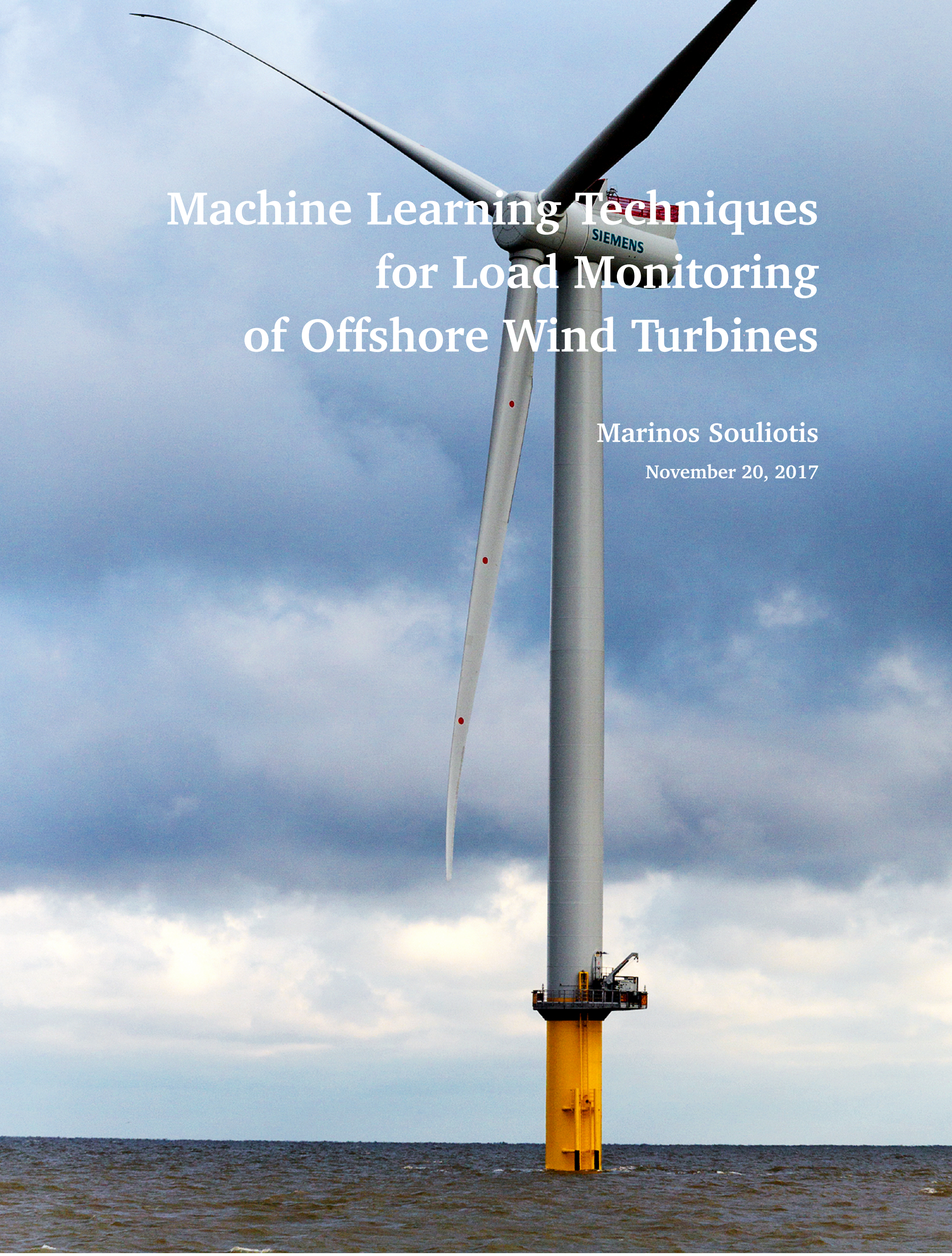


Machine Learning Techniques for Load Monitoring of Offshore Wind Turbines

Marinos Souliotis

November 20, 2017

TU Delft Delft University of Technology

SIEMENS Gamesa RENEWABLE ENERGY

Machine Learning Techniques for Load Monitoring of Offshore Wind Turbines

Master of Science Thesis

MSc Offshore and Dredging Engineering

Author:	Marinos Souliotis	
Committee:	Prof.dr. Andrei Metrikine	TU Delft - Chairman
	Dr. Eliz-Mari Lourens	TU Delft
	Dr.ir. Karel van Dalen	TU Delft
	Ir. Dominik Fallais	TU Delft
	Ir. Vasilis Michalopoulos	Siemens Gamesa Renewable Energy
	Ir. Hendrik Kramers	Siemens Gamesa Renewable Energy

November 20, 2017

- CONFIDENTIAL -

Summary

In the race for cost reduction in the offshore wind industry, support structure optimisation leading to weight reduction plays a prominent role. The fatigue limit state (FLS) is often the driving consideration for the support structure design. However, once the turbines are constructed and operational no information about the actual in situ loading is collected. The load component that is examined in this work is the bending moment of the monopile at the mudline section, an unfavourable location for sensor installation and maintenance. Monitoring the monopile loads can offer an accurate knowledge of its consumed and remaining fatigue lifetime, which in turn has potential benefits of lifetime extension, feedback on current design practices or can provide room for implementation of more aggressive turbine operation modes if remaining lifetime is identified.

Within an offshore wind farm not all turbines are subjected to the same loading. Several factors like different geometrical characteristics of turbines, wind turbulence and wake effects, soil conditions, hydrodynamic loading and idling conditions induce variation to the individual loading of the turbines. For this reason, a turbine or cluster-specific internal load monitoring scheme is investigated in this thesis. In order to make such a scheme efficient for a large number of turbines, data-driven approaches are examined and specifically artificial neural networks (ANNs) and linear regression (LR).

The purpose of this thesis is to examine whether it is possible to estimate the actual loading of an offshore wind turbine monopile at the seabed critical location by utilizing measured turbine signals and/or sensor measurements and machine learning techniques instead of a structural model.

The data used in this research is simulated with the in-house aeroelastic code of Siemens Gamesa Renewable Energy – Bonus Horizontal axis wind turbine Code (BHawC) – and includes operational, as well as idling fatigue load cases. 10-Minute statistics of signals that are collected in turbines are used as inputs to the data-driven models, whereas 10-minute damage equivalent loads (DELs) of monopile bending moments are used as targets. In the case where the training set properly reflects

the entirety of variation in environmental conditions, promising results are achieved. A proper training set can be obtained by using simulation results from a validated BHawC model.

The results of this study indicate that it is possible to accurately estimate the loading history of the monopile in the form of bending moment DELs using feedforward ANNs with a single hidden layer. The mean absolute error is $< 3.5\%$ for both the fore-aft (FA) and side-side (SS) directions for normal operation load cases and the residual error is $< 0.5\%$ for both directions when the inputs to the ANNs are suitable statistics of the standard signals that are measured in wind turbines. In the idling load cases a mean absolute estimation error $< 3.5\%$ in both directions is achieved using only statistics of nacelle accelerations. In both cases, a training data set that properly encapsulates the variety of environmental conditions is used for the ANNs. Additionally, sensitivities of load estimation with ANNs with respect to training set size, eigenfrequency deviation and input signal noise are investigated.

A different type of sensors, inclinometers, when installed at tower bottom (TB) offer a highly accurate estimation of monopile loads at the seabed. TB rotation angle is highly correlated with the target load. 10-minute equivalent ranges of rotation signals at TB can be used as inputs to a piecewise LR scheme in order to accurately estimate DELs. When the 10-minute intervals are additionally binned based on whether the mean wind speed is below, around or above the turbine rated wind speed, this method gives mean absolute errors smaller than 2.5%. As a result, TB inclinometers are recommended as highly useful sensors for purposes of load monitoring. However, factors such as the actual accuracy level and durability of such sensors compared to alternatives should be additionally considered.

The data-driven methods examined in this thesis show potential not only for online internal load monitoring, but also for fast estimation of past load histories using recorded data. Once the model has been trained and tested, making new estimations with the trained model is computationally highly efficient. Moreover, more than one wind turbines of the same cluster can be possibly monitored with a single trained ANN if there is not a substantial difference in their properties, like natural frequencies. Lastly, an important assumption when using the data-driven models that are examined in this work is that the turbine properties remain unchanged between training and estimation phases, e.g. very limited impact of scouring.

Acknowledgements

I would like to thank Eliz-Mari Lourens for trusting me with this interesting and innovative topic and the support throughout this thesis. Dick and Vasilis for their coaching during these nine months in the Siemens Wind Power department in the Hague and all the Siemens colleagues for their help and the nice moments. Professor Andrei Metrikine for giving me directions in critical parts of this work and his motivational teaching and Dominik Fallais for his detailed and valuable feedback. I would also like to thank Karel van Dalen and Federico Pisano for being in my committee and reviewing my final report.

I am grateful to my family and Theodora for their support during the last two years. Last but not least, I want to thank all the friends I made in the Netherlands for the good times.

Contents

List of Figures

List of Tables

Nomenclature

Latin symbols

b_j^h	bias of hidden neuron j	$-$
b_k^o	bias of output neuron k	$-$
E	network error vector	$-$
f	frequency	Hz
F_s	sampling frequency	Hz
H_s	significant wave height	m
J	Jacobian matrix	$-$
L	objective function	$-$
m	exponent for the calculation of equivalent load cycles	$-$
M_{ml}^{FA}	monopile bending moment at mudline in fore-aft	kNm
M_{ml}^{SS}	monopile bending moment at mudline in side-side	kNm
N	Hessian matrix approximation with damping terms	$-$
N_{ref}	number of reference cycles for DEL calculation	$-$
P_{el}	electrical power output	MW
R	Pearson correlation coefficient	$-$
r	number of hidden layer neurons	$-$
R_{max}	maximum range of signal	$-$
s	number of network outputs	$-$
S_{XX}	power spectrum of time signal coefficient	$-$
T	vector of target values	$-$
T_R	return period of 10-minute simulation	$days$
U^{FA}	wind speed fore-aft component	m/s
U^{SS}	wind speed side-side component	m/s
V	output vector of hidden layer	$-$
var	variance	$-$
V_b	output vector of hidden layer including unit column	$-$

v_j	output of hidden neuron j	$-$
W^h	input-hidden layer weight matrix	$-$
w^h_{ij}	weight from input i to hidden layer neuron j	$-$
w^o_{jk}	weight from hidden neuron j to output k	$-$
W^o	hidden-output layer weight matrix	$-$
X_b	network input vector including unit column	$-$
X_i	feature array i	$-$
$X_{i,stand.}$	standardised feature array i	$-$
$\ddot{x}_{na}$	side-side nacelle acceleration	m/s^2
$\ddot{x}_{tb}$	side-side tower bottom acceleration	m/s^2
Y	output vector of network	$-$
y_k	output k	$-$
$\ddot{y}_{na}$	fore-aft nacelle acceleration	m/s^2
$\ddot{y}_{tb}$	fore-aft tower bottom acceleration	m/s^2
Z	input vector of hidden layer	$-$
z_j	input of hidden neuron j	$-$

Greek symbols

δ_w	network weight update vector	$-$
ΔS^m_{eq}	10-min equivalent range	$-$
$\epsilon_{mu,abs}$	mean absolute normalised error of test set	$-$
ϵ_{mu}	mean normalised error of whole set	$-$
θ^p_b	blade pitch angle	rad
θ^{na}_p	nacelle pitch angle	rad
θ^{na}_r	nacelle roll angle	rad
λ_n	spectral moment of order n and cyclic frequency f	$-$
μ	mean value	$-$
σ	standard deviation	$-$
ψ	neuron activation function	$-$
Ω_{gen}	generator rotational speed	rad/s

Abbreviations

ANN	Artificial Neural Network
BHawC	Bonus Horizontal axis wind turbine Code
DEL	Damage Equivalent Load
DLC	Design Load Case
FA	Fore-Aft

FE	Finite Element
FLS	Fatigue Limit State
LM	Levenberg - Marquardt
LR	Linear Regression
LS	Least Squares
MBM	Mudline Bending Moment
ML	Machine Learning
NN	Neural Network
OWF	Offshore Wind Farm
OWT	Offshore Wind Turbine
PDD	Probability Density Distribution
PDF	Probability Density Function
PLC	Programmable Logic Controller
PSD	Power Spectral Density
SCADA	Supervisory Control And Data Acquisition
SS	Side-Side
SSE	Sum of Squared Errors
TB	Tower Bottom
TF	Transfer Function
TP	Transition Piece

Chapter 1

Introduction

1.1 Motivation

New designs for offshore wind turbines (OWTs) are continuously transitioning towards higher power and larger size in order to more efficiently harvest the vast wind resources that offshore sites offer. Although the size of new offshore wind farms is increasing, higher turbine power means less OWTs necessary to achieve the desired park power/production, contributing to a decreased cost of electricity (fewer monopiles, less installation costs, etc.). This increase in OWTs power also increases the individual turbine contribution to the total wind farm energy production. Monitoring techniques can be employed to continuously assess the condition of the turbines. One category of these techniques is aimed at monitoring the structural loads and the corresponding fatigue damage accumulation due to cyclic loading at the desired location.

Support structure optimization leading to weight reduction can be an important source of cost reduction in the offshore wind industry. The fatigue limit state (FLS) is often a driving consideration for support structure design. However, once the turbines are constructed and operational no information about the actual in situ loading is collected. Monitoring loads in fatigue sensitive zones can provide significant benefits both in making decisions for the specific monitored wind farm, as well as for assessing the accuracy of the original fatigue design. An important decision can be for example to extend the farm lifetime in case the actual loads in situ are lower than those designed for. On the contrary, if the actual loads prove to be greater than the design loads, one could aim to adjust power production in order to achieve the lowest cost of energy over the farm lifetime. A relevant method is presented by Kusiak [9]. Furthermore, if good accordance between designed and actual loads can be proved for a number of projects in the designer's portfolio, this can be beneficial for future projects by reducing a part of the technical risk and assuring better financing conditions.

However, the critical fatigue points of the structure are sometimes not easily accessible, such as the monopile section at mudline level. This makes measuring in that location economically non-attractive, especially for long term periods. Therefore, economic advantages could result from the development of a remote sensing technique for load monitoring. Such a technique will use measurements at other points of the structure, preferably signals that are already collected in wind turbines (standard signals, such as electrical power output, rotor rotational velocity and nacelle accelerations) to estimate the target loads (e.g. monopile overturning moment at the mudline).

1.2 Literature review

This section contains an overview of selected literature relevant to the topic of load monitoring for wind turbines. Two different approaches (physics-based and data-based models) to load monitoring are briefly described, a selection of existing work on the topic is presented and lastly conclusions about the strengths and weaknesses of each approach are drawn.

1.2.1 Introduction

Load monitoring techniques can be generally divided in physics-based and data-based models. These two types differ in the methodology they use to estimate the target quantities and both have advantages and limitations. The main aspects of each technique are presented in the following.

- Physics-based models

  Physics-based models use physical laws and parameters to describe physical systems and are most commonly used in engineering. One example is the modelling of solid structures using the finite element method. These models offer significant insight into the examined system, however in complex systems they require many parameters to be known, some of which are either not possible to determine or their estimation is highly uncertain. Additionally, when modelling complex systems, issues such as discretization arise. In the finite element example, in order to have an accurate model the structure has to be discretized with a proper grid. This is a part of a process called verification. Experiments are often used to confirm the accuracy of analytical or numerical models (validation process). Methods for estimating the structural response at locations different than the ones measured include e.g. modal expansion or various types of Kalman filters. A comparison between three such methods was carried out in Maes et al [13].

- Data-based models

  On the other hand, data-based models are based only on observed input and output of the system, and are thus often characterized as ”black-box” models. In other words they act as implicit transfer functions between input and response and require little or no information on the system parameters. They use principles of statistics and in some cases probabilities to model the available data and make estimations for new data. There are numerous types of data-based models, with some being more or less suitable depending on the specific problem at hand. Neural networks are a popular example of such a data-based model that has been used in various engineering tasks.

