Harnessing Data for Wind Turbines: Machine Learning Digital-Enabled Asset Management Strategies

CDT Thesis

Innes Murdo Black
CDT Student
Athanasios Kolios
Supervisor
Renewable Energy Marine Structures
Naval Architecture, Ocean
Marine Engineering
University of Strathclyde, Glasgow

June 12, 2024

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

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by the University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Abstract

As interests in offshore wind farms continue to grow, so does the demand to reduce the cost of energy (COE). Maintenance cost and downtime can reduce the COE through greater information on offshore wind assets concerning the operational loads and structural integrity. This has had a significant impact on the interests of digital-enabled asset management (DEAM) using digital twins. Digital twins' technologies can replicate operational assets computationally, providing more information and increasing one's confidence in operations and maintenance (O&M). DEAM is a multi-disciplinary field and making advances in this field requires aspects of multiple modelling domains, this thesis aims to develop this and help aid in the future of DEAM. The work carried out in the thesis has been validated against operational data recordings from offshore structures. This provides value and confidence to the results of the state-of-the-art models for real-world engineering systems.

This research presents a portfolio of four research areas that have been published in a variety of peer-reviewed journal articles and conference papers. The areas are: 1) A proposal for standardisation of pre-processing data. Current standards have not addressed how to deal with data for machine learning, and this paper aims to begin this discussion with an example. This work implements a trend condition monitoring model that makes predictions on the power of an offshore structure using supervisory control and data acquisition (SCAD) data. There are 5 different machine learning (ML) models used and the data is validated using unused data with the modelling errors quantified. 2) A novel approach to dealing with the limitations of small data sets. This is an innovative way of transferring information from a homogeneous population to increase the accuracy of an artificial neural network (ANN). The ML model is a comparison of a conventional ANN compared to the proposed hard-parameter transfer ANN model. The ML model makes a classification of the error signature from the gearbox using both SCADA data and condition monitoring system (CMS) data. The validation of the comparison uses unseen data during the training process and the errors are measured. 3) Is a case study on Wikinger offshore wind farm population homogeneity where the operational and environmental conditions are compared for all three wind turbines. This case study provides a framework to follow when investigating an offshore wind farm population. This uses operational data from three wind turbines with both SCADA, CMS data, and processed data from RAMBOLL. The outcomes from this paper are used to determine the type of ML model used in the last study. 4) Is the model development of a population-based structural health monitoring (PBSHM) model. This study investigates three domain adaptation techniques suited to strong homogeneous populations. The ML model takes SCADA data as an input and predicts the damage equivalent moments (DEM) on the jacket foundation structure. To validate the PBSHM model data from a structure that is not used during the training of the model is used to quantify the precision of the model.

The individual contributions of the developments in each of the constituent areas relate to an overall improvement in modelling approaches that are necessary for DEAM and aid in the realisation of true digital twins. All the areas relate to offshore wind ML and are related to operational data. The link between the measured data and the individual models aid in gaining more information and greater insights into the O&M.

Contents

List of Figures

List of Tables

List of Tables

Chapter 1

Introduction

1.1 Background Scenario and Motivation for Research

The wind industry is rapidly growing, specifically offshore wind and has established itself as the leading provider of sustainable energy [54]. This is not just the case in Europe, where it originated [117] [74], but also in emerging markets such as North America [124], [8] and in the Asian markets too [222], [9], [110]. Projections from the industry show that global offshore wind will reach 130-140 GW by 2030 and potentially 1500 by 2050 [98]. The rapid growth of the wind industry increases the competition, and as consequence, this will drive developers to reduce the cost of energy for market dominance in current, emerging, and new markets.

Three key performance indicators (KPI) that are necessary to achieve market dominance: are the levelised cost of energy (LCOE). This is the sum of the capital expenses (CAPEX) to manufacture and install the asset and the operational expenses (OPEX) for running and maintaining the asset. The values of these KPIs are dependent on the specific project, the design philosophy, and the installation environment. In an environment characterised by a historical adherence to conservative practices, there is a commitment to risk reduction through the integration of innovative technologies. This approach involves a hereditary inclination toward caution, manifesting in the implementation of increased safety factors for components and a more frequent schedule of inspections to uphold continual safety standards. The philosophy of conservatism is necessary for both ensuring the asset achieves the design life, but also to increase the CAPEX and OPEX. Meaning the LCOE can be reduced with increased confidence on the design, current loads, and prediction of future loads.

There is a strong motivation towards the digitisation of wind turbines and gathering meteorological information from wind farms. These data are pivotal for operational monitoring, building, and validating analytical modes to increase confidence on the design by measuring the loading on the structures. There are recommendations on sensors, and sensors placements [7]. Incorporating this on an offshore structure can be incredibly valuable, and as more information is gathered from assets more methods can be developed to increase the confidence of offshore structures and reduce the LCOE.

Digital-enabled asset management systems (DEAMs) aim to utilise the data gathered from an entire wind farm to make data-reinforced decisions. One concept in particular, digital twin, is to replicate the operational asset as accurately as possible within the digital environmental scope. This can be accomplished with a combination of data acquisition from the operational asset as well as developing a modelling framework. Provided the models can accurately behave close to the real asset they will provide valuable insight to the structural health.

1.2 Problem Statement

In the last decades, Structural Health Monitoring (SHM) has emerged as an effective means to assess the structural and mechanical systems, aiming to increase the amount of information on the structure by reducing the LCOE and substituting the traditional time-based maintenance method with a cost-effective condition-based strategy. Within this context, the condition of structures is to be assessed with the aid of a monitoring system that can record the responses of the asset, and subsequently diagnose damage or irregularities. But most importantly, notifying the operator to take the necessary action. Such strategies may be significantly more efficient than the traditional methods, but this comes with the requirement of sophisticated data processing techniques that can distil the measured information and extract the condition-sensitive indicators.

In short, SHM follows two methodologies [8], [98]: The model-based approach, a physics-based model is constructed based on the individual geometry and topology of an asset, utilising the operational conditions the parameters are updated through the sensor data, most commonly accelerometers, inclinometers or strain gauges [158]. The second approach is a data-driven where the models of the system are not based on physical laws, instead machine learning tools are applied to learn within the measured data. Both techniques have their advantages. Physics-based models have the potential to simulate a variety of operational and environmental conditions to high accuracy. However, the model must regularly be validated to ensure confident results [74]. Also, as technologies progress and the complexity increases, determining the ambiguity of the results becomes increasingly more difficult. On the other hand, when conducting a data-driven approach, complex behaviour can be learned without having to define a model from physical principles. Furthermore, addressing the ambiguity can be easily incorporated.

Unfortunately, machine learning algorithms usually require large data sets to be recorded during system operation to infer reliable predictions. Supervised learning techniques require measured signals to be comprehensively labelled to describe what each signal represents. Unsupervised techniques require large quantities of unlabelled measured signals. Furthermore, machine learning tools offer minimal insight into the underlying physics of the problem and they are considered black box models [21].

Analysis of SHM in engineering should require a combination of physical models and data-driven approaches, as valuable information can be derived from both methodologies. This presents the requirement of large amounts of data, sophisticated data processing techniques, an expensive physical model, and the development of a machine learning model. Surely incorporating all of this at once is an expensive solution?

The advantage of implementing a low-cost monitoring campaign is that only a select few wind turbines in the wind farm need to be instrumented. This results in a significant reduction in the amount of model updating required for the physical models in the wind farm. Furthermore, it generates a highly effective population-based structural health monitoring (PBSHM) campaign. The PBSHM model is developed using both the physical models and machine learning techniques generating a Gray box model for an entire wind farm with a fraction of the effort and cost in comparison to producing physical models for the entire wind farm.

1.3 Aims and Objectives

This research aims to develop machine learning model approaches related to offshore wind condition monitoring as well as develop a process on how to conduct a PBSHM campaign. The contribution to academia investigates the components of DEAM, using machine learning techniques with data from several different wind turbines verifying the models using the predictions and the real data. This is conducted by firstly addressing concerns about how data is pre-processed for machine learning models in wind energy. The contribution to knowledge from this highlights that current standards are falling behind on this aspect and that model accuracy can be improved with effective pre-processing. A byproduct of this contribution is that with transparent and standardised pre-processing DEAM models will be more easily compared against one another.

Within DEAM a challenge when utilising machine learning is sophisticated data. Following from the pre-processing another aspect is availability and addressing the limitations of small data-sets. For effective machine learning, one needs large amounts of data. The data processing aspect of this challenges this idea and identifies that transfer learning can aid in improving the accuracy of alarm assessment of an offshore wind farm.

The main contribution to knowledge is the development of a low-cost DEAM model. This is conducted in two parts, the first stage investigates the type of PBSHM problem, and the population form for this task uses sensor data from the foundation of a jacket structure that is processed to determine the damage equivalent moments at the interface of the jacket and the transition piece for three structures. Stage two applies the knowledge of the population form gained from the case study and develops a variety of models using conventional machine learning and domain adaptation techniques. The final model is verified using the data from a structure that is not used in the development of the model to verify the low-cost monitoring campaign effectiveness.

The thesis aims to deliver the following objectives:

- Conduct a comprehensive literature review of offshore wind DEAM techniques and provide references. This includes the SHM techniques and the machine learning models used.

- Begin the discussion on why pre-processing should be standardised in offshore wind and why it should be more transparent in academia.

- Provide a solution to address the issue of limited data sets. This is conducted by developing a time-series model that makes predictions of the state of the wind turbine using transfer learning. The model is compared to a traditional machine learning method to highlight the improvements made.

- Conduct a case study on the limitations of PBSHM for a low-cost monitoring campaign.

- Develop a time domain PBSHM model and compare the effectiveness of several models to determine the most suitable method for this task.

1.4 Structure of the Thesis

With the main body of this thesis made up of machine learning modelling for the application of DEAM, this thesis approaches both challenges and does so in four chapters. The first chapter consists of a literature review of current SHM techniques using both SCADA and vibration-based methods. This also covers some of the machine learning techniques that are commonly used. The second chapter conveys the concerns and solutions when using data. This is further broken into three sections; The first part deals with SCADA, the second focuses on the types of sensors needed for DEAM, and the last part is pre-processing. Pre-processing involves conducting a sensitivity study using five different machine learning methods for power trend condition monitoring. The study compares raw data with pre-processed data.The third chapter is a case study on the Wikinger wind farm where the CMS is used to determine the damage equivalent loads on the foundation of the structure, and this is used to investigate the effects on the population of the wind farm with regards to the operation and environment by combining the damage equivalent moments with the documentation and SCADA data. Finally, the project presents two DEAM tools. The first is a solution to the limited data problem in machine learning. This is indicative of a newly installed wind turbine with a short amount of operational time. Hard-parameter transfer learning was used with this model, and it makes predictions of the operational state of the wind turbine, providing error signals on the health of the wind turbine. The second DEAM tool is a PBSHM model. This is a general classifier on the damage equivalent moments for the structures on the Wikingier wind farm. There is a collection of machine learning models used for this and a comparison is made.

This thesis is a collection of work that has been published in peer-reviewed journals. The components of this thesis are from the following publications:

- Literature Review
  - **Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management**, Innes Murdo Black, Mark Richmond, Athanasios Kolios, 2021, International Journal of Sustainable Energy, Taylor & Francis
- Data Collection and Processing
  - **Standardisation of Wind Turbine SCADA Data Suited for Machine Learning Condition Monitoring**, Innes Murdo Black, Athanasios Kolios, 2022, Proceedings in Marine Technology and Ocean Engineering, Trends in Maritime Technology and Engineering
- Case Study of an Operational Wind Farm
  - **Population-Based Structural Health Monitoring: Investigation into the Heterogeneity of an Offshore Wind Farm**, Innes Murdo Black, Moritz Werther Häckell, Athanasios Kolios, 2023, Renewable energy,
- Digital Enabled Asset Management of Wind Turbines Results

- **Long term findings**, Simon Sielder, Innes Murdo Black, Carolin Sophie Wenedelborn, Moritz Werther Häckell, Debora Cevasco, 2022, ROMEO, D4.5
- **Deep Neural Network Hard Parameter Multi-Task Learning for Condition Monitoring of an Offshore Wind Turbine**, Innes Murdo Black, Deborah Cavesco, Athanasios Kolios, 2022, Journal of Physics
- **Population Based Structural Health Monitoring: Homogeneous Offshore Wind Model Development**, Innes Murdo Black, Moritz Werther Häckell, Athanasios Kolios, 2022, Renewable energy, (Submitted)

Chapter 2

Literature Review

Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management, Innes Murdo Black, Mark Richmond, Athanasios Kolios, 2021, International Journal of Sustainable Energy, Taylor & Francis

2.1 Introduction

The total capacity in Europe of installed wind power sat at 18,499 MW in 2018 which is an increase of 2,649 MW from 2017 [212]. The industry has tended away from small clusters of wind turbines where maintenance is more accessible and the overheads of sending a team for regular intervals were not expensive. For offshore wind farms, the cost of maintenance relative to the levelised cost of energy (LCOE) is significantly increased compared to onshore. It is reported in the North Sea that the operations and maintenance (O&M) cost between 20%-25% of the LCOE compared to around 12% onshore [167], [188]. The impact of the offshore environment coupled with increasingly expanding machines means that the maintenance strategy of planned, scheduled or responsive regime incorporates a more proactive, predictive methodology. The key contributor to this shift in the industry is the intelligent monitoring of structural health termed condition monitoring.

Condition monitoring systems (CMS) are being developed by several operators.

Companies monitor several parameters including; vibrations, oil quality and temperatures in some of the main assemblies [118]. This information is used to infer the health of the assets to determine the remaining useful life or to determine if scheduled maintenance is required based on the monitored irregularities. There is an additional cost of implementing supervisory control systems, which have deterred operators in the past but the financial benefit has eradicated scepticism [39]. All large utility-scale offshore wind turbines have supervisory control and data acquisition (SCADA) systems to govern their performance. SCADA systems provide a magnitude of information over the operational life of a turbine providing updates at a resolution of 10 minutes.

This paper is a continuation of M. Luengo and A. Kolios. [118], which carries out a detailed review of condition monitoring systems, following the statistical pattern recognition paradigm. Developing this idea, this report seeks to understand the types of maintenance procedures implemented in offshore wind engineering. Focusing on current detection methods incorporating machine-learning techniques. Currently, there is a variety of review papers on how condition monitoring is beneficial when implemented correctly [231], [13]. Others look at how predicting the structural health of a component using machine learning methods can determine scheduled maintenance [168], [169]. Most of which are looking at vibration methods or SCADA methods individually. This article investigates this but develops on the idea that they can complement the results from machine learning methods when used together. One of the major drawbacks of vibration-based methods is that the results are difficult to interpret without the help of an expert. Combining both sets of information complements the analysis for easier insight and implementation on improving offshore wind turbine operational management.

The remainder of this Thesis is constructed as follows: A methodology for the systematic review is portrayed in section 2. In sections 3, and 4 include a review of 5 different machine learning methods and condition monitoring strategies. Finally, the remaining sections with a discussion and conclusion separately.

2.2 Method

This review conducts a qualitative systematic review providing an exhaustive summary of current systems in place that tackle offshore wind monitoring and maintenance. The purpose of this review style is to restrict confirmation bias.

The systematic review procedure is developed based on [1]. This focuses on assessing what methods have been applied towards maintenance and monitoring. This study is only looking into 5 different regression-type models. Refining this search helped manage the large volume of literature and streamline the process. The top 5 most commonly integrated regression methods are implemented. Observing that there are other regression methods such as; Tobit, Cox, Poisson, Lasso, and linear to name a few but they are less commonly applied in recent papers.

The process is as follows; both Google Scholar and Scopus were used to initialise the database, searching specifically for condition-based maintenance for offshore wind turbine operational management. Then a limitation of the top five most used types of machine learning applied to SCADA data, and all vibration-based methods were included. **Secondly, a limit from 2013** is set, removing duplicates from both searches. After this process was completed the initial screening process, with titles and abstracts checked against a predetermined criterion for relevance. The thorough search screening for this paper ranks papers with higher citations having more recognition and importance in the community. At this point, a full paper consideration is taken. Noting, some significant works older than 2013 and others with limited citation works that have a large impact in this field are still included. Following the exhaustive search, a review of the most relevant information will be transcribed.

2.3 Machine Learning

Machine learning is the application of artificial intelligence (AI) that provides systems the ability to experience without explicitly being programmed. Nevertheless, machine learning is referred to as an area of artificial intelligence that is solely concerned with identifying patterns from data for predictions on unseen information. One of the most

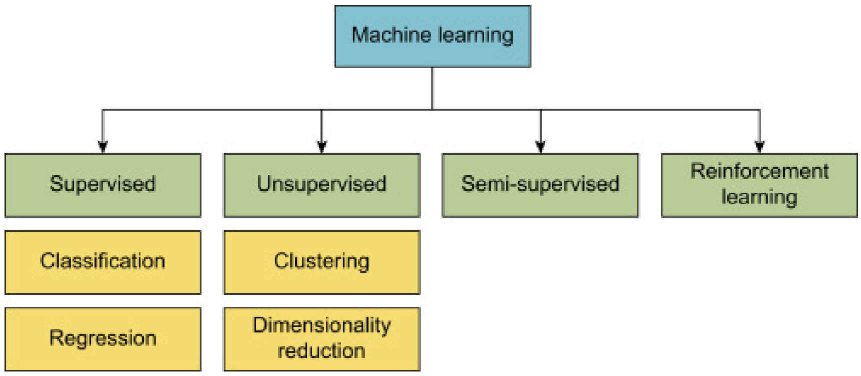


Figure 2.1: Machine learning utilises four techniques: supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised and unsupervised further categorise into classification/regression and clustering/dimensionality reduction

common definitions, by T. M. Mitchel [125] 'A computer program is said to learn from experience E concerning some task T and some performance measure P, if its performance on T, as measured by P improves with experience E'. In this section the types of machine learning methods, see Figure 2.1 for a breakdown of the different types of learning. These will be discussed, detailing the nuances and how they are incorporated into the industry. Following that, a more in-depth discussion of 5 specific regression models that are both most commonly applied and widely applicable to time-series data predictions is discussed. These are support vector regression, K-Nearest Neighbour, Bayesian Network, Gaussian Process Regression, and Artificial Neural Networks.

Machine learning demands to learn relevant patterns from data to make predictions. There are various ways in which this can be achieved from a vast variety of learning algorithms to select from. Common taxonomy organises the different approaches into learning styles, based on this there are main categories of learning; supervised, unsupervised, semi-supervised and reinforcement. Supervised learning can be further coupled into classification or regression depending on either categorical or continuous targets respectively. The main task of unsupervised learning involves clustering and reducing the dimensionality of the input information.

2.3.1 Supervised Learning

For supervised learning, the algorithm must have input variables and the target variables could be; the severity of symptoms, the presence of credit card fraud or future clinical outcomes. The aim here is to develop an algorithm to determine the optimal function that captures the relationship between input and output target variables. This type of learning is often related to learning with a teacher, in this situation the teacher knows the correct answer and corrects the algorithm when a mistake is made. Therefore, requiring an iterative process of predictions and adjustment until the output prediction and targeted value have reached a maximal efficiency. The performance is estimated by comparing targeted values against unseen information.

Classification

The main aim of classification is to predict group membership, labels or classes, from observations. This type of algorithm is common in brain disorder research. Classification is advantageous for this problem since it can be broken into a categorical decision; for example, should the patient be medicated with A, B, or C. In this case, the algorithm learns to distinguish the patient with a particular disease from healthy controls. There are a plethora of published articles that have used neuron-imaging data to determine disorders [9-14]. A small niche is using motor signals to identify Parkinson's disease [107]. This motion can be extended to a multi-disorder diagnosis, where the algorithm can predict the probability of a patient having a particular disease [171].

Regression

For a regression problem, the aim is to determine a score on a continuous scale. Some problems cross over with classifier and regression algorithms, but the outcome is continuous rather than a categorical variable. This characteristic is useful, when, predicting outcomes such as stock market trends or predicting meteorological trends, both follow a continuum.

There is a collection of survey articles related to the energy industry, for example, [199], applies regression models to determine the solar radiation to make predictions for photo-voltaic panels, [84] similarly makes predictions on wind power forecasting [223], has applied various methods to determine consumer electric power consumption.

The level of functioning is completely different, in the clinical trials for classification models the output is distinguished between true and false. For the regression models applied to the energy market, the output has a quantitative value that can be ranked.

2.3.2 Unsupervised Learning

As opposed to supervised learning, in an unsupervised learning environment, there is no target value. The aim is to uncover underlying structures in data. There are two main approaches to unsupervised learning: clustering and dimensionality reduction.

Clustering

Clustering is a technique that helps us identify meaningful subgroups within large datasets, like scientific papers. It essentially sorts individual items (e.g., articles) into smaller, distinct groups based on their similarities. Think of it like organising books by theme: papers on similar topics end up together. This method isn't limited to text analysis. Image recognition and pattern recognition, like identifying objects in photos, also benefit from clustering techniques like grid mapping. Interestingly, even "noisy" data from unreliable sensors can hold value. Research [232] has shown that re-evaluating such data using clustering can sometimes lead to surprising and effective results.

Dimensionality Reduction

In a situation where the number of features substantially outnumbers the number of observations, a dimensional reduction can be useful. The greater amount of features increases the visual complexity, dimensional reduction is a process where reducing the number of random features under consideration and replacing it with a principal set.

A dimensional reduction study on three-dimensional shape retrieval carried out in [205] aims to gather the most pertinent shapes from two-dimensional images. Another study takes conventional household objects in CAD, and uses the ResNet library in Python, which is a neural network [25], to retrieve the original shape and alleviate computational demand, omitting redundant information. Shape descriptors are commonly used in text recognition software, A.A.Mohmamad [126] has reviewed various Dimensionality reduction text recognition methods. The handwritten text has considerable amounts of ambiguity and redundancies. These methods seek to retrieve the most relevant information to determine the text.

2.3.3 Semi-Supervised Learning

As the title suggests, when target variables are only available for a portion of the data. Semi-supervised learning addresses the issue by allowing the model to integrate the available unlabelled features for supervised learning. This approach is effective when it is impractical or expensive to attain the measured target data.

2.3.4 Reinforcement Learning

Reinforcement learning aims to build a system whereby it can learn from the interactions in its environment, much like the operant conditioning Sutton & Barto [184]. For this type of learning, the algorithm's conduct is shaped by a sequence of rewards and penalties, which are dependent on whether the decision is towards the final goal, that is set by the programmer. As opposed to supervised learning, where examples are given to model the behaviour, reinforcement learning is allowed to develop its path, based on trial and error. This is one of the most promising areas of machine learning for autonomous control of vehicles. A model-free reinforcement algorithm developed by Chen-Huan Pi [142] has a quad-copter tracing a predetermined path subjected to disturbances. The algorithm rewards the system when the trajectory is as efficient at following the predetermined path subsequently minimising the error.

Machine learning is an emerging topic within artificial intelligence that is gaining momentum in all research fields, with each industry developing different topics within the AI field. The nature of this, for the most part, is open source which allows an easy transfer of knowledge between industries. Machine learning is concerned with identifying patterns from data and subsequently using these patterns to make predictions for unseen data. This is a stark contrast from inferential statistics that seek explanatory power. It is important to understand that there are significant challenges concerning machine learning, it requires a larger number of observations, whereas some statistical inferences require less. This issue is being addressed with major developments in transfer learning or domain adaptation [202], among other methods.

2.4 Models

Machine learning is a growing area with a magnitude of methods to select from. In the previous section, only the basic procedure is explained. Supervised learning is a common type of machine learning for condition-based maintenance. The following section will provide a more detailed discussion of classification and regression models, introducing the fundamental concepts behind them. Before digging into the details of these models it is important to divulge the process of dealing with raw data.

Pre-processing temporal data is mandatory to improve the final model's accuracy. The mote of errors could be, incomplete data, noisy or inconsistent. The following process can be implemented:

- Feature extraction involves cases where data sources from multiple databases may need to be integrated into a single data set. In the case of offshore wind, one may want to extract vibrational features from the metalogical data.

- Data reduction involves reducing the size of the data through a subset selection, feature selection or data transformation. This can be handled using row reduction and column reduction. This generally improves the efficiency of the algorithm by removing irrelevant records, improving the quality of the data.

- Data cleaning is a widely studied practice that can be carried out using varying methods for time-series data [27] and [62] have discussed potential options. Alternatively, for high dimensional data [10] explores other methods. It involves potentially removing noisy data points or replacing them. This handles missing and incorrect entries.

- Data splitting, consists of three sets; training set, which is used for learning. The Validation set is used to tune the parameters of the model and the testing set is used to assess the performance of the fully trained model. This is most commonly used in artificial neural networks, however, it is not employed in all ML processes.

This process is carried out to provide more accurate and efficient results. The following subsections discuss the underlying features of; artificial neural networks, support vector regression, K-nearest neighbors, Dynamic Bayesian network, and Gaussian regression.

2.4.1 Artificial Neural Network

Artificial neural networks (ANN) have been developed since the 1940s by McCulloch and Pitts [121] and the learning rule of ANN is based on simple neuron-like logic called a perceptron. A single neuron is a universal approximator of a smooth surface. The connection strength between artificial neurons is called 'Hebbian learning' [83] ANNs took off in the 1980s when; Werbos [206] Parker, and Rumelthart, Hinton and Williams [160] all worked on backpropagation; hence, systematic learning can be achieved. The human brain is a vast network of connections coupled with sensory receptors that perform the simple task of receiving and passing on signals. If the signal is strong enough it passes the information on, otherwise, it is rejected. This is the building block for the ANN, with the structure depending on the architecture. Dendrites are assimilated to the inputs of the signal, collating the information along the cell body. The myelinated axion is the functioning body of the neuron processing the information and at the axon terminal, the information is passed on to the subsequent neuron. Learning in ANNs is commonly carried out using backpropagation optimisation. Originally it was considered that a synaptic signal was either 1 or 0. For an optimisation problem, discontinuous trends are difficult to solve. Rao and Rao [150] considered the synaptic strength to be a continuous fixed function. Recently have been developments implementing the self-adjusting threshold discussed in [207]. One of the most significant advantages of ANNs is the ability to predict non-linear, complex behaviour effectively. This success is directly proportional to the selected instances. When presented with incomplete information ANNs can still produce results. ANNs are dependent on the samples, and with

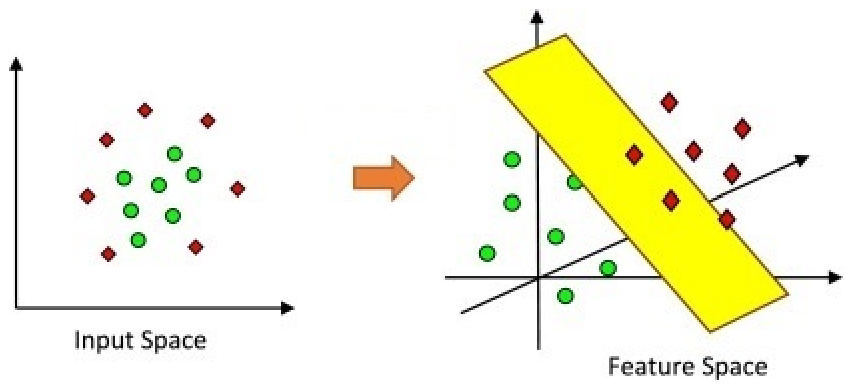


Figure 2.2: The transformation from input data into feature space for a greater fit [69]

limited information, the accuracy is hindered. This also means that large quantities of data are necessarily imposing a hardware requirement. But there are developments in domain adaptation, [202]. This method transfers knowledge from other related data sets to improve the performance of machine learning methods when there are limited amounts of data for specific tasks.

2.4.2 Support Vector Regression

Due to some of the shortcomings of ANN, the so-called support vector regression (SVR) was developed in the 1990s for non-linear regression and classification problems by V. Vapnik [197]. There are three main reasons for SVR's success; reliable training efficiency with small samples, the robustness against error models and its computational efficiency compared to other methods such as ANN [119] [201]. In the training process of machine learning techniques, it is often assumed that the larger the sample, the error function will decrease. This can result in what is known as over-fitting [52], where a complicated function is designed to reduce the loss of the estimated target and desired output. This yields very accurate results in the training stage, but the estimation is poor. This is not consistent with SVRs.

SVRs aim to generate a function by separating the variables with a hyper-plane, see figure 2.2 for as an example. To determine the two classes separating the hyperplane a suitable regularisation model is required for the training stage. Vapnik-Chervonekis developed the basis of SVR [43], the Kernel function [123], this function transforms the input space into a feature space, also known as Hilbert space. Hilbert space is a mathematical structure used to represent points, similar to how a coordinate system (like the X and Y axes on a graph) can represent points in 2D space. However, a Hilbert space can be infinite-dimensional, allowing it to represent much more complex data [179]:

$$y = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \tag{2.1}$$

where $\alpha_i \& \alpha_i^*$ are Lagrange multipliers for N training variables and b is a constant real number. For this instance, a Gaussian radial bias kernel function can be used with σ, a tuning parameter:

$$k(x_i, x_j) = exp\left(-\frac{\parallel x_i - x_j \parallel}{2\sigma^2}\right). \tag{2.2}$$

2.4.3 K-Nearest Neighbour

Similar to the previous machine learning methods, the K-nearest neighbour (K-NN) presumes that the current time series will exist in the future. This method differs from the other methods since the training of the algorithm is considered 'lazy learning', where learning is a generalisation of the input data as opposed to training the algorithm before receiving queries in 'eager learning'. The objective is therefore to determine the present value based on a generalisation of past values. This is evolved from Kantz and Schreiber [101].

The algorithm aims to determine the Euclidean distance from the present and past variables, while some other less common distances can also be used [133]. Locating the closest distance of the variables from the learning data to determine the output. The distance is given by:

$$D(x_i, x_j) = \sqrt{\sum_{l=1}^{N} (a_l(x_i) - a_l(x_j))^2} \tag{2.3}$$

where $a_1(x)$ denotees the 1^{st} feature instance of x. To determine the K-NNs the value $\hat{f}(x_q)$ as the estimate of $f(x_q)$ is the most frequent value of f among the K-NN can be represented as:

$$\hat{f}(x_q) \leftarrow argmax \sum_{i=1}^{k} \delta(\upsilon, f(x_i)) \tag{2.4}$$

with $\upsilon(a, b) = 1$ if $a = b$ else $\upsilon(a, b) = 0$. This is one of the simplest machine learning methods. Since there is no training necessary it can be considered faster at making decisions. However, with large multidimensional data sets, the algorithm does not perform well since it is more difficult to determine the Euclidean distance.

2.4.4 Dynamic Bayesian Network

Dynamic Bayesian networks (DBN) are directed graphs representing a set of variables and their condition dependence [138]. BNs constitute both quantitative and qualitative analysis. The quantitative section involves prior and conditional probabilities for each node and the qualitative section involves The clear advantage of implementing a low-cost monitoring campaign is that only a select few wind turbines in the wind farm need to be instrumented. This results in a significant reduction in the amount of model updating required for the physical models in the wind farm. Furthermore, it generates highly effective population-based structural health monitoring (PBSHM) campaign graphs. Each node represents a random variable. The arcs in a graph or diagram indicate the causal relationships or dependencies between the nodes. The joint distribution between variables of the vector is X, and the network is represented by:

$$p(X) = \prod_{l=1}^{N} p(x_i|pa(x_i)) \qquad \forall x_i.....x_N \in \Omega x_i.....x_N \tag{2.5}$$

with values of variable x_i and $pa(x_i)$ denoting an instant of the parents of X.

BNs are a set of conditional independence statements; this is the main consideration when building the model. This can be determined by employing the rules of d-separation. Using this method the joint probability distribution for a set of random variables can be determined using the chain rule. Dynamic BNs are an extension of BNs; typically they partition the variables into input, hidden and output variables for a state-space model. This can tackle discrete-time stochastic models.

2.4.5 Gaussian Process Regression

A Gaussian process (GP) is a multi-variable probabilistic approach applied to regression and classification machine learning problems. GP algorithms are stochastic processes, that can model multivariate infinite vectors. The process can be defined for a random set of variables x as mean $m(x)$ and covariance $k(x, x')$ [151]:

$$f(x) = GP(m(x(x), k(x, x'))) \tag{2.6}$$

GP regression determines marginal Gaussian distributions from the training data and can describe non-linear trends as an output function. Essentially GP is a non-parametric generalisation of a joint-normal distribution of an infinite set of input variables, defined by the mean and Kernel function [152]. A typical common Kernel is the squared exponential kernel used in [193]:

$$k(x_i, x_j) = exp\left(1 - \frac{1}{2}d\left(\frac{x_i}{l}\frac{x_i}{l}\right)^2\right) \tag{2.7}$$

With l being a length scale. The function determines a probabilistic accuracy, meaning that the output from this process has a predetermined confidence interval. This is advantageous as there is an uncertainty associated with the output function. The Gaussian process is also versatile since there is a variety of covariance functions for varying problems however, this does raise the issue of implementing the correct Kalman filter [81].

Machine learning is continually adapting, with huge amounts of investment from large international organisations and research institutions. This study has only selected several models that are suited towards time-series problems but there are vast amounts of other models for different challenges. There is no consensus on what is the most desirable methodology, there are guides and books but no concrete standardisations for confident, accurate model implementation.

Table 2.1: Summary of references relating to the Machine learning methods used in the specific types of Monitoring and Maintenance Techniques

Monitoring For Maintenace Techniques	**Type of Machine-Learning Method**					
	ANN	**SVR**	**K-NN**	**DBN**	**GRP**	**Other**
Vibrational	[58]	[234]	[61]			[145] [175] [68] [106] [17] [77] [190] [130] [226] [233] [109] [25]
Trending	[78] [204] [115]	[100]	[100]		[177] [163]	[216] [217]
Clustering	[30] [103] [47] [120] [221]	[80] [50] [227]	[129] [191]	[85] [200]	[229] [230]	[50] [131]
Normal behaviour modeling	[38] [65] [20] [170]					[14] [185]
Damage Modeling	[18] [48]	[112] [153] [154]	[235]	[157]	[28] [166]	[146] [23] [57] [97] [173]
Alarm Assessment	[143] [96] [214]	[108] [215]		[203] [228]		[147] [60] [72]
Performance Monitoring	[164]					[139] [159] [165] [140] [180] [16] [176]

2.5 Data Driven Decision Making For Wind Turbine Operational Maintenance

The maintenance of an offshore wind farm can be categorised into two approaches; preventative and corrective maintenance. The latter applies to the run-to-failure approach, which inherently risks cascading failures with the potential of catastrophic loss. This approach is also likely to cause greater downtimes since planning for maintenance is followed by the failure event.

In contrast, the preventative-based maintenance philosophy aims to repair a component before failure, this is broken into two subcategories: calendar-based maintenance and condition-based maintenance. Calendar-based maintenance is performed by annual or semi-annual visits or scheduled replacements based on the operational life of the component. A more effective approach is condition-based maintenance, where the components are repaired based on the health of the part. This strategy aims to predict failure before it occurs so that the scheduling can be planned earlier in a corrective manner to reduce downtime. Producing accurate predictions of the remaining useful life of an asset is a complex task, especially so in a complex system with multiple failure modes.

There are three main requirements of a condition-based maintenance system according to Wiggehuzen et al, [208]:

- Detection of failure mechanism
- Measurable criteria
- Detection of time

These three steps have been interpreted in a variety of different ways, and the creativity within the maintenance sector has resulted in a variety of different models. The following sections will look into Acoustic and SCADA models for the detection of failure mechanisms and how to use the measurable criterion.

