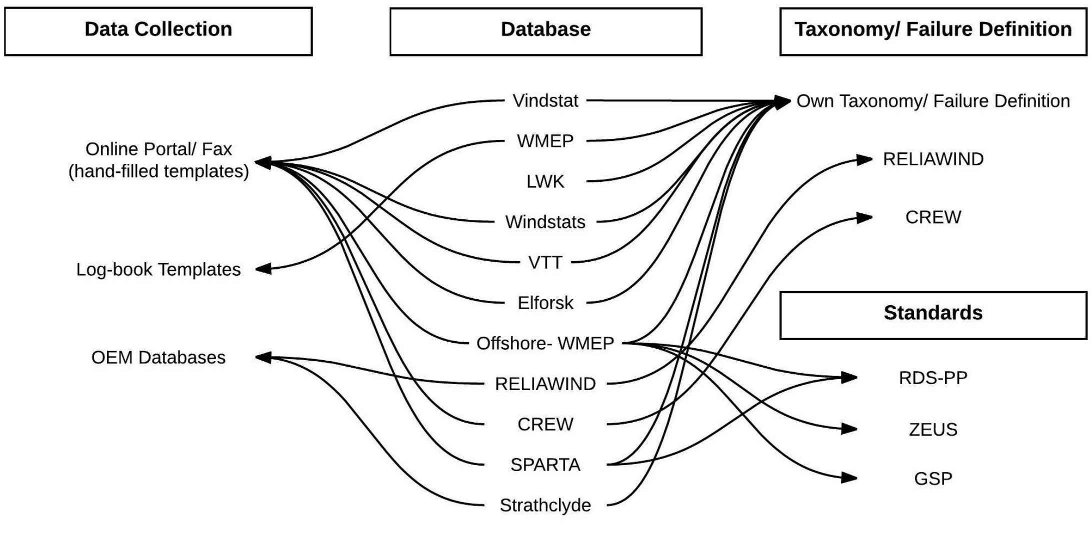


Figure 2.13: Database data collection and standards or guidelines followed.

Building a trustworthy data management system could help avoid the above issues. A study by KPMG [99] has shown that the majority of surveyed organizations do not trust their data and analytics for business operations insights. At the same time, an earlier study [100] has revealed almost all of the respondent executives believe that data and analytics would be at least somewhat important to their current growth strategies.

A number of challenges for the offshore wind industry and the individual organizations are summarized in the following subsections.

2.3.2.1 Industrial Challenges

Some key challenges for the offshore wind industry are outlined below:

- No uniform terminology for failure definition and component classification is being used.
- There is a lack of data sharing and scepticism. As it was identified from the industry review, the companies that are participating in one of these projects are only willing to share very little information, while looking for the highest achievable benefit.
- No data management standards and guidelines are present. Table 2.3 summarizes several guidelines and standards that have been used or could be modified for the offshore wind industry. It should be emphasized that there is no data management standard at present for the offshore wind industry.

2.3.2.2 Organizational Challenges

Some key challenges for the individual organizations are outlined below:

- Data anarchy, which is usually defined as a system where no one is responsible for overseeing data as a cross-functional business asset, there is a lack of data policy and a lack of processes in place for using the data as the basis for decision making.
- Insufficient & inconsistent reporting of daily operational actions, which results in data gaps and poorly recorded maintenance information that could obstruct fault diagnostic activities.
- Uncertainty on how to deal with inhomogeneous data, due to the lack of a data management system for the offshore wind industry.
- Absence of data centricity, which results in work duplication and errors in the processed data.

- Inflexible and unadaptable data workflows, which do not allow the addition of new monitoring equipment or new inspections techniques in the every day operations.

2.3.3 Operational Offshore Wind Workflow

A typical offshore wind farm workflow is illustrated in Figure 2.14, as it was interpreted by observing the operations and speaking to the staff at Teesside offshore wind farm. Most of the offshore wind farms are following a similar model, where the operator and the OEM or maintenance service provider are splitting the operational actions. The operator is responsible for the High Voltage (HV) side and the maintenance provider for the Low Voltage (LV) side. The HV side includes everything that is below the door level

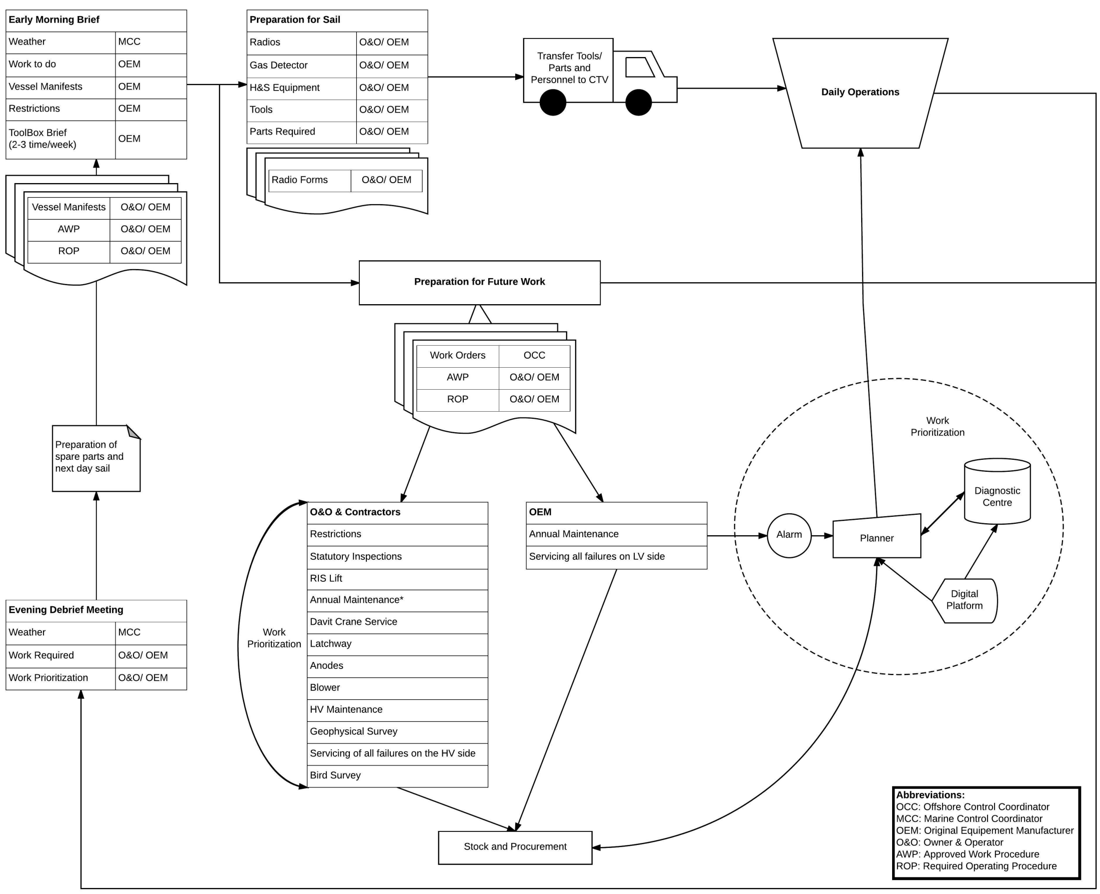


Figure 2.14: Typical offshore wind farm workflow, showing the operations performed during a typical day from the wind farm operator and the service provider.

of the turbine tower and usually includes all the environmental and geophysical surveys that are required. The LV side includes everything that is on top of the door level of the turbine tower. The operational day usually starts with a brief to the technicians, then the tools are loaded to the vessel and the operations begin. During the operations, the technicians are in contact with the operational team and the marine coordinator in order to report any issues. If any issues arise, the maintenance provider will try to tackle them and order any parts if needed. The day usually ends with a debrief meeting, in which the next day is planned. Throughout the operations, there is paperwork involved when assigning work procedures, and a lot of manual work prioritization, based on people's experience. By accurately and detailed logging and digitalizing of all the maintenance information, the physical paper work can be reduced and tools can be built to prioritize actions and automate maintenance scheduling. This is a simple model that involves a lot of paperwork, which makes it inflexible and encourages work duplication, due to the early stage of the majority of the projects, there are not any major issues on the operations. The processes can seem effective, as the work is performed on time and all major failures have been tackled on time, but their efficiency can be improved. With the introduction of new processes and techniques, this workflow could be streamlined, automated, flexible and adaptable to any future additions or changes.

2.4 Data Analytics

Another way to achieve cost reduction is by analysing the operational data more effectively. This can be addressed with the implementation of different analytics techniques. Data analytics is the science of examining raw unprocessed data in order to draw conclusions from it. Data analytics could be either descriptive, diagnostic, predictive or prescriptive. In the offshore wind context, this could mean the following:

- Descriptive – for example visualising the offshore wind turbine's condition from monitoring systems.
- Diagnostics – for example explaining the root cause of a fault or an alarm, reducing the time a technician needs to spend on the turbine in order to investigate it.
- Predictive – for example estimating the future behaviour of a sensor reading or forecasting the failure of a component well in advance.
- Prescriptive – for example providing information to determine the repair process to fix a fault.

The different data analytics techniques are further described in the following subsections.

2.4.1 Descriptive

Descriptive analytics are the most commonly used type of analytics, as they visually describe the generated data from the asset. These could be the visualization of the SCADA or CMS systems or some basic reliability statistics, including the ones presented in Section 2.2.5. Despite their simple nature, they can give a lot of information about current and past state of the machine and visualize any trends. This eases the comparison and classification of different turbine states.

2.4.2 Diagnostics

Diagnostic analytics are trying to investigate the reason why a fault has occurred. This is usually done by implementing a Root Cause Analysis (RCA). RCA is a process designed for use in investigating and categorizing the root causes of events with safety, health, environmental, quality, reliability and production impacts [101].

The importance of RCA can be seen from different industry stakeholders' perspectives:

- From the OEM side, it can improve its products and reduce the associated costs.
- Owners and operators can obtain a better and independent assessment of the product, discover mitigation techniques and assess the performance of their assets.
- From an engineering perspective, the design of the components could be improved as the failures and their causes are becoming known and will contribute in more robust designs.

The necessity of data integration and especially SCADA and CMS systems for an accurate RCA has been emphasized by [12], in order to provide good potential for fault detection.

2.4.3 Predictive

Predictive analytics bring together advanced analytics capabilities spanning ad-hoc statistical analysis, predictive modeling, data mining, text analytics, optimization, real-time scoring and machine learning. These tools help organizations discover patterns in data and go beyond knowing what has happened to anticipating what is likely to happen next [102]. Predictive analytics is driven by predictive modelling. It can be understood to be more of an approach rather than a process. Predictive analytics and machine learning go hand-in-hand, as predictive models typically include a machine learning algorithm. These models can be trained over time to respond to new data or values, delivering the results the business needs. Predictive modelling largely overlaps with the field of machine learning [103].

2.4.3.1 Machine Learning

Machine learning (ML) is defined as a set of methods that can automatically detect patterns in data and then use uncovered patterns to predict future data or to perform other types of decision making under uncertainty [104]. ML is usually split between supervised and unsupervised learning. In supervised learning, the goal is to learn a mapping from inputs x to outputs y, given a labelled set of input-output pairs, also known as training data set. Supervised learning can also be split in classification, where discrete responses are predicted, and regression, where continuous responses can be predicted. Unsupervised learning only has inputs in the form of a training data set, and the goal is to find patterns in the data, most commonly by clustering techniques.

Several ML models have been developed and implemented for wind energy applications. Primarily these techniques are applied to weather forecasting and wind turbine power outputs, but also for predicting and diagnosing wind turbine failures. This thesis focuses on the wind turbine prediction and diagnosis aspects of ML. It considers subassemblies that are critical, i.e. contribute significantly to failure rates, cost of repair and downtime. For offshore wind turbines, the most critical components have been identified to be the gearbox, generator, pitch/yaw systems and blades [73]. These are the ones with the most frequent failures and the highest downtime.

The most commonly used ML algorithms in the wind turbine literature include:

- Support Vector Machines (SVM); they use kernels, which calculate the distance between two observations. Kernels can be linear and non-linear functions. The algorithm finds a decision boundary that maximizes the distance between the closest members of separate classes.
- k-Nearest Neighbours (kNN); they are a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output could be either a class member or the property value for the object.
- Regression; it is the simplest form of algorithm, attempting to fit a straight hyperplane to the dataset.
- Decision Trees; they learn a hierarchical trend by splitting the dataset into separable branches that gradually understand and learn non-linear relationships.
- Artificial Neural Networks (ANN); are based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. They use hidden layers between inputs and outputs in order to model complex relationships between the data.
- Gaussian mixture; is a probabilistic model that assumes all the data points

are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.

- k-means; is a general purpose algorithm that makes clusters based on geometric distances between points. The clusters are grouped around centroids, causing them to be globular and have similar sizes.

A representative sample of the different ML algorithms used for different subassemblies is shown in Table 2.4. For wind turbine gearboxes, the majority of the literature has used ML for fault diagnosis, prediction and misalignment, for generators, pitch and yaw systems literature has focused on fault identification and finally for blades the focus has been on damage, fatigue and icing detection.

Table 2.4: Machine learning literature overview.

ML Model	Gearbox	Generator	Pitch/Yaw	Blades
SVM	[105, 106]	[107]	[108]	[109]
kNN	[110]			[111]
Regression	[112]	[113]		[109]
Dec. Trees	[114]			[115]
Neural Nets	[116, 117]	[118]	[108]	[119]
Gaussian	[120]			
k-means	[121]			[122]

The majority of papers in the literature concentrate on gearboxes and blades. Gearboxes are very well instrumented, which makes it easier to diagnose and predict faults. Blade damage has also been a key focus of research, due to their high impact on annual energy production and their high contribution to turbine downtime [73].

It is impossible to find an algorithm that can accurately solve all the problems, as the choice of algorithm depends on many factors, such as the size and structure of the data. As a result, different algorithms are tested in order to evaluate their performance before their implementation using a test dataset. The main advantages and disadvanteages of these algorithms are summarized in Table 2.5. In general, the algorithms with the highest overall accuracy and more complex decision making process will require the largest number of data and training time. On the contrary the ones with the more linear decision making will require less data and less training time. The size of the required dataset is very hard to estimate and it depends on the application and the data classes. In order to evaluate this in practice learning curves can be used, using resampling methods on small datasets and by adding confidence intervals to the final results.

Table 2.5: Advantages and disadvantages of machine learning models.

ML Model	Advantages	Disadvantages
SVM	Modelling of non-linear decision boundaries with many kernels to choose from. They are robust against overfitting.	Memory intensive, hard to tune as the right kernel needs to be selected, and don't scale well to larger datasets.
kNN	Good with noisy training data and large datasets	Memory-intensive, perform poorly for highly-dimensional data, require meaningful distance function to calculate similarity.
Regression	Quick and easy to understand and implement and can be easily updated	Bad performance on non-linear relationships and complex patterns. Polynomial fits can be added, making it time consuming.
Decision Trees	Can learn non-linear relationships and are fairly robust to outliers. Ensembles perform very well in practice.	Unconstrained trees are prone to overfitting as they keep branching until they memorize the training data, which can be avoided by using ensembles.
Neural Networks	Adaptable architecture and reduced need for feature engineering due to the hidden layers.	Not suitable as generic purpose algorithms as they require large amounts of data, which makes them computationally intensive to train and they require domain expertise to tune.
Gaussian	Fast and agnostic.	Large datasets are needed.
k-means	Fast, simple, and flexible.	Number of clusters is user defined, which is not always easy.

2.4.3.2 Machine Learning Model Implementation Challenges

The objective for the ML models is to be used and tested in the maintenance application, but very often there is a gap between theory and engineering application, which has been attributed to the following reasons:

- ML models are potentially difficult to understand, to interpret and to build. Many algorithms have complex statistical and mathematical models, that are challenging to understand for the O&M engineers.
- Many papers focus on mathematical methodologies only. A lot of solutions developed have very theoretical aspects, which makes them less appealing to professionals.
- Manufacturers, software developers and operators that have applied ML models efficiently in offshore wind farms, might not be willing to disseminate work in the public domain due to intellectual property rights. As a result, the majority of work available in the public domain is academic, with less industrial input.
- ML models are usually very bespoke, referring to particular failure modes and components. Especially supervised learning algorithms are very case specific and can some times only identify a particular failure mode and failure location. It is thus difficult to choose a generic and flexible algorithm.
- ML models might not always be necessary, as they might be overcomplicating the analysis. In certain occasions simple visualization, trending or reliability analysis might be sufficient to investigate the condition of the asset/ component.
- ML models usually focus on components and failure modes that have been better understood over the years and in systems that are very well instrumented. This means that systems that are not that well monitored are usually less understood and less investigated by the literature.

The above challenges for the implementation of the ML models in the offshore wind workflow have not been addressed yet by existing research in the public domain. The BS ISO 17359:2018 standard on condition monitoring and diagnostics of machines [123] has addressed the key steps that need to be followed when setting up condition monitoring equipment and how the information received can be used during the maintenance activities. Although these guidelines are a good starting point for industry-wide use, they do not address the implementation of diagnostic and predictive models in the workflow and some of the suggestions are very generic and not applicable to offshore wind.

2.4.4 Prescriptive

This is the final phase of data analytics, which is looking at what actions need to be taken in order to prevent an upcoming failure. It is the step that is not only seeking the correct implementation of the machine learning model described in Section 2.4.3.2, but also explores the fault mechanism and repair, by determining what tools/expertise is needed.

2.4.5 Review of Commercial Data Management and Analytics Tools

A market study was carried out in order to identify the best commercial solutions that could provide data management and analytics capabilities, as well as in order to identify any gaps in the market. A series of meetings, demos and trials has been conducted with 12 shortlisted companies.

Following are a number of shortlisted tools, out of the originally 90 identified. The different levels of engagement with the tool developers could include a detailed live demonstration of the software, use of the software with simulated or actual data from other sites, or direct use of data from Teesside offshore wind farm within the tool. The capabilities of the different tools are briefly described below. The tools have been anonymized for confidentiality reasons.

Tool A It is a start up that has developed expertise in the aerospace and chemical industries. They apply semi-supervised machine learning techniques, with a focus on anomaly detection. Their approach characterises the normal activity of the asset, by training the models with different parameters and tries to flag any deviation from it. This allows them to have a high level status of the asset, which will allow them to investigate the flag issue better. They have previously worked with SCADA and CMS data, but they have little knowledge of the wind industry.

Tool B It is a specialised tool, offering analysis of CMS data. It has been successfully applied to the nuclear industry and can cope with a variety of CMS data. It performs monitoring and diagnosis of failures and it provides alarms for trending signals.

Tool C It is one of the most complete wind turbine data management, analytics and planning tools in the market. They have the competitive advantage of having data from an OEM's fleet of onshore wind turbines. The tool utilizes most of the information received from the asset, including SCADA, maintenance logs, alarms, metocean data and creates custom-made key performance indicators (KPIs). This includes reliability information of components and health indication metrics, and then provides diagnostic solutions. The tool has also the capability to prevent false alarms generated by the SCADA system. Moreover, it has predictive analytics capabilities for critical component failures and is capable of flagging alarms of potential failures, before the SCADA system ones. Finally, the tool has the ability to create maintenance scheduling and prioritise daily maintenance operations, by incorporating metocean information and availability of spare parts.

Tool D It is a wind farm monitoring system that offers live wind farm and turbine statuses, as received from the SCADA sensors and alarms. Furthermore, it allows parameter plotting and turbine alarm analysis, as well as detailed component breakdown and relationships. Another feature of the tool is the management system, which helps in creating work orders and scheduling them, by using information received from weather forecasting sources.

Tool E It is a platform that allows the user to visualize in real time the performance of the assets, through the SCADA data and alarms, by showing raw sensor information and different KPIs, as well as the work orders performed on the assets. Moreover, it allows the user to add new tasks, schedule and prioritise work for the asset. In the last couple of years, the company is trying to add a predictive analytics module, with the use of artificial neural networks. The tool has also recently used artificial neural networks for predicting failures using data from the SCADA system.

Tool F It is a complete data management and analytics platform. It offers real time fleet monitoring of SCADA and CMS data in an integrated platform, by showcasing different KPIs and turbine alarms. It uses a model based method in order to provide early warnings of component failures. A physical model is running in parallel with the actual component in order to estimate a non-measured quantity, based on real-time data. The model is based on a Kalman filter to establish a virtual sensor for bearing friction. The behaviour of the model is compared with measurements to reveal degradations. The deviation between the actual measurement and the twin signal is tracked. This method has the advantage that the model does not require any information about the physical component to be monitored. The tool also uses deviation tracking and artificial neural network techniques to predict failure of other components.

Tool G It is a tool providing maintenance analytics, mainly from SCADA and metocean data. By enhanced power curve analysis, the tool is able to provide diagnostic solutions for a faulty or underperforming turbine. Their knowledge as an OEM of the sensor information, as well as all the relationships, when data are missing or data quality is poor, allows them to identify key power loss instances and root causes of wind turbine errors, with a limited amount of SCADA data. Their previous case studies have shown detection of poor data quality and underperformance of wind turbines, caused by yaw, pitch, torque and due to wake losses.

Tool H It offers condition monitoring solutions, specialising in issues with the main bearing, the gearbox and the generator. Focusing on the analysis of vibration data from the CMSs. The software can connect directly to current CMS and SCADA servers and provide insights and health monitoring of components. Their experience with machine vibrations allows them to understand the tonality of the different gears and bearings, along with their sidebands and related harmonics.

Tool I It is a platform developed by an OEM that allows the development of different modules for predictive analytics that is used in a variety of industries, including gas and wind turbines. It includes modules for anomaly detection, data exploration and pre-processing, text analysis, text mining, time series analysis, feature engineering, machine learning, optimization methods, predictive models, quality control, signal processing and statistical methods. This platform allows the analysis of different datasets from the user. At the same time, the company is building a bespoke model for wind turbines, by incorporating all the above features. Moreover, it is trying to include data from the field, through their digital worker platform, as well as create digital twins for critical subassemblies, such as the yaw system.

Tool J This start-up has developed a tool that has been used in different industries, including manufacturing and telecommunications. It has the ability to perform prescriptive analytics, by using a wide range of timeseries and textual information, by utilizing application specific supervised and unsupervised machine learning techniques. The company has also acquired a large amount of wind data, ranging from 6 months to 3 years from 1000 wind turbines, from a developer in Scotland, where they have used it to train their predictive analytics platforms. Through their case study, they managed to predict failures more than a week in advance, with an 80% accuracy.

Tool K This tool has been developed by a wind farm operator. It uses realtime SCADA and environmental information for live asset performance monitoring of onshore wind turbines. It is mainly an asset management tool that allows the user to view and compare in real time a large number of operational data received by different wind farms. It uses a standardised taxonomy in order to be able to compare similar components of different turbines in the tool. They are also trying to integrate vibration, oil particle counter, oil condition and prognostic health monitoring data in future developments of the tool.

Tool L This SME has developed a solution where it incorporates readings from vibration, temperature, acoustic and current sensor and along with information from maintenance actions and alarms, it is able to define the status of the equipment. This solution has been successfully tested mainly in the manufacturing industry and solar energy industries. At the same time, by applying different methods, such as anomaly detection, unsupervised and supervised learning, it can detect abnormal operation of the asset, which is then flagged to the user who is able to provide feedback. The feedback informs the tool whether the inspection has identified the failure, after a visit has been made to the faulty component. This allows them to constantly improve their service and improve their machine learning algorithms used.

2.4.5.1 Criteria

Table 2.6 shows the different criteria in which the tools have been shortlisted. These include the following tools' capabilities:

i. Handling wind turbine data and information.
ii. Managing and automating big data processes.
iii. Providing predictive analytics solutions for different subassemblies.
iv. Having an easy to use and navigate visualization.
v. Being flexible, allowing to build custom made KPIs.
vi. Having diagnostic capabilities, either manual or automated ones, allowing RCA of turbine failures and alarms.
vii. Providing an integrated solution, being able to handle different data types and streams at once and make correlation between them.
viii. Providing an integrated platform for all the wind turbine data sources.

Table 2.6: Evaluation criteria for different data management and analytics tools.

	i.	ii.	iii.	iv.	v.	vi.	vii.	viii.
Tool A		X	X			X		
Tool B		X	X	X	X	X		
Tool C	X	X	X	X	X	X	X	
Tool D	X	X		X	X			
Tool E	X	X	X	X	X		X	
Tool F	X	X	X	X	X	X	X	
Tool G	X	X	X	X		X		
Tool H	X	X			X		X	
Tool I	X	X	X	X		X	X	
Tool J			X		X	X		
Tool K	X	X		X	X	X	X	
Tool L		X	X	X	X	X		

2.4.5.2 Ranking

As it can be seen from Table 2.6, none of the software tools assessed is able to fullfill criteria (viii). Meaning that none of the tools in the market is able to provide a holistic Internet of Things (IoT) platform that can incorporate all the data sources generated from the wind turbine. The majority of the tools would focus either on SCADA or CMS outputs, some of them are also treating both but without necessarily making a link between them, and not including any environmental data or work order information. This means that valuable information about the asset might be missing, such as the reliability and maintenance history of the assets, or any other type of historical information that could help in interpreting the results better.

There are two tools in the market that fullfill 6 out of the 7 criteria set. These are tools C and F. Tool C has also the advantage of having access to a large number of wind turbine data, which have been used for training purposes of the ML models.

The majority of the tools that are applying ML are focusing on neural network techniques. This has a significant limitation for new market entrants in the field as a very large number of training data is needed in order to create a robust model. On the contrary, the big advantage of these techniques is that they have numerous applications and once the company builds their expertise it can be transferred to other systems.

2.5 Research Gaps and Opportunities

This state-of-the-art review presented an academic and industrial perspective on offshore wind O&M and data-driven decision making. Several research gaps have been identified throughout the review and are summarized below.

2.5.1 Research Gaps

This thesis seeks to contribute to the following five research gaps as they have been identified from the previous sections:

Process improvement Development of improved methodologies and tools for automated data collection processes.

Data management Development of a data management framework for integrating textual and sensor data.

Reliability Thorough reliability assessment by understanding wind turbine failure discrepancies within the farm and by quantifying the effect of environmental conditions of turbine components.

Data analytics Combining SCADA and CMS data for failure forecasting and presenting a holistic framework for ML model maintenance implementation.

O&M Present applications of risk-based operations in offshore wind farms, generate long-term reliability computations for O&M planning and provide representative maintenance planning models and realistic data-driven cases.

2.5.2 Cost Reduction Opportunities

This thesis is aiming to bring together the different O&M components outlined in Figure 2.2 in order to propose innovative ways for data-driven decision making. Also, by incorporating the research gap identified in the literature review, innovative tools and methodologies can be developed that will help reduce the offshore wind farm associated costs and increase the turbine's reliability. In more detail:

- Through process improvement and automation, a lot of the processes previously performed manually will be carried out instantaneously and procedures will become more efficient, which will significantly reduce the man hours.
- With improved data management processes, the data can be better aggregated and integrated, in order to create a more robust and holistic analysis. This can save time and create a better understanding of the assets by generating more meaningful insights, providing more accurate information and helping identifying the root cause of the problem rapidly.
- Enhanced reliability information and data analytics can help identifying the potential failure of the asset more quickly, which can improve the understanding of the assets resulting in more targeted maintenance activities and reduced asset downtime.
- Data-driven and risk-based operational planning can help in significantly reducing the OPEX costs of offshore wind farms, by generating more evidence-based strategies, able to test and predict more realistic scenarios.

2.6 Overarching Approach to Research

This section outlines the methods, case study results and the links between the different chapters and sections in the thesis. Moreover, it presents the wind farm that the case studies are based on.

2.6.1 Overall Methodology

The overall methodology and how the case studies are linked with the different methodology sections are shown in Figure 2.15.

The methodology is split in two chapters; Chapter 3 outlines the data management and analytics techniques followed and Chapter 4 the data utilization capabilities, connecting data to O&M decision making. The research methodology begins by presenting the ideal offshore wind farm workflow and by recommending approaches to standardize the way that data are being collected and how they are treated (Section 3.1). These are needed in order to streamline the inputs for the data management system that is then introduced (Section 3.2). Once these are provided; reliability analysis and ML models are shown for the integrated data sources (Sections 3.3 and 3.4). By combining the knowledge from the above topics, an ML O&M decision support framework (Section 4.1), a RBI methodology (Section 4.2) and an O&M strategy tool (Section 4.3) are presented.

The methodologies are then used to generate different case studies, the results of which are presented in Part III. These are grouped into two chapters, the Operational Analysis (Chapter 5), which shows results by analysing the operational data and the Maintenance Optimization (Chapter 6), presenting results that provide direct input and decision making to the operational personnel, allowing them to improve the way that operations are performed. The former includes work order analysis (Section 5.1), reliability analytics (Section 5.2), failure RCA (Section 5.3) and an ML approach for a gearbox failure (Section 5.4). The latter includes turbine alarm prediction and correlation with environmental conditions (Section 6.1), O&M planning enhanced reliability inputs from operations (Section 6.2) and risk-based inspection for the turbine's transition piece (TP) (Section 6.3).

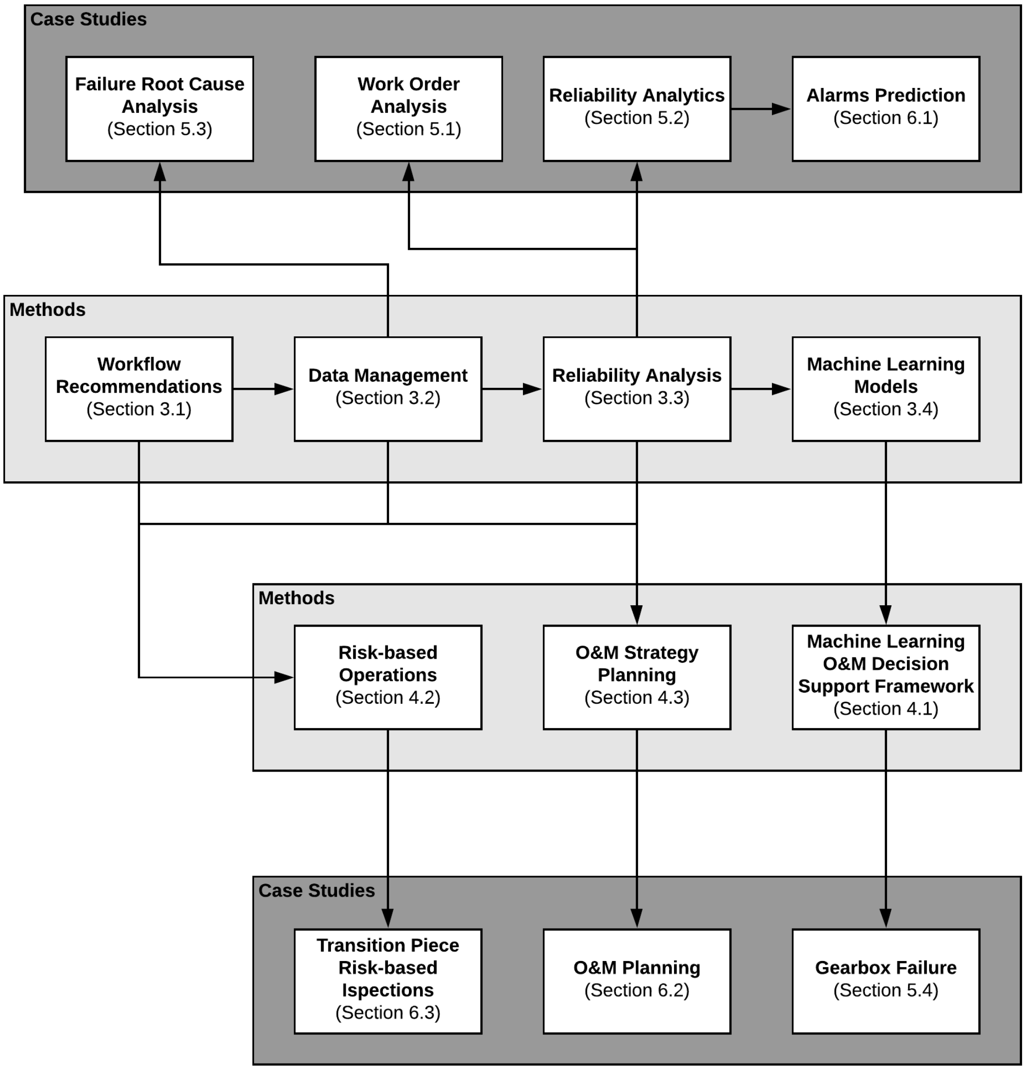


Figure 2.15: Link between methodology and case study sections.

2.6.2 Teesside Offshore Wind Farm

Teesside offshore wind farm has been used as the test site for all the methodologies and case studies presented in this thesis. It is a 27 turbine wind farm located 1.5 km north of the coast of Redcar. It comprises of Siemens 2.3 MW turbines, with a rotor diameter of 93 m. Installation of the farm began in January 2013 and was completed in June 2013. By April 2013, 17 turbines had been installed with three turbines operational and connected to the UK electricity grid. The turbines were installed in 3 rows of 9 turbines in water depths varying from 5-16 m. The layout of the site is shown in Figure 2.16. Turbines 1- 9 belong to row A, which is the furthest offshore, 10- 18 to row B in the middle and 19- 27 to row C, which is the closest to the shore. The closest port, where the daily operations take place is Teessport, is located around 7.5 km away from the closest wind turbine.

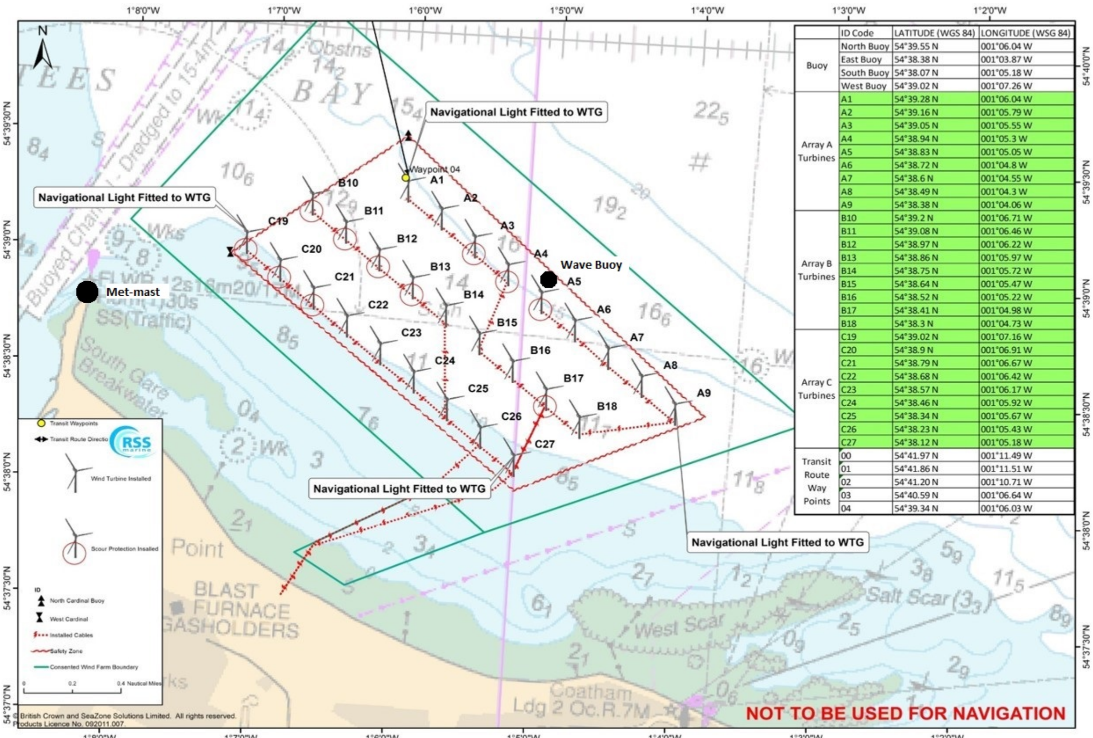


Figure 2.16: Teesside offshore wind farm layout [124].

The wind farm was developed by EDF Energy (Northern Offshore Wind) Ltd. and operated by Teesside Wind Farm Ltd., wholly owned by EDF Renewables. As of June 2018 EDF Renewables owns a 51% of the share of the wind farm. Siemens, the wind turbine OEM, provided a comprehensive maintenance service for the first 5 years of wind farm operation, while the turbines were under the warranty period. Since August 2018 service operations are being performed by EDF EN Services UK, the service arm of EDF Renouvelables. EDF Renewables UK have two maintenance contracts now; one for the windfarm (EDF ENS UK) which looks after the turbine and the foundation with the exception of major works related to the foundation such as the corrosion management and a HV contract to maintain the substation and export cables with EDS HV.

A wave buoy is located between turbines A4 and A5 and a met mast at the tip of the small peninsula east of the wind farm. They are both used to monitor the environmental conditions on site in order to decide whether an operation could take place or not. Moreover, the wind data are also used for estimating yield or performing comparisons of commercial contract offers for availability calculations.

The turbines also generate data from the following sources:

- SCADA data, producing 149 tags per turbine.
- CMS data, with 7 sensors per turbine's drivetrain.
- SHM systems on two foundations, A3 and A13.
- Work orders, maintenance & inspection reports and videos.

A dedicated CTV is always available on site to perform the daily maintenance operations. There is no nearby heliport and due to the small distance from the turbines to the port, there has not been a need for a helicopter. For heavy-lift operations jack-up vessels have been used, due to the shallow water depth.

2.6.3 Future Wind Farms

Although this thesis uses Teesside offshore wind farm as a test site, the main aim of the research project is to create transferable lessons learnt from Teesside offshore wind farm for the benefit of future offshore wind farms of the EDF Group that are under development. Thus, any recommendations, suggestions, methodologies and case studies introduces in this work are presented such that they can be adopted by wind farm developers, operators and service providers to suit their specific needs.

Part II

Methods for Data-driven O&M Decision Making

Chapter 3

Data Organization & Data Analytics

This chapter presents techniques, methodologies and recommendations developed, which aim to reduce operational costs of offshore wind farms through improved data collection, operational workflow optimization, integrated data management, reliability analysis and ML techniques.

3.1 Workflow Recommendations

Due to the rapid growth of the offshore wind industry, there is a need for standardizing data collection processes and improving the existing data management framework to enable operators to better manage offshore wind assets and processes for long-term maintenance scheduling. To achieve this, the main steps that offshore wind owners and operators could implement are shown below. These recommendations have been created as a result of the industrial placement outcomes. A more detailed explanation of the different steps is shown in the following subsections of this section.

1. Negotiate with OEMs to gain access to as many data as possible at the pre-construction phase. It is important to negotiate an optimum number of datasets and data streams, by carefully considering the available sensors at the different OEMs turbines. The required types of data and the resulting KPIs are presented in the following subsections.

2. Set-up a scalable, centralized, robust and flexible data management system that will allow better understanding of the operations in the long-term and to trace back information where needed. It is very important for the owners/ operators to have such a system in place, as it makes them independent from any OEM and will allow them to have a better understanding of the operations taking place during the warranty period from the OEM. This allows them to challenge operational decisions, as well as to have a critical judgement over the choice of the service providers and contractors, once the warranty period is over.
3. Log all the performed and planned maintenance information in detail and export relevant metadata, which will facilitate data-driven decision making, based on reliability metrics from the individual assets. This can be used either for challenging the services providers or in order to create in house tools and analysis.

An overview of the required actions to implement these steps, and the associated benefits are shown below.

3.1.1 Data Workflow Improvements

A typical offshore wind farm daily workflow was presented in the literature review, Figure 2.14. The most significant inefficiencies with this type of workflow is the data collection method, which results in unrecorded and uncorrelated information. Figure 3.1 shows the overall process proposed for the management of the data collection forms. These processes have been created by evaluating the industry state of the art, presented in Section 2.3, as well as understanding the processes followed at Teesside offshore wind farm. Initially, a work request is created in the system, which can be assigned either to an existing or to a new work order. Once this work request has been approved, then the maintenance action can be performed. Once the maintenance action is complete, a report is written and any future maintenance is planned.

In order for this process to work efficiently a system is proposed, shown in Figure 3.2. This proposed system is novel as it combines conventional computerized maintenance management systems (CMMS) with the relevant data sources generated by the different sensor systems. It shows a way to efficiently and effectively collect operational and maintenance data, highlighting the link to the different database sources. This includes

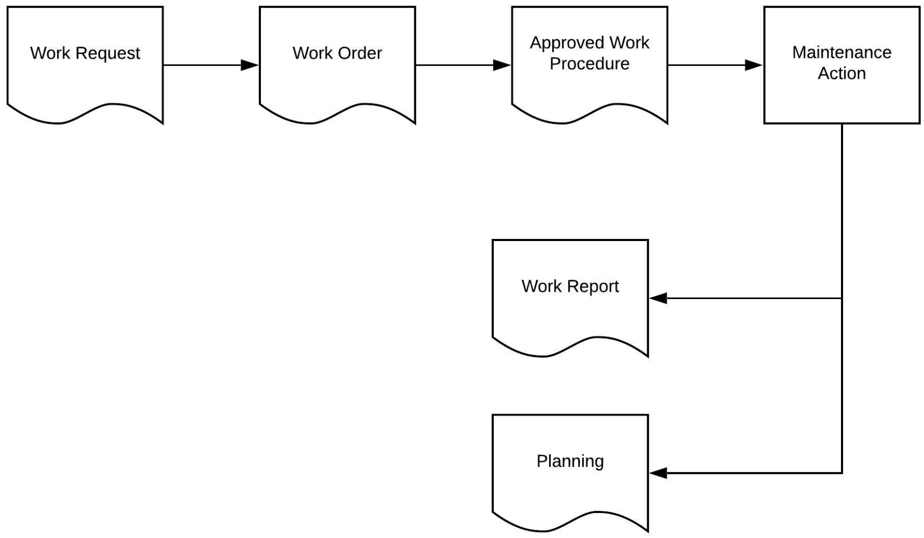


Figure 3.1: Data collection workflow overview.

fill-in the blank, dynamic and multiple choice forms for the technicians and maintenance personnel, as it allows the generation of printable reports at all stages of the process. The benefits of such a system is that it is flexible in terms of integrating different digitised data sources, which can be accessed from anywhere when needed. It also allows predefined questions and links to exist in databases leading to a more accurate process for capturing operational and maintenance information.

The system works as follows: once the maintenance personnel have selected the relevant maintenance type carried out, the system will require the detection method (for condition-based and corrective maintenance), where it can be linked either to an alarm from the SCADA system or to a report. Then the cause, effect and removal of malfunction is required, with the following options:

- Causes include: high wind, grid failure, lightning, icing, malfunction of control system, component wear or failure, component loosening, other cause (needs to be specified), unknown cause.
- Effect of malfunction: overspeed, overload, noise, vibrations, reduced power output, causing reoccurring damages, plant stoppage, other consequence (needs to be specified).

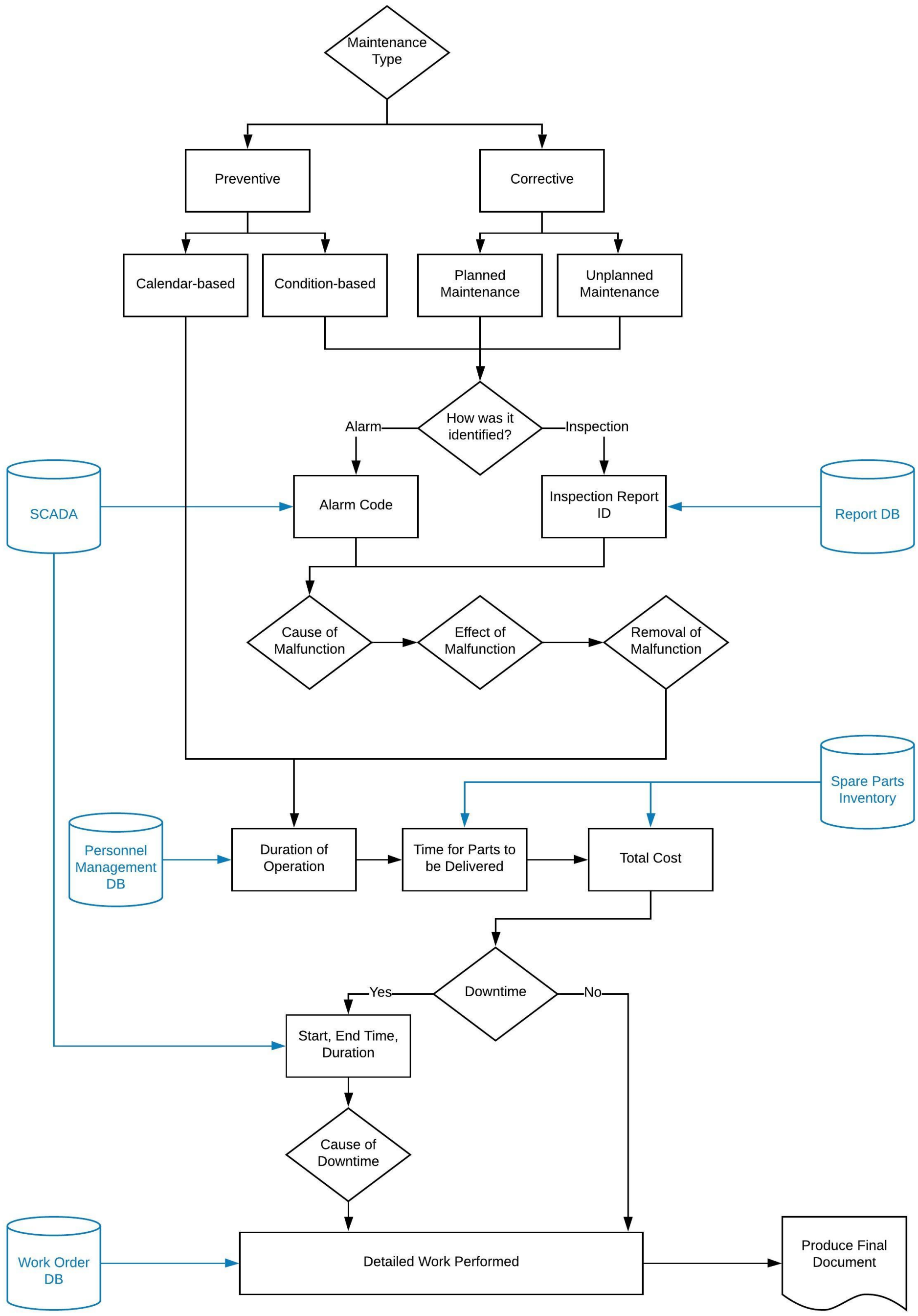


Figure 3.2: Proposed data collection workflow.