1.2.2 Selected literature

The most extensive source of knowledge in this topic, in the case of an onshore wind turbine is the work of Cosack [5]. A fatigue load monitoring system using artificial neural networks is introduced in order to estimate the loading at various critical turbine locations, such as the blade root in- and out-of-plane moments, the tower base longitudinal moment, tower top nodding moment and the rotor thrust and torque. In order to infer the actual loading of the wind turbine using a data driven method standard signals are used, such as: electrical power, generator speed, wind speed and pitch angle conventionally measured for operational control, as well as accelerations at turbine top.

For the design of a system that is used to estimate the loads mentioned in the previous paragraph, the most important disturbances that influence loads and standard signals have to be known. These include: mean wind speed, vertical wind shear, vertically inclined flow, yaw error, air density, overall turbulence intensity, mass imbalance and soil/foundation stiffness. If these disturbances influence loads, they should manifest themselves in the standard signals that are used as inputs to the model, or else the transfer function will be unable to capture this relationship. Both damage equivalent load ranges and probability density distributions (PDD) of load ranges are examined as targets for the neural network. Using PDD instead of damage equivalent load ranges allows for a more detailed fatigue load analysis, however the number of parameters that are estimated by the ANN and describe the 10-minute PDD is higher.

The investigated methods are first tested with simulations carried out with the aeroelastic code Flex5, considering the National Renewable Energy Laboratory (NREL) 1.5 MW WindPACT (Wind Partnership for Advanced Component Technologies) wind turbine. Then, the methods are validated on the basis of measured data from campaigns carried out at the prototype of the Multibrid M5000 turbine and then at two Nordex N80 turbines that are closely located to each other. The estimation methods that were developed with simulation data provided promising results in the case of the Multibrid 5000 turbine. Accurate estimations were achieved for the estimation both of equivalent loads and load range distributions of all examined loads. Small networks with a single hidden layer proved to be sufficient for an accurate estimation. A trend towards lower accuracy in estimating equivalent load ranges than equivalent load magnitudes was evident. The developed empirical approach for fitting observed load histograms with standard distributions and subsequently predicting their parameters was successfully applied with a small number of common distributions.

Next, networks trained with data acquired from one N80 turbine are used to predict loads for a neighboring turbine of the same type. For the prediction of rotor thrust and tower base moment the accuracy of the estimation did not deteriorate much. However, in the case of the rest of the loads, such as blade root moments, errors between predictions and measurements were significant. This is despite the fact that the turbines are identical and located only 600m away from each other. This shows that the actual loads that the different turbines were subjected to, or the actual turbine structures had some deviation from each other. Further investigation is necessary in order to assess the transferability of networks between turbines.

In the work of Obdam et al [18] the Flight Leader Concept for Wind Farm Load Counting, that was developed based on data from onshore wind turbines at ECN Windturbine Testpark Wieringermeer (EWTW), is applied in an offshore wind farm for the first time (OWEZ). OWEZ is the first Dutch OWF and is located near Egmond aan Zee village and consists of 36 Vestas V90 pitch-controlled variable speed turbines, with a rated power output of 3 MW. SCADA data including nacelle wind speed, rotor rotational speed, blade pitch angle, electrical power output, nacelle yaw angle and optionally wave height and direction is collected on all 36 turbines, while 2 of them are additionally equipped with mechanical load measurement devices and will act as the Flight Leaders of the wind farm. Also data from a wave buoy and a metereological mast of the farm are available. The Flight Leader Concept includes using ANN to establish the transfer function between available SCADA parameters and load indicators on blade (flapwise and edgewise) and tower (north-south and east-west) on the 2 fully instrumented "Leader" turbines. Once these transfer functions have been created, the respective SCADA data from each of the rest 34 turbines can be used as input to them in order to estimate the loading indicators for these uninstrumented turbines. In order to validate the method's performance the ANN are trained on data solely from the first instrumented turbine and subsequently SCADA data from the second one is used to estimate its load indicators, which is then compared with the actual measured loads. The performance is significantly lower than for the onshore wind farm, with the inclusion of wave characteristics as inputs to the estimation not improving the accuracy. The performance is characterized reasonable and further study is proposed using more OWEZ data. Finally, it is suggested that the concept could be potentially modified for OWF application if future studies show promising results.

In the case of OWTs there is an extra source of loading for the structure: hydrodynamic loading. Sometimes hydrodynamic loads are dominant and using only the commonly measured wind turbine signals (especially when excluding nacelle accelerations) as network inputs yields large inaccuracies. This phenomenon possibly implies lack of adequate knowledge of the hydrodynamic loading of the turbines. Smolka et al [27] examines whether sea state measurements are necessary in these cases for an accurate estimation or if they can be substituted by using extra acceleration sensors mounted on different locations on the turbine. The locations examined are the tower top, tower bottom, the foundation structure at mean sea level height or below the sea level. In this study equivalent load ranges are used to describe the fatigue loads (instead of load distributions) and a new type of input is used, namely rainflow counted equivalent ranges of some of the standard signals. Three common sub-structure types, monopile, jacket and tripod are examined in the context of this study. It is concluded that accelerometers can replace wave height measurement and even a higher estimation accuracy might be achievable.

In Smolka and Cheng [26] a database corresponding to one year of measurement data collected by an offshore AREVA M5000 turbine is used to investigate the dependency between estimation error and provided inputs in order to determine the most important signals and statistical parameters. Also, the trade-off between the cost of running longer measurement campaigns and accepting a larger estimation error is explored. This originates from the nature of neural networks, which cannot extrapolate from what has been presented to them during training. It is demonstrated that different input configurations for the neural network can lead to similar estimation accuracy. It is also shown that a minimal duration requirement for the measurement campaign can be estimated based on the sample data spread in mean wind speed and turbulence intensity.

1.2.3 Conclusions

In general the available methods for remote sensing of wind turbine loads can be divided in two main categories: physics-based and data-based. A comparison between the two model types is presented in Table 1.1. ANNs are used as the reference type of model for data-driven models.

In the context of this work only data-based models will be examined, also called machine learning (ML) techniques. The reason for choosing this type of models is that they are in general significantly simpler and faster to build and deploy than physics-based models. The reason for that is that hardly any parameters of the modelled system have to be known and the model can be built based only on observed input and output of the system (in the specific case time series). Although in physic-based models the system output would not be necessary to make new estimations, it is generally used to validate the model. For example, if a finite element (FE) model is created for the turbine structure it has to be validated that this model is accurately representing the real structure. So in this case measured accelerations can be com-

Table 1.1: Comparison between physics-based and data-driven models for load monitoring

Category	**Physics-Based**	**Data-Based (ANNs)**
Description	• Information stored in physical parameters and laws of the model • Very specific input data required for predictions • Based on time domain or linearized frequency domain models	• Measured input and load (target) data required for set-up • Implicit transfer function is stored in the algorithm
Advantages	• No measured loads required (recommended for validation) • Simultaneous load estimates at all locations of the model • Can extrapolate to any load conditions for which the model is valid	• Automated training and semi-automatic network topology optimization possible • No detailed information on the wind turbine parameters and physical laws required • Computationally efficient estimation, especially for DELs
Disadvantages	• Detailed structural information required • Suitable for observable systems only • Post-processing of time series required • Structural model might not be totally accurate	• Insight in TF not straightforward (black-box model) • Extrapolation to input not presented during training not possible • Separate network for each critical point • Trial and error approach for set-up

bined with the finite element model to estimate MBMs, but these estimations have to be compared with actual measured moments to ensure the model is accurate. As a result, in both physics- and data-based models some (limited) output measurements are necessary.

Physics-based models such as modal expansion use time-series of measured response data, such as tower accelerations and a FE model to estimate time series of other quantities, for instance displacements, moments and accelerations. in any node of the model. This ability to estimate signals of interest at multiple structure locations is an advantage of these methods. However, the form of the estimated quantity is a time series which sometimes has to be further processed. Moment time series have to be processed with the rainflow counting method to give Damage Equivalent Load (DEL) values for fatigue calculations. If such models are used for several wind turbines a huge amount of data will be produced and processed continuously, which induces significant computational efforts. On the other hand, ML techniques can be used to establish a black box relation directly from input to a specific DEL output, making the process much faster. Finally, in some ML algorithms, such as ANNs, the calculation from input to output is a fast matrix multiplication as opposed to more complex operations required when using physical models. All the aforementioned reasons make it attractive to examine ML techniques as a solution to efficiently monitoring the loading of multiple wind turbines and whole wind farms.

1.3 Research question

As was mentioned in the last paragraphs of Subsection 1.2.1, data-driven models that use ML techniques offer significant benefits when employed for the creation of a load monitoring system for multiple OWTs and entire farms. They are faster and simpler to build than the competing physics-mased models, such as modal expansion algorithms and typically require less computational effort and storage. However it is of vital importance to assess the accuracy of data-driven models to conclude whether it reaches a satisfactory level. This depends mainly on three factors:

- Choosing a suitable format for the load quantities that need to be estimated.
- Choosing inputs to the data-driven models that appropriately reflect the cause of variation in the load quantities, thus leading in an accurate estimation.
- Selecting an appropriate type of data-driven model among the possible alternatives.

These three tasks form together the main research question of this thesis which can be summarised as:

"Is it possible to accurately estimate the actual loading of an offshore wind turbine monopile at the mudline critical location by utilizing standard turbine signals and/or sensor measurements and machine learning techniques instead of a structural model?"

1.4 Methodology and structure

In this Section the structure of the thesis is presented by summarising the contents of each Chapter and Appendix.

Chapter 1 contains an introduction to the topic of load monitoring that is examined in this thesis. The motivation for this study is explained, including the potential benefits from a load monitoring system. Next, a brief overview of the existing relevant literature is carried out and some conclusions on the currently available scientific methods are drawn. The most important aspects of data-based and model-based approaches are presented and a comparison is made. Finally, a clear research question to be answered by this thesis is stated. Only data-driven approaches will be examined in the context of this thesis.

In Chapter 2 an introduction to the Siemens Wind Turbine SWT-6.0-154 OWT examined in this study is done. The targets of the estimation procedure for a load monitoring system are loads and in this study the overturning moment at the monopile mudline section will be used. These are in the form of Damage Equivalent Loads (DELs) over 10 minute intervals. The method to obtain these DELs from the moment time-series that result from the BHawC simulations is shown. Then, the turbine signals that will be used as input for the data-based models and their locations on the structure are presented. These signals are output from the BHawC simulations mentioned above. They will be pre-processed to statistical parameters, which will be the actual input to the models. The chosen parameters that are relevant and will be mostly used in the following chapters are shown.

In Chapter 3 a brief introduction to machine learning (ML) models and Arfificial Neural Networks (ANNs) in particular is done. An overview of the different types of problems where machine learning can be used is presented in Section 3.1 and the general procedure of setting up a ML model is described. In Section 3.2 information on different types of ANNs and activation functions is provided. The theoretical background behind feedforward ANNs is elaborated and the various algorithms that are available for network weight optimization are listed. Then, the Levenberg-Marquardt algorithm is presented.

In Chapter 4 ANNs are applied for the load estimation problem of monopile overturning moments in the mudline section in fore-aft (FA) and side-side (SS) directions. As load output of the ANNs a DEL accumulated in a period of 10 minutes is used. The approach is illustrated in Figure 1.1. First, in Section 4.1 information relevant to the data used in the case study is presented and the performance metrics that will be used for the assessment of the estimation accuracy are introduced. In Section 4.2 a selection of available input and target variables are plotted together in order to explore the data and discover useful correlations. Next, in Section 4.3 all available combinations of available signals are tested as input to the ANN and the performance of different combinations is presented. The main difference of offshore with respect to onshore WT is the hydrodynamic loading, mainly due to waves. The effect of significant wave height H_s on the monopile moment DELs in FA and SS directions is illustrated in Section 4.4 and the difference is attributed to the more complex structural behavior in FA direction. The inclusion of the 10-minute sigificant wave height H_s as an input to the ANNs is also examined. Lastly, in Section 4.5 ANNs are applied in order to estimate moment damage equivalent loads during idling conditions.

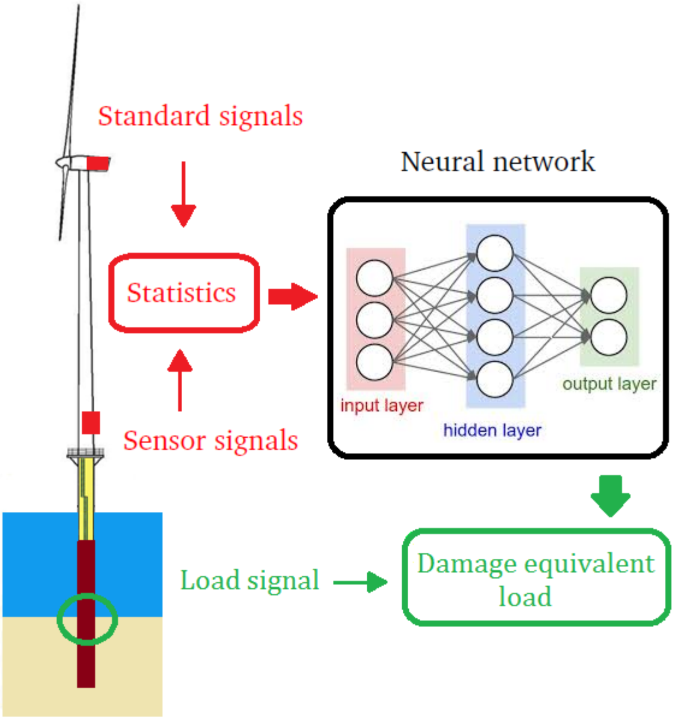


Figure 1.1: Overview of the method for load estimation with use of artificial neural networks.

Chapter 5 presents three sensitivity studies for the application of ANNs for load monitoring. A natural limitation of ANNs is that they cannot extrapolate outside the range of inputs that have been presented during training. Extrapolation introduces significant errors to the estimation. This is shown in Section 5.1, where training sets of different sizes are created based on the likelihood of occurence of the individual 10-minute simulations within the wind farm's design lifetime. Those sets of varying size are used to train the ANN models, which are then used to predict the remaining test set samples. The impact of training set size on the estimation accuracy is assessed. In Section 5.2 the sensitivity of the estimation accuracy of ANNs to deviations in the turbine's first eigenfrequency is examined. ANNs are first trained on data from a turbine with a specific eigenfrequency and are subsequently tested on four wind turbines with the same characteristics, whose eigenfrequencies deviate from the original. In Section 5.3 acceleration signals from nacelle and TB are contaminated with artificial noise. A network is trained and tested on statistics of those noisy acceleration signals and its performance is compared with the respective noiseless case.

Linear regression models are examined in Chapter 6. The linear correlation between nacelle acceleration standard deviation and target DELs can be used to attempt a supervised extrapolation within the normal operation range. A brief overview of the theory behind Linear Regression (LR) method is presented in Section 6.1. In Section 6.2 the linear relationship between the standard deviations of nacelle accelerations and the respective moment DEL is exploited to examine using a simple linear regression model instead of ANNs. In Section 6.3 the simulations of the operational load case DLC 1.2 are binned based on the 10-minute mean wind speed in 3 bins: below, around and above rated. Then separate LR models are used for each bin, in order to approximate the underlying function more locally. Finally, in Section 6.4 the rotation angle (pitch & roll) signal at tower bottom (TB) is considered and its correlation with the MBM signal is shown. When it is also entered in the linear regression scheme in the form of equivalent magnitude dS_{eq} the results exhibit a high increase in accuracy.

The most important conclusions of this study are presented in Chapter 7. In addition, some recommendations for the practical implementation of a data-driven load monitoring scheme are given. Finally, some relevant topics are recommended for future work.

Appendix A includes the method for obtaining the power density spectrum of an arbitrary time signal and the spectral moments of various orders, which contain valuable information about the signal's frequency content.

The regression plots for the eigenfrequency sensitivity study of Section 5.2 are presented in Appendix B.

Chapter 2

Fatigue Loads and Measured Signals of Wind Turbines

In this Chapter the specific wind turbine that will be used throughout this study is introduced. Information about the turbine model and its location in Westermost Rough offshore wind farm is provided in Section 2.1. The targets of the estimation procedure are internal loads of the monopile and more specifically the bending moment at the mudline section. The commonly used types of load formats are described in Section 2.2. The parameters that will be used as inputs to the ANNs are signals that are typically available through the SCADA system. Also, additional signals that can be acquired by installing extra sensors will be considered. The signals will be processed and suitable statistical parameters will be extracted in order to be used as inputs to the estimation. An overview of the signals and the statistical parameters that are considered in this work is given in Section 2.3.