2.5.1 Acoustic Condition Modelling

In the context of an offshore wind turbine, with mechanical, electrical and structural components. There have been various signals and monitoring tools to determine the structural health of the turbine.

The identification of dynamic system responses has been carried out qualitatively since the introduction of acoustic modelling [189], recently with the emergence of condition monitoring, the maturation and cost reduction of digital computer hardware [41]. Condition monitoring is becoming more attractive and the offshore wind industry is becoming an emerging topic. Monitoring of rotating machinery is most competitive in terms of profitability and reliability [187], [145]. The failure detection systems depend on pattern recognition related to displacement, velocity and acceleration time histories.

The sensors are commonly placed on the housing or shafts of the components during the operation. DNV-GL [5] codes on condition monitoring state that the minimal amount of sensors of geared turbines shall include at least one vibration sensor on the main bearing, two on the generator bearing and 5 for the gearbox sensor. Sensors are predominantly mounted on the housing of the component. The most common sensor used is a Piezo-electric accelerometer because of the large bandwidth, ranging from 0.1 Hz to 30 Hz BSI 13373-1 [3]. Piezo-electrics sensors, unfortunately, suffer from roll-off at lower frequencies.

Raw vibration analysis is no arbitrary task. The data as a whole needs to be analysed in detail. Common methods are; time-domain analysis, for instance, Hilbert Transform. Statistical analysis, and frequency domain techniques, such as the Fast-Fourier Transform or time-frequency domain techniques like the wavelet transform.

Fast-Fourier-Transform has been the most applied technique to obtain frequency spectrum by converting the time domain signals into the frequency domain. Specific harmonics are directly correlated to the gradation or faults of moving parts. Qiao, W. and D. Lu [145] have explored this for faults specific to wind turbines. Fast-Fourier transforms are effective in stationary signals but, may result in indistinct solutions in non-stationary environments. Offshore wind, and especially floating offshore wind the nacelle oscillates. The IEC-64001-25 [4] has suggested binning vibration measurements over the power band of the operational wind turbine, or using log mean composition and applying synchrosqueezing transform [175], [68].

Envelope analysis is used in signal processing to detect fault frequencies that may not be represented in the spectrum produced by Fast-Fourier-Transform such as shock impulses [174]. A Band-pass filter is applied to the time domain signals centring on the desired energy region. The amplitude is demodulated in the filtered time-domain signal, extracting this repetition rate of impact. Applying this process during Fast-Fourier-Transform characteristic impact frequencies and their modulations, sidebands, are determined.

Cepstrum involves taking the inverse Fourier transform of the logarithmic power spectrum. This methodology has been applied in auto-correction ANSIs, which is just performed on the logarithm of the power spectrum, In Cepstrum the correction is mainly focused on the lower harmonics [68].

For rotating machinery, fault detection is usually determined by distinguishing specific harmonics or side bands. Envelope analysis is carried out to specifically locate sidebands using amplitude demodulation. Cepstrum analysis is carried out to distinguish between different harmonic groups [148]. The combination of this collection of techniques can lead to good failure detection. They can identify various forms of failures and are incorporated into commercially available solutions on vibrational analysis for offshore wind turbine condition monitoring. These tools often require experts to interpret the results to determine whether the information is indicative of a fault. Efforts have been put to fully automate fault detection with vibrational analysis using features such as sideband energy, [106], deep learning convolution networks [17] or using a system that uses vibrations analysis combined with wind and rotor speeds [77]. There is a need for a more generic approach of fault detection utilising machine learning methods, and vibrational analysis with the minimal amount of human intervention possible.

Offshore wind turbine condition monitoring is an emerging field with various tools emerging. Roller bearing failure detection using vibrational signals by extracting features with support vector regression modelling [61]. Also, another the same problem using artificial neural networks [234]. Gearbox failures are of interest since it has one of the highest down times, these failures have been diagnosed using Von-Kalman filters during operation and in non-stationary conditions [58] also including complex wavelet transformations [190]. Another approach has developed a tool using the angular velocity data [130], investigating jerk', using the rate of change of acceleration [226].

Condition monitoring systems have been incorporated in several industries for years, Sky-wise [11] in the Aerospace or Pulse for the Nuclear industry. The Wind industry operators have only recently introduced dedicated condition monitoring systems. The reason this development requires time is largely dependent on the initial cost being high [216], resulting in ambiguity on returns. The benefit of this investment will take years to impact the cost-benefit analysis. Another aspect is the probability of false alarms, hence, unnecessary costs from scheduled maintenance. As some insurance companies dictate that condition monitoring is mandatory the offshore wind industry will develop more effective mechanisms, according to a study carried out in 2014 by Yang. et al, [217] only 60%-80% accurate diagnosis is necessary to provide adequate returns to justify the implementation.

2.5.2 SCADA Modelling

Condition monitoring systems based mostly on vibrational analysis can be costly, the appeal of SCADA-based systems for condition monitoring is largely due to the sensors already being installed to track the normal operations of the vessels. This subsection will discuss approaches implemented in the industry using SCADA data for failure detection and condition monitoring. There are six main approaches; trending, clustering, normal behaviour modelling, damage modelling, alarm assessment, and performance monitoring. SCADA data typically records the meteorological data, component temperatures, control variables and electrical characteristics of horizontal offshore wind turbines. The exact configuration is dependent on the manufacturer and operator but is most consistent with Table 2.1. SCADA systems typically collect raw data at 1 Hz (once per second), but for storage and analysis purposes, this data is often aggregated and represented as statistical summaries taken every 10 minutes. Some operators can offer maximums, minimums and standard deviations on each time step. Some forms of valuable information are starts, stops, alarm logs [70], oil pressure levels coupled with filter status [217] and vibrations.

There is no standardisation implemented in the wind industry or any industry. The general trend seems to be more sensors in modern offshore wind turbines for more data.

Data cleansing is a technique involving the transformation of raw data into an understandable format. For instance, the data, whether it is missing, inconsistent or noisy data, especially in the offshore wind industry. This has a large influence on meaningful reliability assessment [56]. Two common methods of addressing this issue; Fuzzy-set theory [127]. This aims to distinguish the gradual assessment of membership from elements using indicator functions. Dempster–Shafer theory [59] is a framework for reasoning with uncertainty by applying understood connections. These theories can help mitigate the issue of vague data but do not remove the problem.

There is information that can be utilised from the offshore oil industry or the onshore wind industry that can be utilised in setting up meaningful databases, taking into consideration the different environmental conditions into account [224]. Another issue is the changing technologies of a fast-developing industry so the richness of data or cost aspects needs to be tackled in each of the following models.

Trending

As soon as optimising maintenance strategies grew in the wind industry, so did structural health monitoring. One key area of interest in SCADA systems is temperature measurements. Thermal dynamics of components can directly relate to the efficiency of the system. In the offshore wind, the gearbox has one of the highest down times causing this keen interest.

A study carried out from 2002 to 2007, [209], applies SCADA-based monitoring methodologies consisting of trending methods. Implementing regressions on scatter diagrams of temperature, power, and as three-dimensional graphs including the ambient temperature. Manual interpretation of the filtered results was proven to be beneficial for determining anomalous behaviour. [216] have concluded that when the gearbox of a horizontal operational wind turbine efficiency decreases, the gearbox temperature will rise (compared to the ambient temperature), with an expected 6 months before failure. Wilkinson et al, [210] investigated condition monitoring methods. One of which was a trending method that compared the temperature difference of separate wind turbines on the same site and tried to determine faults. The authors dismissed the efforts rendering the approach too inaccurate due to the environmental condition disparity between each wind turbine. It could be argued that this method is still valid but a more complex approach is necessary to irradiate the temporal nature of turbines. The binning method, where an average is taken, was applied to; the wind speed, generator speed and output power by Yang et al, [217]. In this case, the trending methods were applied to historic data and current information to detect levels of damage, the added value is the differing scales of damage dictated by the damage mode and dependent parameter.

The trending method applied to SCADA parameters can observe the development of failure using past data compared to the present information. There are several studies indicated here where the results are case-dependent. Particularly in studies using temperature data which is case-specific and requires manual interpretation. Attempts to visualise the information have not provided more insight. If trending methods are to be used for maintenance, the difficulties of interpretation and variance on individual offshore wind turbines will need to be addressed. If not, this will likely result in uncertainty and the possibility of false alarms.

Clustering

Visualising trends can be problematic, especially in a wind farm with turbines operating individually under dissimilar meteorological conditions. The evolution of trending was clustering SCADA data, where algorithms began to use classification methods to determine either 'normal', 'faulty' or 'error' observations.

One of the original implementations of clustering using artificial neural networks was carried out by Catmull [30], incorporating a self-organising map interpretation of SCADA information. The methods built clusters through the organisation of neurons on a regular grid for the training process such that the neighbouring neurons have similar inputs. To visualise the clustering information a unified distance matrix is used combined with projections of patterns. In this case, only normal operational signals are implemented for training. The fault detection operated by determining the distance between input data and the best-matched neuron, quantisation error, for abnormal behaviour detection. Similarly, a study by Kim [103] included failures in the training of the algorithm and subsequently was able to determine individual failures

Reviewing clustering methods highlights again that interpreting this information still requires human intervention. Also developing a tool that requires the faults to have been previously recorded is not always possible. The Advantages of clustering are not dissimilar from trending hence more suitable methods are necessary.

Normal Behaviour Modelling

Normal Behaviour Modelling (NBM) encompasses the previous detection methods in normal operations of an offshore wind turbine but has incorporated a significant development where the tool aims to model the desired parameter during the training phase empirically. There are two main concepts; full signal reconstruction, where only the signals no other than the target are used to predict the desired target and Auto-Regressive with eXogenouos input modelling, which includes the target value but also incorporates previous values of the target to be used.

There is a magnitude of authors investigating the validity of artificial neural networks' ability to monitor offshore wind turbines. G. Ciulla et al, [38] have investigated the power curve of a Senvion MM92 aero generator from a 2.05 MW wind turbine. They were able to produce results with deviations below 1% for the producibility and below 0.5% for the power curve.

Garcia et al, [65] devised a predictive maintenance scheme called SIMAP using Auto-Regressive with eXogenouos artificial neural network. This involved gearbox bearing temperatures and oil coolant temperatures. Measuring the difference from the coolant temperature before and after the model makes a prediction coupled with selected inputs. Using cross-correlations and the impulse response analysis the scheme can produce results with a confidence level of 95% for the upper and lower bounds.

There is no information on the topology of the artificial neural network from this study.

Another approach using adaptive neuro-fuzzy logic by Amber B et all, [14] for online estimation of the wind speed from the tip speed ratio. This study is carried out with the NREL 5 MW offshore wind turbine. The wind speed and rotational speed of the rotor are used in the estimator, the neuro-fuzzy logic generates Boolean logic rules. For this example, only 216 rules were determined. This study was able to produce results with a testing error of around 6%-7%.

A large number of studies have detailed how NBM is suitable for detecting failure. The concept of evaluating residual measured values minus modelling signals provides an easier interpretation of failure indicators than trending. The large dependency of vast amounts of data for the training stage, coupled with the data pre-processing stage can produce undetected problems or false alarms. There has been a variety of different methodologies discussed but to determine the most effective a more comprehensive study is necessary to evaluate the best solution. Besides, there is no clear universal strategy or consensus on input and output parameters for NBM.

Damage Modelling

Effective modelling can be carried out when an entire system can be observed at every detail, unlike NBM, where most of the modelling is considered a 'black box'. Damage modelling compares the desired signals with empirical models of normal operation. The interpretation of information involves physical models that can better represent the damage development and provide accurate results.

An electro-thermal analysis of a doubly-fed induction generator in a wind turbine using a geared transition conducted in [146] used thermal-dynamic theory and combined it with temperature trending methods. The study made a few assumptions consisting of the rotor aerodynamics being steady-state and the drive-train being considered as rigid body dynamics. This case study used SCADA information of a 1.5 MW wind turbine with gear teeth failures, vibration faults, and generator winding imbalances. The diagnostic rules determined faults for power transmission efficiency, generator winding, and lubricant temperature.

Switching generators require cooling to sustain the operation. Borchersen and Kinnaert [23], devised a physics-based numerical model of the coil temperatures. The model was developed using Kalman filters of the actual system. The damage is determined using the values from the physical model on the three coils coupled with a cumulative sum algorithm. This study has used three years of historical data for 43 offshore wind turbines with an 88% fault detection success rate.

The comparison from measured signals and physical turbine damage models has shown success when applied to failure detection, but significant challenges still exist with regards to the accuracy. Due to the lack of studies conducted on varying failure modes and from different turbines, the full extent of this method of condition monitoring is not yet established.

Alarm Assessment

Interpreting the information from complex failure detection strategies is one of the leading causes of discontinuity from development to delivering a condition monitoring system. There are a variety of tools constructed to provide better insight into the outputs of SCADA control alarms or Normal Behaviour Modelling. In general, alarms are usually broken into system operation, environmental and communication/software to indicate system malfunctions.

A fault diagnostic tool for a wind turbine developed by Yingning Qui, and Yahuni [147] used SCADA alarm material and is based on Dampster-Shafer theory. Dampster-Shafer theory is a multidimensional probability theory that takes each alarm as an item of evidence which supports different possible failures. The system verification was carried out using battery failures producing false negatives and false-positive rates with an accuracy of 76%.

An analysis for detecting false alarms conducted by Alberto Pliego Marugán [143] uses artificial neural networks. There are three trained ANNs in this approach, one of which deals with vibration information, the second deals with SCADA data and both of which are fed into a final model which also includes the alarm information. The system verification model involved a fuzzy logic-based methodology, where fuzzification was applied to each input and output variable and was critically analysed in a fuzzy data set. The final results of this tool ranged from 80-90% false alarm detection.

Evaluating alarm assessments for condition monitoring has shown benefits for fully autonomous fault detection systems. There is a severe lack of information regarding industrial algorithms for alarm assessment in offshore wind. The lack of clarity, highlighted the large discrepancy in the accuracy of outcomes, for the results status code consistency to increase these articles underline how necessary it is for more work in this area to reach full condition monitoring autonomy.

Performance Monitoring

Monitoring the operations of an asset is essential to review its performance. Equally, the performance of the turbine can be evaluated in terms of health, assessing the degradation in terms of health has opened up an opportunity to assess the performance in greater detail.

Addressing the performance of an operational wind turbine has been standardised in the IEC 64001-12 [228], discretising the power curve into bins of 0.5 m/s wind speeds and calculating the mean power value for each bin. Applying this method does not consider the non-linear power-to-wind relationship [94]. There are a variety of probabilistic methods [139], [159], [165] and non-parametric [140] that encapsulate the power curve of a wind turbine. This aspect is crucial to effective power modelling.

A robust diagnosis of a wind turbine pitch failure conducted by Sales-Setién [180] uses power estimation to diagnose the fault. This method involves a pitch misalignment estimation using a statistically based fault scheme. This model incorporates the wind and wake model to determine the power curve, the fault detection is conducted using a model-based observer including a closed-loop. One of the main characteristics of this scheme is the performance improvement by building a bank of observers that lead to the decision to create consensus. This system can determine pitch errors at 0.1 deg at low wind speeds and 1 deg at high.

Another area that affects the output power is the yaw alignment, this is another area that has significant developments for performance monitoring. Dongran Song et al, [164] detail two favourable yaw control systems. The China Ming Yang 1.5 W wind turbine is assessed in this case. Two yaw control systems are constructed using low-pass filters. An Auto-Regressive Integrated Moving Average is used to predict future wind speed and direction. A Kalman filter is then used to determine the manoeuvre of yaw rotation. This method increased the performance by 15% however, the frequency of this SCADA data is every 10 seconds and would benefit from a higher frequency of results.

Performance monitoring is beneficial, increasing the efficiency of the operational wind turbine and assessing the condition of the turbine. Research has indicated that SCADA data is suitable for this task but a higher resolution is necessary for a more optimal tool. Again there is a limited amount of resources from the industrial sector and the standardisation is not implemented in most academic papers.

Table 2.2: Summary of the benefits and limitations for the 5 individual regression machine learning methods

Method	Capabilities	Limitations
Artificial Neural Network	**No Prior knowledge needed:** since the algorithm is based only on historical data. **Processing incomplete data:** A trained network can still produce results with incomplete knowledge. The accuracy is dependent on the importance of the information. **Can be updated in real-time:** The program can run while updating the firmware reducing downtime.	**Unexpected behaviour:** One of the major drawbacks of an ANN. When an ANN produces a probing result there is no insight on how these anomalous results are produced. **Hardware dependence:** ANNs accuracy is directly dependent on the amount of information used to train the network. The network requires a large amount to learn correctly **Black Box:** it is incredibly difficult to understand the network within this type of model.
Support Vector Regression	**Rationalisation and training efficiency:** As opposed to other methods the probability of local optima during training is unlikely since Quadratic programming is formulated in the development. This is rewarded with the goodness of fit with unseen information. **Outlier Suppression:** Unlike other methods, SVRs consider trade-off parameters when processing information	**Parameter selection:** One of the major drawbacks for SVRs is the correct implementation of the Kernel function. The function represents the data in the Hilbert space. Other parameters need to be adjusted to reach the desired fitting. **Computational Effort:** SVRs usually have good performance with limited data. With an increased magnitude of data, the time required to solve the dual optimisation problem and Lagrange multipliers increases significantly.
K-Nearest Neighbour	**Simple and effective:** This is a well understood methodology. The decision rule is extensively researched. The number of classes reduces the error to twice that of a Bayes probability. **No assumptions:** Unlike other methods K-NN is a non-parametric and therefore requires no presumptions to be made. **No training necessary:** This methodology does not explicitly build a model. It uses previous information to infer new data.	**Slow algorithm:** As the dataset increases the inverse will happen to the efficiency. Especially with increased dimensionality, the accuracy of the algorithm declines. **Outlier sensitivity:** The K-NN algorithm is hypersensitive to outliers as the neighbours are selected based on the selected distance criteria. **Unkonwn amount of clusters:** It is difficult to determine the right amount that produces the most meaningful results.
Dynamic Bayesian Network	**Overfitting control:** Since Bayesian statistical methods are encapsulated in the model local maximums can be avoided. **Unaffected by parameterization:** DBN looks for interdependence hence they are not adversely affected by more variables.	**Unknown amount of arcs:** For many cases the number of arcs and nodes could be in the thousands hence a closed-form solution cannot be achievable.
Gaussian Process Regression	**Directly captures uncertainty:** In regression models the prediction is given as a value and a distribution. This is not directly captured in other methods. **Prior knowledge implementation:** When designing the model, if the path of the trend is known a specific kernel function that correlates can be implemented.	**Computationally demanding:** For more than a few hundred data points the time to process this information is longer than other methods due to the cubic inversion issue.

.

2.6 Discussion

There have been a variety of different machine learning methods' main features discussed. The five popular methods' strengths and limitations for; ANN, SVR, GPR, DBN, and K-NN are compared in Table 2.

Given the capabilities and limitations of the above models, it is difficult to advise a specific type. They can all provide reasonable results for condition-based maintenance. Whether you are looking to understand the contributions towards the outcome or not would be a consideration among many. One main significant capability all models have is the ability to develop individual predictions from historical data, something that cannot be carried out with conventional statistical methods or physical models. This comes with the cost of computational effort in comparison to statistical data fitting. Conversely, physical models can be too computationally expensive and hinder productivity but can provide more insight.

The majority of discussion has strictly tackled the individual ML models applied to condition monitoring, but in more recent studies which include multi-agent systems, combinations of algorithms used to gain improved results, is gaining traction in the offshore wind industry. Xiuxing Yin [220] effectively applied this technique using multiple ML methods gaining impressive results in predicting the trust of the operational wind turbine.

Another interesting aspect is bringing combinations of condition monitoring methods. Multi-agent systems also share features that are used to determine different objectives. M'hammed et al, [162] applied both clustering techniques measuring the vibrations of multiple components and an alarm assessment to determine the most effective maintenance strategy.

The challenges that arise from condition monitoring systems of offshore wind turbine systems are discussed. Currently, vibration-based condition monitoring depends on advanced signal processing methods coupled with expert insight. Concerning SCADA-based condition monitoring, there are various techniques addressed. The principal methods in addition to novel ideas are gathered and presented as follows:

- **Trending** methods have demonstrated an effective ability at detecting anomalies. The specific cases have highlighted those specific configurations and interpretations that are necessary for different machines. Automated trend monitoring systems are unlikely to provide accurate and adequate alarms.

- **Clustering** is a more effective method than **trending** at determining the differentiating from normal operation and anomaly but still has similar limitations. This also requires historical failure information to develop an effective diagnostic tool and it is unlikely that the full extent of faults is determined.

- **Normal Behaviour Modelling** is the main focus for SCADA data condition monitoring because of the simplicity of anomaly detection, by using the trained model modelled versus the measured variable. This article has detailed machine learning methods that can be used to determine failures effectively. But there is no clear comprehensive technique described as the best. From the variety of cases reviewed, it was difficult to determine with certainty if the fault detection technique is lacking or the method used to determine the behaviour. The specific type of configuration of the model also remains unclear to determine if specific issues like overfitting are present. There is also a lack of published standards to understand the performance of individual models to benchmark tools.

- **Damage Modelling** is effective at specifically looking at the physical cause of failure for offshore wind turbines. This will be no easy feat since reliable and accurate damage models encompassing all failure modes are difficult. There is limited information on this with limited studies being published. Inferring different models for varying turbines for different environments cannot be examined yet.

- **Alarm Assessment** combined with probabilistic methods and or physical rules has provided promise at reducing the frequency of alarms to more a critical catalogue of alarms. The studies have shown limited industrial developments that back this. Expert systems with fuzzy interference working together can interpret complex information and deliver easily understood information.

- **Performance Monitoring** tools effectively determine the difference in power output from the environmental conditions, operations and health of the turbine. This tool is necessary to determine the performance of the offshore wind turbine and aid the optimisation of power extraction and maintenance.

2.7 Conclusion

This article has given a systematic review of how monitoring for maintenance can be carried out, detailing specific measures that can be taken for condition-based maintenance. This article has addressed the types of machine learning, providing a review of models suited to condition maintenance for offshore wind turbines. Bringing both of these components together is a trend across all industries, the offshore wind industry is catching up to the oil sector or aerospace. There are consistencies across all sectors; there are no design principles for implementing these tools and there is no consensus on the best-suited machine learning methodology.

Monitoring and maintenance carried out on an offshore wind turbine are implementing vibrational analysis, but limited studies are incorporating the vibration analysis coupled with SCADA data to determine faults. This journal paper has discussed in detail how machine learning can be used to detect faults with a variety of approaches. Another powerful tool of machine learning is that specific faults can be categorised into alarms for immediate action to be taken.

A common issue when implementing data-driven approaches is the prepossessing of the data. Since there are no specific standards on how this is carried out, each individual producing a ML model may have dissimilar results when using the same architecture. This could be a result of many different approaches to anomaly detection, the variation in the opinion of experts, or the magnitude of dimensional reduction among some possibilities.

Lastly, current methods usually require experts to implement and understand the results of the monitoring systems. Exploring the automaton of condition monitoring systems could remove the complexity of the results decreasing the ambiguity on the condition of the asset.

Chapter 3

Data Collection and Processing

3.1 Data Processing

Standardisation of Wind Turbine SCADA Data Suited for Machine Learning Condition Monitoring, Innes Murdo Black, Athanasios Kolios, 2022, Proceedings in Marine Technology and Ocean Engineering, Trends in Maritime Technology and Engineering

3.1.1 Introduction

Europe installed 14.7 GW of new wind capacity in 2020, which is 6% less than in 2019 and 19% less than expected, but COVID-19 struck. That is not to say that Wind energy in Europe is slowing down, quite the opposite. A realistic expectation for installed capacity across Europe over the next five years is 105 GW [53]. That is a doubling of the annual installation rate for offshore wind from 3 GW to 5.8 GW. The United Kingdom is expected to install a capacity of 18 GW where 15 GW will be offshore. Wind energy reduces the cost of energy (COE) and becomes a more desirable investment. A significant portion of the investment is operations and maintenance accruing to 30% of the overall cost [42]. This cost is decreasing, one reason being, trends towards direct-drive wind turbines. Wind turbines with no gears from the blades to the generator, consequently have zero gear losses and have the advantage of less maintenance and repairs but, they are larger [44]. Another incipient machinery fault detection before they become a catastrophic failure. Unexpected failure is directly related to wind turbine downtime and loss of revenue. Hence, effective condition monitoring (CM) for optimal maintenance actions is critical.

There are two main schools of thought for wind turbine maintenance: corrective, passive, or preventative, active. Corrective maintenance involves fixing issues on the occurrence of failure. Preventative maintenance is performed before the actualisation of failure. This can be further described by; scheduled maintenance, where fixed intervals based on the product's expected life are carried out. Another is Condition-based maintenance involves continual structural health monitoring of components to optimise maintenance schedules. Condition-based maintenance intends to prevent failures, reduce scheduled maintenance, and reduce maintenance costs [198].

Condition-based maintenance can be carried out in the following ways: trending, damage modelling, alarm assessment, performance modelling, and clustering see [22] for a breakdown of theses. Digital-enabled asset management performs the former tasks using computational methods. One requirement for computational intelligence is well-organised data. The handling of data in current IEC 61400 1 is limited only to the acquisition and implementation of condition monitoring but no clear standardisation of data for the implementation of computational methods is addressed to date. That is the purpose of this paper. To begin the discussion of standardising pre-processing for machine learning methods of wind turbine condition monitoring.

This study proposes a comprehensive data organisation process to enhance the accuracy of wind turbine condition monitoring models. To demonstrate the effectiveness of our pre-processing approach, we develop a trending condition monitoring model based on MET MAST data and a wind turbine's operational output power. The model's performance will be assessed by comparing the predictions obtained using pre-processed data with those using raw data. Section 3.1.1 introduces the scope, purpose, and motivation of this paper. Section 3.1.2 will provide some research background starting with current data acquisition and supervisory control (SCADA) standards for horizontal wind speed and power. Also, a detailed overview of trending condition fault detection. Section 3.1.3 pre-processing procedure. Section 3.1.4 details the machine learning models used for the trend condition monitoring. Section 3.1.5 demonstrates the improvement of the pre-processing of the raw data in a case study using SCADA data and METMAST data from a wind farm, applying the data sets to the fault detection models. Rounding things off in section 3.1.6 with a conclusion.

3.1.2 Formulation of the problem

Industrial standards are set by the IEC 61400 design requirements. Another contributor is the DNV GL, equally as important, but has a different philosophy. But both do not to this date, sufficiently address the concerns of modern machine learning model development for digitally enabled asset management. The challenge of standardising pre-processing is broken down into two sections, firstly how SCADA is gained and secondly, how it is implemented in digital-enabled asset management.

Supervisory Control and Data acquisition

To perform effective digital-enabled asset management of a wind turbine requires information during the operational lifetime. Fortunately, a significant effort has been placed into gathering data throughout an operational wind turbine from oil temperature readings of the gearbox to measuring the vibrations throughout the wind turbine blade. This has helped, those who have access to data, to apply machine learning methods for a range of purposes. Mostly to reduce the operational expenditure and increase the return on investment.

The magnitude of the power extraction is based on two aspects, the operational input, and the meteorological factors. The focus of this case study is on the pre-processing procedure and not how all data is gathered, only the wind speed and power are considered to highlight the process. However, for performance-based condition monitoring more meteorological and operational features could be included to improve the model. The acquisition of information is the same, just the instrument, and the validation setup vary. Current standards for gathering meteorological data are described in [90].

The set-up for wind speed measurements must be at the hub height, and it is recommended in [90] to include wind shear and veer, since these are sources of uncertainty

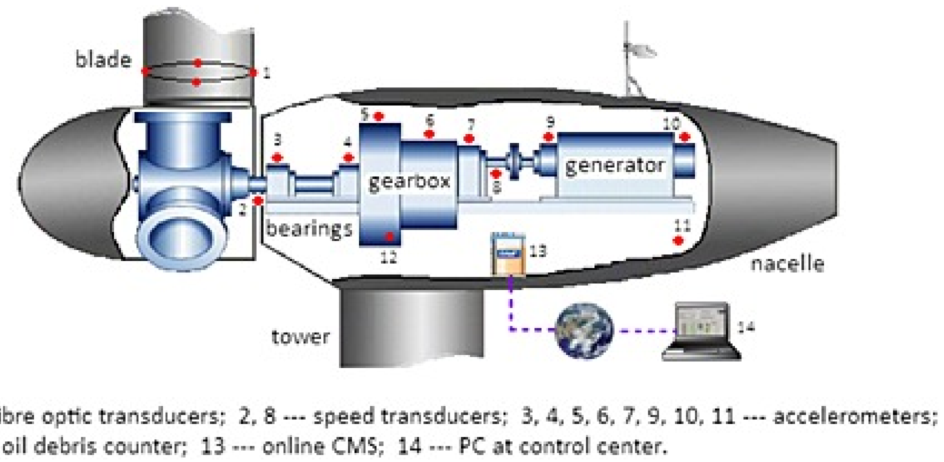


Figure 3.1: A simplified graphic of classic sensor placement on the nacelle for a wind turbine [219]

for horizontal flow. The rotor equivalent wind speed can be measured to reduce this uncertainty, but it may require another sensor configuration [90]. The wind speed is calculated using a cup anemometer for all power performance measurements and can be mounted in a variety of ways. There are specific requirements for the positioning of monitoring system's mountings. Figure 3.1 is a classical representation of the various sensors and their placement within the nacelle. In this example, the anemometer is placed on top of the nacelle.

An important aspect of all measurements from the supervisory control and data acquisition systems is the calibration. The sensor must be calibrated before installation in a wind tunnel to check that it can maintain the validity of its measurement period. If required, it is recommended in [90] to perform post-calibration. Comparing initial results with the new information from the site and altering accordingly.

Certain features are more sensitive than others, for example, the air temperature does not fluctuate as much as say the wind, hence this is represented by higher fidelity measurements with wind speed being at least 1 Hz and the air temperature being of less importance. This information is aggregated into 10-minute periods derived from the continuous string of data. When the data is aggregated it must include statistical information such as mean, standard deviation, min, and max value. Some data rejection measures are included to ensure data obtained from the operation is only included, such

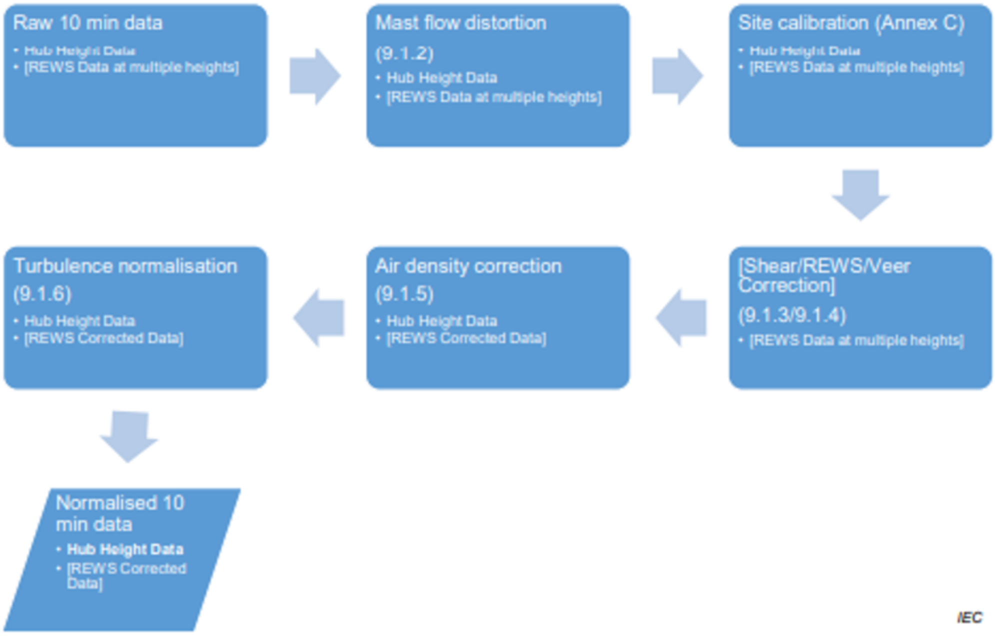


Figure 3.2: Process flow for wind speed pre-processing carried out in the IEC. 61400-11, [90]

as external conditions in particular wind speeds out of the operational range, manual shutdown, or specific atmospheric conditions.

After aggregating into 10-minute intervals, the data is normalised in five processes. These stages aim to improve the accuracy of the wind speed. The process is detailed in Figure 3.2. It includes references to each process. This entire process is to reduce the uncertainty from the wind turbulence in a wind farm.

An accurate estimation of the power is important for the power trending of the operational wind turbine. The net electric power measurement device is usually a power transducer. This establishes the voltage and current on each phase which can then be calculated into the active power. A class of 0.5 or higher must be implemented for MW wind turbines. Measuring the accuracy of the transducer can be calculated by following the procedure detailed in the IEC standards [92].

The data acquisition for all features is the same where higher frequency data is reduced to 10-minute intervals and the statistical attributes are calculated. The process is as follows, the sensor must be of a specific quality to ensure that the readings are of a certain standard. Once the sensor is placed on the operational wind turbine it will then be post-calibrated to ensure the readings are as expected. Recordings are taken, usually 1Hz, and then the method of bins is applied. For complex parameters, a normalisation procedure will be included. Wind speed for example. For more information on SCADA data on mechanical loads and Acoustic noise signals see [91] and [89] respectively.

Trend Condition Monitoring Models

Trend monitoring systems evaluate the short-term and long-term trends in, the performance, oil temperatures, and mechanical behaviour of wind turbines. Computational models are then used to determine these trends and they are compared against the actual value from the SCADA systems.

In the paper, [211], it has successfully applied SCADA-based condition monitoring using an artificial neural network (ANN) applying the temperature of the oil in the drive train to reveal the patterns in the data. In this example, the data is just normalised. Another project carried out in [31] uses a self-organising map, this is an unsupervised learning ANN technique. There is no mention of the pre-processing procedure other than using the industry-standard 10-minute average data for 21 features. The self-organising map is used to determine the abnormality in the trends of, generator temperatures, reactive power, and gear winding temperatures.

The study from [192], performs a trend analysis on a direct drive wind turbine. In this instance the handling of the data is more complex, there is a variety of techniques used. Firstly, the bootstrap method is used for outlier detection, then the feature importance is determined using a random forest algorithm. After these methods are applied the most relevant features are used to train a deep neural network for the trend analysis.

SCADA monitoring can be carried out in a variety of ways. [211], uses the torque/rotational speed of the high-speed shaft, the RPM ratio from the low speed to the high-speed shaft, and the wind speed to determine the output power of the operational wind turbine using an adaptive ANN. In this project, the data is; re-sampled to 1 Hz from 100Hz, clustered into regions using a self-organised map algorithm, and then split for training and testing purposes.

The Standards on trend analysis for wind turbines are limited. DNVGL-SE-0439 details the assessment of the validity of condition monitoring but no procedure. The IEC 61400-25 series on communications for monitoring and control of wind power plants offers insight into the transfer of information from a server to a client. IEC 61400 25-3, discusses a binning process again for vibrations on the power to aid in the process of alarms to deal with the no-linear data. Again, there is no beginning-to-end condition monitoring process described.