- Removal of malfunction: remote repair (control reset or change in control parameters), minor repair (select subsystem), major repair (select subsystem).

A duration of operation needs to be defined, with the repair time and duration of transfer being filled by the personnel management database. If any parts have been consumed, the details can be taken from the spare parts database and the associated costs can be calculated. Finally, if the maintenance action resulted in any downtime, it can be linked to the SCADA system for the duration and a justification and duration from technician can be added as follows: maintenance, weather conditions, parts unavailability, personnel unavailability, vessel unavailability, out of business hours. A report can then be generated by the system containing all the above information.

3.1.2 Data Collection

The following data sources section presents the key sources of data that operators should have access to and describes the associated data collection guidelines. A robust data collection system can give wind farm owners and operators a competitive advantage against any OEM, as they will have a centralized system with data coming from different systems and OEMs.

3.1.2.1 Data Sources

Following are some key data sources that the operators should negotiate access to and can provide data insights and ease decision making:

- **SCADA** system, which is standard on every wind turbine. A list of the most important SCADA tags are presented in Table 3.1. The method for prioritising the SCADA tags is explained in Section 3.2. Note that the tag names and sensors could vary with turbine make and type.
- **CMS** due to the wider bandwidth and higher sampling frequencies, it can detect faults more accurately and give more precise and early warning on the location of the fault [12], which can help to avoid full subassembly replacement. The main bearing, blade bearing, all stages of the gearbox (for geared turbines), the inner and the outer parts and the driven and non-driven ends of the generator can be instrumented with vibration sensors as part of the CMS.
- **Turbine Alarms** are important in fault detection, as they are the first sign of a degradation and can be used in conjunction with the SCADA and CMS sensor readings to determine the cause of why the alarm occurred.

- **Work Orders**, if logged in detail and related with operational work procedure forms they do not only provide failure rates, but also useful information to plan future maintenance activities.
- **Inspection reports** can provide useful information on the condition of different turbine subassemblies. They can be used to support risk-based maintenance and inspection activities, which can reduce the O&M costs.
- **Personnel Management Database** can show the duration of the transfers and performed maintenance actions, which is important for future maintenance planning.
- **Spare Parts Database** can support logistical planning and reduce ordering time.

These data sources could be used in an integrated multi-parameter monitoring platform in order to get the full benefits of an integrated and flexible database architecture during daily operational activities. It is also very important in order to make the connection between the different sensor warning and parts replacement to create a common terminology.

Table 3.1: SCADA tag data selection.

ActiveEnergyExport	Gearbox_Oil_Pressure	Main_Bear_Temp
ActivePower	Gearbox_Oil_Temp	Nacelle_Temp
Ambient_Temperature	Generator_Bearing_Temp	Nacelle_Yaw
Availability	Generator_Fan_Temp	PitchRefPosition
Blade_A_PitchAngle	Generator_Speed	PowerFactor
Blade_B_PitchAngle	Generator_Temp	ReactiveEnergyExport
Blade_C_PitchAngle	Generator_Torque	ReactivePower
Brake_Temp	Grid_Fault_Time	Rotor_Speed
Converter_Grid_Temp	Hydr_Oil_Pressure	Tower_Freq
Converter_Current	Hydr_Oil_Temp	Tower_Humidity
Converter_Volt	HSGen_Temp	Transf_Temp
Converter_Temp	HSRot_Temp	Wind_Direction
Gearbox_Bearing_Temp	IMSGen_Temp	Wind_Speed
Gearbox_Temp	IMSRot_Temp	WT_Status

3.1.2.2 Guidelines

In a scalable system, new KPIs might need to be generated or indicators might need to be improved, depending on the changing needs of the stakeholders. Figure 3.3 shows the data guidelines that can be followed when it needs to be presented or shared. These guidelines have been created by modifying the UNESCO IOC 73. It is important to initially define the end, which could be anything from as generic as reducing O&M costs or as specific as MTBF. Then the correct data sources can be selected. The data will then need to be well understood in order to use the correct vocabularies, ontologies and thesauri. Subsequently, the data will need to be filtered, extracted, metadata will need to be generated if needed and be integrated with the rest of the information. Finally, they can be archived and disseminated. This generic process can be implemented to build the whole data management framework, or just to add new data sources and KPIs in an existing system.

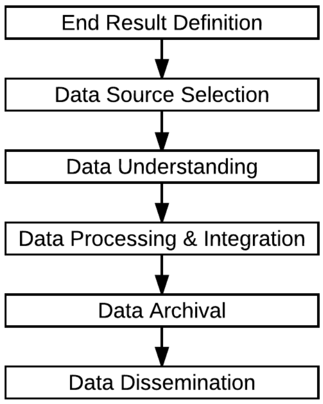


Figure 3.3: High level data management guidelines.

Building on the guidelines from IEA Task 33 [85], that were outlines in the literature review, some further recommendation can be proposed in order to automate, improve and streamline the above processes:

1. Log all information on failure root cause and detection, in order to be able to construct a diagnostic database.
2. Standardise terminology by following previously set guidelines, as presented in Table 2.2, with drop-down options in all the reporting forms.
3. Combine operational worksheets where possible, to reduce the possibilities for work duplication.
4. Digitise all operational forms for easier archival, retrieval, editing and sharing.

5. Archive the operational information in an easily extractable format, such as .csv files, in order to be able to automatically create metadata.
6. Automate the extract, transform, load (ETL) processes.
7. Automate the data extraction and metadata generation processes, to be used for generation of KPI and other analytics.

Some of the key benefits resulting from the implementation of the above steps are presented below:

- Reduce operational costs, as information will be easily accessible in an integrated way and KPIs will be easily visualised.
- Increase confidence in the data, due to standardised terminology and reduced human interaction with the data.
- Generate data centricity through an integrated platform, as explained in 4.1.3 below.
- Allow customisable data analytics for the different stakeholders.

3.1.3 Data Architecture

Once the above data collection recommendations have been implemented, the workflow can be substantially improved, streamlined and automated. The next step is to build a robust data architecture. An integrated system for the architecture is shown in Figure 3.4, that creates a data centric model with three layers of information and data access for operational, asset management and research activities.

The data can be extracted from the data sources, transformed and loaded into a data warehouse, via an automated ETL process. The transformation process includes separate manipulation of the structured data via a relational database system (RDBS) and the unstructured information via a NoSQL database for faster processing. Useful information from the unstructured data, such as actions from operational reports, can then be extracted and linked to the unstructured information. Moreover, all the unstructured data can automatically create metadata that will be stored in the RDBS and linked to their original files. After the ETL process has been completed, the data can be integrated and visualized in the data warehouse. Filtered and selected information from the data warehouse can be shared via a visualization platform with the operational team. The maintenance activities could then be fed back into the data warehouse. The additional information added from the operations can then be shared in the asset management platform. The research teams can have access to all the data to support the development of data analytics and failure prediction/ identification techniques. This data centric model allows all the different stakeholders to have access to the most up-to-date data versions and ensures the awareness of all the available data sources. The scalability of this solution is crucial, as it will allow generation of new KPIs and encourage data sharing within the organizations, while tackling the challenges presented in the state of the art review, Section 2.3, regarding the data sharing within the industry.

3.1.4 Implementation

Several of the above recommendations have been implemented using a historical data set from Teesside offshore wind farm. Some of the key achievements from the implementation are listed below:

- Implementation of a .csv data reporting structure; the information can be automatically extracted. This allows easy reporting of reliability information and other KPIs from daily operational status reports. Previously, more than 100 man-hours were needed in order to manually extract 1 year worth of specific O&M information from the operational reports, without including any post-processing time.
- The integration of all SCADA data allows the trending and visualisation of different tags for individual turbines and for the whole wind farm. Beforehand, the process involved manual downloading of the tags and the individual integration and plotting, further explained in Section 3.2.
- The integration of environmental and monitoring sources with the maintenance work orders has allowed a holistic understanding of the asset and the operations performed. This allowed the correlation between data sources and implementation of studies, such as RCA.

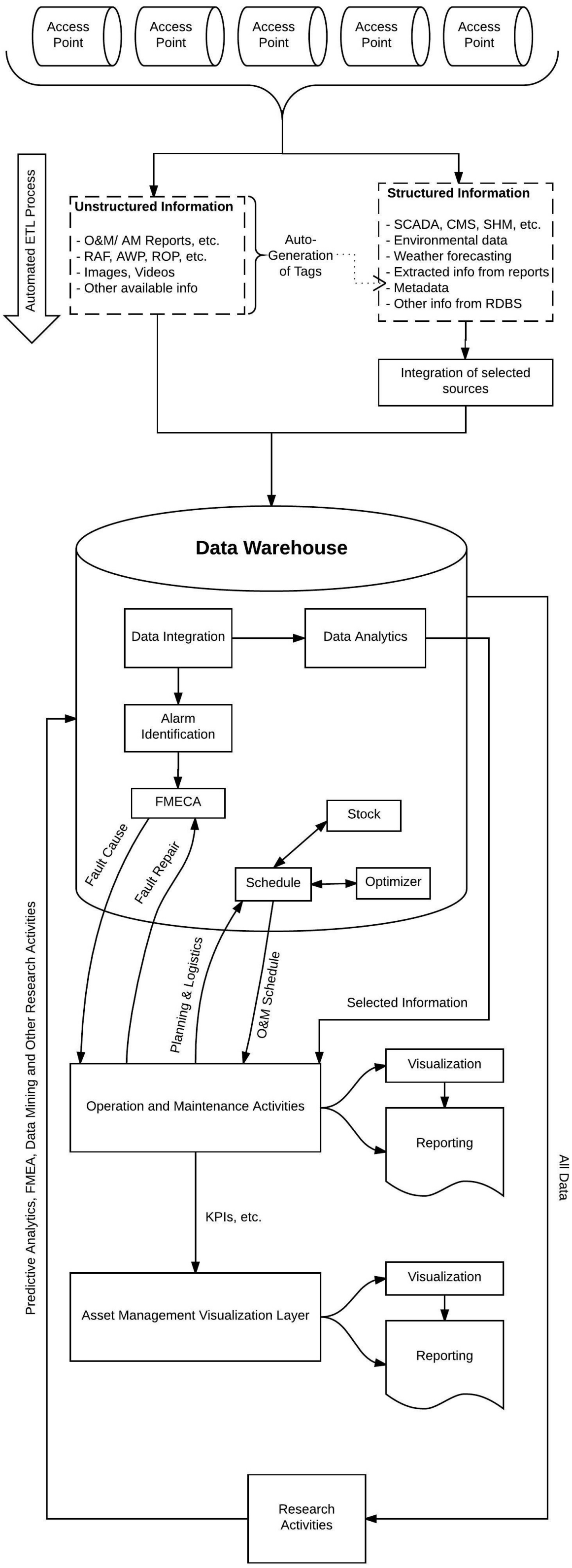


Figure 3.4: Proposed data architecture framework.

3.2 Data Management

The methodology followed in order to perform the fundamental steps for data integration are presented in this section. It includes the selection of the data sources, their filtering and the integration techniques used for the utilization of the wind farm data. Such a process is unique but this section tries to propose a new way of data management techniques for the benefit of more robust and holistic data analysis that will lead to operational cost reduction, compared to the ones identified in Section 2.2.3.

The procedure that was followed includes the integration of data from different sources, after an initial transformation process described in the subsections below. An RDB schema was formed to allow quick visualization solutions. A schematic showing an overview of the process that was followed can be seen in Figure 3.5, with more in depth analysis in the following subsections.

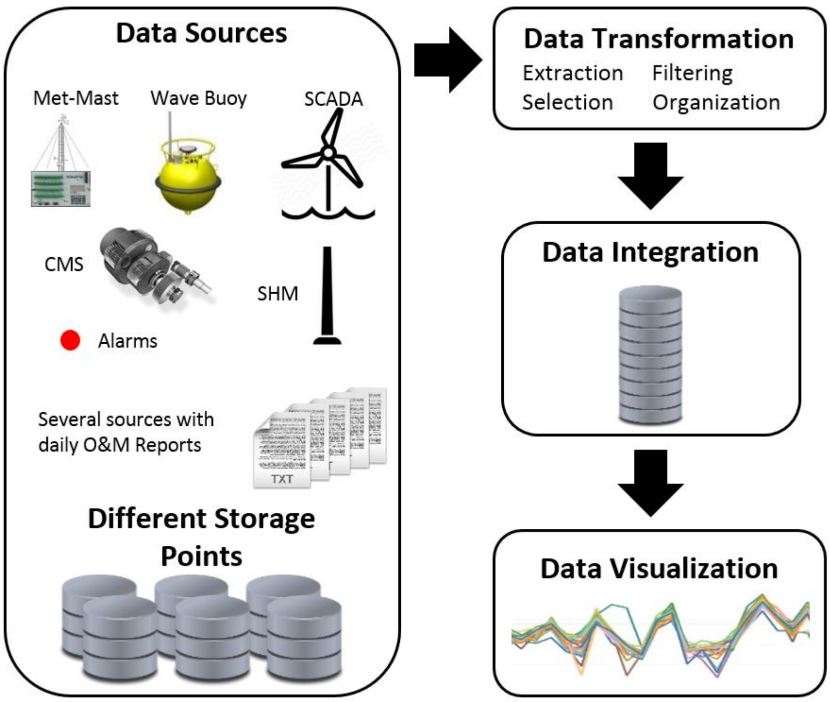


Figure 3.5: Data management flowchart.

3.2.1 Data Sources

The data sources used in this initial study include 24 months of operational data from Teesside offshore wind farm. The data were individually retrieved from different access points and they were individually saved to a local server, before any further analysis. The data sources include a met mast, a wave buoy, SCADA, CMS, SHM, alarms and all the different operational data. The operational data include the planning minutes of daily operations, the actions from the owner's side, the actions from the service provider and logs from an operations & personnel management system. The data are generated from at least 15 different sources and are being collected by a variety of different methods, including online databases, direct access to servers and e-mails. This shows the importance and the difficulty of creating a single point of access for all this inhomogeneous information.

3.2.2 Data Transformation

The next step was to transform the data and convert them into the required format, in order to be able to integrate them more easily. This process consists of four main steps:

i. Selection A decision matrix was initially created in order to select the data needed for the analysis for the SCADA and CM systems and by using the guidelines outlined in Section 3.1. The criteria for selection included availability, accessibility, failure rates/ reliability, forecasting and other key performance indicators. The operational data consist of a selection of different data sources that include the four operational sources, along with the SCADA alarms, compiling an O&M summary document. More detail on the content of the summary document and description for each component can be found in Table 3.2. The environmental sources did not need any pre-processing at this stage.

ii. Extraction This step was mainly including the O&M reports that do not include structured information. The other sources were already saved on a local server, after they had been downloaded from the different access points. The set of O&M reports and alarm data were then extracted and transformed into a structured format in order to be manipulated by the RDB, as explained above.

iii. Filtering Once all the required information had been collected and extracted, it was assessed for consistency. The met mast, wave buoy, SCADA, CMS, SHM have

Table 3.2: O&M summary selection document.

Document Content	Explanation
Date	Date of performed operation- in order to be able to group them by date.
Alarm Timestamp	Timestamp of when the alarm was triggered- if any. This helps in defining condition based maintenance or component failure.
Alarm Duration	Duration of the alarm (No need to include end time, since it can be calculated from the timestamp and the duration).
Alarm Tag	Alarm number, as it appears on the SCADA system and its explanation.
Weather	Short explanation of the weather conditions as described on the maintenance logs- to understand environmental limitations.
Location	Where the operation took place (Wind turbine number/ wave buoy/ met mast etc).
Vessel	Vessel name used for the operation.
Time of Arrival at Turbine	Time when the technicians arrived at the turbine.
Time of Departure from Turbine	Time when the technicians departed from the turbine.
Duration of Operation	(Departure Time from Turbine) - (Arrival Time at Turbine)- in order to categorize operations and calculate their average time.
Number of Technicians	Number of technicians that carried out the operation- important input for cost and O&M models.
Performed Action Title	Short name for the operation performed- to be able to categorize them easily (eg. Annual Service, Fan Replacement, etc.).
Performed Action Info	More detailed information for the operation. It could include description of the operation and more detail on the repair/ replacement.
Status	Status of the action- completed on the day or still in progress.
Operation Type	Scheduled/ Unscheduled, Condition- Based maintenance, Annual Service.
Failure	Characterization of the operation as a failure or not.
Assembly	Name of the assembly where the failed component belongs [70].
Subassembly	Name of the assembly where the failed component belongs [70].
Component	Name of the component [70].

been filtered in order to remove duplicate values stored in the individual columns and avoid any non-numerical values from sensor readings. The filtering was performed using Bash commands, as it was proven to provide faster results. Figure 3.6, shows the time that is needed to filter one year worth of data for an individual turbine. As it can be seen, there is a significant time gain when Bash is used, for all the required operations compared to Python (two popular languages for data cleaning). The files included 2 column tables with an average of ∼16,000 rows each. This translates into 26 seconds of time saving for the filtering of both CMS and SCADA data for every year worth of data. The alarms also had to be filtered in order to remove any repetition on the same turbine and to separate any remote/ local/ manual stops or upgrades of the turbine's software with any alarm that indicates failure or degradation of components. A sample of the algorithm used for filtering is shown in Algorithm 1.

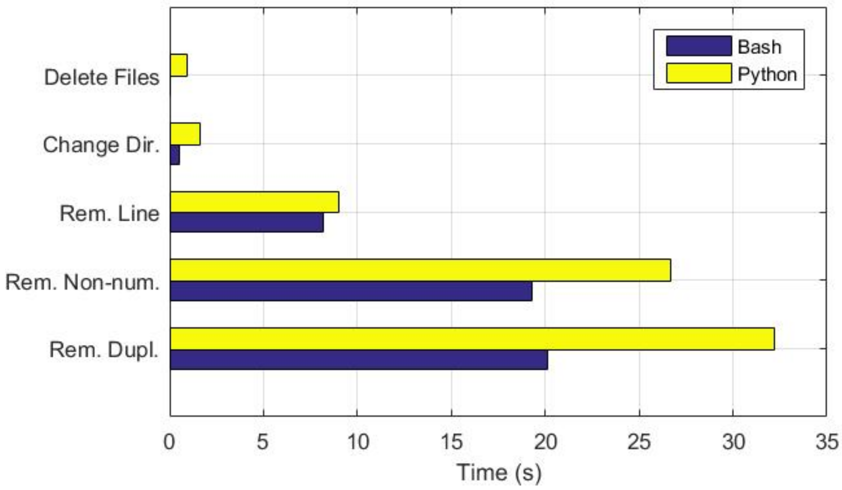


Figure 3.6: Filtering comparison of Bash and Python.

Bash command would be expected to work more efficiently when dealing with files directly, as was shown in this case, since it is shell scripting that is able to deal with the system's files directly and processes them faster than Python or any other programming language. However, a mixing of programming language would work better in this case as Bash would not be able to handle complex commands and calculations as easily as a programming language would do.

iv. Organization The data sets were then organized and prepared to be imported into the database. Furthermore, the O&M data sets were organized by adding the relevant components and subassemblies in order to compare it with literature, as shown in Table 3.2. At this stage some high-level programming was used in MATLAB, in order to quickly relate the action, with the time of arrival at the turbine, the time of departure, the duration of the operation and the number of technicians, as shown in Table 3.2.

Algorithm 1 Data cleaning

```
Define directory where all files are located
for Each individual file (v1) do
    Remove duplicate values from both columns
    Save file with new name (v2)
end for
for Each individual of the the v2 files generated do
    Remove duplicate values from the timestamp column
    Save file with new name (v3)
end for
for Each individual of the the v3 files generated do
    Remove any non-numerical values from the value column (for numerical sensor
  readings) and replace with NULL
    Store the timestamp and the non-numerical value into a new column
    Save file with new name (v4)
end for
for Each individual of the the v4 files generated do
    Remove any blank rows as they won't be read by SQL
    Save file with new name (v5)
end for
for All files do
    Delete all the v1, v2, v3 and v4 files
end for
```

3.2.3 Data Integration

Once all the data had been transformed into the required format, a relational schema was built, in order to import and integrate them into the DB. A high-level overview of the schema is shown in Figure 3.7 and the code to compile the different tables is shown in Algorithm 2. The schema has been designed so that it is flexible, scalable and adaptive to any additional sensors or data sources that will need to be added in the future, or when the DB will be updated. The initial data tables were created, using the column that contains the date and time (DateTime/ Timestamp) information as a primary key. This guarantees that all the time values inserted will be unique, if not an error will be returned. In that case, the data will need to be filtered again or checked. Using the timestamp as a primary key, saves computational time, as less data will be processed, and could save up to 20% storage space per table. Moreover, it allows an easier integration of the tables with missing or different timestamps, through their primary keys. The information from the different sensors is initially stored in separate tables that have been downloaded from the data source. They are then integrated together, by data source type, turbine number and finally combined all together. This creates two large pools of tables, the individual turbine one and the final integrated one. This approach guarantees an easier and faster data analysis for the different sub-categories on a sensor, turbine and on a farm level.

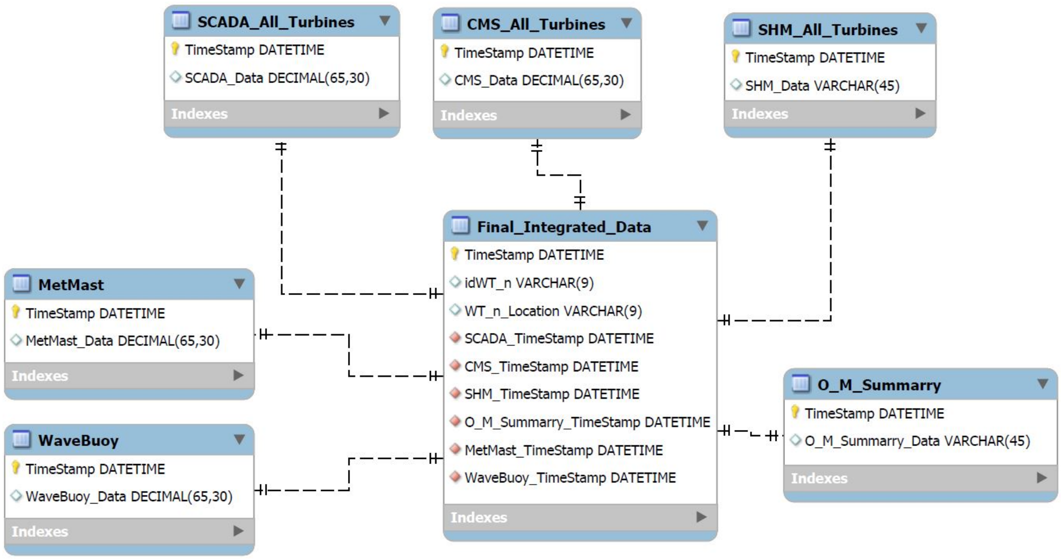


Figure 3.7: High-level data integration schema.

The integration was done using MySQL by initially combining the tables using their primary key, which allows the existence of only one primary timestamp column and not allowing any duplicates, and then by joining all the tables accordingly. *Union* and *left outer join* functions were then used. This allows a full outer join to be replicated in MySQL, with the fastest possible way, making sure that all the data will be included.

An example of the final code can be seen in Algorithm 2. Following this algorithm, all the tables including all the different data sources were added together, linked by their timestamp, as shown in Figure 3.7.

Algorithm 2 Data integration

Require: At least two tables, each with at least a timestamp column and one with the different readings.

```
CREATE TABLE "Integrated Table"
SELECT a.PrimaryKey,
        "Table 1"."Column 1"
        "Table n"."Column n",
FROM (
    SELECT PrimaryKey from "Table 1"
UNION
    SELECT PrimaryKey from "Table n"
) a
LEFT OUTER JOIN
        "Table 1" ON PrimaryKey = "Table 1".PrimaryKey
LEFT OUTER JOIN
        "Table n" ON PrimaryKey = "Table n".PrimaryKey
;
```

3.2.4 Data Quality Check

A quality check was implemented for the different O&M data sets extracted from daily logs, SCADA, CMS data and the alarms, in order to ensure that the information provided is correct and that there are no human or code errors.

The information from the O&M reports were manually checked. All the days in every month were checked, by looking at the equivalent operational logs and the information reported in the DB. This process showed that there were no human errors, during the manual extraction process.

In terms of the timeseries data, the final integrated values, that were filtered and loaded into the DB, were tested with the initially extracted ones from the different servers. The original number of values were then recorded. Using Bash script, the number of each of the non-numerical values in the SCADA data were counted, as they have been originally removed and replaced by NULL, before being imported into the DB. Similarly the same code was used for the other tags, by changing the directory and the "File" name. Then, the *ismember* function was used in MATLAB, initially for the DateTime column, in order to make sure that all the timestamps exist. Once this was proven, both columns were checked. By adding the count of true values from the latter process with the one of the non-numerical values it should add up to the number of the original values, as indicated by the original data. During the filtering process, duplicate DateTime values had been removed. In that case these values would appear as a true. An example of the script is provided Algorithm 3.

Algorithm 3 Data quality check

$O \leftarrow$ Read original file
$DB \leftarrow$ Read the database table generated
for All the rows, check the timestamp column **do**
    Check if $O(:,1)$ is a member of $DB(:,1)$
    Count instances
end for
for All the rows, check both columns **do**
    Check if $O(:,:)$ is a member of $DB(:,:)$
    Count instances
end for

3.2.5 Data Visualization

The way that the data integration was performed, allows rapid visualization using appropriate big data visualization software. In this case, Tableau was chosen as by the time that the thesis was written it was the most popular business analytics tool and it was widely used within the company. Since licenses were available and senior management was familiar with the software it was chosen in this case. There are open source alternative libraries that work with software like Python, such as Boken, Shiny or Plotly, but they would require coding experience and would not be easily used and edited by individuals at different functions within the organization. Other visualization software could be used At this stage, the data were filtered further, by removing *NULL* values, and more layers were added to group them for easier and faster analysis. An example of the visualization capabilities is shown in Figure 3.8, where the readings of the SCADA tag "Generation RPM" are shown, along with the different count of alarms and the work order information for one turbine. The x-axis shows the days in the month, the left y-axis shows the number of alarms and the right y-axis shows the revolutions per minute for the turbine's generator.

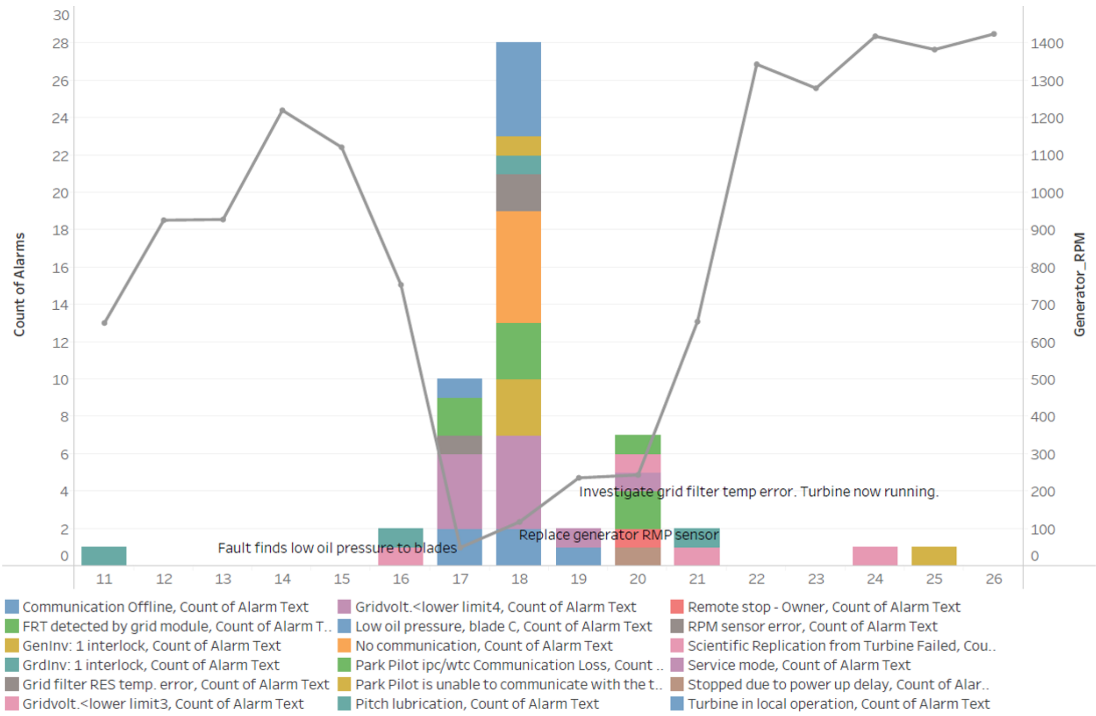


Figure 3.8: Data integration visualization example.

3.3 Reliability Analysis

3.3.1 Failure Rates

The failure rates of the wind farm were calculated by using Eq. 2.2, defined in the state-of-the-art section. Work order information was used, once categorized and grouped into taxonomies.

3.3.2 Turbine Reliability and Environmental Conditions

Another way to analyse the faults or failures that occurred in a wind turbine is by investigating the turbine alarms. If they are analysed appropriately, they can provide the exact timestamp that a turbine fault occurred, which can give a better insight about the root cause of the fault. A detailed description of the steps followed for this study is shown below.

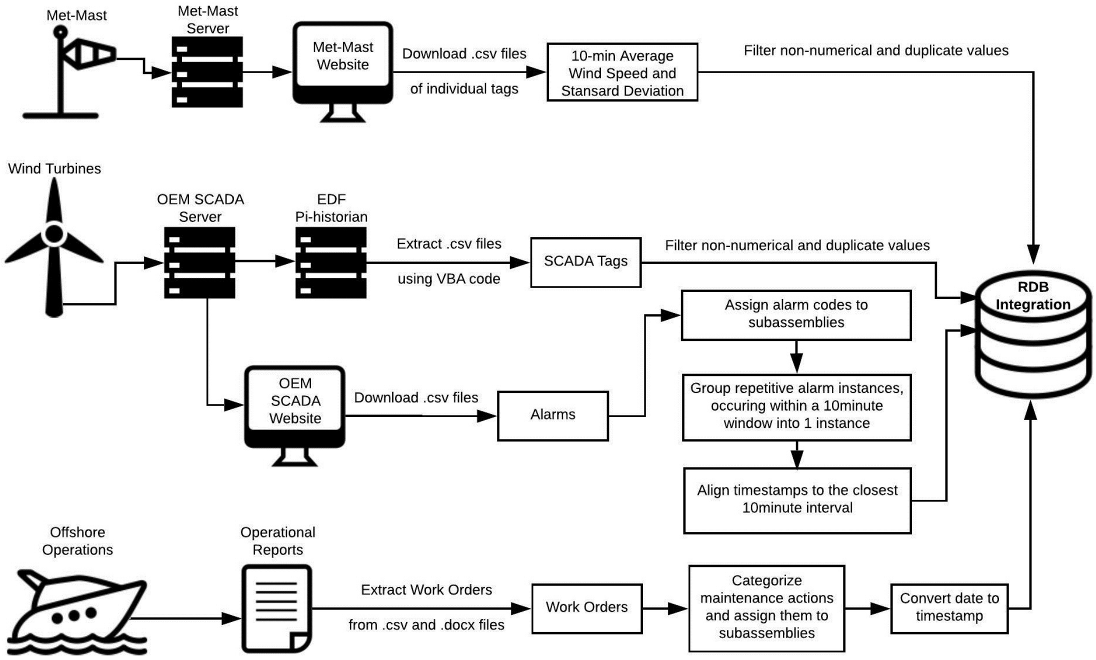
3.3.2.1 Data Integration

The framework presented in Section 3.2 was used. A flow chart of the data handling process is shown in Figure 3.9a and data analysis in Figure 3.9b, which are further explained in the following sections.

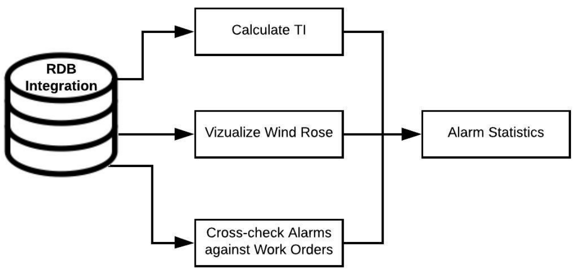
Data Sources The data sources used in this study include 4 years of operational data from Teesside. These include met mast, SCADA, alarms and maintenance operation description from the maintenance logs. The majority of results presented represent 4 years of operations, apart from the met-mast data, which only represents a period from September 2015- September 2017.

The data sources used in this study are the most common ones for offshore wind turbine operators. Other available data sources have not been included in this study for the following reasons:

- Wave buoy data are not incorporated in this study, as they only provide one point measurement and since high frequency vibration data of the foundation are not available, they would not be able to provide any useful reliability insights for this study.



(a) Data collection and preprosessing



(b) Data post-processing

Figure 3.9: Flowchart of data integration and analysis.

- CMS; vibration data for the drivetrain are available for all the turbines at Teesside. However, these data are only available post-processed, providing measurements once or twice a day and the original timeseries measurements are not available. This makes the proposed analysis impossible, as correlation would be difficult between CMS and weather 10-min average timeseries.
- Inspection reports and cost data; these data sources could not be provided due to commercial confidentiality reasons.

Data Transformation The first data handling step was to manipulate the data and convert them into the required format, preparing it for later integration. This process is formed of the following steps:

- Extraction of the information included in the maintenance logs and of the relevant SCADA and met mast information.
- Filtering of all the data. Once all the required information had been collected and extracted, it was assessed for consistency. The met mast and SCADA data have been filtered in order to remove duplicate values stored in the individual columns and avoid any non-numerical values from sensor readings. The filtering was performed using Bash commands as shown in Algorithm 1.
- Grouping of turbine alarms. For reoccurring alarms within half an hour from the previous one, the alarms were grouped together and considered as a single instance.
- Identification of critical alarms; The alarms that have led to a further investigation, either by sending personnel offshore to inspect or to repair the potential failure, or by shutting the turbine down and investigating the issue remotely are classified as critical alarms.
- Alignment of work order, alarm and SCADA sensor timestamps; The alarm timestamps were rounded to the closest 10-min due to their complex timestamps, in order to ease the integration with the SCADA data. The work orders were then assigned to the alarm timestamp indicating that the turbine is shut down for maintenance. Every time that the turbine is in local control for a maintenance activity, a work order is logged.
- Organization of the data sources. The data were then prepared to be imported into the database. Furthermore, the O&M data were organized by categorizing the actions to failures and adding the relevant components and subassemblies. The alarms had to be filtered in order to group any repetition on the same turbine and to separate any remote/ local/ manual stops or upgrades of the turbine's software with any alarm that indicates failure or degradation of components.

Data Integration The data were then integrated with the method explained in Section 3.2.3.

Turbulence Intensity, Wind rose and Temperature difference In order to calculate the turbulence intensity (TI), the data received from the SCADA system have been used and the TI has been calculated by Eq. 3.1.

$$TI = \frac{\sigma_U}{U} \tag{3.1}$$

where U is the wind speed and σ_U the standard deviation of the wind speed. In wind farm calculations, the turbulence intensity is usually estimated by the information received from the met mast. However, those measurements would not provide the TI of the individual turbines, which can be estimated by using the SCADA data. The accuracy of the above equation with offshore wind field data have been proven by [125], on the same turbine model. The wind rose was also calculated with the data received from the met mast, as in that case it was easier to have one point of reference for the wind farm, than individual wind roses for the different turbines.

The temperature difference of the ambient and the internal nacelle temperature was also calculated.

3.3.3 Forecasting Turbine Alarms

In order to further test the analysis presented above, a forecasting tool was designed, by incorporating the data shown in Figure 3.9a.

3.3.3.1 Data Post-processing

The analysis of the data was performed in two stages. Initially, the data were visualized from the database using Tableau, which helped to visually interpret the information and create some quick data aggregation. Then, the data were transferred to Matlab for statistical analysis and implementation of the alarm prediction tool. A visual representation of the process is shown in Figure 3.10.

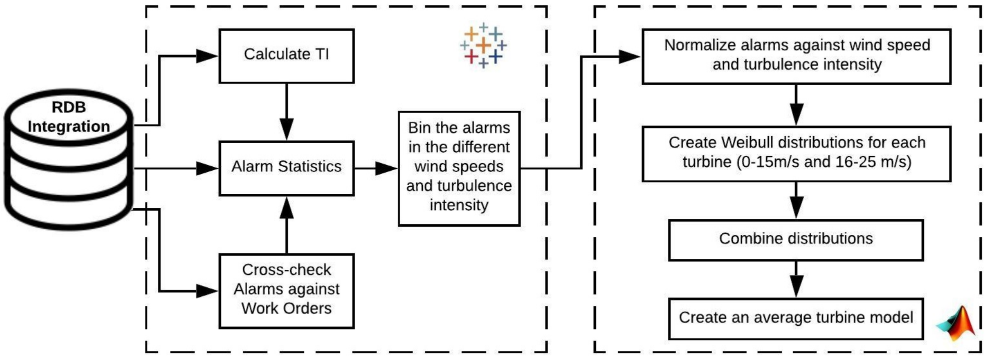


Figure 3.10: Alarm data analysis.

Data Visualization The data were initially visualized using Tableau. An initial sanity check was performed in order to make sure that the work orders and the alarms match. Once this was completed, the alarms were binned against the different wind speed and TI values for the different turbines. TI was calculated by using Eq. 3.1.

Reliability analysis The binned data were then transferred to MatLab. The data were normalized against wind speed and TI to account for the environmental conditions, i.e. the alarms were normalised regarding the given occurrence of each bin. An example of normalized and non-normalized data is shown in Figure 3.11. Two different Weibull

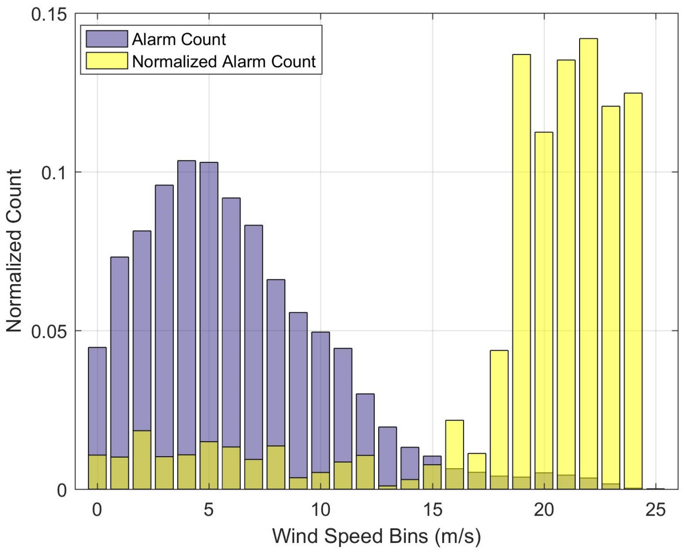
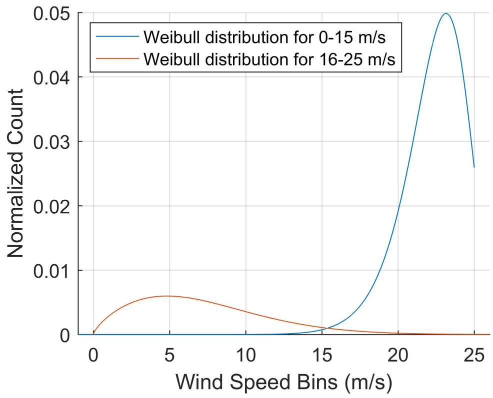


Figure 3.11: Alarm count against wind speed for normalized alarm count data against wind speed and non-normalized ones for wind turbine number 13.

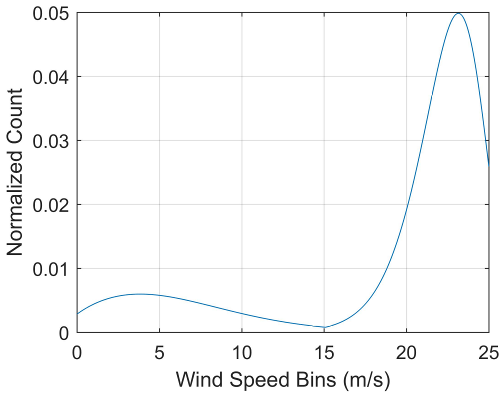
distributions were generated for the data from 0-15 m/s and 16-25 m/s. The range of the distributions was decided due to the shape of the data, shown in Figure 3.11 and it could vary depending on the analysis. This essentially creates two different probabilistic models for lower and for higher wind speeds. After that, the distributions were combined into a single one and they were averaged for the whole wind farm, as shown in Figure 3.12. This allows the generation of a generic model for the farm. The data shown are normalized against the frequency of the wind speed, meaning that for example for every 10-min average 23m/s recordings, there is a 5% chance that there will be an alarm. The normalization of the alarms against the wind speed was done by dividing the alarm count by the number of wind speed instances (i.e. wind speed occurrence) in every bin. This seeks to evaluate the alarms. The combination of the distributions was done by neglecting the tails and only taking only into account the more dominant part of the distribution, as shown in Figure 3.12b. This was done, as the tails of the distribution do not represent the observed data, as the dataset was split in two parts depending on the wind speeds. A chi-square goodness-of-fit test was used for the 2 Weibul distributions and the p values calculated were 0.1625 for the 0-16 m/s and 0.1108 for the 17-25 m/s distributions. These p values included the tails of the distributions; when using the combined Weibull distribution a p value of 0.9435 was calculated. For the TI, a Rayleigh distribution has been used, as shown in Figure 3.13, with a goodness of fit of 0.1259. The generic steps for the distribution generation are shown in Algorithm 4.

Algorithm 4 Generic distribution generator

```
Input: binned failure rates for individual subassemblies and generic turbine one
for all failure rate inputs do
    create distributions
    repeat
        test distribution goodness of fit
    until distribution satisfies fit requirements
end for
    return Weibull and Rayleigh distributions for the wind farm and subassemblies
```



(a) 2 Weibull distributions



(b) Combined Weibull distributions

Figure 3.12: Weibull distributions for a generic farm model.

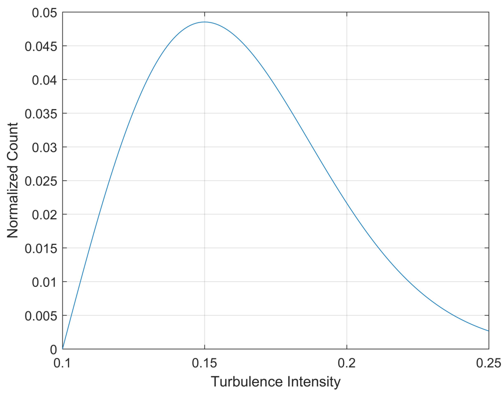


Figure 3.13: Rayleigh distribution for TI.

3.3.3.2 Tool

The tool reads as inputs time series of 10-min average wind speed and standard deviation data. The TI for the farm is then calculated, Eq. 3.1, by averaging all the individual TI values. The tool then reads the failure alarm data for the different subassemblies, in this case this includes 5 years of historical data collected from Teesside offshore wind farm. Then the data are analyzed as described in Section 3.3.3.1. The tool can then generate a future estimate of the alarms for the different subassemblies and an overall one for the wind farm. A generic step by step pseudocode is presented in Algorithm 5.

Algorithm 5 Generic alarm prediction tool algorithm

```
Input: met mast 10-min average and standard deviation timeseries
Calculate TI
Input: Weibull and Rayleigh distributions for the wind farm and subassemblies
for individual subaasemblies do
    initialize probability for subassembly
    for met mast parameters do
        select met mast parameters
        select equivalent parameters from Weibull and Rayleigh distributions
        assign probabilities for wind speed and TI
        average the two probabilities
        add to previous probability
    end for
    Multiply final probability with the number of turbines
end for
    return wind farm and subassembly number of alarms expected
```

3.3.3.3 Testing

Once the 10-min average wind speed and standard deviation data have been correlated with the fault alarms, a prediction can be made based on the provided predicted wind speeds. Once the data from all the turbine sensors have been aggregated and the average turbine model has been built, the tool can be tested. In order to test the capabilities of the tool, the inputs from the met mast were used, as shown in Figure 3.14. This gives values for a single data point, making it easier and less complicated to test, as just one source of information is needed in order to test the tool.

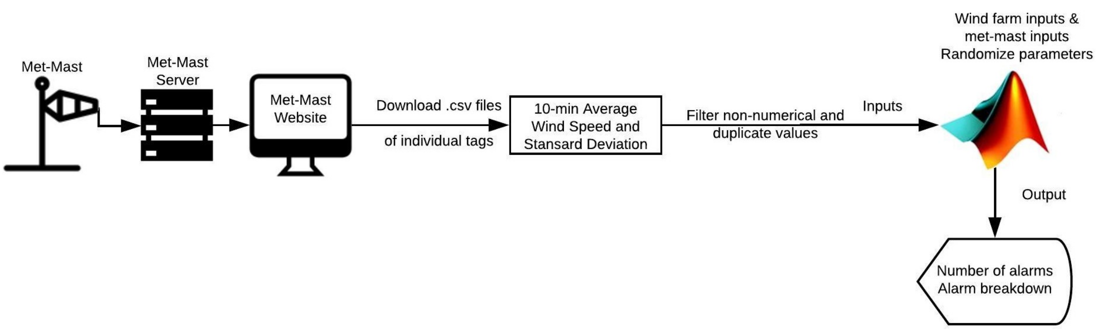


Figure 3.14: Alarm forecasting model inputs.