2.1 The SWT-6.0-154 wind turbine

The wind turbine model examined in this study is the SWT-6.0-154. The specific wind turbine is located in the Westermost Rough OWF, approximately 8 km off the coast of Withernsea in the northeast of England. The farm location is shown in Figure 2.1.

The farm consists of 35 turbines of this type, with a total rated power of 210 MW. A brief overview of the most important characteristics of the SWT-6.0-154 model is presented in Table 2.1. An schematic representation of the turbine with the most important parts highlighted is shown in Figure 2.2.

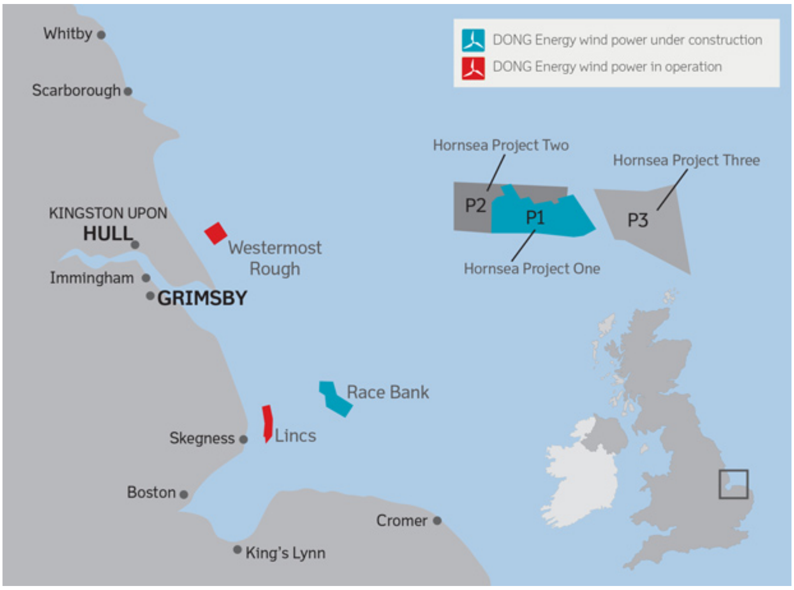


Figure 2.1: Location of Westermost Rough offshore wind farm in the UK map [29]

2.2 Fatigue load formats

Fatigue loads are of great importance for the design of the wind turbine structures and affect almost all turbine components, such as the blades, gearbox, tower and foundation. The reason for that is that they are almost constantly subject to dynamic loading from wind and in the case of offshore turbines also from waves. This results in an accumulated number of loading cycles in the order of 10^8 to 10^9 over the turbine lifetime, according to Cosack [5]. Fatigue loads can be described with several different formats, of which uncompressed load time series in terms of stress histories are the format containing the full load information. Load time series can be compressed to different load formats, however with a degree of information loss. Two such common load formats are load range distributions (or multi-stage distributions) and equivalent load ranges (also called damage equivalent loads DELs or single-stage distributions). Only the latter format is examined in this work.

2.2.1 Target load signals

Although fatigue is related to the stress time history of the examined section, in this work the overturning moment time history and the respective condensed formats will be examined (at monopile mudline level). Stresses can then be directly calculated if the internal forces and moments, as well as the section properties are known. For

Table 2.1: Characteristics of the SWT-6.0-154 offshore wind turbine

Rotor	
Type	3-bladed, horizontal axis
Diameter	154 m
Speed range	5-11 rpm
Blade	
Material	Glass Reinforced Epoxy Resin Composite (GRE)
Blade length	75 m (B75)
Generator	
Type	Synchronous, PMG, Direct Drive
Generator	
Cut-in wind speed	3-5 m/s
Nominal power at	12-14 m/s
Cut-out wind speed	25 m/s
Maximum 3 s gust	70 m/s (IEC version)
Weights (approximately)	
Towerhead mass	360,000 kg
Tower	
Type	Cylindrical and tapered tubular
Hub height	103.26 m
Monopile	
Height	60 m
Water depth	26.5 m
Diameter	6.5 m

example, the normal stress at any point of the monopile section depends on the axial force and the two bending moments around two perpendicular section axes. The examined load signals that will be used as targets for the estimation with data-driven models are shown in Table 2.2.

Table 2.2: Examined target wind turbine load signals

Number	Symbol	Target signal	Unit	Signal location
1	$\mathrm{M}_{\mathrm{ml}}^{\mathrm{FA}}$	Bending moment FA	kNm	Monopile mudline
2	$\mathrm{M}_{\mathrm{ml}}^{\mathrm{SS}}$	Bending moment SS	kNm	Monopile mudline

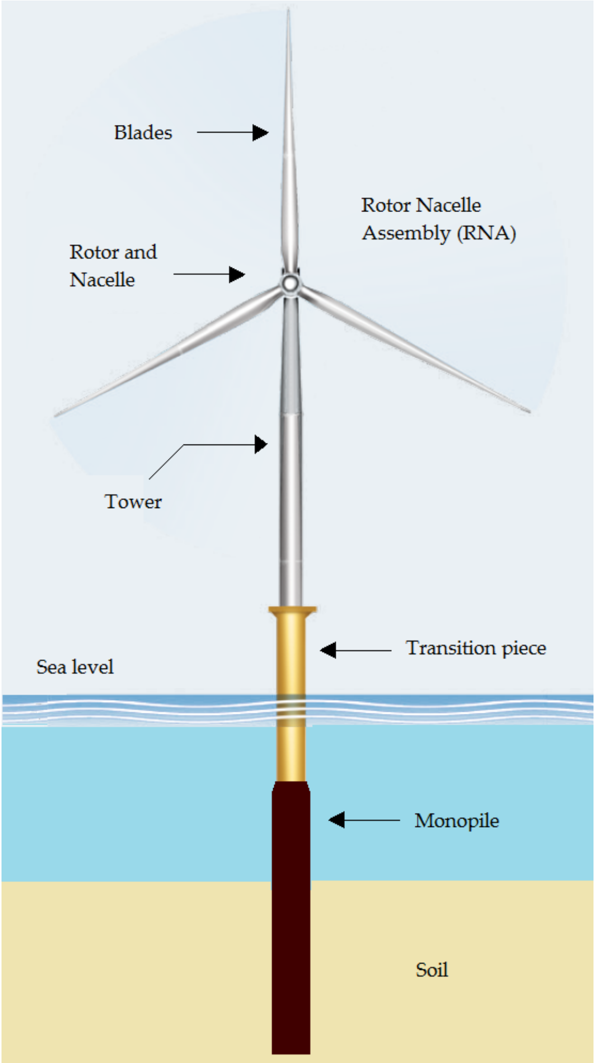


Figure 2.2: Illustration of an offshore wind turbine on a monopile foundation

During the 10-minute simulations the yaw angle of the rotor-nacelle-assembly is varying. As a result, the FA and SS directions with respect to the support structure fixed coordinates are not constant during this interval. For this reason the mean value of the yaw angle during the 10 minutes and the respective FA and SS directions of the monopile are considered. All applicable signals are transformed to this selected reference system.

2.2.2 Load range distributions

The method proposed for obtaining the load range distributions is briefly described here. First, the measured load signal is processed using the rainflow counting method, see Downing and Socie, [6]. This method returns all load cycles encountered in the time-series, which includes full cycles and possibly some half-cycles. Subsequently a histogram of the computed load cycle ranges is constructed with respect to a number of predefined load amplitude bins. The vertical axis contains the number of occurences of each bin over the sampled period and the horizontal axis the amplitude bins. From the histogram, the discrete PDF of the load cycle amplitudes can be obtained by scaling the histogram such that its total area is equal to 1.

2.2.3 Equivalent load ranges

The individual load ranges ΔS_i resulting from rainflow counting can be further processed to a single equivalent load range by applying an exponential relationship between load cycle ranges and damage caused to the specimen, assuming a linear damage accumulation (Palmgren-Miner rule). The exponent m used is a material constant and its inverse $\frac{1}{m}$ is the slope of the SN-curve for the specific material. Its typical value for welded steel is 4 and this value is used for the DEL calculation of the monopile bending moments. In the generic case where there are n load ranges ΔS_i with different cycle numbers N_i, they can be combined to a single equivalent load range $\Delta S_{eq,Nref,m}$ which for the reference number of cycles N_{ref} gives the same total damage:

$$\Delta S_{eq,Nref,m} = \sqrt[m]{\frac{\sum_{i=1}^{n} \Delta S_i^m N_i}{N_{ref}}} \tag{2.1}$$

This number describes the damage induced to the specimen over the examined time period in the simplest way possible. This simplification can lead to deviations from the theoretical damage caused by multi-stage distributions, mainly when elasto-plastic behavior occurs at large load magnitudes. Equivalent load ranges calculated for a specific number of reference cycles N_{ref1} can be transformed to a different number of cycles N_{ref2} by applying the relationship:

$$\Delta S_{eq,Nref2,m} = \Delta S_{eq,Nref1,m} \sqrt[m]{\frac{N_{ref1}}{N_{ref2}}} \tag{2.2}$$

Additionally, k equivalent ranges that are derived for the same exponent m and number of reference cycles N_{ref} can be combined in a single equivalent load.

$$\Delta S_{eq,Nref,m} = \sqrt[m]{\sum_{i=1}^{k} \Delta S^m_{eq,Nref,m,i}} \tag{2.3}$$

An example of the three load formats is given in Figures 2.3a to 2.3c, where a load time history is first processed with the rainflow counting method and reduced to a histogram of load ranges, with the x axis containing the values of the bin edges and y axis the number of occured cycles for each bin of ranges. With some further assumptions the influence of these ranges is combined in a single equivalent load range for a number of reference cycles N_{ref}, as shown in the lower right part of the figure.

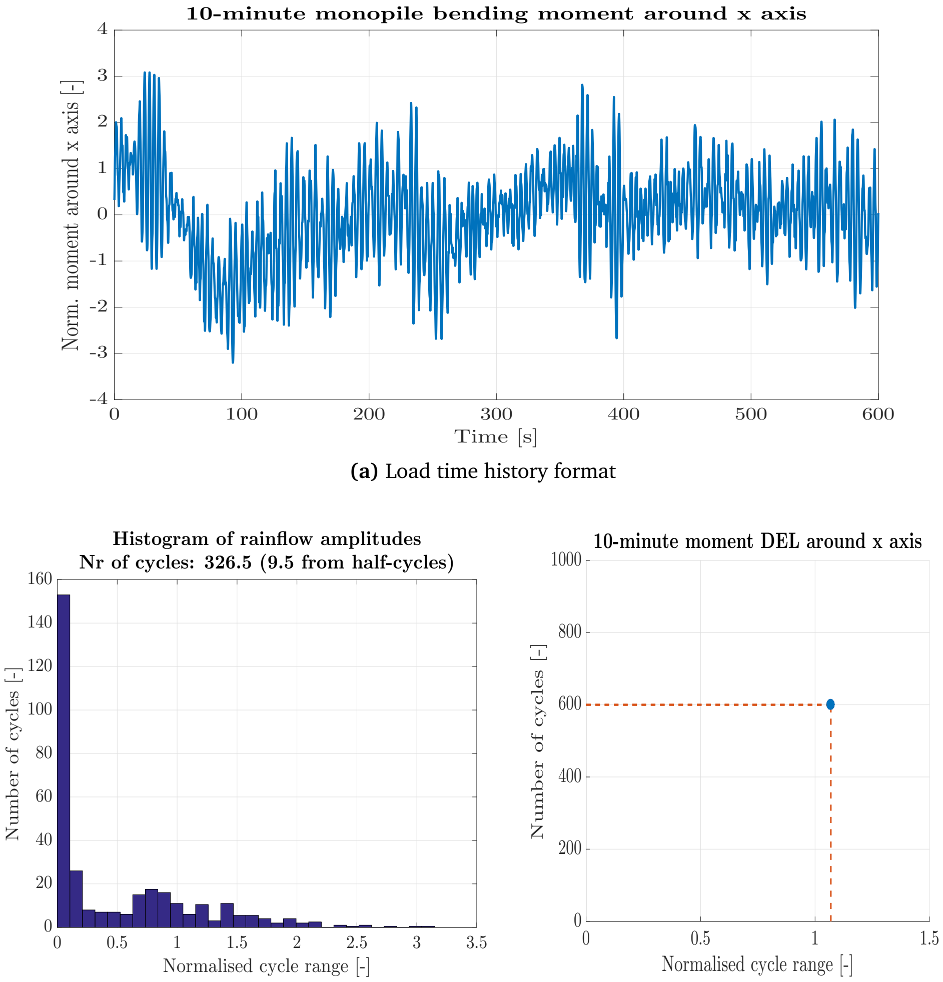
2.3 Input parameters – statistics of measured turbine signals

The input for the estimation procedure will be selected from a variety of available signals. For this thesis simulated response measurements and standard signals generated with the BHawC aeroestic software package will be used. The input signals that are examined in this study are listed in Table 2.3. For convenience a numeral identifier is assigned to each input.

Table 2.3: Measured wind turbine signals that will be examined as inputs to the data-driven models. Signals 1-5 are named ”standard signals” throughout this work and are typically collected in wind turbines.

Number	Symbol	Input signal	Unit	Signal location	Standard
1	$\mathbf{\Omega_{gen}}$	Rotational velocity	rad/s	Rotor	yes
2	$\theta^{\mathbf{P}}_{\mathbf{b}}$	Pitch angle	rad	Blade	yes
3	$\mathbf{P_{el}}$	Electrical power	MW	Nacelle	yes
4	$\mathbf{\ddot{y}_{na}}, \mathbf{\ddot{x}_{na}}$	Accelerations FA & SS	m/s^2	Nacelle	yes
5	$\mathbf{U^{FA}}, \mathbf{U^{SS}}$	Wind speed FA & SS	m/s	Nacelle	yes
6	$\mathbf{\ddot{y}_{tb}}, \mathbf{\ddot{x}_{tb}}$	Accelerations FA & SS	m/s^2	Tower bottom	no

These input signals will be processed in order to extract relevant statistical parameters that can be used as inputs to the estimation procedure. The statistical parameters, derived from time and frequency domain representations of the input and output signals, are listed in Table 2.4.



(a) Load time history format

(b) Load range histogram format

(c) DEL format

Figure 2.3: Three different load formats with decreasing level of information: load time history, load range histogram and equivalent load range.

Table 2.4: Statistical parameters of input signals

Parameter name	**Symbol**
Time domain	
mean value	μ
variance / st. deviation	var / σ
maximum range	R_{max}
10-min equivalent range	ΔS_{eq}^{m}
Frequency domain	
spectral moment of order n	λ_n
Used spectral moments	$\lambda_{-2}, \lambda_{-1}, \lambda_{-0.5}, \lambda_1, \lambda_2, \lambda_3$

The spectral moments of various orders are derived from the signal power spectral density (PSD). A detailed description of the computations of the PSD and the spectral moments of arbitrary order of the input signals is presented in Appendix A.

Chapter 3

Artificial Neural Networks

This chapter is an introduction to machine learning (ML) algorithms with specific attention given to artificial neural networks (ANNs) which are flexible function approximators that can be used for a variety of tasks. In Section 3.1 the different uses of ML are briefly presented: supervised or unsupervised learning and regression or classification. Additionally, the procedure of choosing a ML model and optimizing its parameters to fit specific data is illustrated. In Section 3.2 the mathematical model of a feedforward ANN is provided and the Levenberg-Marquardt algorithm for the network weight optimization is presented. Feedforward ANNs will be used to establish data-driven models for load monitoring in Chapter 4.

3.1 Machine learning models

The term machine learning is attributed to the subfield of computer science whose objective is to explore the study and construction of algorithms that can learn from and make predictions on data, rather than being explicitly programmed for a specific task. Such algorithms, do not follow strictly static program instructions but make data-driven predictions or decisions by utilizing a number of input and output observations of the underlying system that is to be modelled.