Trend condition monitoring can be carried out in a variety of ways. Understandably, there are no standard models since the process of predicting values is rapidly changing with machine learning methods. The one area that is of concern is the handling of the data from server to model implementation. The articles discussed have included a pre-processing procedure and have not. How can one determine if that model is only suitable for that one data set, a phenomenon called over-fitting, or in fact, is that model universally applicable? Standardising the pre-processing of dirty data and normalising it will aid in overcoming this phenomenon and streamline the condition monitoring model pipeline.

3.1.3 Pre-processing Procedure

Computational algorithms require the presence of data in a mathematically feasible format, pre-processing is how that is achieved. Data processing techniques consist of data reduction, data projection, and missing data treatment. Data reduction aims to reduce the size of the data set through feature selection or case selection. Data projection intends to change the appearance of the data by scaling the features. Missing data treatment includes deleting missing values, and/or imputing them with estimates. These techniques will follow the process in Figure 3.3.

Effective SCADA data indexes are timestamped, and if it abides by the IEC standards it will be in 10-minute intervals at least. Depending on the provider the nacelle information, acoustics, mechanical loads, or METMAST data may be in different files. Combining all these elements in a way that every row aligns with a consistent timestamp guarantees a cohesive set of relevant parameters. In cases where multiple data sets have different frequencies, the entries are compiled so that the index remains uniform

Depending on the task knowing the property of the column may be important, for supervised learning, removing unnamed columns may be helpful, for unsupervised techniques discovering patterns and information is the main task, and removing columns will hinder this.

To deal with impurities (NANs, missing values, and anomalies) there are plenty of techniques [128] discuss these in detail. Missing data treatment deletion methods include listwise or pairwise deletion. Pairwise the most popular according to [86], but this can lead to mass loss of information and may impact the results [196]. Alternatively, imputation methods, according to [182] it is more effective since data is not lost. This can be carried out by; K- nearest neighbours, mean imputation, hot-deck imputation, cold deck imputation, regression imputation.

This paper includes the use of multivariate imputation using K-nearest neighbours, [71]. The method applied implements the Minkowski distance to impute a value for the missing number or NaN for that feature by clustering data that has the lowest distance to one another. The Minkowski distance has the form:

$$X = (\sum_{i=1}^{n} |x_i - y_i|^p)^{1/p} \tag{3.1}$$

The Minkowski distance, also known as L2 when $p = 2$, is the distance between the two arbitrary points of $x_i - y_i$. Sensors do not have 100% up-time and are subject to failure and replacement. This technique deals with missing data and or NaN values.

Recently there has been significant work put into outlier detection, it can be carried out using; statistical methods, regression methods, or kernel-based methods. [178] provides an in-depth analysis of these techniques.

In this process, some insight has been included. For example, by setting hard limits on some of the variables such as wind speed or power. Another is knowing when a feature is a scalar and it can only have a positive value. Sensor errors are expected hence, removing these entries and then applying the missing data treatment can improve the data quality.

Data projection is the process of normalising the data to exact, specified, and repeatable conditions. Scaling the values according to a defined rule such that all the features have the same degree of influence and thus the method is immune to units. Normally intervals of [0,1] or [-1,1] are used for the target. [86], has observed that [0,1] is mostly used and that is what has been implemented in this process. Calculated by:

$$[0,1]interval = \frac{x_i - min_x}{max_x - min_x} \tag{3.2}$$

Feature selection is the process of selecting a subset of properties listed in the data. The data set may be of significance or irrelevance to the evaluation of the targeted outputs assuming that the data included has an impact on the accuracy of the model. In general, there are two methods filters and wrappers. Both evaluate the preset criteria independently before the machine learning method. Some of the most common methods comprise statistical methods such as Pearson's correlation [99] or decision trees such as xgboost [36] or LightGMB [55]. To deal with the noisy data from an operational wind turbine A RandomeForest decision tree is implemented to perform the feature selection. The information gained from the RandomeForest is calculated by:

$$G(T,X) = Entropy(T) - Entropy(T,X) \tag{3.3}$$

Where T is the target and X is the feature split in its simplest form. Entropy is a measure used to evaluate the impurity or disorder in a set of data points within a decision tree. The equation implemented is:

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)C_{right(j)}} \tag{3.4}$$

Where ni_j is the impotence node, wj is the weighted number of samples reaching j, Cj is the impurity of the node, C_{left} is the child node, and $C_{right(j)}$ is the child node from the right split. This will provide a number from 0-1 with the higher the gain the more significance that feature has on the desired output parameter.

The final technique for pre-processing is data splitting. This is one of the main procedures on how to evaluate the final model. Splitting the data into training, validation, and testing. The potion of these depends mostly on the size of the data set.

The process flow for the implementation of all the techniques is addressed in Figure 3.3. This considers both supervised and unsupervised machine learning methods applied to operational wind turbines. There is scope for alternative techniques within the framework, as all standard models should. But the main takeaway from this standard process is how the data is concatenated, the known-known impurities removed, and the known unknowns are controlled using feature selection and imputation. Finally, the data is scaled to provide a level playing field for all machine learning methods and continuity for other developers to work from.

3.1.4 Machine Learning Models

There are vast and diverse amounts of regression machine learning methods. Starting with linear regression to more complex methods using transfer learning techniques. In this article 5 different methods have been included for their diversity, and different approaches to determine similar results highlight that this process will aid in achieving improved results. Starting with an artificial neural network, RandomeForrest, Gaussian process regression, xgboost, K-nearest neighbours, and finally Support Vector Machine. This section will briefly describe these processes and the significant technique used for that process. The ANN is built using the Keras Application Programming Interface (API) [2], and the rest of the models use the Scikit-learn API [141].

Artificial Neural Network

Artificial neural network (ANN) is a technique that takes inspiration from the functioning of the human brain. Where the neutrons work in parallel passing and storing information depending on the synaptic weights(activation function). The concept was first introduced by, [122]. Back-propagation [161] revolutionised ANNs since it reduced computational time and efficiency.

More recently, the optimisation of activation functions improved, with different

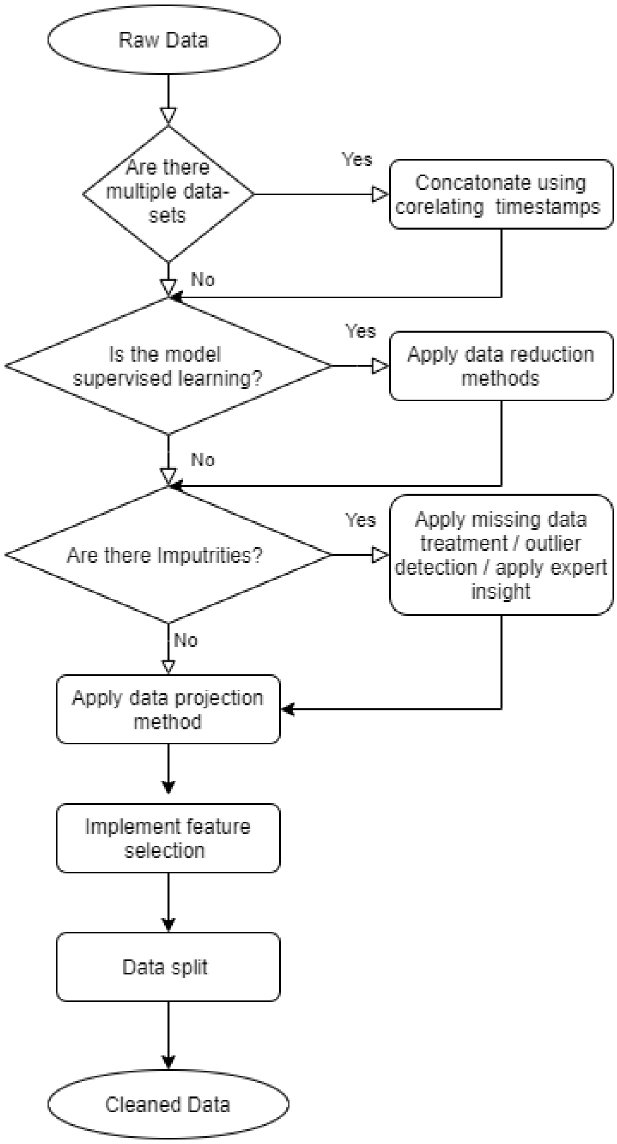


Figure 3.3: Flowchart detailing the procedure of taking raw data and applying techniques to clean data.

functions representing them, and the optimiser determining them. This article uses the adam method for stochastic optimisation for the activation functions described, [104]. Adam, the adaptive learning method computes the learning rates for different parameters. Derived from the adaptive moment m_n to the n^{th} power,

$$m_n = E(X_n) \tag{3.5}$$

X is a random variable. To estimate this, firstly, moving averages of the gradient based on mini-batches are calculated using:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{3.6}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{3.7}$$

Where v_t and m_t are moving averages, g_t is the gradient on the current mini-batch. $B_1, 2$ are the new additions from previous methods. These hyper-parameters have default values of 0.9 and 0.99 respectively. Since m_t and vt are estimates of the first and second moments, they have the following property:

$$E(m_t) = E(g_t) \tag{3.8}$$

$$E(v_t) = E(g_t^2) \tag{3.9}$$

As you expand the value of t a fewer number of values of the gradients contribute to the overall value. As they get multiplied by a smaller beta. This pattern is captured by:

$$m_t = (1 - \beta_1) \sum_{i=0}^{t} \beta_1^{t-i} g_i \tag{3.10}$$

Introducing a bias correction term for the Momentum equations:

$$E(m_t) = E(g_i)(1 - \beta_1^{t-i}) + \epsilon \tag{3.11}$$

This bias estimator is not just true for just Adam optimisation, this estimator holds for SGD with Momentum and RMSprop among some methods.

Support Vector Machine

One of the most influential approaches to supervised learning is the support vector machine [24], [40]. One key innovation is the 'Kernel trick'. This consists of observing that many machine learning methods can be written as the dot product between examples. The kernel applied in this report is the radial basis function kernel [179]:

$$K(x, x') = exp\frac{|x - x'|^2}{2\sigma^2} \tag{3.12}$$

Where two examples x and x', are represented as feature vectors for an input space.

K-Nearest Neighbours

K-nearest neighbours are the first of the non-parametric models implemented in this paper, where the complexity is a function of the training set size, unlike linear regression, for example, this has a fixed-length vector of weights, the nearest neighbour regression model simply stores the X and y from the training set. When asked to classify a test point x, the model looks up the nearest entry in the training set and returns the associated regression target. In other words:

$$\hat{y} = \sqrt{(x_i - y_i)^2 + (x_{i+1} - yi + 1)^2} \tag{3.13}$$

The algorithm can also be generalised to distance metrics other than the L 2 norm, such as learned distance metrics [152].

Gaussian Process Regression

The Gaussian process regression (GPR) is based on [152]. This incorporates Bayesian interference, cross-validation, and Gaussian noise. The kernel used is the squared exponential with Gaussian noise, as stated by, [136] this is an effective method for extracting results for operational wind power regressions. It is described by:

$$K_{se}(x,x') = \sigma_f^2 exp\frac{|x-x'|^2}{2\sigma^2} + \sigma_f^2(x,x') \tag{3.14}$$

where σ_f and σ are defined as hyper-parameters that represent the signal variance, and is a characteristic length scale that signifies how quickly the co-variance decreases with the distance between successive data points.

Light Gradient Boosting Decision Tree

The last method implemented is the light gradient boosting decision tree (LGBM), this method is developed based on the work from [102], on xgboost. But it differs in the growth of the trees. It is leaf-wise growth as opposed to level-wise tree growth. The main principle of LGMB is to have the minimum objective loss, $obj(0)$:

$$obj(0) = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3.15}$$

Where K is the number of trees, f is the ensemble of decision trees. Each tree in the ensemble contributes to the overall mode, hence, f is all possible classification and regression trees (CART) - a tree ensemble model consists of a set of CART. The loss function used in this model is the $L2$, with the regularisation term Ω:

$$\omega(f) = \tau T + \frac{1}{2}\lambda \sum_{j-1}^{T} \omega_J^2 \tag{3.16}$$

Where T is the number of leaves, w is the vector score of the leaves, tau and lambda are tuning hyper-parameters that require optimising. The growth of the tree structure is dependent on the gain:

$$gain = \frac{1}{2}\frac{G_L^2}{H_L-\lambda} + \frac{G_R^2}{H_R-\lambda} + \frac{(G_R+G_L)^2}{H_R+H_L+\lambda} - \tau \tag{3.17}$$

This formula can be decomposed as 1) the score on the new left leaf 2) the score on the new right leaf 3) The score on the original leaf 4) regularisation on the additional leaf. We can see an important fact here: if the gain is smaller than tau, one would do better not to add that branch.

3.1.5 Case Study

To validate the proposed framework for standardisation of SCADA data, a case study on an operational wind turbine is presented. The wind turbine considered is the Gamesa G97-2.0 MW IIA/III, designed for low to medium winds. The data set of the wind turbine consists of 136 features, 39 of which are related to the meteorological mast data METMAST. The other 97 features consist of condition monitoring sensor data, CMS data, and parameters such as low-speed and high-speed generator RPMs. The data used in this report is of 10-minute internals with 102,082 time steps (709 operational days). One can only assume that this data was taken from a higher frequency and processed into 10-minute averages as discussed in section 3.1.2, since most sensors can gather at a frequency greater than every 10 minutes. Only the power produced by the generator and the METMAST are included in this report for the power estimation model.

The data for this wind turbine has two significant modes operating around 1 MW and 2 MW, these two states are highlighted in the histogram distributed on the Y-axis. As expected, the wind speed distribution represents a standard normal distribution. The main concern is the distribution of the data points within the power. The power data should represent a shape close to a Weibull distribution factors such as wind shear or operational philosophy may deter the distribution from that. In this case, there are two prominent modes, with incredibly noisy data indicative of poor quality data.

The distribution of data in Figure 3.4 and the potential cause of outliers are discussed in detail in Section 3.1.2. The challenges with this problem are the quality of the data and the processes used to rectify it may not be suitable for other models. The second aspect is, what strategy should be applied when there are multiple data streams of information being concatenated together. And lastly, how should one deal with the errors that are present in the transfer from the sensor to the server and then to the client, some of the errors are highlighted in Table 3.1 for this data set.

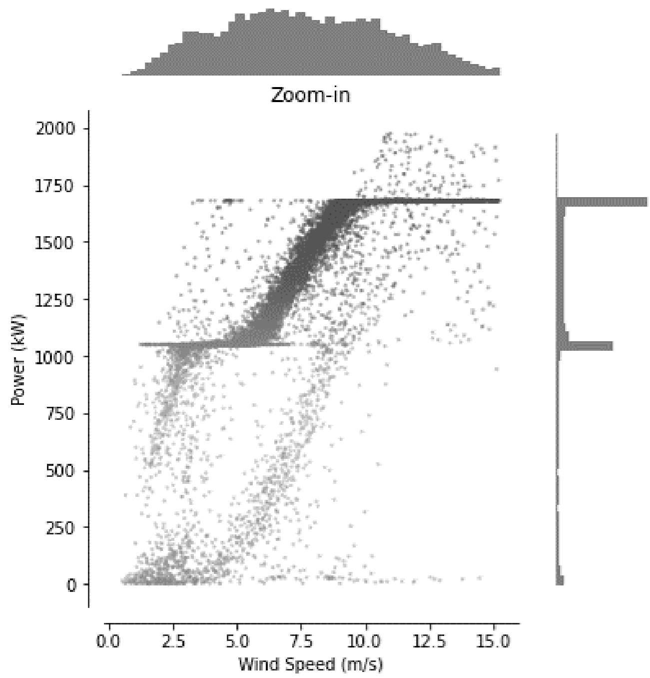


Figure 3.4: Scatter plot, with histograms on the x and y axis of unprocessed data, highlighting the issue of this data set.

Table 3.1: Table of a few features from the SCADA data

Feature	**Missing**	**Missing(%)**
Instantaneous air pressure 10M calc.(mmHg/mb)	18658	18.28
Instantaneous relative humidity 10M calc. (%)	19338	18.94
Instantaneous temperature 10M calc.(ºC)	19800	19.4
Maximum direction 10M calc. (Height1) (°)	18658	18.28
Maximum horizontal speed 10M calc. (Height4) (m/s)	18658	18.28
Minimum horizontal speed 10M calc. (Height1)(m/s)	18658	18.28
Unnamed: 0	102082	100
......	......	......
......	......	......

The results section will answer all these issues and highlight the need for a consensus on how to deal with SCADA data for machine learning algorithms.

Error Assessment

In this case study the regression results are assessed by the $R2$ score, which represents the variance in the dependent variables that is predictable from the independent variables. This has the form:

$$R2 = 1 - \frac{\sum_i (y_i - \hat{y})^2}{\sum_i (y_i - \mu)^2} \tag{3.18}$$

Where, y_i is the real real values, μ is the mean, y_i, and $\hat{y}$ is the predicted value from the model.

3.1.6 Results

The flow of this section will follow the process highlighted in Section 3.1.3 Figure 3.3. The trending monitoring model will apply all the processes and exclude all of them to compare the results. The input parameters for the model are the METMAST data and the output data is the power.

The first set addresses the issue of the separate two data sets. Fortunately, this data set is timestamped. In this situation, the two data streams are taken every 10 minutes on every 10 minutes of the hour. Making this process a simple search, removing any entries that do not correlate, and concatenating them.

The trending condition monitoring technique used involves a regression, supervised learning machine learning model. Hence, unnamed features are removed.

Preprocessing, particularly handling impurities, offers significant opportunities for innovation. This study employs diverse techniques, recognising that optimal approaches may vary across datasets. Acknowledging the potential existence of newer methods and emphasising the value of transparently documenting both the specific impurity-handling strategies employed and the rationale behind them.

The first step is dealing with missing entries, this report implements K-NN a multi-

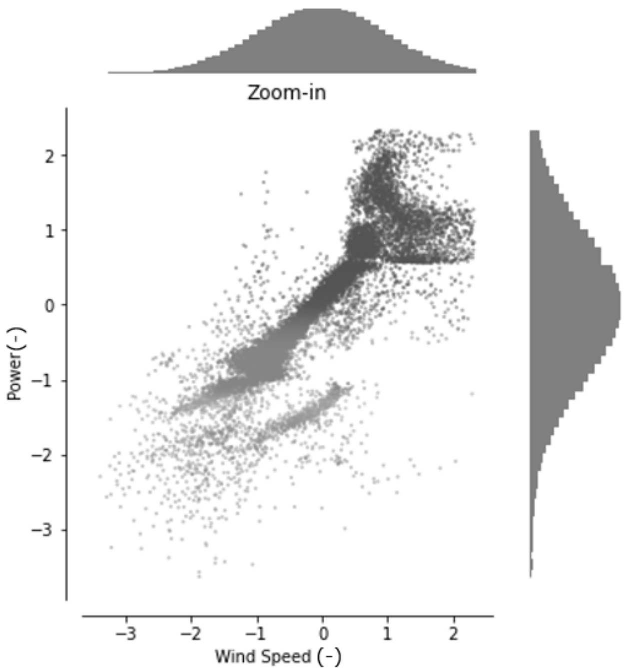


Figure 3.5: Scatter plot, with histograms on the x and y axis after processing.

variate imputation method. As opposed to this, missing entries must be removed from the raw data. Resulting in a smaller data set. For machine learning methods, the larger the data set tends to improve the performance of the model.

Dealing with outliers is important. It must be done with care, as one does not want to remove too much variance in the data such that the model overfits and cannot provide effective results from new, unseen data. The first procedure is removing all vector data from all the scalier features. Feature projection is especially important in the case of this data, with the power having two modes with significantly more data points. When training models this can lead to reduced performance. Figure 3.5 highlights how projection methods can redistribute data to improve performance. Where the projection method has removed the two modes and redistributed the values from -3 to 3. This has improved the ability of the models to learn the task. This technique is only applicable to this type of model since the data is in 10-minute intervals. When implementing this into a continuous deployment environment new data points can be incorporated into the previous data set and then converted to the quantile Gaussian

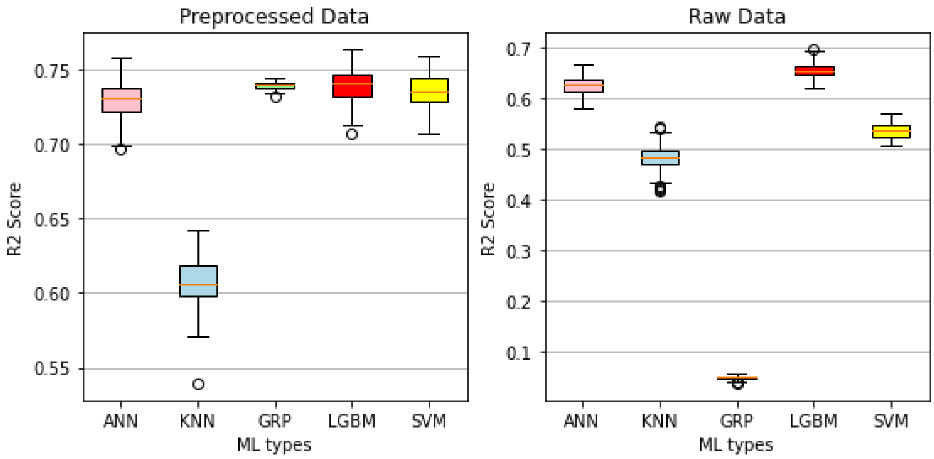


Figure 3.6: Direct comparison of raw vs clean data detailing the distribution of R2 scores after the Monte Carlo simulation.

distribution technique with ample time. In higher frequency models such as some financial trading models or vibration signals sub 1 Hz, the computational effort may not be practicable.

The last step is crucial for the evaluation. This is a mandatory step but the ratio of splitting the data should be dependent on the size of the data set. The ratios implemented for training validation and testing are 0.7, 0.1, and 0.2 respectively.

The purpose of standardising wind turbine SCADA data for machine learning is to transform the data into a format that can be used more effectively. Condition monitoring can be as simple as finding patterns in data that do not conform to normal behaviour. The Monte Carlo simulation is carried out over 100 iterations for each of the machine learning models to determine the r2 score variance. The pre-processing improves the accuracy with which the data is observed across all models, highlighted in Figure 3.6. The most significant improvement is the GRP improving from less than 10% accuracy to 73%.

3.1.7 Conclusion

Establishing the power trend for a wind turbine is complex due to the unpredictable nature of wind and the operational modes defined by the turbine. However, this paper introduces a trend condition monitoring method using a global model, which employs METMAST data as input and power as output data. Five machine learning techniques, ANN, K-NN, SVM, LGBM, and GPR have seen enhancements through the standardisation process, significantly improving both the accuracy and variance for this data set.

The work in this paper has focused on trending condition monitoring, and the techniques implemented on the standardisation of SCADA data. This work will apply to machine learning methods within the wind turbine SCADA data realm.

This paper has highlighted that there is an inconsistency in the application of pre-processing SCADA data for operational wind turbines. This paper addresses this issue, the process should follow the predetermined process outlined in section 3.1.2 and this should be documented. This paper has highlighted a method from beginning to end that should be implemented as it will improve the accuracy of your results, highlighted by applying this to multiple modeles. However this process is not the final form, the steps taken are necessary but the techniques in each process are interchangeable. The purpose of this procedure and study is to highlight the need for this process in standards so that the wind turbine industry has consistent data pre-processing, and the focus can move on other aspects, such as model development, removing doubt by highlighting the effective procedure. Collaborative machine learning model development, since there will be a standardised base for all model development. And lastly, transparency in model evaluation since the procedure of how the data is organised will be less ambiguous.

Chapter 4

Case Study of an Operational Wind Farm

Population-Based Structural Health Monitoring: Investigation into the Heterogeneity of an Offshore Wind Farm, Innes Murdo Black, Moritz Werther Häckell, Athanasios Kolios, 2022, Renewable energy, (Submitted)

4.1 Introduction

In conventional structural health monitoring (SHM), a model is developed using data recorded from an individual wind turbine [6]. It is expected to facilitate generalisations on future measurements for that specific system. However, for a single structure, difficulties are presented in gathering comprehensive data as typically only a fraction of the information is available for a given structure. Hence, only a small percentage of the environmental and damage conditions can be assessed. If a framework can transfer complete information from one structure to another in the population, this will allow for diagnostic inference on the second structure without requiring the same data sources.

The population-based approach to structural health monitoring aims to transfer valuable knowledge between groups of similar systems, whether the characteristics of the system are the same or similar within the population that is heterogeneous or homogeneous respectively. This type of population will define the degree of transferable knowledge that can be transferred, and by what means. This work delves into the behaviour of a seemingly identical wind turbine system. While individual turbines may appear alike, we propose classifying them into two categories: those exhibiting similar characteristics (homogeneous) and those demonstrating a wider range of behaviours (heterogeneous). This distinction will help us better understand the system's overall performance and identify potential factors contributing to variability within the seemingly uniform population.

By determining the type of population, homogeneous or heterogeneous, a model can be made to represent the behaviours of the population and infer information on damage between systems. The representation of structures developed in this document is designed firstly to quantify the degree of similarity between the structures and secondly to facilitate the transfer of knowledge via machine learning.

One of the main concepts of conducting PBSHM is that of knowledge transfer. This process is crucial for assorted reasons. Firstly, conventional methods of data-driven SHM using supervised, unsupervised, or semi-supervised machine learning methods assume that the test and training data are drawn from the same distribution. This assumption is questioned in PBSHM as each member of the population will have discrete characteristics (and hence distribution) because of, e.g., environmental variations, manufacturing, and assembly differences, and/or operational conditions. Therefore, conventional methods begin to fail when models transfer information between systems. For example, one SHM model trained on a 5 MW offshore wind turbine will begin to fail when making predictions on a 1 MW onshore wind turbine since the dynamics are different even though they are of a similar form. Pinpointing unique variations within seemingly identical turbines enables crafting a universal and accurate general model.

To establish an effective transfer of knowledge in the context of a wind turbine substructure, one must start by focusing on the responses of a wind turbine from the complex environmental and operational factors. These factors are assessed in the context of operation, material, topology, and geometry as a basis to explain the observed dynamical behaviour differences throughout the population of wind turbines. Some studies have built approaches to deal with PBSHM; the series [26], [75], [66] carries out a detailed investigation on how to tackle the issue. Where the first paper [26] is the introduction to PBSHM, the second [75] introduces the idea of heterogeneous populations and how to find similarities within a population form. The last study in the series implements transfer learning techniques applied to scaled models of buildings to highlight how PBSHM can be conducted effectively.

This study investigates the potential challenges of PBSHM using SCADA and SHM data from a wind farm. Firstly, the fixed boundary conditions from the geometry, topology, operational conditions, and materials are critiqued along with the SHM technique of estimating the fatigue damage equivalent load on the jacket structure. The study observes the effects of environmental and operational conditions on the individual structures of the population and compares this against several wind turbines from the same farm. Lastly, there are some honourable mentions of challenges that are currently out with the monitoring campaign such as scour, marine growth, and corrosion that need to be discussed in the context of PBSHM.

4.2 Population-Based Structural health monitoring

For PBSHM it is relevant to define the contextual difference between homogeneous and heterogeneous populations. This syntax is borrowed from graph theory work from [49] where the names clearly explain how structures can be represented by attributes. To determine whether two systems are similar enough for knowledge transfer, it is unpracticable to consider every property or dimension of the structure - e.g., comparing the geometrical similarity of two structures using 3D, finite element (FE) or computer-aided design (CAD) models of the structure directly would be computationally inefficient. For our desired goal, it is more efficient to consider only the properties and dimensions that have a significant effect on the transferability of knowledge.

The differences in wind turbine forms are a result of manufacturing tolerances and various installation sites which result in diverse design requirements. To contextualize the potential differences, one main area with creative freedom is with the geometry of the design which varies based on the boundary conditions of the location. The final

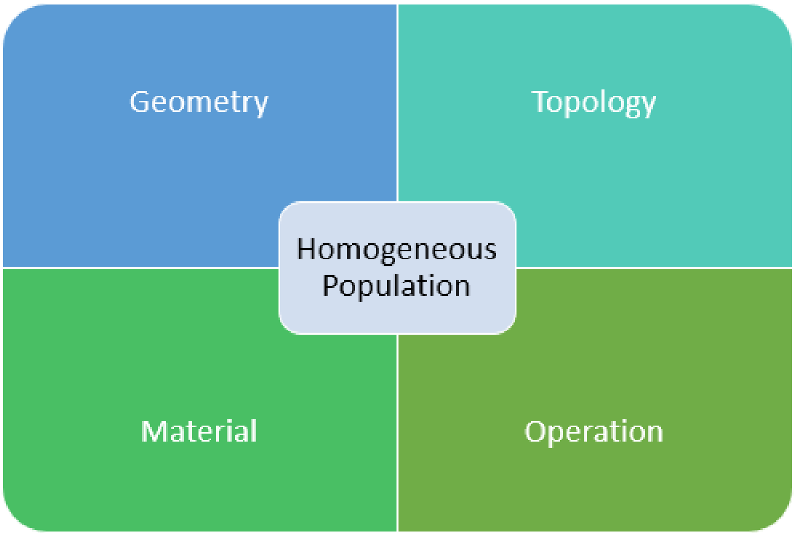


Figure 4.1: Categories of heterogeneous populations within the PBSHM framework, in the centre, where all 4 categories share sufficient similarity, a homogeneous population exists. Noting that all 4 attributes can influence each other separately to create independent heterogeneous populations.

three differences are the topology of the wind turbine (WT): how it operates in context to another WT, where it is in comparison to the other WTs, and the most uniform material which is currently across the designs. Figure 4.1 highlights how the four classes are intertwined, where one class may overlap with another forming a different class of heterogeneous populations. These topics are discussed in the following sections.

4.2.1 The Data

The measurement data used as an input for the low-cost monitoring technique takes high-frequency 25Hz CMS data that is processed into 10-minute averages. Figure 4.2 shows the WTGs available for this study that are both equipped and unequipped with strain gauges (SG) and the associated DEM. The three positions with SG are WT 1,2,3. The box plot emphasizes the 25th percentile, 75th percentile, 95th percentile, and the 5th percentile. Additionally, the black dots represent instances where the values approach one. While these are typically considered outliers, in this thesis, their inclusion has been maintained throughout all aspects of the study.

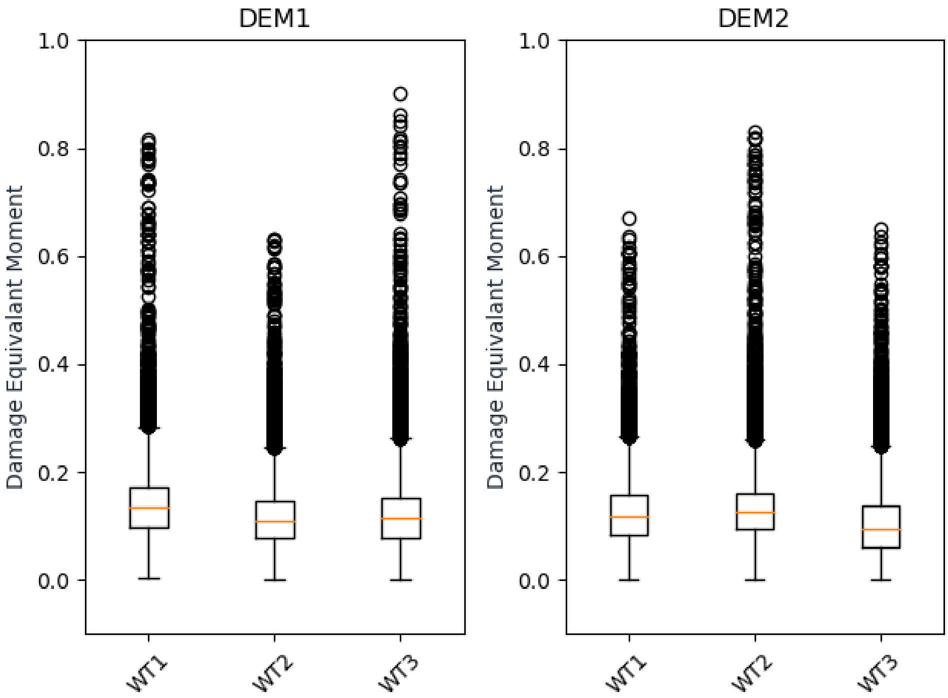


Figure 4.2: Normalized damage equivalent moments box plot of the two orthogonal directions for all three of the wind turbines

SCADA systems are equipped on all WTGs and, depending on the feature, the resolution varies, but it is delivered in the form of 10-minute averages with statistical measurements that include the mean, minimum, maximum, and standard deviation. This encompasses meteorological information at the hub height, such as wind speed, wind direction, temperature, and pressure. The SCADA data also covers the operational signals such as power production, pitch angle of the individual blades and the rotor rotational speed.

To increase the value of a low-cost monitoring program, transferring knowledge that is unavailable in other wind turbines can provide insight and confidence into other assets. If one can infer knowledge accurately on another WT, then one can save money by installing strain gauges on a fraction of the WTG. Based on this principle, the population form is the DEM where only 3 WTG have the CMS strain gauges installed.

4.2.2 Definition of the Population Form

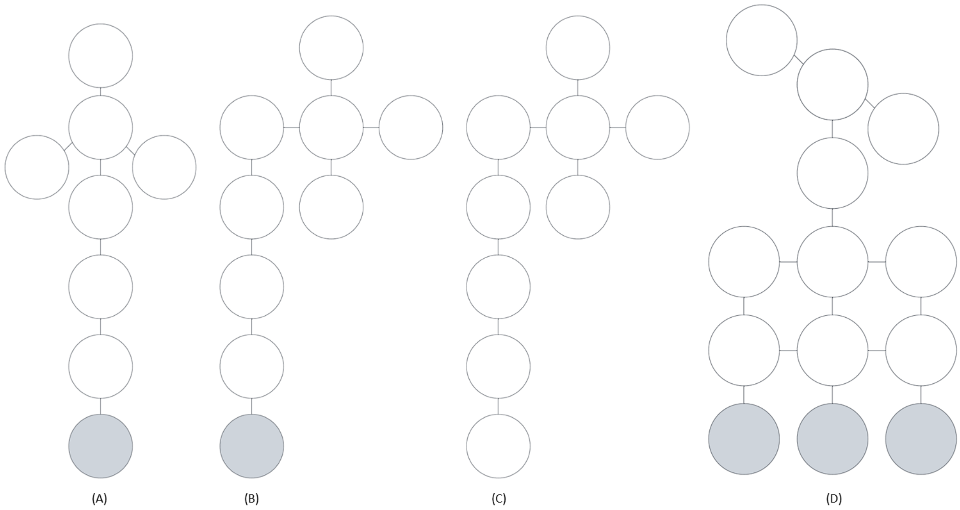


Figure 4.3: Graphical representations of structures, with lines representing edges, circles signifying nodes, and Gray circles representing ground nodes [26]

To begin this discussion, it is important to address the semantics of this study as this lays the foundation of all the discourse ahead. The population form is important as this provides the basis of if and how the information can be transferred. A graph can be used to describe a structure in a simplified representation, this can then inform a relevant measure of similarity between structures. This, in turn, can then be used to determine the level of similarity between systems within a population.

Homogeneous populations are conceptually of structural equivalence. This implies that the graphs used to represent the structures within the populations are topologically equivalent, with nodes in the same position. Examples of graphs represent structurally equivalent and inequivalent wind turbines. For example, 4.3a could be a monopile, so could 4.3b with a different design, 4.3c could be a floating structure as it has no ground node, and 4.3d could be a jacket structure. Figures 4.3a and 4.3b are nominally identical with the same ground node and associated graph. Conversely, Figure 4.3c shows a topologically similar graph, however, it is not structurally similar. Lastly, Figure 4.3d is both topologically and structurally dissimilar to Figures 4.3 a, 4.3b, and 4.3c.