3.4 Machine Learning Models

An effective way for diagnosis and prognosis of wind turbine failures is to use ML algorithms. To demonstrate this capability, this section describes the data, the pre-processing and the failure detection and diagnosis techniques used to identify the spalling at the inner part of the planet stage bearing of an offshore wind turbine gearbox at Teesside offshore wind farm.

3.4.1 Gearbox Data Description

In order to avoid catastrophic failures of critical components, wind turbines are commonly fitted with RMS's; this becomes even more critical for offshore wind turbines. Thus, all modern offshore wind turbines have SCADA systems and most of them also include CMSs. The examined offshore wind turbines have both SCADA, CMS and particle counter installed.

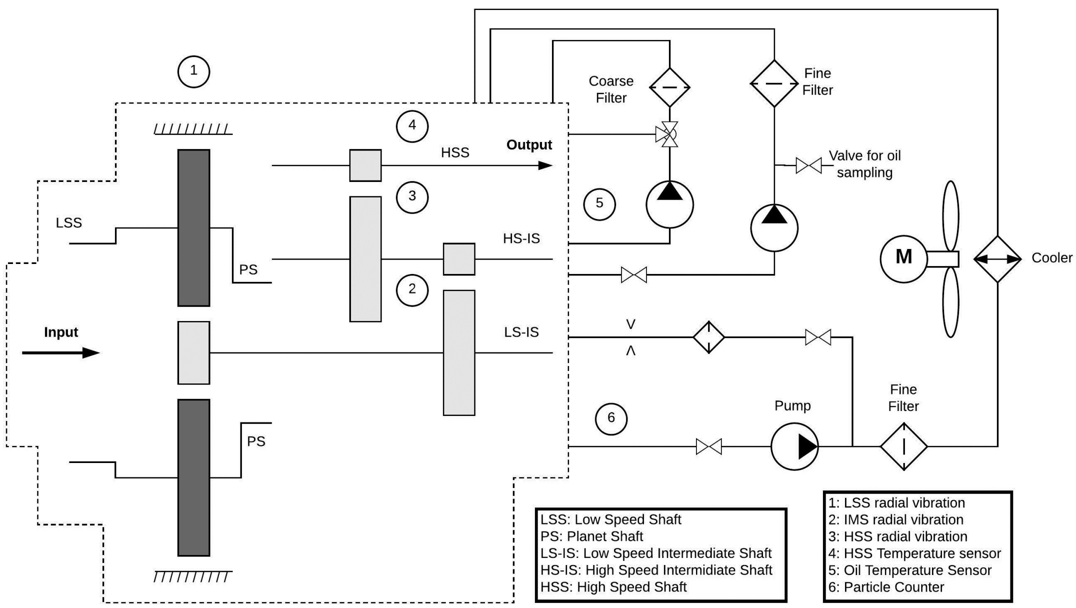


Figure 3.15: Schematic of a 2.3 MW 3-stage planetary/ helical gearbox and its cooling system.

A schematic diagram of the 3-stage gearbox examined for this study is shown in Figure 3.15, along with the available fitted sensors. These include three single axis accelerometers, one at each stage of the gearbox and an oil particle counter, as part of the CMS, as well as two temperature sensors measuring the high speed shaft (HSS) and the oil temperatures, as part of the SCADA system. Moreover, in this study SCADA sensors for the active power and the rotor speed have been taken under consideration. All the available data used in this study are up to 3 years prior to the gearbox replacement. The analysis and interpretation of the SCADA sensor data is usually straight-forward as it is captured in timeseries. The active power and wind speed data analysed are captured in 30-second average sampling and the temperature and rotor velocity data in 10-min average instances. The CMS data provided are pre-processed by the monitoring equipment and generated in lower sampling frequencies and within a specific time period and active power range. Hence, their sampling rate is dependent on the performance of the turbine and can vary between a few hours and a couple of days. The different analysis methods provided for the CMS data are FFT, cepstrum, envelope and rms. These methods were explained in Section 2.2.1.2.

CMS are expected to diagnose failures sooner and more precisely than SCADA, in locations where both systems are present. CMS could detect anomalies up to one year in advance, whereas SCADA systems up to three months prior to failure [12], this is due to the higher bandwidth of the signals, which allows a more detailed analysis of any anomalies in the vibrations of the structure or the machine, easing the fault diagnosis and prognosis. This difference is crucial for the O&M planning teams, to schedule the required maintenance operations.

3.4.2 Data Pre-processing

An integrated database has been created, incorporating the sensor readings from the SCADA, CMS, alarms and maintenance actions, received from the turbines, as presented in Section 2.3.

For the wind measurements received from SCADA, only the turbine with the replaced gearbox was investigated, wind turbine (WT) 14, due to uncertainties in the wind sensor calibration of the other turbines. Initially, certain instances were filtered out, by using conditional statements to remove the equivalent SCADA timestamps, for a duration of 2 years and 9 months, from the information received by the SCADA alarms and the maintenance logs, including: (i) Yaw system, (ii) Pitch system, (iii) Generator faults, (iv) Electrical and grid faults, (v) Sensor failures, (vi) Environmental conditions, (vii) Maintenance operations. This was performed, in order to remove any unrelated time instances. No gearbox related SCADA alarms or maintenance actions, apart from the scheduled routine inspections, had been activated during the lifetime of the faulty wind turbine, only regarding the gearbox oil level. Moreover, in order to examine the data further, the power curves and the rotor velocity data were binned, as indicated by the IEC 61400-12-2 standard [126]. The mean values of the normalized wind speed and normalized power output for each wind speed bin were calculated according to the standard.

CMS data pre-processing is not necessary, as the analysed data are generated and provided by the hardware supplier. They are also provided at different power ranges (0-920, 921-1150,1151-1403, 1404-1656, 1657-1909, 1910-2185, 2186-2415, 2416-3000) kW. For this study, the 2186-2415 kW power range has been selectedas it is close to rated power and gives a better indication for the gearbox's degradation. For noisy signals, a Savitzky–Golay filter with a high order number has been applied. This filter fits a set of data points to a polynomial in the least-squares sense.

3.4.3 Failure Detection and Diagnosis

Understanding, diagnosing and predicting the failure modes occurring on critical assemblies is important for reducing lead time for component delivery, as well as increasing asset availability. This is done remotely, by monitoring the information received from the SCADA and CMS. However, an inspection might be required to identify the failure root cause in more detail.

The different SCADA data readings, associated to the gearbox failure diagnosis used are the active power, rotor velocity, HSS temperature and gearbox oil temperature. The oil temperature against the square rotor velocity is also shown, because as indicated by [127], the gear stage inefficiency is proportional to the change in temperature over the squared of the rotor velocity, with the later being equal to the gear velocity at the planetary stage. This means that when a fault occurs on a gear stage, the temperature difference should increase in response to an efficiency reduction. Furthermore, the gear oil temperature for different time instances is binned and compared. The SCADA readings can give an initial indication of the failure location, which needs to be further investigated by the vibration data.

For the CMS data, the planet bearing readings are used for the analysis. It is difficult to understand if the signal represents a faulty or healthy turbine, without the exact machine frequencies and amplitudes held by the OEM. Thus, the different dates need to be plotted in order to understand the signal trending and how the failure is progressing in a waterflow diagram for the envelope and FFT spectrums of the planet stage. Moreover, the cepstrum rms of the planet stage and particle counting signals are examined.

3.4.4 Machine Learning Algorithms

3.4.4.1 SCADA

As there was no SCADA generated alarm related to any gearbox component, it would be interesting to investigate the possibility of creating warnings from the SCADA sensor readings that can trigger further investigation in similar future situations. Due to the nature of the data, different classification learning algorithms were selected and trained, including SVM, ensemble classifiers, decision trees and kNN. The input data used for the models include the active power, wind speed, rotor velocity, HSS temperature and gear oil temperature. The data were labelled as “healthy” or “warning” states based on the data generated before and after the gearbox replacement. For the different states, the training data were randomly selected on a 75%/ 25% of training/ test data, as suggested by [128]. The output of the models is the performance of the different classifiers.

Classification Algorithms

SVM SVM can be used when there are two classes of data, in this case the healthy and the unhealthy state. The data could be either separable or nonseparable. Separable data can be separated by a linear hyperplane, which separates the two states. For nonseparable data, the data could either have a soft margin that will separate the majority, but not all data points or with the use of a nonlinear transformation, using kernels [129]. In this case the sensor data received are considered nonseparable.

Training data are given for m observations as $T = \{X, y\} = \{x_i, y_i\}_{i=1}^{m}$, where $x_i \in R^n\}$, $y_i \in \{+1, -1\}$. In the simplest case of separable data, the two classes are separated by an n-dimensional decision boundary which is the result of an $n + 1$-dimensional hyperplane. Since multiple hyperplanes can exist, the desired one would be the one with the maximum margin. A learning method is needed that will optimize over the parameters w = [w1, . . . , wn] $\in R^n$ and $b \in R$ to find the optimal hyperplane, which is given by:

$$d(x, w, b) = w^T x + b = \sum_{i=1}^{n} w_i x_i + b \tag{3.2}$$

After defining the hyperplane, the data x can be classified as follows:

$$x = \begin{cases} Class1 & \text{if } d(x, w, b) > 0 \text{ (i.e. its associated } y = +1) \\ Class2 & \text{if } d(x, w, b) < 0 \text{ (i.e. its associated } y = -1) \end{cases} \tag{3.3}$$

For nonseparable data, as classes may overlap, it will not be feasible to find a (w, b) pair that will satisfy the class constraints. Consequently, a soft margin is required and the problem becomes:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i \tag{3.4}$$

such that $y_i[w^T + b] \geq 1 - \xi_i$ and $\xi_i \geq 0$ where ξ is a slack variable and C is a penalty parameter, often called a box constraint as it keeps the allowable values in a bounded region.

Some classification problems do not have a simple hyperplane, thus there is a variant of the mathematical approach that retains the simplicity of the SVM separating hyperplane using kernels. In this case, there is a class of functions $G(x1, x2)$ in a linear space S_l and a function ϕ mapping x to S_l (the dot product takes place in space S_l such that):

$$G(x1, x2) = < \phi(x1), \phi(x2) > . \tag{3.5}$$

The functions include polynomials, for a positive integer p, Eq. 3.6, Gaussian, Eq. 3.7 and multilayer perceptron, Eq. 3.8.

$$G(x_1, x_2) = (1 + x_1' x_2)^p. \tag{3.6}$$

$$G(x_1, x_2) = e^{(-||x_1 - x_2)||^2)} \tag{3.7}$$

$$G(x_1, x_2) = tanh(p_1 x_1' x_2 + p_2) \tag{3.8}$$

kNN A kNN classifier searches for the k points in the training set that are nearest to the test input x and counts how many members of each class are in this set and returns that empirical fraction as the estimate [104]:

$$p(y = c|x, D, K) = \frac{1}{K} \sum_{i \in N_K(x,D)} \mathbb{1}(y_i = c) \tag{3.9}$$

where $N_K(x, D)$ are the indices of the k nearest points to x in D and $\mathbb{1}(e)$ is the indicator function defined as shown below:

$$\mathbb{1}(e) = \begin{cases} 1 & \text{if } e \text{ is true} \\ 0 & \text{if } e \text{ is false} \end{cases} \tag{3.10}$$

Given an mx-by-n data matrix X, which is treated as mx (1-by-n) row vectors x1, x2, ..., xmx, and an my-by-n data matrix Y, which is treated as my (1-by-n) row vectors y1, y2, ...,ymy, the various distances between the vector xs and yt are defined as follows:

Eucledian distance

$$d_{st}^2 = (x_s - y_t)(x_s - t_t)' \tag{3.11}$$

Mahalanobis distance

$$d_{st}^2 = (x_s - y_t)C^{-1}(x_s - t_t)' \tag{3.12}$$

C is the covariance matrix

Minkowski

$$d_{st} = \sqrt[p]{\sum_{j=1}^{n} |x_{sj} - y_{tj}|^p} \tag{3.13}$$

Cosine distance

$$d_{st} = 1 - \frac{x_s y_t'}{\sqrt{(x_s x_s')(x_t x_t')}} \tag{3.14}$$

Decision Tree A decision tree predicts responses to data, by following the decisions in the tree from its root down to the leaf node, where the response of the classification is given as true or false [130, 131].

A tree can learn by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset until the subset at a node has all the same value of the target variable, or when splitting no longer adds value to the predictions.

Algorithms for constructing decision trees usually work top-down, by choosing a variable at each step that best splits the set of items. A commonly used metric for the expected error rate is the Gini's index:

$$I_G(\hat{\pi}) = 1 - \sum_{C} \hat{\pi}_c^2 \tag{3.15}$$

Ensemble Bagged Trees Ensemble methods combine several decision trees to produce more accurate predictive results. Bagging is the procedure for combining different classifiers that have been constructed using the same data set and is an acronym for bootstrap aggregating. It combines classifiers in order to improve an unstable classifier.

The predicted class for an observation is the class that yields the largest weighted average of the class posterior probabilities, in this case the classification scores, computed using selected trees only.

For each class $c \in C_t$ and each tree $t_d = 1, ..., T$, $\hat{P}_t(c|x)$ is computed, which is the estimated posterior probability of class c with observation x using tree t. C_t is the set of all distinct classes in the training data.

$$\hat{P}_{bag}(c|x) = \frac{1}{\sum_{i=1}^{T} a_t I(t_d \in S)} \sum_{i=1}^{T} a_t \hat{P}_t(c|x) I(t_d \in S) \tag{3.16}$$

where a_t is the weight of trees t, S_t is the set of indices of selected trees that comprise the prediction and $I(t \in S_t)$ is 1 if t is in the set S_t, and 0 otherwise.

The predicted class is the one that yields the largest weighted average:

$$\hat{y}_{bag} = \arg\max_{c \in C} \{\hat{P}_{bag}(c|x)\} \tag{3.17}$$

3.4.4.2 CMS

In terms of CMS signals, the only indication for degradation was noticed at the cepstrum rms values of the planet bearing readings. Thus, it was selected for further analysis and model predictions. An autoregressive (AR) model was used for predicting the future trend of the rms signal [132]. The model's parameters are estimated using variants of the least-squares method, by only using a historical data series. Seven hundred time instances have been used for training, with the last 300 instances representing the forecasted and actual future data that were generated.

The AR models are mainly used in financial models to forecast time series. The AR models are used in cases where past observations might predict current observations. A compact way to write the AR models is the following:

$$A(q)y(t) = e(t) \tag{3.18}$$

where q is the delay operator, $y(t)$ the output at time t and $e(t)$ the white-noise disturbance value.

3.4.4.3 General machine learning algorithm structure

A generic algorithm for using many of the machine learning models presented in the previous sections is shown in Algorithm 6.

Algorithm 6 Generic machine learning algorythm implementation

```
Separate training from test data
Input the different parameters
Perform feature selection
Perform dimensionality reduction
Choose and run the algorithm
    Train the model
    repeat
        Improve model accuracy by tuning parameters
    until high accuracy is obtained
    return Final model's accuracy
for New data do
    Run the model
end for
return Condition of the component
```

3.4.4.4 Measuring accuracy

The effectiveness of the algorithms is measured through their Accuracy (ACC) and F_1 score. ACC, given by Eq. 3.19, takes into account only the predictions that the model identified correctly, whereas the F_1 score takes into account both takes both false positives and false negatives, making it more suitable metric if there is an uneven class distribution.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.19}$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{3.20}$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

Chapter 4

Data Utilization

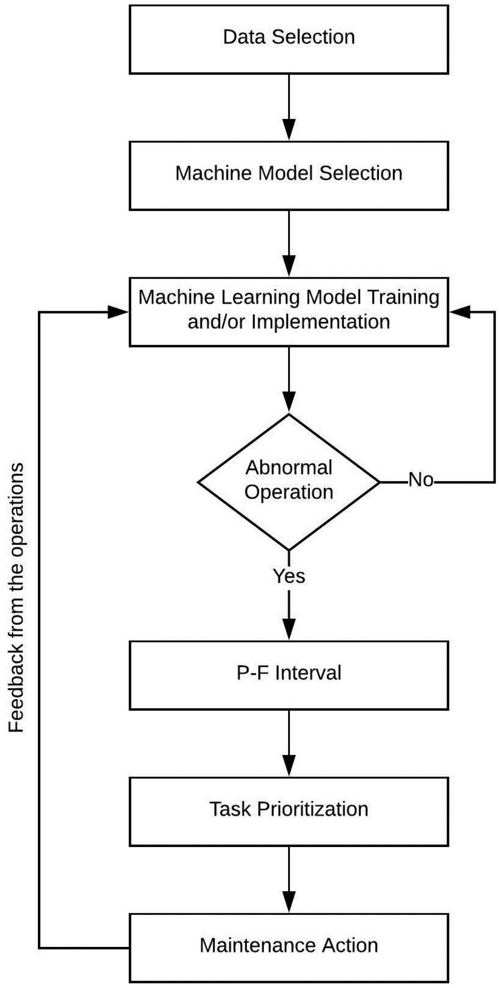
This chapter presents techniques aiming to reduce the operational costs by utilizing further some of the methodologies presented in Chapter 4. This includes a decision support framework for maintenance optimization, risk-based inspection and the use of an in-house O&M strategy tool.

4.1 Machine Learning O&M Decision Support Framework

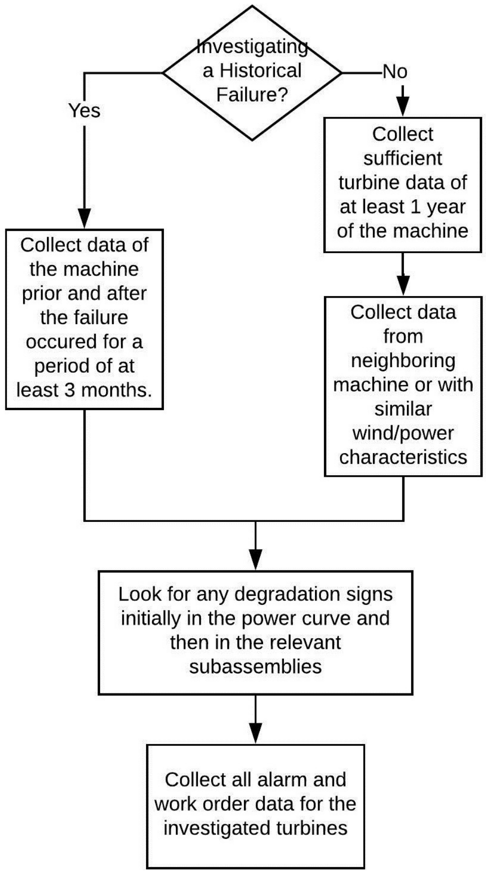
Better implementation and embedding of the ML models, presented in Section 3.4.4, in the operational workflow is needed in order to maximize the cost reduction opportunities that they can offer. An overview, showing the main elements, of the decision support framework (DSF) presented in this study is shown in Figure 4.1a. A more detailed diagram of the framework, is presented in Figures 4.1b, 4.3, 4.4 and 4.5, where the three stages of data selection, ML algorithm selection and ML algorithm and maintenance execution stages are presented. The main elements of the framework are discussed in the following sections.

4.1.1 Data Selection

Data selection is the first and most critical part of any ML study. At this stage, the user will have to select the correct and most representative data sources to use in the model depending on the failure mode investigated. Understanding the data selected, as well as being able to assess the quality of the data is key for a robust and accurate analysis. A



(a) DSF Overview



(b) Data selection

Figure 4.1: DSF overview and data selection.

period of at least 3 months prior to failure, which can form the training dataset, and 3 months after the component has been replaced or the issue has been resolved, which will allow the generation of a healthy-state dataset. It is better to collect healthy-state data after the failure has been resolved, as the duration of the failure consequences cannot always be traced back in time. In the case where there are no historical failure data, it is recommended to use a period of at least 1 year of data, which will show the seasonal variation experienced by the environmental conditions. Additionally, it is suggested to also collect data from neighbouring turbines, in order to be able to have a reference case and make the required comparison. An initial stage of identifying the degradation signs is proposed, by investigating the wind turbine's power curve, before looking at the individual subassemblies. Moreover, the collection of the relevant turbine alarms and work orders is key, in order to have a clear understanding about the turbine's condition and be able to understand or label the instances. This process is shown in Figure 4.1b. A data selection example for a gearbox planet bearing failure is shown in Figure 4.2.

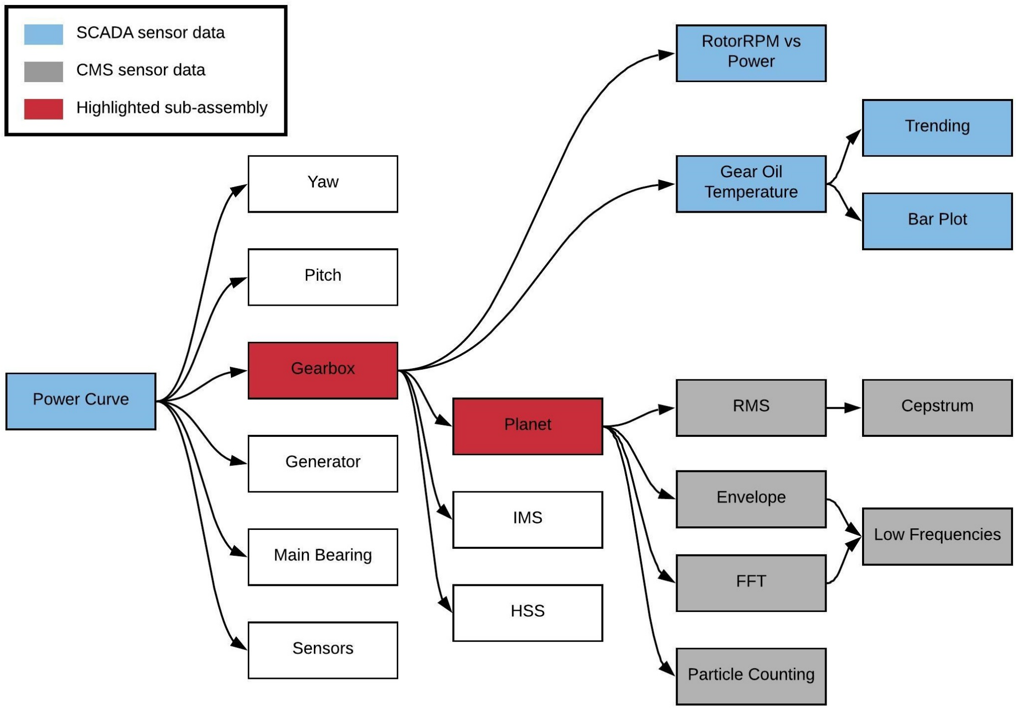


Figure 4.2: Data selection for a gearbox failure.

4.1.2 Machine Learning Model Selection

A selection of the most commonly used ML algorithms, that were identified in Section 2.4, can be seen in Figures 4.3 and 4.4. The decision tree has been created based on [104, 128, 133, 134]. A description of the major supervised learning algorithms was described in Section 3.4. The main categories include:

- Predicting category; if the data are labelled, a supervised learning algorithm should be used, otherwise an unsupervised learning algorithm can be selected.
- Predicting quantity; it can lead to regression algorithms.
- Investigating; it can be supported by implementing dimensionality reduction techniques to reduce the noise in the data and make relationships more apparent.

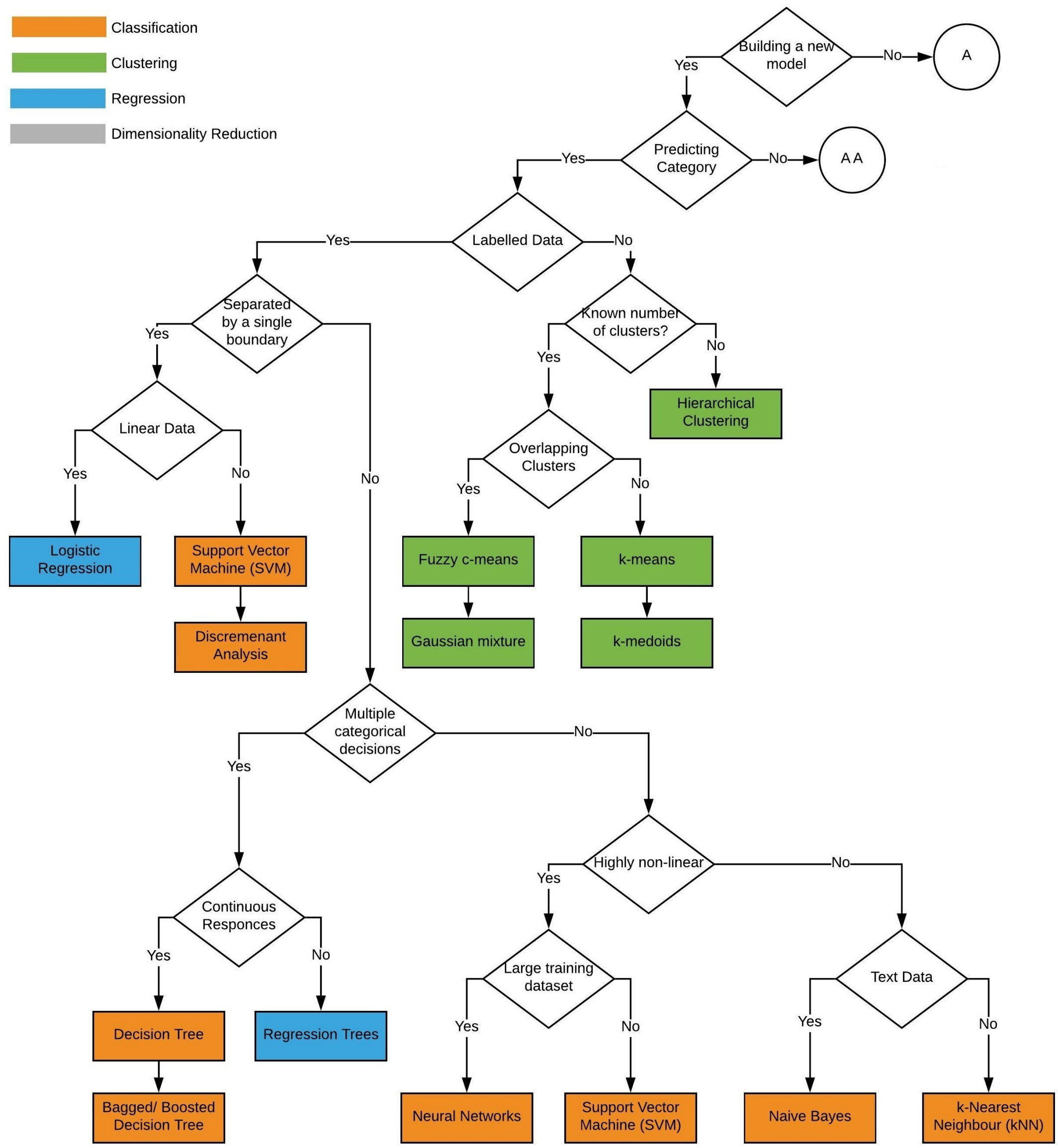


Figure 4.3: DSF: Machine learning algorithm selection (predicting category).

4.1.3 Machine Learning Model Training and Implementation

Once the appropriate algorithm has been selected and modified accordingly, it can be tested on the provided data, as shown in Figure 4.5. If needed, and in order to reduce computational time and complexity, a dimensionality reduction can be performed, as well as a feature extraction that is usually required for vibration signals such as FFT or rms analysis.

Once the model has been defined and completed, it can be subsequently improved

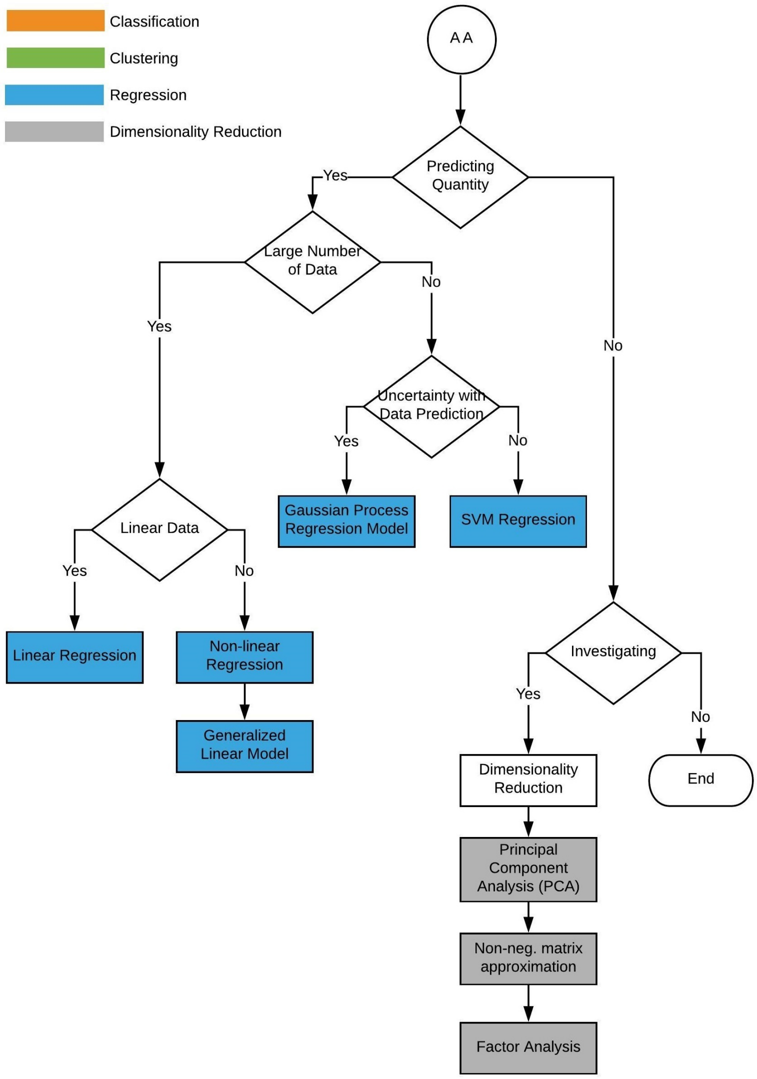


Figure 4.4: DSF: Machine learning algorithm selection (predicting quantity).

regarding its accuracy and goodness of fit, for example through hyperparameter tuning techniques, which are identifying the parameters that need to be altered in order to achieve the optimum result.

4.1.4 P-F Interval

The estimation of the remaining useful lifetime (RUL) of the component can be a useful result of a ML model. This can be enhanced with the use of a P-F interval diagram, which is defined as the interval between the occurrence of a potential failure and its decay into a functional failure [135].

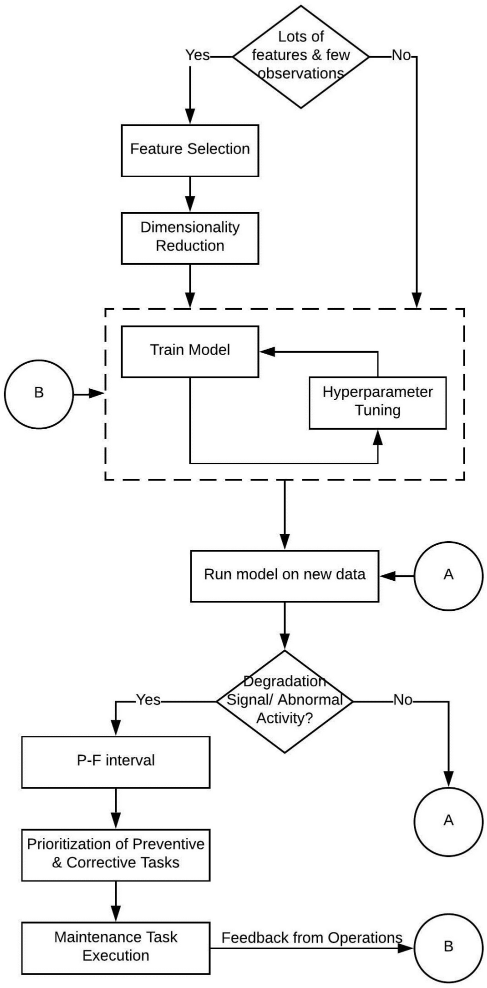


Figure 4.5: DSF: Machine learning algorithm and maintenance execution.

4.1.5 Maintenance Task Prioritization

Maintenance can be either preventive, which contains the planned maintenance actions performed at specific time intervals, or corrective, including any maintenance actions initiated by an unexpected or predictive wind turbine fault. Corrective maintenance is often the dominant type of maintenance strategy, being prioritized over preventive one. Whilst this may be operationally acceptable for current offshore wind projects, future multi-GW assets are likely to require a different, more formalized and data-driven approach. From operational experience of the authors, offshore wind farm service providers are not routinely using any tools or methodologies to prioritize maintenance activities and usually the prioritization is performed ad-hoc and based on engineering knowledge. Priority criterion functions, in the form of penalty functions, have been proposed by [136, 137, 138] and can be used for maintenance activity prioritization and grouping. Both cases are very important for the offshore wind applications, as access to turbines is challenging and the minimization of the turbines' downtime is key. The penalty function proposed by Dekker [136] is shown in Eq. 4.1.

$$h_i(\Delta T) = c_i^r\Big(\frac{x_i^* + \Delta T}{\lambda_i}\Big)^{\beta_i} + c_i^r\Big(\frac{x_i^* - \Delta T}{\lambda_i}\Big)^{\beta_i} - 2c_i^r\Big(\frac{x_i^*}{\lambda_i}\Big)^{\beta_i} \tag{4.1}$$

with $-x_i^* < \Delta T < x_i^*$ and where c_i^r is the failure repair cost, λ_i the scale parameter of the Weibull distribution (in days), β_i the shape parameter of the Weibull distribution, x_i^* the optimal replacement age (in days), calculated by Eq. 4.2.

$$x_i^* = \sqrt[\beta_i]{\frac{(c_i^p + S_c)\lambda_i^{\beta_i}}{c_i^r}} \tag{4.2}$$

where S_c is the set-up cost and c_i^p the cost of the activity carried out at any moment against the same cost.

Table 4.1: Example inputs for cost functions.

Action	S_c	λ_i	β_i	c_i^p	c_i^r	t_i
A	1	350	2	10	10	20
B	10	200	2	1000	500	25
C	2	200	2	100	50	30

By substituting the values from Table 4.1 in Eq 4.1 and 4.2, the different activities' cost functions can be visualized, Figure 4.6. t_i is the initial planning moment, which is predefined for the scheduled maintenance activities and can be estimated with the use of the P-F interval for the corrective actions. t_i also represents the optimal execution time of the activity, with any deviation from it resulting in a penalty cost. Action A could be an activity related to a component inspection that is less urgent, B a major component replacement that needs to be done as soon as possible, and C a minor component replacement. By visualizing the cost function, a task prioritization can be facilitated. If, for example, action A cannot be implemented at day 20, due to weather conditions, it could still be shifted to day 34, after the other two actions have been completed, which cannot be the case for actions B and C due to their high penalty costs. Finally, the visualization of the penalty cost function could aid in maintenance activities' grouping that are performed at the same or at neighbouring turbines.

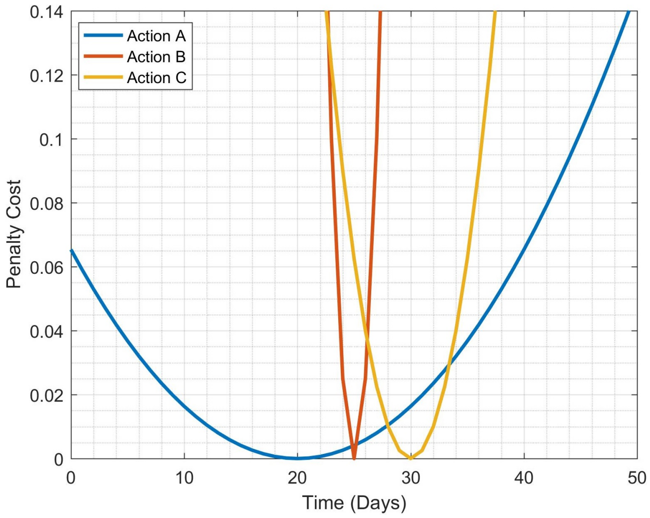


Figure 4.6: Penalty cost functions example.

4.1.6 Maintenance Action

After the maintenance activities have been prioritized, the resources can be allocated and the weather window is found, the maintenance activities can be performed. A proposed data workflow for integrating scheduled and unscheduled maintenance data into the data analysis and maintenance planning is shown in Figure 4.7, modified from how Boeing is utilizing the operational data [139, 140]. The framework shows a streamlined way of integrating the operational data into the analysis, with a centralized data-driven decision making. During scheduled maintenance, if there is a defect in the components and a non-routine task is executed, it needs to be logged and its root cause to be identified. If the findings are unrelated, data mining might help in order to better understand the issue. In the case where no findings have been detected, a report can be generated and the scheduled maintenance interval can be revised if needed. For unscheduled maintenance tasks, once the task has been executed, data mining and RCA of the failures can be carried out to understand the fault better. The data collected can also be used for supervised ML models. If a model has been used to diagnose or predict a failure for that maintenance task, it needs to be improved accordingly from the field data. Similarly, the scheduled maintenance intervals can be revised, once the failure has been fully understood.

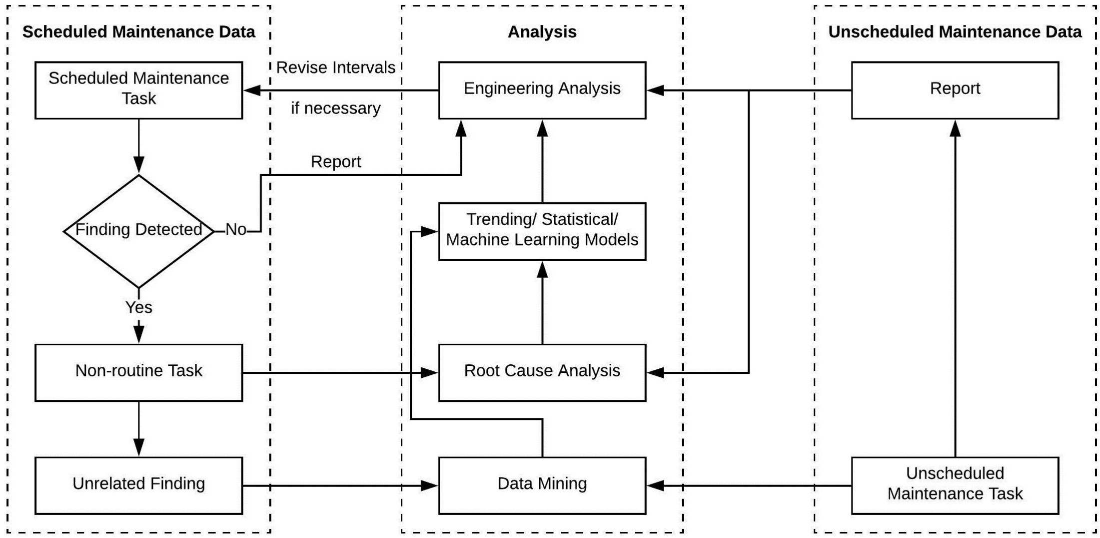


Figure 4.7: Scheduled and unscheduled maintenance data workflow.

4.2 Risk-based Operations

This section proposes an RBI framework that is later being tested for TP inspections. It has been created based on common approaches in other industries [141, 19, 142, 134]. An applied RBI technique is proposed that comprises of four main stages, as shown in Figure 4.8, which create a loop in order to improve future inspection and maintenance planning. The individual phases are described in more detail in the following subsections.

4.2.1 Initial Decision and Evaluation

The initial decision and evaluation stage includes the sample size, the location of the turbines and the identification of the components that need to be inspected. It is common practice to select an inspection sample. This is usually a large number during the first inspections and it can be reduced later on, as this is a dynamic decision process that needs to be re-evaluated at every inspection interval. Once the turbine sample and location have been defined, a risk matrix can be created for all the inspected

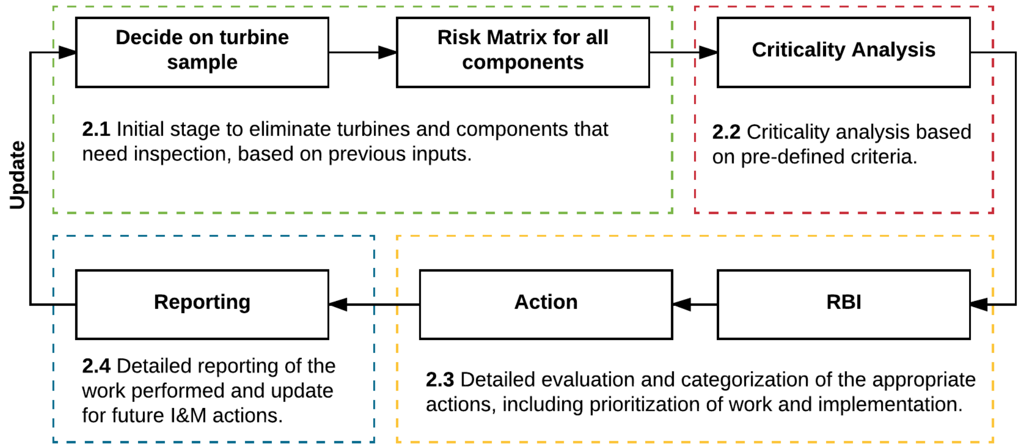


Figure 4.8: High-level overview of the risk-based framework.

components. A typical risk matrix used is shown in Table 4.2. This type of risk matrix was chosen as it is the industry standard, used by the operational teams of EDF- this way the results of this methodology can be easily adopted. The numbers on the matrix are just for reference purposes, which are used later on in the model. This is an initial screening step to eliminate those components that will not need to be inspected. The information can be generated by considering previous inspection reports and understanding the potential failures and their impact. This process significantly reduces the amount of time needed for the RBI, as fewer components will be considered.

Table 4.2: Risk Matrix

		Impact				
		V.Low	Low	Medium	High	V.High
Probability	V.High	8	11	19	21	25
	High	7	10	14	20	24
	Medium	3	9	13	17	23
	Low	2	5	12	16	22
	V.Low	1	4	6	15	18

4.2.2 Criticality Analysis

Once the initial screening stage is complete, a criticality analysis is performed for the remaining components, as developed and shown in Figure 4.9.

The Cost-Risk Priority Number (C-RPN) is calculated by taking under consideration the difference of the failure cost (C_D), the severity of effect (S) or severity of restriction (R) that the failure could create, the occurrence probability (O) and detection method (D), shown in Eq. 4.3 and 4.4. S, O, D and R, are integers in the range 1-10 given by definitions in Table 4.3. Depending on the consequences of the severity in terms of production effects or restriction issues, the S/R values can be decided. In most of the cases it is clear if the severity affects the operations or the power production of the wind farm, but if there is a case where both are affected, it is suggested to take under consideration the highest C-RPN number.

$$C\text{-}RPN = C_D \times S \times O \times D \tag{4.3}$$

$$C\text{-}RPN = C_D \times R \times O \times D \tag{4.4}$$

C_D (Eq. 4.5) is the difference between the total loss without corrective action (C_a) and the total expected loss after the corrective action (C_n). If the value is negative, the implementation of the maintenance action should be reconsidered, by initially checking if there are any restrictions or safety issues that might be caused if the issue is not fixed.

$$C_D = C_a - C_n \tag{4.5}$$

A guidance on how to define the different S, R, O and D values is given in Table 6.12. The initial thresholds have been specified by [143]. The risk priority number (RPN) criteria have been further refined in order to reflect the offshore wind turbine operations, RMS's, non-destructive testing (NDT) inspections and potential restrictions. RMS include any type of SCADA, CMS and SHM systems.

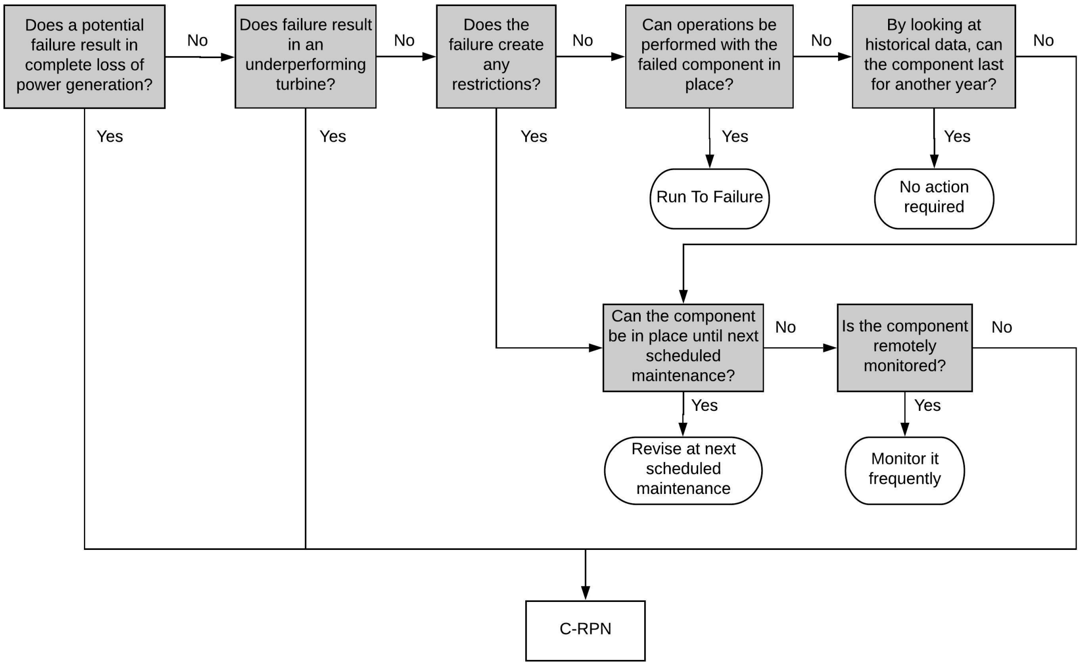


Figure 4.9: Criticality analysis steps.