3.1.1 Applications of machine learning models

Machine learning is particularly useful in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible. One widely known example that is used as an introductory problem in ANNs is the task of recognizing hand-written digits (e.g. MNIST dataset). It is a task which has proven too complex to program in an explicit way. However, using an ANN with multiple hidden layers that have learnt by examples have demonstrated excellent performance with accuracy higher than 99% in predicting the correct digit, Liu [11]. The ANN for this specific case has learnt to predict correct digits using training samples of vectors with all the pixel values of the digit picture (input) and the respective digit label (target). Digit recognition is one example of supervised learning tasks.

Supervised learning is the process of inferring an underlying function from available labeled data. Each labeled data sample is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). For instance in the task of inferring the function which relates the location, area and age of a house to the price of the house, the labeled data should include an input vector consisting of three values: $X = [location, area, age]$ and an output value which is a single scalar in this case: $y = price$. These X and y values are instances of three independent variables for X and a dependent variable for y. Each variable can be continuous, discrete or categorical. Categorical variables are binarised, so for example the employment status with 2 potential cases (employed and unemployed) would be described by 2 separate binary variables (one for employment and one for unemployment). An employed person would correspond to $[1, 0]$ and an unemployed person to $[0, 1]$.

Supervised learning tasks can be further divided into regression and classification tasks; this distinction is made based on the type of the dependent target variable y. A task is referred to as a regression task if the target variable y is a continous numeric variable. One example is the problem of estimating house prices, since the price of a house can take values in a continuous range of real numbers. On the other hand, when the task is to assign a sample in one of several available categories the process is called classification. For example the problem of distinguishing between ”spam” and ”non-spam” email is a classification task. The binarised format described in the previous paragraph is used for the target value in this case: 0 for ”non-spam” and 1 for ”spam”.

In unsupervised learning the samples are unlabeled, which means they consist only of input and no corresponding label. The target is to explore the data and discover a structure in it such as hidden patterns or groupings. Cluster analysis is a common unsupervised learning task.

3.1.2 Creating a machine learning model

The process of setting up and using a machine learning model requires several steps. The four main steps are illustrated in Figure 3.1. The first step that corresponds to Block $\#1$ is to examine the available data. Data can be stored in spreadsheets, databases, binary files, or big data systems. Very often preprocessing of the data is necessary before using it for machine learning. Data sets often contain missing values or outliers due to sensor malfunction or damage. As a result, a series of preprocessing steps are required for finding, removing, and cleaning data in order to prepare it for use. It is also important to visualize data before applying any model. Visualisation can provide useful insight in the data and reveal structure, e.g. correlations or problematic data like outliers. It can also provide information on what type of model suits the problem. For instance, a non-linear problem cannot be properly described with a linear model and will probably have low accuracy.

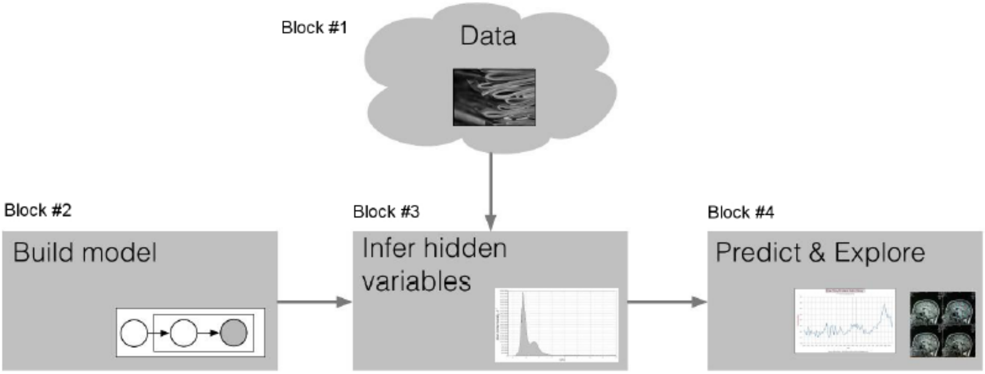


Figure 3.1: Overview of the procedure of setting up a machine learning model [19]

Another data preprocessing technique involves transforming the data in a consistent way to make them suitable for specific types of models and to increase computation efficiency. Data can be standardized in order to have a zero mean ($\mu = 0$) and standard deviation equal to 1 ($\sigma = 1$). Each independent variable is standardized according to the Equation 3.1:

$$X_{i,stand.} = \frac{X_i - \mu_i}{\sigma_i} \tag{3.1}$$

Some algorithms such as ridge regression require standardized data and may give less accurate results in case non-standardized data is used. For other algorithms it might not be necessary to standardise data, but it is generally a common practice. In ANNs the $mapminmax$ transformation is most common for both inputs and outputs, where features are transformed in the $[-1, 1]$ interval. This transformation is also the default pre- and postprocessing method in the Matlab 2016a Neural Network toolbox [8] that is used in this thesis. Its formula is given below. Y_i is the transformed and X_i the original vector. $Y_{max} = 1$ and $Y_{min} = -1$ for the case of transformation to the $[-1, 1]$ interval.

$$Y_i = \frac{(Y_{i,max} - Y_{i,min})(X_i - X_{i,min})}{(X_{i,max} - X_{i,min})} + Y_{i,min} \tag{3.2}$$

Model selection is illustrated in Block $\#2$ and refers to choosing from a variety of available models that are suitable for the task and building the respective algorithm. There is no model that is optimal for all problems so the choice is problem specific. Models differ in complexity, ranging from simple linear regression models to more complex models such as neural networks. Also some models are of probabilistic nature and can provide confidence intervals for the estimations, such as a Gaussian Process model. Another distinction that is important for computational efficiency is between ”lazy” and ”eager” learning. Lazy learning models require the whole training dataset in order to make predictions for new data, which increases storage demands. However they are able to approximate the target function locally which generally increases accuracy. Examples of lazy learning models are GP and nearest-neighbor algorithms. In eager learning the system generalizes the training data before receiving queries, resulting in faster prediction of new values. Eager learning methods include ANNs and linear regression.

The third step shown in Block $\#3$ of Figure 3.1 is to infer the hidden variables of the model for the specific problem that is examined. Depending on the model type there are different variables called ”model parameters” that contain the model information. They add abstraction to the model and are tuned for a specific problem in a procedure called training of the algorithm. A simple example is the offset w_0 and slope w_1 that fully describe a straight line in 2 dimensions. These are the 2 model parameters that are optimized for a Linear Regression (LR) model in the simplest case of a single independent variable. The optimization procedure for the generic LR model is given in Section 5.2. For the optimization of the model parameters an objective function is introduced and this quantity is optimized (either minimized or maximized). The sum of squared errors (SSE) is a common objective function that leads to the widely used least squares criterion. The objective function optimization can both: be analytically tractable or based on iterative algorithms and provide global or local optima.

LR weight optimization returns a global minimum which is found in an analytical way and is given by Equation 6.5. ANNs use the same type of SSE objective function. However the network weights are optimized by means of an iterative algorithm, such as the LM algorithm, which will be presented in Subsection 3.2.3. Additionally, since generally the obtained minimum can be local rather than global, multiple optimization instances should be carried out in order to obtain the best performing training and corresponding network parameters. For the training of the model, the data set is generally divided into a training set, a validation set and a test set, as shown in Figure 3.2.

- training set
  The training set data is presented to the model so it can learn from it. It contains labeled data that has been collected and result from the underlying distribution

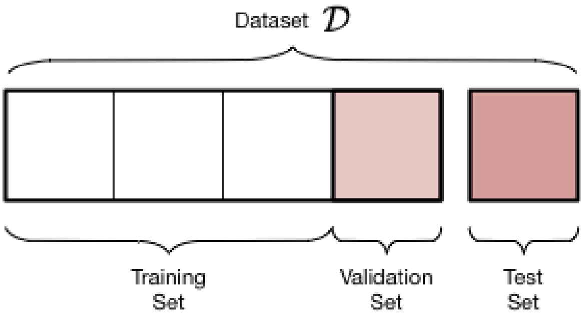


Figure 3.2: Division of the data set in training, validation and test sets [20].

of the problem. In ANNs the training set is used to adjust the network weights.

- validation set
  A validation set may be used by some algorithms like ANNs for early training stopping to minimise overfitting. The network weights are not adjusted based on this data set, but it is used to verify that any increase in accuracy over the training data set actually yields an increase in accuracy over a data set that has not been shown to the network before (network has not trained on it). If the accuracy over the training data set increases, but the accuracy over the validation data set stays the same or decreases, then the neural network is overfitting and training is stopped.

- test set
  The testing set is not presented during training and is used to assess the performance of the trained model. It provides information on how well the model generalizes beyond the training data. The testing dataset also consists of labeled data but only the input is presented to the model, which makes a prediction called output. This output of the model is compared with the known correct value (label) that is called target. The model performance is assessed on the basis of this comparison between output and target.

As soon as the model has been tested on the test set and a satisfactory performance has been achieved, the model can be used to make predictions for new unlabeled input data. This new data is assumed to be generated from the same underlying distribution that the training data came from. In other words, the system that is described by the ML model is supposed not to change. If this is not the case in reality, the applicability of the model should be reconsidered.

3.2 Introduction to ANNs

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The ANN information processing structure consists of a large number of interconnected basic processing elements (neurons) that collectively work to solve specific problems. ANN models have been used for a variety of complex tasks, such as time series prediction, speech recognition and image classification. In the context of this thesis they will be used for the task of function approximation. In this Section, after a brief presentation of biological neurons, an introduction to both the structure and the training of the ANN model will be made.

3.2.1 Biological neuron

A biological neuron or nerve cell is a type of cell that processes and transmits information through electrical and chemical signals [24]. A typical neuron consists of a cell body (soma), dendrites, and an axon. Neurons can connect to each other to form neural networks. Signals are transmitted between neurons via specialized connections called synapses. A typical biological neuron is illustrated in Figure 3.3.

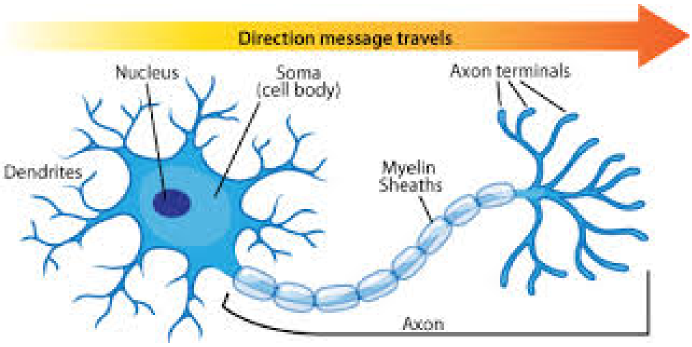


Figure 3.3: Illustration of a typical biological neuron [2].

Neurons are electrically excitable and maintain voltage gradients across their membranes. Dendrites are the branched projections of a neuron that receive the incoming electrochemical stimulation received from other neural cells and propagate it to the cell body, also called soma, of the neuron from which the dendrites project. Electrical stimulation is transmitted onto dendrites by upstream neurons, usually through their axons. The connections between upstream axons and downstream dendrites, named synapses, can be located at different points throughout the dendritic tree. Another function of dendrites is to integrate these synaptic inputs and to determine the extent to which action potentials are produced by the neuron. If there is a change in voltage that exceeds a certain threshold, an all-or-none electrochemical pulse, called an action potential is generated, which travels rapidly along the cell's axon, and activates synaptic connections with other cells when it arrives. Multiple interconnected neurons form neural circuits and neural networks, which are basic components of the nervous system and determine how an organism perceives the world and its reactions to external influences.

3.2.2 Mathematical model of neural networks

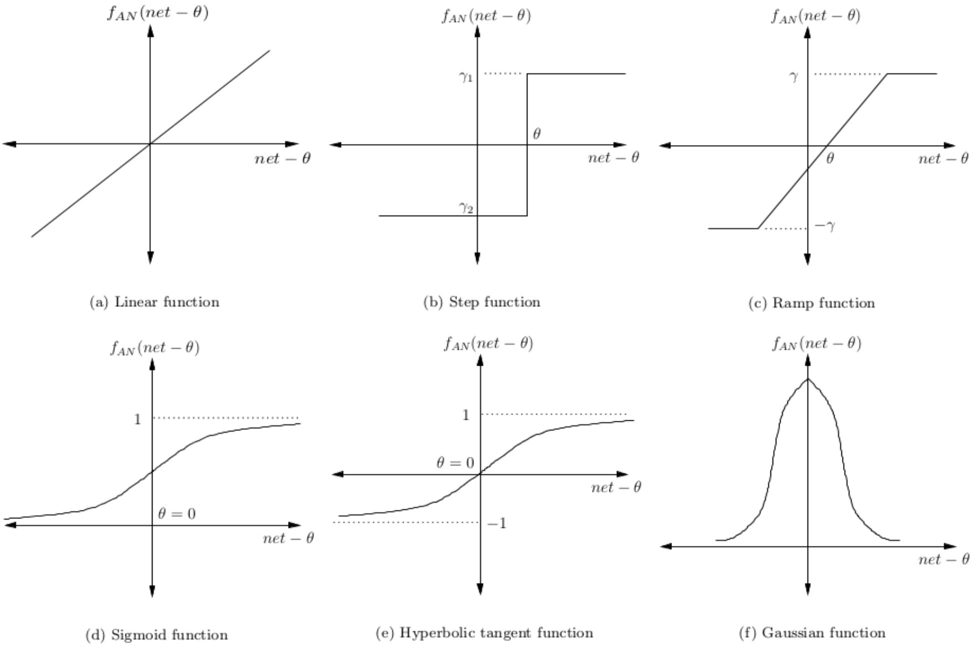
Artificial neural networks (ANNs) are computational models that can describe nonlinear, multivariable, static and dynamic systems. Their function imitates that of biological neural networks and they can be trained by providing them with input – output observations of the underlying system, rather than being explicitly programmed. For that reason, they are generally considered black box, data-driven models opposite to physical models for example. In Leshno et al [10] it is shown that a standard multilayer feedforward network with a locally bounded piecewise continuous activation function can approximate any continuous function to any degree of accuracy if and only if the network's activation function is not a polynomial. The first type of AMM was the perceptron algorithm, invented in 1957 at the Cornell Aeronautical Laboratory by Frank Rosenblatt [22]. It consisted of a single layer of neurons (single layer perceptron), which made it capable of only learning linearly separable patterns. After several years it was recognised that feedforward neural networks with two or more layers (multilayer perceptron) had far better capabilities of representing complex, nonlinear patterns than perceptrons with a single layer.

In analogy with the biological neural networks, multilayer feedforward ANNs consist of several layer of neurons, interconnections between the neurons and weights assigned to these interconnections. Artifical neurons are mathematical models of biological neurons and form the basic building block of an ANNs; these neurons mathematically process and transmit an input to an output. A simple model of an artificial neuron consists of a function with several inputs and one ouput. Inputs are the equivalent of the biological neuron's dendrites and the output is the equivalent of the axon. Each input is weighted by an individual weight factor, representing the synaptic strength. The sum of the weighted inputs, often plus a bias term, is passed through a nonlinear activation function and the resulting value is the output of the artificial neuron. Various types of activation functions exist, some of which are shown in Figure 3.4. Commonly used activation functions are threshold, sigmoidal and tangent hyperbolic functions. Recently, the rectified linear unit (ReLU) activation function has attracted a lot of interest in the field of deep learning [17]. In this thesis the hyperbolic tangent sigmoid activation function will be used, given by Equation 3.3.

$$\sigma(z) = \frac{2}{1 + \exp^{-2z}} - 1 \tag{3.3}$$

Generally activation functions map the neuron's output into a certain interval like

$[0, 1]$ or $[-1, 1]$ [4].



(a) Linear function (b) Step function (c) Ramp function

(d) Sigmoid function (e) Hyperbolic tangent function (f) Gaussian function

Figure 3.4: Common activation functions for artificial neural networks [21].

ANNs are created by interconnected neurons arranged in layers. A typical structure consists of one input and one output layer and one or more hidden layers. ANNs are categorized according to the type of neuron interconnections as follows:

1. Feedforward networks, which may consist of several layers of neurons and information flows in only one direction, namely from input to output. These networks do not contain cycles or loops.

2. Recurrent networks, where feedback can be introduced either internally in the neurons, to other neurons in the same layer or to neurons in preceding layers.