Definition 1 - A population of member structures is homogeneous if individual members are pairwise structurally equivalent, with material, geometric, and physical parameters δ (i.e. graph attributes) that can be considered to be random draws from an underlying base-distribution $p(\delta)$. As a result, the distribution $p(\delta)$ describes population variation in terms of the parameter σ. If any pair of members is not structurally equivalent, the population is heterogeneous [49].

Within the context of homogeneity, it is also important to consider the case of strong homogeneous populations. In this situation where the density $p(\delta)$ connected with the population is of high correlation such that each member of the population is an identical system, this can be considered strongly homogeneous. In this sense, considering a population of systems with the same model, but subject to manufacturing tolerances, a conventional method of SHM can be applied [34], [156].

Fatigue Damage Equivalent Moments

There are a vast number of model and feature spaces that can be applied to represent a population form. For a wind turbine, the form could be wind turbine power curves to frequency responses. But, in this case, the form is fatigue damage equivalent loads for the jacket support structure. The entire population in this study has the same geometry and material but with small deviations in topology due to the location.

The condition monitoring system calculates the forces from strain gauges on the foundation of the structure. From the forces, the damage equivalent loads are produced. The two-phase operation is as follows:

Phase 1 - Calculation of forces from strain.

1. Run dynamic ROSA simulation, processing acceleration data [149]
2. Extract stress at selected element via Fatima [149]
3. Calculate strain using Hooke's law
4. Calculate forces with internal functionality
5. Compare forces with extracted moments

Phase 2 - Calculation of DEM from forces

1. Gather applicable force location from the sensor location
2. Calculate the cyclical forces at that sensor
3. Apply ASTM E1049-85 rain-flow cycle counting algorithm, to summate the loads over time [15]
4. Apply a scale factor to force accumulation.
5. Sum the damage accumulation over the cycles to calculate the DEM

4.3 Methodology

This segment discusses a process of determining the level of knowledge sharing within a wind farm and the degree of transferability within the population form. This process is about collating the relevant information, and the necessary processes required to determine the heterogeneity of the population form. The framework investigates the four elements that define the population form: geometry, operation, topology, and material.

Starting with the documentation on the individual wind turbines within the population form, it is impractical to cross-examine the geometry of each of the wind turbines using finite element analysis to reduce the elements of the main components into a hierarchy of shapes. For example, reducing a jacket connection pile component into the geometry of a beam of cylindrical shape and comparing this to a monopile foundation which would have the same geometry and shape. This would indicate that there is a degree of transferable knowledge, which does not require an exhaustive finite element comparison.

To accompany the knowledge from the geometry, Young's modulus could be used to class the material of the components into a hierarchy. The details and levels of the hierarchy are dependent on the documentation available. The first level would be material class, then material within that class and then properties. The material

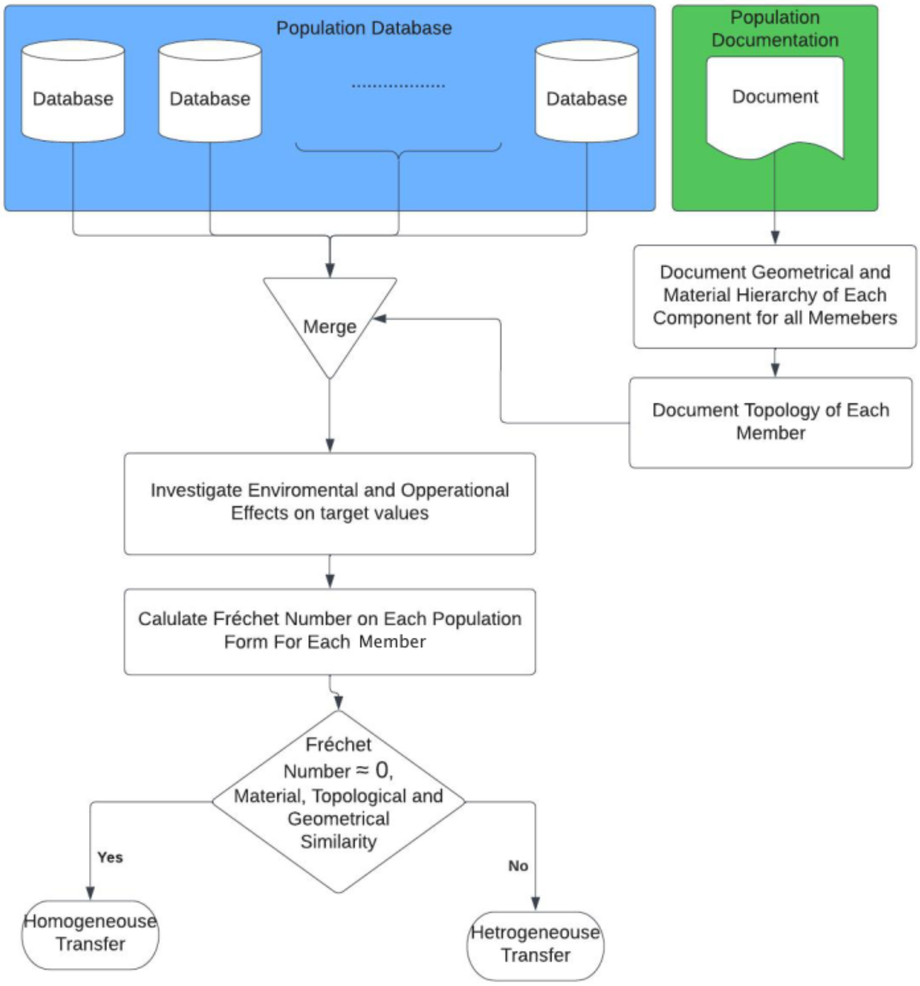


Figure 4.4: This is a guide on the process of determining the degree of transferability of an operational wind farm. The blue section contains all the buckets of data, this could be SCADA data and/or CMS for individual wind turbines. The green segment is all the design documentation available on the individuals within the population.

properties determine whether it is possible to make inferences of damage assessment and classification with labels between two structures. Two materials of the same material class will experience similar failure modes and may exhibit the same material responses giving more confidence in the classification of damage. The greater the similarity between the materials the more likely the assessment of damage will be the same between the two structures.

Merging the knowledge gained documentation with the dataset available is how to empirically determine the degree of transferability. Generating plots of the population form concerning the operational context highlights the operator's philosophy. If the operational modes are similar, then a general model may work. Conversely, this may require that the elements of curtailment may need to be dealt with separately to increase the accuracy of knowledge sharing. By discretising the population form of an operational wind turbine into modes this may reduce negative transfer.

The second element that synergises with the documentation is the comparison of environmental effects on the individual wind turbines within the population. Knowing the location of the individual wind turbines within the wind farm and generating plots of the features from the SCADA data and the form, will highlight how the dynamics vary throughout the wind farm. One important investigation of the environment is the turbulence intensity from different wind directions and how this relates to the similarity of the population form. Based on this, the degree of similarity may be visible. But to reinforce the judgment, the Fréchet number can be used to empirically determine the degree of similitude to the population form.

In the case where the documentation highlights a strong degree of similarity within the population form and the Fréchet number is small, this would indicate that the population is heterogeneous. This means that there is a high probability of accurate knowledge sharing within the population. Conventional SHM techniques may be applicable but if not, domain adaptation methods will provide strong results. Conversely, if the Fréchet number is large this indicates that the population is heterogeneous and that conventional SHM techniques will fail. Domain adaptation techniques must be implemented, and the degree of accuracy will depend on the size of the dataset, the model implemented and the degree of heterogeneity.

4.4 Population Based Structural Health Monitoring Documentation Based Case Study

Building on the established methodology, this section analyzes both data and documentation to determine structural similarities. To highlight the process comparisons are made to other WTG not associated with this wind farm. This section covers the operational, material, topological and geometrical considerations that are addressed in the resources available.

Table 4.1: Wind turbine geometrical foundation comparison.

Foundation Type					
Jacket Structure			Pile Cap Foundation		
Component	Geometry	Shape	Component	Geometry	Shape
Transition piece	Complex	Transition Piece	Pile Cap	Beam	Cylindrical
Grouted Connection	Beam	Cylindrical	Reinforcement	Plate	Cylindrical
Piles	Beam	Cylindrical	Piles	Beam	Cylindrical
Jacket Leg	Beam	Cylindrical	Drilling Template	Plate	Cylindrical
Jacket Brace	Beam	Cylindrical	Tower adaptor	Plate	Cylindrical
Jacket Node	Complex	Node	Tension Anchors	Beam	Cylindrical
J-tube	Beam	Cylindrical	Bolts	Beam	Cylindrical
Boat Landing	Complex	Boat Landing			
Tower Platform	Plate	Rectangular			

Table 4.2: Wind turbine foundation material comparison.

Foundation Type					
Jacket Structure			Pile Cap Foundation		
Component	Material	Material Class	Component	Material	Material Class
Transition piece	Steel	Metal	Pile Cap	Concrete	Ceramic
Grouted Connection	Concrete	Ceramic	Pile Cap Reinforcement	Steele	Metal
Piles	Steel	Metal	Piles	Concrete	Ceramic
Jacket Leg	Steel	Metal	Drilling Template	Steel	Metal
Jacket Brace	Steel	Metal	Tower adaptor	Steel	Metal
Jacket Node	Steel	Metal	Tension Anchors	Steel	Metal
J-tube	Steel	Metal	Bolts	Steel	Metal
Boat Landing	Steel	Metal	Pile Reinforcement	Steele	Metal
Tower Platform	Steel	Metal			

4.4.1 Geometry

Obtaining the geometrical information may not be possible for all structures within a population. It is necessary to define a hierarchy of properties that defines the properties of the irreducible element for increased levels of accuracy. Abstracting the significant properties and dimensions of an irreducible element representation of the structure attempts to simplify the geometry. This can be performed for various structural components that have a well-defined structural dynamic behaviour such as beams, and plates. However, some cases are too complex for this application. Table 4.1 is a reduced hierarchy of geometry class, followed by the shape. The irreducible element further defines the major dimensions of the element which echoes the fact that certain transfer learning approaches are valid at a less explicit level than others when transfer learning is applied. Similarities become better defined at a lower level of hierarchy and improve the chance of success.

Geometry is important, for example, when determining the overall dynamic behaviour of structures when both systems have the same materials, geometry and topology. One can assume that the dynamic behaviour should be the same (or remarkably similar) between two systems. When there are differences in dynamic behaviour it may be because of damage.

The complexity of geometry and the degree of uncertainty within the match will influence the certainty of inference to the target domain. Within transfer learning, a regression model which is trained on the source domain is used in another target domain (the target domain is the system to which one wishes to transfer the knowledge). Transfer learning only works if there are similarities between the source to the target. Hence, transfer learning approaches used in SHM will require similar structures, otherwise a negative transfer and differences in source to target will cause misleading results [76]. For complex geometry, a negative transfer may be unavoidable.

4.4.2 Material

The geometry and topology of the structures are strongly linked with the transfer of damage detection and damage labelling. Material properties conversely determine the assessment and classification labels. Again, there is a hierarchy of material properties that can be used to describe irreducible elements as, like all other aspects of the detailed design, material properties may not be available. Material class is considered such as ceramics and metals. The next set of descriptions would be the specific class such as steel or brass, both belonging to the metal class. The finest levels of description would be Young's Modulus, density, or material grade to name a few. To determine if two structures are similar all the details must be known throughout the hierarchy, otherwise, there will be a degree of uncertainty.

The material properties determine whether it is possible to make inferences of damage assessment and classification with labels between two structures. Two materials of the same material class will experience similar failure modes and may exhibit the same material responses giving more confidence in the classification of damage. The greater the similarity between the materials the more likely the assessment of damage will be the same between the two structures. For example, materials from the same metal class, such as aluminium and steel, will suffer from the same damage class, corrosion. However, there is no guarantee that the change in material properties will suffer the same extent of damage assessment because of corrosion. If the material is identical then the damage assessment and classification are trivial. Table 4.2 highlights the differences in the material class of similar components. An offshore transition piece on a jacket structure has the same purpose as a pile cap of an onshore foundation, but they have dissimilar materials and may exhibit different responses, failure modes and damage assessments.

4.4.3 Topological

Scour, corrosion, and marine growth are environmental effects that have been considered; these are global effects which are currently not implemented in the monitoring campaign. Some methods can be used to monitor these influences but come with a significant financial burden and are not included in the SCADA data. To shed some light on the potential magnitude of heterogeneity throughout the population on individual WTs this section considers the four attributes of homogeneity.

The dynamic behaviour of an operational wind turbine's structure is largely based on the topology. The topology is determined by connections of the physical elements that construct the entire structure (individual) with its environment. The topology has a strong link to the damage location and without topology, structures cannot have a fully consistent location labelling. If the topology of two structures is the same, then the damage location labels can be applied consistently across both. Incorporating the material within the description means that one can determine whether inferences on the material damage are possible.

The complex class can provide some flexibility when the topology of the structures is similar. If the application is comparing the dynamic behaviour of two different wind turbines, it may be more useful to consider the entire gearbox as one complex element. Depending on what behaviour we want to capture and extend and the type of damage we want to detect, we need to select topology so that the differences in configuration can be considered local. However, if the local topology is relevant to the application of locating damage, then it may not be beneficial to generalise components and so breaking the components into parts may be advised.

Corrosion

The design of an offshore steel jacket assumes that corrosion is an inevitable phenomenon, both internal and external. The effects of corrosion reduce the cross-sectional area of the elements with other local effects such as pitting. This is a global phenomenon and, in terms of the population, should be a constant effect across the individual constituents. Earlier studies in ROMEO WP.4.4 have highlighted the effects of corrosion on the dynamic properties of the jacket structure, as shown in Figure 4.5, where the torsional mode shape received a 10% reduction in frequency.

Marine Growth

Paradoxical to corrosion, marine growth adds additional mass to the structure and can introduce an additional surface roughness to the exterior members. Altering the surface roughness of the members in the WP 4.4 study concluded that this has a minimal effect

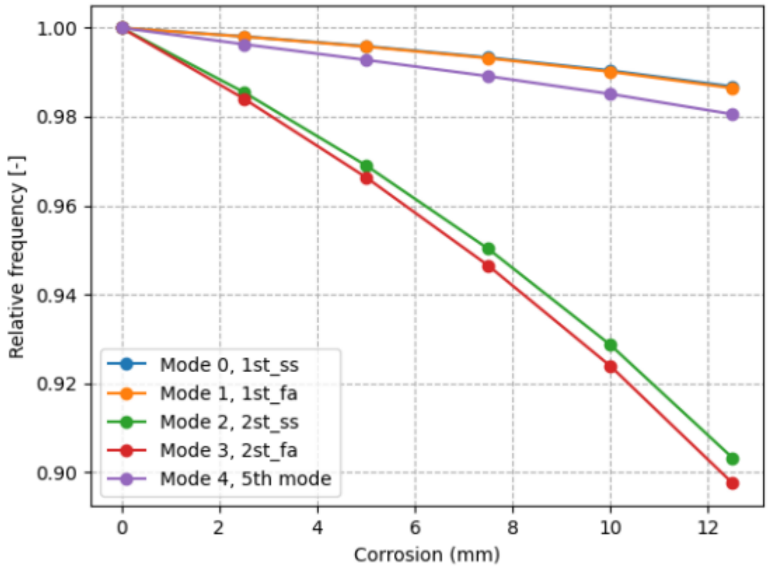


Figure 4.5: Dynamic properties variation for proportional changes in the global corrosion profile.

, [29]

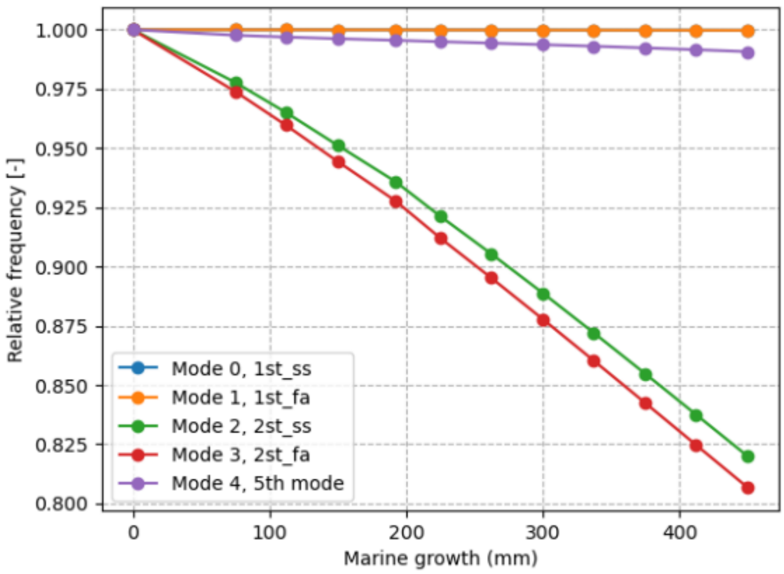


Figure 4.6: Dynamic properties variation for proportional changes in the global marine growth profile, [29].

on the structural dynamics of the foundation, in this case, the maximum effect on the torsional mode is 0.1%, see Figure 4.6.

Scour

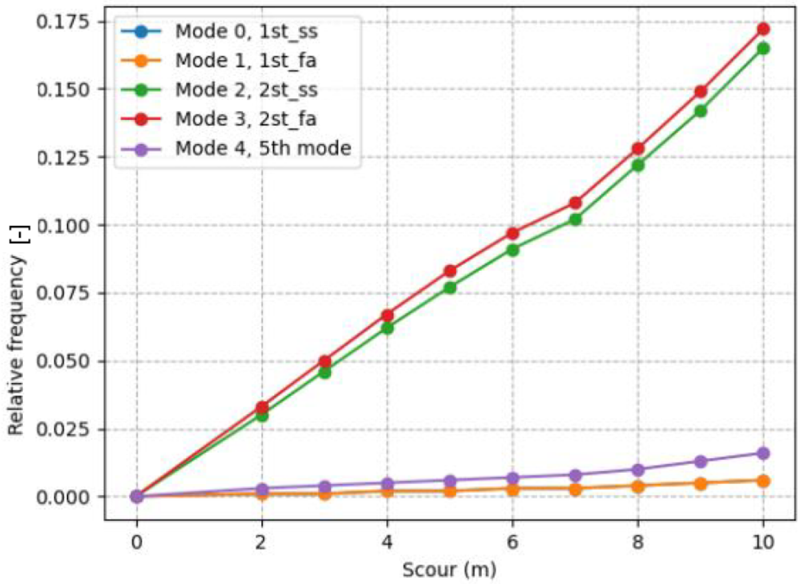


Figure 4.7: Dynamic properties variation for proportional changes in the global scour profile, [29].

Scour is a phenomenon that removes sediment from the foundation components at the seabed and gets carried away by the momentum of the seawater. This causes a reduction in the soil around the area of the foundation reducing the length of the piled foundation. The reduction in soil results in reduced mass and stiffness. Figure 4.7 is taken from the study on the effects of scouring where up to 3.2m of soil is removed and the results indicate that this has a minimal effect on the dynamics of the structure. In this scenario, the main effect is the second moment with a 3% reduction in frequency. An open-access study on the effects of scour is detailed in this study [144], where low levels of scour go unnoticed.

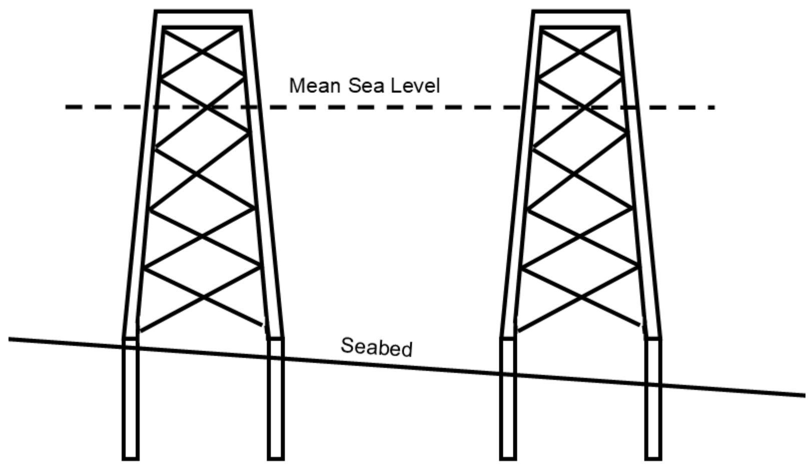


Figure 4.8: 4-leg jacket foundation for different water depths, [29].

Water Depth

Water depth and geological conditions play a significant role in the design of offshore structures. The wind farm has implemented a jacket structure and depending on the water depth there will be varying degrees of foundation pile exposed. However, the length of the pile is constant throughout the entire wind farm (as demonstrated by Figure 4.8). The complex dynamics of a jacket structure are hard to estimate exactly without running an FE analysis but if one were to reduce the foundation pile to a beam then a simplified judgment could be made. One would expect a longer beam to experience increased levels of fatigue damage over the lifespan.

4.5 Population Based Structural Health Monitoring Data Driven Considerations

4.5.1 Operation

When investigating the operational and environmental effects of a small population of wind turbines from the wind farm one would expect that since they all have geometrical homogeneity, material homogeneity and topological strong homogeneity we would get remarkably similar distributions for these properties. And we do, as highlighted in

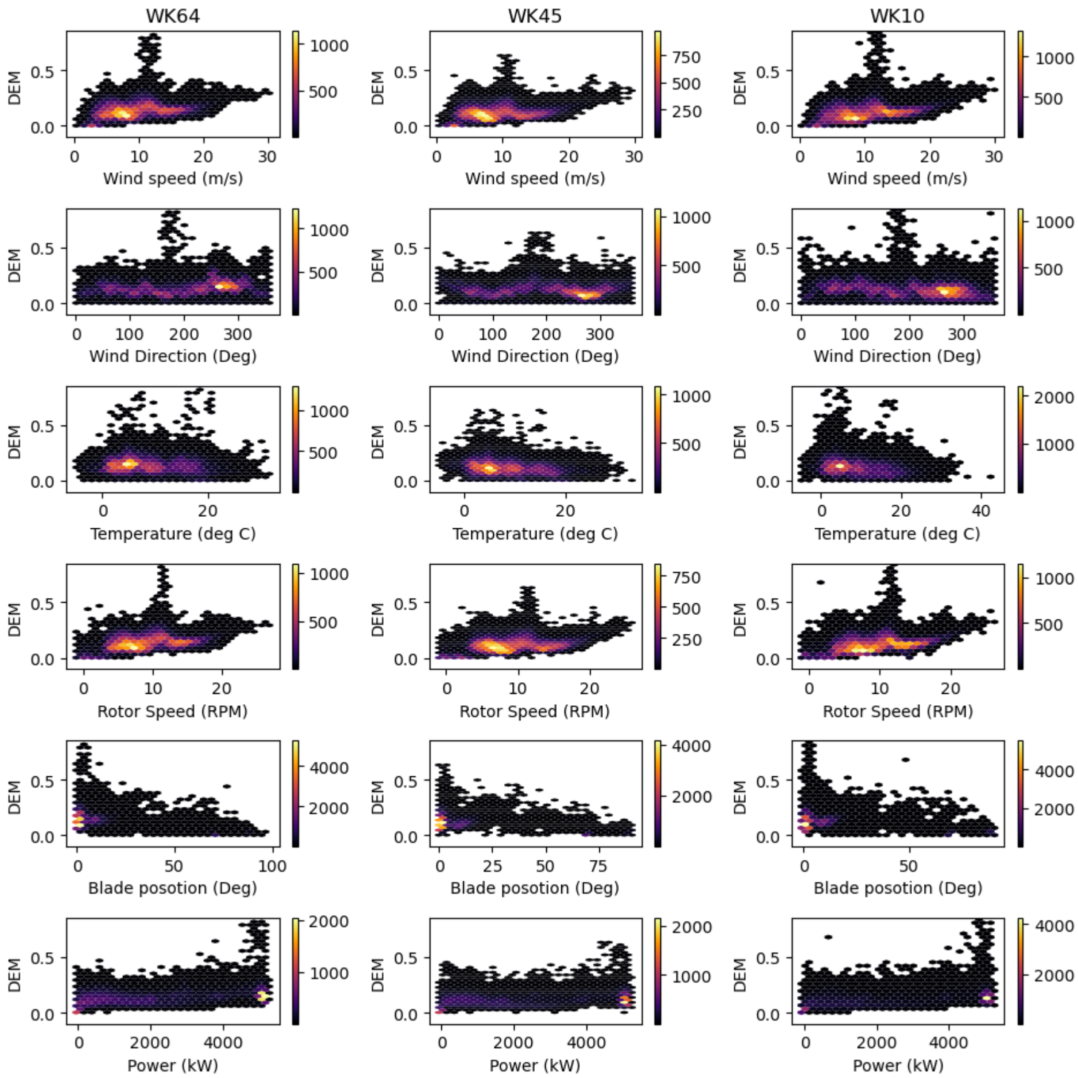
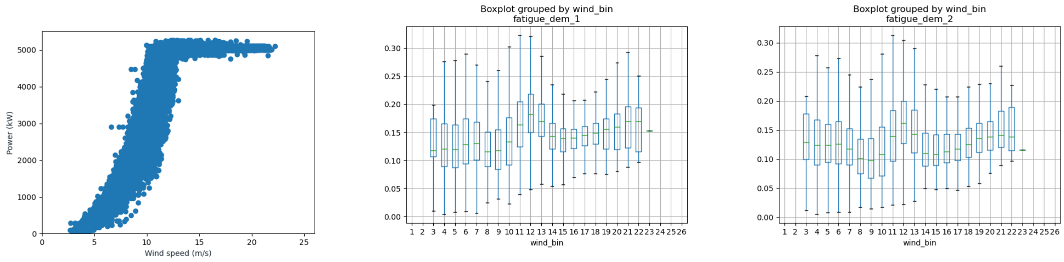


Figure 4.9: This series of figures highlights the operational and environmental effects on the population of the wind farm. This includes the power, blade position, wind speed, rotor speed, and temperature against the DEM

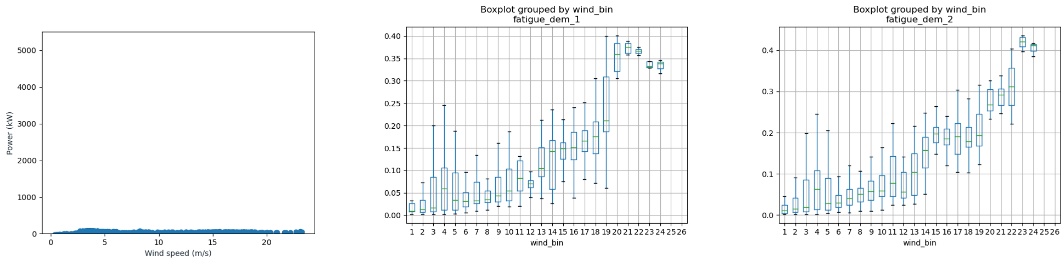


(a) Normal operation power curve

(b) DEM 1 for the normal operation

(c) DEM 2 for the normal operation

Figure 4.10: The operational influence on the DEM on the two orthogonal directions



(a) Stalled operation power curve

(b) DEM 1 for the stalled operation

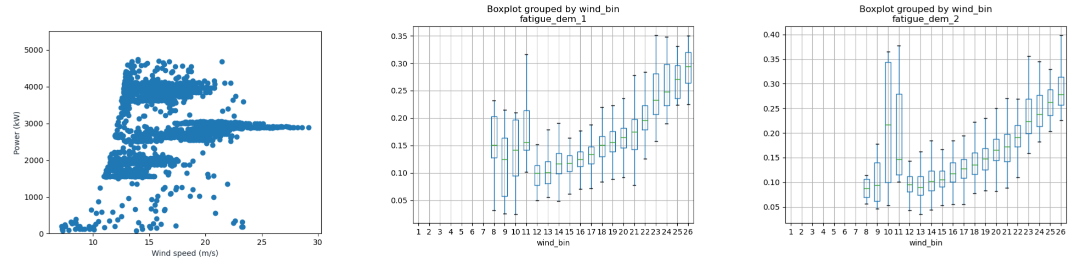
(c) DEM 2 for the stalled operation

Figure 4.11: The operational influence on the DEM on the two orthogonal directions

Figure 5.6, but we have variations in the magnitudes of these effects on the fatigue damage results for the individual wind turbines. It is difficult to determine if this is related to the sea depth as WT3 is installed in the deepest location out of the three wind turbines, but the distribution of the data is akin to that of the other three.

An interesting observation from the damage equivalent moments (DEMs) on all three wind turbines based on the direction of the wind is that the damage rate is increased. WT1 and WT3 have two peaks in the DEM vs wind direction graph and WT2 experiences one. This phenomenon is apparent in the wind speed and temperature graphs too. Figure 4.13 displays the wind farm location and the general topology of the water depth with WT2 and WT1 being in a closer cluster as opposed to WT3. The deviation in the results is one of the main reasons that these features would be mandatory in model development.

The operational condition of the wind turbine has a significant effect on the dy-



(a) Curtailed modes power curve

(b) DEM 1 for the curtailed operation

(c) DEM 2 for the curtailed operation

Figure 4.12: The operational influence on the DEM on the two orthogonal directions

namics of the structure. In the entire population of the wind farm, individual wind turbines can be in a different operational state, leading to diverse structural dynamics. Curtailments are implemented for the planned reduction of power output. The wind farm exhibits multiple curtailments of various reduced power programs, maximum coefficient of power and lastly parked, where there is zero power extraction as represented in figure 4.10, 4.11, and 4.12.

The main challenge when dealing with operational conditions in PBSHM is that different operators may have different philosophies on how to best gain the most out of the entire wind farm as implementing curtailments leads to increased heterogeneity. Conversely, if the same philosophy is implemented for the entire population, then separate ML models can be used to tackle the consistent modes.

Observing the different operational modes, there is a different distribution of the DEM for each wind bucket. It must be noted that the splitting of the operational modes is not based on SCADA-referenced signatures but rather on manual interoperation of the data. A large scatter is observed in each of the individual wind buckets. In [88] this has been attributed to the environmental effects, such as turbulence, and must be included in the general model. Secondly, creating a general model that encompasses all three of these modes will increase the complexity of the predictive function.

4.6 Heterogeneity analysis of an offshore wind farm

When assessing the homogeneity of a population it might be useful to consider some of the environmental and operational effects. The previous chapter briefly mentioned the main components of determining the degree of heterogeneity. This chapter will begin the discussion on the main components that will enforce this judgment from observable data in the monitoring campaign. Another aspect that is equally important and must be highlighted is the limitations on defining the degree of homogeneity in the context of an operational wind farm. Some of the material, geometrical and topological differences within classes are difficult to capture with current measurement campaigns such as corrosion, scour and marine growth. This section will discuss the measurable effects and how this affects the heterogeneity of the population.

The population is based on the wind farm located off the island of Rugen in the Baltic Sea with a total installed capacity of 836 MW, where individual wind turbines were manufactured with a capacity of 5 MW each. The preferred choice of foundation used across the entire population is the jacket structure. The main measurable differences across the entire wind farm are the location and the water depth. Other differences such as corrosion, scour, and marine growth are not monitored under the current campaign. The following subsections will discuss current hypotheses on the dynamics of the structure based on these differences.

4.6.1 Population Heterogeneity

Table 4.3: Fréchet Distance for all three wind turbines

Fréchet Distance			
WT1	0.00		
WT2	13.48	0.00	
WT3	14.48	14.92	0.00
	WT1	WT2	WT3

This section discusses the numerical method of determining the similarity of a population form. For a given domain in the metric space, the Fréchet distance is a measure of similarity between the two domains. This considers the location and the ordering of the points along the points of both domains. For homogeneous populations the value of the Fréchet distance will be 0, strong homogeneous values will be small, and heterogeneous will be large. The value is dependent on the length of the instances in the domain and the magnitude of the values. The Fréchet distance is calculated by:

$$\Delta = \|\nu_s - \mu_t| \,_2^2 + T_r(|\Sigma_s + \Sigma_T - 2(\Sigma_s \cdot \Sigma_T)^{1/2}) \tag{4.1}$$

Where Σ_s, Σ_T are the mean of two matrices X_s, X_T and Σ_s, Σ_T is the covariance of the matrix X_s, X_T. After applying this to the DEM for each of the wind turbines, the results in Table 4.3 indicate that the population form is homogeneous with values close to 0. This is a numerical representation, and it is limited in the scope of interpretation on the homogeneity. This is a technique that mathematically reinforces the decision after observation of the data. However, in the following section, the location is the main source of heterogeneity for this population form and breaking this down into more detail will provide greater insight into the small deviations in heterogeneity as observed in Table 4.3.

Table 4.4: Statistical values of both DEM for all three wind turbines.

DEM 1				DEM 2		
	WT1	WT2	WT3	WT1	WT2	WT3
Mean	0.1193	0.1193	0.1193	0.1049	0.1049	0.1049
Std	0.0629	0.0629	0.0629	0.0623	0.0623	0.0623
Min	0.0015	0.0015	0.0015	0.0017	0.0015	0.0017
Max	0.9024	0.9024	0.9024	0.6494	0.6494	0.6494

To contextualize this, statistical measures have been calculated based on the normalized DEM which are represented in Table 4.4. This includes the mean, standard deviation (Std), minimum and maximum of the DEM in both orthogonal directions. There is an almost identical nature to the three WTGs; WT2 does have higher and lower maximums and minimums, but both the mean and standard deviation are identical to four significant figures.

4.6.2 Location based Heterogeneity

Wind turbines extract energy to produce electricity, therefore, the wind leaving the wind turbine must have less energy content than the wind upstream of the turbine. Consequently, the wind downstream will be more turbulent and have a reduced speed. As the wind downstream begins to flow it will eventually return to a free stream condition. However, wakes, if intersected by a wind turbine downstream, are said to be shadowed by the wind turbine producing the wake. There are two main effects: one is a reduction in wind speed and the second is increased turbulence and potentially increased mechanical loading.

Studies on the effects of wakes on wind farm operations [73] indicate that array efficiency is dependent on the spacing between the turbines and the nature of the wind regime. When wind turbines extract energy kinetic from the wind increases the disturbance in the air flow. This disturbance reduces time and is one of the contributing factors to wind turbine layout optimisation.

The varying degrees of efficiency on the individual wind turbines could affect the homogeneity throughout the wind farm. If there is an increased amount of mechanical loading, then over the lifespan one could assume that there may be a difference in accumulated fatigue damage over the lifespan.

Table 4.5: Fréchet Distance for all three wind turbines in 4 different wind directions.