4.2.3 Risk-based Inspection Implementation

Once the criticality analysis (CA) has been performed and the high RPN operations have been defined, RBI can be applied (Figure 4.10). The framework is evaluating the severity and importance of the operation. Low importance is an occasion that can be fixed easily with basic tooling and as a result, the action can be immediately planned. In the cases of medium and high importance, further processes are followed in order to identify the root cause of the problem via a Failure Modes, Event and Criticality Analysis (FMECA). This will ensure that the causes are well understood and preventive measures are implemented in order to avoid any similar future failures. Finally, the maintenance actions can be planned.

4.2.4 Reporting

The final step is the reporting of the executed inspection and maintenance information. This is a very important step and the information has to be recorded in such a way so that it can be easily retrieved and related to other sources. At the same time, the information needs to feed forward to update the current status of the assets and be able to advise future inspection and maintenance (I&M) decisions.

Table 4.3: RPN detailed criteria ranking.

#	**Severity (S) of Effect**	**Occurrence (O) Prob.**	**Likelihood of RMS detection (D)**	**Likelihood of NDT det. (D)**	**Severity of restriction (R)**
1	No effect	<1 in 1,500,000	RMS will certainly detect failure cause/ failure mode.	The failure can be also identified by inspection during operations.	No restrictions.
2	Very minor effect on power production.	1 in 150,000	Very high chance the RMS will detect a failure cause/ failure mode.		Very minor restrictions.
3	Minor effect on power production.	1 in 15,000	High chance the RMS will detect a cause of failure/ failure mode.		Minor restrictions.
4	Small effect on power production, repair not required.	1 in 2,000	Moderately high chance the RMS will detect a failure cause/ failure mode.	The failure might also be identified by inspection during operations.	Few restrictions, repair not required.
5	Moderate effect on power production, repair required.	1 in 400	Moderate chance the RMS will detect a failure cause/ failure mode.		Moderate restrictions, repair required.
6	Component performance is degraded.	1 in 80	Low chance the RMS will detect a failure cause/ failure mode.	Visual inspection is needed to detect the failure.	Severe restrictions, repair required to continue operations.
7	Component is severely affected. Turbine may not operate.	1 in 20	Very low chance the RMS will detect a failure cause/ failure mode.		
8	Component is inoperable with loss of primary function.	1 in 8	Remote chance the RMS will detect cause of failure/ failure mode.	Need NDT monitoring equipment to detect the failure.	Restrictions do not allow safe access to the turbine.
9	Failure involves hazardous outcomes.	1 in 3	Very remote chance the RMS will detect a failure cause/ failure mode.		
10	Failure is hazardous and occurs without warning. Turbine operation is suspended.	>1 in 2	RMS will not detect a failure cause/ subsequent failure mode.		

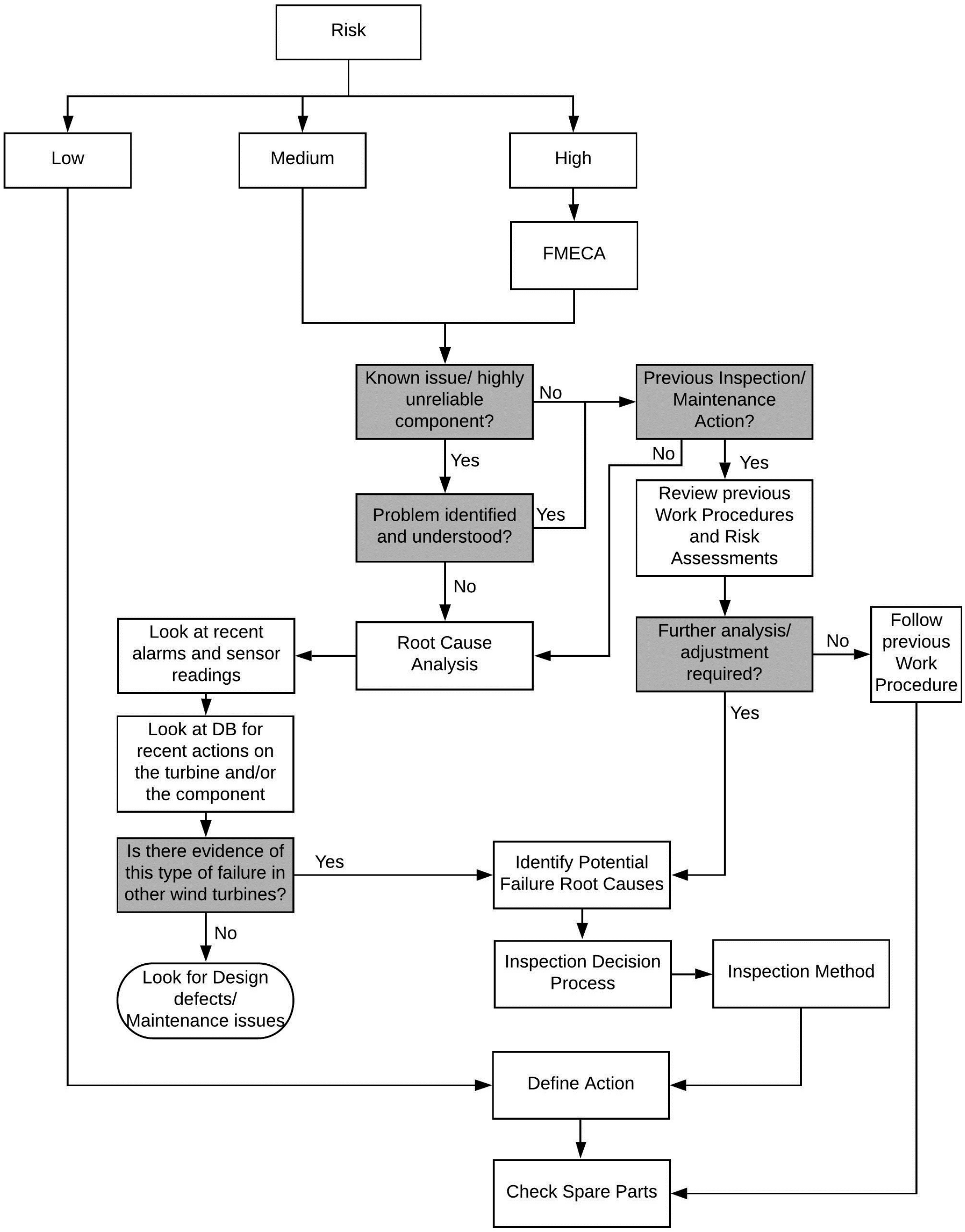


Figure 4.10: RBI plan detailed breakdown.

4.3 O&M Strategy Planning

This section presents the use of an O&M strategy planning tool required to run different wind farm lifetime availability studies. The reliability inputs for the tool include data from literature used to run different scenarios, enhanced with observed data from Teesside offshore wind farm, following on from Section 3.3.

4.3.1 Reliability Data Analysis

This methodology aims to consider the different reliability parameters of the individual turbines in the farm, thus a data-informed approach is used by enhancing the literature parameters with site specific ones. This is achieved by analysing 5 years of operation of turbine alarms from the site. Turbine alarms are used and the process followed in Figure 3.9 is used to create the operational data statistics.

4.3.2 O&M Tool Description

The different scenarios are run using an integrated asset management tool for strategic planning of O&M activities. The tool is comprised of 6 modules; weather, reliability, wind turbine, logistic, project and economic, as shown in Fig. 4.11. The weather and reliability modules are two separate engines generating the weather conditions and the failures that are the intervention instances, which are then simulated by the other four modules. The tool provides as outputs the different associated OPEX costs, reliability and availability information of the assets and statistics about the different operations taking place. It runs on a MCS for all the input parameters and all simulations in this case study have used 10,000 repetitions.

A description of the different modules is outlined in the following subsections.

4.3.2.1 Weather generation

The tool allows different types of weather models to be used; one that reads directly the historical data provided, another one that utilizes the historical data to create a synthetic weather series, using a Neural Networks model and another one that uses historical data to create a synthetic weather series using Markov Switching

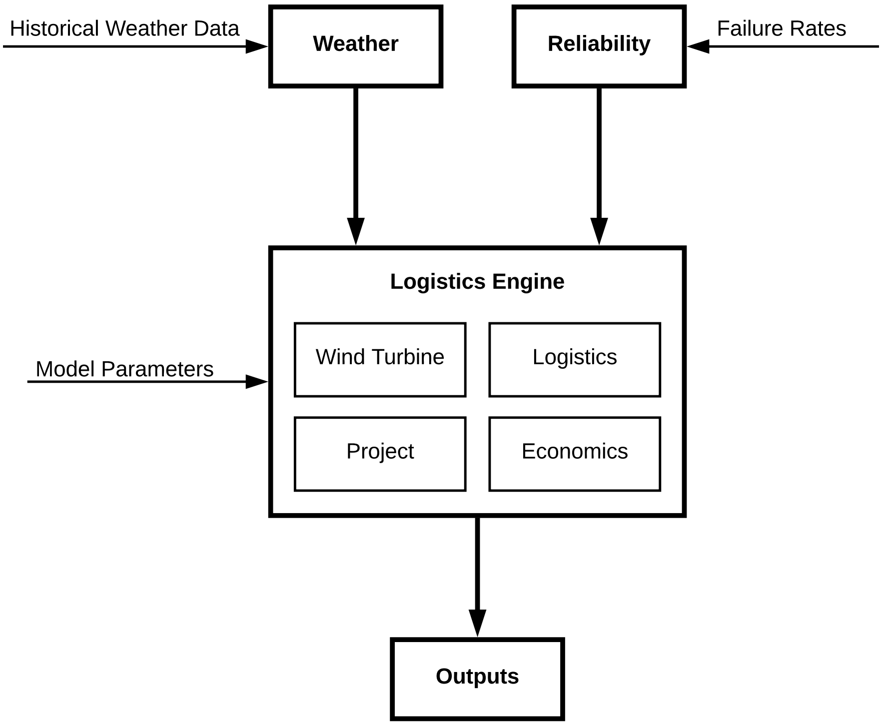


Figure 4.11: Overview of the asset management tool.

Autoregressive Model (MS-AR). The parameters used for the weather generation are the wind speed and the wave height of the site. The synthetic time series can create some stochastic weather outputs, which can provide better results when simulating multiple future scenarios. For this study, the MS-AR model is used, as it has previously been validated for the selected site. It uses the homogeneous MS-AR model with Gaussian innovations, embedded within the METIS Matlab simulation tools. To cope with the variation in the offshore weather conditions at each project, it is vital that a stochastic weather model is adaptable to account for seasonal or monthly variations to produce realistic weather time series. Thus, this model was used and configured to produce monthly realisations of wind speed and wave height, which is conditioned into one full continuous time series. The generated time series were cross examined against a typical offshore operation, to validate the adopted methodology and MS-AR model configurations and can be found at [144]. By demonstrating that the model is capable of reproducing operational characteristics such as the average length of operational weather windows and monthly workability (i.e. the amount of time above the threshold available to complete the operations), it supports the case to implement the methodology and embedded MS-AR models within offshore simulation tools such as the one presented in this section.

4.3.2.2 Reliability

It is used for generating probabilistic or calendar based events. In the first case, the events are generated to simulate unexpected failures of the components and the latter to simulate preventive maintenance tasks that have been scheduled for a specific period of the year. The generated failures can be designed for as many components as the user requires to every level of detail; from a component up to the system level. The failures are not linked and are governed by the individual input parameters. The tool is able to handle different reliability distribution types; exponential, Weibull, Dirac and a user defined one. Once the failure mode has been triggered, a maintenance task is activated; when the maintenance task is complete the repaired or replaced component can be modelled to be as good as new, as bad as old or to have a new reliability value.

4.3.2.3 Wind Turbine

The wind turbine modules are then linked with the different reliability components triggering the maintenance actions. It is possible to determine the nominal power of the turbine, as well as the capacity factor for every month. These values are used for estimating the total power generation, as well as the economic parameters required.

4.3.2.4 Project management

The project management module is used to specify the different tasks taking place and are linked to the reliability components. This allows the modelling of all the tasks required for an offshore maintenance operation. For example, the transit in and out of the vessel as well as the duration of the task or of individual sub-tasks. It is possible to add restrictions to the tasks, such as weather limitations, as well as technicians' shift time.

4.3.2.5 Logistics

The logistics module comprises all the resources needed to perform an operation and it is linked to the individual tasks. This includes the available vessels, personnel, equipment and consumables needed for each operation.

4.3.2.6 Economic

All of the operations taking place are linked to a cost, which are then discounted to the present value and aggregated to provide the total OPEX costs.

4.3.3 Model Outputs

In order to compare the results, two different metrics are used; time-based availability (TBA) [145] and production-based availability (PBA) [146]. TBA is the fraction of the time that the turbine is available (i.e. it is not shut down for any maintenance or due to a failure) over the total operational time of the asset, as shown in Eq. 4.6. PBA is the fraction of the actual energy produced by the asset over the potentially energy expected (i.e. including the energy lost when the turbine was shut down), shown in Eq. 4.7. These are calculated in the tool, by providing the average monthly capacity factors at Teesside offshore wind farm, which are undisclosed due to confidentiality reasons.

$$TBA = \frac{TimeAvailable}{TotalTimeConsidered} \tag{4.6}$$

$$PBA = \frac{EnergyActuallyProduced}{EnergyPotentiallyExpected} \tag{4.7}$$

Part III

O&M Case Studies

Chapter 5

Operational Analysis

This chapter presents the results of four case studies, outlined below, that are based on the methodologies described in Chapter 3. The purpose of the case studies is to validate and demonstrate the value of the methodologies developed.

- **Work Order Analysis** This section presents the simplest type of analysis that is possible with the operational data. Work order statistics can provide an initial insight into the health status of the turbine and are a metric used for long-term reliability forecasting for subassemblies.
- **Reliability Analytics** A more detailed approach for analysing reliability information is presented in this case study, aiming to better understand the failure mechanisms and how the environmental parameters affect the offshore wind turbine reliability.
- **Failure Root Cause Analysis** The use of integrated data from different sources is investigated in order to perform failure RCA.
- **Gearbox Replacement** An insight into the different data for diagnosing and predicting a gearbox failure in advance is shown. The data from the SCADA and CMS systems are used in order to build and test ML models.

5.1 Work Order Analysis

This section presents interesting reliability insights and conclusions that have been derived by analysing the work orders of an offshore wind farm.

5.1.1 Failure Rates

The normalized failure rates calculated for an 18 month period, between the 2nd and 3rd year of operation of Teesside offshore wind farm can be seen in Figure 5.1. The work orders were manually extracted from the operational reports and then for each failure a subassembly was assigned to it, as described in Section 3.2.2. As it can be seen, the majority of the failures occur for auxiliary systems, which reach almost 30% of the failures if "Other Components" and "Service items" are grouped together. These include lifts, lights, service cranes, latchways, etc. The gearbox is the second biggest contributor with 14%, followed by the electrical components, sensors, heaters/coolers, power supply/ converter and controls, ranging from 5-9%. The rest of the subassemblies contribute less than 5% each to the total failures. The results have been plotted against the ones from a wind farm consisting of ~350 wind turbines, during their 2nd and 3rd year of operation, as presented in [62]. The study results show that the majority of the failures (~15-17%) appear on the Generator and at the Pitch/ Hydraulic Systems and the Gearbox subassemblies. At a first glance in the comparison of the failure rates, a difference in the distribution of failures within the different subassemblies can be noticed. Although it was attempted to follow the same taxonomy as [61], some components could have been placed in different subassemblies, thus this could explain some of the differences. Moreover, [61] uses 350 turbines from different sites, with different environmental conditions, generators and most probably manufacturers. The disadvantage of such a model is that it is grouping turbines together, assuming similar failure rates amongst them. As discussed in the literature review, it has been reported that different capacity/ manufacturer turbines have different failure rates. This is evident in Figure 5.2, where a difference in failure rates can be seen even in neighbouring turbines.

Considering the individual subassemblies (Figure 5.1), the generator has the

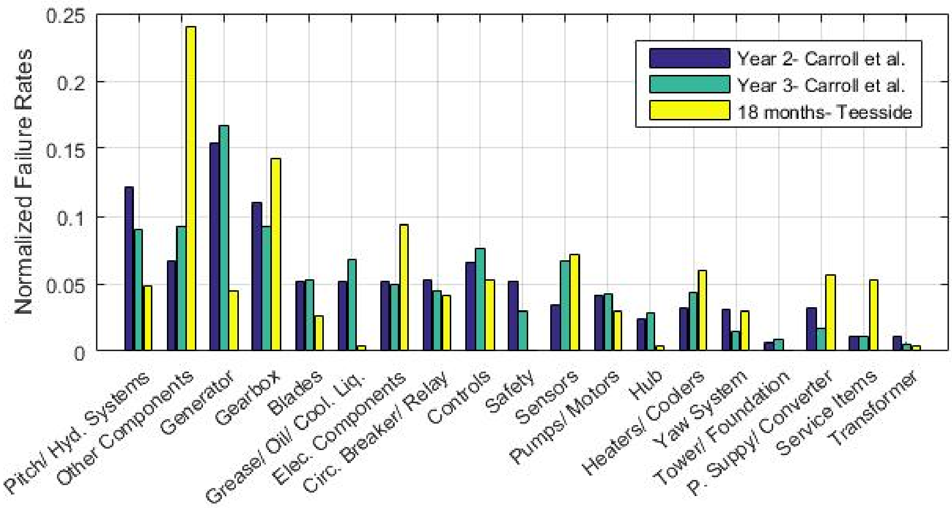


Figure 5.1: Normalized failure rates for subassemblies from work orders. Comparison of 18 months of data from Teesside with [62].

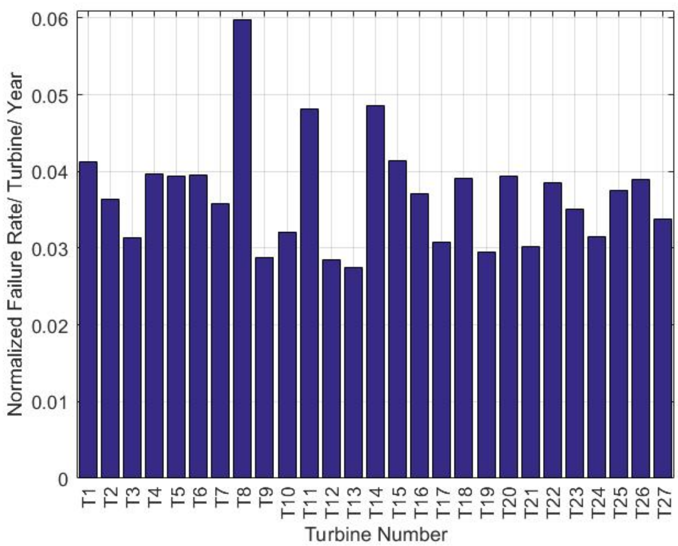


Figure 5.2: Normalized failure rates for the the different turbines generated using work orders.

majority of failures rates in [62], whereas at Teesside, the majority of failures appear on the "Other Components". The difference of the two studies can be traced back to the different generator types, although there is not much information provided on the generator type, it can be seen from the breakdown of the components at [61] that

1/3 of the generator failures are on the slip ring. At the Teesside case, the generator used is asynchronous and does not use slip rings, which reduces the failure rates. The LWK DB has also shown that the onshore turbines with asynchronous generators have lower failure rates [12]. Pitch and hydraulic mechanisms have the highest contribution to failure rates throughout the lifetime of the turbines as shown in [70, 61]. For the first 2-3 years of the project they have the second highest contribution to failure rates, which is not the case for Teesside. In general, Other Components, Gearbox, Electrical Components, Power Supply/ Converter and Service Items have proportionally higher failure rates at Teesside. All the other subassemblies have proportionally similar or less failure rates.

Differences could also be due the fact that not all of the turbines were installed and connected to the grid at the same time, so they are not exactly the same age. Furthermore, the fact that Teesside is relatively close to the shore, which results in lower average wind speeds throughout the year of ∼7.5 m/s and a capacity factor of 37%, which is at the low levels of the average wind speeds presented by [61], could mean that all the electrical and mechanical components will have higher life expectancy and consequently lower failure rates.

5.1.2 Wind Speed and Failures

A correlation between the failure rates and the average wind speed using the work order data was attempted, shown in Figure 5.3. The failure rates per turbine, per season, plotted against the average wind speed recorded from two sensors mounted on each of the turbines' nacelles are shown.

Some correlation between wind speed and failure rates can be seen, as other studies have shown in the past for onshore and offshore wind turbines [12, 61, 147]. This result was possible, only when the inspection failures were excluded. This is important, since failure/ wear of the components that are observed during inspections, do not represent the environmental conditions of when the failure appears, as the inspections usually occurs between March and September, when the average wind speeds fall to 6-7 m/s. When these failures are incorporated in the failure rates, this correlation is not visible.

A limitation of such an analysis is that at least in the presented case study, the correlation is not very clear. The reason for that is due to the fact that the time that the operations are taking place is not the same as the time when a failure has occurred. These two can be apart from a few hours to days or even months, which means that the weather conditions different. This indicates that it might be hard to observe such a correlation by using the data from the work orders.

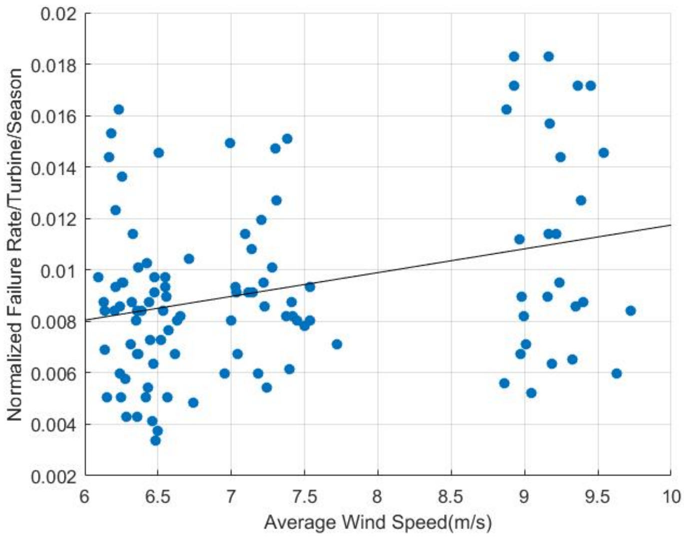


Figure 5.3: Normalized failure rates against average wind speed.

5.2 Reliability Analytics

In order to address the limitations of using the work order information to generate reliability statistics, as shown in the previous section, the alarms are analyzed, as explained in Section 3.3. This approach is followed in order to better correlate the effect of environmental conditions, such as wind speed, TI and temperature, as well as the location of the turbine to the farm with the overall reliability of the individual assets and their subassemblies.

In order to investigate the effect of TI and wind speed on the wind turbines, the actual, post-construction turbine spacing in the wind farm was investigated. Spacing between rows at Teesside offshre wind farm is 6.13D, within the rows 3.44- 3.76D and diagonally between different turbine rows, starting from 6.24D, where D is the turbine rotor diameter. Figure 5.4 shows the wind rose for over two years, collected from the site's met mast, located on land around 12.5D east from T19. It can be seen that the majority of the wind is coming from the land, from the southwest direction.

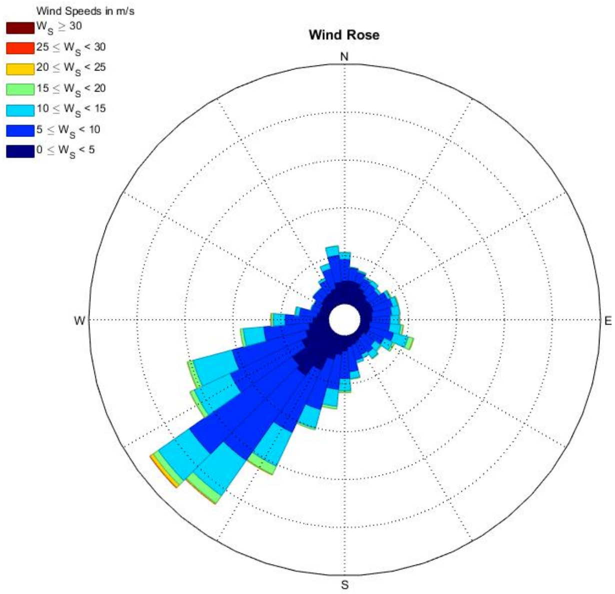
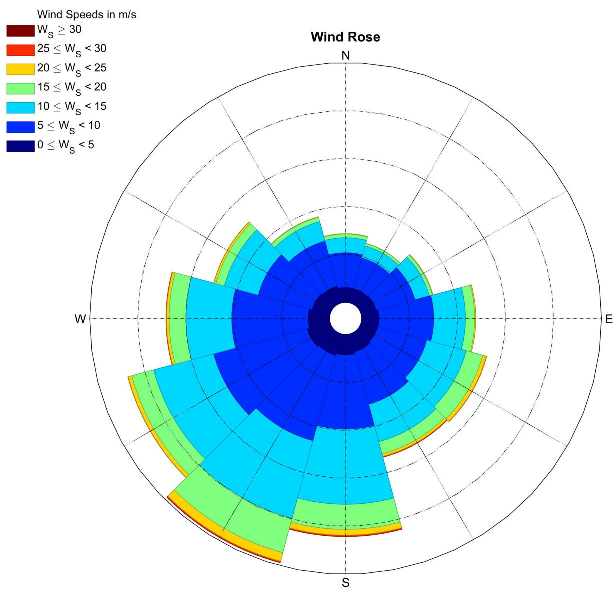
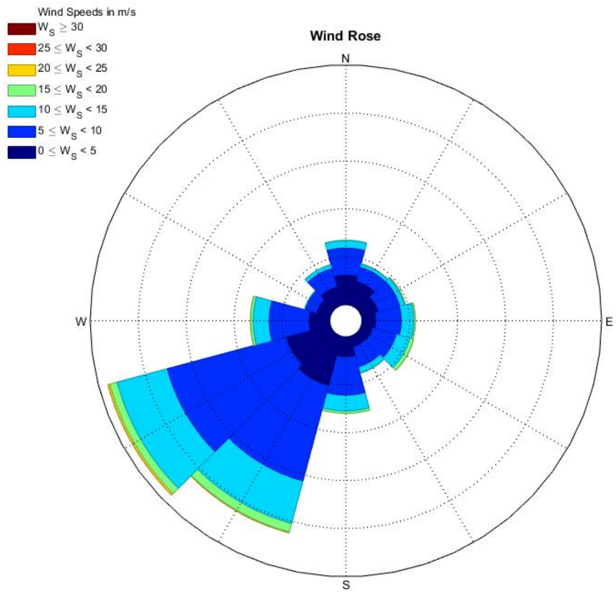


Figure 5.4: Met mast wind rose from September 2015- September 2017 on 10° intervals.

A comparison of the pre-construction and operational wind speeds, shown in Figure 5.5, can reveal potential differences between the pre-construction and operational measurements. The pre-construction data were measured at 20 m. above the average mean sea level and extrapolated to 80 m. height. A directional sector analysis was conducted on the data to determine suitable boundary layer power law wind shear exponent values to use for translating the measured values with the turbine hub height wind data. There is a significant difference between the wind speed and wind direction measurements before and after the construction of the wind farm. For both pre and post-construction cases, the data show dominant winds from the southwest. For the pre-construction case, the wind data have a wider spread across different directions, which could have influenced the wind farm layout. It is assumed that the design layout of the farm was decided with the available pre-construction data. If the provided data were closer to the observed ones from operations, a different wind farm layout might have been chosen.



(a) Pre-construction



(b) September 2015- September 2017

Figure 5.5: Met mast wind rose from pre-construction and operational readings on 30° intervals.

5.2.1 Turbulence Intensity and Wind Speed

Figure 5.6, shows the normalized, average TI for 4 years of operation for all the wind turbines. It can be clearly seen that WT13 has the highest TI value. In contrast, turbines 9 and 19 have the lowest TI values. The different turbine wind speeds, averaged for 4 years of operation, can also be seen in the same figure. As expected from the definition of TI, it is directly proportional to the lower wind speeds. As an effect, WT19 has the highest average wind speed and WT13, the lowest.

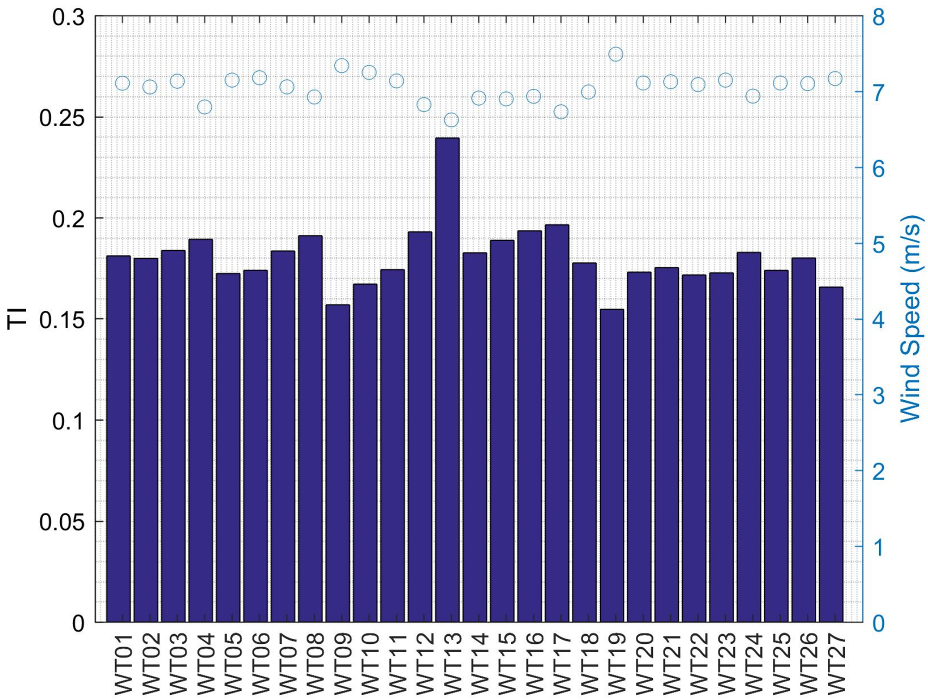


Figure 5.6: Normalized TI values for the different turbines.

The wake effects can be deduced from the TI parameters [125]. According to Figures 2.16, 5.4 and 5.6 the TI is higher in row B, due to the wake effect from row A. Due to the wind farm design, the wake effects are not affecting row C, as they are directed towards the spacing between the turbines at the dominant wind direction, thus the TI is lower at WT19 to WT27.

Figure 5.7 shows the behaviour of the TI for the different wind speeds. WT27 has been selected to visualize the TI, as it is one of the turbines with average wind speed and TI values. TI values are very high at low wind speeds and keep decreasing, until

11-12 m/s wind speeds and then again they start increasing until around 21 m/s, where they stabilize. Figure 5.8 shows the TI values for the operational turbine wind speeds from the cut-in (4 m/s) to the cut-out speed (25 m/s) [148], where the TI profile can be better observed. For both figures the average and the 90th percentile are shown, as most of the components are designed as an average of the 90th percentile [149].

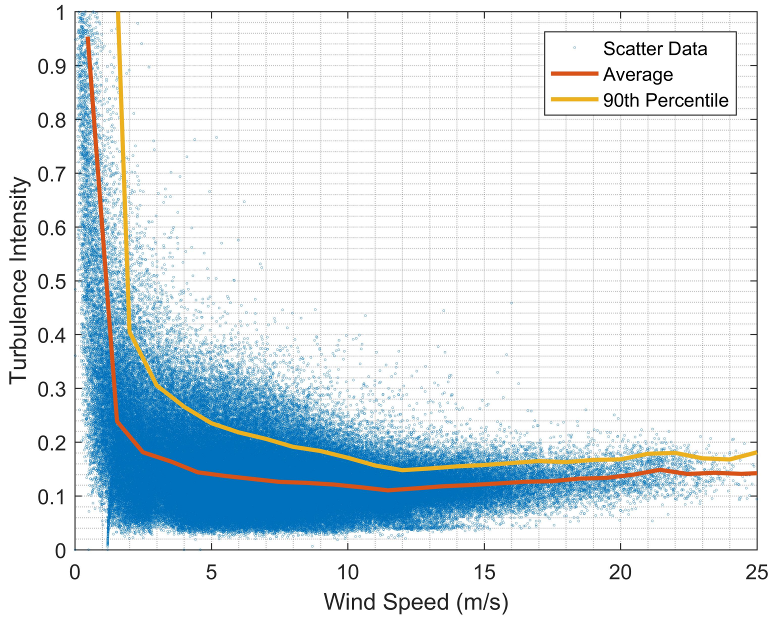


Figure 5.7: TI against wind speed, showing all the scatter data, as well as the average and the 90th percentile values.

5.2.2 Ambient and Nacelle Temperature

Figure 5.9 shows the temperature difference of the nacelle and the ambient temperature for all the turbines. WT15 is the one with the highest temperature difference and row A the one with the highest temperature difference. The difference for summer and winter seasons typically varies between 14 and 16 ^{o}C respectively. The temperature difference values during winter are higher compared to the ones during summer, with the standard deviation of the values also being higher. For both summer and winter time, WT15 has the highest temperature variation.

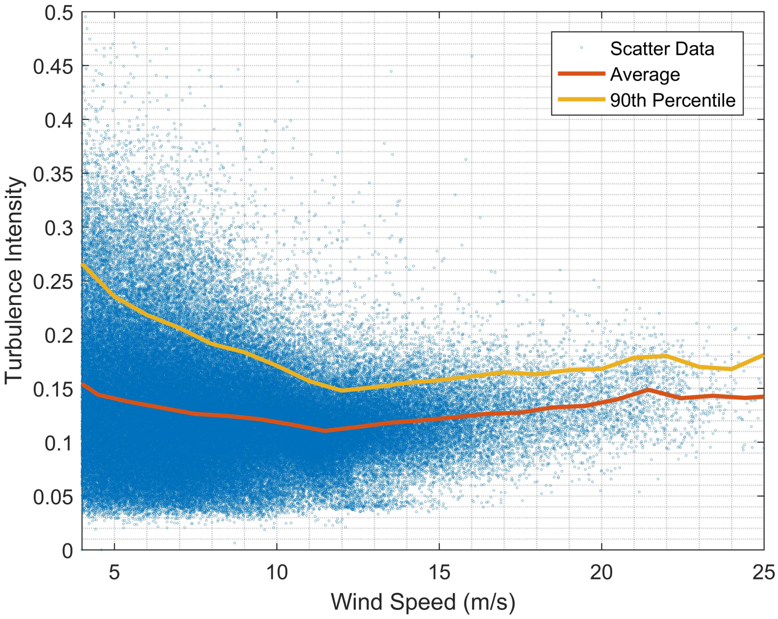


Figure 5.8: TI against wind speed for turbine operational wind speeds, showing all the scatter data, as well as the average and the 90th percentile values.

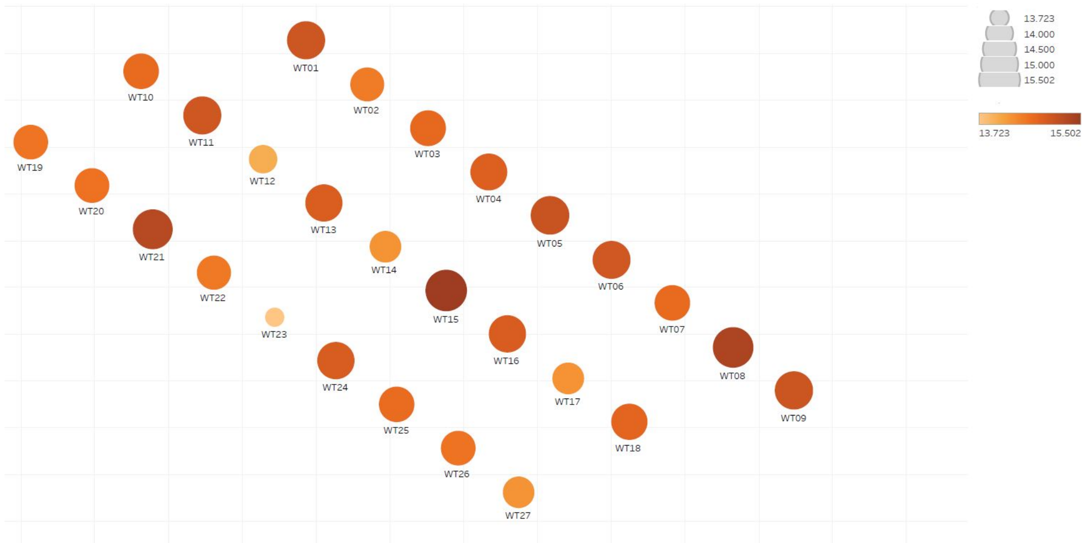
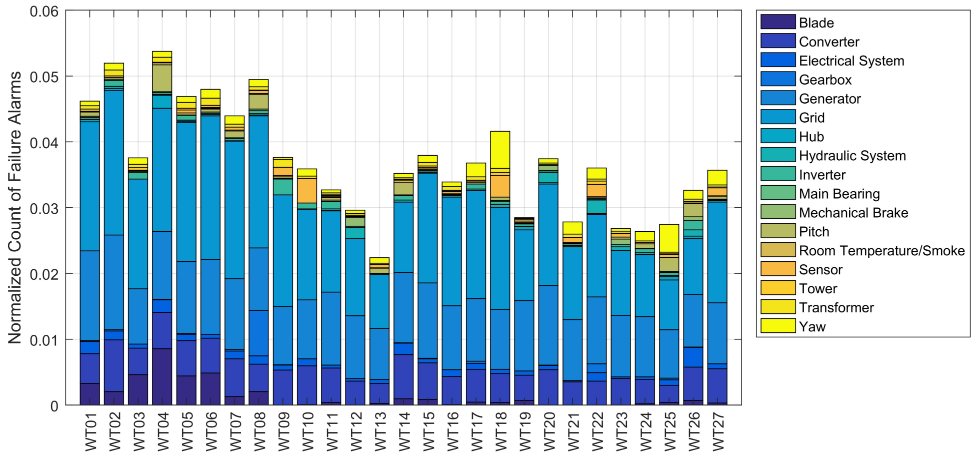


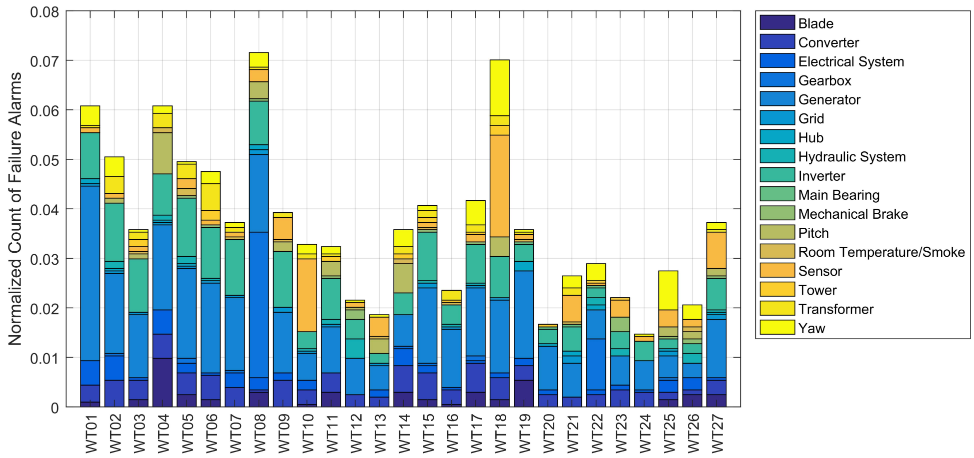
Figure 5.9: Temperature difference map of the turbines. Bubble and color size indicate the temperature difference values ranging from 13.7 to 15.5 ^{o}C.

5.2.3 Fault Alarms

Figure 5.10 provides a breakdown of the different subassemblies' fault alarms. These are shown for all the failure alarms and for the ones that have been active for more than 2 hours. This time period was chosen after expert advice; assuming that if it is a false alarm or one that can be remotely reset, this would have been done within two hours from the time that it initially appeared. In Figure 5.11, the data are also grouped by wind farm row. The grid issues account for a large portion of the alarms,



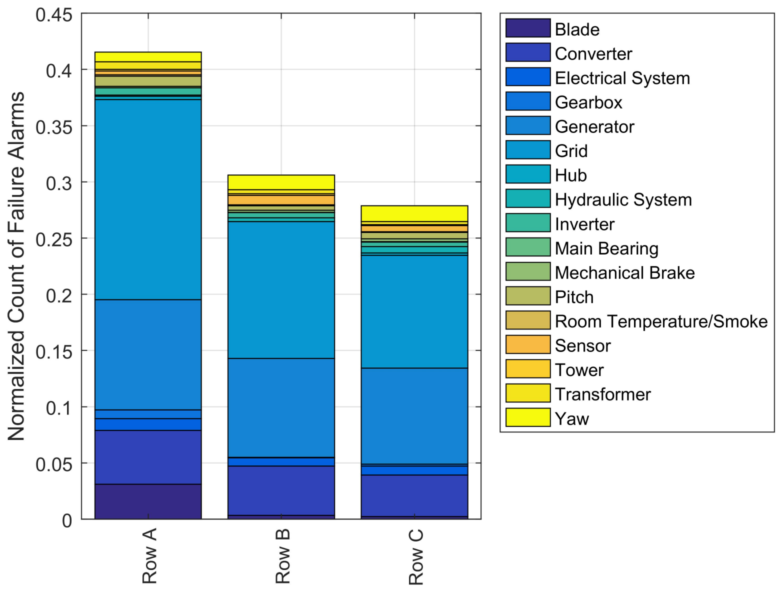
(a) All failure alarms.



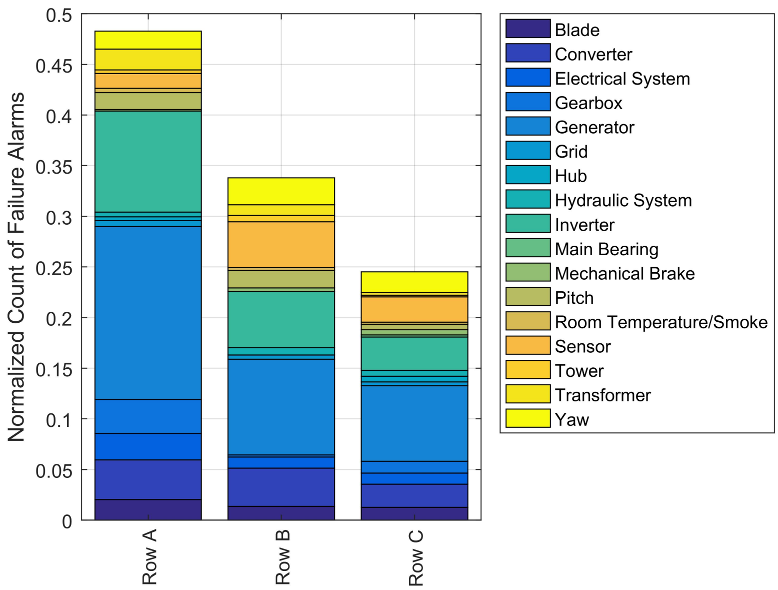
(b) Failure alarms longer than 2 hours in duration.

Figure 5.10: Count of failure alarm breakdown in different subassemblies for the different wind turbines.

for the ones less than 2 hours. For alarms longer than 2 hours, the grid issues are minimal and the generator issues are dominant. Moreover, it is easier to identify other



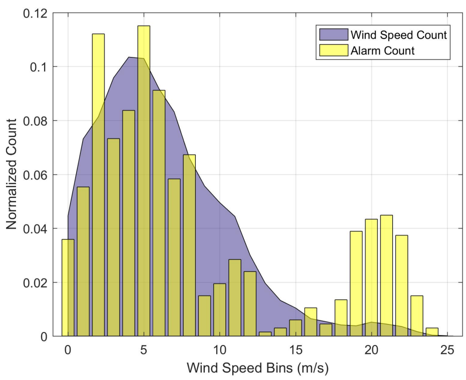
(a) All failure alarms.



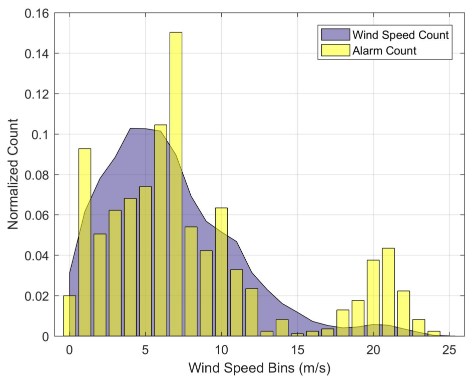
(b) Failure alarms longer than 2 hours in duration.

Figure 5.11: Count of failure alarm breakdown in different subassemblies for the different wind farm rows.