3.2.3 Training of ANNs

As mentioned earlier, the neural network is a model whose parameters are estimated based on known input and output of the system it describes. The process of learning the model parameters is called training of the ANN. All the information that is learnt

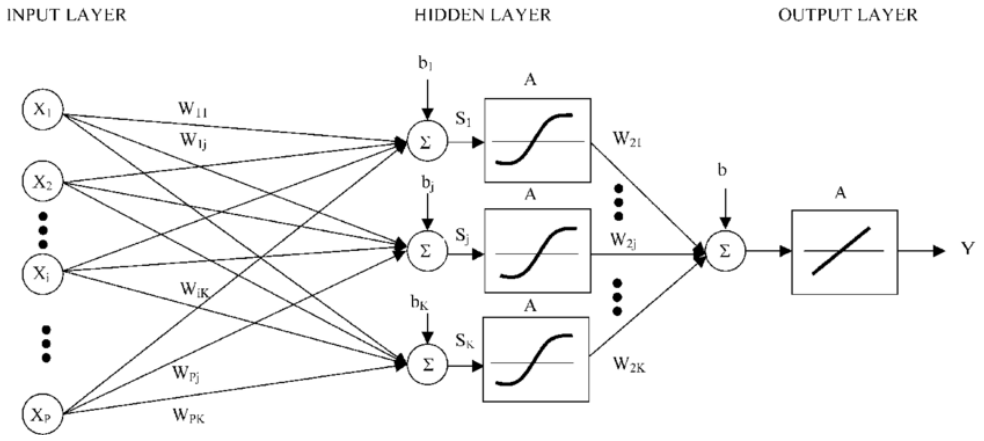


Figure 3.5: Architecture of a typical feedforward artificial neural network with one hidden layer [30].

by an ANN is contained in its topology, namely the number of layers, the number of neurons of each separate layer and the activation functions of the neurons, as well as the weights of the connections between neurons. The training can be defined as the adaptation of weights in the multi-layer network such that the error between the desired output and the network output is minimized. The sum of squared errors (SSE) is introduced as a cost function for the optimization, thus the training procedure is a nonlinear optimization problem with respect to the networks weights. There is a variety of training algorithms, which can be used, each one with their strengths and weaknesses. Some examples are: Error backpropagation (first-order gradient), Newton and Levenberg - Marquardt algorithms (second-order gradient), Bayesian regularization, Conjugate gradients and others.

Error backpropagation was first proposed as a method for training ANNs in 1974 by Werbos [32]. In 1986 Rumelhart, Hinton and Williams [23] showed experimentally that this method can generate useful internal representations of incoming data in hidden layers of neural networks. The term backpropagation is not only associated with first-order gradient methods. In this Subsection the LM algorithm will be presented, also called LM backpropagation. It is commonly used for the non-linear LS optimization of the ANN objective function. For a more detailed presentation of the ANN mathematical model the reader can refer to Babuska [3] and for the LM algorithm to Lourakis [12]. The training phase can be divided in two computational phases:

1. Feedforward computation. Starting from the input layer, the input data is forwarded to the the first hidden layer and summed for each neuron, with a possible

addition of a bias term b_j. This sum is then passed through the activation function and the output is computed, which in turn is passed to the next hidden layer and the same procedure is followed, until the final output of the network is computed. More analytically, the procedure for an ANN with one hidden layer is as follows. First, the input z_j for each of the hidden layer neurons is calculated:

$$z_j = \sum_{i=1}^{p} w_{ij}^h X_i + b_j^h, \quad j = 1, 2, ..., r \tag{3.4}$$

where w_{ij}^h and b_j^h are the weights and bias of neuron j respectively and p is the number of inputs of the network, r the number of hidden layer neurons. This formula also applies in the case of networks with multiple hidden layers. For hidden layers other than the first, the term X_i can be substituted by the output from the previous hidden layer of neurons. Next, th input z_j is passed through the hidden layer neuron in which the output is computed based on the type of activation function ψ.

$$v_j = \psi(z_j), \quad j = 1, 2, ..., r \tag{3.5}$$

Then, the outputs y_k of the network are calculated as follows:

$$y_k = \sum_{j=1}^{h} w_{jk}^o v_j + b_k^o, \quad k = 1, 2, ..., s \tag{3.6}$$

where w_{jk}^o and b_k^o are the weights and biases used to compute the kth output respectively and s is the number of outputs. The same steps can be written in matrix notation, where X_b and V_b include one unit vector each in order to account for the bias terms.

$$Z = X_b W^h V = \psi(Z) Y = V_b W^o \tag{3.7}$$

2. Weight adaptation. The values Y that the network outputs at the end of the feed-forward computation are compared with the known, correct output values (targets) of the training dataset T (in this case the DEL values). The difference between the two arrays is the error of the network E.

$$E = T - Y$$

An objective function has to be established to optimize the weights of the network, either by minimizing or maximizing. In this case, the objective function to be minimized is the sum of squared errors (SSE). It has the following formula for arbitrary number of outputs l.

$$L(w) = \frac{1}{2} \sum_{m=1}^{N} \sum_{k=1}^{s} \epsilon_{mk}^2 = trace(EE^T) \tag{3.8}$$

$$w = [W^h W^o]$$

in which N is the number of samples in the training dataset. Considering a single sample of the training dataset, the objective function can be written as:

$$L(w) = \frac{1}{2}\sum_{i=1}^{l} \epsilon_l^2 = \frac{1}{2}\|\epsilon_l\|_2^2 = \frac{1}{2}\|t_l - y_l\|_2^2 \tag{3.9}$$

Considering the Equations 3.4, 3.5 and 3.6 it becomes clear that the objective function L in Equation 3.8 is a function of the weights w of the network that are used during the feedforward computation. As a result, the training of the ANN is a nonlinear optimization problem with respect to the network weights w that include both hidden layer weights W^h and output layer weights W^o. The weights are optimised such that the error in Equation 3.8 is minimised. There are several training algorithms available, however the standard algorithm is the LM algorithm and this one is presented here.

If a function f is considered, that maps a weight vector $w \in \Re^m$ to the network output $y = f(w)$, where $y \in \Re^n$, the LM algorithm is based on a linear approximation of f in the neighborhood of w. Considering $\|\delta_w\|$ to be small, the function f can be approximated by a Taylor series expansion as:

$$f(w + \delta_w) \approx f(w) + J\delta_w \tag{3.10}$$

in which J is the Jacobian of the network with respect to the weights w and is given by:

$$J(w) = \nabla f(w) = \left[\frac{\partial f(w)}{\partial w_1}, \frac{\partial f(w)}{\partial w_2}, ..., \frac{\partial f(w)}{\partial w_M}\right] \tag{3.11}$$

To optimize the network weights for one sample in the training dataset, the cost function $L(w) = \frac{1}{2}\|\epsilon_l\|_2^2$ has to be minimized, which is equivalent to minimizing:

$$L(w) = \|\epsilon_l\|_2 \tag{3.12}$$

The algorithm is initialized at a starting point by drawing a random weight vector w_0 and it subsequently produces a series of vectors $w_1, w_2, ...$ that converge to a local minimizer w^+ of L. By substituting the equality $y_l = f(w + \delta_w)$ and the Equation 3.10 in Equation 3.12 and expanding, the quantity to be minimized becomes:

$$\|t - f(w + \delta_w)\|_2 \approx \|t - f(w) - J\delta_w)\|_2 = \|\epsilon - J\delta_w)\|_2 \tag{3.13}$$

This is a linear least squares problem and its solution δ_w is attained when $J\delta_w - \epsilon$ is orthogonal to the column space of J. This can be written as:

$$J^T(J\delta_w - \epsilon) = 0 \Rightarrow \quad J^T J\delta_w = J^T\epsilon \tag{3.14}$$

The matrix $J^T J$ is an approximation to the matrix of second order derivatives, the Hessian. In the LM algorithm, a term $\mu > 0$ is added to the diagonal elements of $J^T J$, which is called damping term. With the introduction of this damping term μ, the LM algorithm is able to swap between a slow descent approach when away from the minimum and a fast convergence when close to the minimum. It can additionally cope with cases where the Jacobian is rank deficient and the matrix $J^T J$ is singular. If the updated weight vector $w + \delta_w$ leads to a decrease of the cost function L, the update is accepted and the next iteration takes place with a decreased damping term μ. On the other hand, if the cost function is not decreased, the damping term is progressively increased until a decrease in L is achieved and only then the next iteration begins. Rewriting Equation 3.14 results in:

$$N\delta_w = J^T \epsilon \tag{3.15}$$

where $N = J^T J + \mu I$. The algorithm is terminated at the occurence of one of the following events:

- The error $\epsilon^T \epsilon$ drops below a chosen threshold e_1.
- The relative change in the magnitude of the weight update δ_w falls below a chosen value e_2.
- The magnitude of the gradient $J^T \epsilon$ becomes smaller than a threshold value e_3.
- A chosen number of algorithm iterations is reached.

3.3 Conclusions

In this Chapter an introduction to ML models and the range of their applications in supervised and unsupervised learning was done. The 4 basic steps of creating a ML model were presented, including the role of the training, validation and test data sets:

1. Data set preparation, visualization and division in training, test (and possibly validation) sets.
2. Choice of a ML model. This choice can be influenced by a variety of factors. For instance some ML models such as ANNs are efficient for big data sets in terms of features and/or samples, while other models like Gaussian Processes include an uncertainty indication for each prediction.
3. Tuning of the specific model parameters in order to fit the data in the training procedure. Evaluation of the generalization to new data, using the test data set.

4. Making predictions for new data in the supervised learning case (e.g. regression or classification) and exploring data in unsupervised learning tasks (e.g. clustering).

In Section 3.2 the mathematical model of feedforward artificial neural networks and the Levenberg-Marquardt algorithm for network optimization have been presented. Summarizing the most important characteristics of ANNs:

1. ANNs are inspired from the biological neurons and neural networks that form the nervous systems of most living organisms. They consist of artificial neurons organized in interconnected layers.

2. They are nonlinear black-box models when nonlinear activation functions are used. They can capture nonlinear relationships between input and output and approximate arbitrarily nonlinear multivariable functions [10].

3. A variety of iterative optimisation algorithms can be used to train the network to the specific problem data. The term training is used to describe the adaptation of the connection weights, which is a nonlinear optimization problem that returns local optima. In this work the second order gradient based Levenberg-Marquandt algorithm will be used to perform the nonlinear optimization problem.

4. ANNs belong to the eager-learning models that generalize the training data before new queries are made for prediction. In other words, after the model has been trained, the training set is no longer necessary for new predictions, since all the information is contained in the network architecture and weights. This makes new predictions fast to evaluate.

Chapter 4

Application of Artificial Neural Networks in Load Monitoring

In Chapter 4 ANNs will be applied for load monitoring of an OWT and particularly the estimation of equivalent load ranges of monopile overturning moments at seabed level. Information about the data sets used in this thesis is presented in Section 4.1. Furthermore, the values that will be used for specific ANN parameters are listed and the performance metrics that will be used to assess accuracy are introduced. In Section 4.2 some selected variables of the dataset are visualised. Conclusions drawn from the visualisation can potentially be used to improve the input data configuration. Additionally, they can indicate potential for application of a different model. Section 4.3 contains the results of the ANN application for estimation of equivalent load ranges of the monopile bending moment at the seabed for an OWT in normal operation conditions. Only standard signals and additional accelerometers at the tower bottom are considered as potential network inputs. A sensitivity analysis of the output accuracy for different input combinations is presented.

In the next Section, 4.4, the differences in complexity of the turbine behaviour in the FA and SS directions are related to the accuracy of the estimation in those directions. The role of wave-induced loading is taken into account in assessing this phenomenon and the inclusion of the significant wave height as input to the ANNs is examined. Section 4.5 showcases the application of ANNs for the estimation of seabed overturning moment equivalent load ranges during idling conditions. Lastly, in Section 4.6 the results of this Chapter are summarized and the most important conclusions are presented.

4.1 Case study details

Important information about the origin of the data set used throughout the present Chapter is given in this Section. Furthermore, some artificial neural network parameters that are kept constant in this work are listed and the performance metrics that are used to assess the estimation accuracy of the data-driven models are introduced.

4.1.1 Data set details

For this thesis standard signals and structural response data from design simulations of the integrated turbine and support structure are used. A single OWT located in the Westermost Rough park is examined. The design has been carried out by the Siemens Wind Power division in The Hague (the Netherlands). Time domain FE simulations for various loading and operational scenarios are obtained using the inhouse aeroelastic tool ”BHawC”; BHawC allows for multiple output channels in the post-processing phase, like electrical power, blade pitch angle, wind speed at the hub and also displacements, rotations and accelerations on tower, foundation and blades, internal forces and moments. High resolution time series are generated for the output channels.

In the context of this thesis these simulation results are treated as virtual measured data for testing the application of ANNs for the purpose load monitoring. Simulation results from BHawC contain no noise and will be first used in that pure form in order to examine if a transfer function between input and output can be established with proper accuracy. In case of an application to a real turbine, either simulated or real measurement data can be used for the ANN training phase. However real measurement data has to be used for the actual monitoring application once the model has been trained. Actual measured data from the field always include some noise, since measuring equipment and methods are not perfect. A noise sensitivity study is presented in Chapter 5.

At first, the normal operation state of the turbine, with presence of waves and wind-wave misalignment is considered, named DLC 1.2 in relevant standards. For more information on design load cases the reader can refer to [1]. The total sampled 10-minute simulations that will be used for the training amount to 11232, with varying environmental conditions to properly resemble the conditions that the structure will encounter over its lifetime. Numerous variations of wind and wave directions, yaw error as well as seeding are captured. The mean wind speed varies from 3 m/s (cut-in wind speed) to 28 m/s (cut-off wind speed) with a step of 1 m/s.The most important loads that act on the OWT are wind and wave loads. Each mean wind speed bin includes several hundreds of simulations with different seeds and directions of both wind and wave, yaw errors, etc. The significant wave height H_s of the deployed normal sea states is correlated with wind speed, according to the site-specific environmental conditions.

The data set from simulated results contains a large number of observations and will be used to determine what kind of input signals and which statistical parameters are necessary to properly reflect the variation of the target loads and effectively obtain accurate predictions. The examined load signals and their damage equivalent load (DEL) format that will be used as targets for the ANNs in this Chapter are described in Section 2.2. The measured turbine signals and their statistics that will be used as inputs for the ANNs are described in Section 2.3.

In addition, the same wind turbine is examined for idling conditions (DLC 7.2). In this case, the blades are pitched to idling position and the turbine is not producing electricity. The 10-minute mean wind speed varies from 1 to 32 m/s with steps of 1 m/s. The aerodynamic damping is limited due to the state of the turbine. This load case consists of 2304 samples in total, all of which are used in this study. The same approach as in normal operation case is followed, with monopile bending moments at the mudline acting as the target loads. However the number of examined input signals are limited to four in this case, since the pitch angle and electrical power signals have zero variation during idling conditions and are not relevant. Pitch angle influences the turbine aerodynamic loads during idling conditions, but its value is constant not only over each 10-minute duration, but also over all different simulations. Thus, it induces no variation to the data set and is disregarded.

4.1.2 ANN implementation details

The Matlab Neural Network Toolbox is used to create, train and test feedforward ANNs in this work [8]. Networks with one hidden layer have been used for the load estimation and generally a small number of less than ten hidden neurons has been proven sufficient for accurate estimation. Information about the ANN architecture like number of hidden layers, number of hidden neurons and activation functions will be given for each specific case study at the respective part of the report. For the training of the ANNs the LM algorithm is used in all cases. The default Matlab values for the LM algorithm are used and the most important of them are presented in Table 4.1. For further information the reader is referred to the Matlab documentation for the LM algorithm.

4.1.3 Estimation accuracy metrics

Three metrics for assessing the accuracy of the estimated output values with respect to the correct target values are introduced in this Subsection. A metric that quantifies the linearity of the correlation between outputs and targets is the Pearson's product-moment correlation coefficient R, which is given by Equation 4.1. Throughout this

Table 4.1: Parameters of the Levenberg-Marquardt training algorithm and processing function that are used in this work.