Wind Direction	Turbulence Intensity		
	WT1	WT2	WT3
N	0.165	0.171	0.148
E	0.101	0.138	0.156
S	0.147	0.133	0.160
W	0.165	0.171	0.148
Global	0.144	0.140	0.127

To assess the similarity of the structure's heterogeneity as a function of wind direction Table 4.5 highlights the fatigue damage equivalent loads for four different wind directions, which can be used as a direct comparison of the location-based heterogeneity. As expected, there are various degrees of similarity from the different wind directions. The observation is that with increased heterogeneity one can attribute this

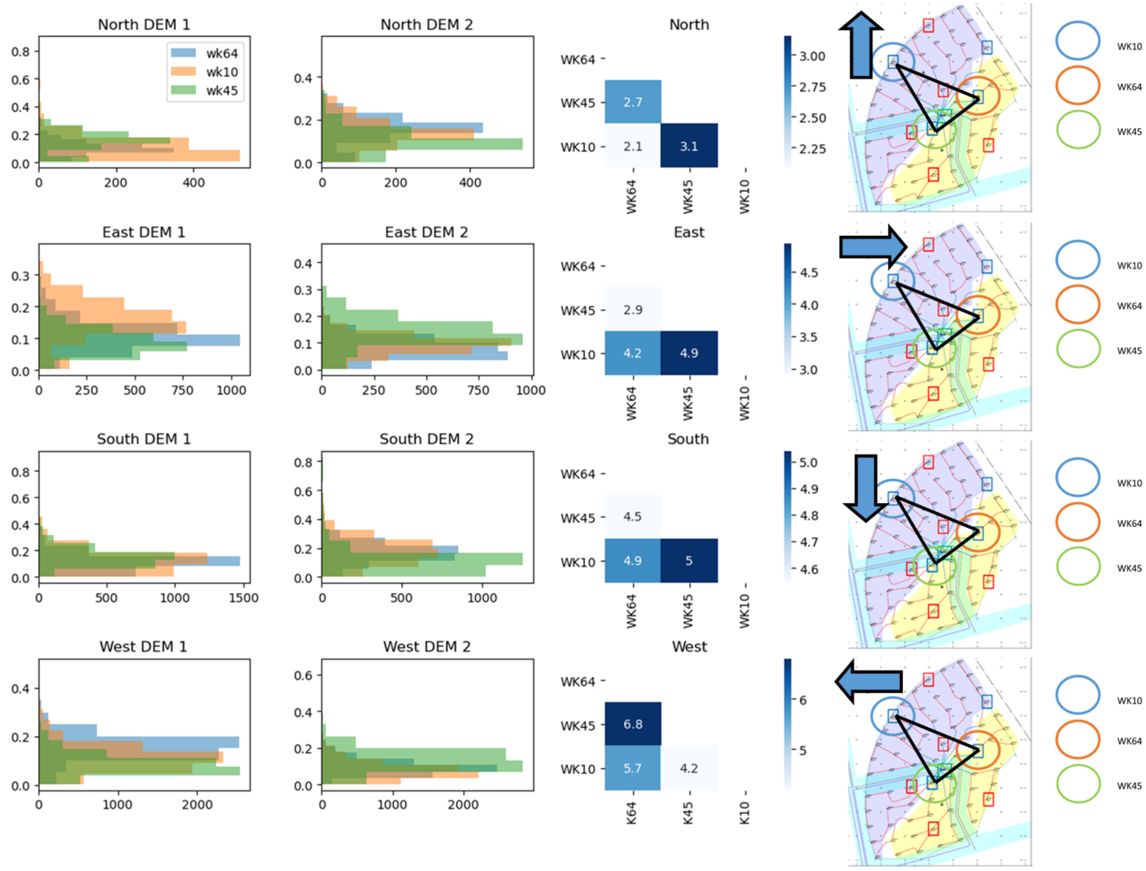


Figure 4.13: A Fréchet distance heat plot of the damage equivalent moments from three of the wind turbines from four different wind directions, N- north, E- east, S - south, and W – west, over 12 months.

to the turbulence intensity and the wind direction with westerly winds having the highest turbulence intensity for all three wind turbines and an average highest Fréchet correlation. Consequently, any fatigue model should include the turbulence intensity in the calculations.

4.7 Conclusion

Population-based structural health monitoring is dependent on the idea that knowledge between structures can be transferred. Models and data can be shared directly for structures within the homogeneous population. However, in structures of a heterogeneous population, it is imperative to determine the structural similarities to determine whether it is possible to transfer such knowledge.

To compare the degree of similarity within a population, the elements can first be converted into an irreducible element model, where the abstractions from the structures determine the degree of transferability. With matching properties, this strengthens the consistency of the labels required for transfer learning. In the situation where you have matching topology, operational conditions, and geometry this gives matching location labels. Matching the material properties gives consistent damage classifications or assessment labels. Hence, documenting the similarity of all four aspects is important in determining whether the transfer of knowledge is possible.

In the context of the wind turbine foundation DEM by introducing the Fréchet number one can mathematically quantify the similarity of the population form, but this must be accompanied by an observation to determine the degree of homogeneity. In this study the different operational cases of the wind turbine experience different rates of damage at different wind speeds. Moreover, the environmental effects from different wind directions, combined with the turbulence intensity influence, have a significant impact on the degree of similarity of the DEM and need to be taken into consideration.

Out with the monitoring campaign, three challenges that affect the individual structural integrity of offshore wind turbines have been modelled using finite element analysis and it is concluded that these have a minimal effect on the structural integrity. Hence, one could only assume that these will have a minimal effect up to the design guidelines. Exceedance over the lifespan of the wind turbine may lead to larger deviations and model updating may be necessary in determining if these events, at the later parts of operation, need to be accounted for.

When making a comparison of the environmental and operational effects on the population in the wind farm this highlighted that there is a strong homogeneous correlation. The Fréchet number is low for the entire dataset and the plots reinforce the judgment for the damage equivalent moments on the foundation of the structure.

Chapter 5

Applied Digital Enabled Asset Management

Deep Neural Network Hard Parameter Multi-Task Learning for Condition Monitoring of an Offshore Wind Turbine, Innes Murdo Black, Deborah Cavesco, Athanasios Kolios, 2022, Journal of Physics

5.1 Hard-parameter Transfer

5.1.1 Introduction

Consistency is a great way of de-risking offshore wind, but unfortunately, some failures are inherent. As the popularity of wind turbines increases globally so do the varying effects which make the failures of the structure more difficult to predict. Furthermore, new technologies and optimised designs can suddenly fail due to quality or stress-related failures, respectively. This type of failure is termed infant mortality failure, causing a potentially significant loss in revenue, especially if employed in the offshore wind sector.

Anomalies detection and failure modes diagnosis can be utilised to identify the structural health of offshore wind turbine components, by using intelligent computing such as artificial neural networks. Richmond et. al [155] performs a stochastic assessment of the aerodynamics of an offshore wind turbine using an artificial neural network among other machine methods to determine wind speeds and directions. By extending this work, one could make assumptions on the fatigue life of an offshore wind turbine. Bao et al. [19] utilise a one-dimensional convolutional neural network to determine the occurrence of damage to the support structure of an offshore wind turbine. In this example, the examination looks at localised damage to a jacket support structure under regular waves with an incredibly high accuracy of 98.4%.

Artificial neural networks have been used to examine the 'life percentage' of an offshore wind turbine based on the failure time distributions. Yang et al. [218] apply a two-level failure probability procedure for the gearbox, the rotor, the pitch mechanism and the generator. This simplified method has shown that condition-based maintenance schemes can reduce the cost of preservation compared to classic time-based maintenance, where regular intervals are used to assess the asset. A breakdown of convectional models suited to wind turbine maintenance is discussed in [22]. The components of the drive-train have the potential highest impact on the maintenance cost of the next generation of offshore wind farms [32]. These are among the most expensive components for an offshore wind turbine and are continuously undergoing remodelling where innovations are to accommodate bigger power outputs [33] with larger loads. Condition Monitoring (CM) signals, in combination with high-frequency Supervisory Control and Data Acquisition (SCADA) data have been extensively used to train machine learning (ML) models to predict failures in the drive-train components [181] [186]. In [181], Stetco et al. document the state-of-the-art ML methods and processes for the wind turbine condition monitoring. Tautz-Weinert and Watson [186] examine and discuss the effectiveness of the numerous ML methods based on the type and amount of data available. However, the information available from the literature mainly focuses on training the algorithms based on the availability of relatively big sets of data – generally for at least more than three years. The authors, thus, identified a knowledge gap in the field of offshore wind applications regarding the investigation of a detection algorithm suitable for small data-sets.

Conventional Machine Learning Versus Multi-Task Learning

Conventional ML typically involves optimising for a particular task $T = (y, f(x))$, where y is the output feature domain and $f(x)$ is the predicative function made up of X feature data. The model is trained for a single task, this generally may achieve acceptable performance for a single domain $D = (x, p(x))$ of marginal probability distribution $p(x)$, but by focusing on one signal task we ignore information that may help us do better on other metrics relating to that task. By sharing representations of a global task trained on the source domain D_s and target domain D_t, with a similar probability distribution, we may be able to better represent our general task. This is Multi-Task Learning.

A benefit of multi-task learning is that knowledge can be transferred. Inductive transfer learning is a subcategory of transfer learning. Where an accurate model is usually trained on the source domain D_s to determine the hypothesis space, this article implements the hard-parameter transfer technique of inductive transfer. The hypothesis space generated by hard parameter transfer can help improve the target task results. Particularly, if there are small amounts of data or class labels for the target task. Performing inductive transfer learning using source domain data to train the general model and applying the target task data for fine-tuning can lead to a more accurate model [134].

The multi-task learning (MTL) theory has been employed across many fields of application requiring a supervised prediction of one or more classes; two-stage facial recognition [213], quality assessment of fetal head ultrasonic assessment [114], or for bandwidth allocation for multiple mobile users [87]. The MTL models come in a plethora of forms: joint learning, learning to learn, and learning with an auxiliary task, are among some of the names that have been allocated to its predictive assignment. To generalise the need for its application, it can be stated that a MTL approach is worth being investigated as soon as the problem requires optimising for more than one task.

Problem statement and scope of the analysis

The availability of data from offshore wind turbines is limited, with CMS data on the generators is less accessible. When wind turbines are installed this restricts the amount of data available. This is a sensitive period where novel failure events are more likely to happen in the early stages of the structure (infant mortality failures). Additionally, low-cost monitoring campaigns might be preferred to reduce the cost related to the hardware and the storage of data, limiting the number of assets with SCADA or CMS systems. To address this issue, this work aims to apply a data-driven multi-task learning approach to monitoring the health status of an offshore wind turbine gearbox.

The purpose is to demonstrate how a hard parameter transfer model can achieve greater results than a conventional machine learning model when applied to a limited amount of training data. Therefore, the scope of this paper is to make a step towards understanding of the setup of a suitable monitoring algorithm based on CMS with a small set of data of an offshore wind turbine gearbox.

The remainder of this article provides a discussion of the literature and the theoretical basis the of MTL method in Section 5.1.1. In Section 5.1.2, the methodology of this study is introduced, with details on the methods applied for the data pre-processing, the model training, and the evaluation metrics used to determine the effectiveness of the detection. Section 5.1.3 includes the results highlighting the main findings and showing the clear context for which MTL is effective, to finally closes with a discussion and conclusion.

5.1.2 Methodology

This section introduces the methodology for training and comparing models predicting, in binary form, the status of an offshore wind turbine gearbox. The data collected, pre-processing, and their division into data sets for the training of a conventional and a MTL algorithm are introduced. The overall workflow for the training of the models is described in detail. Finally, the metrics used for the evaluation of the models are defined.

Data Collection

The analysis presented in this paper is built on time-series data from one MW offshore wind turbine, in normal operation. The signals from the SCADA and condition monitoring (CM) systems consist of eight monitoring channels, recorded with a ten-minute resolution. These channels include meteorological information, the operational data of the wind turbine, and the vibration data from the gearbox, with the associated flag warning raised in case of anomalies. This latter provides the (binary) label targeted in the training of the classification models.

There are two data sets, the source domain data contains 31804 time-steps (220 operational days) collected from a turbine. From another turbine of the same population, considered homogeneous. The target domain data has a reduced length of 8141 time-steps These two data sets are, thus, representative of an existing wind turbine and a newly installed turbine. Both wind turbines have the same form and hence similar distributions in each of the identical features. The source domain data has alarm signatures for a total of 19% of the data set and the target domain data has a similar portion with 18%.

Data Pre-processing

The data sets used for the training and testing of a machine learning algorithm generally require some pre-processing to ensure satisfactory performances of the prediction model. The application in this paper consists of a two-step procedure. First, a data cleaning process is performed to remove outliers and missing values. For the treatment of missing values, time instances of the database are removed from the analysis if over 50% of the data is missing. The outlier removal is processed by removing vector instances where the values should be scalar. For the remainder of the data, the K-nearest neighbours (K-NN) imputation method is applied [194].

The last step is to split the data into training, and testing sets. In the specific of this experiment, 80% of the data is employed for the training of the models, while the remaining 20% is used for the final testing.

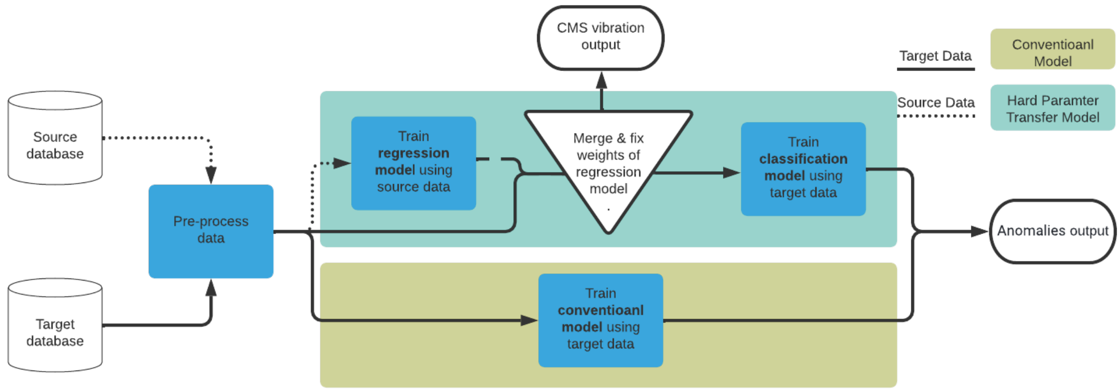


Figure 5.1: Flowchart detailing the data flow and construction of the conventional model on the left and the multi-task learning model on the right.

Models

The MTL model is built up in two training stages, with information being leveraged from two data sets. The knowledge acquired from the source domain data D_s with a greater number of samples is used to train, a feature extractor (regression model). Taking the SCADA data, and making predictions on the vibration data from the gearbox. A classifier is connected to this first model and uses a hard parameter transfer to merge the last neurons from the artificial neural network to the first layer of the convolutional neural network (classification model). The weights of the neurons of the regression model are fixed, and the remaining weights of the model are trained using the small target domain D_t data set. The model is developed using the Keras API [2].

The conventional model takes the same architectural form as the regression and the convolutional neural network classifier together, but it is just trained by only using the target domain data D_t, a smaller data set. Therefore, the model receives the SCADA data as the input, and it outputs the anomalous behaviour of the gearbox. Both models have the same desired task but one model uses hard-parameter transfer to transfer knowledge and increase the amount of knowledge representing the anomalous behaviour since it has two outputs the CMS vibration predictions and the binary error message.

Training of the models

Three distinct models are trained for this paper. One is the feature extraction model, another is the classification model used for the MTL procedure, and the last is a conventional classification model.

The transfer learning model is built up of three blocks, as can be observed in Figure 5.1. The feature extractor; takes in input from the meteorological data, the wind turbine operational data, and the gear oil temperature, and then outputs the gearbox vibration features. This model is built up with a deep neural network (DNN) architecture. In particular, it consists of 14 sequential layers all implementing a rectified linear unit (RELu) activation function. Utilising a uniform variance scaling [82] allows the neural network to train extremely deep rectified models directly from scratch. The optimiser for the regression feature extraction is called 'Adam', which is derived from the adaptive moment estimation [104]. This is effective for noisy, nonlinear data.

Hard parameter MTL is carried out using the classification model highlighted in Figure 5.1. This convolutional neural network (CNN) architecture is the same as the classifier used for the conventional model. Both classifiers feature one convolutional layer of width 64 and similarly utilise the RELu activation function. This is followed by a drop out of 0.5 to further four layers of a one-dimensional convolution, which implements the sigmoid function - commonly used in classification models. The kernel initialisation of the weights for the CNN uses uniform variance scaling. A standard gradient descent with Nesterov Momentum is employed to improve the accuracy while dealing with noisy data from the vibration signals, with a learning rate of 0.1 and a momentum of 0.9. Lastly, the cross-binary entropy loss function is implemented to distinguish the gearbox status class.

For the consistency of the comparison, the conventional model takes the overall same architecture of both the feature extraction and the classification model combined. The main differences between the conventional model to the hard parameter transfer model are:

- The binary cross entropy optimiser is applied to the whole model.

- The entire model is trained in one process and one data set.
- The model only has one output stream of information representing the gearbox status.

Evaluation metrics

The regression results are evaluated using the mean absolute percentage error $MAPE$, which represents the average absolute percentage error for each period minus the actual values divided by actual values.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{\hat{y}_i - y_i}{y_i}| \tag{5.1}$$

where $\hat{y}_i$ is the forecasted value, y_i is the actual value and n is the number of samples. The metrics of the classification models are calculated using $F1_{score}$ and the $Accuracy$. The $F1_{score}$ metric conveys the balance between the precision, the true positive predictive value, the recall, and the true positive rate, by calculating their harmonic mean. The accuracy represents the percentage the model correctly calculates.

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \tag{5.2}$$

$$F1_{score} = \frac{t_p}{t_p + 0.5(f_p + f_n)} \tag{5.3}$$

The true positive t_p, the false positive f_p and the false negative f_n are used to calculate the F1-score and are going to be explicitly reported to judge the quality of the classification. W Yang et. al. [218] conducted a study on wind turbine monitoring and indicated that predicting more than 60% of wind turbine faults will reduce the cost of operations and maintenance. This has been implemented as a threshold for our models.

5.1.3 Results

The correlation plot in Figure 5.2 highlights Pearson's correlation of the variables to one another. The set of features proves the potential advantages of hard perimeter transfer learning. In machine learning, the higher the correlation to the data the increases the chances of the the predicative function describing the task. Linear correlations, while useful, are limited by their assumptions of linearity, sensitivity to outliers, inability to infer causality, influence of hidden variables, danger of extrapolation, and invalidity for agreement assessment, [105]. It can be observed that from the "Power Bin" to "Wind Speed", the features have no relation to the gearbox "Error". On the other hand, the correlation between the vibration data and the gearbox status is higher. This increased correlation helps improve the predictive function.

Regression model

This model takes the SCADA data from the "larger" source domain data set and makes predictions on the CMS vibration data after pre-processing. The training stage is carried out over 1000 Epochs having a total of 388,803 trainable parameters. To validate the accuracy of the model new, unseen data from the wind turbine is fed into the model producing a: MAPE = 27.00%, an accuracy = 99.92% and R2 = 68.61%. This model is the highest-performing model in the process. The predictions from the model are highlighted in Figure 5.3. With an R2 score of 68.61%, this is a more than acceptable indicator of anomalous behaviour. The plot highlights the feature extractor's ability to do so.

Implementing a MTL model, this information would normally be fed into the model or in the worst case be ignored. This procedure utilises the information to make strong estimations on the erroneous behaviour and it could be used in other maintenance methods in a multi-agent manner for predictive trending methods. Where one could monitor the trends of the vibrations more closely to determine long-term effects.

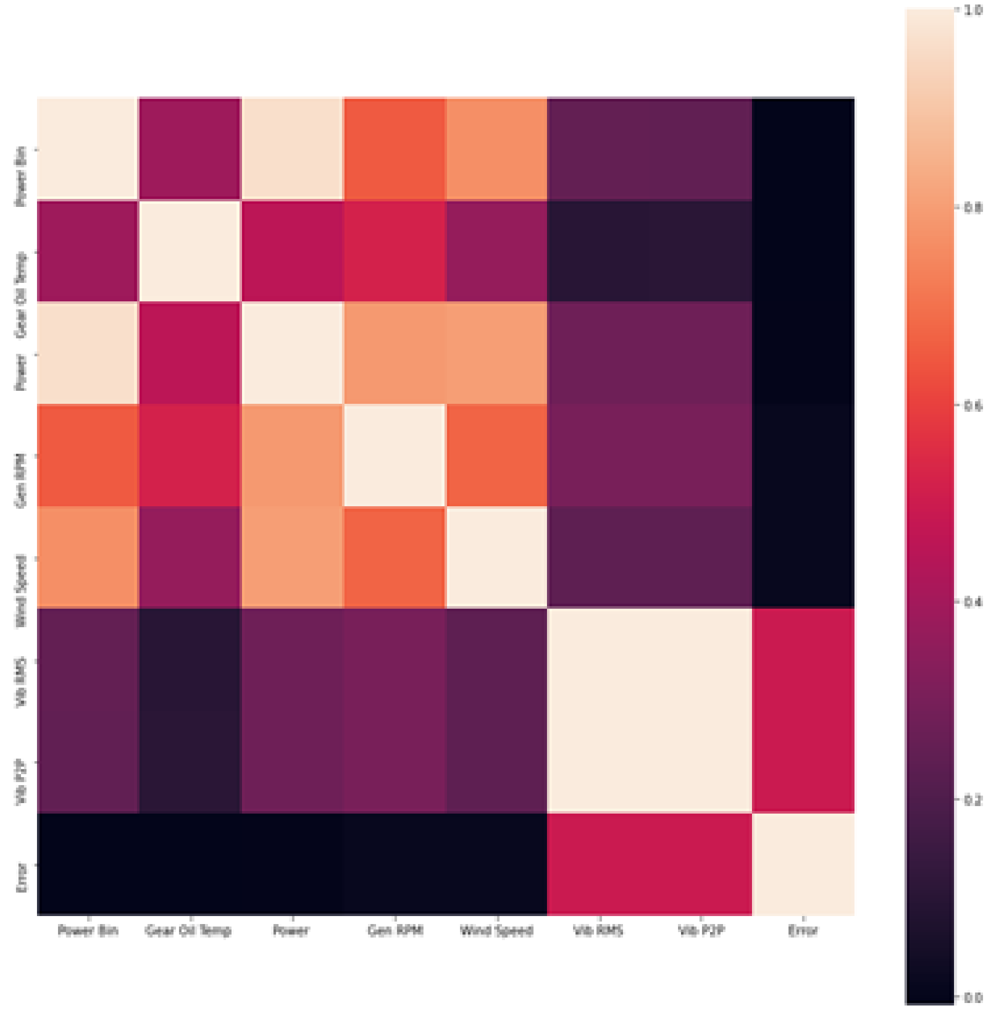
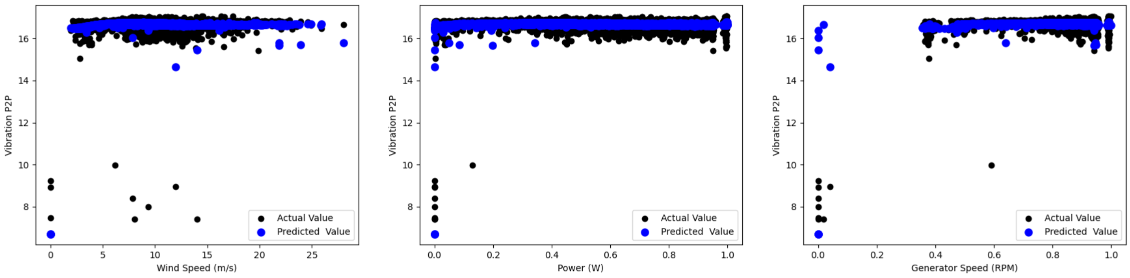


Figure 5.2: Linear correlation for all the features, the first five features are the inputs for both the regression model and the traditional model. The vibration signals are the outputs of the regression aspect hard parameter transfer model, and the Error is the output of the full hard parameter transfer model and the conventional model.



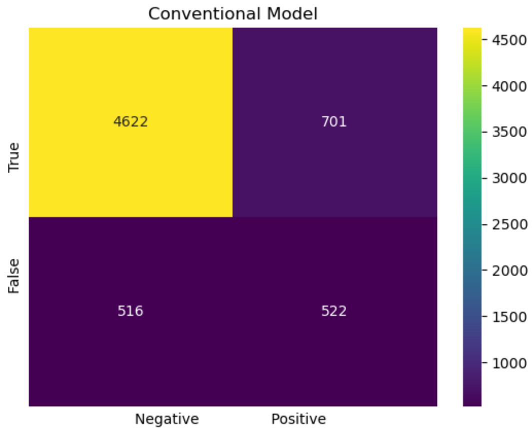
(a) Vibration vs Wind (b) Vibration vs Power (c) Vibration to Gen Speed

Figure 5.3: Regression model, trained on the large data set, used as the basis for the MTL. This figure details the ability to determine the vibration signals from some of the input features.

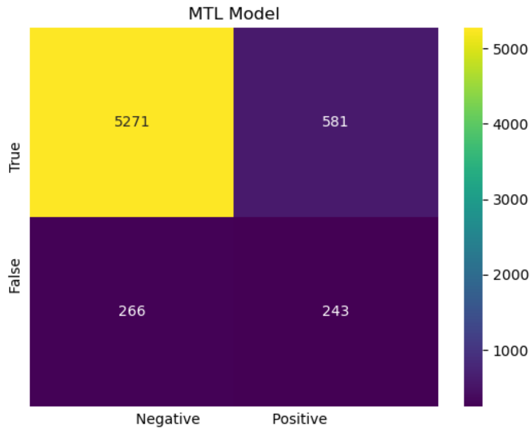
Multi-Task Learning Model

Similarly, the MTL classification model participates in training over 1000 Epochs of 83, 329 trainable parameters. This model uses significantly fewer data points compared to the feature extractor of 8441 producing an accuracy of 91.29%, and an F1-score of 69.54%. The classification for the anomaly detection utilising MTL is more than acceptable, Yang et al [218] indicate that predicting more than 60% of wind turbine faults will reduce the cost of operations and maintenance. The classification results are displayed in Figure 5.4. With anomaly detection, it is expected that the true negative is the median. The most detrimental prediction is a false positive, and this presents one significant portion of this model that would need improvements.

For condition monitoring, a false positive reading would highlight to the operator that the machine is currently running 'normally' but is in a state of potential error. In the moment of a failure event, a chain reaction of issues can lead to catastrophic failure. To prevent this one might call the machine to turn off. However, this is not possible with a false positive, the machine running in a state where there is an error but the observation from the model is contradicting reality. Alternatively, if the anomaly signature is not so serious the machine could be asked to operate at a reduced rate. False negatives do not present as much of a risk as they will not lead to catastrophic failure. Both scenarios are not ideal, both cause a loss in earnings and the whole objective with condition-based maintenance is to optimise the up-time of the wind turbine.



(a) Conventional classification test results



(b) MTL classification test results

Figure 5.4: Classification results from the MT model, highlighting the true negative, false positive, false negative, and true positive rates of the test data set.

Model Comparison

Table 5.1: Model comparison of the conventionally trained model vs the hard-parameter MTL model

	Accuracy	F1 Score
Conventional Model	83.76%	57.56%
Hard Parameter MTL Model	91.29%	69.54%

The conventional model undertook the same treatment as the hard parameter transfer model, with 1000 Epochs. However, the model has suffered from an over-fitting of the result. It is unable to predict many of the errors leading to failure. A comparison of the results from the conventional compared to the MTL model is made in Table 5.1. This highlights that with a reduced data set, the results can be improved by implementing hard parameter transfer with homogeneous data. The conventional model training is hindered because there are simply not enough data points and the correlation of the data is not as great to fully converge the model to surpass the MTL model accuracy.

One significant aspect that is not explored in this paper, is varying the sample size of the data. The current understanding of machine learning is that the larger the data and the network the greater the accuracy. With this current understanding, and the results displayed here, there must be a crossover point where the benefit of transfer learning will diminish, and no longer be a suitable candidate to determine the anomaly detection of an offshore wind turbine. An examination of this crossover point where conventional methods are more accurate would provide a time frame where this technique is recommended.

There are two ways in which time-series data varies, firstly the frequency, and how many samples are taken over a specific period. The second is the length. Both contribute to the size of the time series data used to train the model. In the case of this report, there are 8451 points over two years. This limiting factor will determine the level of improvement from conventional model training to hard parameter transfer.

5.1.4 Conclusions

This paper has successfully highlighted how MTL accelerates the accuracy of data-driven condition monitoring of an ANN with limited data. This is a novel approach to offshore wind energy but is consistent with other areas where this methodology is implemented. The main observation is that infant mortality failure can be quickly detected, and scheduled maintenance can be planned.

By implementing the two different cost functions the model is better suited at extracting the features and classifying through reduced noise and overfitting compared to the traditional method. One observation is that this model is only suitable for a limited period. The longer the wind turbine is in operation, there is less probability that the components will fail, coupled with an increased amount of data.

One of the most significant properties of the multi-task learning method is the increased amount of usable information. You have the vibration and error information both of which can be used in tandem to authenticate the diction process. One detects patterns for anomaly detection and another makes time-series predictions. This is opposed to the classical method which has only one output, the error evaluation, which is useful but ideally more information is advisable in maintenance.

A progression from this concept would be to: investigate the time frame in which MTL is most advisable for maximum accuracy of a condition monitoring system during the infant mortality period. Secondly, comparing varying types of MTL models to determine a baseline model for continuity in the industry.

5.1.5 Population Based Structural Health Monitoring

Population Based Structural Health Monitoring: Homogeneous Offshore Wind Model Development, Innes Murdo Black, Moritz Werther Häckell, Athanasios Kolios, 2022, Renewable energy,

5.1.6 Introduction

Moving beyond detecting damage on a single structure, to diagnosing damage in an entire population raises the issue of acquiring data relating to each of the structures. One of the main concerns with this is the large cost associated with obtaining the information necessary to determine any damage to the structures. Population-based structural health monitoring (PBSHM) seeks to reduce this cost by developing methods that share the information between the structures. The concept of PBSHM is introduced in [26], [75], [66]. If the population of the structures is homogeneous, where the structures are nominally identical, then it may be possible to establish a general model which is common across all structures. Conversely, even if the models are heterogeneous, dissimilar structures, it may be possible to transfer select types of damage across the structures. The most promising technology that will allow for this transfer of information is found in the machine learning discipline of Transfer Learning.

The standard of similarity between the structures is indicative of the level of knowledge transfer between the structures. This can be achieved through quantifying how the structures are similar, and where the similarities lie. This determines what type of machine learning approach is necessary. For this study, the geometry, topology, operation, and material of the offshore structures are the same. The main observation from this is the population is of a strong homogeneous nature. Within a strongly homogeneous population, all the structures have the same material, geometry, and topology (this refers to the components for all the parts in the structure). This implies that all the structures are the same model and brand in the wind farm. The variation within the wind farm is due to the operational state, the location of the wind turbine and manufacturing defects.

The five references provided are focused on the area of renewable energy, the challenges, and solutions in implementing transfer learning. [113] proposes a strategy to tackle small data sets using parameter-based transfer learning. The authors suggest a new model based on transfer learning for wind turbine diagnosis with small-scale data. The model can take the operational information from other wind turbines into account. [37] propose a framework using unsupervised TrAdaBoost learning on SCADA data for WT fault diagnosis. The main observation is that TrAdaBoost shows its superior performance in dealing with data imbalance and different distributions. [67] focuses on the application of machine learning algorithms for structural health monitoring and highlights the importance of domain adaptation in improving the performance of these algorithms. The authors propose a hybrid machine learning model that shows improved performance on several populations of experimental and numerical structures. [172] introduces a transfer-learning-based approach to include physics into data-driven normal behaviour monitoring models. An artificial neural network with an auto-encoder is used in this study to study one month of raw SCADA data. [95] proposes a control strategy for deep transfer learning for fault detection on rotating machinery. The paper applies integrated signal processing on vibration signals. The main observation is the performance is significantly improved by reducing negative transfer and less data is required using this technique than standard deep learning.

The motivation for this study is to provide a solution for low-cost monitoring, where only a few wind turbine generators (WTG) are instrumented with sensors as opposed to the entire fleet. A low-cost monitoring strategy for offshore wind can provide numerous benefits to wind farm developers and operators. Implementing a cost-effective monitoring system can reduce the overall cost of monitoring and increase the reliability of the wind farm. The real-time data provided by a low-cost monitoring system can also aid in the early detection of performance issues and prompt maintenance, resulting in reduced downtime and improved maintenance practices. The monitoring data can also be used to better understand the wind resource and its variability, leading to improved wind farm design and operation. Furthermore, real-time monitoring can provide early warning of potential safety issues, improving safety for workers and maintenance personnel on offshore wind farms. The consequence of a low-cost monitoring strategy is that there will have to be assumptions made on structures that are not instrumented, so the low-cost technique is developed based on a general classifier as opposed to individual models for a monitoring strategy that has the entire wind farm instrumented.

This work focuses on the case of strong homogeneous transfer, with four different machine learning models under consideration. Three models employ supervised domain adaptation techniques, a subcategory from the transfer learning branch, and the last model uses ensemble learning. Transfer learning is one approach to improve the performance of the learner by transferring between different domains. Domain adaptation assumes that there is labelled data in the source domain that can be utilised to aid in the regression of the target domain, by mapping the two domains into a common latent space on which the data distributions are coincident. There are assumptions of domain adaptation, where the input and output feature dimensions are consistent in the source and target domain. This means that Structure One must have the same features as Structure Two in the wind farm. The former method of using an ensemble technique aims to improve the final prediction by grouping the views from the regression models and taking the consensus.

5.1.7 Background

PBSHM involves mapping data and labels from different structures within the population so that a general classifier can be inferred across the entire population. As a result, asset management can potentially be performed digitally for any individual in the population. This section intends to define applicable forms of PBSHM.

For PBSHM it is pertinent to define the contextual difference between homogeneous and heterogeneous populations. This syntax is borrowed from graph theory where the names clearly explain how structures can be represented by attributes. To determine whether two systems are similar enough for knowledge transfer, it is unpracticable to consider every property or dimension of the structure - e.g., comparing the geometric similarity of two structures using 3D, finite element (FE) or computer-aided design (CAD) models of the structure directly would be computationally inefficient. For this study's desired goal, it is more efficient to consider only the properties and dimensions that have a significant effect on the transferability of knowledge.

Differences within the population occur for a magnitude of reasons, and structures are deemed different due to various properties. This can lead to groups of heterogeneous populations. This approach will highlight the four main sources of differences within a structure which are geometry, topology, material, and operation:

- Geometry links to the shape and size of the structure within the population.
- Topology depicts the construction, connection, and location of the components in the structure.
- Material relates to the different, materials classes and specific materials with the associated properties for the structure in the population.
- Operation refers to the different states the operator can curtail the asset too.

Adapting the definitions from graph theory for PBSHM, where a topologically homogeneous population is defined as a group of structures where the geometry σ_m and the material σ_m properties for the nodes and edges of the associated graph can be taken from the base distribution $p(\sigma_m)$, the probability mass of the distribution $p(\sigma_m)$, defines the small differences between the individuals within the population. A strong homogeneous population would have a uni-modal distribution with low dispersion for the geometrical, topological, and material properties. With the strictest, perfect, form of homogeneous composition the underlying distribution of the population is identical. The latter is uncommon, but this assumption can be made if you want to apply conventional Machine Learning (ML) methods trained on one structure and apply this to another. Applying these conventions to the population and categorising the individuals within the population helps determine the difficulty of transferability.

Notable differences in the observable data may occur outside the structural properties of the individuals within the population beyond the categories previously discussed. These differences relate to how the data acquisition and any processing to obtain the features are conducted. A classic example of this would be sensor placement. This will lead to differences in the distribution of the data even though it is placed in the 'same' position. Manufacturing and installation differences will also contribute to small variations in the homogeneity of the data distribution.