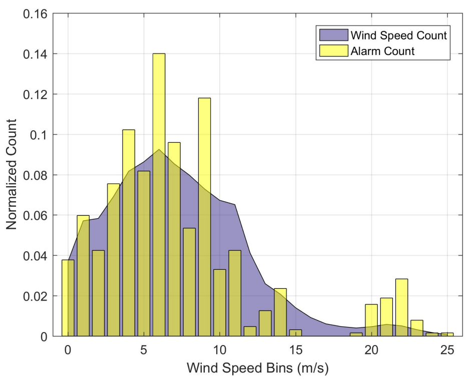
failures that turbines have faced, such as sensor, yaw, electrical system, gearbox and inverter. In both cases row's A turbines experience the majority of alarms, followed by rows B and C. Wind turbines 4, 2 and 8 encounter the majority of failure alarms for all the durations, whereas for longer than 2 hours, turbines 8 and 18 experience the majority of alarms. Wind turbines 13 and 24 have the lowest number of alarms in both instances. A correlation of alarms and wind speed is shown in Figure 5.12, where the



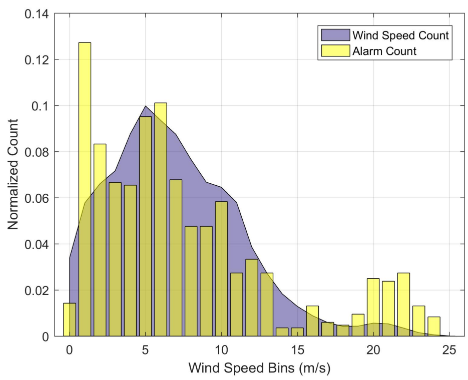
(a) WT13



(b) WT17



(c) WT19



(d) WT09

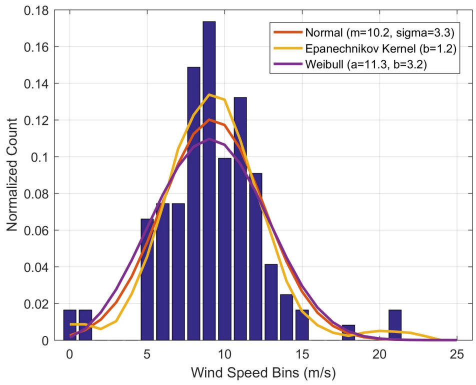
Figure 5.12: Normalized overlapped plots of the wind speed and failure alarm count for a selection of turbines.

normalized wind speed and alarm counts against the wind speed bins are plotted, for wind turbines 13, 17, 19 and 9. The figure shows two different normalized plots in one, in order to emphasize the effect of wind speeds on the failure alarms. The turbines selected are the ones with the highest TI and wind speed values for the 4 years of operation. Although the wind speed frequencies at the highest wind speeds are lower, the alarms are proportionally much more frequent, which indicates an initial correlation of the alarms and the wind speed.

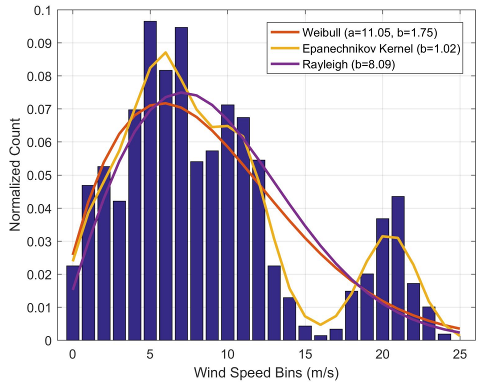
5.2.4 Wind Turbine Subassemblies

To further investigate the correlation between wind speed and TI with system alarms and failures, the different operational alarm profiles are plotted in Figures 5.13 and 5.14 against the equivalent average wind speed bins. These include temperature increases, generator, electrical systems, gearbox, transformer, communication systems and sensor faults. Electrical systems, transformer, communication systems and sensors experience a higher number of failures at lower wind speeds, whereas temperature, generator and gearbox failures, seem to occur more frequently at the higher wind speeds. Weibull, epanechnikov kernel, Rayleigh and normal distributions are fitted to the data and are plotted in the figures, along with their constants. Weibull and Rayleigh distributions were chosen because they are commonly used in reliability data analysis and can be picked up by researchers and practitioners. The kernels were needed in order to fit all the complexities of the data in a single distribution. The data shown in Figures 5.13 and 5.14 do not represent the relative occurrence of the wind speeds, as shown in Figure 5.12, but seek to show the actual expected failure alarm distribution for the analysed wind farm whilst operating under warranty within the first few years.

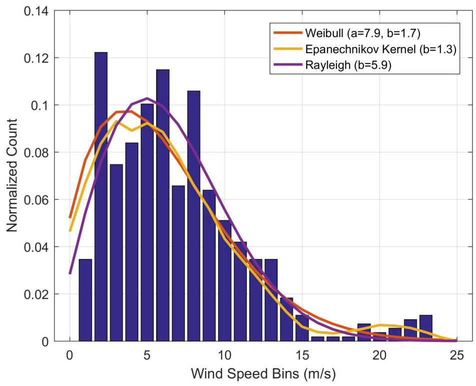
Gearboxes experience most of the failure alarms during the most frequent wind speed conditions. Generators experience two peaks of alarms at 4-6 m/s and at 19-22 m/s. Electrical systems encounter the majority of the failure alarms at low wind speeds and transformers at low and at the turbine's rated wind speeds. Sensors experience failure alarms at low wind speeds and communication systems have a spread of alarms from 4-18 m/s. Converters experience a peak during the higher wind frequencies and a smaller one during high wind speeds ranging from 18-23 m/s. Inverters have a failure alarm peak at 1 m/s and then almost a normal distribution from 3-12 m/s. Yaw systems have a spread of alarms through almost all the operational wind speeds of the turbines and pitch systems have most of them during the lower wind speeds, up to 8



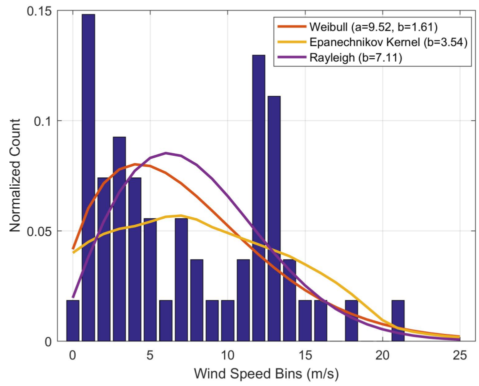
(a) Gearbox



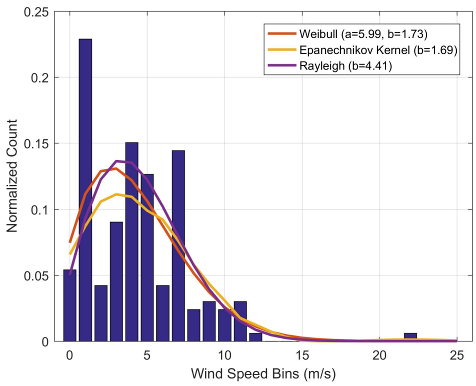
(b) Generator



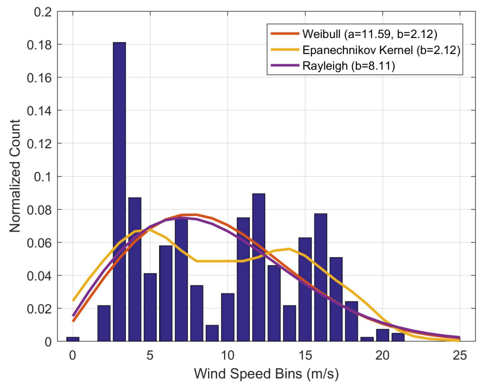
(c) Electrical Systems



(d) Transformer

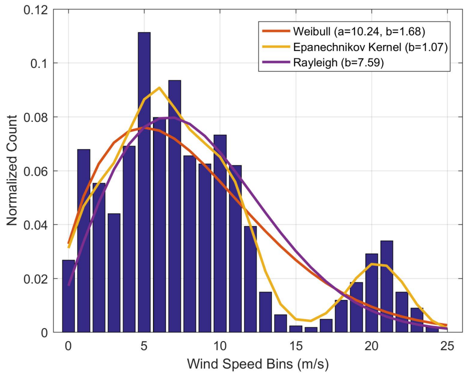


(e) Sensors

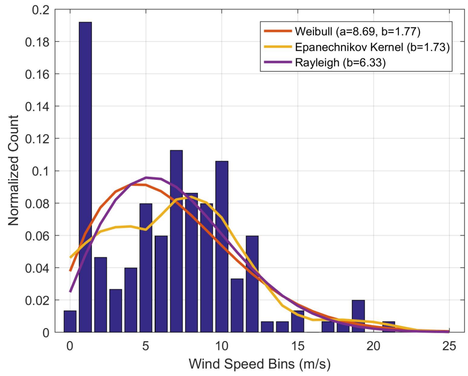


(f) Communication Systems

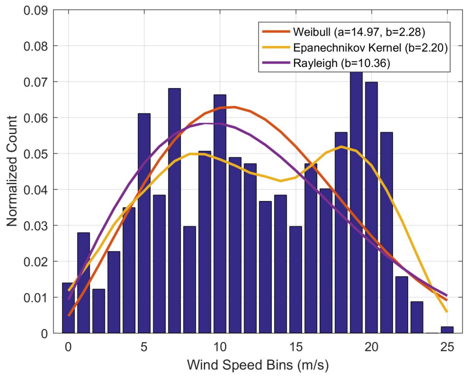
Figure 5.13: Wind turbine subassemblies' failure alarms against wind speed bins.



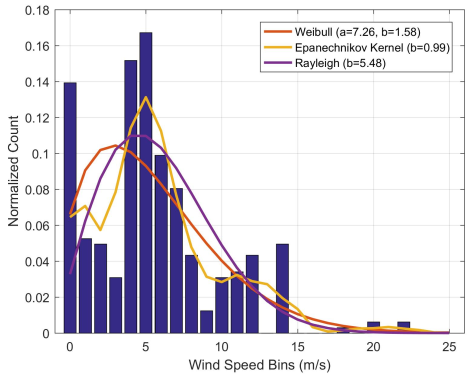
(a) Converter



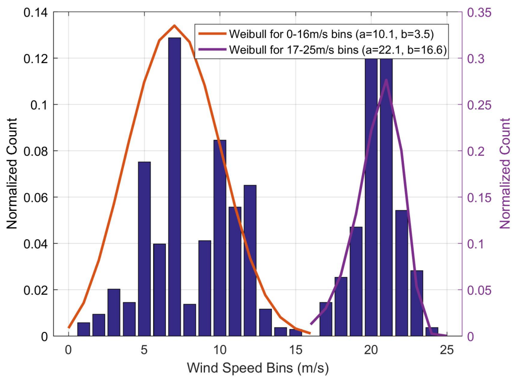
(b) Inverter



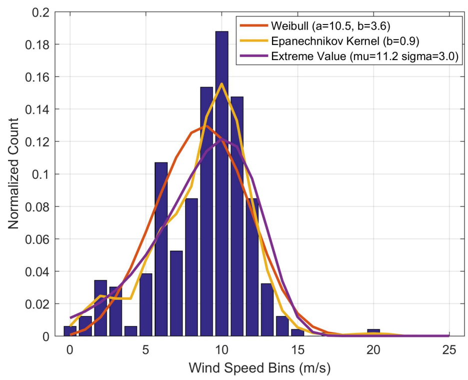
(c) Yaw



(d) Pitch



(e) Temperature

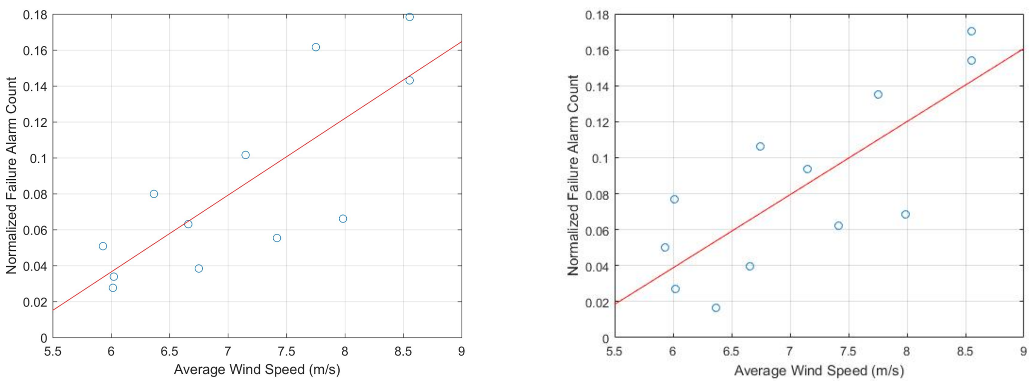


(f) Vibration

Figure 5.14: Wind turbine subassemblies' failure alarms against wind speed bins.

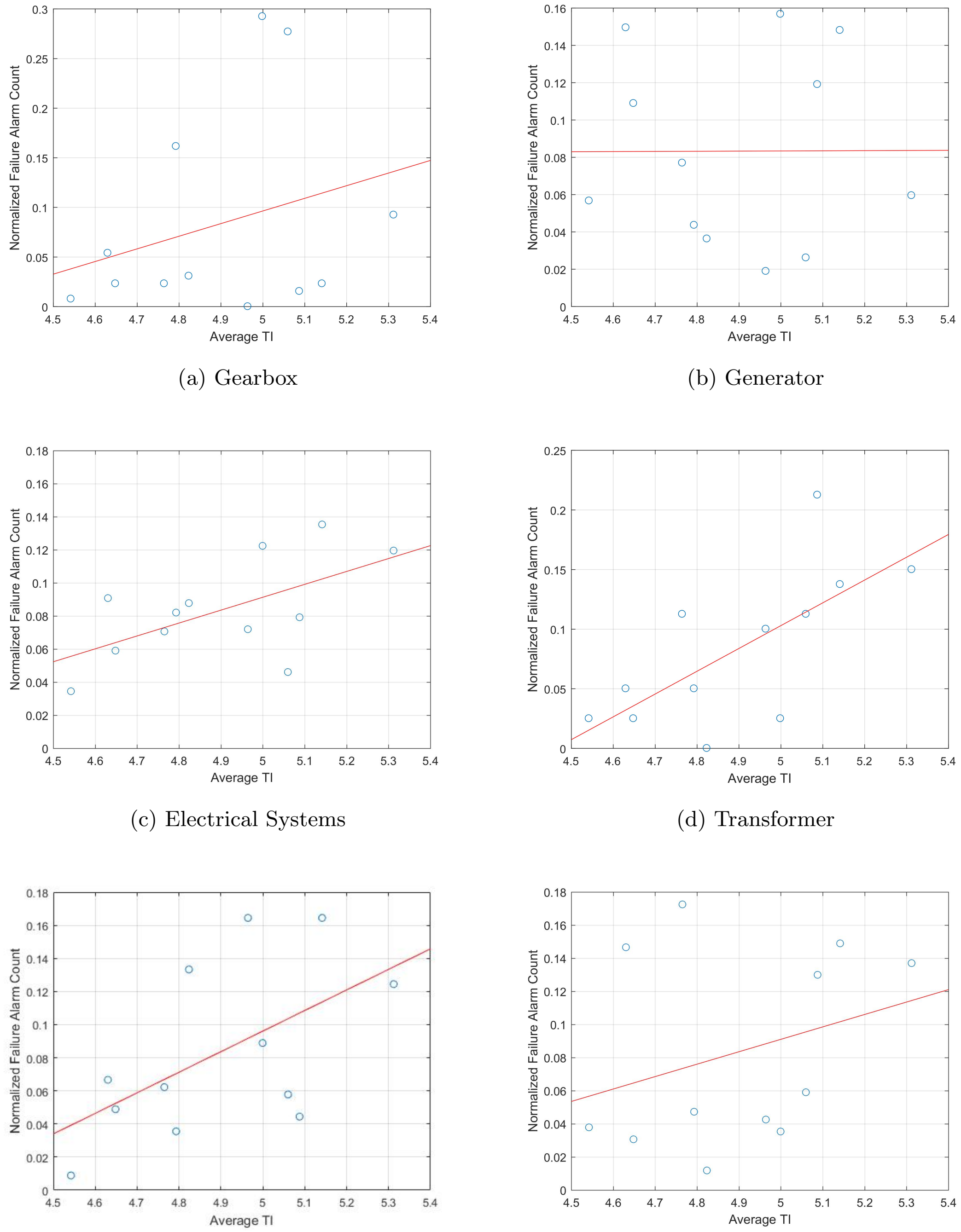
m/s. Temperature alarms have two peaks at 7 and 20-21 m/s, whereas vibration alarms have almost a normal distribution spread with a peak at 10 m/s.

The correlation of the yaw system, tower and blade vibrations against wind speed are shown in Figure 5.15. Figures 5.16 and 5.17 show the correlation of all the subassemblies against TI. The data are averaged for the different months of the year, against the average wind speed of that month over four years of operation; resulting in 12 points per graph represented the average of each month for these four years, attempting to also show the seasonality of the results. The coefficient of determination (R^2) for the different plots are shown in Tables 5.1 and 5.2. The R^2 values indicate the variability of the results and the strength of the correlations shown in Figures 5.15- 5.17. A good correlation between the wind speed and the yaw system and vibration failure alarms can be observed, with R^2 values greater than 0.6. Electrical systems, transformer and sensors have the highest R^2 scores and the best correlations with TI regarding the gradients, compared to the rest of the subassemblies. The R^2 values are relatively low though: 0.32-0.48. Other systems, such as the gearbox, communication systems, converter, inverter, pitch and temperature, display a gradient that indicates a possible correlation with TI, but their R^2 are much lower, showing greater variability of the results. Since the plots present the values on a monthly basis the variability in the R^2 values is attributable to the seasonality of the results.



(a) Yaw System (b) Tower and Blade Vibration

Figure 5.15: Correlation between monthly average wind speed with turbine failure alarms for different subassemblies.



(a) Gearbox

(b) Generator

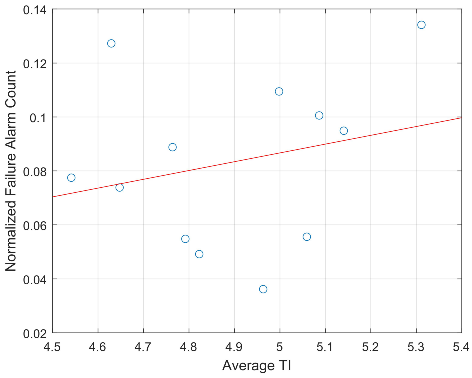
(c) Electrical Systems

(d) Transformer

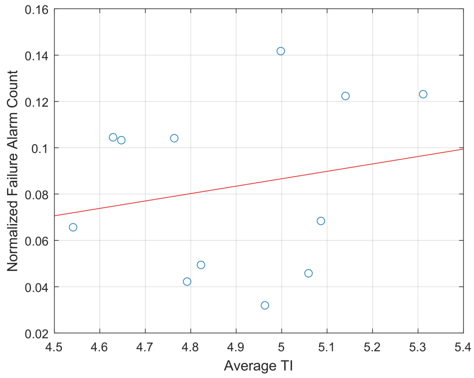
(e) Sensors

(f) Communication Systems

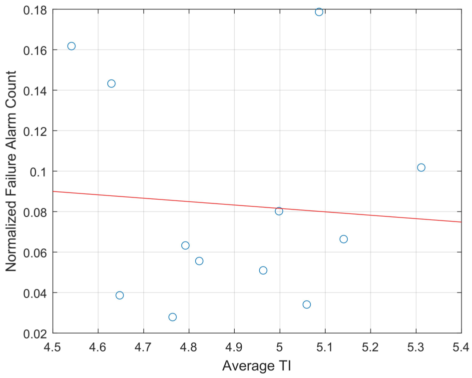
Figure 5.16: Correlation between monthly average TI with turbine failure alarms for different subassemblies.



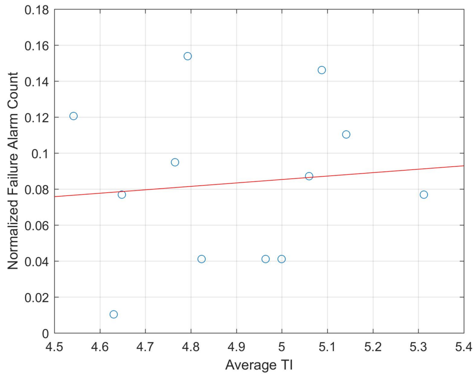
(a) Converter



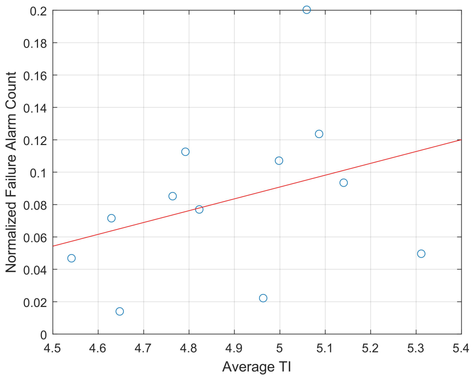
(b) Inverter



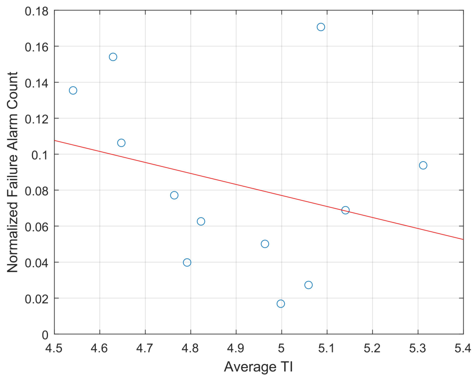
(c) Yaw



(d) Pitch



(e) Temperature



(f) Vibration

Figure 5.17: Correlation between monthly average TI with turbine failure alarms for different subassemblies.

Table 5.1: Coefficient of determination (R^2) for the correlation plots in Figure 5.15.

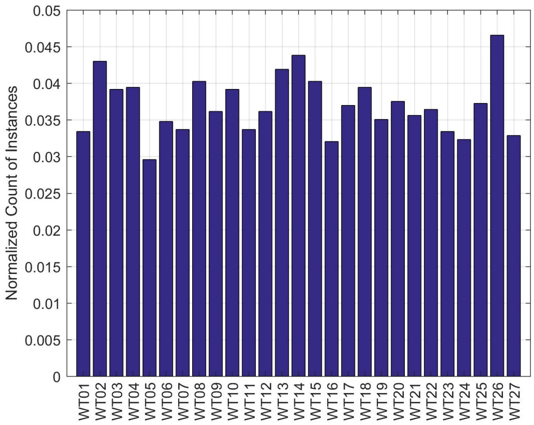
	R^2
Yaw System	0.6325
Vibrations	0.6137

Table 5.2: Coefficient of determination (R^2) for the correlation plots in Figures 5.16 and 5.17.

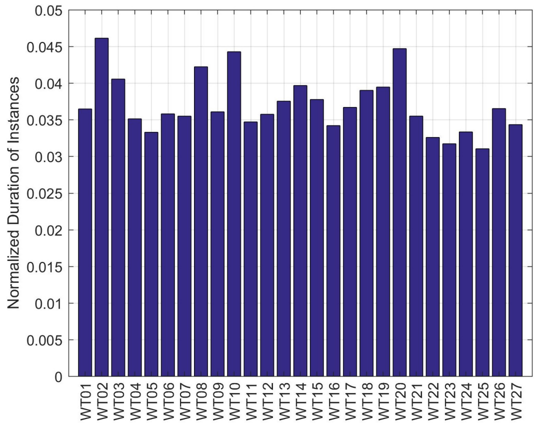
	R^2
Gearbox	0.0821
Generator	0.00001
Electrical Systems	0.356
Transformer	0.4854
Sensors	0.3177
Communication Systems	0.0914
Converter	0.0595
Inverter	0.0406
Yaw	0.0059
Pitch	0.01
Temperature	0.1148
Vibration	0.0831

5.2.5 Wind Turbine Downtime

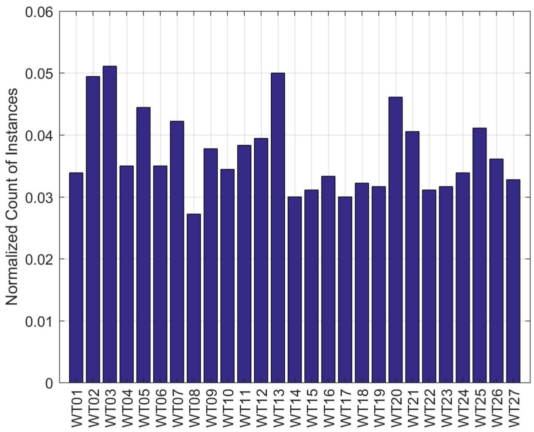
The turbines' contribution to downtime is also examined. Figure 5.18 shows the downtime due to turbine maintenance activities (for which the turbine was shut down), as well as the external stops due to identified issues. Both cases are shown as a count of instances and as duration values. The data shown are normalized to 1, in order to remove the actual downtime duration, which is classed as confidential. The presented, normalised downtime considers the same 4 years of operation as the rest of the data sets. For each subfigure, the data for the individual turbines and for the different downtime causes (maintenance downtime or external stop) are divided by the aggregated downtime duration for all the turbines and for the individual cause. Turbines 26, 14 and 2 are the ones visited more often and turbines 2, 20 and 10 are visited for the longest period of time. In terms of external stops, turbines 3, 2 and 12 have the most stops and turbines 3, 2, 11 and 13 have the longest total stop durations.



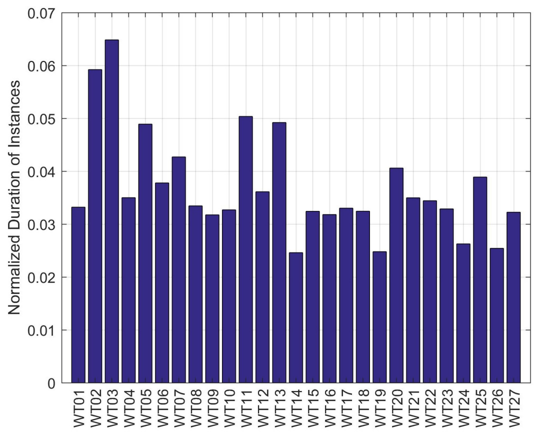
(a) Maintenance Downtime- Count of Instances



(b) Maintenance Downtime- Duration of Instances



(c) External Stop- Count of Instances



(d) External Stop- Duration of Instances

Figure 5.18: Turbine downtime

5.2.6 Reflection on Findings

The TI values of the different turbines, do not vary significantly in the farm, as shown in Figure 5.6, apart from WT13 that has the highest value. This can be explained by investigating the wind farm layout, Figure 2.16, and wind rose, Figure 5.4, which show that WT13 is the only one exposed to the wake effects of three turbines (WTs 21, 22 and 23) from the three most frequent wind directions. WT09 and WT19 have the lowest TI values, which can be similarly explained from the wind rose, as they are directly exposed to the wind from the majority of the wind directions. Moreover, the middle row has the highest overall TI, due to the design of the wind farm, turbines at row C are in front of the ones at row B, in the direction of the wind's highest frequency counts. Row A is not directly behind the flow coming from row B, which allows the turbines to avoid some of the wake effects generated, thus the lower TI values. This gives a better indication and highlights the importance of the preliminary wind farm design studies. If turbines are spaced correctly, they should experience similar TI values across the farm. Another indicator is the comparison of the pre-construction and operational wind roses, shown in Figure 5.5, that reveals the difference between the predicted and the measured wind speed and wind directions. This could be due to issues with the assumptions and extrapolation method used during the design stage, which shows the importance of careful and reliable wind measurements during the wind farm's planning period.

There are no significant ΔT temperature differences between the turbines in the wind farm, with only 1.8 ^{o}C difference between the highest and the lowest average temperatures. It is interesting to note that row A has the highest temperature difference, followed by rows B and C. By examining the seasonal variation, it can be observed that during winter the variation and temperature difference between the turbines is greater, possibly due to the higher levels of energy production. The average temperature difference values could be used during the operational phase of the wind farm, to inform the asset managers regarding any unexpected changes in the temperature, flagging individual turbines where appropriate.

The distribution of the alarms within the different turbines, shown in Figure 5.10 provides a lot of operational detail and is a useful summary for the gathered operational reliability data. Both subfigures are very informative, as they show the variation of alarms experienced at an offshore site, as well as the alarms that persist for longer than 2 hours and cannot be remotely reset. The majority of grid related issues are usually dealt remotely. Similarly, blade related alarms can be remotely reset. Any electromechanical issues, such as on the generator, converter, inverter, sensor, gearbox, sensors, yaw systems, cannot be usually reset and need to be further investigated. Turbines 4, 2 and 8 have the most failure alarms for all the durations and turbines

8 and 18 have the most failure alarms for duration longer than 2 hours. This is due to individual failures of the turbines in different subsystems. WT08 has the highest portion of gearbox failures, whereas WT18 suffered from sensor failures. This shows the variance of the different assets of the same manufacturer and in the same wind farm, in terms of failure rates expected. It is also interesting to note that WT13, the one with the highest TI, has one of the lowest number of failure alarms, which could be due to the lowest wind speed and the lowest overall generated power. The overall failure alarms for the different rows, Figure 5.11, indicate that row A experiences the most failure alarms, followed by rows B and C.

By normalizing the wind speed counts against the turbine failure alarm counts and plotting them at the different wind speed 302 bins for the most vulnerable turbines, shown in Figure 5.12, it can be seen that turbine alarms are highly affected by the wind speed. At low wind speeds, up to 14 m/s, the frequency bins of the alarms follow the ones of the wind speed. At wind speeds >18 m/s there is a significant increase of the alarm frequency, which is not the case for the wind frequency.

Considering the subsystems, shown in Figures 5.13 and 5.14, the behaviour of the temperature warnings, generator and gearbox seem to strongly follow the wind speeds, with increased failure rates at the highest temperatures. On the contrary, electrical systems, transformer, communication systems and sensors, seem to follow the averaged TI curve, shown in Figure 5.7. Some of the above correlations are further investigated in Figures 5.15, 5.16 and 5.17, where a good correlation for yaw system and tower/blade vibrations failure alarms with wind speed can be observed. Similarly for electrical systems, transformer and sensors failure alarms, a correlation with TI can be seen. For the subassemblies where the R^2 values are lower, Figures 5.13 and 5.14 might be more informative and a further analysis might be needed in the future with larger datasets.

Turbine downtime is a useful indication of system failure consequences. For the different rows, the majority of the downtime is caused by the turbine at the edges of the row that are more exposed to the environmental conditions, whereas the ones inside the wind farm, seem to experience less downtime. This indicates that the more exposed turbines are more likely to experience faults compared to the more sheltered ones. This can also be confirmed from the fact that row A turbines experience longer maintenance intervals, compared to the other two rows.

The presented results are informative both for the development as well as for the operational stages of a wind farm. In an O&M setting, the forecasted wind speeds could be used as an indicator and predictor on the failure alarms that will be expected. This can help during the operational planning, for the teams that need to be in standby during the different weather conditions, which can result in lower downtime of the assets and thus higher power output, increasing the revenue stream of the operator. In order to verify those results, they would need to be trialled in an operational setting or considered as part of an operational planning tool, to investigate this viability and cost effectiveness of such a solution. The development of an alarm prediction model could allow the efficient planning of operational personnel and technicians. The presented results are so far based on 4 years of operation and thus can be considered to represent the bottom of the bathtub curve, as they include years 2-5 of the assets operations, which could be informative for the behaviour of the different subsystems. During the design phase, the presented approach and results can be used as inputs for future design of offshore wind farms, which can result up to 5% higher power output in individual turbines if they are spaced out correctly, increasing again the potential revenue of the wind farm owner. The system failures will need to be considered separately, as their behaviour to wind speed and TI varies significantly.

5.3 Failure Root Cause Analysis

Another useful analysis that can be performed by the combination of the different data sources is failure root cause analysis.

5.3.1 Framework

Building on existing knowledge from [12], a proposed RCA framework has been created as shown in Figure 5.19. The framework is based on the data integration methodology explained in Section 3.2. As it can be seen, once the failure rates have been calculated for all the (sub)assemblies/ (sub)systems/ components, then the failure location is identified and failure modes are created from an FMEA. Then the root causes of the failure need to be identified and in order to understand why this failure happened, an RCA needs to be implemented. All the data sources will be available for the specific timestamp that the failure occured, which can show the effect of the failure on other components, or the root cause of the failure from the sensor readings, the maintenance logs or previous alarms. This process can be repeated for different failure components and it can start from any given point on the diagram. For example, an unexpected behaviour can be noticed from a sensor reading at the SCADA, CM, SHM systems,

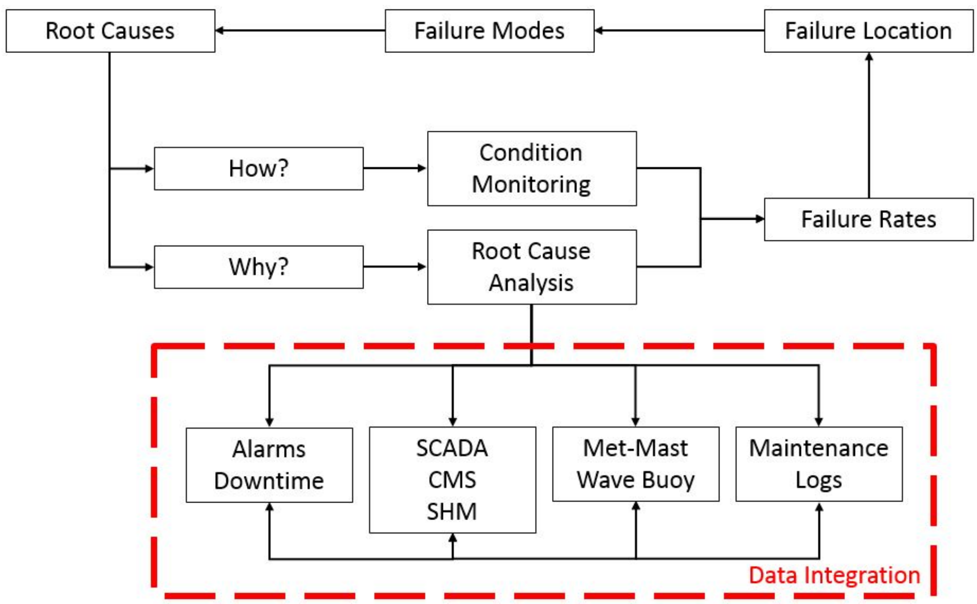


Figure 5.19: Proposed RCA framework.

from an alarm and then traced back through all the other systems in order to identify the root cause.

5.3.2 Yaw Gear Failure

An example of a FRCA was tested on historical data as shown in Figure 5.20 (axes are omitted for sensitivity purposes), which shows different sensor readings, alarms and maintenance logs of an underperforming turbine. The analysed turbine is highlighted and it is compared with the rest of the wind turbines in the farm, as shown in the background. Graphs A and B at the top, show the power factor and blade pitch tags respectively and at the bottom, the alarms and maintenance log occurrence (C) and the nacelle yaw tag (D) are shown. The unit measurements for Figure 5.20 are degrees for the yaw and pitch tags, whereas the power factor is dimensionless. The x-axis indicates the time of the measurements in days. The SCADA tags presented are formed from 10-min average values. The RCA approach outlined above was followed in order to determine the failure root causes.

The turbine's underperformance could be initially observed from any power related SCADA tag, which in this case is the power factor. The faulted turbine is highlighted in Figure 5.20 and since all the turbines are exposed to similar environmental conditions, they are expected to perform with a similar trend and produce similar power factors. Some deviations are expected in the power outputs as explained in Section 5.2 and shown in Figure 5.6, but the overall trends should be similar. As an effect, an outlier can be easily spotted. The most possible failure modes in this case could be on the yaw or the pitch system or a sensor reading issue. The integrated DB system allows the operator/owner to rapidly check throughout the different sensors and system in order to identify the failure root cause. By looking at graphs B and D, it can be seen that these tags for the highlighted turbine are not performing similarly with the rest of the turbines in the farm. Thus, there is a fault in both or in one of these sub-systems. This also reduces the possibility of a sensor error, since it would require the simultaneous failure of several sensors. Then, by following the proposed RCA approach, the failure mode needs to be identified, which can be done through an FMEA. RELIAWIND guidelines can be followed and potential failure modes can be investigated as shown in Table 5.4. Then,

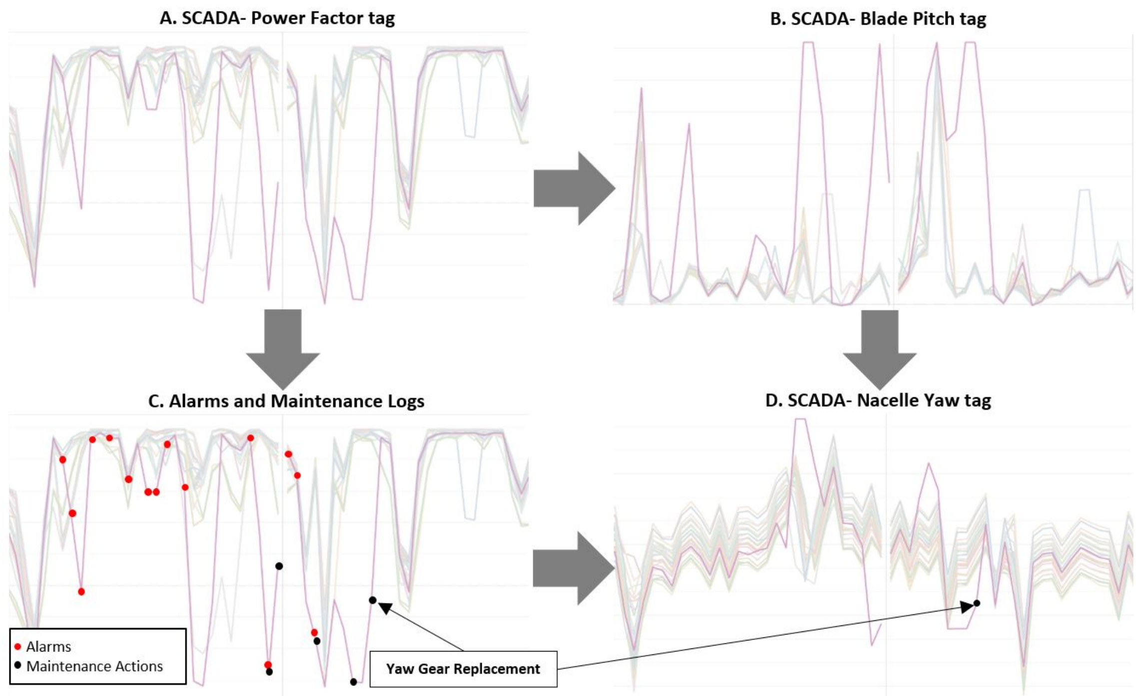


Figure 5.20: Example of FRCA of an underperforming turbine against the rest of the wind farm, showing: A. The power factor(dimensionless), B. Blade pitch (degrees), C. Alarms and maintenance logs on the power factor, D. Nacelle yaw (degrees). The x-axis indicates the time of the measurements averaged in days. The highlighted purple line is the turbine under investigation and background ones are the rest of the turbines in the farm.

by investigating the alarms, it can be seen that several warnings have been triggered that include only yaw errors, such as: "Yaw converter error", "Too many yaw converter errors", "Too many yaw converter warnings". Because of this, visits to the turbine were planned, at the first weather window. It was attempted to restart the turbine in the first 3 visits. At the forth visit, the wind vane was replaced (Failure Mode 2- Table 5.4) and finally the solution was found at the fifth visit to the turbine, where the yaw gear was replaced, which is Failure Mode 1. As it can be seen in (D), after the final repair, the yaw error was resolved and the turbine yaw problem was fixed.

Table 5.3: FMEA example of the 5 most common failure modes for pitch and yaw systems[70].

SubAssembly	Failure Mode 1	Failure Mode 2	Failure Mode 3	Failure Mode 4	Failure Mode 5
Pitch	Electrical	Battery Failure	Pitch motor failure	Pitch motor converter failure	Pitch bearing failure
	Hydraulic	Internal leakage of proportional valve	Internal leakage of solenoid valve	Hydraulic cylinder leakage	Position sensor/ no signal
Yaw System	Yaw gearbox and pinion	Degraded wind direction signal	Degraded guiding element function	Degraded hydraulic cylinder function	Brake operation valve failure

5.3.3 Reflection on Findings

The above FRCA example shows the benefit and use of such an approach to the operators in order to have a holistic view on their assets. If this process had been followed from the beginning, the fault may have been identified on the first turbine visit and some of the associated cost would have been saved. These results indicate that following a FRCA would have required less visits to the turbine and reduced associated costs. The estimated revenue loss from the underperforming turbine was £12,000 of lost production (for a 2.3MW turbine), which could have been avoided if the suggested approach had been followed, offering significant potential for the lifetime of a larger offshore wind farm.

5.4 Gearbox Failure

This section presents the results from the detection, diagnosis and machine models of the investigated gearbox failure, as explained in Section 3.4.1. The case study has been developed after the gearbox has been replaced and uses the learning from it to generate the different models.

5.4.1 Data Selection

The data selection process implemented in this study is summarized in Figure 4.2. Initial turbine degradation signs can be observed at the power curve, when comparing the binned power curve values for 5 months before the gearbox replacement and 2 years prior to that and similar conclusions can be derived from the active power against the rotor velocity data. The data are presented in detail below from the SCADA and CMS systems. To maintain confidentiality, the data presented are normalized and some scales have been adjusted. Each figure was normalized individually, so no correlation between figures is possible.

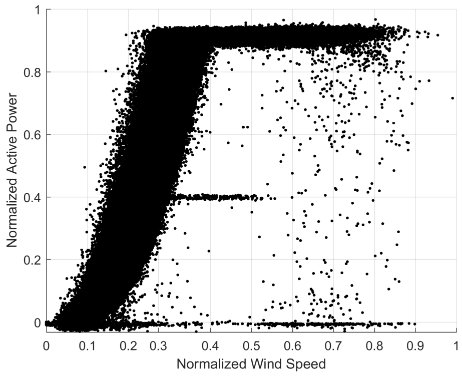
5.4.2 Failure Detection and Diagnosis

The results for the failure detection and diagnosis from SCADA and CMS data are shown below.

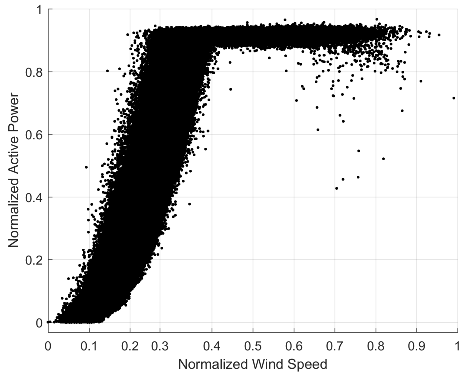
5.4.2.1 SCADA

The original and filtered power curves are shown in Figures 5.21a and 5.21b respectively. The power curve information was filtered further to reflect a period of 5 months prior to the gearbox replacement and the same period 2 years before, as shown in Figure 5.21c. As can be seen, there is no clear underperformance of the turbine at this stage, just a slight deviation to the right for the 2017 data.

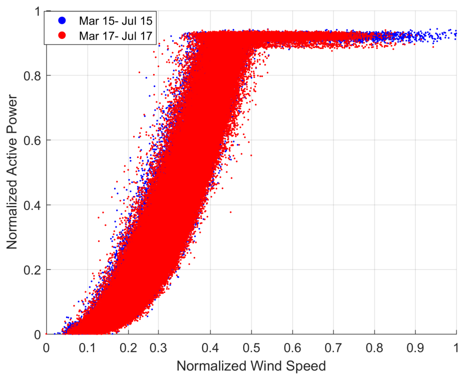
The binned power curve is visualized in Figure 5.21d. By following this method it is more evident that the turbine is underperforming a few months before replacement, for the wind speeds from 0.3 to 0.55. It was chosen to compare the power curve characteristics of the turbine, with a previous healthy state of the same or a



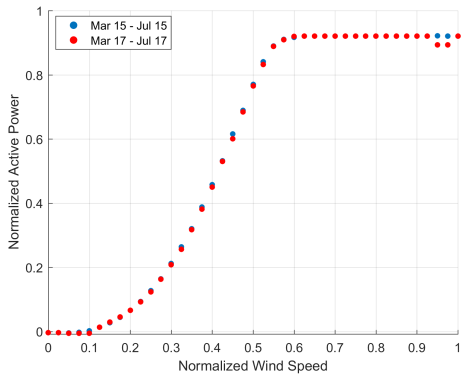
(a) Normalized 30-sec average power curve for 2 years and 9 months.



(b) Normalized filtered 30-sec average power curve for 2 years and 9 months.



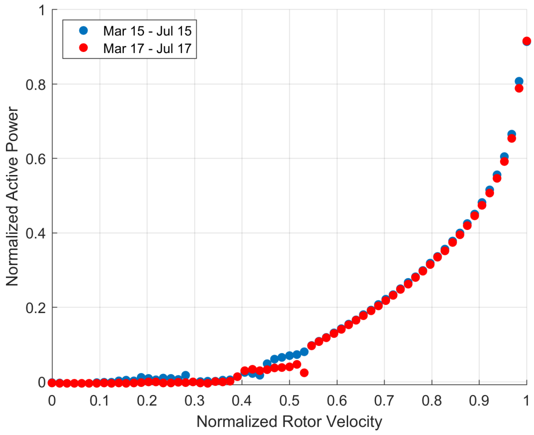
(c) Normalized 30-sec average power curve for March- July 2015 and 2017.



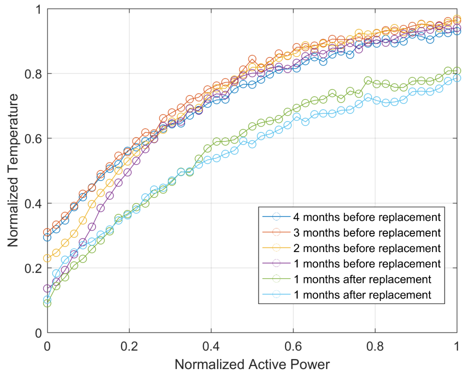
(d) Normalized binned power curve for March- July 2015 and 2017

Figure 5.21: Power curve analysis.