Parameter name	Value
LM optimisation algorithm	
Maximum number of epochs	1000
Minimum gradient	10^{-7}
Maximum damping term	10^{10}
Maximum validation failures	6
Pre- & post-processing	
Processing function	mapminmax (Equation 3.2)

work the correlation coefficient is calculated only for the test set values.

$$R_{Y,T} = \frac{cov(Y,T)}{\sigma_Y \sigma_T} = \frac{E\left[(Y-\mu_Y)(T-\mu_T)\right]}{\sigma_Y \sigma_T} = \frac{E\left[YT\right]-E\left[Y\right]E\left[T\right]}{\sigma_Y \sigma_T} \tag{4.1}$$

where Y stands for the vector of output values of the model and T stands for the vector of target values. The symbols μ and σ stand for the mean value and standard deviation respectively. A second performance index is the mean absolute percentage error $\epsilon_{mu,abs,test}$ of the test samples, which is given by Equation 4.2, where n_{test} is the number of test samples.

$$\epsilon_{mu,abs,test} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left| \frac{Y_i - T_i}{T_i} \right| \tag{4.2}$$

The third index is the mean percentage error $\epsilon_{mu,all}$ of all data samples, including training, validation and test. It can take both negative and positive values and reflects the residual of the individual percentage errors. It is computed according to Equation 4.3, where n_{all} is the number of samples of the full data set (training, validation and test).

$$\epsilon_{mu,all} = \frac{1}{n_{all}} \sum_{j=1}^{n_{all}} \frac{Y_i - T_i}{T_i} \tag{4.3}$$

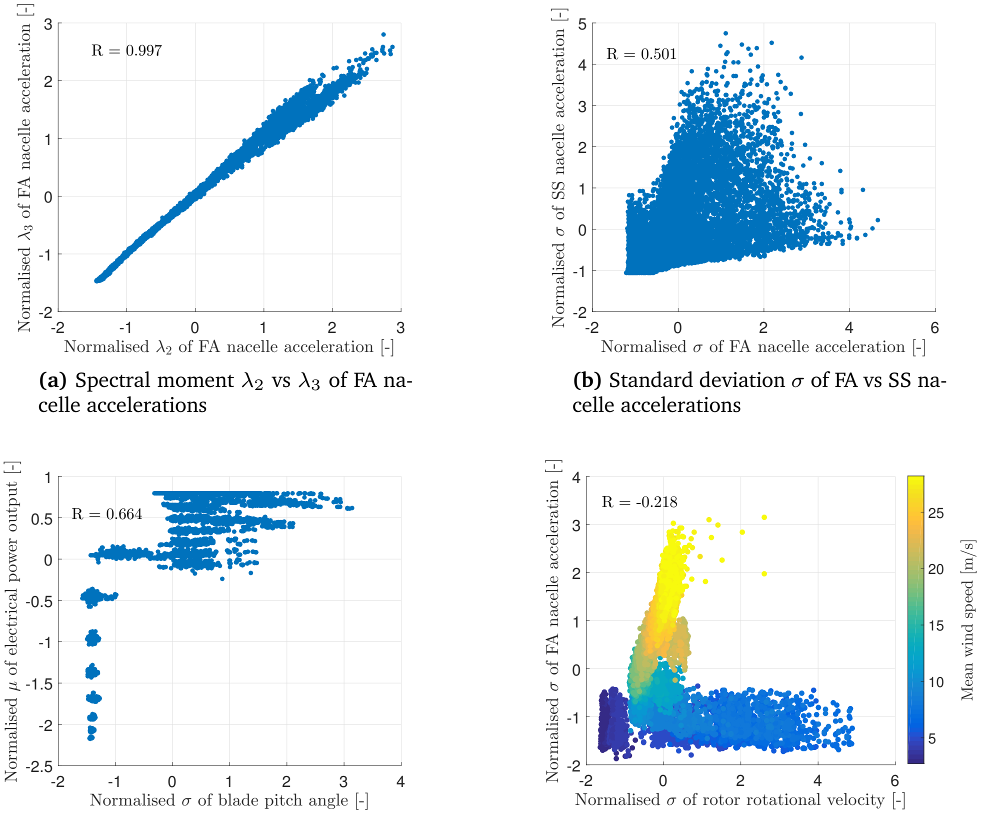
4.2 Dataset exploration

An important step before applying a ML model is to visualise the available input and output data. The different inputs can be plotted against each other, as well as against the output in order to better understand the data structure and to discover correlations. This step is efficient when the number of inputs is relatively small. On the other hand, when the input space is high-dimensional, it is inefficient to visualise all variable correlations but it is a good practice to visualise at least some selected variable combinations. This visualisation process should not be confused with input selection methods. Input selection methods return selected subsets of the available input space, such that most information is retained but redundant variables are left out to increase efficiency. A review of these methods for ANNs is given in May et al. [14]. In Figures 4.1a to 4.1d some selected input variable pairs are plotted together in order to examine the degree of their correlation.

It is generally desirable to use input variables that are uncorrelated. Variables that are highly correlated with others are superfluous, since a single of those variables provides a similar amount of information as the whole set. As a result, extra variables increase computational demands without significant benefit for the accuracy. In Figure 4.1a two highly correlated variables (second and third spectral moments λ^2 and λ^3 of FA nacelle acceleration) are plotted. Their correlation is highly linear, as can be evidenced by the Pearson correlation coefficient of $R = 0.997$. This means that either of the two variables can be ommitted, without losing much information. On the other hand, in Figure 4.1b two less correlated variables are shown with $R = 0.501$. While they are not linearly dependent and each one provides unique information, it is possible that for the estimation of specific target variables one of the two is redundant. This means that the target variable depends only on one of those two input variables.

Proper visualisation of selected variables can provide useful insight in the characteristics of the examined system – in this case the OWT. In Figure 4.1c the standard deviation σ of blade pitch angle is plotted against the mean value μ of the 10-minute power output. At the area of low power mean values μ the pitch standard deviation σ is at its lowest value. This corresponds to the below rated operation domain, where the blade pitch is kept constant at the angle that maximises power output from the (low) wind speed. On the upper part of the graph the above rated operation domain can be recognised. In this domain the power output is kept constant at rated power – 6 MW for the specific Siemens SWT-6.0-154 OWT – by changing the blade pitch angle accordingly. In Figure 4.1d the standard deviation σ of rotor rotational velocity is plotted against standard deviation σ of FA nacelle accelerations color-coded according to the mean μ of the wind speed perpendicular to the rotor plane. Distinct regions can be recognised based on the color coding.

Except for input correlations, input variables can be visualised against target variables. For instance, Figure 4.2a illustrates a linear-resembling correlation between an input and a target variable. Such observations can be valuable, because they can lead to examining and potentially using a simpler data-driven model for the specific task. In fact this linear correlation between the standard deviationσ of nacelle accelerations and the DEL of mudline bending moment is the basis for using LR models in Chapter



(a) Spectral moment λ_2 vs λ_3 of FA nacelle accelerations

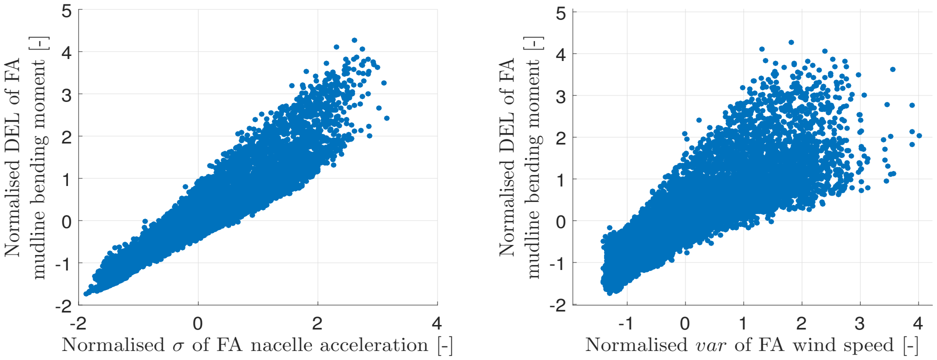
(b) Standard deviation σ of FA vs SS nacelle accelerations

(c) Standard deviation σ of blade pitch angle vs mean μ of electrical power output

(d) Standard deviation σ of rotor rotatonal velocity vs σ of FA nacelle accelerations, color coded based on 10-minute mean wind speed

Figure 4.1: Correlation plots and coefficients of selected statistics of input signals. Each point corresponds to a 10-minute simulation of the turbine under normal operation conditions.

6. In Figure 4.2b the correlation between the variance var of wind speed and the DEL of MBM is shown, which exhibits higher scatter than Figure 4.2a. It can be deduced that nacelle acceleration σ is a better estimator than wind speed var for MBM DELs.



(a) Standard deviation σ of FA nacelle acceleration vs DEL of MBM

(b) Variance var of FA wind speed vs DEL of MBM

Figure 4.2: Correlation plots of selected input and target variables. Each point corresponds to a 10-minute simulation of the turbine under normal operation conditions.

4.3 Load estimation using ANNs for normal operation conditions

The results of the load estimation study with the use of ANNs for the normal operation case are presented in this Section. The accuracy of ML techniques under ideal conditions is examined in order to assess the potential for practical use. The term ”ideal conditions” is used because noiseless data from BHawC simulations are used and because in real application the actual turbine might differ substantially from its finite element digital twin.

4.3.1 Illustration of ANN nonlinearity

In Figures 4.3 and 4.4 the resulting model for the FA and SS directions is illustrated in the form of the learned 3-dimensional nonlinear surface. Only two input quantities have been used in both cases to enable visualisation. These two input quantities are the standard deviations of nacelle accelerations in FA and SS, as defined in Table 2.3. The network characteristics and inputs are summarised in Table 4.2.

Each point on the nonlinear surface represents the output value y that the ANN returns for the respective combination of input values (x_1, x_2), after it has been trained with a part of the available data set (training set). The nonlinearity of the learnt surface illustrates the ability of the ANN to approximate arbitrarily nonlinear multivariable functions [10].

Table 4.2: Parameters of the ANNs and the data set used for the ANN nonlinearity illustration.

Number of hidden layers	Number of neurons	Activation function	Train/Validation/Test (%)
1	8	*tanh*	70/15/15
Input combination	**Statistics**		
4 (Table 2.3)	σ		

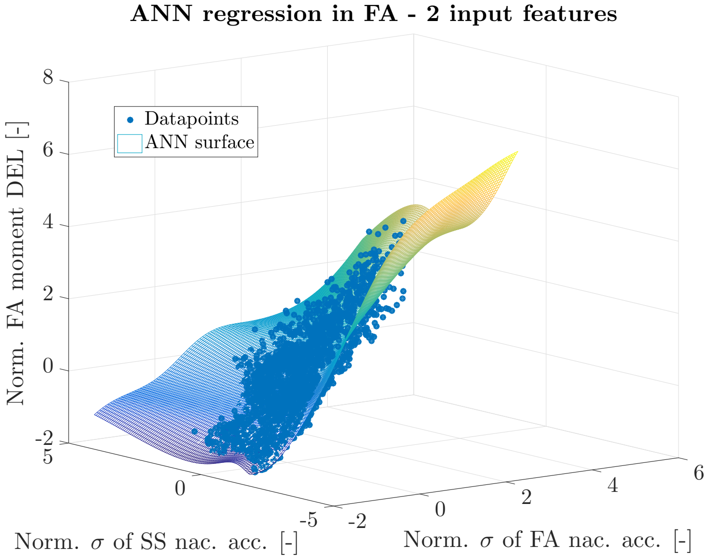


Figure 4.3: Learned nonlinear surface for the FA direction using an ANN with two inputs.

An inherent limitation of the ANN model, as well as of all ML models, is that they are not able to extrapolate and make predictions for input out of the range that has been presented during training. So the network will return an output value for all input vectors that are provided to it, but input vectors might fall outside the range for which the network has been optimised. This is an important distinction. The network estimate in such domains is rather random and different networks can provide completely different estimates even when the same training set is used. Such behavior is shown in Figure 4.4. In the region where the surface is colored yellow the network is extrapolating. These estimates are non-realistic and most importantly are not supported by data. So if the model is making predictions for new input it should be assured that these fall within the input domain which is used for the training of the

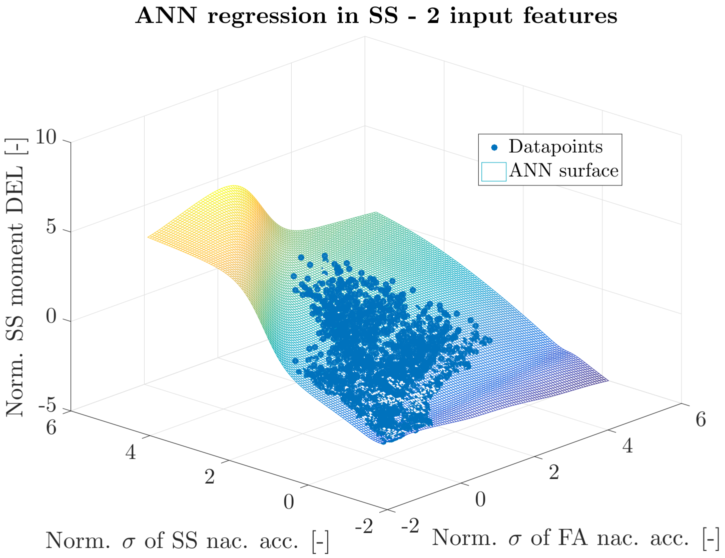
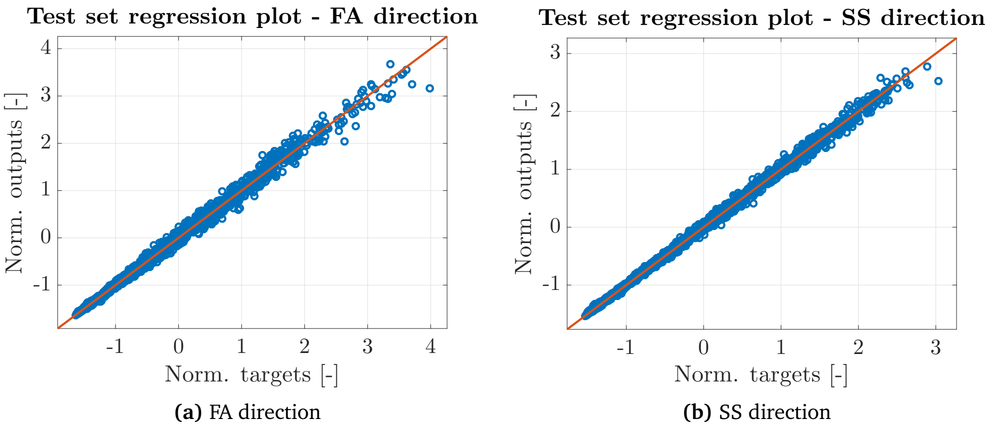


Figure 4.4: Learned nonlinear surface for the SS direction using an ANN with two inputs.

network. Input values that do not agree with this criterion are beyond the model's capabilities and should not be taken into account. One important benefit of using simulated data for the training of the ANNs is that extra simulations can be run after the initial training and then they can be used to expand the training set a posteriori. This method is called ”training in batches”. The already optimised network is optimised further based only on the new data.

4.3.2 Evaluation of ANNs accuracy using standard signals as inputs

A commonly used type of figure for evaluating the prediction accuracy is the regression plot. In regression plots the estimation accuracy is assessed by the degree of similarity between the x and y values, which is visually indicated by the scatter of the points around the line $x = y$, which is plotted with an orange line. In the case of perfect estimation all datapoints would be on that line. The target values that are a priori known (in this case DELs that have been computed with the rainflow counting method) are plotted on the x axis, whereas the output values that are estimated by the network are ploted on the y axis. Figure 4.5a is a regression plot of the DEL moments in the FA direction as estimated by a ANN whose parameters are shown in Table 4.3. The hidden layer activation function is the hyperbolic tangent sigmoid function,



(a) FA direction

(b) SS direction

Figure 4.5: Regression plots of MBM DELs in FA and SS directions, using standard signals and all statistics, including negative spectral moments λ_- as inputs to the ANNs. Results are shown for test set only.

which is presented in Equation 3.3. In this case all the standard signals of the wind turbine have been used for the estimation, which corresponds to input combination 1-5.