5.1.8 Transfer Learning

Transfer learning technologies offer several opportunities for dealing with scenarios where the population form domains and distributions are different for each member when training and evaluating the model [135]. Separate from multi-task learning, where the objective is to lean multiple tasks across different domains [225], transfer learning utilises knowledge from the source to improve predictions on the target task, in our case, using the Damage Equivalent Moments (DEM) from two separate WTGs to create an improved general model. This type of learning is what makes PBSHM achievable. Even when performing in the homogeneous population scenario, variations in the structure, such as location, will lead to differences in the data distributions. Learners trained on one structure will not apply to another structure in the population. Formal definitions of transfer learning and transfer learning technologies are discussed in this section with domain adaptation having an entire subsection.

Definitions

Domain - A domain $D = (\chi, p(X))$ is an object made up of a feature space χ and a marginal probability distribution $p(X)$ over feature data $X = x_{i=1}^{N}$, which is a bounded sample from χ. This is the SCADA data for one structure in the context of this report.

Task – A task $T = (\Upsilon, f(\cdot))$, this would be the DEM for one structure, is an object made up of a label space Υ and predictive function $f(\cdot)(p(x|Y))$ in probabilistic terms, can be inferred from training data $X = x_{x=i,y=i}^{N}$, with $X_i = \sum \chi$ and $y_i = \sum \Upsilon$, noting that both χ and Υ are distributions not individual observations, which are build-up of finite samples sets X and Y. In the case of source domain data-sets $D_s = (\chi_{i,s}, y_{i,s})^N$ and with $x_{i,s} = \sum \chi_s$ and with $y_{i,s} = \sum \Upsilon_s$ and similarly for the target domain $D_t = (\chi_{i,t}, y_{i,t})^N$ and with $x_{i,t} = \sum \chi_t$ and $y_{i,t} = \sum \Upsilon_t$, [135]. Given these artefacts, one can theoretically conduct transfer learning.

Transfer Learning – For transfer learning there must be a given source domain D_s and associated task T_s and a target domain D_t and task T_t. The objective is to improve the target predictive function $f_t(\cdot)$ in T_t by utilising the knowledge from the source, assuming $D_s \neq D_t$ and or $T_s \neq T_t$, [225].

Homogeneous transfer – Homogeneous transfer learning assumes that $D_s = D_t$ and $T_s = T_t$ meaning the attributes are the same. A sub-category of this strong homogeneous transfer, where the domain and task are similar hence, $D_s \cong D_t$ and $T_s \cong T_t$.

Heterogeneous transfer – Heterogeneous transfer learning is when the domain, feature, and task space are non-identical hence, $D_s \neq D_t$, $T_s \neq T_t$, and $x_s \neq x_t$, respectively. It can also assume that $y_s \neq y_t$.

Domain adaptation – domain adaptation is relevant when the inference for the target domain D_t and T_t, and the target predictive function $f_t(\cdot)$ is improved given the source domain D_s and T_s. Assuming $x_s = x_t$ and $y_s = y_t$ but the distributions $p(x_s) \neq p(x_t)$.

To contextualise these definitions in the form of PBSHM for wind turbines, homogeneous transfer learning is a situation where both the source space and target space are the same. This is a situation where the context is the problem between similar assets. This could be where the wind turbines are the same, but have different distribution due to sensor placement, and location, to name a few. Hence $D_s \cong D_t$ and $T_s \cong T_t$. Heterogeneous transfer learning is applied when the features are dissimilar. A situation in the wind turbine industry would be when using data from two different wind turbine designs, e.g., a monopile foundation and a jacket structure. In this case, the features will be dissimilar and the tasks dissimilar, hence $D_s \neq D_t$ and $T_s \neq T_t$.

Transfer Learning Technologies

There is a continually growing variety of transfer learning technologies. This section aims to briefly describe transfer learning, but the focus is on fundamental differences in the approaches of a subcategory called domain adaptation where parameter, instance, and feature-based approaches are described. Visit [63] for a more comprehensive discussion on transfer learning.

Starting with a typical approach of deep learning and artificial neural networks, transfer learning technologies have been developed using fine-tuning. This methodology seeks to learn based on the parameter weights during a particular set of layers in the artificial neural network. The artificial neural network is trained on the domain D_s and some of the layers are fixed. The remaining un-fixed layers are trained using the target domain D_t. Examples of this are conducted in [93] , [64], and [51].

Another approach to transfer learning is knowledge graphs, where the aim is to find objects that define specific entities and their interrelationships. This has been particularly successful in search engines, incorporating semantic searches. Currently, knowledge graphs have been integrated as training data for machine learning models [79], [132].

Like knowledge graphs, the ontologies' goal is to give representations of entities that describe all the interdependences and interactions. Ontologies are useful in outlining knowledge about specific domains. Most importantly ontologies are helpful for explaining concepts and sharing information. If a new project is undertaken ontologies can be reused or transferred to help identify more efficient processes. In the context of PBSHM, an ontology is knowing what types of techniques and methods are most appropriate for one system to another. Ontologies have been explored in multiple industries, including Structural Health Monitoring (SHM) [111], [195], [12].

Domain Adaptation

Domain adaptation is a subclass of transfer learning to transfer the feature space between the source and the target domains, based on the assumption that the marginal distributions of $p(x_s) \neq p(x_t)$ are not the same. This type of technique is primarily used in homogeneous transfer learning where the source domain and target domain are similar. There are three main approaches to domain adaptation which are parameter, feature, and instance-based.

Parameter-based domain adaptation takes the parameters of a trained model, is built using the source domain D_s data and is then adapted to suit a model for the task domain D_t.

Feature-based domain adaptation techniques are designed on the research of common features which have similar attributes to the source T_s and target T_t task. A new feature, often called the encoded feature space, is built with a projecting application, which aims to correct the difference between the source $p(x_s)$ and target $p(x_t)$ distributions. The task is then considered to be in an encoded space.

For instance-based domain adaptation the general principle is to redistribute the labelled training data to correct the differences between the source $p(x_s)$ and target $p(x_t)$ distributions. This re-weighting consists of multiplying, the individual loss of each training instance by a positive weight. The re-weighted training instances are then directly used to learn the task.

Negative Transfer

One of the major drawbacks, when performing transfer learning between WTGs, is if the information is incorrectly detailed from one domain to another as this can reduce the performance of the general learner when compared to the learning from the target domain alone. This phenomenon is known as negative transfer and is most prominent when the source, D_s and the target, D_t domain are most dissimilar, e.g. heterogeneous. The fundamental idea of transfer learning is that there must be some shared information across domains. This may be hard to contextualise when data is unlabelled, or the tasks are dissimilar.

Negative transfer raises the important question: When is it right to transfer knowledge? This motivates the reasoning behind developing a measure of the similarity of structures. The case study provides information on the heterogeneity of the data used in this work. This study reinforces our understanding of the data and helps mitigate the issue of negative transfer as we become aware of where differences in the distributions lie.

5.1.9 Methodology

This study uses data from the Wikinger wind farm, where there is only a select amount of the operational wind turbines that have CMS with strain gauges installed; to be specific, four out of 64. This section aims to develop a variety of models that can perform PBSHM using structures of two structures as the source and target respectively in the pursuit of a general classifier for the entire farm. The population form is the DEM on the foundation of the structure which can be used to determine the fatigue life. This section starts with how the SHM data is gathered, what the population form is for this study, and then what the definition of the population is. This is then used to detail the individual models including how the data is pre-processed, the model development and the error metrics used.

The Data

The measurement data used as input for the low-cost monitoring technique comprises two different frequencies of data 25Hz CMS data, and SCADA data at 10-minute averages. The process of determining the damage equivalent moments (DEM) is in the next subsection The WTGs available for this study are equipped and unequipped with strain gauges (SG). The three positions with SG are WT 1, 2 and 3.

SCADA systems are equipped on all WTGs and, depending on the feature, the resolution varies. This encompasses meteorological information at the hub height, such as wind speed, wind direction, temperature, and pressure. The SCADA data also covers the operational signals such as power production, the pitch angle of the individual blades and the rotor rotational speed.

To increase the value of a low-cost monitoring program, transferring knowledge that is unavailable in other wind turbines can provide insight and confidence into other assets. If one can infer knowledge accurately on another WT, then one can save money by installing strain gauges on a fraction of the WTG. Based on this principle, the population form is the DEM where only four WTGs have the CMS strain gauges installed.

There are a considerable number of model and feature spaces that can be applied to represent a population form. For a wind turbine, the form could be wind turbine power curves to frequency responses. But, in this case, the form is fatigue damage equivalent moments for the jacket support structure. The entire population in this study has the same geometry and material, with small deviations in topology due to the location.

Fatigue Damage Equivalent Moments

The condition monitoring system calculates the forces from the strain gauges on the foundation of the structure. From these forces, the damage equivalent loads are produced. The two-phase operation is as follows:

Phase 1 - Calculation of forces from strain.

1. Run dynamic ROSA simulation [149]
2. Extract stress at selected element via Fatima
3. Calculate strain using Hooke's law
4. Calculate forces with internal functionality
5. Compare forces with extracted forces

Phase 2 - Calculation of DEM from forces

1. Gather applicable force location from the sensor location
2. Calculate the cyclical forces at that sensor
3. Apply ASTM E1049-85 rain-flow cycle counting algorithm [15]
4. Apply a scale factor to force accumulation.
5. Sum the damage accumulation over the cycles to calculate the DEM

Homogeneous Population

A homogeneous population in the context of offshore wind is one where the task distributions are similar for all WTGs which, depending on the features selected, may have similar feature spaces. This makes homogeneous populations ideal candidates for domain adaptation methods, and to demonstrate the effectiveness of transfer learning for PBSHM. This section presents a homogeneous population of three wind turbines located in the Wikinger wind farm. All the structures are of the same design and capacity; hence, they have the same material and geometry. The damage equivalent loads histogram is presented in Figure 5.5.

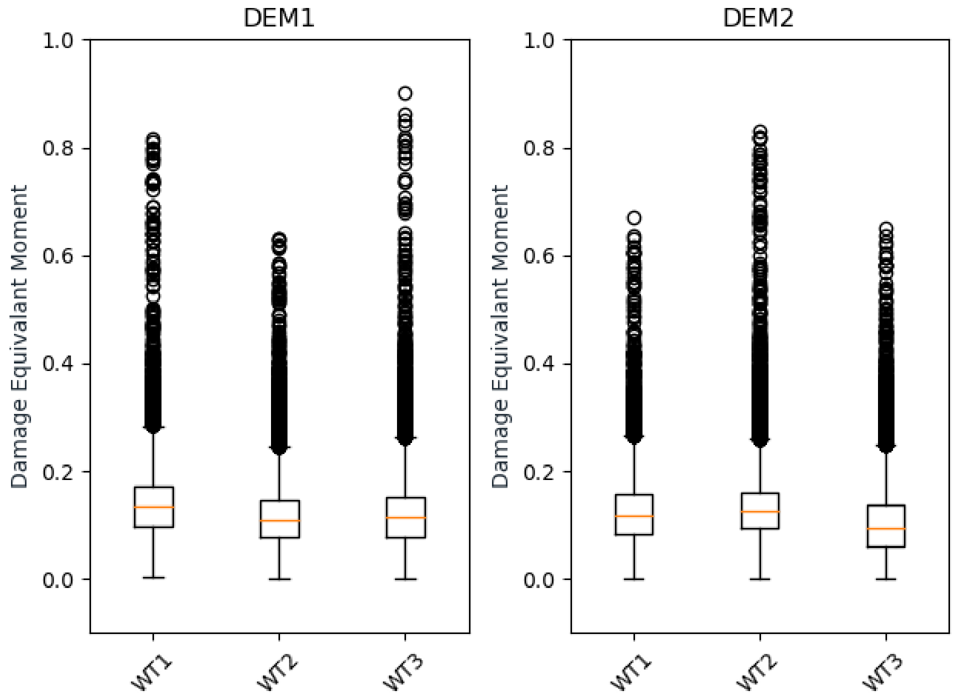


Figure 5.5: Normalised damage equivalent moments box plot of the two orthogonal directions for all three of the wind turbines.

The three structures can be considered as a homogeneous population, as they are structurally similar in their representation and the material and geometry parameters can be described by a uni-modal distribution with low deviation. The SHM problem presented here is due to small deviations in the DEM which arise from the location and the operational context of the individual wind turbines within the wind farm. Figure 3 displays two-dimensional heatmaps of the DEM amplitudes for various operational features for the three wind turbines.

Small deviations in the overall distributions are highlighted in Figure 5.5. One of the contributors to these results is the location as the entire wind farm has deviations on the water depth. The design of the individual WTs does not require the topology and geometry to be altered for single locations but rather in design clusters spanning a range of water depths. One consistency is the height of the transition piece. It needs to be at a constant height across the population, meaning that there are variations in the length of the foundation which influences the dynamics. Another aspect is the operation and metrological differences. Figure 5.6 displays a heatmap of these features for each of the WTs which are influenced by the location and the degree of turbulence intensity based on the direction of the wind. Navigating these deviations in the distribution is the aim of the general model development.

The Fréchet number is a numerical method of determining the similarity of a population form for a given domain in a metric space. The Fréchet distance is a popular measure of similarity between the two domains and is calculated by:

$$\Delta = \|\nu_s - \mu_t|\ ^2_2 + T_r(|\Sigma_s + \Sigma_T - 2(\Sigma_s \cdot \Sigma_T)^{1/2}) \tag{5.4}$$

Where μ_s, μ_T are the mean along the source and target along the first axis, Σ_s and Σ_t are the covariance matrix of the source and target domain datasets. This considers the location and the ordering of the points along the points of both domains. For homogeneous populations, the value of the Fréchet distance will be 0, for strong homogeneous distributions the value will be small, and heterogeneous will be large. The value is dependent on the length of the instances in the domain and the magnitude of the values.

Table 5.2: Fréchet Distance for all three wind turbines

Fréchet Distance			
WT1	0.00		
WT2	13.48	0.00	
WT3	14.48	14.92	0.00
	WT1	WT2	WT3

By applying this to the DEM for each of the wind turbines, the results in Table 5.2 indicate that the population form is homogeneous with values close to 0. This is a numerical representation, and it is limited in the scope of interpretation on the homogeneity. This is a technique that mathematically reinforces the decision after

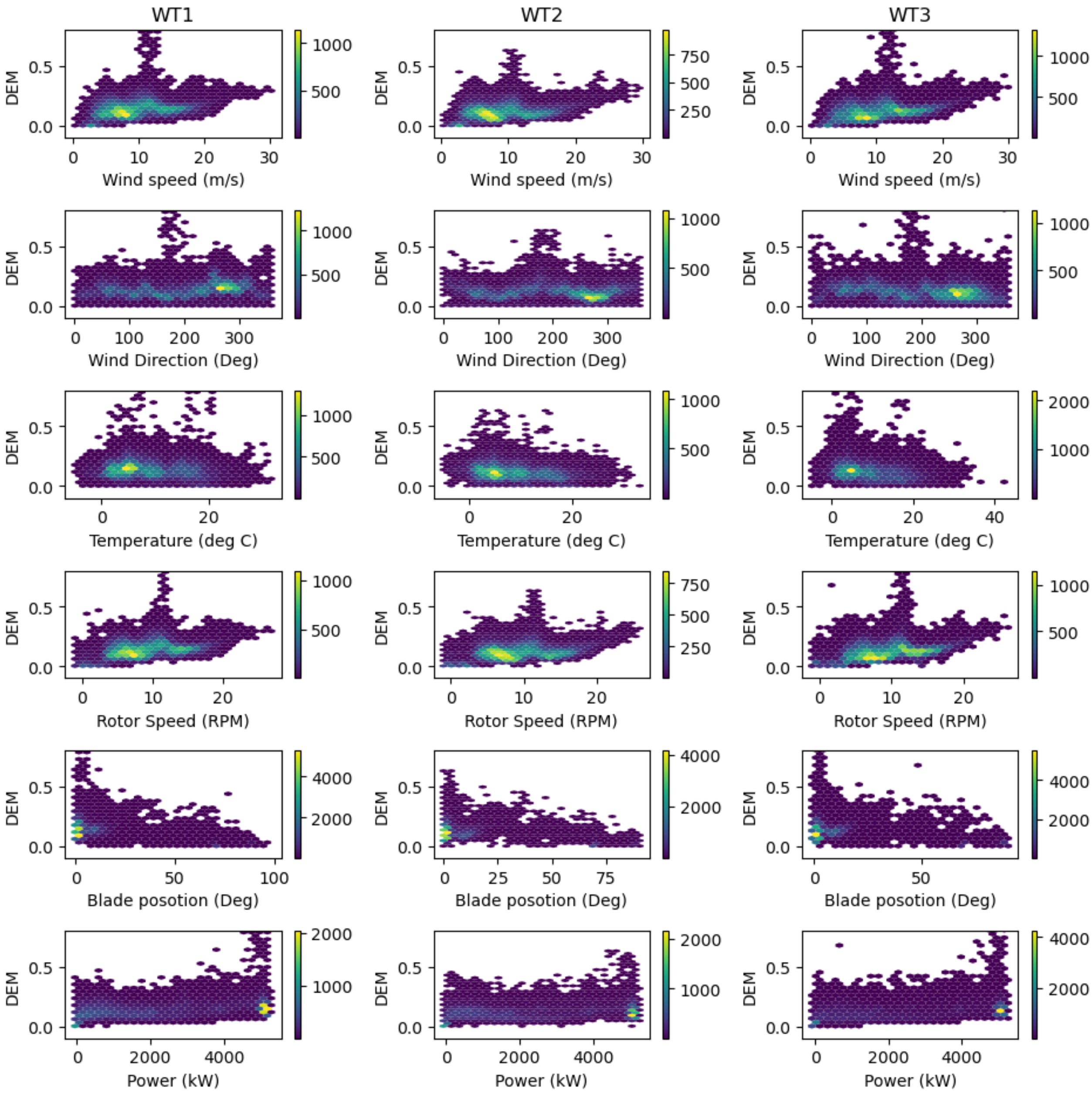


Figure 5.6: This series of figures highlights the operational and environmental effects on the population of the wind farm. This includes the power, blade position, wind speed, rotor speed, and temperature against the fatigue damage equivalent moments in direction 1.

observation of the data. Nevertheless, in the following section, the location is the main source of heterogeneity for this population form and breaking this down into more detail will provide greater insight into the small deviations in heterogeneity observed in Table 1.

Table 5.3: Statistical values of both DEM for all three wind turbines.

DEM 1				DEM 2		
	WT1	WT2	WT3	WT1	WT2	WT3
Mean	0.1382	0.1503	0.1193	0.1233	0.1092	0.1049
Std	0.0634	0.0576	0.0629	0.0588	0.0623	0.0623
Min	0.0015	0.0015	0.0015	0.0017	0.0015	0.0017
Max	0.8158	0.6304	0.9024	0.6669	0.8286	0.6494

To contextualise this, some statistical measures have been calculated based on the normalised DEM and these are represented in Table 5.3. This includes the mean, standard deviation (Std), and minimum and maximum of the DEM in both orthogonal directions. There is an almost identical nature to the three WTGs; WT2 does have higher and lower maximums and minimums, but both the mean and standard deviations are identical to four significant figures.

Models

An Artificial Neural Network (ANN) is an optimal base model suited for stochastic problems to estimate the DEM without direct measurement. Sensors often cannot be placed on the sub-structure, and SCADA data is usually gathered in the tower and nacelle, hence, this does not provide direct information on important parameters on the foundation's structural behaviour. Therefore, a general model using the SCADA data to determine the DEM would provide excellent potential in the application of wind turbine foundation monitoring. This subsection will briefly describe an ANN and the differences that the domain adaptation models make from the original ANN model.

Three domain adaptation methods are implemented in this study that all use a base model of an ANN and are altered based on their specific procedures. However, before these procedures are explained, the original model is developed using the data from one wind turbine only. No transfer learning is carried out to generate this model; only hyperparameter optimisation using WT1.

The architecture of an ANN is built up of hidden layers where each layer has a density of neurons attached to that. Techniques such as dropout can be introduced to aid in removing bias within the architecture, increasing weights (w) to 0 and 1 such that the sum of the weights remains constant. The individual neurons have a synaptic weight associated with them which can be represented as an activation function. For this type of problem, the Rectified Linear Unit (RELU) is used. Weights w are associated with each activation function and the entire network is curated based using the Adam [?] Optimiser.

CORrelation Alignment (CORAL)

CORAL [183] is a feature-based domain adaptation method, intending to minimise the domain shift from the source D_s to the target D_t by aligning the second-order statistics of the source and target distributions. The method transforms the source features to minimise the Frobenius norm [116] between the correlation matrix, the input target data and the transformed input source data. The transformation is described by the following optimisation:

$$\min_{A} ||A^T C_S A - CT||_F^2 \tag{5.5}$$

Where, A is the feature transformation matrix such that C_s and C_t is the correlation matrices of the source and target data, respectively. The solution of this operation can be written in explicit form and the feature transformation is computed in four steps:

$$C_S = Cov(X_S) + \lambda I_P \tag{5.6}$$

$$C_T = Cov(X_T) + \lambda I_P \tag{5.7}$$

$$X_s = X_S C S_S^{-1/2} \tag{5.8}$$

$$X_s = X_S C S_T^{1/2} \tag{5.9}$$

Where λ is the regularisation parameter.

TwoStageTrAdaBoostR2

TwoStageTrAdaBoostR2 algorithm, [137], is an instance-based domain adaptation method suited to regression tasks. This method is characterised by the ‘reverse boosting' principle where the weights of the source instances predicted decrease at each boosting iteration, and one of the instances increases. In the 'two stages' version of TRAdaBoostR2, [137], the algorithm is where the weights of the source and target instances are carried out separately. In the first stage, the weights from the source instances are frozen, but the ones on the target instances are updated according to the classical AdaBoostR2 [45]. In the second stage, the weights of the target instance are now fixed whereas the ones on the source are updated according to TrAdaBoostR2. During each first stage, a cross-validation score is computed with the labelled target data. The cross-validation score is used to determine the most effective estimator within all boosting iterations. This algorithm performs the following steps:

- Normalise the weights $\sum w_s + \sum w_t = 1$
- Fit an AdaBoostR2 estimator $f_s(\,)$ on the source and target labelled data (x_s, y_s), (x_t, y_t) with the respective importance initial weights w_s, w_t. During the training of AdaBoostR2, the weights of w_s are frozen.
- Compute the cross-validation score on (x_t, y_t).
- Compute the error vectors

$$e_s = L(f(X_s), y_s) \tag{5.10}$$

$$e_s = L(f(X_s), y_s) \tag{5.11}$$

- Normalise the vectors

$$e_s = e_s / max_{e \in e_s \cup e_t} \tag{5.12}$$

$$e_s = e_s \; max_{e \in e_s \cup e_t} \tag{5.13}$$

- Update the source and target weight

$$w_s = w_s \beta_s^{e_s} / Z \tag{5.14}$$

$$w_t = w_t / Z \tag{5.15}$$

Where Z is the normalising constant B_s is chosen so that the sum of the weights is equal to $\frac{n_t}{n_t+n_s} + \frac{t}{N-1}(1 - \frac{n_t}{n_t+n_s})$with t the current boosting iteration number. B_s is located with a binary search.

- Return to the first step and loop until the number of boosting iterations is reached.

The general model is selected by the best estimator according to the cross-validation.

RegulartransferANN

RegulartransferANN [35] is a parameter-based domain adaptation method. This assumes that an effective global estimator can be obtained using ladled target data. The aim consists of fitting the neural network on the target data based on the objective function which is regularised by the Euclidean distance of both the source and target parameters:

$$\beta_t =_{\beta_1,....\beta_D} ||f(X_t, \beta) - y_t||^2 + \sum_{i=1}^{D} \lambda_i ||\beta_i - \beta_{Si}||^2 \tag{5.16}$$

Where the estimation function is f with D network layers. B_t is related to the target parameters, β_s is the source neural network parameters:

$$\beta_s =_\beta ||f(X_s, \beta) - y_s||^2 \tag{5.17}$$

The trade-off parameter is λ_i, where training is biased towards source or target domains depending on the associated weighting.

Pre-Processing

The task of training models for predictions that involve multiple data streams from different structures requires coordination so that effective ML modelling can take place. Several issues arose when working with different data streams in ML. This section discusses the process of dealing with these issues by using feature, selection, projection, and data cleaning. The first barrier that prevents effective ML modelling in this problem is that sensors tend to break, just like most industrial components. This meant identifying a suitable time frame, where the maximum period of operational uptime should be met. The optimal period took place between 02/10/2018 and 02/07/2019 (nine months). Secondly, synchronising the CMS data to the SCADA data had to be conducted and matching the instances of 10-minute intervals was the next step.

After synchronising the data matching instances for all WTs, the next stage was to perform data cleaning. This procedure involved implementing previous value imputation in place of missing values and NaNs. At this point, there were zero missing values and over 300 features. The next step was to reduce this to suit the needs of the ML task.

For the feature selection process, the inputs were taken from the SCADA data since this is equipped on all WTGs. The final features comprised hand-picked features and statistically relevant features. The outputs are the DEM in the two orthogonal directions from the CMS systems' strain gauge rings. Now that both the input and output data have been established, the feature projection is implemented where the features are normalised from $0 - 1$.

Lastly, the datasets were split into training and test data to a ratio of 80% to 20%, respectively. The source domain was WT1, and the target domain was WT2 using 80% of the datasets for the domain adaptation model. Testing was conducted using the subsequent 20% of both the source and domain datasets and the entirety of WT3. In summary, SCADA data features were used as inputs, with WT1 as the source domain and WT2 as the target domain, where the general model makes estimations on the DEM in the two orthogonal directions. Testing takes the remaining 20% of the source and target domain and the entire dataset of WT3 to validate the results using the same input and output features for all WTG.

Model Development

The method implemented involves three stages. Stage one involves defining the most suitable ANN and source domain and this became the basis for all the subsequent experiments as the architecture is mimicked, this is developed using the Keras API [2]. Stage two entailed hyper-parameter tuning the remaining experiments using the same architecture from the ANN and application of the three domain adaptation algorithms using the Adapt API [46]. Stage three will investigate the optimal model further by altering the source and domain data.

To establish a standardised process for all the experiments, the pre-processing procedure was applied to all three WTGs as this provides a standard platform for model development and increases the performance of the model. This also provides a consistent feature for training and testing since all three WTG domains consist of the same input SCADA features and output DEMs based on CMS data. To train all the domain adaptation models, the source domain dataset and features are made up of the data from WT1, and the target domain dataset with features from WT2. WT3 is not used in any of the training and is only used for validating the model.

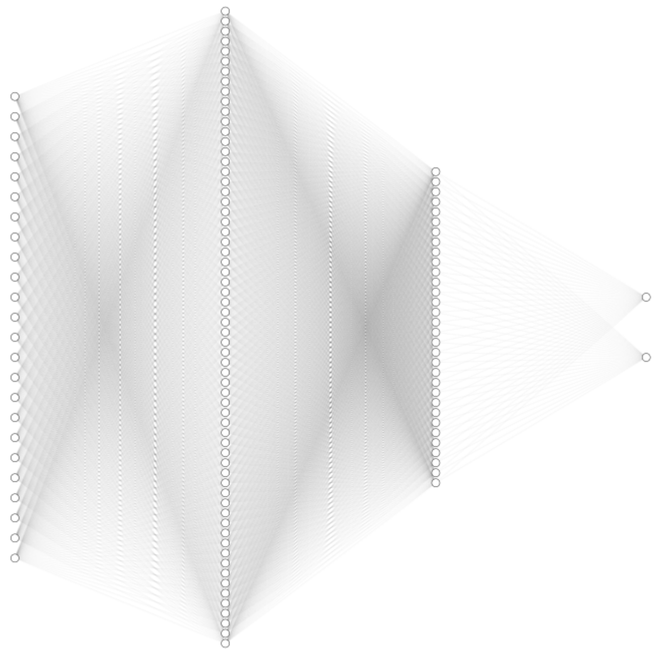


Figure 5.7: An artificial neural network architecture with 2 hidden layers of density 64 and 32 respectively, batch size of 1, and a dropout ratio of 0.2.

In stage one, an optimal model was produced by conducting an exhaustive search iterating the number of hidden layers, the density of neurons on each layer, the ratio of dropout, batch size, and the degree of convolution. The final form of the ANN is displayed in Figure 5.7.

In stage two, the experiments were set up to firstly investigate the performance of each of the techniques and secondly to establish what the most suitable method would be to carry forward for further testing. Techniques from all the main approaches of domain adaptation are implemented which include parameter, feature, and instance-based methods; these are RegularTransferNN [35], CORAL [183], and TwoStageTrAdaBoostR2 [137] respectively. In each instance, the model is tuned using their distinct hyperparameters for each of the methodologies.

The goal of the experiments is to transfer knowledge between all three WTGs using a single general model by taking SCADA data and making inferences on the DEM such that a low-cost monitoring methodology can be applied to the entire wind farm. The RegularTransferNN is the most accurate model on all three metrics for this architecture, with input data and output data for WT1 as the source domain and WT3 as the target domain. Further testing was conducted to investigate what is the most suitable target and source domain by altering them.

Error Assessment

The performance of the regression algorithms is based on how the general classifier can make predictions on DEM for all three wind turbines. In this case, common KPIs are implemented which provide a percentage of the performance.

Mean absolute error (MAE) is a measure of the errors between the paired observations. This is the arithmetic average of the absolute error where $\hat{y}_i$ is the prediction, and y_i is the true value.

$$MAE = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n} \tag{5.18}$$

The coefficient of determination (R^2) this represents the proportion of the variation from the predicate value to the actual value and μ is the arithmetic mean.

$$R_2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \mu)^2} \tag{5.19}$$

Cumulative error ($CFPE$) encompasses the total error for all instances used in the model. Where a conservative result would be a negative % value and an underestimate would have a positive % value. Where m is the power factor, in this case, it is set to 4.

$$CFPE = \frac{\sum_{i=1}^{n}(\sqrt[m]{y_i}^{m}) - \sum_{i=1}^{n} \sqrt[m]{\hat{y_i}}^{m})}{\sum_{i=1}^{n} \sqrt[m]{y_i}^{m})} \tag{5.20}$$

5.1.10 Results

The results section is broken down into three sections in the pursuit of an optimal model. Stage one aims to determine the most suitable ANN architecture for traditional SHM, training on one structure only, and then applying the remaining WTG to this model for comparison. Stage two takes the model from stage one and applies it to all three of the domain adaptation techniques. Stage three alternates the source and target domain adaptation techniques in pursuit of the optimal model.

Stage 1

To demonstrate the applied form of strong-homogeneous populations, three structurally equivalent WTGs are applied to an ANN using standard SHM techniques. The ANN is trained on one structure and then the trained model is applied to the other WTGs. The data used is the test dataset during normal operation for all three WTGs. The results are summarised in Table 5.4 with all three metrics presented.

Table 5.4: Standard SHM approach to PBSHM, table indicating the test results from the ANN trained on one WTG and tested on the resulting two WTGs.

	MAE (%)			CFPE (%)			R2 (%)		
WTG	WT3	WT2	WT1	WT3	WT2	WT1	WT3	WT2	WT1
WT3	0.025	0.031	0.029	-16.12	-29.43	-18.71	0.65	0.51	0.54
WT2	0.037	0.030	0.042	-7.37	-22.80	-31.56	0.31	0.45	0.15
WT1	0.034	0.030	0.029	-17.04	-29.06	-21.81	0.42	0.45	0.55

The normal operation test using the standard approach of training on one structure and then applying the resultant WTG provides varying degrees of accuracy. All the tests, in this case, fall below the threshold for the CFPE of ±10% This is expected as there are small perturbations in label space due to manufacturing tolerances and location-specific effects.

Stage 2

One of the main challenges of PBSHM is performing damage identification on the population with different label spaces $y_s \neq y_t$. However, in this case, the label spaces are strongly homogeneous but we have identified that the general classifier using the standard approach to SHM does not provide adequate inferences from section 1. The results of applying the three domain adaptation techniques using WT1 as the source domain and WT3 as the target domain are displayed in Tables 5.5, 5.6, and 5.7.

Table 5.5: Cumulative error for the optimal models during the model development stage. The test data sets are used to determine the error for both DEMs.

Accumulative Error (%)				
Model	WT1	WT2	WT3	avg
Artificial Neural Network	-21.806	-31.558	-18.712	-24.025
CORR	-23.400	-34.475	-23.239	-27.038
RegularTransferNN	-2.225	-18.216	-5.639	-8.693
TwoStageTrAdaBoostR2	-8.626	-17.240	-4.338	-10.068

Table 5.6: Mean Absolute error for the optimal models during the model development stage. The test data sets are used to determine the error for both DEMs.

MAE (%)				
Model	WT1	WT2	WT3	avg
Artificual Neural Network	0.026	0.043	0.030	0.033
CORR	0.037	0.029	0.034	0.034
RegularTransferNN	0.033	0.030	0.031	0.031
TwoStageTrAdaBoostR2	0.034	0.050	0.039	0.041

Table 5.7: R2 Score for the optimal models during the model development stage. The test data sets are used to determine the error for both DEMs.

R2 Score (%)				
Model	WT1	WT2	WT3	avg
Artificual Neural Network	0.546	0.151	0.538	0.412
CORR	0.637	0.191	0.564	0.464
RegularTransferNN	0.594	0.338	0.553	0.495
TwoStageTrAdaBoostR2	0.431	0.158	0.310	0.300

The goal of the experiments is to transfer knowledge between all three WTGs using a single general model by taking SCADA data and making inferences on the DEM such that a low-cost monitoring methodology can be applied to the entire wind farm. Tables 4, 5, and 6 highlight the variation in the accuracy from the ANN to the three-domain adaptation techniques. The RegularTransferNN is the most accurate model on all three metrics for this architecture, input data and output data, and is the only model that reaches the target CFPE of $\pm 10\%$.

CORAL focuses on aligning second-order statistics of the source and target domains by transforming feature representations. TwoStageTrAdaBoostR2 adapts the AdaBoost algorithm for domain adaptation by assigning higher weights to unclassified target domain instances during boosting. RegularTransferNN utilises neural networks for transfer learning, involving pre-training on a source domain and fine-tuning on the target domain. CORAL and RegularTransferNN have higher potentials through alignment or fine-tuning, while TransferAdaBoost is specifically tailored to adapt the AdaBoost algorithm. In this particular case, the RegularTransferNN was the most suited to this particular dataset. The fine-tuning of the trained ANN provided a better general regression model than the other two. However, this may not always be the case, it is always dependent on the dataset.

Stage 3

In the pursuit of an optimal general classifier for the Wikinger wind farm, the first two stages are set up to determine the optimal choice of model. Stage three investigates what model data sources are best suited to achieve an average optimal score in all three metrics. The aim of this is to determine the sensitivity of the model's accuracy with the input data. If there is a large difference in the results then this may constitute to a large degree of variance by the general model for an entire wind farm.

Table 5.8: Comparison of the test results from the RegulartransferANN using different source and target data sets.