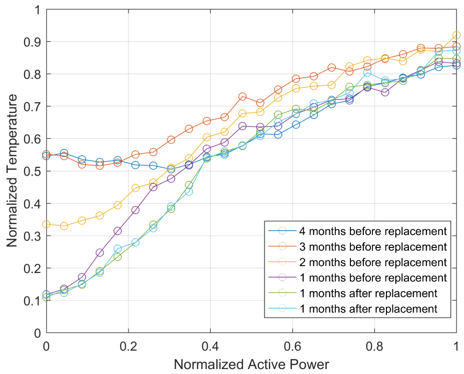
neighbouring turbine, as the provided OEM's power curves might not be representative. Similarly, active power was binned and plotted against rotor velocity data, Figure 5.22a, leading to the same conclusion, of turbine underperformance for the period in 2017. Moreover, as the rotor is directly connected to the gearbox's planetary stage, an initial assumption can be made for the failure location. The different gearbox temperatures have been compared at different time periods, before and after the gearbox replacement, as seen in Figures 5.22b- 5.22f. The oil temperature against the square of the rotor velocity is presented in Figure 5.22d. All the temperature related figures indicate that there is a significant temperature increase, caused by the faulty gearbox. In Figures



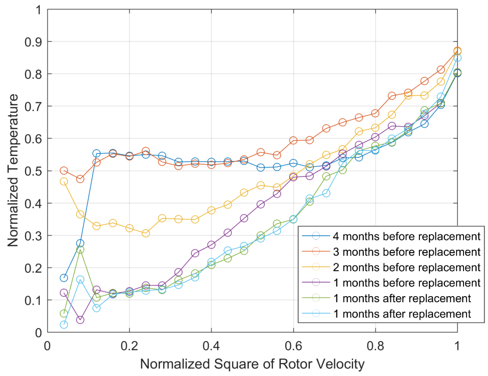
(a) Normalized active power against rotor velocity.



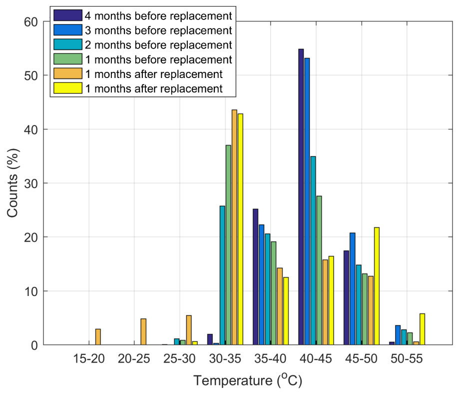
(b) Normalized gearbox high speed stage temperature against active power.



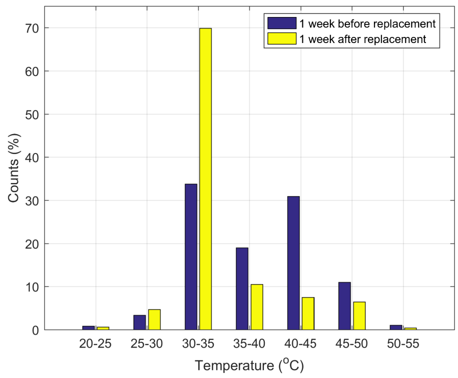
(c) Normalized gearbox oil temperature against active power.



(d) Normalized gearbox oil temperature against square of rotor velocity.



(e) Gear oil temperature bins for different months.



(f) Gear oil temperature bins.

Figure 5.22: SCADA parameters used.

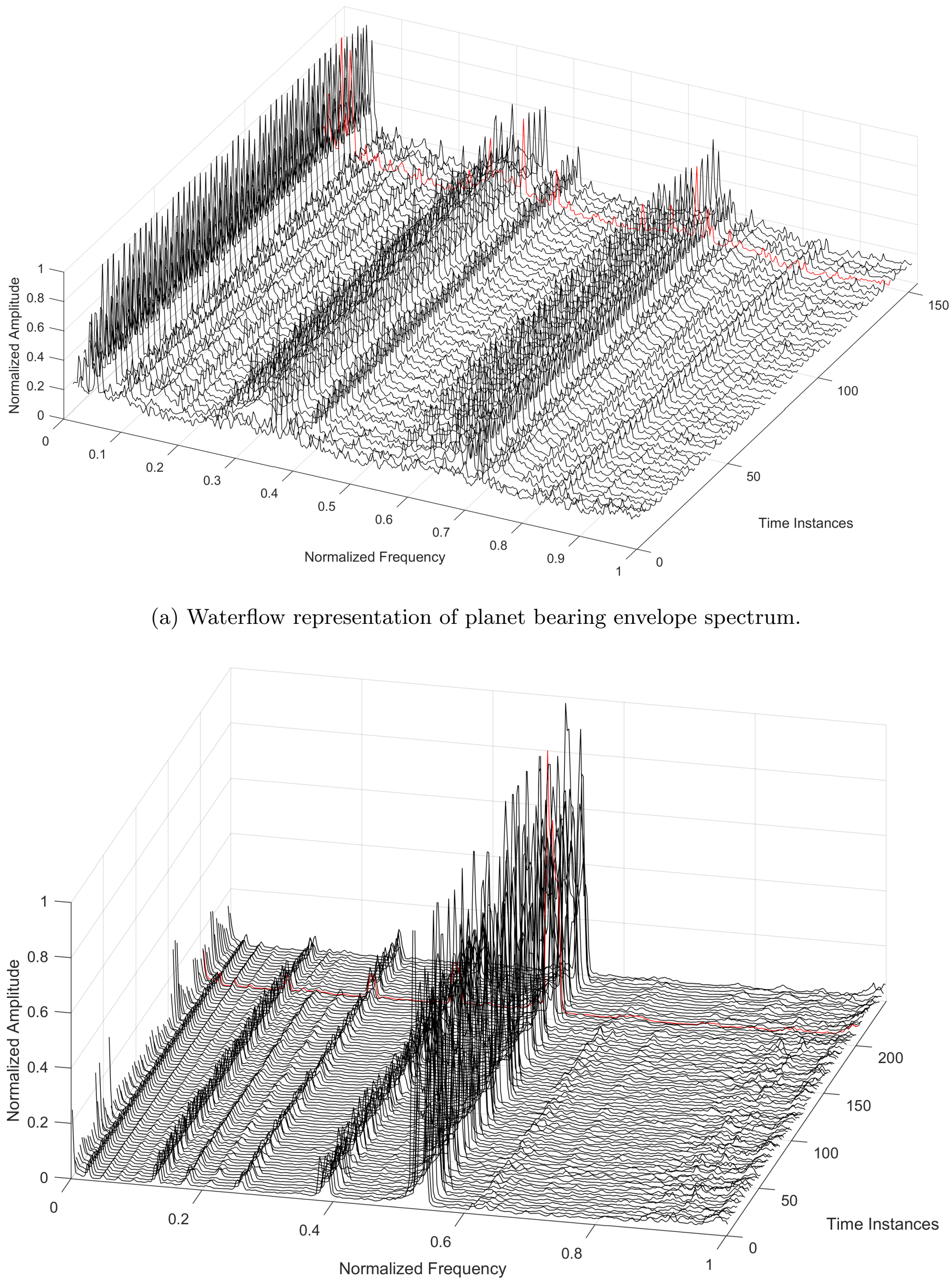
5.22c and 5.22d, the temperature increase is seen at the lowest active power and rotor velocity values, which makes it hard for an alarm system to capture them, but it is still not totally clear, when comparing all the previous and after replacement curves. In the case of the HSS temperature, Figure 5.22b, the temperature increase is more clear throughout the whole power range, showing a 3-4 ^{o}C temperature difference, but keeping it still within the SCADA alarm limits, as no such alarms have been triggered. Figure 5.22e shows a bar plot with the different temperature bins. Although a difference can be noticed for the before and after replacement values, there is a temperature increase at high temperatures, for 2 months after replacement, which is due to the highest energy production during that period. Figure 5.22f makes this difference more apparent, as the environmental conditions were very similar during those instances. Although there is an evident temperature increase at the different gearbox temperature data shown and a reduction in the rotor velocity, the exact location of the failure cannot be precisely identified by only investigating the SCADA data. For this particular case, relying on temperature monitoring is insufficient for early fault detection.

5.4.2.2 CMS

Figures 5.23a and 5.23b show a three dimensional representation of the envelope and FFT values at different time instances. As it can be seen, no obvious changes in the component's frequency response and no visible, critical sidebands are building up.

By examining the signal from the cepstrum analysis, a more pronounced increase in amplitudes can be observed. The filtered rms signal is shown in Figure 5.24, where an increasing trend can be seen after March 2017, which is a sign of degradation.

A constant increase in the particle counting can be noticed in Figure 5.25, by comparing the slopes after and before the replacement. This increase is most indicative when the particle counting against the cumulative energy generation of the turbine is examined. This could provide an early indication of the fault, revealing that material breakout is present.



(a) Waterflow representation of planet bearing envelope spectrum.

(b) Waterflow representation of planet bearing FFT spectrum .

Figure 5.23: Envelope and FFT waterflow representations from the CMS (annotated line represents the gearbox replacement date).

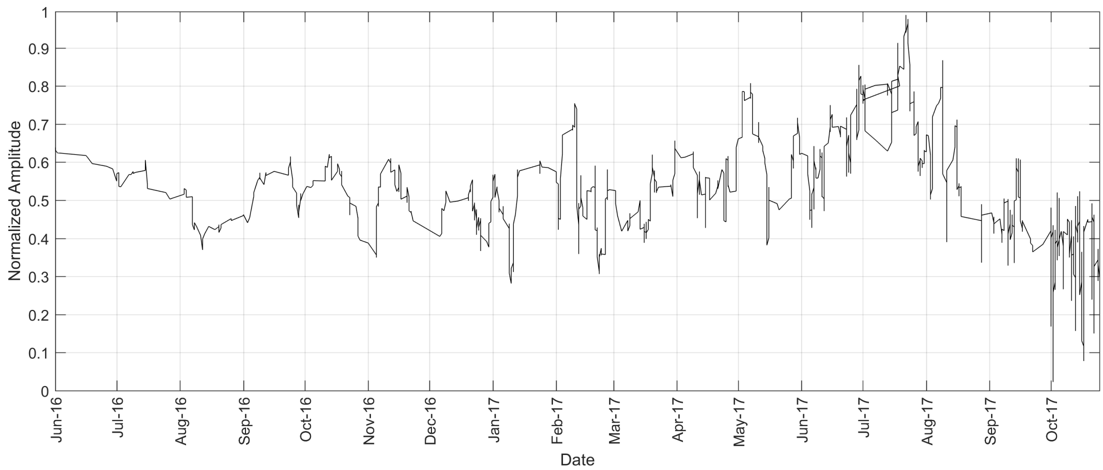


Figure 5.24: Normalized cepstrum rms of the planet bearing against the different dates.

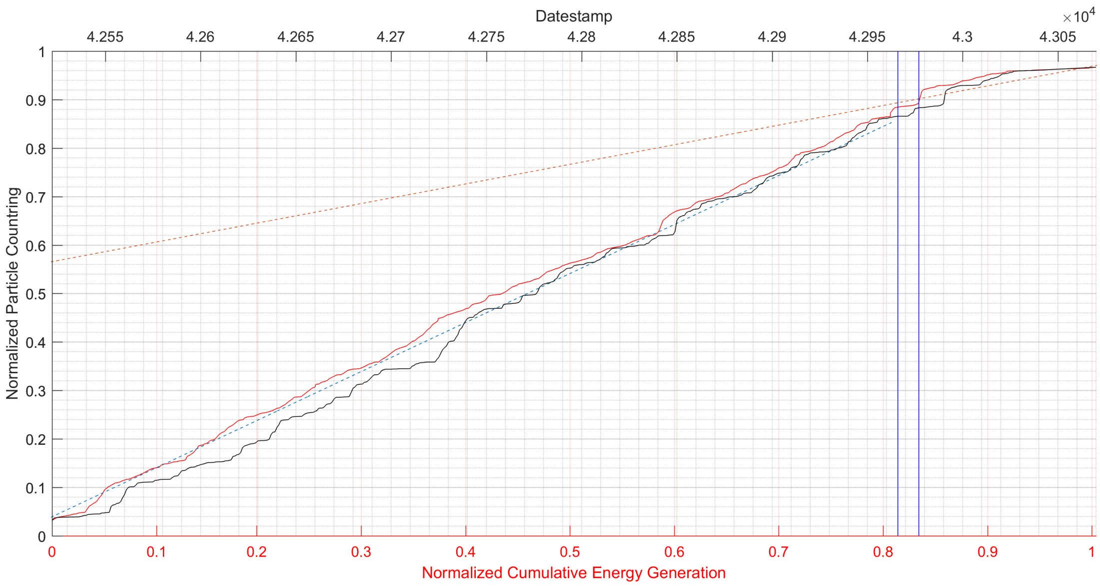


Figure 5.25: Normalized particle counting against date (top x-axis and black line) and normalized cumulative energy generation (bottom x-axis and red line). The vertical blue line indicates the replacement interval. The dotted blue and orange lines indicate the slopes of the particle counting against energy generation lines before and after replacement respectively.

5.4.3 Machine Learning Algorithm Selection and Implementation

By following the decision diagram in Figures 4.3 and 4.4, the machine learning algorithms were selected. The training data were randomly selected for the different categories on a 75/ 25% of training/ test data sets.

5.4.3.1 SCADA

The SCADA data are generated in timeseries. They can be used to predict a category, which in this case would be the healthy and faulty state of the turbine. Since the data for the two states have been collected and the time periods are known, the data can be labelled. As discussed in Sections 3.4 and 4.1, a supervised algorithm can be selected in this case. As the data cannot be easily separated by a single line, as shown in Figure 5.26 and they do not have multiple categorical decisions, they are categorized as highly non-linear. This feature could be hard to identify in this case. Therefore, both an SVM and kNN classifiers are tested. The data used for the algorithms are 10-min average active power, wind speed, rotor velocity, HSS gearbox temperature, gear oil temperature and generator temperature.

For the SVM algorithm the data would need to be separated with a hyperplane. Due to the complex nature of the data, there is no simple hyperplane that can be used as a separating criterion. Consequently, different kernel options were tested in order to optimize the algorithm, including linear, Gaussian, cubic and quadratic. The most precise and efficient method was the Gaussian kernel, based on the algorithms accuracy. After the hyperparameter tuning process, an optimal box constraint of 1 and a scale of 0.26 were selected. A more detailed performance of the classifier can be seen in the confusion matrix, Figure 5.27. The top row of the matrix shows the TP percentage in the population, the FP percentage in the population and the TP over the FP rate for the warning state; the middle row shows the FN percentage of the population, the TN percentage of the population and the FN over the TN rate of the healthy state; the bottom row shows the positive predictive value over the false discovery rate for the warning and healthy states, and the ACC of the model, defined by Eq. 3.19. The accuracy of the algorithm was 95.12%, the F_1 score 96.30%, Eq. 3.20, with a TP rate

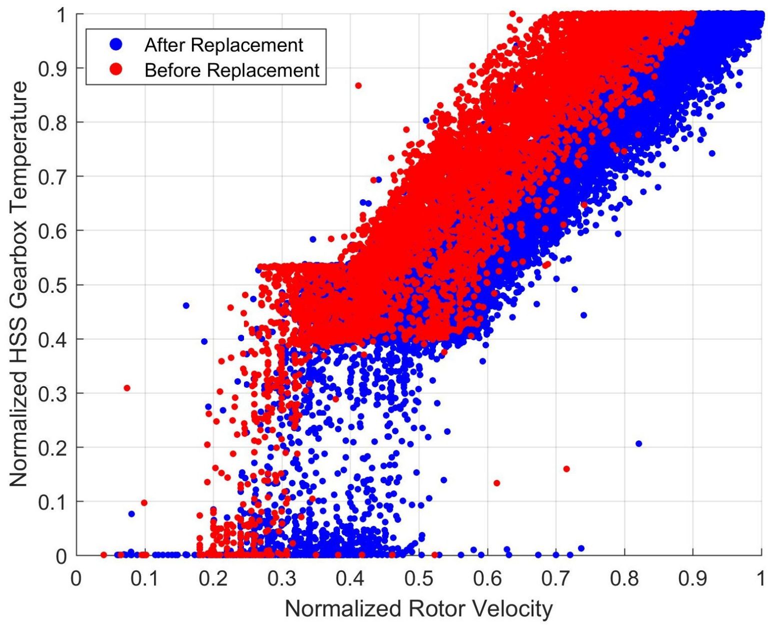


Figure 5.26: Normalised gearbox high-speed-shaft temperature against rotor velocity.

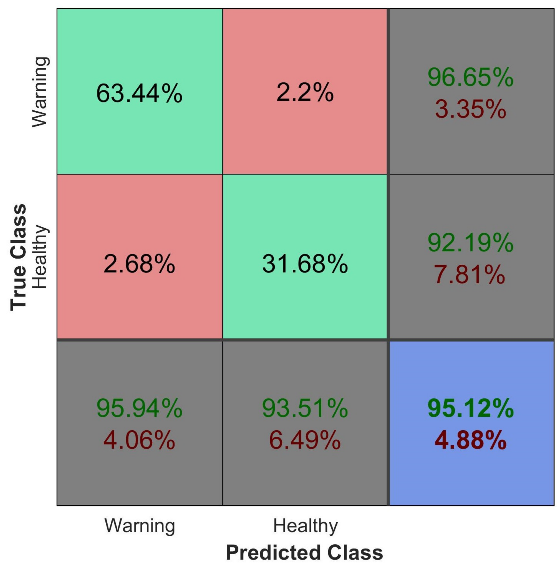


Figure 5.27: Confusion matrix for the SVM algorithm. The grey boxes are showing the corresponding TP, TN, FP & FN rates and blue box the algorithm's accuracy. Green and red boxes show the percentage of correctly and incorrectly classified data for the healthy and warning states.

of 96.65% for the warning and 92.19% for the healthy state and a positive predictive value of 95.94% for the warning and 93.51% for the healthy state.

For the kNN algorithm, difference distance metrics were tested, including Euclidean, Chebyshev, cosine, Minkowski and Mahalanobis. The latter proved to give more accurate results. After the hyperparameter tuning an equal distance and 10 neighbours were chosen as the optimal number. A more detailed performance of the classifier can be seen in the confusion matrix, Figure 5.28. The accuracy of the algorithm was 94.34%, the F_1 score 95.67%, with a true positive rate of 95.20% for the warning and 92.69% for the healthy state and a positive predictive value of 96.13% for the warning and 90.99% for the healthy state.

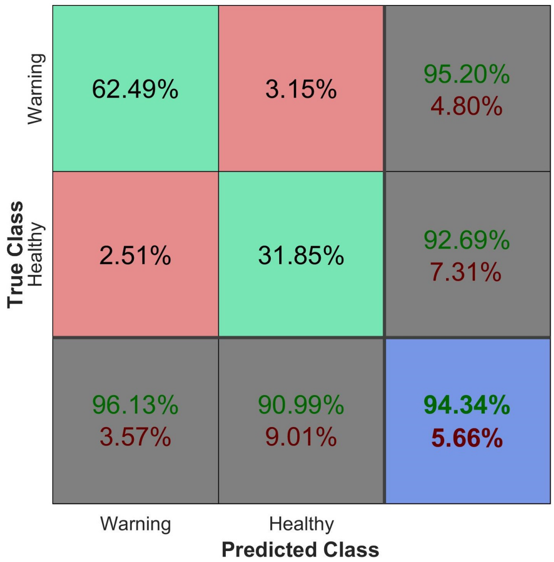


Figure 5.28: Confusion matrix for the kNN algorithm. The grey boxes are showing the corresponding TP, TN, FP & FN rates and blue box the algorithm's accuracy. Green and red boxes show the percentage of correctly and incorrectly classified data for the healthy and warning states.

Both algorithms produced >94% results' accuracy, with the SVM having a higher overall accuracy. More training data sets would help further to increase the confidence in the model. The threshold probability used is 0.5, assuming that a misclassification in both classes would have the same cost, aiming for a sub-optimal result. This varies depending on the associated misclassification the wind turbine operator would accept. The algorithms used were outlined in Section 3.4.4.

In order to further test the accuracy of the DSF presented, a number of additional algorithms were tested, as shown in Table 5.4. As it can be seen from their accuracy results, they 2 algorithms that were selected from the DSF are the ones that give the highest accuracy in the classification of the sensor data.

Table 5.4: Supervised learning algorithms tested with their accuracy.

Algorithm	Specifications	Accuracy
SVM	Gaussian, Scale:0.26	95.12%
kNN	Mahalanobis, NN=10	94.34%
Ensemble	Bagged Trees, Split: 10, learners: 30	93.22%
Decision Tree	Gini's index, max number of splits: 400	90.42%
SVM	Quadratic, box constraint: 1	87.16%

Temperature readings have aided in identifying early warning signs of the gearbox's components. However, a limitation of this analysis is that the environmental temperature outside the nacelle has not been taken into account, as it is expected to have little influence to the findings, when the power output is taken under consideration; this is something that could be considered for future analysis. Moreover, it was made clear that it is not possible to detect the exact fault location from SCADA temperature sensors and a more detailed analysis of the vibration signals is required, which can also be done for cross-validation. The different waterflow representations of the envelope and FFT spectrums show only a very minor increase in the signal sidebands, which does not necessarily represent an evident failure. Thus, the cepstrum rms was used to better visualize the vibration signal increase. This is not always easy, as those systems are in different platforms and inhomogeneous formats, which creates a problem in interpreting and analysing the data in time. This study also investigated different data-driven models for the SCADA and CMS systems, with a high level of accuracy, which could be further tested in order to increase the confidence levels in the results, to reduce unnecessary warnings.

5.4.3.2 CMS

For the vibration data, filtering and feature extraction was needed. A Savitzky-Golay filter with a high order number was applied and feature extracted data were selected, due to the nature of the CMS outputs.

For this case a quantity is predicted, and hence a regression algorithm is chosen for the analysis. An autoregressive model was trained in order to be able to predict future failures. The model's parameters are estimated using variants of the least-squares method, by only using a historical data series. An example of the model's outputs can be seen in Figure 5.29. Seven hundred time instances have been used for training, with the last three hundred instances representing the forecasted and real data that were generated. The model was able to predict the future trend, for the replaced gearbox turbine. The predicted and actual curves for the 300 time instances modelled, that represent 5 months, have the same slope value. The model was tested for all turbines of the farm, giving similar trend results for 26 out of 27 turbines, with an overall accuracy of 96.22%. In the case where the algorithm failed to forecast the instances correctly, it produced a false alarm. The algorithm used can be found in Section 3.4.4.

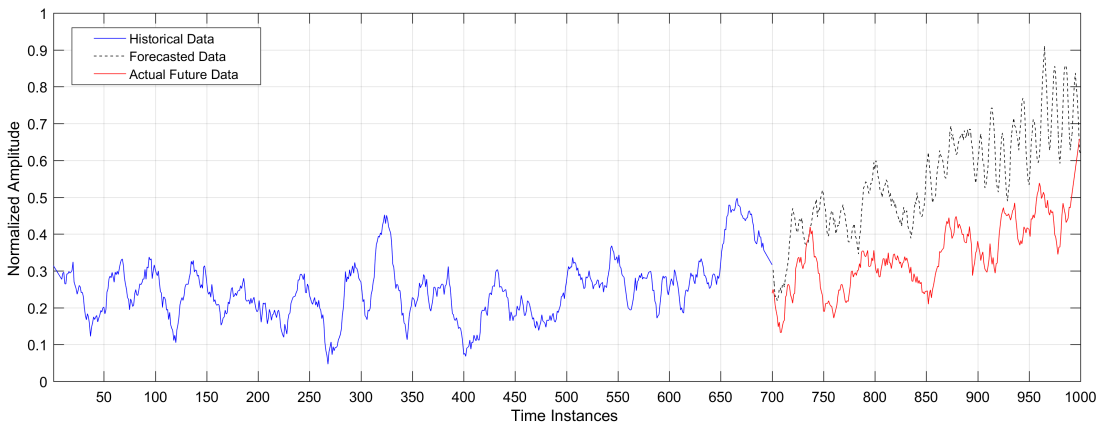


Figure 5.29: Normalized cepstrum rms of the planet bearing, with forecasted and actual sensor values for 300 time instances.

5.4.3.3 Oil Particle Counting

A similar autoregressive model can be applied to the oil particle counter data shown in Figure 5.25. Both signals estimate either the number of particles due to a failed component, which would be constantly increasing in case there is a spalling (for oil particle counters) or the increased vibrations due to a failed component, which will similarly be increasing if there is a failed component (for CMS signals).

5.4.4 Reflection on Findings

This case study has presented the use of all the available operational data from SCADA and CMS systems in order to predict a gearbox failure, through machine learning techniques. It has utilized all the operational remote monitoring information in order to create a predictive analytics model, which has reduced the number of required visits to the turbine for visual inspections and has allowed the prediction of the gearbox failure of at least six months in advance. This has provided the operational personnel with enough time to source the required spare parts, charter the jack-up vessel and schedule the operation well in advance, while the turbine was still operational, without the need to shut it down before the operation would take place. The lack of such a predictive analytics system could result in long downtime of the asset and potentially higher last minute vessel charter and equipment costs.

Chapter 6

Maintenance Optimization

This chapter presents the results from three additional case studies, outlined below, that are based on the methodologies described in Chapter 4 and are building on the case studies presented in Chapter 5. The purpose of the case studies is to validate and demonstrate the value of the methodologies developed for improved and data-informed maintenance operations.

- **Alarms Prediction** A novel tool is presented based on the reliability analytics case study (Section 5.2), aiming to predict the failure alarms that occur in offshore wind farms at different wind conditions.
- **O&M Planning** An in-house O&M planning tool is used in order to compare data-informed reliability inputs and generic failure rates found in literature in order to enhance strategic maintenance and planning decision making.
- **Transition Piece Risk-based Inspections** The risk-based operation methodology developed is applied at a TP inspection planning case study in order to propose a more cost-effective way to perform annual inspections for offshore wind farms.

6.1 Alarms Prediction

This case study is investigating further the results presented in Section 5.2 in order to examine the usefulness of the generated results for maintenance optimization. The methodology followed was presented in Section 3.3.3.

6.1.1 Results

For a 2 year period, the predicted alarm distribution is shown in Table 6.1 and Figure 6.1. The overall accuracy achieved was 99.3% for all the alarms. Over that period, the tool overpredicted the number of alarms by 1.5%. This value is the average of the total number of alarms predicted, the alarms from the different subassemblies and the number of critical alarms. The actual number of alarms is not shown due to confidentiality reasons. Instead, the percentage of the total number of alarms is shown. The rest of the alarms not shown on the table, are the ones that are either automatically reset and resolved or that do not indicating failures, but changes of state of the turbine, such as "Remote Stop" or other warnings such as "High/ Low Wind Speed".

Table 6.1: Alarm prediction tool outputs for the different failure alarms as a percentage of the total turbine alarms.

Alarms	% of Fail. Alarms	% of Critical Fail. Alarms
Converter	28.90	2.75
Electrical	8.15	1.85
Gearbox	2.85	1.66
Generator	26.15	4.02
Inverter	2.36	1.53
Pitch	3.89	1.00
Sensors	2.45	1.50
Transformer	0.89	0.28
Yaw	8.95	0.85

Figure 6.2 also shows a cumulative distribution function for the different fault alarms during the five year period. In general a wide variation is shown amongst the different

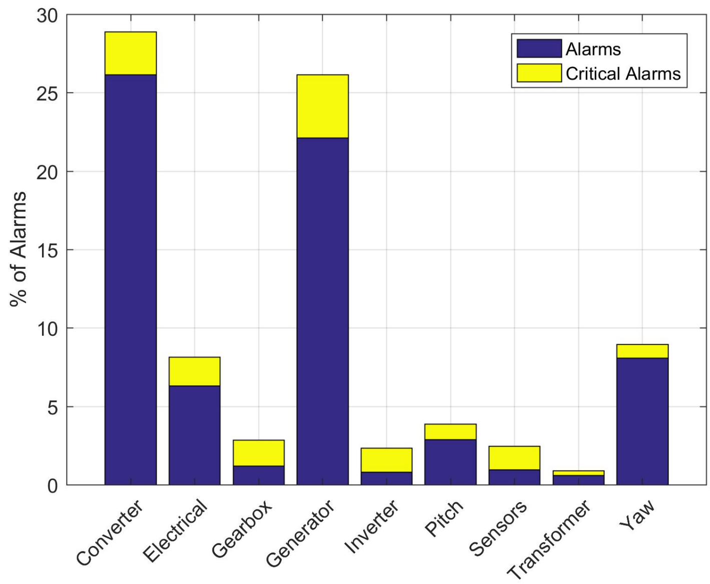


Figure 6.1: Alarm prediction tool outputs for the different failure alarms as a percentage of the total turbine alarms.

subassemblies. It is shown that yaw and gearbox systems are more affected by the higher wind speeds, whereas, sensor, pitch and electrical systems experience more issues during the lower wind speeds and thus during higher turbulence intensity values. Generators, converters, transformers and inverters have a more uniformal failure distribution. A more detailed analysis of these findings were presented in Section 5.2.

6.1.2 Tool Validation

The tool was then tested using the following cases:

- Short-term planning: used for daily operational decision making. It is mainly dependent on the accuracy of the available forecasted data and it cannot exceed 10 days.
- Long-term planning: used for planning a long-term strategy for the farm and given the available wind data, it could provide estimates for a longer period of time, informing future planning and spare part management strategies.

The tool was also tested for high and low wind conditions. The separation of the high

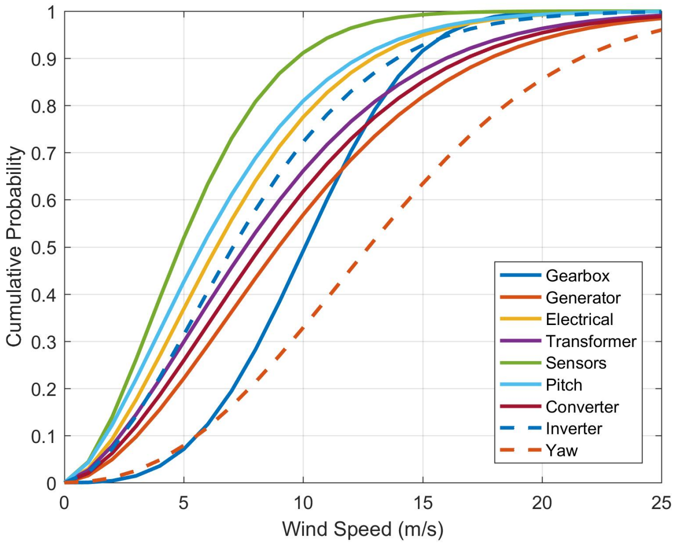
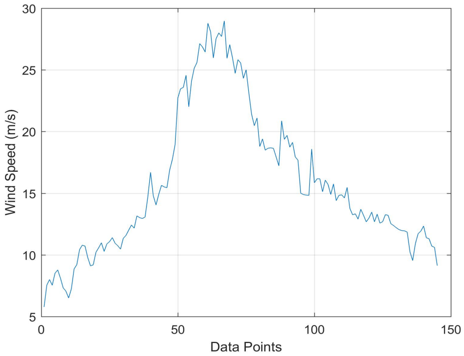


Figure 6.2: Cumulative normalized failure distribution function for the different subassemblies against the wind speed.

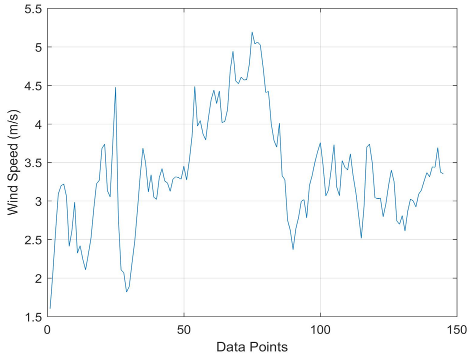
and low wind conditions was chosen due to the data availability. As there were not long enough periods of time for wind speeds well over 10 m/s.

The separation of the high and low wind conditions was chosen due to the data availability. As there were not long enough periods of time for wind speeds well over 10 m/s. All of the cases used are within the same time period used to complete the original distribution, but they only represent a few days, taking into account that the original distribution was created using a 5 year period. Moreover, they are not coming from the same data source, as the Weibull distribution data are from the individual turbine wind anemometers and the ones used for the model's validation are from the met mast.

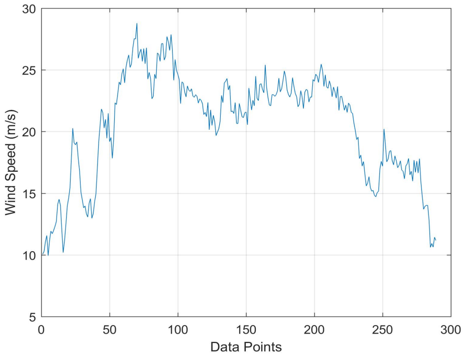
In both cases, a short-term (24h and 48h) and a longer period is tested of 9/10 days, in order to evaluate the capabilities of the tool in shorter and longer term forecasting. The different inputs from the met mast are shown in Figure 6.3 and the overall average accuracy results are shown in Table 6.2. Accuracy for the individual subassemblies has been calculated by using Eq. 6.1. As it can be seen, the overall low wind accuracy is lower compared to the high wind one. For the high wind cases, the accuracy increases with the increase of the period tested, whereas the low wind case is inconsistent.



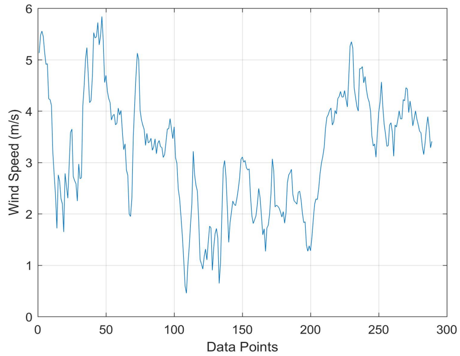
(a) 24h high wind condition



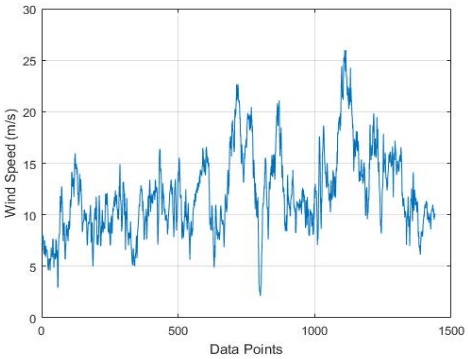
(b) 24h low wind condition



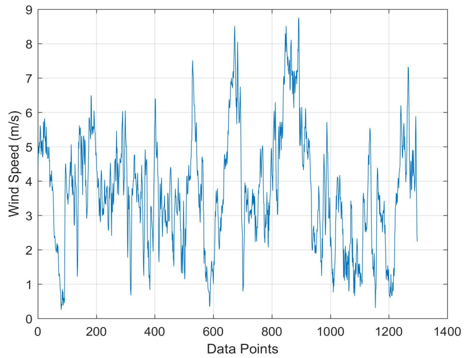
(c) 48h high wind condition



(d) 48h low wind condition



(e) 10-day high wind condition



(f) 9-day low wind condition

Figure 6.3: Met mast wind speed inputs for the alarm prediction tool.

$$Accuracy = \frac{|Real - Predicted|}{Real} \tag{6.1}$$

A more detailed output of the model and a comparison with the actual data is shown in Tables 6.3, 6.4 and 6.5.

Table 6.2: Alarm prediction tool's overall accuracy.

Duration	**Low Wind**	**High Wind**
24h	88.5% (11.5% underestimated)	87.8% (12.2% underestimated)
48h	92.7% (7.3% overestimated)	90.3% (9.7% underestimated)
9/10-day	86.9% (13.1% overestimated)	93.1% (6.9% underestimated)
2 years	**99.3% (0.7% overestimated)**	

Table 6.3: Comparison of actual and predicted number of alarms for a 24h low wind (LW) and high wind (HW) period.

Subassemblies	LW Actual	LW Predicted	HW Actual	HW Predicted
Converter	1	1.47	1	1.48
Electrical	0	0.54	0	0.21
Gearbox	0	0.03	0	0.15
Generator	1	1.07	2	1.96
Inverter	0	0.13	0	0.09
Pitch	0	0.28	0	0.10
Sensors	0	0.22	0	0.02
Transformer	0	0.07	0	0.04
Yaw	0	0.13	1	1.33

6.1.3 Reflection of Findings

The overall analysis of the results indicates that about 1/5- 1/4 of the alarms generated by the turbine needs some sort of action, which could either be a remote reset or an intervention. Around 15% of the overall fault alarms are critical and will need in situ investigation or replacement of the turbine components. Moreover, the large variation of the failures between the different subassemblies against the wind speed are also highlighted.

Table 6.4: Comparison of actual and predicted number of alarms for a 48h low wind (LW) and high wind (HW) period.

Subassemblies	LW Actual	LW Predicted	HW Actual	HW Predicted
Converter	3	2.79	7	7.10
Electrical	1	1.04	1	0.54
Gearbox	0	0.07	0	0.17
Generator	1	2.02	11	10.52
Inverter	0	0.25	0	0.13
Pitch	0	0.55	0	0.27
Sensors	0	0.44	0	0.09
Transformer	0	0.13	0	0.06
Yaw	1	0.24	6	5.04

Table 6.5: Comparison of actual and predicted number of alarms for a 9/ 10day low wind (LW) and high wind (HW) period.

Subassemblies	LW Actual	LW Predicted	HW Actual	HW Predicted
Converter	10	12.84	15	13.99
Electrical	5	4.53	3	2.65
Gearbox	0	0.46	2	1.85
Generator	9	9.62	18	16.79
Inverter	1	1.12	1	0.97
Pitch	2	2.36	1	1.20
Sensors	2	1.84	1	0.42
Transformer	0	0.53	1	0.42
Yaw	2	1.40	9	10.07

The normalized alarms against wind speed, in Figures 3.11 and 5.12 indicate that, there are comparatively more alarms at wind speeds higher than 18 m/s, indicating the influence of high wind speeds on the number of alarms. It is worth noting that at wind speeds 13-15 m/s there is a 'dip' in the number of alarms, compared to the ones experienced at lower wind speeds. These are the wind speeds where the turbine is reaching its rated power. It can be assumed that since these are the rated wind speed velocities, the turbine is designed to operate under those conditions and thus less failure alarms are triggered.

The forecasting capabilities of the tool have shown an overall 99.3% accuracy considering the prediction based on the 2 years of data provided by the met mast. This is a really high accuracy value, which indicates that the results can be potentially used in long-term prediction of failure alarms in strategy and operational expenditure tools. The short-term prediction of 24/48h and up to 10 days have shown accuracies ranging from 83.1% to 93.1%, indicating that the tool does not perform that well for short-term forecasting. This is expected as the models are data-driven and the failures occurrence is a probabilistic process, based also on the physics of failure, which would result in a higher accuracy when dealing with larger data sets. Moreover, short-term wind forecasting depends on the forecasting capabilities of the different weather models, which can add additional uncertainties into the estimation of future failure alarms. In this study we have assumed a perfect forecasting, as historical data have been used. On average, the low wind cases produce less accurate results compared to the high wind ones, probably due to the higher TI values at low wind speeds, which can cause the underestimation of the generated failure alarms. The 48h window provides the best short-term estimate with an accuracy of over 90% for both high and low wind cases, which means that the tool could be used for for a 48h forecasting with a high level of accuracy.

Some limitations of this work are discussed below:

- The training data for the tool include readings from the turbines' anemometers, which can be influenced from the wakes generated by the turbine's rotor. This is an unavoidable uncertainty in the inputs of this study, but is characteristic of the type of operational data that a wind farm owner would have typically access to. The anemometers would give point measurements for every turbine, which can then be correlated to the individual failures. For future uses, the use of nacelle lidar systems can be investigated to reduce this uncertainty in the readings, but the benefits of lidar will have to outweigh its cost.
- The tool does not work for individual turbines. It generates generic farm outputs that indicate the overall expected failure of the assets. Although this information might not specify the turbine, it is still very useful for planning of component replacement and repairs.
- The tool is producing accurate results for the first five years of the asset, after

that some deterioration factors might need to be taken into account to represent more realistic data.

- The tool and methodology presented in this thesis are data-driven, so a considerable amount of data is needed in order to generate accurate distributions for the different turbines. Examples of ageing factors on a wind farm level as a percentage of power output decline over the years can be taken from [150] and for individual components/ assemblies can be then simulated or estimated over time.
- If the tool is used in a live case study, the level of uncertainty in the short-term applications might be higher, due to any uncertainties related to the weather forecasting.
- The tool can only be used for failures detected by the alarm system. This is not always the case though, as failures can occur either without any warning, or can be detected earlier by anomaly detection or trending analysis, as it was presented in the gearbox failure case study (Section 5.4). As an effect, wind farm operators should not exclusively rely on such a tool, but use it to inform their future actions.

The overall tool's performance is very promising, as it shows that with readily available operational SCADA and met mast data and reasonable computation time, the tool can re-run when needed and produce highly accurate results. The majority of the work is spent during the pre-prepossessing, aggregation and categorization of the different data. The tool's performance is also high due to the fact that the results are averaged for the wind farm and are not shown for individual turbines.

Previous work has proven the relationship between wind speed and failure rates for onshore and offshore wind turbines, indicating that in higher wind speeds will result in higher turbine failure rates [73, 84, 151]. This study took those results a step further and quantified them on a component level and also used the environmental data to help predict future failures. These results can then be used in operational strategy and cost prediction tools, such as the one presented in Section 4.3 and in the following case study (Section 6.2). The use of the presented tool will allow prescriptive maintenance to be applied at the offshore wind assets, enabling the right personnel to be available and materials to be in stock, minimizing delays and downtime.

6.2 O&M Planning

The results from the reliability analysis in Section 5.2 are used again in this case study, in order to estimate future maintenance planning scenarios and produce availability calculations. The methodology and the tool that are used for this case study are presented in Section 4.3.

6.2.1 Model Inputs

Table 6.6 shows the generic model input values used for the simulations, taken from [37, 73, 150]. Table 6.7 the overall multiplication factors, as derived from Figure 5.10 values.

Table 6.6: Generic model inputs.

Input	Manual Reset	Minor Repair	Major Repair	Major Replacement
Failure rate /turbine/year	7.5	6.81	1.17	0.29
Ageing Factor/ year	1.69%	1.69%	1.69%	1.69%
Repair time (h)	3	6.67	17.64	116.19
Number of technicians	2	3	3	9
Vessel type	CTV	CTV	CTV	Jack-up

Table 6.7: Overall multiplication factors.

Turbine	Factor	Turbine	Factor	Turbine	Factor
A1	1.25	B10	0.97	C19	0.77
A2	1.4	B11	0.88	C20	1.01
A3	1.01	B12	0.8	C21	0.75
A4	1.45	B13	0.6	C22	0.97
A5	1.27	B14	0.95	C23	0.72
A6	1.3	B15	1.02	C24	0.71
A7	1.19	B16	0.92	C25	0.74
A8	1.34	B17	0.99	C26	0.88
A9	1.02	B18	1.12	C27	0.96

Following is a list of assumptions that are taken under consideration in the model:

- 60 hours of annual maintenance for every turbine, performed during the summer period.

- 10 days of delay for major component replacement, in order to wait for the jack-up vessel to arrive.
- Operations are taking place the whole week (including weekends) from 7:00 until 18:00.
- Operational limits are considered for weather conditions above 1.5 m of significant wave height and 12 m/s of wind speed.
- The failure rates were modelled using a Weibull distribution for the different cases, as shown in Eq. 6.2; where t is the time when the failure occurs, λ_i the scale factor and β_i the shape factor. In the case where ageing is taken under consideration $\beta_i = 1.07$ and the component is considered "as bad as old" after the maintenance has taken place. In the case where ageing is not taken under consideration $\beta_i = 1.0$ and the component is considered "as good as new". The scale factor was then converted to a daily value by dividing the failure rate in Table 6.6 with the number of hours in a year; 8760.

$$R(t) = e^{-(\lambda_i * t)_i^{\beta}} \tag{6.2}$$

- There is always a CTV available on site.
- Turbine is shut down when an operation takes place.
- Turbine is immediately shut down only when a major repair is required, assuming that the turbine cannot operate and is shut down until the technicians arrive and fix the fault. In the case of minor repairs, the turbine does not need to be shut down and when a major replacement is needed, it is assumed that it takes place before the catastrophic failure of the component and it has been predicted well in advance; thus, a shut down is not required.

The different scenarios that have been considered in this thesis are the following:

- One turbine (with and without ageing)
- Teesside wind farm with average reliability values to all the turbines (with and without ageing)

- Teesside wind farm with weather-informed reliability parameters (with and without ageing)

6.2.2 Single Turbine

Figure 6.4 shows the TBA and PBA values for a single turbine over a 20 year period. The turbine considered is turbine A1, with regards to its distance from the port and the generic model inputs are used for the different failure modes. As expected, PBA values are lower than the TBA ones. The asset management tool is accurate enough and able to pick even those small differences in availability and reliability calculations. The failure rate generation results are shown in Figure 6.5 for the different types of failure considered with and without taking into account the ageing factor. The failure rates using the ageing factor are increasing on an average of 1.69%, in order to simulate the degradation. For the cases without ageing, the failure rate values are fluctuating around the average value. As there are no studies indicating the failure rate over the lifetime of the assets, this is considered a sufficient assumption. Only one study has presented a model of the failures of subassemblies through time [55], but with only

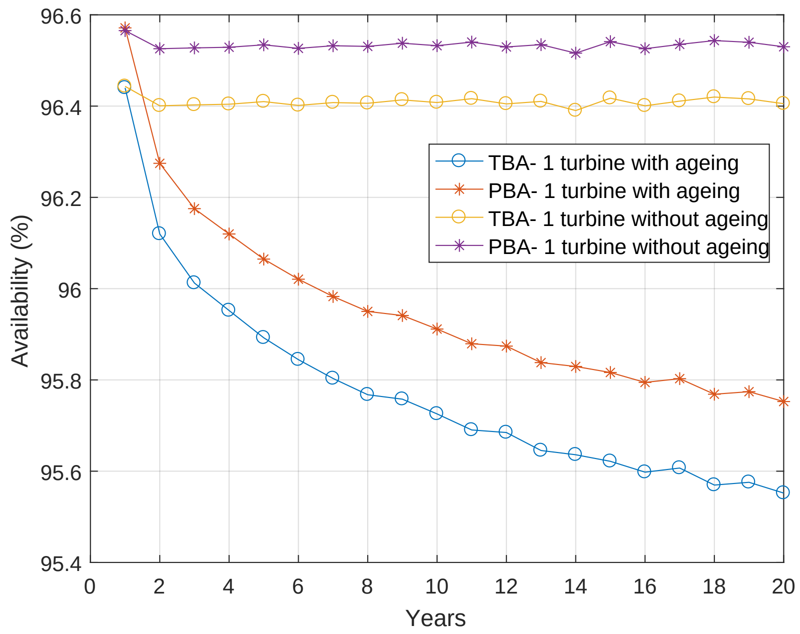


Figure 6.4: Time and production based availability for 1 turbine, with and without ageing for 20 years of lifetime.