Table 4.3: Parameters of the ANNs and the data set used for the standard signals input combination.

Number of hidden layers	Number of neurons	Activation function	Train/Validation/Test (%)
1	8	*tanh*	70/15/15
Input combination	**Statistics**		
1-5 (Table 2.3)	All (Table 2.4)		

The correlation coefficient R has the value 0.995 for the test set in this example. This is considered rather high, indicating that there is a small scatter around the line $x = y$. The estimation has a mean absolute test error of 3.2% and mean error of 0.2%.

In Figure 4.5b the same results are illustrated, but in this case for the SS direction of the wind turbine. In this case a higher coefficient $R = 0.998$ for the test set is found and the scatter of the values is slightly smaller, indicating a good agreement between predicted and target values. The mean absolute test error in this case is equal to 3.1% and mean error of 0.5%. Therefore, the DEL estimation is accurate both in the FA and the SS direction when using all standard signals. It should be mentioned at this point that this similar performance in FA and SS directions is achieved when negative spectral moments of the standard signals are included in the estimation. If only positive spectral moments are used, the estimation is sigificantly less accurate in both directions. It is also less accurate in FA direction than in SS. In the SS direction there is significantly lower aerodynamic damping than in the FA, which constitutes the dynamic behavior simpler to capture through the standard signals. On the other hand, the nonlinearity of the aerodynamic damping combined with significant wave loads make the structural behavior more complex in the FA direction, which could require additional inputs to increase the accuracy of the estimation. Negative spectral moments substantially improve the accuracy, which shows that there is significant content in the low frequencies of the standard signals. The difference between excluding and including negative spectral moments can be seen in Table 4.4.

Table 4.4: Accuracy of DEL estimation with ANNs for the normal operation case, with and without including negative spectral moments of standard signals as inputs.

Input combination	**FA**			**SS**		
	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$
1-5, λ_+	0.974	7.0	0.7	0.991	5.9	1.5
1-5, λ_+ & λ_-	0.998	3.1	0.5	0.995	3.2	0.2

4.3.3 Accuracy of ANNs for different input combinations

In order to examine which input signals have the highest correlation with the target DELs, a brute force method is used to examine the permutation of the five standard signals and the TB acceleration. All possible combinations of the six signals are used to train ANNs for both FA and SS directions and their accuracy is compared. The reason for choosing this brute force method is the small number of possible input signals. All statistics of each individual signal are used in the input combinations, thus this study does not examine the influence of individual statistics. To allow for a proper comparison the training, validation and test sets are identical for all 63 input combinations. The geometry of the ANNs also remains constant in all combinations, according to Table 4.5, although this is not necessary for a fair comparison.

Table 4.6 illustrates the effect of using different combinations of input signals and their statistics as input to the networks. There are in total 63 different combinations of 6 different signals, in which all signals that have both a FA and SS component are considered as one input with double statistics. The first column of the Table shows the input combination, while the rest of the columns show the three performance metrics both for the FA and SS directions. Also, the median value of the metrics from eleven different training sessions has been used in all combinations, since the weights

Table 4.5: Parameters of the ANNs and the data set used in the input sensitivity study for the normal operation case.

Number of hidden layers	Number of neurons	Activation function	Train/Validation/Test (%)
1	8	*tanh*	70/15/15
Input combinations	**Statistics**	**Initialisations per combination**	
Permutation of available inputs (Table 2.3)	All (Table 2.4)	11 – median shown in Table 4.6	

initialisation for each training instance affects the obtained results.

Several conclusions can be drawn for the effectiveness of the various inputs. The single most important input among standard signals is the nacelle acceleration. It achieves overall the highest performance among all cases where a single input is used, with a 4.1% mean absolute error in the FA and 3.0% in the SS. TB acceleration gives comparable accuracy level with errors equal to 3.3% and 4.6% in FA and SS respectively. Nacelle accelerations seem to have a stronger correlation with the SS MBMs than TB accelerations. The opposite applies for the SS direction. However the difference in accuracy is small in both cases. The accuracy that each sensor provides for the FA and SS directions can be related to the modal characteristics of the turbine and the location of the accelerometers. As is shown in Figure 4.6 the first bending mode has its maximum deflection value at the nacelle level, where the nacelle accelerometers are located. On the other hand, the first bending mode has a much lower value at the TB level, lower than the second mode, as is depicted in Figure 4.7. The MBM is generally dominated by the first bending mode, with the higher bending modes contributing significantly less. The different excitation of the structure in the two distinct directions can be the reason why different accelerometers give better estimation for each direction.

When the two acceleration signals (combination 4,6 in Table 4.6) are combined, the mean absolute error is decreased to 3.0% and 2.8%. From the results it can be deduced that accelerations are suitable estimators for the target load values. This confirms the initial intuition that they would provide good accuracy since both accelerations and moments are response quantities of the turbine structure. On the contrary, wind speed and pitch angle are not response quantities, since they describe the loading and the system itself, rather than the system's response to the loading. However, high estimation accuracy can be also achieved by using all available standard signals, without the need for installing TB accelerometers. The combination 1-5 gives mean absolute error of 3.2% and 3.1% in the two directions. This fact shows potential for accurate load monitoring using only the currently measured signals.

Table 4.6: Accuracy metrics for all input signal combinations for the normal operation case.

Input combination	**FA**			**SS**		
	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$
1	0.893	17.6	4.6	0.712	53.3	28.8
2	0.894	18.2	5.4	0.697	72.5	46.1
3	0.902	18.0	4.8	0.967	14.4	4.0
4	**0.994**	**4.1**	0.3	**0.998**	**3.0**	**0.2**
5	0.884	18.4	5.2	0.604	82.9	56.0
6	**0.996**	**3.3**	**0.2**	**0.996**	**4.6**	**0.7**
1, 2	0.897	17.5	4.7	0.759	49.1	25.4
1, 3	0.927	16.2	4.2	0.970	15.5	5.7
1, 4	0.995	3.5	4.2	0.998	2.8	0.2
1, 5	0.895	17.7	4.8	0.743	50.5	27.2
1, 6	0.997	2.8	0.2	0.996	4.4	0.7
2, 3	0.925	16.2	4.2	0.969	14.6	4.2
2, 4	0.995	3.7	0.3	0.998	3.1	0.3
2, 5	0.897	17.5	4.7	0.712	71.6	46.2
2, 6	0.996	3.2	0.3	0.996	4.5	0.8
3, 4	0.994	3.6	0.2	0.998	3.1	0.2
3, 5	0.913	17.0	4.2	0.970	14.2	4.0
3, 6	0.997	2.8	0.2	0.996	4.4	0.8
4, 5	0.994	3.8	0.3	0.998	3.2	0.3
4, 6	**0.997**	**3.0**	0.2	**0.998**	**2.8**	**0.2**
5, 6	0.996	3.3	0.3	0.996	5.0	0.8
1, 2, 3	0.929	15.7	3.9	0.973	14.3	5.2
1, 2, 4	0.996	3.1	0.2	0.998	2.9	0.2
1, 2, 5	0.897	17.4	4.5	0.773	47.9	24.3
1, 2, 6	0.997	2.8	0.2	0.996	4.5	0.7
1, 3, 4	0.995	3.3	0.2	0.998	3.2	0.2
1, 3, 5	0.927	16.0	4.0	0.970	16.7	5.3
1, 3, 6	0.997	2.7	0.2	0.996	4.4	0.9
1, 4, 5	0.995	3.3	0.2	0.998	3.0	0.3
1, 4, 6	0.997	2.5	0.1	0.998	2.7	0.2
1, 5, 6	0.997	2.9	0.2	0.996	4.8	0.9
2, 3, 4	0.995	3.3	0.2	0.998	3.1	0.3
2, 3, 5	0.926	15.9	4.0	0.965	18.2	8.0
2, 3, 6	0.997	2.7	0.2	0.996	4.9	0.9
2, 4, 5	0.995	3.6	0.2	0.998	3.2	0.2
2, 4, 6	0.997	2.9	0.2	0.998	3.0	0.2
2, 5, 6	0.997	3.0	0.2	0.996	4.9	1.0
3, 4, 5	0.994	3.5	0.3	0.998	3.0	0.3
3, 4, 6	0.997	2.4	0.1	0.998	2.7	0.2
3, 5, 6	0.997	2.8	0.2	0.996	5.1	1.2

Input combination	**FA**			**SS**		
	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$
4, 5, 6	0.997	2.8	0.2	0.998	3.0	0.3
1, 2, 3, 4	0.995	3.1	0.2	0.998	3.1	0.3
1, 2, 3, 5	0.930	15.6	3.7	0.972	15.6	4.9
1, 2, 3, 6	0.997	2.6	0.2	0.996	4.6	0.7
1, 2, 4, 5	0.996	3.2	0.2	0.998	3.1	0.2
1, 2, 4, 6	0.997	2.4	0.1	0.998	2.8	0.1
1, 2, 5, 6	0.997	2.7	0.2	0.996	4.9	0.8
1, 3, 4, 5	0.995	3.2	0.2	0.998	3.2	0.2
1, 3, 4, 6	0.997	2.4	0.1	0.998	3.0	0.3
1, 3, 5, 6	0.997	2.9	0.2	0.996	5.5	1.3
1, 4, 5, 6	0.997	2.4	0.1	0.998	2.8	0.2
2, 3, 4, 5	0.995	3.3	0.2	0.998	3.3	0.3
2, 3, 4, 6	0.997	2.5	0.2	0.998	2.9	0.2
2, 3, 5, 6	0.997	2.7	0.2	0.996	5.1	1.2
2, 4, 5, 6	0.997	2.7	0.2	0.998	2.8	0.3
3, 4, 5, 6	0.997	2.5	0.2	0.998	2.9	0.2
1, 2, 3, 4, 5	**0.996**	**3.2**	**0.2**	**0.998**	**3.1**	**0.2**
1, 2, 3, 4, 6	0.997	2.4	0.1	0.998	2.8	0.3
1, 2, 3, 5, 6	0.997	2.6	0.2	0.996	5.0	0.9
1, 2, 4, 5, 6	0.997	2.5	0.2	0.998	3.2	0.3
1, 3, 4, 5, 6	0.997	2.4	0.1	0.998	2.9	0.2
2, 3, 4, 5, 6	0.997	2.5	0.1	0.998	2.9	0.3
1, 2, 3, 4, 5, 6	0.997	2.3	0.1	0.998	2.7	0.2

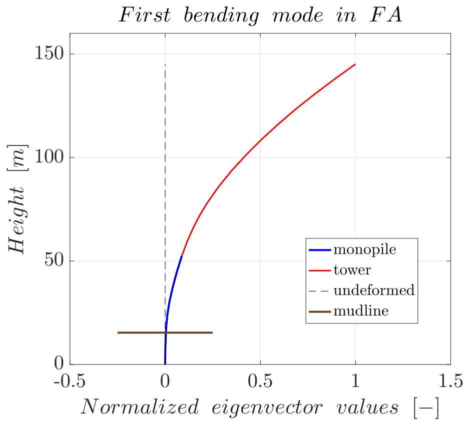


Figure 4.6: The first bending mode of the turbine in the FA direction.

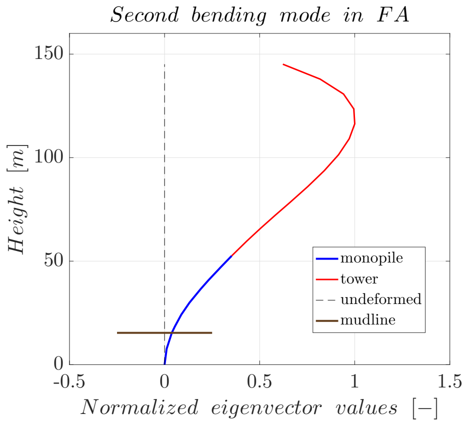


Figure 4.7: The second bending mode of the turbine in the FA direction.

4.4 Influence of hydrodynamic loads on DELs and on estimation accuracy

Hydrodynamic loads from waves play a major role on the internal forces and moments that are developed in the structure and therefore they contribute to the consumed lifetime. It is necessary to include the effect of these loads in the estimation procedure, either in the form of wave measurements or structure response to the wave loads. Wave mesurements are generally costly and often inaccurate if available, therefore emphasis has been given until now on capturing the contribution of wave loads through structural acceleration measurements instead. In this Section the influence of waves on the loading of the turbine is examined in two distinct cases. The first one includes combined aerodynamic and hydrodynamic loading in the FA direction and the second includes mostly hydrodynamic loading in the SS direction.

4.4.1 Influence of hydrodynamic loads on FA and SS DELs

The observation that was made in Section 4.3 that the turbine behavior in FA is more complex needs to be confirmed. The difference in estimation accuracy for the FA and SS directions, as well as the underlying reason for the discrepancy, are explained by Figure 4.8. In this Figure the MBM DELs in both directions are plotted against the standard deviation of the respective nacelle acceleration. Each point on the graph represents the input–output combination of a 10-min simulation. The graph for the FA direction shows a significantly higher scatter than the one for the SS. A color-coding of the data points based on the significant wave height H_s shows that the scatter in the FA can be explained by taking H_s into account. Higher wave heights correspond to higher DELs for the same standard deviation of acceleration (x-axis). In other words, the mixed aero- and hydrodynamic information contained in the FA nacelle acceleration is ”clarified” when further information about the hydrodynamic loads is introduced through the H_s value. On the other hand, the SS nacelle acceleration already contains almost exlusively hydrodynamic information. As a result, further introducing the specific H_s value offers a smaller estimation improvement.

4.4.2 Including wave measurements as inputs to the estimation

The available wave characteristics from the simulations that can be used as extra inputs to the model are: significant wave height H_s, peak wave period T_p and also wave direction. In this Section the inclusion of the 10-minute significant wave height is included as an additional input to the estimation procedure of the ANNs decribed in Table 4.3. The extended input combination is compared to the input combination that was examined in Subsection 4.3.2. The estimation accuracy metrics for the ANNs of the original case and the extra input are shown in Table 4.7.

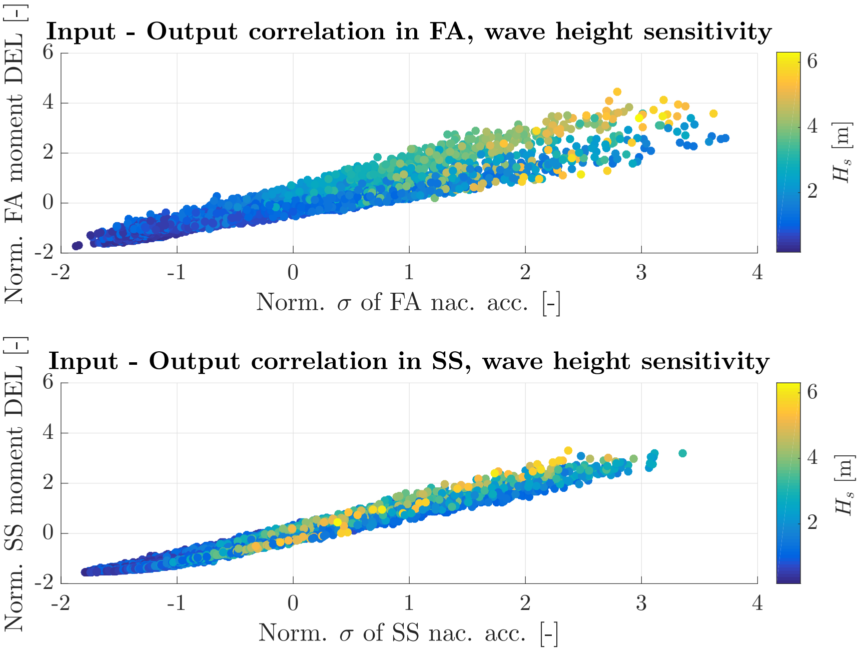
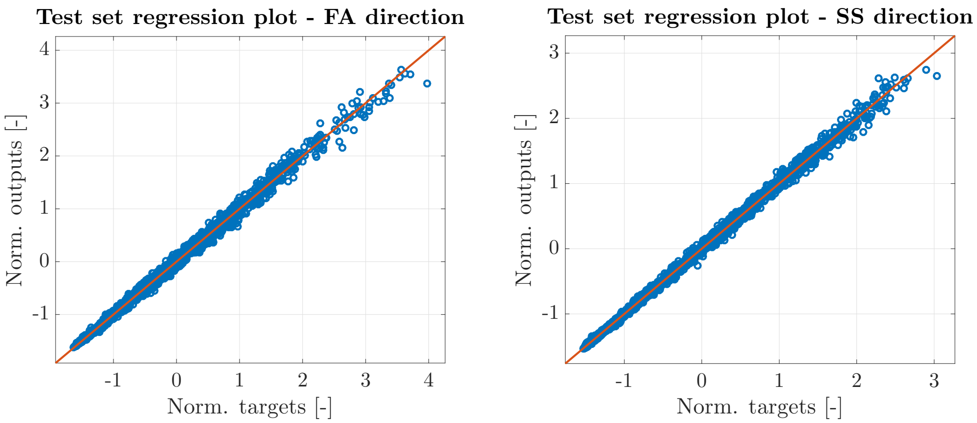


Figure 4.8: Standard deviation σ of nacelle accelleration plotted against MBM DEL in FA and SS directions, color-coded based on 10-minute significant wave height H_s.