	MAE (%)			CFPE (%)			R2 (%)		
	Training Data WK(Source/Target)								
WTG	1/2	2/3	1/3	1/2	2/3	1/3	1/2	2/3	1/3
3	0.031	0.029	0.023	-5.64	-18.02	-20.24	0.55	0.49	0.63
2	0.030	0.029	0.041	-18.22	-25.48	-10.36	0.34	0.57	0.17
1	0.033	0.034	0.025	-2.23	-26.10	-18.58	0.59	0.42	0.69
Average	0.031	0.031	0.030	-8.69	-23.20	-16.39	0.49	0.49	0.50

The selection process for determining the optimal model is to have the least $CFPE$ and the highest $R2$ and MAE. A threshold is placed on the $CFPE$ of $\pm 10\%$ since the purpose of this type of model is to predict the DEM which in turn is used to determine the fatigue life estimations of the structure. A high or low estimation of the accumulation of DEM will link to poor fatigue life estimations. Ideally one would be conservative in the prediction of the accumulated DEM as one would not like to oversell. There is only one setup in this entire process that achieves this, unfortunately, it does not have the highest average $R2$ or MAE as seen in Table 5.8 with the WT1 and WT3 source domain and target domain setup. However, this particular set-up achieves higher results when making inferences on both the source and target domain but is less accurate in the inferences on WT2. Thus, not achieving the desired goal of a general classifier.

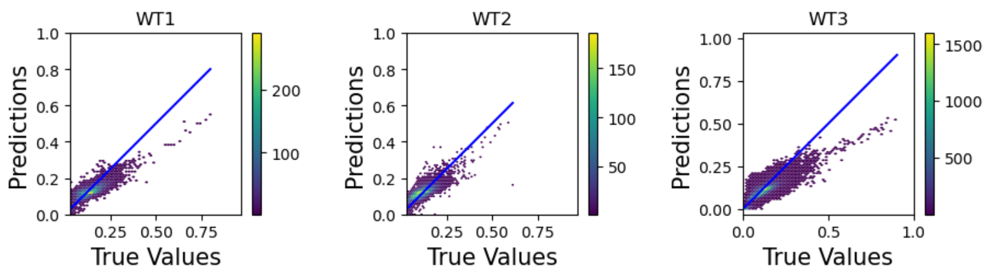


Figure 5.8: Error plot showing the predictions from the general model against the actual value for the DEM 1 of all three WTGs.

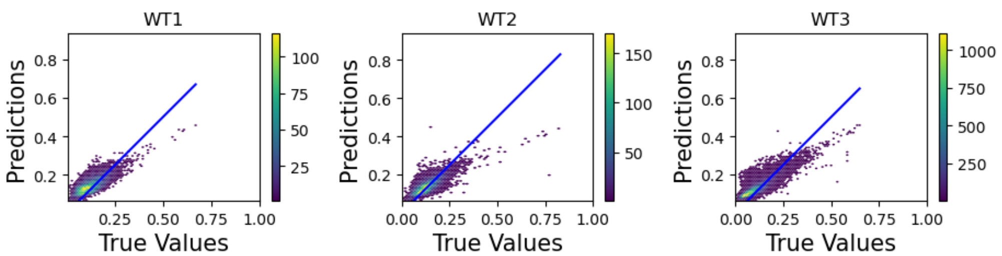


Figure 5.9: Error plot showing the predictions from the general model against the actual value for the DEM 2 of all three WTGs.

In contrast to the high consistency of the MAE of around 3% Figure 5.8-5.9 display the evaluation of the general model estimations. The hyperparameter λ was altered from 0.1 – 0.99, varying the bias general model training from the source domain data to the target. The optimal model presented has a 50/50 split with $\lambda = 0.5$. One aspect of the final general model is the lack of accuracy at high DEM values, where one sees the highest deviation from the real value and is the main constituting factor to the reduced performance of the $CFPE$.

5.1.11 Conclusion

Knowledge transfer is an important process in PBSHM. The benefit of transferring knowledge about structural health from one structure to another within the population is imperative to the progress of low-cost digital-enabled asset management for WTGs. It is important when applying this technique that a process is conducted to determine what similarities exist within the population so that negative transfer can be avoided. In this text two categories of structures have been discussed: homogeneous and heterogeneous.

Implementing transfer learning in the form of domain adaptation has been demonstrated to effectively mitigate problems where both the features and label spaces are consistent. This paper has demonstrated that domain adaptation applies to homogeneous populations where there are small deviations in the geometry due to the water depth, manufacturing tolerances and sensor placement.

CORAL focuses on aligning second-order statistics of the source and target domains by transforming feature representations. TwoStageTrAdaBoostR2 adapts the AdaBoost algorithm for domain adaptation by assigning higher weights to unclassified target domain instances during boosting. RegularTransferNN utilises neural networks for transfer learning, involving pre-training on a source domain and fine-tuning on the target domain. CORAL and RegularTransferNN have higher potentials through alignment or fine-tuning, while TransferAdaBoost is specifically tailored to adapt the AdaBoost algorithm. In this particular case, RegularTransferNN proved to be the most suitable for the dataset, as the fine-tuning of the trained artificial neural network provided a better general regression model compared to the other two techniques. However, it's important to note that the effectiveness of these techniques is highly dependent on the characteristics of the dataset, and the choice of the most suitable approach may vary accordingly.

If a higher accuracy of a model is required to determine the remaining fatigue life, further optimisation measures can be taken: These would include: 1 - separating the general model into discrete models for operational modes; 2 - discreet model development based on the wind direction; 3 - further studies on the feature selection and hyperparameter tuning; 4 - implementation of high-frequency SCADA data for higher order statistics. Measures 1 and 2 may lead to the development of specific models and increase bias within the estimation. As such a less accurate but more general model may produce greater estimations of unforeseen events.

Chapter 6

Discussion

This thesis desires to make advances in areas required for DEAM. This is a multidisciplinary task that has been addressed by tackling the challenges that arise when using data. Stage 1 is how to collect and process the data, stage 2 highlights how one would implement a low-cost DEAM tool. Stage 3 highlights the importance of analysing the available information. Stage 4 provides innovative ways of using data for DEAM. This section discusses the important findings and considerations from each stage.

6.1 Literature

The literature review aspect of this research highlights models that can be used as well as gaps where voids can be filled. It provides an overview of topics related to asset management for both vibration-based techniques and data-driven techniques as well as popular machine learning techniques used in DEAM.

It was most importantly identified that there is a plethora of machine learning models and applications that apply to DEAM in offshore wind structures. Machine learning models offer different properties when training algorithms with different data structures and distributions. Physical models can also capture behaviour in different areas of machine learning with varying degrees of accuracy and complexity. As more structures are fitted with SCADA and CMS systems, researchers and engineers have more opportunities to make advances. Completed models become more within the grasp of becoming operationally ready. Additionally, complications in the machine learning pipeline can become standardised. One of the main observations from the literature review is the increasing interest towards integrating data-driven decision-making in asset management.

One of the outcomes from the literature review is that the creativity and interest in machine learning applied to DEAM is increasing. The number of research articles is growing each year. Unlike Physical models that require expert insight to conduct, machine learning requires knowledge of the data structures only. This can be particularly useful since inspiration when creating models can be gained from other industries.

When conducting a data-driven model, it is useful to begin the process by assessing the data first, examining the extent in which one of the 5 processes; trending, clustering, normal behavioural modelling, damage modelling, alarm assessment or performance modelling can be performed. Secondly, some research on the application is being conducted to provide a benchmark for the model. Lastly, conducting some research in not only the wind industry for the application being created. Inspiration can come from other industries when creating a machine-learning model.

A common pitfall when recreating a preexisting model is that there is no standardised preprocessing for data-driven machine learning. Since there are no specific standards on how it is managed in offshore wind. Recreating a preexisting model with the same architecture may provide dissimilar results. Addressing this will reduce the ambiguity of model performance.

6.2 Data Collection and Processing

The data collection and processing chapter in this thesis forms the first instalment of how data is used in DEAM. The first aspect is on SCADA data. How data is captured and what must be done to it for effective processing. It finishes off with the main advantages of standardising and implementing an effective preprocessing procedure such that it can improve the results of machine learning models.

An important part of machine learning is the quality of data, and a preprocessing procedure must be carried out every time with a DEAM tool. A study was conducted on a trend condition monitoring tool. It takes raw data from the SADA system from one wind turbine using the wind speed and the power output data to produce a power trending tool. The data was applied to 5 machine learning tools; ANN, SVM, GPR, K-NN and LGBX. After performing the preprocessing procedure and making a comparison of the results with raw versus the processed there was a unilateral improvement to the test results.

One of the main objectives of this study was to highlight that there is a gap in academia and also the standards when it comes to carrying out preprocessing. From the standards, other industries have outlined a preprocessing procedure, such as the gas turbine and aerospace industry. Both give a detailed description of how to deal with time-series data for condition monitoring. Offshore wind, sadly has not. The advantage described in this study is that a standard procedure will improve transparency since the format of the data will be predetermined with a standardised process. It will also reallocate the academic and industrial resources towards model development. Another consequence of standardised data is that models can be collaboratively developed if one knows how the data structure works a multi-agent system is more easily tackled.

The trend condition models were investigated by conducting a Mont-Carlo simulation. Repeating the training and testing for each of the machine learning models 100 times and storing the error quantification. It was found that there is a varying degree that the preprocessing aided in improving the results.

6.3 Case Study of an Operational Wind Farm

One of the main limitations of digital-enabled asset management in offshore wind research is the limitation of data availability. Fortunately, this study has a large amount of data available and in this section, it is explored with the main aim of determining how data can be shared for a PBSHM tool. This is conducted by firstly, the fixed boundary conditions from the geometry, topology, operational conditions, and materials are critiqued along with the SHM technique of estimating the fatigue damage equivalent load on the jacket structure. The study observes the effects of environmental and operational conditions on the individual structures of the population and compares this against several wind turbines from the same farm. Lastly, there are some honourable mentions of challenges that are currently with the monitoring campaign such as scour, marine growth, and corrosion that need to be discussed in the context of PBSHM.

One of the main findings from this case study is that contrary to current graph theory, in offshore wind population-based structural health monitoring, geometry, topology and material are not sufficient enough in analysing the degree to which a population is similar. To determine the homogeneity of the population form the operations must be included towards the argument. One of the main reasons behind this view is operators of wind farms have different operational philosophies. This means that there are wind turbines with the same design, but have different operators curtailing the WT at different degrees hence having different damage envelopes.

Building on the previous statement, the next stage of the case study was to define and determine the population form of the Wikinger wind farm. This was a systematic approach involved scouring the design documentation of the wind farm and also analysing the data. A flowchart of this process is detailed in Figure 4.4. The documentation indicated that the material, geometry and operation of all the components were the same however, the topology varied. This came in the form of the location, where the sea depth varied and as one would expect the location of each wind turbine. The main effect expected from the different locations was the environmental conditions on each of the constituent wind turbines. Investigating the data highlights that there is a difference in the distribution of the DEM for north, east, south, and west. The water depth will affect the loading of the support structure however, it is difficult to determine since it is so marginal.

The main observation from the case study is that the population is strongly homogeneous. Where the environmental condition has small effects on the DEM, likewise with the water depth. However, the statistical measurements as well as the small Fréchet number indicate that there is a strong correlation in the distribution of the DEM.

Population-based structural health monitoring is dependent on the idea that knowledge between structures can be transferred. Models and data can be shared directly for structures within the homogeneous population. However, in structures of a heterogeneous population, it is imperative to determine the structural similarities to determine whether it is possible to transfer such knowledge. When making a comparison of the environmental and operational effects on the population in the Wikinger wind farm this highlighted that there is a strong homogeneous correlation. The Fréchet number is low for the entire data set and the plots reinforce the judgement for the damage equivalent moments on the foundation of the structure.

6.4 Applied Digital Enabled Asset Management

This section is built up of two main works; the first is to try to tackle the situation of limited data for machine learning models, one of the bottlenecks for this type of approach. The second is a development from the case study where a variety of models are constructed to create a PBSHM model.

6.4.1 Hard-parameter Transfer

Maintaining a fully operational wind farm is a great way of de-risking the assets, but unfortunately, some failures are inherent within systems. As the popularity of wind turbines increases globally, so do the occurrence and the failures of the structure become more difficult to predict with the increasing complexity of the design. Furthermore, new technologies and optimised designs can suddenly fail due to quality or stress-related failures, respectively. This type of failure is termed infant mortality failure, causing a potentially significant loss in revenue, especially if employed in the offshore wind sector.

The purpose of this study is to tackle the issue of limited data and try to increase the performance of ML methods in the infancy of asset life. By doing so creating more robust and accurate ML models that can aid in DEAM. This was tackled by adopting multi-task learning.

Multi-task learning is when knowledge can be transferred. Inductive transfer learning is a sub-category of transfer learning. Where an accurate model can be trained on the source domain D_s to determine the hypothesis space, this article implements the hard-parameter transfer technique of inductive transfer. Such that a portion of the model is trained using the source domain D_s and the remaining portion is trained using the target domain data D_t.

In this study, there are two separate databases from two different wind turbines. The source domain consists of 31804 time steps (220 operational days), and the target domain data has a reduced length of 8141 time steps. The data has features from the SCADA system and the CMS recordings from the gearbox. This consists of aggregated 10-minute samples for each of the instances and associated errors to that data.

The hard-parameter transfer mode makes predictions of the error from the CMS recordings and a comparison is made from a conventional ML model using the target domain data only and the Hard-parameter transfer model which takes advantage of the heterogeneous nature of both the data sets.

The significant findings from this except for the thesis is that hard-parameter transfer will increase the accuracy of a ML in comparison to traditional ANN methods. In terms of outperforming the traditional method for specifically infant mortality failure, this is a diminishing advantage from the beginning of the installation. As the option continues so will the samples recorded. And as you increase the amount of samples so will the performance of the traditional ANN.

The adoption of hard-parameter transfer also can provide more insight into the system. The main body of the model was built by taking advantage of the Persons correlation of the features. Such that the relation from the SCADA to the vibrations was greater than the relation from the SCADA to the error. By doing so, and utilising the heterogeneous nature of the data. The hard-parameter transfer model trained a portion of the ANN to be accurate in predicting the vibrations. The remaining layers of the ANN were subsequently trained to determine the error. Meaning it can produce values for the vibrations as well as the error. Providing more information and increasing the confidence of the results.

6.4.2 Population Based Structural Health Monitoring

The focus of this work is based on a strong homogeneous population, with a population form of the DEM from the foundation of the jacket structure. There are four different machine learning models under investigation. Three models employ supervised domain adaptation techniques, a subcategory from the transfer learning branch, and the last model utilises ensemble learning. Transfer learning is one approach that has the main purpose of improving the performance of the model by transferring knowledge between different domains. Domain adaptation assumes that there is labelled data in the source domain that can be utilised to aid in the regression of the target domain, by mapping the two domains into a common latent space on which the data distributions are coincident. There are assumptions of domain adaptation, where the input and output feature dimensions are consistent in the source and target domain. This means that Structure One must have the same features as Structure Two in the wind farm. The former method of using an ensemble technique aims to improve the final prediction by grouping the views from the regression models and taking the consensus.

Based on the case study the three structures can be considered as a homogeneous population, as they are structurally similar in their representation and the material and geometry parameters can be described by a uni-modal distribution with low deviation. The SHM problem presented here is due to small deviations in the DEM which arise from the location and the operational context of the individual wind turbines within the wind farm.

The model development of this project involved determining an ANN that could be used for the remaining models since all models use the same architecture. The three domain adaptation techniques either, alter the data sets in for feature-based domain adaptation. The instance-based technique creates multiple instances of the ANN and uses boosting techniques to determine the optimal model. The last model is parameter-based, where the weights of the neural network are optimised using both the source and target domain with an.

Three different metrics are used to asses the model, the coefficient of determination, the accumulative error and the mean absolute error. For the first model, where the ANN is trained using one structure and is tested on the following two. As expected the results are in favour of the source turbine used to train the algorithm, results are located in Table 5.4. The performance of transferring the knowledge is poor in comparison to the domain adaptation techniques. The domain adaptation techniques all create improved general classifiers. For this problem, the champion was the RegTransferANN with the most accurate model in all three metrics when using WT1 and WT2 as the source and target domain respectively.

Chapter 7

Conclusions and Recommendations

7.1 Review of Research Objectives

The primary purpose of this research was to develop DEAM methodologies required for asset management of future offshore wind turbines through the use of data from offshore structures. DEAM is a multidisciplinary, with requiring an understanding of data structures, complex machine learning algorithms and structural mechanics of offshore structures. The components of this research include a comprehensive literature review of DEAM techniques, challenging standards to introduce standardisation on pre-processing of ML models, a solution to the challenge of limited data sets, a case study and implementation of a low-cost monitoring campaign. The outcome of this research is to support DEAM in offshore wind by developing machine learning algorithms such that information can be shared in a wind farm reducing the amount of instrumentation for a low-cost monitoring campaign, and also challenging the difficulties of implementing data-based monitoring.

7.1.1 Literature Review of DEAM

A systematic review of how monitoring for maintenance can be conducted, detailing specific measures that can be taken for condition-based maintenance. This article has addressed the types of machine learning, providing a review of models suited to condition maintenance for offshore wind turbines. Bringing both components together is a trend across all industries, the offshore wind industry is catching up to the oil sector or aerospace. There are consistencies across all sectors; there are no design principles for implementing these tools and there is no consensus on the best-suited machine learning pipelines.

There are; limited studies incorporating the vibration analysis coupled with SCADA data to determine faults. The vast majority of monitoring and maintenance only performs caustic analysis. This journal paper has discussed in detail how machine learning can be used to detect faults with a variety of approaches. Another powerful tool of machine learning is that specific faults can be categorised into alarms for immediate action to be taken.

A common issue when implementing data-driven approaches is the prepossessing of the data. Since there are no specific standards on how this is carried out, everyone producing a ML model may have dissimilar results when using the same architecture. This could be a result of many different approaches to anomaly detection, the variation in the opinion of experts, or the magnitude of dimensional reduction among some possibilities.

Lastly, current methods usually require experts to implement and understand the results of the monitoring systems. Exploring the automaton of condition monitoring systems could remove the complexity of the results decreasing the ambiguity on the condition of the asset.

7.1.2 Standardisation and Pre-processing

Determining the power using a trending method for a wind turbine is challenging with the stochastic nature of the wind and the operational states defined by the wind turbine. By creating a global model, this paper has presented a method of trend condition monitoring that uses METMAST data as input and power as output data. 5 machine learning techniques; ANN, K-NN, SVM, LGBM, and GPR have experienced improvements using the standardisation process with the accuracy and variance improved dramatically for this data set.

The work in this paper has focused on trending condition monitoring, and the techniques implemented on the standardisation of SCADA data. This work will apply to machine learning methods within the wind turbine SCADA data realm.

This paper has highlighted that there is an inconsistency in the application of pre-processing SCADA data for operational wind turbines. This paper addresses this issue, the process should follow the predetermined process outlined in section 3.2 and this should be documented. This paper has highlighted a method from beginning to end that should be implemented as it will improve the accuracy of your results. However I do believe that this should not be the final form, the steps taken are necessary but the techniques in each process are interchangeable. The purpose of this procedure and study is to highlight the need for this process in standards so that the wind turbine industry has consistent data pre-processing, and the focus can move on other aspects, such as model development, removing doubt by highlighting the effective procedure. Collaborative machine learning model development since there will be a standardised base for all model development. And lastly, transparency in model evaluation since the procedure of how the data is organised will be less ambiguous.

7.1.3 Address the issue of limited data sets

This paper has successfully highlighted how MTL accelerates the accuracy of data-driven condition monitoring of an ANN with limited data. This is a novel approach to offshore wind energy but is consistent with other areas where this methodology is implemented. The main observation is that infant mortality failure can be quickly detected, and scheduled maintenance can be planned.

Inspired by studies in infant mortality prediction, where leveraging large datasets with machine learning (ML) proved successful, this work implements two different cost functions in a wind turbine health monitoring model. Compared to traditional methods, this approach effectively extracts relevant features and performs classification with reduced noise and overfitting. However, one limitation observed is that the model's effectiveness might decrease over time. Similar to how infant mortality risk declines as the asset ages, the probability of component failure in a wind turbine also decreases with longer operation till the end of life. With increasing operational time, the model might require adjustments or retraining due to the changing data distribution and lower overall failure rates.

One of the most significant properties of the multi-task learning method is the increased amount of usable information. You have the vibration and error information both of which can be used in tandem to authenticate the diction process. One detects patterns for anomaly detection and another makes time series predictions. This is opposed to the classical method which has only one output, the error evaluation, which is useful but ideally more information is advisable in maintenance.

A progression from this concept would be to: investigate the timeframe in which MTL is most advisable for maximum accuracy of a condition monitoring system during the infant mortality period. Secondly, comparing varying types of MTL models to determine a baseline model for continuity in the industry.

7.1.4 Conduct a case study on the limitations of PBSHM for a low-cost monitoring campaign

Population-based structural health monitoring is dependent on the idea that knowledge between structures can be transferred. Models and data can be shared directly for structures within the homogeneous population. However, in structures of a heterogeneous population, it is imperative to determine the structural similarities to determine whether it is possible to transfer such knowledge.

To compare the degree of similarity within a population, the elements can first be converted into an irreducible element model, where the abstractions from the structures determine the degree of transferability. With matching properties, this strengthens the consistency of the labels required for transfer learning. In the situation where you have matching topology, operational conditions, and geometry this gives matching location labels. Matching the material properties gives consistent damage classifications or assessment labels. Hence, documenting the similarity of all four aspects is important in determining whether the transfer of knowledge is possible.

In the context of the wind turbine foundation DEM by introducing the Fréchet number one can mathematically quantify the similarity of the population form, but this must be accompanied by an observation to determine the degree of homogeneity. In this study the different operational cases of the wind turbine experience different rates of damage at different wind speeds. Moreover, the environmental effects from different wind directions, combined with the turbulence intensity influence, have a large impact on the degree of similarity of the DEM and need to be taken into consideration.

Out with the monitoring campaign, three challenges that affect the individual structural integrity of offshore wind turbines have been modelled using finite element analysis and it is concluded that these have a minimal effect on the structural integrity. Hence, one could only assume that these will have a minimal effect up to the design guidelines. Exceedance over the lifespan of the wind turbine may lead to larger deviations and model updating may be necessary in determining if these events, at the later parts of the operation, need to be accounted for.

When making a comparison of the environmental and operational effects on the population in the wind farm this highlighted that there is a strong homogeneous correlation. The Fréchet number is low for the entire dataset and the plots reinforce the judgment for the damage equivalent moments on the foundation of the structure.

7.1.5 Develop a time-series PBSHM model and compare the effectiveness of several models to determine the most suitable method for this task

Knowledge transfer is an important process in PBSHM. The benefit of transferring knowledge about structural health from one structure to another within the population is imperative to the progress of low-cost digital-enabled asset management for WTGs. It is important when applying this technique that a process is conducted to determine what similarities exist within the population so that negative transfer can be avoided. In this text two categories of structures have been discussed: homogeneous and heterogeneous.

Implementing transfer learning in the form of domain adaptation has been demonstrated to effectively mitigate problems where both the features and label spaces are consistent. This paper has demonstrated that domain adaptation applies to homogeneous populations where there are small deviations in the geometry due to the water depth, manufacturing tolerances and sensor placement.

If a higher accuracy of the model is required to determine the remaining fatigue life, further optimization measures can be taken: These would include. Separating the general model into discrete models for operational modes; Discreet model development based on the wind direction; Further studies on the feature selection and hyperparameter tuning; Implementation of high-frequency SCADA data for higher order statistics. The first two measures may lead to the development of specific models and increase bias within the estimation. As such a less accurate but more general model may produce greater estimations of unforeseen events.

7.2 Contribution to knowledge

This section will delve into the four core objectives of this thesis, listing their novelty, scientific soundness, and the diverse stakeholders they encompass with the associated value.

7.2.1 Objective 1

Conduct a comprehensive literature review of offshore wind DEAM techniques and provide references. This includes the SHM techniques and the machine learning models used.

Novelty

Conducting a thorough analysis of currently available research data regarding the reliability, availability, and maintenance of cutting-edge turbines to critically examine the gaps in maintenance procedures implemented in offshore wind engineering. Exploring the vulnerability of specific components and technologies to key factors, while assessing the structural integrity of offshore wind turbines. Providing valuable insights for the evaluation and validation of innovative alternatives in the future development of wind turbines, thereby serving as a highly informative resource for the qualification of next-generation models.

Scientific Soundness

Conducting a comprehensive systematic literature review that involves a meticulous examination of existing literature and a thorough comparison with relevant studies to provide a detailed and comprehensive analysis. Implementing a transparent review process to ensure the integrity and reliability of the research, while presenting the collected data in a clear and accessible manner

Value and Stakeholders

This objective provides a substantial amount of disclosed information at a high level of detail, making it valuable for the community of researchers and scientists who are initially addressing the topic of operations and maintenance in offshore wind structures. Offering significant insights and practical guidance for operators in terms of new alternatives and next-generation wind turbines and developing effective maintenance strategies.

7.2.2 Objective 2

Begin the discussion on why pre-processing should be standardised in offshore wind and why it should be more transparent in academia.

Novelty

Initiating a discussion on standardising pre-processing in offshore wind emphasises the importance of consistent and comparable research results. Establishing standardised protocols and methodologies enhances the reliability and credibility of findings in this field. Advocating for increased transparency in offshore wind research promotes open and accessible practices. Transparent reporting of methods, data sources, and analysis techniques fosters collaboration, reproducibility, and knowledge sharing among researchers, advancing the field.

Scientific Soundness

Utilising advanced supervised learning algorithms to address detection problems, employing state-of-the-art techniques. Implementing a transparent process for the preparation and analysis of data, ensuring openness and clarity throughout the data handling and evaluation stages.

Value and Stakeholders

Researchers and scientists actively involved in the field of offshore wind energy, preprocessing techniques, and data analysis are key stakeholders. Regulatory agencies and organisations responsible for overseeing offshore wind projects and ensuring compliance with industry standards.

7.2.3 Objective 3

Provide a solution to address the issue of limited datasets. This is conducted by developing a time-series model that makes predictions of the state of the wind turbine using transfer learning. The model is compared to a traditional machine learning method to highlight the improvements made.

Novelty

There are currently no papers on hard perimeter transfer learning applied to offshore wind on Science Direct. This highlights the novelty of this paper, trying to improve infant mortality failures of a wind turbine using this machine learning method. Considering that the accuracy of machine learning methods suffers with limited data this is exactly where this type of model is advantageous.

This algorithm has been applied to other areas such as financial technology, cloud modelling and image processing. However, I do believe that this is a novel idea applying it to this specific area using different features. This model is advantageous for wind turbines where SCADA data is limited, when this model is trained it improves the biases with the transfer of learning and reduces the overfitting of data.

Scientific Soundness

The process will be detailed in depth with visualisations of the neural network structure such that it can be easily implemented. The theory in which the models are built will be accurately detailed for the unilateral understanding of the mathematics. Verification is carried out in the pre-processing stage with the data split. The confidence interval of the model can be quantified using the error metrics discussed.

Value and Stakeholders

This paper provides value to; Researchers and scientists in the field of SHM & PBSHM. Professionals in structural engineering are interested in advancements in monitoring approaches. Stakeholders responsible for managing structures seeking effective monitoring strategies.

7.2.4 Objective 4

Develop a time-series PBSHM model and compare the effectiveness of several models to determine the most suitable method for this task.

Novelty

Comparison of multiple models to determine the most suitable method for the task involving a systematic evaluation and comparison of different modelling techniques, improving the accuracy and reliability of low-cost structural health monitoring. campaign. Developing a time-series PBSHM model incorporates time-dependent data, allowing for a low-cost monitoring solution.

Scientific Soundness

Implementation of state-of-the-art supervised learning algorithms for detection problems. Use of state-of-the-art transfer learning algorithms PBSHM, and is a transparent process in the preparation.

Value and Stakeholders

This provides value to researchers and scientists in the field of SHM & PBSHM. Professionals in structural engineering are interested in advancements in monitoring approaches. Stakeholders are responsible for managing structures and seeking effective monitoring strategies. Organisations are responsible for establishing guidelines and standards for structural health monitoring.

7.3 Future Work

To follow on from these works with a low-cost maintenance perspective, with a focus on DEAM capabilities. Machine learning can be applied to PBSHM in a way that expands the utility of the sharing of information.

The machine learning-based expansion of PBSHM can be developed further by incorporating heterogeneous data. This can involve a situation where an offshore wind farm expands its capacity and implements a different structure. By using a combination of the pre-existing structure and some of the new structure one could develop a general model that could make predictions on the target for the pre-existing or the older structure.

Extending the methods into other areas of the structure, the PBSHM model was only specifically used to determine the damage equivalent loads on the foundation of a jacket structure. However, a wind turbine is made up of multiple critical components that need constant monitoring. The methods used here can be used to extend the knowledge sharing on the entire wind farm by producing general models that can infer information from other critical components, such as damage equivalent moment on the tower, or the transition piece.

One of the exciting areas of machine learning is that new types of models are continually being developed. New optimisation techniques can be used to train artificial neural networks. Implementing the newly developed algorithms for time-series forecasting, hard-parameter transfer, domain adaptation and in general supervised learning. This could test the robustness of the models described in this thesis.

Bibliography

[1] Anastasia ioannou, andrew angus, feargal brennan, (2017), risk-based methods for sustainable energy system planning: A review, renewable and sustainable energy reviews, volume 74, pages 602-615, issn 1364-0321, https://doi.org/10.1016/j.rser.2017.02.082. (http://www.sciencedirect.com/science/article/pii/s136403211730312x).

[2] Keras. `https://keras.io`.

[3] Bs iso 13373-1:2002 condition monitoring and diagnostics of machines— thermography — part 1: General procedures, 2002.

[4] Iec 61400-25-6:2011 wind turbines - communications for monitoring and control of wind power plants - logical node classes and data classes for condition monitoring, 2011. Cited on pages 9, 146.

[5] Guideline for the certification of condition monitoring systems for wind turbines, 2013. Rules and Guidelines, IV: Industrial Services; 4.

[6] Structural health monitoring; a machine learning perspective. *Reference Research Book News*, 28, 2013.

[7] Bsee offshore wind recommendations - guidlines for structural health monitoring for offshore wind turbine towers and foundations. 2017.

[8] U.s. department of energy. 2018 offshore wind technologies market report. 2018.

[9] Norwegian energy partners. 2021.

[10] Charu C. Aggarwal and Philip S. Yu. Outlier detection for high dimensional data. *SIGMOD Record*, 30:37–46, 2001.

[11] Airbus. Skywise digital solutions, 2020.

[12] Stefan Anderlik, Reinhard Stumptner, Bernhard Freudenthaler, and Martin Fritz. A proposal for ontology-based integration of heterogeneous decision support systems for structural health monitoring. In *iiWAS*, 2010.

[13] Estefania Artigao, Sergio Martín-Martínez, Andrés Honrubia-Escribano, and Emilio Gómez-Lázaro. Wind turbine reliability: A comprehensive review towards effective condition monitoring development. *Applied Energy*, 228:1569–1583, 2018.

[14] Aamer Bilal Asghar and Xiaodong Liu. Adaptive neuro-fuzzy algorithm to estimate effective wind speed and optimal rotor speed for variable-speed wind turbine. *Neurocomputing*, 272:495–504, 2018.

[15] ASTM. Standard practices for cycle counting in fatigue analysis. *ASTM*, 2017.

[16] D. Astolfi. A study of the impact of pitch misalignment on wind turbine performance. *Machines*, 7(1), 2019.

[17] Martin Bach-Andersen, Bo Rømer-Odgaard, and Ole Winther. Deep learning for automated drive train fault detection. *Wind Energy*, 21:29–41, 2018.

[18] P. Bangalore and L.B. Tjernberg. An artificial neural network approach for early fault detection of gearbox bearings. *IEEE Transactions on Smart Grid*, 6(2):980–987, 2015.

[19] Xingxian Bao, Tongxuan Fan, Chen Shi, and Guanlan Yang. One-dimensional convolutional neural network for damage detection of jacket-type offshore platforms. *Ocean Engineering*, 219:108293, 2021.

[20] R. Bi, C. Zhou, and D.M. Hepburn. Detection and classification of faults in pitch-regulated wind turbine generators using normal behaviour models based on performance curves. *Renewable Energy*, 105:674–688, 2017.

[21] Innes Murdo Black, Mark Richmond, and Athanasios Kolios. Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management. *International Journal of Sustainable Energy*, 40:923–946, 2021.

[22] Innes Murdo Black, Mark Richmond, and Athanasios Kolios. Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management. *International Journal of Sustainable Energy*, 3 2021.

[23] A. B. Borchersen and M. Kinnaert. Model-based fault detection for generator cooling system in wind turbines using scada data. *Wind Energy*, 19(4):593–606, 2016.

[24] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. ACM Press, 1992.

[25] O. M. Bouzid, G. Y. Tian, K. Cumanan, and D. Moore. Structural health monitoring of wind turbine blades: Acoustic source localization using wireless sensor networks. *Journal of Sensors*, 2015, 2015.

[26] L A Bull, P A Gardner, J Gosliga, T J Rogers, N Dervilis, E J Cross, E Papatheou, A E Maguire, C Campos, and K Worden. Foundations of population-based shm, part i: Homogeneous populations and forms. *Mechanical systems and signal processing*, 148:107141, 2021.

[27] J. P. Burman and M. C. Otto. Census bureau research project: Outliers in time series. (CENSUS/SRD/RR-88114), 1988.

[28] M. Cao, Y. Qiu, Y. Feng, H. Wang, and D. Li. Study of wind turbine fault diagnosis based on unscented kalman filter and scada data. *Energies*, 9(10), 2016.

[29] Innes Murdo Black Moritz Werther Hackell Debora Cevasco Carolin Sophie Wenedelborn, Simon Sielder. Final report on best practice guidelines for

future wf structural condition monitoring using low-cost monitoring. *Horizon 2020*, 2022.

[30] S. Catmull. Self-organising map based condition monitoring of wind turbines. In *EWEA Annual Conference 2011*, 2011.

[31] Simon Catmull. Self-organising map based condition monitoring of wind turbines. 6 2011.

[32] D Cevasco, M Collu, and Z Lin. Om cost-based fmeca: Identification and ranking of the most critical components for 2-4 mw geared offshore wind turbines. volume 1102, 2018. cited By 6.

[33] D Cevasco, S Koukoura, and A J Kolios. Reliability, availability, maintainability data review for the identification of trends in offshore wind energy applications. *Renewable and Sustainable Energy Reviews*, 136:110414, 2021.

[34] D Cevasco, J Tautz-Weinert, U Smolka, and A Kolios. Feasibility of machine learning algorithms for classifying damaged offshore jacket structures using scada data. *Journal of Physics: Conference Series*, 1669:012021, 10 2020.

[35] Ciprian Chelba and Alex Acero. Adaptation of maximum entropy capitalizer: Little data can help a lo. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Languag Processing*, pages 285–292, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[36] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[37] Wanqiu Chen, Yingning Qiu, Yanhui Feng, Ye Li, and Andrew Kusiak. Diagnosis of wind turbine faults with transfer learning algorithms. *Renewable Energy*, 163:2053–2067, 2021.

[38] Giuseppina Ciulla, Antonino D'Amico, V. Di Dio, and Valerio Lo Brano. Modelling and analysis of real-world wind turbine power curves: Assessing deviations from nominal curve by neural networks. *Renewable Energy*, 140:477–492, 2019.

[39] Caitlyn E. Clark and Bryony DuPont. Reliability-based design optimization in offshore renewable energy systems. *Renewable and Sustainable Energy Reviews*, 97:390–400, 2018.

[40] Corinna Cortes, Vladimir Vapnik, and Lorenza Saitta. Support-vector networks editor. *Machine Leaming*, 20:273–297, 1995.