5 years of operational data. The study showed that generator and electrical systems tend to follow the bathtub curve, whereas gearboxes tend to have a linear increase of failures over time. This type of modelling is not used for this study, as it varies between literature findings and turbine manufacturer and it is not clear yet what the most representative failure distribution curve is. Instead, an increased failure rate over time is considered through the ageing parameter since the only available study shows a power degradation of the turbine over its lifetime.

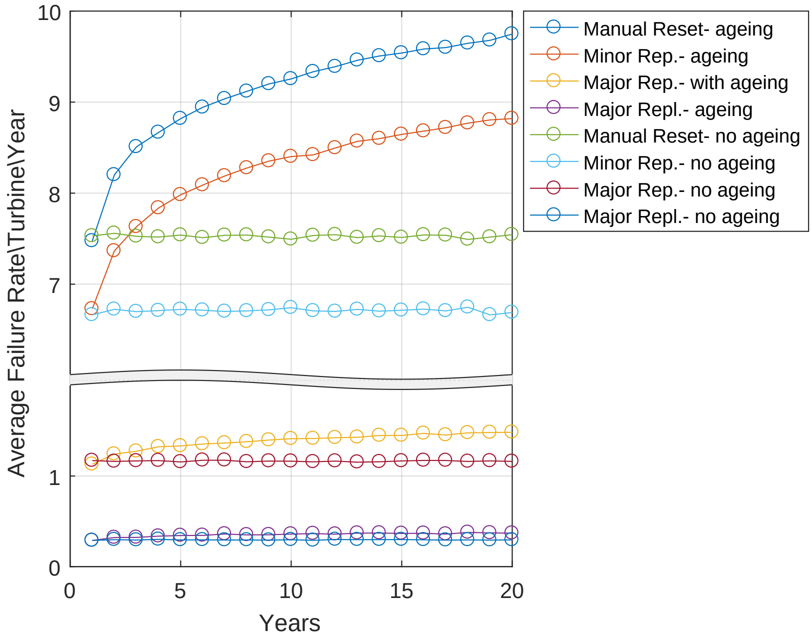


Figure 6.5: Failure rate evolution over time for 20 years for 1 turbine (including manual reset, minor repair, major repair and major replacement with and without ageing).

6.2.3 Wind Farm with Generic Values

Figure 6.6 shows the TBA for the simulated case of the Teesside wind farm, using the generic input values. The simulation with the ageing parameters is shown at the lower part of the figure and without the ageing parameters is shown at the top. All the individual turbines follow the same pattern and the small differences between them are caused due to the different distance from the port which could result in longer or shorter downtime periods. By comparing the wind farm results with the ones from the individual turbine model, turbine A1 has a 0.5% lower overall availability when it was

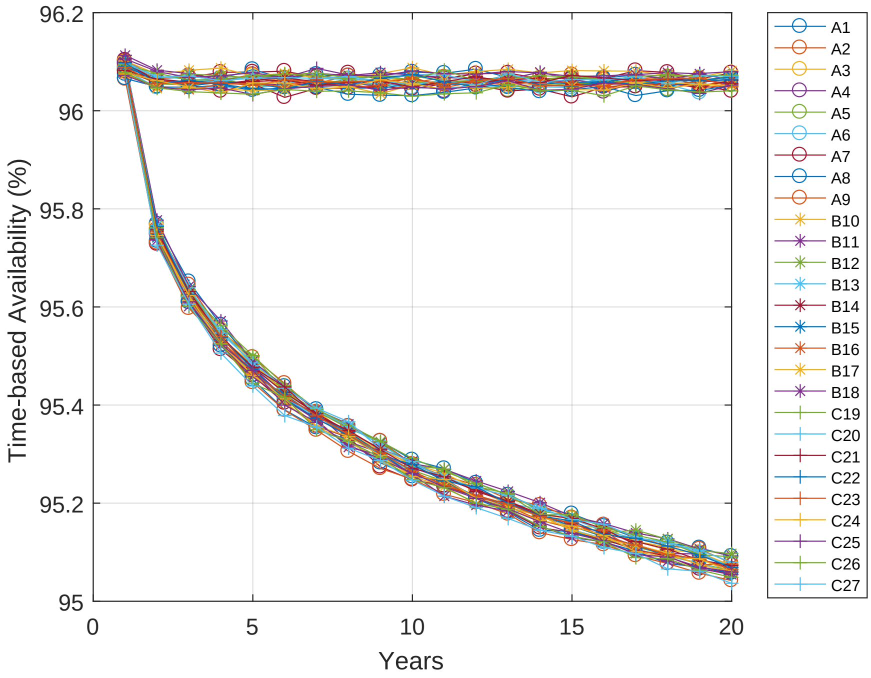
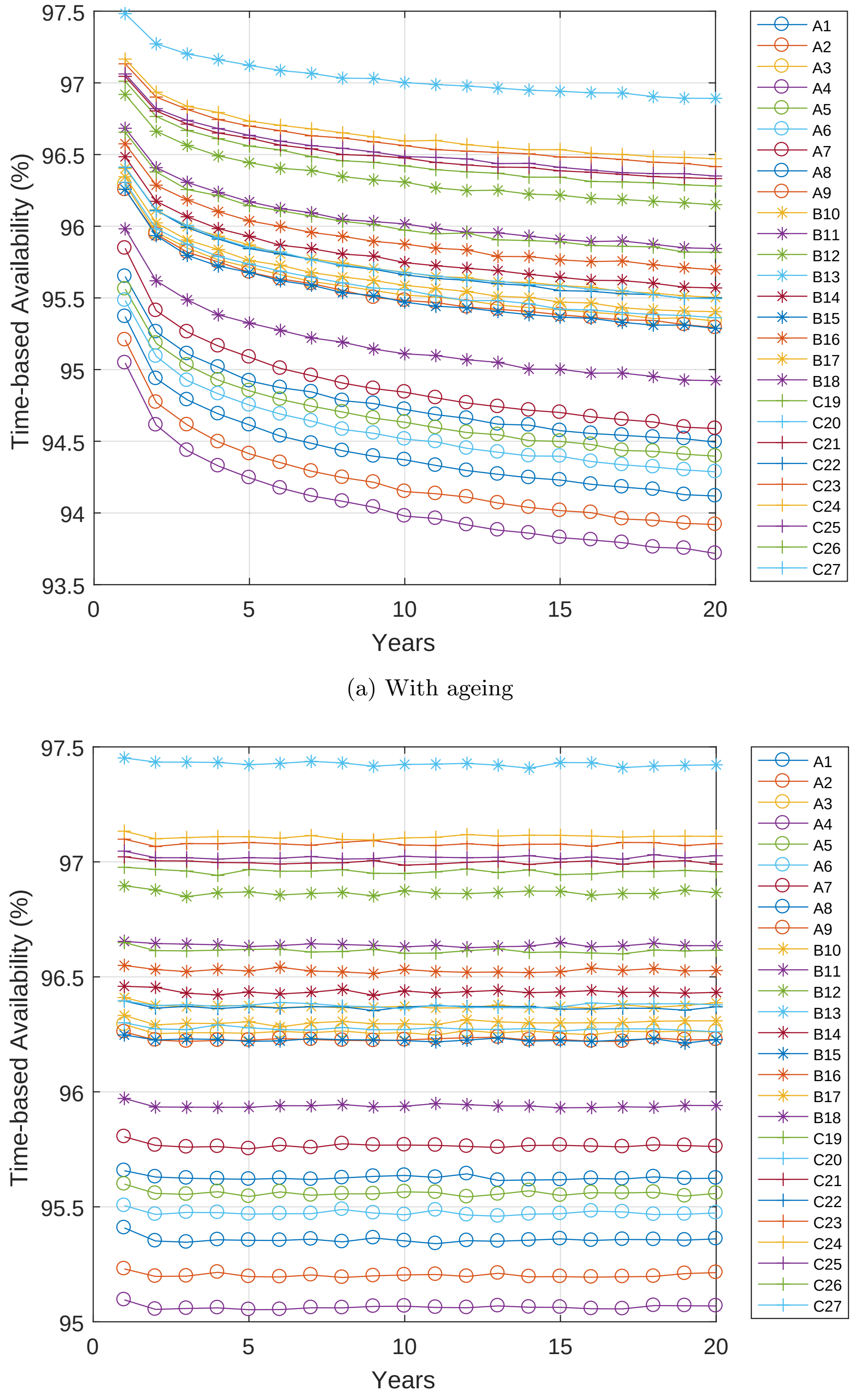


Figure 6.6: Time-based availability for 27 turbines over 20 years for the generic wind farm case. The top 27 lines correspond to the cases without ageing and the bottom ones with ageing.

by itself compared to when it was simulated on the farm. This is due to the resource unavailability, as the technicians and/ or the CTV might be on another turbine. The overall availability difference could be translated to around 22 GWh difference for the 20 years of operation, which could be around £6.5m. of OPEX costs.

6.2.4 Wind Farm with Data-informed Parameters

Figure 6.7a shows the TBA for the simulated case of the Teesside wind farm, taking under consideration the data-informed parameters from Table 6.6 and the ageing factor. Figure 6.7b has the same input parameters, but without ageing. In both cases the turbines perform according to the different input parameters and the multiplication factor used. There were no added significant delays from any unexpected events and all the turbines followed the same trend. There are a few overlaps in the TBA values for example between turbines A1 and A5 as well as turbines C20, A3, C19 and B15. This is possibly due to any failure or weather downtime or any vessel delays.



(a) With ageing

(b) Without ageing

Figure 6.7: Time-based availability for 27 turbines with weather conditions for 20 years.

6.2.5 Comparison

For the different cases in the previous subsections, the average time and production based availability values are shown in Figure 6.8. The difference between the highest and lowest availability values are shown in Table 6.8.

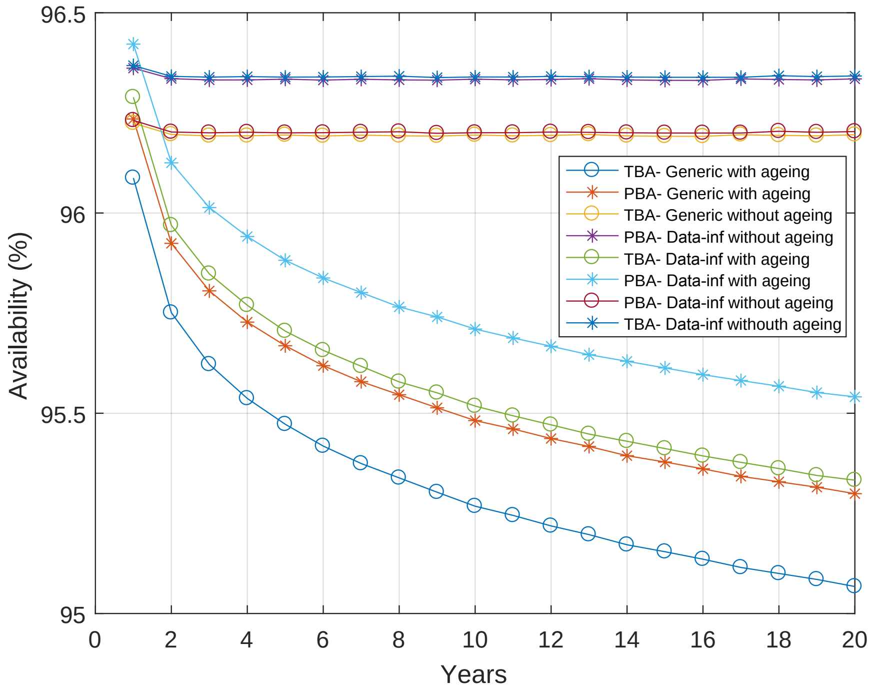


Figure 6.8: Comparison of average time based and production based availability values for 27 turbines with and without ageing over 20 years of operation.

Table 6.8: Availability comparison for the different cases.

Cases	**Highest**	**Lowest**	**Difference**
Generic failure rates- no ageing	C27	B11	0.04%
Generic failure rates- with ageing	C27	B11	0.05%
Turbine specific failure rates- no ageing	A4	B13	2.48%
Turbine specific failure rates- with ageing	A4	B13	3.16%

It is interesting to see that in the case where the data-informed parameters were taken into consideration, the TBA and PBA values are higher. This could be due to the fact that as the failures are defined by different failure rate values, the repair actions will not be required at the time, and thus the resources can be managed more effectively. It shows that in the case where generic values are used for OPEX modelling, the overall values are more pessimistic and an average increase of 0.22% in production and in revenue can be simulated with the data-informed parameters. This difference could also be lower or higher, depending on the weather parameters that are chosen each time to run the simulation, as they are stochastic. Similarly the TBA values for the data-informed case were higher on average by 0.25% compared to the generic ones. The overall availability difference could be translated to around 10 GWh difference for the 20 years of operation, which could be around £3 m. of OPEX costs.

There is very limited literature regarding offshore wind farm availability values. An early study [74] indicated that Round 1 UK offshore wind farms experienced a TBA of 80.2%. The study was performed before Teesside was built. A more recent study from ORE Catapult [152], which included all the offshore wind farms in the UK, indicated that the 50th percentile of the wind turbines have an average PBA of 96.31%, with the 90th percentile being at 98.57% and the 10th percentile being at 90.28%. This consists of a range of wind turbines and wind farm sizes and distances from the port. Comparing the results of this study with the ORE Catapult's report, Teesside's PBA values seem reasonable, as the wind farm is close to the shore and an intervention can be made easily. As an effect the overall PBA values are very close to the 50th percentile values from the report.

6.2.6 Reflection on Findings

This comparison indicates that the currently publicly available data can be used by offshore wind farm developers as an estimate, as they represent the average cases and it is a good starting point. Once the operators have a more detailed understanding of the asset, they can refine those data with their own parameters in order to make more realistic predictions. It also shows the importance of accurate O&M modelling and the sensitivity of the different reliability input parameters which could have changes in the final availability values of the individual turbines of up to 3% in the presented case study. Such differences could result in large uncertainties in the final OPEX values that would need to be addressed during the design phase of the farm.

6.3 Transition Piece Risk-based Inspection

Based on the outlined approach, in Section 4.2, a case study for RBI for offshore wind TPs is presented here for Teesside offshore wind farm. A bespoke model is initially built and eight different inspection strategies are considered.

6.3.1 RBI Tool

The proposed model is shown in Figure 6.9. It includes inputs from different sources that are presented in the following subsections. Inputs include estimated costs, comments from operations and inspection reports and data from the structural health monitoring systems and any corrosion measurements. Inputs are not weighted, but they can be prioritized via the criticality analysis or the triggering of immediate actions. These are combined in the model where a sequential MCS is used to generate inspection scenarios that would be required through the lifetime of the structure. The model runs a two state Markov chain from the generated inputs, by considering the mean time for a required inspection and its mean duration time, as defined by the input findings. The time for a required inspection is defined by the occurrence probability. Similar Markov chain models have been used by literature in order to estimate the reliability and availability of engineering systems [153, 154].

6.3.1.1 Cost Data

Some basic estimated cost data have been assumed for onshore (including planning and report writing) and offshore operations (including vessel hire, transport and personnel cost). An overview to some of the assumptions used to model the costs can be found in Table 6.9.

6.3.1.2 Comments from Operations

Another input is the assumption that certain inspections can be performed and logged by the maintenance personnel during scheduled operations. These are the ones identified in Table 6.12 with $D<5$, such as signs of external biofouling, paintwork, or internal

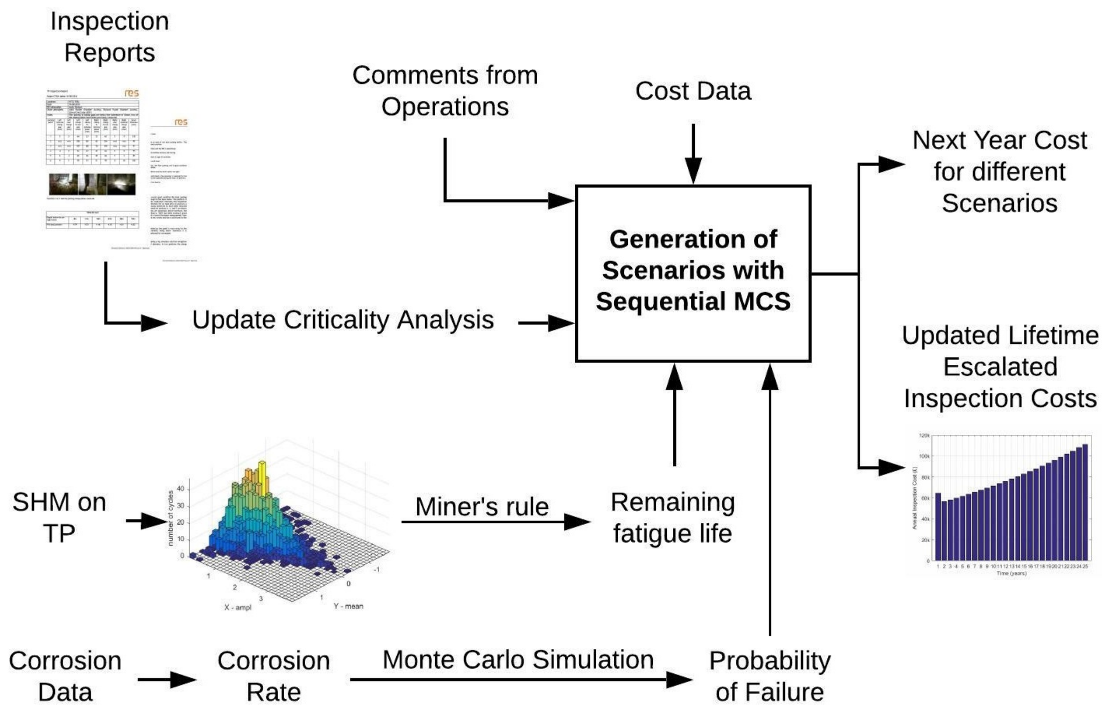


Figure 6.9: Overview of the proposed model for TP RBI planning.

Table 6.9: Overview of assumption for baseline cases used to model the costs.

Inputs Name	Input Values
Number of Days	3
Number of Turbines Inspected	6
Number of Technicians	4
Average Inspection time per turbine	3 h
Distance from port	7 km
Distance between turbines	0.5 km
Transfer time per turbine	10 min
Vessel mob/ demob time	30 min

lighting. They can then be used alongside with the inspection reports, either to update the information received from them, or to report new issues.

6.3.1.3 Inspection Report Information

With the current knowledge from the latest TP inspection activities, the different inspected components have been positioned on the risk matrix (Table 4.2). This is a first quick check to access the severity of the different component inspections, without moving to their C-RPN number. Table 6.12 defines the S/R, O and D values for the different components. The final values, used in Eq. 4.3 and 4.4 to prioritise the inspection activities in the model, are shown in Table 6.10 and 6.11. Time has also been assigned to each of the tasks, which allows the model to run more accurate inspection time estimates, when some of the tasks are not performed.

Table 6.10: Risk matrix (RM) and S/R, O and D ranking for internal TP inspections.

	Access ladders	**Lighting**	**Ventilation**	**Lightning protect.**	**Fall arrest**	**Flooring**	**Hatches**	**Air-tight platform**	**Air-tight hatch**	**Cable hang off**	**Belzona coating**	**Flange welds**	**Grout thickness**
RM	16	20	16	16	16	23	16	16	16	16	20	11	11
S/ R	8	8	8	10	7	5	7	6	6	6	8	4	4
O	1	1	3	1	1	1	1	1	1	1	1	1	1
D	5	3	6	8	6	5	8	2	5	3	3	8	9

Table 6.11: Risk matrix (RM) and S/R, O and D ranking for external TP inspections.

	Turbine ID sign	**Marine growth**	**Paint work**	**Access ladder**	**Access gates**	**Structure bolts**	**Safety signage**	**Platform**	**Cabling**	**Navigation aids**	**Life buoy**
RM	1	20	13	15	6	6	4	4	15	15	15
S/ R	1	7	5	2	5	6	4	5	8	8	8
O	1	8	7	1	1	1	1	1	1	2	1
D	1	1	3	2	2	8	3	2	8	8	1

6.3.1.4 Structural Health Monitoring

The SHM system in place can flag any immediate actions needed on the structure and also allow making predictions of the fatigue lifetime of the different inspected components. Usually a representative sample of turbines is fitted with SHM systems, where the harshest environmental conditions are expected. This study, is using existing SHM systems and performs an analysis of the strain gauge information received. A state of the art review of damage detection methods has been completed by [155]. SHM systems can operate in parallel with the scheduled I&M operations for contingency and in order to be able to identify any unexpected failure modes. This could trigger either an immediate inspection or allow prioritization of the inspection activities in the coming inspection interval. The relevant strain gauge sensors are monitored on the TP. The data collected were converted into stresses and the rainflow diagram was constructed in order to be able to calculate the number of cycles, as indicated by [156]. Finally, by using the Miner's damage hypothesis [157], the total damage of the structure was calculated, as shown in Eq. 6.3.

$$\sum_{i=1}^{k} \frac{n_i}{N_i} = C \tag{6.3}$$

Where n_i is the number of cycles accumulated at stress σ_i, N_i is the average number of cycles to failure at the ith stress and C is the accumulated damage, where 1 indicates failure. The expected lifetime of the inspected component was then calculated by dividing the time interval used by the accumulated damage.

6.3.1.5 Corrosion Data

The corrosion data were used in order to determine the probability of failure (PoF) of the structure due to corrosion. An approach introduced by [158] was followed. The remaining life (RL) of the structure was initially calculated by using Eq. 6.4 [159].

$$RL = \frac{t_{actual} - t_{required}}{CR} \tag{6.4}$$

Where t_{actual} is the current thickness of the structure, $t_{required}$ the minimum allowable thickness defined by the design reports and CR the corrosion rate of the structure. t_{actual} can be calculated at any given year y by Eq. 6.5, where t_0 is the original designed structure thickness.

$$t_{actual} = t_0 - CR \times y \tag{6.5}$$

From the design reports [160], t_actual=4200mm with a corrosion allowance of 7.5mm was used. An indicated CR=0.3mm/year was considered from the standards [46]. These data are recommended to be updated by experimental or field studies and observations. In this study a more conservative CR=0.4mm/year was considered in order to reduce any associated risks. CR and t_actual are presumed to have normal distributions, assuming a uniform corrosion [161] with a standard deviation of 0.1 and 1 respectively. The PoF can be finally calculated by using an MCS to define the values of Eq. 6.4, when the criterion shown in Eq. 6.6 is met. A graphical representation of the results is shown in Figure 6.10.

$$PoF = P(RL < 0) \tag{6.6}$$

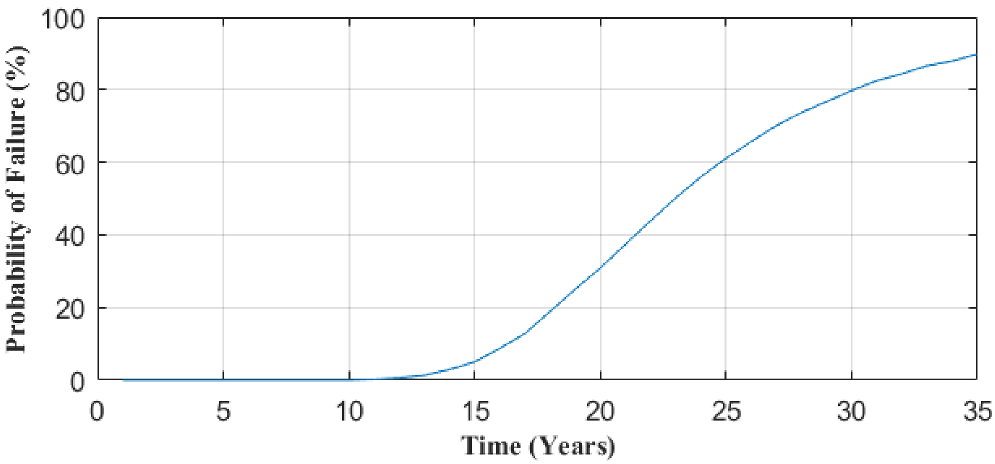


Figure 6.10: Estimated annual probability of failure of the TP.

6.3.2 Scenarios

Eight different test cases have been considered. Baseline 1 and 2 represent the typical inspection plans that wind farm operators are following. Baseline 1 considers inspection of a sample of 6 out of the 27 turbines, throughout the lifetime of the wind farm, where the same TP components are being inspected annually. Baseline 2 assumes that the inspection operations are performed in combination with other maintenance operations, so in that case, the transport costs are shared between the two. In reality, the actual estimated costs would be somewhere in the middle, as it is not always possible to plan both inspection and maintenance operations at the same time. Both Baseline 1 and 2 represent very low risk inspection strategies, since all the components are thoroughly inspected every year. These two cases are visualized in Figure 6.11, where the escalated annual costs at a rate of 3% are shown. At Year 1, a one-off tooling and equipment cost is assumed for all cases.

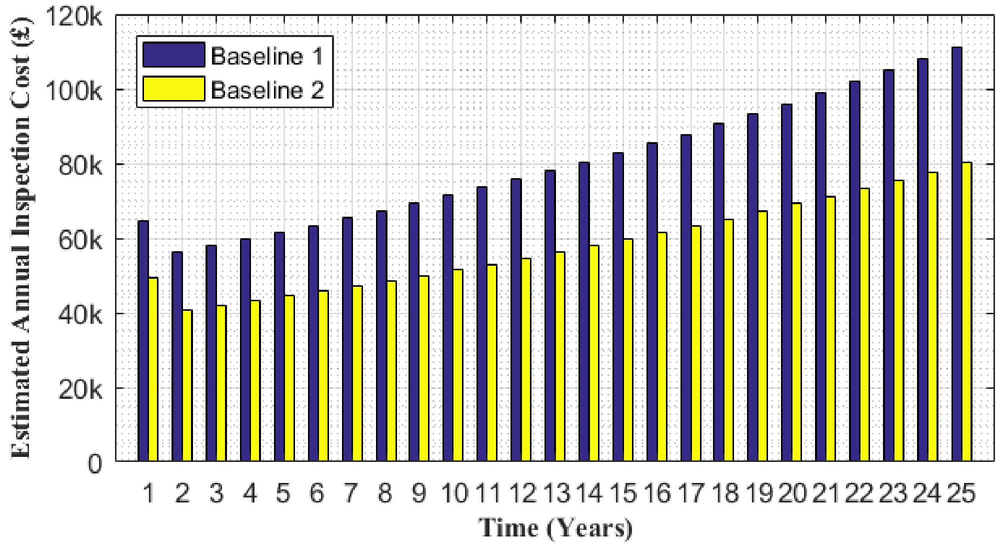


Figure 6.11: Annual escalated costs for B1 and B2 inspection plans.

Below is a brief description of the assumptions considered in the different scenarios:

Scenario 1 Full internal and high RPN of failure external inspections for the first years, until PoF of the structure exceeds 1%.

Scenario 2 Full internal inspections are performed, whereas the external inspections are completed during regular operations.

Scenario 3 Similar to Scenario 2, with reduced internal inspections, assuming there is a SHM system in place on 2 representative turbines.

Scenario 4 Same as Scenario 3, but with bi-annual inspections.

Scenario 5 Same as Scenario 4, with a reduced turbine sample of 3.

Scenario 6 Same as Scenario 4, with an increased turbine sample of 9.

The estimated lifetime costs for the different scenarios can be found in Table 6, along with the qualitative risk levels for Years 1- 12 and 13- 25, when PoF of structure >1%.

Table 6.12: Overview of assumption for baseline cases used to model the costs.

Scenarios	Cost (£ m.)	Risk level Years 1- 12	Risk level Years 12- 25
Baseline 1	2.005	Very Low	Very Low
Baseline 2	1.447	Very Low	Very Low
Scenario 1	1.411	Low	Very Low
Scenario 2	1.040	Low	Very Low
Scenario 3	0.839	Low	Very Low
Scenario 4	0.438	Medium	Low
Scenario 5	0.434	High	Medium
Scenario 6	0.442	Medium	Low

Finally, another output of the model is a cost estimate for the different test cases considered in the following years. Since the model is designed to be updated on an annual basis, there is no need to show a longer period of cost estimates, as these values would alter depending on the inputs and will not be representative. An example for years n and $n+1$ is shown in Figure 6.12. Scenarios 4- 6 are bi-annual and thus not shown at year $n+1$.

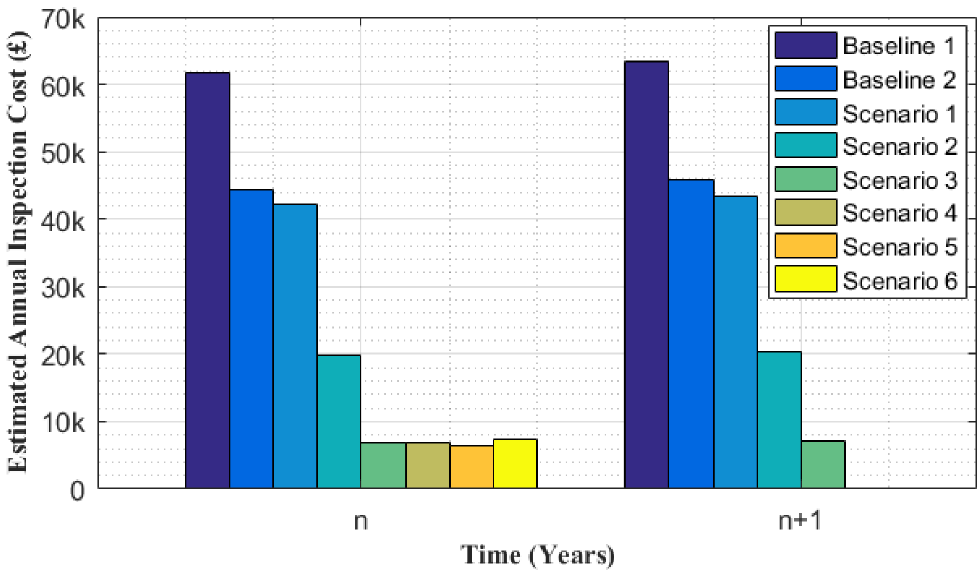


Figure 6.12: Example cost output for years n and n+1.

6.3.3 Reflection on Findings

Different scenarios are tested to understand the influence of various operational parameters to the overall TP inspection costs. At the same time, a trade-off is presented between lower costs and higher risk that should be investigated when selecting an inspection strategy. From Table 6.12, it is evident that:

- Baseline 1 and 2 are the highest cost and lowest risk cases.
- Scenario 5 is the lowest cost and highest risk case.
- Scenario 3 is the most balanced case, offering a reduced cost and low or very low risk. This is achieved with the introduction of SHM systems.

Moreover, the next year cost estimates shown in Figure 6.12 can provide a better understanding of the short-term cost benefits for the different scenarios. It can also allow the end-user to implement different scenarios at each inspection interval, by evaluating the costs and associated risks.

These test cases are just an example and have been created in order to highlight the importance of utilising operational data for upcoming inspections. It is evident from the results of the case studies that there are significant cost reduction opportunities when following risk-based operations. Furthermore, by introducing the restriction metric R in the C-RPN equation, safety becomes a priority when maintenance actions are ranked. In the case of Teesside offshore wind farm, it was shown that:

- Encouraging reporting of failed components during regular maintenance operations can reduce inspection costs by up to 26%, just by minimising work duplication, while maintaining low risk levels, as comparison of Scenarios 1 and 2 costs indicates.
- Introduction of SHM can decrease inspection costs by up to 19%, as shown in Scenario 3.
- Reducing or increasing the turbine sample size during inspection, from 6 to 3 or 9, does not have a significant effect on cost, as shown in Scenarios 4, 5 and 6 respectively.

Limitations of this work could include the early stages of the offshore wind industry, which would result in unexpected failure mechanisms and lower estimated costs for the lifetime of the asset. To tackle this issue, frequent monitoring and analysis of the SHM systems is suggested. Furthermore, this study does not provide a detailed analysis of all the potential failure mechanisms and assumes the use of a commercial SHM system. The potential cost and level of detailed analysis that can be invested in such systems is up to the user and there is always a cost-benefit analysis that needs to be considered. As an effect, only the benefits in daily observations and measurements are presented, as this study makes the case for a holistic approach when it comes to I&M decision making.

Part IV

Concluding Remarks

Chapter 7

Discussion and Conclusion

This chapter describes the implications of the results presented and how these can be used alongside the developed methodologies by the offshore wind farm owners and operators in order to create more effective data-driven decisions.

7.1 Summary of Results

All the chapters in Part III presented the results of the different O&M case studies and had their own discussion either embedded into the description of the results or as a separate section in the end. This section brings together and combines these findings, by providing a consistent discussion. An overview of how the different sections of the thesis are linked together is shown in Figure 7.1.

7.1.1 Process Improvement

The literature review has indicated a lack of standards and guidelines for data collection and taxonomies. Section 3.1 has suggested methodologies and recommendations for data workflow improvement, data collection, data architecture and the implementation of a proposed infrastructure. Some of the process improvements suggested have been tested, providing promising results; including the automation of operational data collection, which eases and streamlines the analysis of the data, allowing more time for data mining rather than data formatting and collection.

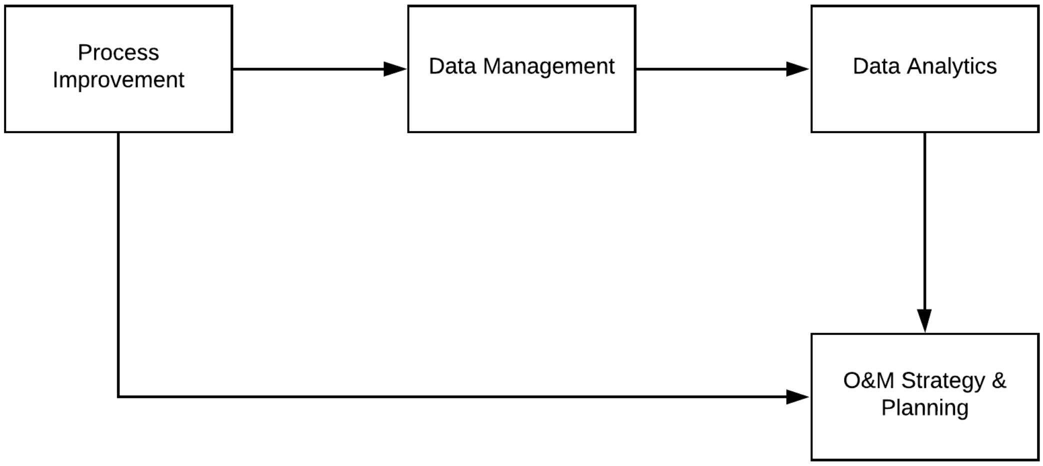


Figure 7.1: Overall section connections.

7.1.2 Data Management

The review of the state of the art in data management for wind turbines concluded that there is a lack of data management guidelines in the offshore wind industry and very few industrial and academic solutions have addressed the full data streams generated by an operational wind farm. A framework for combining both structured and unstructured data was needed in order to speed up the analysis and fault diagnosis. Therefore, a methodology was suggested for integrating and visualizing structured and unstructured datasets for an operational offshore wind farm, as presented in Section 3.2. The methodology has been implemented showing a much quicker navigation between the different data sources that are at a single point of access instead of multiple ones. This has allowed the use of root cause analysis for the turbine failures, as well as eased the reliability analysis.

7.1.3 Data Analytics

Once a streamlined process and a robust data management system are in place, data analytics techniques can be applied. This thesis presented a data analytics framework that includes reliability analysis and predictive analytics.

Reliability Analysis It was highlighted in Section 2.2.5 that there was a need of better understanding the reliability characteristics of wind turbines and that there was a lack of reliability data for offshore wind farms in the literature. This led to the presentation of failure rate statistics for Teesside offshore wind farm in Section 5.1, as well as more detailed reliability analysis in Section 5.2. It was indicated that wind turbines need to be treated as individual assets, as even the same turbine under the same conditions can perform differently at a different location in the wind farm. In order to better understand this relationship the turbine alarms were used, as they can be linked to the weather timestamp and provide a better understanding of the assets compared to the traditional work order analysis. Results showed the correlation of several subassemblies with the environmental conditions, leading to the development of a data-driven alarm predicting tool in Section 6.1.

Diagnostic and Predictive Analytics Another method that was identified as a key component for cost reduction was the use of diagnostic and predictive analytics' techniques, as they allow the forecasting of failures well in advance giving more time to the operator for planning and avoiding any unexpected downtime of the asset. The different ML techniques, as well as their different uses were presented in Sections 2.4 and 3.4. It was also indicated from Section 2.2 that not many studies fully utilized both SCADA and CMS data for ML. A case study was presented for using both SCADA and CMS systems for fault diagnosis and prognosis on a failed gearbox at Section 5.4. Moreover, a forecasting of turbine alarms was presented in Sections 3.3.3 and 6.1.

7.1.4 Maintenance

The data were then used in order to help and optimize the maintenance operations. A framework was implemented for incorporating the machine learning models into the maintenance decision making, aiming to streamline the processes, Section 4.1. The framework was tested on choosing the right machine learning model for the gearbox replacement case study, Section 5.4. An RBI framework was also generated in Section 4.2 and was applied on a TP inspection case study in Section 6.3. Finally, an O&M strategy tool was utilized in order to test the differences in availability, by using different reliability data to estimate the lifetime availability of an offshore wind farm and the sensitivity of different reliability parameters.

7.2 Building a data-driven strategy for O&M

Big data and analytics are at the top of the agenda for many companies across a range of different industries, sectors and levels within the organizations. However, making use of the full potential of the companies' data has always been a challenge. The energy industry is one of the many industries that are struggling to take the most of the generated asset data.

This focus on data is creating a lot of hype, along with the implementation of artificial intelligence and machine learning techniques. It is generally believed that with the use of big data techniques, companies will be able to transform their processes and have substantial performance benefits. According to research presented in the Harvard Business Review, companies that inject big data and analytics into their operations show productivity rates and profitability that are up to 6% higher than those of their competitors [162]. More specifically for the the energy sector, the International Energy Agency has released a report on Digitalization & Energy, where the average cost savings that are expected by the adoption of digital data and analytics up to 2040 are the following [163]:

- 5% reduction in the O&M costs.
- 5% increase in the plant's and networks' efficiency.
- 5 year life extension for plants and networks.

Although there seems to be a lot of potential both on the reduction of CAPEX and OPEX costs, as well as increasing efficiency, there is often a lack of implementation of those solutions. This is due to two main reasons; first, there is lack of trust from decision makers in investing in such solutions and second, different tools have been implemented in the past without providing any promising outcome (or they did not meet the expectations). Thus, it is important for data-driven tools and methodologies to prove their benefits with quick wins and with proof of concepts.

It is also very important that there is a desire from the business to drive an integrated approach to data aggregation, model building and implementation. This will allow organizations to follow integrated approaches across the different business units, avoiding work duplication and preventing the common trap of starting with the data and simply asking what they can do for you. This integrated approach has to be initiated by senior decision makers in the organizations in order to create a uniform approach for handling data across the business. In order for companies to improve their performance and move to applied data-driven strategies, they need to make improvements in the following three areas.

7.2.1 Data Selection

The different data sources available in an offshore wind farm are presented in Section 2.2. The volume of data and information collected at an operational wind farm is growing rapidly. The opportunities to create new insights with the existing or new types of data is also increasing, as a significant number of companies are entering the market, as described in Section 2.4.5, and due to the less costly computational power and storage that is widely available. Better and more advanced data insights and analytics give companies a better view of their operational performance, allowing them to act strategically upon their future decision making, focusing on actions that reduce OPEX and increase turbine efficiency, performance and reliability. As an effect, it is crucial for the whole wind turbine value chain to create new ways to make use of existing data sources as well as exploring new opportunities.

Use of available data The large number and the different format of existing data sources is making it hard for wind farm developers to fully utilize all the available data. Thus, section 3.1.2 presented guidelines on more efficient data collection, proposing a structured way of recording and efficiently collecting the right data and information for different uses. Moreover, in Section 3.2 a data management framework was presented, allowing the operational personal to integrate and visualize the operational, helping them to understand the value of the different data sources and the relationships between them, making use of all the generated data from the farm.

Innovative thinking Wind farm operators and OEMs need to also investigate the potential of new sources of data. It is important to understand the missing data sets and invest in new technologies that can add significant value to fault diagnosis and prognosis and that will identify faults quicker and more accurately. This could lead to the generation of digital clones of the machines, allowing better analysis and interpretation of the operational data through the use of artificial intelligence. It will also reduce the need for human intervention, which will subsequently reduce the operational cost and increase the safety of the personnel.

IT infrastructure Existing IT architecture may prevent the accessibility and integration of data and information. Additionally, managing unstructured data often remains beyond traditional IT capabilities and it was made apparent from Section 2.4.5, as most of the commercial companies are only using timeseries data to provide their analysis. A streamlined data flow is very important in order to achieve real-time decision making. The framework presented in Section 3.1.3 tackles those challenges and proposes a new way for data centricity, with multiple users and capabilities. The time for implementing those solutions centrally within a large company will require significant effort. EON has been developing for years a data lake project, which is being tested in their offshore wind fleet [164]. Similarly EDF R&D has been working on a data lake project for the nuclear fleet, which is now also implemented for wind turbines [165]. Due to the significant time commitment that is required for such projects, companies can address short-term needs by prioritizing their requirements. This can be done by connecting the most important data for analysis, following by all the required filtering and integration processes and then working on any missing information that might be required. This can allow companies to create their own proof of concepts, as well as being able to identify their gaps and work with companies that can help them tackle these challenges. This approach will allow the creation of a flexible IT infrastructure that will be able to support existing tools and incorporate external solutions by promoting swift analysis, innovation and collaboration.

7.2.2 Model Development

Data are valuable, but performance improvement and cost reduction come from analytics models that allow prediction of faults, optimization of operation and advanced insights about the condition and performance of the wind turbines. The most effective approach in designing a model usually originates with identifying the opportunity or the value it can add and determining how the model can help towards this goal; it will seldom start with the data.

Avoid research for the sake of research Unfortunately, not all models follow this approach. The research community tends to follow a data mining approach, which will produce inconsistent results and usually raise more research questions. Owning large datasets allows operators to run multiple statistical models to identify hidden patterns, but can often provide little benefit if the results do not provide any actionable items; this is one of the risks of running generic models in a data lake. A pure data-mining approach would most of the time lead to an endless search for what the data really say.

Create proof of concepts More targeted approaches can generate quicker results and can give an overview of whether the approach would work or not. This will allow model developers to find out whether it is worth investing their time and money in such tools. The approach that can be followed is to create proof of concept models that can rapidly and effective address an operational challenge. This thesis has presented multiple proof of concept methodologies and tools in Chapters 3- 6 that each of them address an existing operational challenge and present the benefits and limitations of following those approaches. Similar tools and methodologies can be developed by using similar approaches to the ones presented in this thesis. It is recommended to follow the "fail fast" approach when creating proof of concept tools with different prototype stages and engagement of the relevant stakeholders at all stages of development in order to understand the usefulness of the proof of concept and tailor it towards the customers' needs or stop developing it at an early stage if deemed unhelpful.

7.2.3 Improved Decision Making

A considerable challenge with data analytics is that operational personnel and even asset managers might not understand or trust the models and the results that they have generated. This can be due to a mismatch between the needs of the different stakeholders, as well as they might not directly align with how the operational or asset management personnel arrive at decisions. Thus, engineering and R&D departments can sometimes be detached from the rest of the business. Models are usually designed from experts in computational and mathematical modelling rather than the people in the operations, or with minimum input from the operational personnel. The three recommendations below can help data analysts to generate usable models that can improve operational decision making.

Create tools that follow company's needs Newly developed tools could fail because they might not be reflecting the user's decision making norms. In order for artificial intelligence tools to be implemented and be trusted they will need to be able to mimic and decide on the behalf of their users and be in sync with the user's decision making norms. Thus, all the stakeholders need to be involved at every stage of the process in order to ensure the inputs and the assumptions are reasonable and that the tool will meet all the needs of the end users. This process is followed for the development of the O&M strategy tool, presented in Section 4.3, in order to ensure that the tool will reflect as accurately as possible the operational conditions, as well as the strategic decision making of the asset managers.

Deliver simple KPIs By necessity, data-driven models require a large number of operational data and sophisticated algorithms in order to better represent the real case scenarios. This can be challenging for the end users, as not all people are experts in advanced statistical modelling and big data processing techniques. Consequently it is important to be able to summarize key findings using simple vizual indicators and KPI metrics. These can be used to flag potential alarm instances, potential failures and to show the health index of the turbines, based on predictive and reliability tools such as the ones presented in Sections 3.3 and 3.4. The final aim of this approach is to use simple tools to deliver complex analytics.

Develop analytics capabilities to exploit data Even when the models have simple graphical interfaces, users will need to upgrade their analytical skills and view data-driven tools as the core for solving daily operational and long-term strategic problems and identifying new areas for future development. Processes will also need to be in place in order for people to take data-driven decisions, as shown with the examples of the decision support framework for ML model selection in Section 4.1 and with the data-driven framework for risk-based operations, presented in Section 4.2.

It is evident that the current way of treating data has to change in order to be able to meet the continuously increasing volume and to exploit them efficiently in order to reduce cost and increase performance of the assets. It is important for wind farm operators to have centralized IT infrastructures in an effort to foster innovation and collaboration both within and outside the company. Data should be treated centrally and not on an individual basis, creating a common space for engineers and developers to produce advanced analytics and proof of concepts. There are already standards and good practises that exist in other fields, as identified in the literature review, Section 2.3, it is just a matter of implementing them and embedding them in all of the company's processes. Moreover, relationships needs to be enhanced and silos have to be broken within companies so that engineers, analysts, asset managers and decision makers are having a common approach to using data in everyday decisions. The wind industry needs to transition to predictive and data-driven maintenance in order to realise the required operational cost reductions. Finally, digital strategies need to be enhanced, by encouraging the development of digital clones of the assets, moving towards the removal of human intervention, which will increase safety and reduce operational costs even further.

7.3 Moving Forward

The results, methods and case studies could be interpreted and used in different ways from the various offshore wind O&M stakeholders. Wind turbine OEMs will have a competitive advantage over owners and operators regarding the turbine itself, its components, the maintenance required and its reliability. On the other hand offshore wind farm owners and operators will have a better knowledge and experience relation to the turbine's foundations and the general planning of the operations.

This thesis has presented the results from an owner/ operator perspective, showcasing the challenges and the insights in the O&M of offshore wind farms. It is important to interpret the findings of this study as case studies of an existing operational assets of 2.3 MW machines that can be used for 10 MW+ machines. The results and methodologies have been designed and presented in a way so that they can be easily scaled up to larger sites and higher capacity wind turbine models. This data infrastructure will need to be implemented well in advance, before the deployment of the new Round 3 and Round 4 offshore wind farm projects in the UK and all the new developments around the globe. The suggested data and analysis framework can help in generating centralized tools that will help meet the ambitious cost reduction targets set by offshore wind farm developments and help moving towards a subsidy-free future for offshore wind.

Organizations are constantly exploring new opportunities to share data in order to foster innovation and tackle specific challenges through hackathons. Several initiatives have been launched in the last few years. Examples include ORE Catapult's Platform for Operational Data [166], hosting data from their Levenmouth turbine, ENGIE with their open data wind farm, providing SCADA data from an operational wind farm for free [167] and Orsted, which is sharing some operational data for two of its offshore wind farms [168]. Operators would also require and work towards quantifying the value of their generated data, both internally and externally. It is important to not only understand how these data can be used for data-driven decision making, but also how organizations can benefit by sharing data for a fee externally. This could not only generate a revenue stream for the data owners, but it will also foster innovation and allow more companies to enter the wind industry, which may result in more bespoke services that can offer targeted solutions to wind turbine owners, operators and OEMs.

Finally, it seems that industry players are trying to diversify by including more services in order to create new revenue streams and get a deeper understanding of the wind sector.

- Large wind operators and utilities are entering the operational services market, in order to take the most of their assets after the warranty period. Companies like EDF, EnBW and EON all have their own service arms, trying to get a better and more holistic understanding of their assets, as well as aiming to take a share from the services market, dominated by OEMs and third party providers.
- Similarly, OEMs are trying to expand more in the analytics space, in order to provide more accurate insights about their turbines to their customers and be able to provide more competitive maintenance services. Examples include the acquisitions of Utopus Insights by Vestas and of NEM solutions by Gamesa [169, 170].

7.4 Recommendations for Further Work

Within the timeline of the project it is not possible to fully evaluate all the presented case studies. Validation of the case studies and methodologies presented with more operational data is suggested, along with the following recommendations for future industrial and academic work.

7.4.1 Industrial Work

Data Sharing During this thesis and the presented work at conferences and journals, there has been a lot of work dissemination of data and insights regarding the reliability and O&M aspects of an offshore wind farm. Similar pieces of work have been conducted, mainly by the academic community, but this should be the new norm for the industry to keep expanding and innovating. Wind farm developers and owners/ operators should create an environment for open access data or data that can be shared for an administrative or infrastructure fee. This can accelerate the development of products and services from the start up or the small and medium-sized enterprises (SME) community that can help in the reduction of operational costs. It needs to be understood that data sharing does not decrease the competitive advantage of companies that are sharing their data. In reality it places them at the forefront of innovation and ahead of the curve. If the legal and contractual frameworks are in place, the companies can utilize the products and services produced by the academic and SME community for their own benefit, i.e. furthering their competitive advantage. In order to facilitate and better implement the data sharing initiatives, a framework needs to be introduced allowing the incorporation of different data sources.

Collaboration Cross-collaboration between companies in the offshore wind, or with companies in other sectors that are facing similar difficulties and are not direct competitors is very important. This can be done via common research projects either through UKRI, InnovateUK or EU Horizon funding for example or through joint industry projects (JIP), where organizations that are facing similar challenges are contributing to a common funding pot where results and data are usually anonymized by an independent party such as ORE Catapult of the Carbon Trust. Recent examples include the Horizon 2020 ROMEO project, aiming to create reliable and data-driven operations for offshore wind farms [171] and the Floating Wind JIP aiming to investigate the challenges and opportunities of developing commercial-scale floating wind farms [172]. This can increase the industry understanding around data and tackle common challenges, by creating tools, methodologies, guidelines that can later on be developed into standards. Collaborative platforms can also be created for wind farm owners to share exchange data and information more effectively. Lessons learnt workshops could be organized more often from wind farm stakeholders to share experiences and exchange ideas, like the ones created by ORE Catapult.

7.4.2 Academic Work

Academic work could focus on better developing reliability models in order to understand the physics of failure and development of digital twins of the models by combining experience of data-driven model development with physics of failure to better enhance the predictive analytics models. This could include:

- Accurate diagnosis of the turbine failures.
- Understand further the relationship between environmental conditions and turbine failures.
- Investigation of deep learning models, given the availability of large datasets.
- Treat the data for the wind farms holistically and create frameworks to analyse data from the whole wind data value chain and not only from O&M.
- Development of digitalization platforms, allowing users to share and analyse data centrally.

Bibliography

[1] International Energy Agency. *Market Report Series: Renewables 2018*. Technical Report. 2018.

[2] Wind Europe. *EU agrees 32% renewable energy target for 2030*. Available at https://windeurope.org/newsroom/news/european-institutions-strike-overnight-deal-on-32-percent-binding-renewable-energy-target/ (2018/08/17).

[3] IRENA. *Renewable Energy: A Key Climate Solution*. Abu Dhabi, UAE, 2017.

[4] Stuart Westmore. *Redcar Turbines*. Available at https://skwestmore.com/2015/06/11/redcar-turbines/ (2018/08/17).

[5] GWEC. *Annual Market Update 2017 Global Wind Report*. Technical Report. 2017.

[6] European Wind Energy Association. *Wind in our Sails The coming of Europe's offshore wind energy industry*. Technical Report. 2011.

[7] Joshua S Hill. *2018 Saw 51.3 Gigawatts Of New Wind Added Globally*. Available at https://cleantechnica.com/2019/02/27/2018-saw-51-3-gigawatts-of-new-wind-added-globally/ (2019/03/24).

[8] Brendan Coyne. *Eon takes offshore wind hit as Supreme Court rejects Robin Rigg appeal*. Available at https://theenergyst.com/offshore-wind-contract-dispute-settled/ (2018/08/17).

[9] Patrick Smith. *Vibrations strike down EDF turbines*. Available at https://www.windpoweroffshore.com/article/1302035/vibrations-strike-down-edf-turbines (2018/08/17).

[10] Anastasia Ioannou, Andrew Angus, and Feargal Brennan. "A lifecycle techno-economic model of offshore wind energy for different entry and exit instances". In: *Applied Energy* 221 (2018), pp. 406–424. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2018.03.143. URL: http://www.sciencedirect.com/science/article/pii/S0306261918304811.

[11] Joe Phillips et al. *A Guide to UK Offshore Wind Operations and Maintenance*. United Kingdom: Scottish Enterprise and The Crown Estate, 2013.

[12] Peter Tavner. *Offshore Wind Turbines: Reliability, availability and maintenance*. Energy Engineering. Institution of Engineering and Technology, 2012. URL: http://digital-library.theiet.org/content/books/po/pbrn013e.

[13] European Energy Innovation. *Innovative Technologies To Lower Cost Of Energy Offshore Wind Farms*. Available at `http://www.europeanenergyinnovation.eu/Articles/Spring-2017/Innovative-technologies-to-lower-cost-of-energy-offshore-wind-farms` (2018/09/06).

[14] Carbon Trust. *O&M moves in-house: DONG Energy*. Available at `http://www.windpower-international.com/features/featureom-moves-in-house-dong-energy-4623672/` (2019/07/10). 2015.

[15] *The Future of onshore wind operations and maitenance*. Report. Accenture, 2017.

[16] National Aeronautics and Space Administration. *RELIABILITY-CENTERED MAINTENANCE GUIDE For Facilities and Collateral Equipment*. USA, 2008.

[17] Philipp Lyding, Stefan Faulstich, and Paul Kuhn. *Establishing a common database for turbine failures*. Germany: Fraunhofer IWES, 2011.

[18] DNVGL-ST-0262. *Lifetime extension of wind turbines*. Standard. Norway: Det Norske Veritas AS, 2016.

[19] Dai Richards. *Risk Based Maintenance*. ABB Industries. 2002.

[20] DNV-RP-G101. *Risk Based Inspection of Offshore Topside Static Mechanical Equipment*. Standard. Norway: Det Norske Veritas AS, 2010.

[21] Wellington A Soares, Vanderley De Vasconcelos, and Emerson G Rabello. "Risk-Based Inspection in the Context of Nuclear Power Plants". In: *International Nuclear Atlantic Conference - INAC 2015*. Sao Paolo, Brazil, 2015. ISBN: 9788599141069.

[22] John Dalsgaard Sorensen. "Framework for risk-based planning of operation and maintenance for offshore wind turbines". In: *Wind Energy* 12.5 (2009), pp. 493–506. ISSN: 10954244. DOI: `10.1002/we.344`.

[23] *Design of Offshore Wind Turbine Structures*. Report. Norway: Balfour Beatty Rail Technologies, 2012.

[24] "The risk concept-historical and recent development trends". In: *Reliability Engineering and System Safety* 99 (2012), pp. 33–44. ISSN: 09518320. DOI: `10.1016/j.ress.2011.11.006`.

[25] ISO 31000. *Risk management- Principles and guidelines*. Standard. Geneva, Switzerland: International Organization for Standardization, 2009.

[26] Sam Mannan. *Lees' Loss Prevention in the Process Industries*. Vol. 1. 2012, pp. 618–686. ISBN: 9780123971890. DOI: `10.1016/B978-0-12-397189-0.00013-6`. URL: `http://www.elsevier.com/wps/find/bookdescription.cws%7B%5C_%7Dhome/703660/description%7B%5C#%7Ddescription%7B%5C%%7D5Cnhttp://www.sciencedirect.com/science/article/pii/B9780123971890000136`.

[27] Louis Anthony Cox. "What's wrong with risk matrices?" In: *Risk Analysis* 28.2 (2008), pp. 497–512. ISSN: 02724332. DOI: `10.1111/j.1539-6924.2008.01030.x`.

[28] P. Thomas, R.B. Bratvold, and J.E. Bickel. "The risk of using risk matrices". In: *SPE Economics and Management* 6.2 (2014). ISSN: 21501173.

[29] Terje Aven. "Practical implications of the new risk perspectives". In: *Reliability Engineering and System Safety* 115 (2013), pp. 136–145. ISSN: 09518320. DOI: `10.1016/j.ress.2013.02.020`.

[30] Terje Aven and Jan Erik Vinnem. *Risk Management: With Applications from the Offshore Petroleum Industry*. 2007, 211 pages. ISBN: 9781846286520.

[31] Fanny Douard, Célia Domecq, and William Lair. "A probabilistic approach to introduce risk measurement indicators to an offshore wind project evaluation - Improvement to an existing tool ECUME". In: *Energy Procedia*. Vol. 24. 2012, pp. 255–262. ISBN: 9781622764365. DOI: `10.1016/j.egypro.2012.06.107`.

[32] Iain Allan Dinwoodie and David McMillan. "Operational strategies for offshore wind turbines to mitigate failure rate uncertainty on operational costs and revenue". In: *IET Renewable Power Generation* 8.4 (2014), pp. 359–366. ISSN: 1752-1416. DOI: `10.1049/iet-rpg.2013.0232`. URL: `http://digital-library.theiet.org/content/journals/10.1049/iet-rpg.2013.0232`.

[33] Matthias Hofmann and Iver Bakken Sperstad. "NOWIcob-A tool for reducing the maintenance costs of offshore wind farms". In: *Energy Procedia*. Vol. 35. 2013, pp. 177–186. ISBN: 18766102 (ISSN). DOI: `10.1016/j.egypro.2013.07.171`.

[34] Ole Erik V. Endrerud, Jayantha P. Liyanage, and Nenad Keseric. "Marine logistics decision support for operation and maintenance of offshore wind parks with a multi method simulation model". In: *Proceedings - Winter Simulation Conference*. Vol. 2015-January. 2015, pp. 1712–1722. ISBN: 9781479974863. DOI: `10.1109/WSC.2014.7020021`.

[35] Tom Obdam et al. "O&M Cost Estimation & Feedback of Operational Data". In: *Wind Farm*. Rijeka: IntechOpen, 2011. Chap. 2. DOI: `10.5772/17011`. URL: `https://doi.org/10.5772/17011`.

[36] "Verification and Benchmarking Methodology for O&M Planning and Optimization Tools in the Offshore Renewable Energy Sector". In: *37th International Conference on Ocean, Offshore and Arctic Engineering*. Madrid, Spain, 2015.

[37] Iain Dinwoodie et al. "Reference Cases for Verification of Operation and Maintenance Simulation Models for Offshore Wind Farms". In: *Wind Engineering* 39.1 (2015), pp. 1–14. ISSN: 0309-524X. DOI: `10.1260/0309-524X.39.1.1`. URL: `http://journals.sagepub.com/doi/10.1260/0309-524X.39.1.1`.

[38] Iver Bakken Sperstad et al. "Testing the robustness of optimal access vessel fleet selection for operation and maintenance of offshore wind farms". In: *Ocean Engineering* 145 (2017), pp. 334–343. ISSN: 00298018. DOI: `10.1016/j.oceaneng.2017.09.009`.

[39] Wenxian Yang et al. "Wind turbine condition monitoring: Technical and commercial challenges". In: *Wind Energy* 17.5 (2014), pp. 673–693. ISSN: 10991824. DOI: `10.1002/we.1508`. arXiv: `arXiv:1006.4405v1`.

[40] John D. Sørensen and Jens N. Sørensen, eds. Woodhead Publishing, 2011, pp. 585–598. DOI: `https://doi.org/10.1533/9780857090638.backmatter`.

[41] Germanischer Lloyd. *Guideline for the Certification of Condition Monitoring Systems for Wind Turbines*. Standard. Germany: Germanischer Lloyd, 2007.

[42] DNV GL. *DNVGL-SE-0439 Certification of condition monitoring*. Standard. 2016.

[43] Katharina Fischer, Francois Besnard, and Lina Bertling. "Reliability-centered maintenance for wind turbines based on statistical analysis and practical experience". In: *IEEE Transactions on Energy Conversion* 27.1 (2012), pp. 184–195. ISSN: 08858969. DOI: 10.1109/TEC.2011.2176129.

[44] Grahm Juhl. *TCM User Manual*. Confidential. 2012.

[45] Strainstall. *Installation of Structural Health Monitoring System*. Internal Report. 2014.

[46] DNV-OS-J101. *Design of Offshore Wind Turbine Structures*. Standard. Norway: Det Norske Veritas AS, 2014.

[47] Trinh Hoang Nguyen et al. "A framework for data integration of offshore wind farms". In: *Renewable Energy* 60 (2013), pp. 150–161. ISSN: 09601481. DOI: 10.1016/j.renene.2013.05.002. URL: http://dx.doi.org/10.1016/j.renene.2013.05.002.

[48] Trinh Hoang Nguyen et al. "Offshore Wind Metadata Management". In: *Int. J. Metadata Semant. Ontologies* 9.4 (Oct. 2014), pp. 333–349. ISSN: 1744-2621. DOI: 10.1504/IJMSO.2014.065445. URL: http://dx.doi.org/10.1504/IJMSO.2014.065445.

[49] Paulo B. Goes. "Big Data and IS Research". In: *MIS Quarterly* 47 (3 2014), pp. 3–8.

[50] Ibrahim Abaker Targio Hashem et al. "The rise of big data on cloud computing: Review and open research issues". In: *Information Systems* 47 (2015), pp. 98–115. ISSN: 0306-4379. DOI: https://doi.org/10.1016/j.is.2014.07.006. URL: http://www.sciencedirect.com/science/article/pii/S0306437914001288.

[51] J. Helsen, G. D. Sitter, and P. J. Jordaens. "Long-Term Monitoring of Wind Farms Using Big Data Approach". In: *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*. 2016, pp. 265–268. DOI: 10.1109/BigDataService.2016.49.

[52] Rick Cattell. "Scalable SQL and NoSQL data stores". In: *ACM SIGMOD Record* 39.4 (2011), p. 12. ISSN: 01635808. DOI: 10.1145/1978915.1978919.

[53] Feng Tian et al. "Implementing a scalable XML publish/subscribe system using relational database systems". In: *SIGMOD 04 Proceedings of the 2004 ACM SIGMOD international conference on Management of data* (2004), pp. 479–490. ISSN: 07308078. DOI: 10.1145/1007568.1007623. URL: http://portal.acm.org/citation.cfm?id=1007568.1007623.

[54] British Standard Institute. *BS 5760-4:2003 Reliability of systems, equipment and components. Guide to the specification of dependability requirements*. Standard. 2003.

[55] F. Spinato et al. “Reliability of wind turbine subassemblies”. In: *IET Renewable Power Generation* 3.4 (Dec. 2009), pp. 387–401. ISSN: 1752-1416. DOI: 10.1049/iet-rpg.2008.0060.

[56] IEC 60050-191:1990. *International Electrotechnical Vocabulary*. Standard. Geneva, CH: International Electrotechnical Commission, 1990.

[57] OREDA. “OREDA Offshore Reliability Data Handbook”. In: Det Norske Veritas Industri Norge AS Hovik, 2002.

[58] Michael Wilkinson et al. “Long-Term Monitoring of Wind Farms Using Big Data Approach”. In: *European Wind Energy Association Annual Event*. Brussels, 2011.

[59] Iain Dinwoodie et al. “Reference Cases for Verification of Operation and Maintenance Simulation Models for Offshore Wind Farms”. In: *Wind Engineering* 39.1 (2015), pp. 1–14. DOI: 10.1260/0309-524X.39.1.1. eprint: https://doi.org/10.1260/0309-524X.39.1.1. URL: https://doi.org/10.1260/0309-524X.39.1.1.

[60] F. Spinato et al. “Reliability of wind turbine subassemblies”. In: *IET Renewable Power Generation* 3.4 (Dec. 2009), pp. 387–401. ISSN: 1752-1416. DOI: 10.1049/iet-rpg.2008.0060.

[61] James Carroll, Alasdair McDonald, and David McMillan. “Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines”. In: *Wind Energy* 19.April 2013 (2016), pp. 1107–1119. ISSN: 1099-1824. DOI: 10.1002/we. eprint: arXiv:1006.4405v1. URL: http://onlinelibrary.wiley.com/doi/10.1002/we.1608/full.

[62] James Carroll et al. “Offshore wind turbine sub-assembly failure rates through time”. In: *Scientific Proceedings*. European Wind Energy Association, Nov. 2015, pp. 112–116. ISBN: 9782930670003.

[63] P J Tavner, J Xiang, and F Spinato. “Reliability Analysis for Wind Turbines”. In: *Wind Energy* 10.July 2006 (2007), pp. 1–18. DOI: 10.1002/we204.

[64] J. Bard, S. Faulstich, and P. Lyding. “The German Wind Turbine Reliability Database- WME”. In: *8th Deep Sea Offshore Wind R&D conference (EERA DeepWind’2011)*. 2011.

[65] P. J. Tavner and F. Spinato. “Reliability of Different Wind Turbine Concepts with Relevance to Offshore Application”. In: *European Wind Energy Conference, Brussels* (2008).

[66] Peter Tavner et al. “Influence of Wind Speed on Wind Turbine Reliability”. In: *Wind Engineering* 30.1 (2006), pp. 55–72. ISSN: 0309-524X. DOI: 10.1260/030952406777641441.

[67] Anders Stenberg. “Analys av Vindkraftsstatistik i Finland”. MA thesis. Stockholm: Aalto University, 2010.

[68] L. Pettersson et al. *RAMS- database for Wind Turbines- Pre-study*. Report. Elforsk, 2010.

[69] *WIND-POOL Wind Energy Information Data Pool*. Report. Kassel, Germany: Fraunhofer IWES, 2016.

[70] Gamesa. *Reliability focused research on optimizing Wind Energy systems design, operation and maintenance: Tools, proof of concepts, guidelines & methodologies for a new generation*. Spain: RELIAWIND, 2007.

[71] Charles Carter et al. *Continuous Reliability Enhancement for Wind (CREW)-Program Update*. Report. USA: Sandia National Laboratories, 2016.

[72] *System Performance, Availability and Reliability Trend Analysis- Portfolio Review 2016*. Report. UK: ORE Catapult, 2017.

[73] James Carroll, Alasdair McDonald, and David McMillan. "Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines". In: *Wind Energy* 19.April 2013 (2016), pp. 1107–1119. ISSN: 1099-1824. DOI: `10.1002/we`. eprint: `arXiv:1006.4405v1`. URL: `http://onlinelibrary.wiley.com/doi/10.1002/we.1608/full`.

[74] Y. Feng, Peter J. Tavner, and H. Long. "Early experiences with UK round 1 offshore wind farms". In: *Proceedings of the Institute of Civil Engineers* 163.4 (2010), pp. 167–181. DOI: `10.1680/ener.2010.163.4.167`.

[75] Christopher J. Crabtree. "Operational and Reliability Analysis of Offshore Wind Farms". In: *European Wind Energy Association Annual Event*. Copenhagen, Denmark, 2012.

[76] Rebecca Martin. "Sensitivity Analysis of Offshore Wind Farm Availability and Operations & Maintenance Costs Subject to Uncertain Input Factors". PhD thesis. University of Strathclyde, 2015.

[77] Johan Ribrant and Lina Margareta Bertling. "Survey of failures in wind power systems with focus on Swedish wind power plants during 1997-2005". In: *IEEE Transactions on Energy Conversion* 22.1 (2007), pp. 167–173. ISSN: 08858969. DOI: `10.1109/TEC.2006.889614`.

[78] Peter Tavner et al. "Influence of Wind Speed on Wind Turbine Reliability". In: *Wind Engineering* 30.1 (2006), pp. 55–72. DOI: `10.1260/030952406777641441`.

[79] P. J. Tavner et al. "Study of weather and location effects on wind turbine failure rates". In: *Wind Energy* 16.2 (2013), pp. 175–187. ISSN: 10954244. DOI: `10.1002/we.538`. arXiv: `arXiv:1006.4405v1`.

[80] S. Faulstich, P. Lyding, and P. J. Tavner. "Effects of wind speed on wind turbine availability". In: *European Wind Energy Conference and Exhibition 2011, EWEC 2011*. 2011, pp. 85–88. ISBN: 9781618399915. URL: `http://www.scopus.com/inward/record.url?eid=2-s2.0-84870153053%7B%5C&%7DpartnerID=tZOtx3y1`.

[81] Maik Reder, Nurseda Y. Yürüsen, and Julio J. Melero. "Data-driven learning framework for associating weather conditions and wind turbine failures". In: *Reliability Engineering and System Safety* 169 (2018), pp. 554–569. ISSN: 09518320. DOI: `10.1016/j.ress.2017.10.004`.

[82] Matti Niclas Scheu et al. "Influence of statistical uncertainty of component reliability estimations on offshore wind farm availability". In: *Reliability Engineering and System Safety* 168 (2017), pp. 28–39. ISSN: 09518320. DOI: `10.1016/j.ress.2017.05.021`.

[83] D.J. Willis et al. "Wind energy research: State-of-the-art and future research directions". In: *Renewable Energy* 125 (2018), pp. 133–154. ISSN: 0960-1481. DOI: `https://doi.org/10.1016/j.renene.2018.02.049`. URL: `http://www.sciencedirect.com/science/article/pii/S0960148118301939`.

[84] Yingning Qiu et al. "Wind turbine SCADA alarm analysis for improving reliability". In: *Wind Energy* 15.8 (2012), pp. 951–966. DOI: `10.1002/we.513`.

[85] Berthold Hahn et al. "Recommended practices for wind farm data collection and reliability assessment for O&M optimization". In: *Energy Procedia*. Vol. 137. 2017, pp. 358–365. DOI: `10.1016/j.egypro.2017.10.360`.

[86] IEC 61400-25-1:2006. *Wind turbines- Part 25: communications for monitoring and control of wind power plants*. Standard. Geneva, Switzerland: International Electrotechnical Commission, 2006.

[87] RDS-PP. *pplication Guideline Part 32: Wind Power Plants*. Standard. Essen, Germany: VGB PowerTech e.V., 2014.

[88] ISO 14224. *Petroleum, petrochemical and natural gas industries- Collection and exchange of reliability and maintenance data for equipment*. Standard. Geneva, Switzerland: International Organization for Standardization, 2016.

[89] *Technical Guidelines for Power Generating Units Part 7: Maintenance of power plants for renewable energy- Category D2: State-Event-Cause code (ZEUS)*. Report. Berlin, Germany: FGW e.V., 2013.

[90] *Generating Availability Data System Mandatory Reporting of Conventional Generation Performance Data*. Report. Atlanta, GA: North American Electric Reliability Corporation, 2011.

[91] IEC 61400-26-2:2014. *Wind turbines- Part 26-2: Production-based availability for wind turbines*. Standard. Geneva, Switzerland: International Electrotechnical Commission, 2014.

[92] Mark Mosley et al. *The DAMA Guide to The Data Management Body of Knowledge*. 2009, p. 406. ISBN: 9780977140084.

[93] *Manual and guides 73- Guidelines for a Research Data Management Plan*. Report. Paris, France: Intergovernmental Oceanographic Commission of UNESCO, 2016.

[94] ISO/IEC 27001. *Information technology- Security techniques- Information security management systems- Requirements*. Standard. Geneva, Switzerland: International Organization for Standardization, 2013.

[95] IEC 61850. *Communication networks and systems for power utility automation- Part 7-3: Basic communication structure- Common data classes*. Standard. Geneva, Switzerland: International Electrotechnical Commission, 2010.

[96] ISO/IEC 10032. *Information technology- Reference Model of Data Management*. Standard. Geneva, Switzerland: International Organization for Standardization, 2003.

[97] ISO 15926-4. *Industrial automation systems and integration- Integration of life-cycle data for process plants including oil and gas production facilities- Part 4: Initial reference data*. Standard. Geneva, Switzerland: International Organization for Standardization, 2007.

[98] N Turner. *Taming the waves: the foundations of effective data management*. DAMA. 2016.

[99] *Buidling trust in analytics- Breaking the cycle of mistrust in D&A*. Report. Switzerland: KPMG International Cooperative, 2016.

[100] *Going Beyond the data: Achieving actionable insights with data and analytics*. Report. Switzerland: KPMG International Cooperative, 2014.

[101] James J. Rooney and Lee N. Vanden Heuvel. "Root Cause Analysis For Beginners". In: *Quality Basics* (July 2004), pp. 45–53. URL: https://www.env.nm.gov/aqb/Proposed_Regs/Part_7_Excess_Emissions/NMED_Exhibit_18-Root_Cause_Analysis_for_Beginners.pdf.

[102] IBM. *Predictive analytics*. Available at https://www.ibm.com/analytics/predictive-analytics (2018/08/28).

[103] SAS Institute. *Predictive analytics and machine learning*. Available at https://www.sas.com/en_gb/insights/articles/analytics/a-guide-to-predictive-analytics-and-machine-learning.html (2018/08/28).

[104] Kevin P. Murphy. *Machine Learning A Probabilistic Perspective*. Cambridge, Mass: MIT Press, 2013.

[105] F. Cheng et al. "Rotor-Current-Based Fault Diagnosis for DFIG Wind Turbine Drivetrain Gearboxes Using Frequency Analysis and a Deep Classifier". In: *IEEE Transactions on Industry Applications* 54.2 (2018), pp. 1062–1071. ISSN: 0093-9994. DOI: 10.1109/TIA.2017.2773426.

[106] Yancai Xiao et al. "Research on Misalignment Fault Isolation of Wind Turbines Based on the Mixed-Domain Features". In: *Algorithms* 10.2 (2017). ISSN: 1999–4893. DOI: 10.3390/a10020067. URL: http://www.mdpi.com/1999-4893/10/2/67.

[107] Sheng-Yu Niu, Bo-Wen Liu, and Xin-Yan Zhang. "Research on Fault Warning of doubly Fed Wind Power Generator based on LS-SVM". In: *the 3rd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2017)*. 2017, pp. 2621–2625. DOI: 10.2991/eeeis-17.2017.23.

[108] Tinghui Ouyang, Andrew Kusiak, and Yusen He. "Predictive model of yaw error in a wind turbine". In: *Energy* 123 (2017), pp. 119–130. ISSN: 03605442. DOI: 10.1016/j.energy.2017.01.150.

[109] Taylor Regan, Christopher Beale, and Murat Inalpolat. "Wind Turbine Blade Damage Detection Using Supervised Machine Learning Algorithms". In: *Journal of Vibration and Acoustics* 139.6 (2017), pp. 1–14. DOI: 10.1115/1.4036951.

[110] Deng-Fa Lin, Po-Hung Chen, and Mike Williams. "Measurement and Analysis of Current Signals for Gearbox Fault Recognition of Wind Turbine". In: *Measurement Science Review* 13.2 (2013), pp. 89–93. URL: https://content.sciendo.com/view/journals/msr/13/2/article-p89.xml.

[111] Georgios Alexandros Skrimpas et al. “Detection of icing on wind turbine blades by means of vibration and power curve analysis”. In: *Wind Energy* 19.10 (2016), pp. 1819–1832. ISSN: 10991824. DOI: 10.1002/we.1952. eprint: arXiv:1006.4405v1.

[112] Hong Shan Zhao et al. “Research on early Fault Prediction of Wind Turbine Gearbox”. In: *Advanced Materials Research* 608 - 609 (2012), pp. 522–528.

[113] Andrew Kusiak and Anoop Verma. “Analyzing bearing faults in wind turbines: A data-mining approach”. In: *Renewable Energy* 48 (2012), pp. 110–116. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2012.04.020. URL: http://www.sciencedirect.com/science/article/pii/S0960148112002613.

[114] Siwen Zhu and Bin Jiao. “Fault Diagnosis of Gearbox of Wind Turbine Based on Improved Decision Tree Algorithm”. In: *Advances in Intelligent Systems Research* 134 (2017), pp. 329–331.

[115] A. Joshuva. and V. Sugumaran. “A data driven approach for condition monitoring of wind turbine blade using vibration signals through best-first tree algorithm and functional trees algorithm: A comparative study”. In: *ISA Transactions* 67 (2017), pp. 160–172. ISSN: 00190578. DOI: 10.1016/j.isatra.2017.02.002.

[116] L. Wang et al. “Wind Turbine Gearbox Failure Identification With Deep Neural Networks”. In: *IEEE Transactions on Industrial Informatics* 13.3 (June 2017), pp. 1360–1368. ISSN: 1551-3203. DOI: 10.1109/TII.2016.2607179.

[117] Jae Yoon et al. “Vibration-based wind turbine planetary gearbox fault diagnosis using spectral averaging”. In: *Wind Energy* 19.9 (), pp. 1733–1747. DOI: 10.1002/we.1940. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/we.1940. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1940.

[118] Pedro Ponce, Arturo Molina, and Brian MacCleery. “Integrated Intelligent Control and Fault System for Wind Generators”. In: *Intelligent Automation & Soft Computing* 19.3 (2013), pp. 373–389. DOI: 10.1080/10798587.2013.778038. URL: https://doi.org/10.1080/10798587.2013.778038.

[119] Nikolaos Dervilis et al. “Machine Learning Applications for a Wind Turbine Blade under Continuous Fatigue Loading”. In: *Key Engineering Materials* 588 (2013), pp. 166–174. ISSN: 1662-9795. DOI: 10.4028/www.scientific.net/KEM.588.166. URL: http://www.scientific.net/KEM.588.166.

[120] X. Wang, J. Zhou, and Peng Guo. “Wind turbine gearbox forecast using Gaussian process model”. In: *The 26th Chinese Control and Decision Conference (2014 CCDC)*. May 2014, pp. 2621–2625. DOI: 10.1109/CCDC.2014.6852616.

[121] Zhipeng Feng and Ming Liang. “Complex signal analysis for planetary gearbox fault diagnosis via shift invariant dictionary learning”. In: *Measurement* 90 (2016), pp. 382–395. ISSN: 0263-2241. DOI: https://doi.org/10.1016/j.measurement.2016.04.078. URL: http://www.sciencedirect.com/science/article/pii/S0263224116301580.

[122] Jialin Tang et al. "A Pattern Recognition Approach to Acoustic Emission Data Originating from Fatigue of Wind Turbine Blades". In: *Sensors* 17.11 (2017), p. 2507. ISSN: 1424-8220. DOI: 10.3390/s17112507. URL: http://www.mdpi.com/1424-8220/17/11/2507.

[123] BS ISO 17359:2018. *Condition monitoring and diagnostics of machines. General guidelines.* Standard. UK: British Standards Institution, 2018.

[124] RSS Marine. *RSS Marine Notice- No36:2013- Teesside Offshore Wind Farm-Construction Information.* Sept. 1, 2013. URL: http://www.pdports.co.uk/Documents/Navigational%20Information/Notices-to-Mariners/Teesside%20Offshore%20Wind%20Farm%20Construction%20Notice%2036.pdf (visited on 05/01/2018).

[125] Tuhfe Göçmen and Gregor Giebel. "Estimation of turbulence intensity using rotor effective wind speed in Lillgrund and Horns Rev-I offshore wind farms". In: *Renewable Energy* 99 (2016), pp. 524–532. ISSN: 18790682. DOI: 10.1016/j.renene.2016.07.038.

[126] IEC 61400-12-2:2013. *Wind turbines - Part 12-2: Power performance of electricity-producing wind turbines based on nacelle anemometry.* Standard. Geneva, Switzerland: International Electrotechnical Commission, 2013.

[127] Yanhui Feng et al. "Monitoring wind turbine gearboxes". In: *Wind Energy* 16.5 (2013), pp. 728–740. DOI: 10.1002/we.1521. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/we.1521. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1521.

[128] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12 (Nov. 2011), pp. 2825–2830. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=1953048.2078195.

[129] Lipo Wang. *Support Vector Machines: Theory and Applications.* Vol. 177. Netherlands: Springer, 2005.

[130] J. Ross Quinlan. "Induction of decision trees". In: *Machine learning* 1.1 (1986), pp. 81–106.

[131] L. Rokach and O. Maimon. "Top-down induction of decision trees classifiers - a survey". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 35.4 (Nov. 2005), pp. 476–487. ISSN: 1094-6977. DOI: 10.1109/TSMCC.2004.843247.

[132] George E. P. Box et al. *Time Series Analysis: Forecasting and Control.* 5th. 2015, pp. 55–71.

[133] MathWorks Inc. *Machine Learning with MATLAB.* Tech. rep. 2016.

[134] W A Soares, V D Vasconcelos, and E G Rabello. "Risk-based inspection in the context of nuclear power plants". In: *Proceedings of the International Nuclear Atlantic Conference - INAC.* Sao Paolo, Brazil, Oct. 2015.

[135] J. Moubray. *Reliability-centered maintenance.* New York: Industrial Press, 1997.

[136] Rommert Dekker. “Integrating optimisation, priority setting, planning and combining of maintenance activities”. In: *European Journal of Operational Research* 82.2 (1995). OR Models for Maintenance Management and Quality Control, pp. 225–240. ISSN: 0377-2217. DOI: https://doi.org/10.1016/0377-2217(94)00260-J.

[137] R.E. Wildeman, R. Dekker, and A.C.J.M. Smit. “A dynamic policy for grouping maintenance activities”. In: *European Journal of Operational Research* 99.3 (1997), pp. 530–551. ISSN: 0377-2217. DOI: https://doi.org/10.1016/S0377-2217(97)00319-6.

[138] R.E. Wildeman and R. Dekker. “Dynamic Influences in Multi-component Maintenance”. In: *Quality and Reliability Engineering International* 13 (1997), pp. 199–207.

[139] Boeing. *Aero QRT 03*. Tech. rep. 2011.

[140] Boeing. *Aero QRT 01*. Tech. rep. 2008.

[141] NASA. *Reliability-centred maintenance guide*. Tech. rep. 2008.

[142] DNV. *Recommended proactive DNV-RP-G101 - risk-based inspection of offshore topsides static mechanical equipment*. Standard. 2010.

[143] Chrysler Corporation, Ford Motor Company, and General Motors Corporation. *Potential failure mode and effects analysis (FMEA) - reference manual*. Tech. rep. 1995.

[144] J Paterson et al. “Assessing marine operations with a Markov-switching autoregressive metocean model”. In: *Proc IMechE Part M: J Engineering for the Maritime Environment (Under Review)* ().

[145] IEC TS 61400-26-1:2011. *Wind turbines - Part 26-1: Time-based availability for wind turbine generating systems*. Standard. Geneva, Switzerland: International Electrotechnical Commission, 2011.

[146] IEC TS 61400-26-2:2014. *Wind turbines - Part 26-2: Production-based availability for wind turbine generating systems*. Standard. Geneva, Switzerland: International Electrotechnical Commission, 2014.

[147] G Wilson and D McMillan. “Quantifying the Impact of Wind Speed on Wind Turbine Component Failure Rates”. In: *European Wind Energy Conference and Exhibition*. 2014.

[148] European Wind Energy Association. *Wind in our Sails The coming of Europe’s offshore wind energy industry*. Technical Report. 2011.

[149] IEC 61400-1-3:2005. *Wind turbines - Part 1: Design requirements*. Standard. Geneva, Switzerland: International Electrotechnical Commission, 2005.

[150] Iain Staffell and Richard Green. “How does wind farm performance decline with age?” In: *Renewable Energy* 66 (2014), pp. 775–786. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2013.10.041. URL: http://www.sciencedirect.com/science/article/pii/S0960148113005727.

[151] Kevin Leahy et al. "Automatically Identifying and Predicting Unplanned Wind Turbine Stoppages Using SCADA and Alarms System Data: Case Study and Results". In: *Journal of Physics: Conference Series* 926 (Nov. 2017), p. 012011. DOI: 10.1088/1742-6596/926/1/012011.

[152] *System Performance, Availability and Reliability Trend Analysis - SPARTA: 2017/18 Portfolio Review*. Report. UK: ORE Catapult, 2018.

[153] R Jansen and R Karki. *Sustainable energy optimisation in a smart microgrid*. Singapore: Springer, 2017, pp. 111–132.

[154] R. A. Gonzalez-Fernandez and A. M. Leite da Silva. "Reliability Assessment of Time-Dependent Systems via Sequential Cross-Entropy Monte Carlo Simulation". In: *IEEE Transactions on Power Systems* 26.4 (Nov. 2011), pp. 2381–2389. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2011.2112785.

[155] Chia Chen Ciang, Jung-Ryul Lee, and Hyung-Joon Bang. "Structural health monitoring for a wind turbine system: a review of damage detection methods". In: *Measurement Science and Technology* 19.12 (Oct. 2008), p. 122001. DOI: 10.1088/0957-0233/19/12/122001. URL: https://doi.org/10.1088%2F0957-0233%2F19%2F12%2F122001.

[156] A. Niesłony. "Determination of fragments of multiaxial service loading strongly influencing the fatigue of machine components". In: *Mechanical Systems and Signal Processing* 23 (Nov. 2009), pp. 2712–2721. DOI: 10.1016/j.ymssp.2009.05.010.

[157] M A Miner. "Cumulative Damage in Fatigue". In: *Journal of Applied Mechanics* 3 (1945), pp. 159–164.

[158] Ujjwal R. Bharadwaj, Vadim V. Silberschmidt, and John B. Wintle. "A risk based approach to asset integrity management". In: *Journal of Quality in Maintenance Engineering* 18.4 (2012), pp. 417–431. DOI: 10.1108/13552511211281570.

[159] *API 570 - piping inspection code: in-service inspection, rating, repair and alteration of piping systems*. Report. Washington, DC: API Publishing Services, 2016.

[160] EDF Energy Renewables. *Teesside offshore wind farm - corrosion and cathodic protection*. Report. UK: unpublished confidential document, 2012.

[161] TWI Limited. *Guidelines for use of statistics for analysis of sample inspection of corrosion*. Report. Norwich, UK: Crown, 2002.

[162] Andrew McAfee and Erik Brynjolfsson. *Big Data: The Management Revolution*. Harvard Business Publishing, Oct. 2011. URL: https://hbr.org/2012/10/big-data-the-management-revolution.

[163] International Energy Agency. *Digitalization & Energy*. Technical Report. 2017.

[164] Thomas Pump. *E.ON Climate & Renewables Internet of Wind Setting the Pace in Renewable Energy*. Available at https://www.industryofthingsworld.com/wp-content/uploads/2018/10/Day-01_Stream-4_Thomas-Pump_E.ON_.pdf (2019/05/30). 2018.

[165] Marie-Luce Picard, Jean-Marc Rangod, and Christophe Salperwyck. *A Data Lake and a Data Lab to Optimize Operations and Safety within a nuclear fleet*. Available at `https://www.slideshare.net/HadoopSummit/a-data-lake-and-a-data-lab-to-optimize-operations-and-safety-within-a-nuclear-fleet` (2019/05/30). Hadoop Summit, San Jose. 2016.

[166] ORE Catapult. *Platform for Operational Data (POD)*. Available at `https://ore.catapult.org.uk/operation-performance/strategic-programmes/pod/` (2019/07/10). 2018.

[167] ENGIE. *Welcome to ENGIEs first open data windfarm*. Available at `https://opendata-renewables.engie.com/pages/home/` (2019/07/10). 2017.

[168] Orsted. *Offshore Operational Data*. Available at `https://orsted.com/en/Our-business/Offshore-wind/Offshore-Operational-Data` (2019/07/10). 2018.

[169] Orsted. *Vestas to acquire Utopus Insights, a leading energy analytics and digital solutions company*. Available at `https://www.vestas.com/en/media/company-news?n=1679383` (2019/07/10). 2018.

[170] Paul Dvorak. *GAMESA boosts its maintenance business with NEM Solutions*. Available at `https://www.windpowerengineering.com/business-news-projects/gamesa-boosts-maintenance-business-nem-solutions/` (2019/07/10). 2017.

[171] ROMEO. *What is ROMEO project*. Available at `https://www.romeoproject.eu/` (2019/07/10). 2019.

[172] Carbon Trust. *Floating Wind Joint Industry Project (JIP)*. Available at `https://www.carbontrust.com/offshore-wind/floating/floating-wind-jip/` (2019/07/10). 2019.

Appendix A

Research Outputs

A.1 Journal Papers

Koltsidopoulos Papatzimos, A., Thies, P.R., Dawood, T. (2019). Offshore Wind Turbine Fault Alarm Prediction. *Wind Energy*, **22**(12), 1779-1788.

Koltsidopoulos Papatzimos, A., Thies, P.R., Dawood, T. (2019). Data-Driven Operational Reliability Assessment for Offshore Wind Farms. *Reliability Engineering & System Safety*, (resubmitted following review).

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2018). Data Insights from an Offshore Wind Turbine Gearbox Replacement. *Journal of Physics: Conference Series*, **1104**, 012003.

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2018). Cost effective, risk-based inspection planning for offshore wind farms. *Insight - Non-Destructive Testing and Condition Monitoring*, **60**(6), 299-305.

- **Re-published in:** (2018) *NDT World*. **21**(3). pp. 25-32. DOI:10.12737/article_5b8cf232b2c920.76414744 (in Russ.).

A.2 Conference Papers

Koltsidopoulos Papatzimos, A., Thies, P.R., Lonchampt, J., Joly, A., Dawood , T., (2019) Data Informed Lifetime Reliability Prediction for Offshore Wind

Farms. In *Proc. of 2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 17-19 Jun, San Francisco, CA.

Lonchampt, J., Joly , A., Dawood , T., Koltsidopoulos Papatzimos, A. (2019) An integrated asset management model for offshore wind turbines. In *Proc. of the Twenty-ninth (2019) International Ocean and Polar Engineering Conference*, 16-21 Jun, Honolulu, USA.

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2018) Operational Data to Maintenance Optimization: Closing the Loop in Offshore Wind O&M. In *Proc. of the ASME 2018 1st International Offshore Wind Technical Conference (IOWTC)*, 4-7 Nov, San Francisco, USA.

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2017). On Risk-based Inspections for Offshore Wind Farms: A Case Study. In *Proc. of 56th Annual British Conf. of Non-Destructive Testing (NDT 2017)*, The British Institute of Non-Destructive Testing (BINDT), 4-7 Sept, Telford, UK.

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2017). An Integrated Data Management Approach for Offshore Wind Turbine Failure Root Cause Analysis. In *Proc. of 36th International Conference on Ocean, Offshore and Arctic Engineering (OMAE)*, OMAE2017-62279, 25–30 June, Trondheim, Norway.

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2017). Towards automated and integrated data collection-standardising workflow processes for the offshore wind industry. In *Proc. of Offshore Wind Energy 2017*, 6-8 June, London, UK.

A.3 Additional Conference Presentations

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2018). Wind Turbine Component Reliability Assessment. *BIS 6th Wind Power Big Data and IoT Forum*, 4-5 Dec, Berlin, Germany.

Koltsidopoulos Papatzimos A. (2018). Lessons learnt from the application of data management and analytics techniques in offshore wind farms. *IEA Wind Topical Expect Meeting #92 Wind Energy & Digitalization*, 4-5 Oct, Dublin, Ireland.

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2018). Reducing OPEX Through Integrated Data Management. *IMechE Onshore and Offshore Wind Engineering 2018*, 23-24 May, Manchester, UK.

Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R. (2017). Making the most of your SCADA data. *WindPower Monthly: Wind Farm Data Management & Analysis Forum*, 17-19 Oct, Hamburg, Germany.