[41] C. J. Crabtree, D. Zappalá, and P. J. Tavner. Survey of commercially available condition monitoring systems for wind turbines. Technical report, Durham University School of Engineering, Computing Sciences and the SUPERGEN Wind Energy Technologies Consortium, 2014.

[42] Christopher J Crabtree, Donatella Zappalá, and Simon I Hogg. Wind energy: Uk experiences and offshore operational challenges.

[43] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, UK, 2000.

[44] Aldo da Rosa. *Wind Energy*. Elsevier, 1 2013.

[45] Wenyuan Dai, Qiang Yang, Gui-Rong Xue, and Yong Yu. Boosting for transfer learning. In *ICML '07*, 2007.

[46] Antoine de Mathelin, François Deheeger, Guillaume Richard, Mathilde Mougeot, and Nicolas Vayatis. Adapt: Awesome domain adaptation python toolbox. *arXiv preprint arXiv:2107.03049*, 2021.

[47] E.P. de Moura, F.E. de Abreu Melo Junior, F.F. Rocha Damasceno, L.C. Campos Figueiredo, C.F. de Andrade, M.S. de Almeida, and P. Alexandre Costa Rocha. Classification of imbalance levels in a scaled wind turbine through

detrended fluctuation analysis of vibration signals. *Renewable Energy*, 96:993–1002, 2016.

[48] N. Dervilis, M. Choi, S.G. Taylor, R.J. Barthorpe, G. Park, C.R. Farrar, and K. Worden. On damage diagnosis for a wind turbine blade using pattern recognition. *Journal of Sound and Vibration*, 333(6):1833–1850, 2014.

[49] Reinhard Diestel, Alexander Schrijver, and Paul D Seymour. *Graph Theory*. 2007.

[50] M.A. Djeziri, S. Benmoussa, and R. Sanchez. Hybrid method for remaining useful life prediction in wind turbine systems. *Renewable Energy*, 116:173–187, 2018.

[51] Sattar Dorafshan, Robert J. Thomas, and Marc Maguire. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Construction and Building Materials*, 2018.

[52] Richard O. Duda, Peter E. Hart, and David G. Stork. Pattern classification. 2002.

[53] Wind Energy. Wind energy in europe 2020 statistics and the outlook for 2021-2025. 2020.

[54] Peter Enevoldsen and George Xydis. Examining the trends of 35 years growth of key wind turbine components. *Energy for Sustainable Development*, 50:18–26, 2019.

[55] Guolin Ke et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 2017.

[56] Charles R. Farrar and Keith Worden. An introduction to structural health monitoring. *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, 365:303–315, 2007.

[57] Z. Feng and M. Liang. Fault diagnosis of wind turbine planetary gearbox under nonstationary conditions via adaptive optimal kernel time-frequency analysis. *Renewable Energy*, 66:468–477, 2014.

[58] Zhipeng Feng, Sifeng Qin, and Ming Liang. Time–frequency analysis based on vold-kalman filter and higher order energy separation for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. *Renewable Energy*, 85:45–56, 2016.

[59] Refaul Ferdous, Faisal Khan, Rehan Sadiq, Paul Amyotte, and Brian Veitch. Handling and updating uncertain information in bow-tie analysis. *Journal of Loss Prevention in the Process Industries*, 25:8–19, 2012.

[60] R.M. Fernandez-Canti, J. Blesa, S. Tornil-Sin, and V. Puig. Fault detection and isolation for a wind turbine benchmark using a mixed bayesian/set-membership approach. *Annual Reviews in Control*, 40:59–69, 2015.

[61] D. Fernández-Francos, D. Martínez-Rego, O. Fontenta-Romero, and A. Alonso-Betanzos. Automatic bearing fault diagnosis based on one-class -svr. *Computers & Industrial Engineering*, 64(1):357–365, 2013.

[62] A. J. Fox. Outliers in time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(3):350–363, 1972.

[63] Magda Friedjungova and Filipe Joaquim Bernardino Jorge Quix Christoph Jirina, Marcel. An overview of transfer learning focused on asymmetric heterogeneous approaches. In *Data Management Technologies and Applications*, Cham, 2018. Springer International Publishing.

[64] Yuqing Gao and Khalid M. Mosalam. Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civil and Infrastructure Engineering*, 33(9):748–768, 2018.

[65] M. C. Garcia, M. A. Sanz-Bobi, and J. del Pico. Simap: Intelligent system for predictive maintenance application to the health condition monitoring of a wind turbine gearbox. *Computers in Industry*, 57:552–568, 2006.

[66] P Gardner, L A Bull, J Gosliga, N Dervilis, and K Worden. Foundations of population-based shm, part iii: Heterogeneous populations – mapping and transfer. *Mechanical systems and signal processing*, 149:107142, 2021.

[67] P. Gardner, L.A. Bull, N. Dervilis, and K. Worden. On the application of kernelised bayesian transfer learning to population-based structural health monitoring. *Mechanical Systems and Signal Processing*, 167:108519, 2022.

[68] B. Geropp. Envelope analysis - a signal analysis technique for early detection and isolation of machine faults. *IFAC Proceedings Volumes*, 30(18):977–981, 1997. Cited on page 9.

[69] Raoof Gholami and Nikoo Fakhari. Support vector machine: Principles, parameters, and applications. pages 515–535, 2017.

[70] Jamie L. Godwin and Peter Matthews. Classification and detection of wind turbine pitch faults through scada data analysis. *International Journal of Prognostics and Health Management*, 4:1–11, 2013.

[71] Jacob Goldberger, Geoffrey E Hinton, Sam Roweis, and Russ R Salakhutdinov. Neighbourhood components analysis. *Advances in neural information processing systems*, 17:513–520, 2004.

[72] E. Gonzalez, M. Reder, and J.J. Melero. Scada alarms processing for wind turbine component failure detection. *Journal of Physics: Conference Series*, 2016.

[73] F González-Longatt, P Wall, and V Terzija. Wake effect in wind farm performance: Steady-state and dynamic behavior. *Renewable Energy*, 39:329–338, 2012.

[74] Joel Gordon. *Mechanisms of growth in Offshore wind power: Feasibility of 2030 targets in the North Seas Energy Cooperation*. PhD thesis, 06 2020.

[75] J Gosliga, P A Gardner, L A Bull, N Dervilis, and K Worden. Foundations of population-based shm, part ii: Heterogeneous populations – graphs, networks, and communities. *Mechanical systems and signal processing*, 148:107144, 2021.

[76] Lin Gui, Ruifeng Xu, Qin Lu, Jiachen Du, and Yu Zhou. Negative transfer detection in transductive transfer learning. *International journal of machine learning and cybernetics*, 9:185–197, 2017.

[77] J. M. Ha, H. Oh, J. Park, and B. D. Youn. Classification of operating conditions of wind turbines for a class-wise condition monitoring strategy. *Renewable Energy*, 103:594–605, 2017.

[78] M.W. Hackell, R. Rolfes, M.B. Kane, and J.P. Lynch. Three-tier modular structural health monitoring framework using environmental and operational condition clustering for data normalization: Validation on an operational wind turbine system. *Proceedings of the IEEE*, 104(8):1632–1646, 2016.

[79] Takuo Hamaguchi, Hidekazu Oiwa, Masashi Shimbo, and Yuji Matsumoto. Knowledge base completion with out-of-knowledge-base entities: A graph neural network approach. *Transactions of the Japanese Society for Artificial Intelligence*, 33, 2018.

[80] J. Hang, J. Zhang, and M. Cheng. Application of multi-class fuzzy support vector regression classifier for fault diagnosis of wind turbine. *Fuzzy Sets and Systems*, 297:128–140, 2016.

[81] Jouni Hartikainen and Simo Särkkä. Kalman filtering and smoothing solutions to temporal gaussian process regression models. In *2010 IEEE International Workshop on Machine Learning for Signal Processing*, pages 379–384. IEEE, 2010.

[82] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. pages 1026–1034, 2015.

[83] Donald O. Hebb. The organization of behavior: A neuropsychological theory. 1949.

[84] Justin Heinermann and Oliver Kramer. Machine learning ensembles for wind power prediction. *Renewable Energy*, 89:671–679, 2016.

[85] J. Herp, N.L. Pedersen, and E.S. Nadimi. Wind turbine performance analysis based on multivariate higher order moments and bayesian classifiers. *Control Engineering Practice*, 49:204–211, 2016.

[86] Jianglin Huang, Yan Fu Li, and Min Xie. An empirical analysis of data preprocessing for machine learning-based software cost estimation. volume 67, pages 108–127. Elsevier, 11 2015.

[87] Liang Huang, Xu Feng, Cheng Zhang, Liping Qian, and Yuan Wu. Deep reinforcement learning-based joint task offloading and bandwidth allocation for multi-user mobile edge computing. *Digital Communications and Networks*, 5:10–17, 2019. Artificial Intelligence for Future Wireless Communications and Networking.

[88] Jan Häfele, Clemens Hübler, Cristian Guillermo Gebhardt, and Raimund Rolfes. A comprehensive fatigue load set reduction study for offshore wind turbines with jacket substructures. *Renewable Energy*, 118:99–112, 2018.

[89] IEC. 61400-11 acoustic noise measurement techniques. 2012.

[90] IEC. 61869-2 instrument transformers - part 2: Additional requirements for current transformers. 2012.

[91] IEC. 61400-13 wind turbines. measurement of mechanical loads. 2016.

[92] IEC. 61400-12-1 power performance measurements of electricity producing wind turbines. 2017.

[93] Athanasios Kolios" "Innes Murdo Black, Debora Cevasco. Deep neural network hard parameter multi-task learning for condition monitoring of an offshore wind turbine. In *Torque 2022*, 2022.

[94] International Electro-technical Commission. *Wind turbines - Part 12-2: Power performance of electricity-producing wind turbines based on nacelle anemometry*, September 2013. Cited on pages 32, 39, 44, 46, 48, 49, 74.

[95] Faras Jamil, Timothy Verstraeten, Ann Nowé, Cédric Peeters, and Jan Helsen. A deep boosted transfer learning method for wind turbine gearbox fault detection. *Renewable Energy*, 197:331–341, 2022.

[96] I. Jlassi, J.O. Estima, S. Khojet El Khil, N. Mrabet Bellaaj, and A.J. Marques Cardoso. Multiple open-circuit faults diagnosis in back-to-back converters of pmsg drives for wind turbine systems. *IEEE Transactions on Power Electronics*, 30(5):2689–2702, 2015.

[97] 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. *ISA transactions*, 67:160–172, 2017.

[98] Martin Junginger, Atse Louwen, Nilo Gomez Tuya, David de Jager, Ernst van Zuijlen, and Michael Taylor. Chapter 7 - offshore wind energy. In Martin Junginger and Atse Louwen, editors, *Technological Learning in the Transition to a Low-Carbon Energy System*, pages 103–117. Academic Press, 2020.

[99] Pearson K. Notes on regression and inheritance in man. *Nature*, 30(773):240–242, 1894.

[100] J. Kang, X. Zhang, and T. Jin. Tracking gearbox degradation based on stable distribution parameters: A case study. In *2015 IEEE Conference on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application*, 2015.

[101] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, 2 edition, 2004.

[102] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. volume 30. Curran Associates, Inc., 2017.

[103] K. Kim, G. Parthasarathy, O. Uluyol, W. Foslien, S. Sheng, and P. Fleming. Use of scada data for failure detection in wind turbines. In *ASME 5th International Conference on Energy Sustainability*, 2011. Available: [9](https://research-hub.nrel.gov/en/publications/use-of-scada-data-for-failure-detection-in-wind-turbines-2).

[104] Diederik P Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. 2015.

[105] W. Kirch. *Pearson's Correlation Coefficient.* Springer, Dordrecht, 2008.

[106] Sofia Koukoura, James Carroll, and Alasdair McDonald. Wind turbine intelligent gear fault identification. In *Annual Conference of the PHM Society*, pages 1–7, St. Petersburg, United States, 2017.

[107] Nikolaos Koutsouleris, Eva M. Meisenzahl, Stefan Borgwardt, Anita Riecher-Rössler, Thomas Frodl, Joseph Kambeitz, et al. Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain*, 138(7):2059–2073, 2015.

[108] K. Leahy, R.L. Hu, I.C. Konstantakopoulos, C.J. Spanos, and A.M. Agogino. Diagnosing wind turbine faults using machine learning techniques applied to operational data. In *2016 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2016.

[109] S. Lee and S. Lee. Numerical and experimental study of aerodynamic noise by a small wind turbine. *Renewable Energy*, 65:108–112, 2014.

[110] Aitong Li and Yuan Xu. The governance for offshore wind in japan. *Energy Procedia*, 158:297–301, 2019. Innovative Solutions for Energy Transitions.

[111] Ren Li, Tianjin Mo, Jianxi Yang, Shixin Jiang, Tong Li, and Yiming Liu. Ontologies-based domain knowledge modeling and heterogeneous sensor data integration for bridge health monitoring systems. *IEEE Transactions on Industrial Informatics*, 17:321–332, 2021.

[112] Y. Li, S. Liu, and L. Shu. Wind turbine fault diagnosis based on gaussian process classifiers applied to operational data. *Renewable Energy*, 134:357–366, 2019.

[113] Yanting Li, Wenbo Jiang, Guangyao Zhang, and Lianjie Shu. Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data. *Renewable Energy*, 171:103–115, 2021.

[114] Zehui Lin, Shengli Li, Dong Ni, Yimei Liao, Huaxuan Wen, Jie Du, Siping Chen, Tianfu Wang, and Baiying Lei. Multi-task learning for quality assessment of fetal head ultrasound images. *Medical Image Analysis*, 58:101548, 2019.

[115] X. Liu, M. Li, S. Qin, X. Ma, and W. Wang. A predictive fault diagnose method of wind turbine based on k-means clustering and neural networks. *Journal of Internet Technology*, 17(7):1521–1528, 2016.

[116] Nick Lord. Matrix computations, 3rd edition. *The Mathematical Gazette*, 83(498):556–557, 1999.

[117] John Martin and Doug Ramsey. The economics of wind energy. *Journal of Applied Corporate Finance*, 21(2):100–109, 2009.

[118] Maria Martinez Luengo, Athanasios Kolios, and Lin Wang. Structural health monitoring of offshore wind turbines: A review through the statistical pattern recognition paradigm. *Renewable and Sustainable Energy Reviews*, 64:91–105, 10 2016.

[119] M. Martinez-Ramon and C. Cristodoulou. Support vector machines for antenna array processing and electromagnetics. page 126, 2006.

[120] P. Mazidi, M. Du, L. Bertling Tjernberg, and M.A.S. Bobi. A performance and maintenance evaluation framework for wind turbines. In *2016 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2016 - Proceedings*, 2016.

[121] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[122] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[123] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London*, 209:415–446, 1909.

[124] Andrew D Mills, Dev Millstein, Seongeun Jeong, Luke Lavin, Ryan Wiser, and Mark Bolinger. Estimating the value of offshore wind along the united states' eastern coast. *Environmental Research Letters*, 13(9):094013, sep 2018.

[125] T. M. Mitchell. Machine learning. 45(37):870–877, 1997.

[126] A.A. Mohamed. An effective dimension reduction algorithm for clustering arabic text. *Egyptian Informatics Journal*, 2019.

[127] Kambiz Mokhtari, Jun Ren, Charles Roberts, and Jin Wang. Application of a generic bow-tie based risk analysis framework on risk management of sea ports and offshore terminals. *Journal of Hazardous Materials*, 192:465–475, 2011.

[128] I. Myrtveit, E. Stensrud, and U.H. Olsson. Analyzing data sets with missing data: An empirical evaluation of imputation methods and likelihood-based methods. *IEEE Transactions on Software Engineering*, 27:999–1013, 2001.

[129] T. Möller, D. Langenkämper, and T. Nattkemper. Wind turbine segmentation performing knn-clustering on superpixel segmentations. In *Proceedings of SPIE - The International Society for Optical Engineering*, 2019.

[130] A. R. Nejad, P. F. Odgaard, Z. Gao, and T. Moan. A prognostic method for fault detection in wind turbine drivetrains. *Engineering Failure Analysis*, 42:324–336, 2014.

[131] E. Neu, F. Janser, A.A. Khatibi, and A.C. Orifici. Fully automated operational modal analysis using multi-stage clustering. *Mechanical Systems and Signal Processing*, 84:308–323, 2017.

[132] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104:11 – 33, Jan-01-2016 2016.

[133] A. O'Hagan. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 40(1):1–24, Sep 1978.

[134] Emilio Soria Olivas, José David Martín Guerrero, Marcelino Martinez-Sober, Jose Rafael Magdalena-Benedito, and Antonio José Serrano López, editors. *Handbook of Research on Machine Learning Applications and Trends*. IGI Global, 2010.

[135] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[136] Ravi Kumar Pandit, David Infield, and Athanasios Kolios. Gaussian process power curve models incorporating wind turbine operational variables. *Energy Reports*, 6, 11 2020.

[137] David Pardoe and Peter Stone. Boosting for regression transfer. In *ICML*, 2010.

[138] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988.

[139] Morten D. Pedersen and Thor I. Fossen. Efficient nonlinear wind-turbine modeling for control applications. In *IFAC Proceedings Volumes*, volume 45, pages 264–269, 2012.

[140] Morten D. Pedersen and Thor I. Fossen. Efficient nonlinear wind-turbine modeling for control applications. *IFAC Proceedings Volumes*, 45(2):264–269, 2012.

[141] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[142] Chen-Huan Pi, Kai-Chun Hu, Stone Cheng, and I-Chen Wu. Low-level autonomous control and tracking of quad rotor using reinforcement learning. *Control Engineering Practice*, 95:104222, 2020.

[143] A Pliego Marugán and FP García Márquez. Advanced analytics for detection and diagnosis of false alarms and faults: A real case study. *Wind Energy*, 22:1622–1635, 2019.

[144] L J Prendergast, C Reale, and K Gavin. Probabilistic examination of the change in eigenfrequencies of an offshore wind turbine under progressive scour incorporating soil spatial variability. *Marine Structures*, 57:87–104, 2018.

[145] W. Qiao and D. Lu. A survey on wind turbine condition monitoring and fault diagnosis - part ii: Signals and signal processing methods. *IEEE Transactions on Industrial Electronics*, 62(10):6546–6557, 2015.

[146] Y. Qiu, W. Zhang, M. Cao, Y. Feng, and D. Infield. An electro-thermal analysis of a variable-speed doubly-fed induction generator in a wind turbine. *Energies*, 8(5):3386–3402, 2015.

[147] Yingning Qiu, Yanhui Feng, and David Infield. Fault diagnosis of wind turbine with scada alarms based multidimensional information processing method. *Renewable Energy*, 145:1923–1931, 2020.

[148] Spectra Quest. Rotating machinery faulty diagnosis techniques - envelope and cep-strum analyses. Technical report, Spectra Quest, Inc., Richmond, VA, USA, 2006.

[149] Ramboll. *ROSAP-Ramboll Offshore Structural Analysis Package, Version 53.* Ramboll Esbjerg, Denmark, 2018.

[150] V. Rao and H. Rao. Neural network and fuzzy logic. 1998.

[151] C. E. Rasmussen and C. K. I. Williams. *Gaussian Process For Machine Learning.* The MIT Press, 2006.

[152] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[153] T. Regan, C. Beale, and M. Inalpolat. Wind turbine blade damage detection using supervised machine learning algorithms. *Journal of Vibration and Acoustics, Transactions of the ASME*, 139(6), 2017.

[154] T. Regan, R. Canturk, E. Slavkovsky, C. Niezrecki, and M. Inalpolat. Wind turbine blade damage detection using various machine learning algorithms. In *Proceedings of the ASME Design Engineering Technical Conference*, 2016.

[155] M Richmond, A Sobey, R Pandit, and A Kolios. Stochastic assessment of aerodynamics within offshore wind farms based on machine-learning. *Renewable Energy*, 161:650–661, 2020.

[156] Mark Richmond, Ursula Smolka, and Athanasios Kolios. Feasibility for damage identification in offshore wind jacket structures through monitoring of global structural dynamics. *Energies*, 13:5791, 4 2020.

[157] B. Rocher, F. Schoefs, M. François, and A. Salou. Bayesian updating of probabilistic time-dependent fatigue model: Application to jacket foundations of wind turbines. In *7th European Workshop on Structural Health Monitoring, EWSHM 2014 - 2nd European Conference of the Prognostics and Health Management (PHM) Society*, page 1264, 2014.

[158] Rafael V. Rodrigues and Corinne Lengsfeld. Development of a computational system to improve wind farm layout, part i: Model validation and near wake analysis. *Energies*, 12(5), 2019.

[159] T.J. Rogers, P. Gardner, N. Dervilis, K. Worden, A.E. Maguire, E. Papatheou, and E.J. Cross. Probabilistic modelling of wind turbine power curves with application of heteroscedastic gaussian process regression. *Renewable Energy*, 148:1124–1136, 2020.

[160] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning internal representations by error propagation. 1:318–362, 1986.

[161] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323, 10 1986.

[162] M'hammed Sahnoun, P. Godsiff, David Baudry, Anne Louis, and B. Mazari. Modelling of maintenance strategy of offshore wind farms based multi-agent system, 2014. Journal of Intelligent Manufacturing, doi: 10.1007/s10845-015-1171-0.

[163] L. Saidi, J. Ben Ali, M. Benbouzid, and E. Bechhofer. An integrated wind turbine failures prognostic approach implementing kalman smoother with confidence bounds. *Applied Acoustics*, 138:199–208, 2018.

[164] Ester Sales-Setién and Ignacio Peñarrocha-Alós. Robust estimation and diagnosis of wind turbine pitch misalignments at a wind farm level. *Renewable Energy*, 146:1746–1765, 2020.

[165] R. Saravanakumar and D. Jena. Nonlinear estimation and control of wind turbine. In *2013 IEEE International Conference on Electronics, Computing and Communication Technologies*, pages 1–6, 2013.

[166] A. Sarrafi, Z. Mao, C. Niezrecki, and P. Poozesh. Vibration-based damage detection in wind turbine blades using phase-based motion estimation and motion magnification. *Journal of Sound and Vibration*, 421:300–318, 2018.

[167] Matti Niclas Scheu, Lorena Tremps, Ursula Smolka, Athanasios Kolios, and Feargal Brennan. A systematic failure mode effects and criticality analysis for offshore wind turbine systems towards integrated condition based maintenance strategies. *Ocean Engineering*, 2019.

[168] Matti Niclas Scheu, Lorena Tremps, Ursula Smolka, Athanasios Kolios, and Feargal Brennan. A systematic failure mode effects and criticality analysis for offshore wind turbine systems towards integrated condition based maintenance strategies. *Ocean Engineering*, 2019.

[169] Matti Niclas Scheu, Lorena Tremps, Ursula Smolka, Athanasios Kolios, and Feargal Brennan. A systematic failure mode effects and criticality analysis for offshore wind turbine systems towards integrated condition based maintenance strategies. *Ocean Engineering*, 176:118–133, 2019.

[170] Meik Schlechtingen, Ilmar Ferreira Santos, and Sofiane Achiche. Wind turbine condition monitoring based on scada data using normal behavior models. part 1: System description. *Applied Soft Computing*, 13:259–270, 2013.

[171] Hugo G. Schnack, Mireille Nieuwenhuis, Neeltje E. M. van Haren, Lucija Abramovic, Thomas W. Scheewe, Rachel M. Brouwer, et al. Can structural mri aid in clinical classification? a machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage*, 84:299–306, 2014.

[172] Laura Schröder, Nikolay Krasimirov Dimitrov, David Robert Verelst, and John Aasted Sørensen. Using transfer learning to build physics-informed machine learning models for improved wind farm monitoring. *Energies*, 15(2), 2022.

[173] M. Shafiee. A fuzzy analytic network process model to mitigate the risks associated with offshore wind farms. *Expert Systems with Applications*, 42(4):2143–2152, 2015.

[174] S. Sheng. Wind turbine gearbox condition monitoring round robin study - vibration analysis. Technical report, National Renewable Energy Lab. (NREL), Golden, CO, 2012. Cited on pages 9, 10, 143.

[175] S. Sheng. Report on wind turbine subsystem reliability - a survey of various databases. Technical Report NREL/PR-5000-59111, National Renewable Energy Laboratory, 2013.

[176] S. Simani. Overview of modelling and advanced control strategies for wind turbine systems. *Energies*, 8(12):13395–13418, 2015.

[177] G.A. Skrimpas, C.W. Sweeney, K.S. Marhadi, B.B. Jensen, N. Mijatovic, and J. Holbøll. Employment of kernel methods on wind turbine power performance assessment. *IEEE Transactions on Sustainable Energy*, 6(3):698–706, 2015.

[178] Abir Smiti. A critical overview of outlier detection methods. *Computer Science Review*, 38, 11 2020.

[179] A. J. Smola and B. Sch. A tutorial on support vector regression, 2003. Canberra.

[180] Dongran Song, Jian Yang, Xinyu Fan, Yao Liu, Anfeng Liu, Guo Chen, and Young Hoon Joo. Maximum power extraction for wind turbines through a novel yaw control solution using predicted wind directions. *Energy Conversion and Management*, 157:587–599, 2018.

[181] Adrian Stetco, Fateme Dinmohammadi, Xingyu Zhao, Valentin Robu, David Flynn, Mike Barnes, John Keane, and Goran Nenadic. Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*, 133:620–635, 2019.

[182] K. Strike, K. El Emam, and N. Madhavji. Software cost estimation with incomplete data. *IEEE Transactions on Software Engineering*, 27, 2001.

[183] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. *CoRR*, abs/1511.05547, 2015.

[184] R.S. Sutton and A.G. Barto. Introduction to reinforcement learning. 135, 1998.

[185] Mandar Tabib, M Salman Siddiqui, Adil Rasheed, and Trond Kvamsdal. Industrial scale turbine and associated wake development -comparison of rans based actuator line vs sliding mesh interface vs multiple reference frame method. *Energy Procedia*, 137:487–496, 2017.

[186] Jannis Tautz-Weinert and Simon J Watson. Using scada data for wind turbine condition monitoring - a review. *IET Renewable Power Generation*, 11:382–394, 2016.

[187] P. J. Tavner. *Offshore Wind Turbines: Reliability, availability and maintenance.* Institution of Engineering and Technology, 2012.

[188] Peter Tavner. Offshore wind turbines: Reliability, availability and maintenance. 2012.

[189] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb. Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies*, 7(4):2595–2630, 2014.

[190] W. Teng, X. Ding, X. Zhang, Y. Liu, and Z. Ma. Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform. *Renewable Energy*, 93:591–598, 2016.

[191] W. Teng, X. Wang, C. Dang, and W. Shi. Dynamic equivalent model of the wind farm based on the k nearest neighbor. In *Asia-Pacific Power and Energy Engineering Conference, APPEEC*, 2016.

[192] Wei Teng, Hao Cheng, Xian Ding, Yibing Liu, Zhiyong Ma, and Haihua Mu. Dnn-based approach for fault detection in a direct drive wind turbine. *IET Renewable Power Generation*, 12, 7 2018.

[193] Hanany Tolba, Nouha Dkhili, Julien Nou, Julien Eynard, Stéphane Thil, and Stéphane Grieu. Ghi forecasting using gaussian process regression: kernel study. *IFAC-PapersOnLine*, 52:455–460, 2019.

[194] O Troyanskaya, M Cantor, G Sherlock, P Brown, T Hastie, R Tibshirani, D Botstein, and R B Altman. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17:520–525, 2001. cited By 2271.

[195] George Tsialiamanis, D. Wagg, I. Antoniadou, and K. Worden. An ontological approach to structural health monitoring. 10 2020.

[196] Bhekisipho Twala and Michelle Cartwright. Ensemble missing data techniques for software effort prediction. *Intelligent Data Analysis*, 14, 5 2010.

[197] V. Vapnik. Statistical learning theory. 1998.

[198] K. Verbert, B. De Schutter, and R. Babuška. Timely condition-based maintenance planning for multi-component systems. *Reliability Engineering and System Safety*, 159:310–321, 3 2017.

[199] Cyril Voyant, Gilles Notton, Soteris Kalogirou, Marie-Laure Nivet, Christophe Paoli, Fabrice Motte, and Alexis Fouilloy. Machine learning methods for solar radiation forecasting: a review. *Renewable Energy*, 105:569–582, 2017.

[200] H. Wang, H. Wang, G. Jiang, J. Li, and Y. Wang. Early fault detection of wind turbines based on operational condition clustering and optimized deep belief network modeling. *Energies*, 12(6), 2019.

[201] Lipo Wang. Support vector machines: Theory and applications. 2005.

[202] Q. Wang, G. Michau, and O. Fink. Domain adaptive transfer learning for fault diagnosis. pages 279–285, 2019.

[203] Y. Wang, D.G. Infield, B. Stephen, and S.J. Galloway. Copula-based model for wind turbine power curve outlier rejection. *Wind Energy*, 17(11):1677–1688, 2014.

[204] Y. Wang, Y. Liu, L. Li, D. Infield, and S. Han. Short-term wind power forecasting based on clustering pre-calculated cfd method. *Energies*, 11(4), 2018.

[205] Zihao Wang, Hongwei Lin, Xiaofeng Yu, and Yusuf Fatihu Hamza. A dimensional reduction guiding deep learning architecture for 3d shape retrieval. *Computers & Graphics*, 81:82–91, 2019.

[206] Paul J. Werbos. Beyond regression: new tools for prediction and analysis in the behavioral sciences. 1974.

[207] T. Werner, O. Bichler, E. Vianello, and B. Yvert. Modulation device and method, artificial synapse comprising said modulation device, short term plasticity method in an artificial neural network comprising said artificial synapse. 2019.

[208] E. Wiggelinkhuizen, T. Verbruggen, H. Braam, L. Rademakers, J. Xiang, and S. J. Watson. Assessment of condition monitoring techniques for offshore wind farms. *Journal of Solar Energy Engineering*, 130(3):031004–1–031004–9, 2008.

[209] Edwin Wiggelinkhuizen, Theo Verbruggen, Henk Braam, Luc Rademakers, Jianping Xiang, and Simon J. Watson. Assessment of condition monitoring techniques for offshore wind farms. *Journal of Solar Energy Engineering*, 130(3), 2008.

[210] M. Wilkinson, K. Harman, T. van Delft, and B. Darnell. Comparison of methods for wind turbine condition monitoring with scada data. *IET Renewable Power Generation*, 8(4):390–397, 2014.

[211] Michael Wilkinson, Brian Darnell, Thomas Delft, and Keir Harman. Comparison of methods for wind turbine condition monitoring with scada data. *IET Renewable Power Generation*, 8, 5 2014.

[212] WindEurope. Wind energy in europe: Outlook to 2023. 2019.

[213] Wang Xiaohua, Peng Muzi, Pan Lijuan, Hu Min, Jin Chunhua, and Ren Fuji. Two-level attention with two-stage multi-task learning for facial emotion recognition. *Journal of Visual Communication and Image Representation*, 62:217–225, 2019.

[214] C. Yang, J. Liu, Y. Zeng, and G. Xie. Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model. *Renewable Energy*, 133:433–441, 2019.

[215] H. Yang and M. Huang. Integrating auto-associative neural networks with hotelling t2 control charts for wind turbine fault detection. *Energies*, 8(10):12100–12115, 2015.

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

[217] W. Yang, P. J. Tavner, C. J. Crabtree, Y. Feng, and Y. Qiu. Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, 17(5):673–693, 2014.

[218] Wenxian Yang, Peter J Tavner, Christopher J Crabtree, Y Feng, and Y Qiu. Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, 17:673–693, 2014.

[219] Wenxian Yang, P.J. Tavner, Christopher Crabtree, Yanhui Feng, and Yingning Qiu. Wind turbine condition monitoring: Technical and commercial challenges. *Wind Energy*, 17, 05 2014.

[220] Xiuxing Yin and Xiaowei Zhao. Big data driven multi-objective predictions for offshore wind farm based on machine learning algorithms. *Energy*, 186:115704, 2019.

[221] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, and Z. Zhang. Lstm-efg for wind power forecasting based on sequential correlation features. *Future Generation Computer Systems*, 93:33–42, 2019.

[222] Cheng-Dar Yue, Che-Chih Liu, Chien-Cheng Tu, and Ta-Hui Lin. Prediction of power generation by offshore wind farms using multiple data sources. *Energies*, 12(4), 2019.

[223] F. Zemene and V. Khedkar. Survey on machine learning based electric consumption forecasting using smart meter data. *International Journal of Computer Applications*, 180(6):46–52, 2017.

[224] X. Zhang, L. Sun, H. Sun, Q. Guo, and X. Bai. Floating offshore wind turbine reliability analysis based on system grading and dynamic fta. *Journal of Wind Engineering and Industrial Aerodynamics*, 154:21–33, 2016.

[225] Yu Zhang and Qiang Yang. An overview of multi-task learning. *National Science Review*, 5(1):30–43, 2018.

[226] Z. Zhang, A. Verma, and A. Kusiak. Fault analysis and condition monitoring of the wind turbine gearbox. *IEEE Transactions on Energy Conversion*, 27(2):526–535, 2012.

[227] H. Zhao, Y. Gao, H. Liu, and L. Li. Fault diagnosis of wind turbine bearing based on stochastic subspace identification and multi-kernel support vector regression. *Journal of Modern Power Systems and Clean Energy*, 7(2):350–356, 2019.

[228] H. Zhao, H. Liu, W. Hu, and X. Yan. Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renewable Energy*, 127:825–834, 2018.

[229] H. Zhao, Q. Wu, Q. Guo, H. Sun, and Y. Xue. Distributed model predictive control of a wind farm for optimal active power control part i: Clustering-based wind turbine model linearization. *IEEE Transactions on Sustainable Energy*, 6(3):831–839, 2015. Available: [5](https://orbit.dtu.dk/files/107595979/DMPC$_{P}art_{22}0150226_{h}zhao_{q}w_{f}inal.pdf$).

[230] H. Zhao, Q. Wu, Q. Guo, H. Sun, and Y. Xue. Distributed model predictive control of a wind farm for optimal active power control part ii: Implementation with clustering-based piece-wise affine wind turbine model. *IEEE Transactions on Sustainable Energy*, 6(3):840–849, 2015. Available: [5](https://orbit.dtu.dk/files/107595979/DMPC$_{P}art_{22}0150226_{h}zhao_{q}w_{f}inal.pdf$).

[231] P. Zhou and P.T. Yin. An opportunistic condition-based maintenance strategy for offshore wind farm based on predictive analytics. *Renewable and Sustainable Energy Reviews*, 109:1–9, 2019.

[232] Erzhou Zhu, Yuanxiang Zhang, Peng Wen, and Feng Liu. Fast and stable clustering analysis based on grid-mapping k-means algorithm and new clustering validity index. *Neurocomputing*, 363:149–170, 2019.

[233] W. J. Zhu, W. Z. Shen, E. Barlas, F. Bertagnolio, and J. N. Sørensen. Wind turbine noise generation and propagation modeling at dtu wind energy: A review. *Renewable and Sustainable Energy Reviews*, 88:133–150, 2018.

[234] A. Ziaja, I. Antoniadou, T. Barszcz, W. J. Staszewski, and K. Worden. Fault detection in rolling element bearings using wavelet-based variance analysis and novelty detection. *Journal of Vibration and Control*, 22(2):396–411, 2016.

[235] L. Ziegler, N. Cosack, A. Kolios, and M. Muskulus. Structural monitoring for lifetime extension of offshore wind monopiles: Verification of strain-based load extrapolation algorithm. *Marine Structures*, 66:154–163, 2019.