Table 4.7: Accuracy of ANNs with and without adding significant wave height as input to the estimation, in addition to standard signals.

Input combination	FA			SS		
	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$	R[-]	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$
1-5	0.998	3.1	0.5	0.995	3.2	0.2
1-5 & H_s	0.996	2.9	0.1	0.998	3.2	0.3

Performance metrics show a mixed behavior. Some metrics are improved, such as $\epsilon_{mu,abs,test}$ and ϵ_{mu} for the FA direction. On the other hand, R for FA is slightly decreased. In general, both combinations give approximately the same accuracy level in both directions. Thus, including H_s as input does not benefit the accuracy substantially if all standard signals and statistics are available. The regression plots for both directions are illustrated in Figures 4.9a and 4.9b. Only the plots for the test set are shown.



(a) FA direction

(b) SS direction

Figure 4.9: Regression plots of MBM DELs in FA and SS directions using significant wave height in addition to standard signals as inputs to the ANNs. Results are shown for test set only.

4.5 Load estimation using ANNs for idling conditions

In this Section, the procedure of determining the most important input signals for load estimation as well as the accuracy of the respective ANNs is presented. For this assessment the idling load case (DLC 7.2) is used to generate simulation data for the examined Siemens SWT-6.0-154 OWT. The examined idling conditions consist of 2304 different 10-minute simulations. The time series of these simulations are preprocessed and the data set is obtained, divided in inputs and targets. During the idling conditions as simulated in BHawC, some of the available signals that are listed in Table 2.3 have zero variation over the simulation length. These two signals are the pitch angle θ_b^p and the electrical power output P_{el}. The latter also has a zero mean value, while this is not the case for θ_b^p. However the mean value of θ_b^p is constant over all different 10-minute simulations, thus inducing no variation in the data set. As a result the reduced input and output signal table for DLC 7.2 is the following Table 4.8.

The data set is then further divided in 70/15/15 training/validation/test sets respectively. All three sets are kept constant for all input combinations to have a proper comparison. Feedforward ANNs with one hidden layer and 8 hidden neurons are used, in accordance with DLC 1.2 case, see Table 4.9. 11 different initializations are done for each input combination which results in 11 ANN instances and respective performances. The median of the 11 iterations for each performance metric is shown in the Table 4.10.

Table 4.8: Examined input and target wind turbine signals for the idling operational condition DLC 7.2.

Number	Symbol	Input signal	Unit	Signal location	Standard
1	$\mathbf{\Omega_{gen}}$	Rotational velocity	rad/s	Rotor	yes
4	$\mathbf{\ddot{y}_{na}, \ddot{x}_{na}}$	Accelerations FA & SS	m/s^2	Nacelle	yes
5	$\mathbf{U^{FA}}, \mathbf{U^{SS}}$	Wind speed FA & SS	m/s	Nacelle	yes
6	$\mathbf{\ddot{y}_{tb}, \ddot{x}_{tb}}$	Accelerations FA & SS	m/s^2	Tower bottom	no
Number	**Symbol**	**Target signal**	**Unit**	**Signal location**	
1	$\mathbf{M^{FA}_{ml}}$	Bending moment FA	kNm	Monopile mudline	
2	$\mathbf{M^{SS}_{ml}}$	Bending moment SS	kNm	Monopile mudline	

Table 4.9: Parameters of the ANNs and the data set used in the input sensitivity study for the idling load case.

Number of hidden layers	Number of neurons	Activation function	Train/Validation/Test (%)
1	8	$tanh$	70/15/15
Input combinations	**Statistics**	**Initialisations per combination**	
Permutation of available inputs (Table 4.8)	All (Table 2.4)	11 – median shown in Table 4.10	

The accuracy of the results for the different input combinations highlight once again the importance of the nacelle acceleration signals. Nacelle accelerations are the most critical signals for a good estimation of the monopile moment DELs also in the idling load case. Even if only nacelle acceleration is used (input number 4) the Pearson coefficient value is $R > 0.995$ and the mean absolute error is $\epsilon_{mu,abs,test} < 3.5\%$ in both directions. The other 3 signals do not add much information and the installation of accelerometers at TB cannot be justified based on these results. It is important to notice that in the absence of nacelle accelerations, such as in the case of a damaged accelerometers, the estimation of DELs is highly inaccurate, assuming there are also no TB accelerometers.

4.6 Conclusions

From the correlation plots of the input and target statistics some useful observations have been made. Some statistics of input signals exhibit a strong linear correlation,

Table 4.10: Accuracy metrics for all input signal combinations for the idling case.

Input combination	FA			SS		
	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$
1	0.860	119.1	116.6	0.985	30.8	19.1
4	**0.995**	**3.2**	**-0.7**	**0.997**	**3.2**	**1.7**
5	0.907	48.8	53.0	0.986	22.3	11.1
6	0.994	5.7	3.1	0.996	7.3	4.1
1, 4	0.996	2.9	0.4	0.997	3.6	0.9
1, 5	0.908	55.2	56.0	0.988	20.5	11.2
1, 6	0.994	6.4	2.7	0.996	7.2	2.0
4, 5	0.996	2.8	0.7	0.997	4.0	0.0
4, 6	0.996	2.5	-0.2	0.997	3.4	1.6
5, 6	0.994	7.1	4.5	0.996	7.9	3.9
1, 4, 5	0.996	3.3	0.8	0.997	4.0	1.2
1, 4, 6	0.996	3.0	1.2	0.997	3.9	2.5
1, 5, 6	0.994	6.9	4.0	0.996	9.2	4.3
4, 5, 6	0.996	3.1	0.3	0.997	4.2	3.5
1, 4, 5, 6	0.996	3.1	0.3	0.997	5.0	1.1

such as the 2nd and 3rd spectral moments of nacelle acceleration. One of these two spectral moments can be considered superficial. On the other hand, a linear correlation between input and output quantities can offer the potential for using a simpler model, such as linear regression. In this problem a linear correlation between standard deviations σ of nacelle accelerations and DEL of monopile moments is identified. Based on this finding linear regression models are examined in Chapter 6.

In Section 4.3 conclusions are drawn for the effectiveness of various input signal and statistics combinations for the task of load estimation during normal operation conditions. In Subsection 4.3.2 it is shown that including negative spectral moments of all input signals as inputs to ANNs increases estimation accuracy. Negative spectral moments reflect mostly low frequency content of the power density spectrum. As a result, the increase in accuracy implies that low frequency content of the input signals is correlated with the DELs. In Subsection 4.3.3 is performed an examination of the permutation of all input signals with respect to the estimation accuracy. Nacelle acceleration is identified as the most important input signal. It shows the best accuracy among all cases where a single input is used, with a 4.1% mean absolute error in the FA and 3.0% in the SS. TB accelerations show a similar accuracy with a smaller error of 3.3% in FA and a higher error of 4.6% in SS. When both inputs are used the accuracy is significantly increased to 3.0% and 2.8% and no other combination of two signals can yield such a level of accuracy. The combination that uses only standard signals (1-5) also yields good accuracy results, with mean absolute error of 3.2% in FA and 3.1% in SS. it shows potential, because it does not include extra sensors than the ones that are already installed in most wind turbines.

If nacelle acceleration sensors are damaged and a combination of the remaining standard signals is used (1,2,3,5) the error would rise to approximately 15% in both directions. This fact showcases the importance of nacelle acceleration signals for OWTs. Although decent results have been achieved in the case of onshore wind turbines without using nacelle accelerations in Cosack [5], the same levels of accuracy cannot be achieved in the offshore environment, where hydrodynamic loading is important. The reason for this is that the remaining standard signals (1-3,5) contain important information about the wind induced loading but they contain little information about the wave induced loading. The combined hydrodynamic and aerodynamic loading that affects mostly the FA direction leads to a more complex behavior than in the SS, which is showcased in Section 4.4. If wave information is included as input to the estimation in the simple form of 10-minute significant wave height H_s, accuracy is increased only marginally. Finally, the idling load cases DLC 7.2 are examined and the importance of nacelle accelerations is showcased once more, since it is the only one of the standard signals that leads to accurate predictions of moment DELs. Mean absolute estimation error of less than 3.5% in both directions is achieved using only nacelle accelerations.

Chapter 5

Sensitivity studies for load estimation using ANNs

In Chapter 4 ANNs were used to establish implicit TF from statistics of input signals to DELs for monopile bending moments. In all the cases considered 70% of the data set was used for training the ANNs, resulting in a training set that properly reflected the entirety of the data set. Results about the estimation accuracy were found promising and ANNs are a good choice when a proper training data set is available. One of the limitations of all ML models, including ANNs, is that predictions in areas of the input domain, where there is no data available in the training set are not reliable. This inability of ANNs to extrapolate is showcased in Section 5.1, where several sets with increasing size are used for training and the accuracy of trained ANNs is assessed. In Section 5.2 the sensitivity of trained ANNs to deviations in the turbine's first eigenfrequency is examined. ANNs trained on a turbine with a specific eigenfrequency are tested on four wind turbines with the same characteristics, whose eigenfrequencies deviate from the original. In Section 5.3 time signals from nacelle and TB accelerometers are corrupted with artificial noise. An ANN is trained and tested on statistics of those noisy acceleration signals and its performance is compared with the respective noiseless case.

5.1 Training set sensitivity

In this Section a sensitivity study of the estimation accuracy of ANNs with respect to the size and completeness of the training data set is presented. It aims to examine how the size and level of completeness of the training set affects the estimation accuracy. The motivation for this is that an actual measurement campaign in situ is costly and can be generally carried out for a limited amount of time. The fact that a limited set of training data is used implies that it is likely that not all input combinations are contained in that particular set; input combinations which are not present in the data set are referred to as novel inputs, and are the cause for extrapolation errors. For these cases the model should not be used for making predictions.

5.1.1 Training set construction

The data set consists of 4896 10-minute simulations of the operational load case DLC 1.2. These simulations are originally used for the wind turbine fatigue design and contain a variety of environmental conditions to reflect the different loading scenarios that the turbine is expected to encounter during its lifetime. For a more detailed description the reader is referred to Section 4.1.1. Each 10-minute simulation has its own occurence weight, which is the number of times that this 10-minute scenario is expected to occur over the (design) lifetime. These weights are calculated based on the statistical distribution of the specific site's environmental conditions. The difference between this case study and Chapter 4 is that here only simulations with non-zero weights for the Westermost Rough site are included in the data set, which reduces its size from 11232 to 4896 samples.

Nine training sets with different sizes are created based on different values of a parameter that is introduced for this case study. This parameter is the return period T_R in days of the simulations. T_R is a measure of frequency of occurence and will be used to create 9 datasets with increasing size. It is defined as the ratio between the expected number of occurences of a 10-minute interval during the turbine lifetime and the complete operational lifetime of the turbine counted in 10-minute units. This gives the return period in 10-minutes, which is then multiplied with the number of 10 minutes in a day to get the return period in days. Simulations are sorted with increasing T_R and for each training dataset only simulations with T_R smaller than a certain value are considered. Specifically, return periods T_R smaller than 15, 30, 45, 60, 90, 120, 180, 360 and 720 days are considered in each case. Of these simulations that fulfil the return period criterion 70% are used as training set, 15% as validation test and the rest 15% form the test set together with the rest of the simulations that have higher return period than the limit. The method followed is shown in Table 5.1.

Table 5.1: Division of samples in training, validation and test sets depending on their individual return periods.

$\mathbf{T_R \leq T_{Rmax}}$			$\mathbf{T_R > T_{Rmax}}$		
Training	Validation	Test	Training	Validation	Test
70%	15%	15%	0%	0%	100%

The different considered return periods T_R and the resulting sizes of the training, validation and test data sets are shown in Table 5.2.

Table 5.2: Sizes of training, validation and test data sets for nine different maximum return periods T_{Rmax} of training set simulations.

Return period $\mathbf{T_{Rmax}}$ [days]	**Set Size [samples]**		
	Training	Validation	Test
15	544	117	4235
30	1025	220	3651
45	1247	267	3382
60	1399	300	3197
90	1590	341	2965
120	1701	365	2830
180	1886	404	2606
360	2218	475	2203
720	2556	548	1792

As higher return periods are considered the number of training samples increases. It should be mentioned that using the return period as the deciding parameter for creating training sets is useful for this study where simulation results are used as data. When real data from sensors are available, different training sets can be created by considering varying periods of measurements as training sets and leaving the rest as test sets. Such a study with measured data has been carried out in Seifert et al. [25], with the difference that the examined the load quantity is the flapwise blade root bending moment. The accuracy of the load estimation with ANNs is assessed as a function of the specific measurement period that is used as training set.

After the nine different data sets have been created, the number of test samples that have features outside the range presented in the training samples are calculated. A sample is considered to be outside the training range when at least one of its features is outside the range of the respective feature in the training set. The percentage of extrapolating samples of the test set size is illustrated in Figure 5.1. For each return period T_R ten seeds are drawn for the training, validation and test sets and the results are shown as statistics using box plots. From the graph it can be deduced that the percentage of test samples that are outside the training range is decreasing as the T_{Rmax} increases. In the case where simulations with $T_R \leq 1 \;\; year$ are considered, only 10% of the rest of the simulations are outside the training range. For $T_R \leq 2 \;\; years$ this rate drops to 5%. Between T_{Rmax} of 45 and 60 days, as well as between 90 and 120 days, the percentage is not decreasing, possibly indicating that only few simulations have T_R between these values. A similar study with measured data, like the one in Seifert et al [25], but for overturning moment as target load would be useful and is advised for future work.

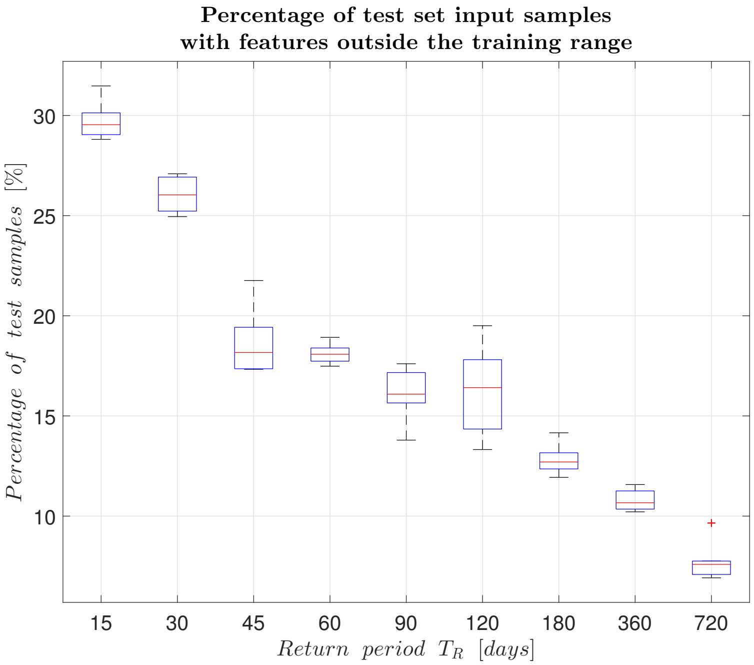


Figure 5.1: Percentage of test set samples which have features outside of the respective range presented during training (extrapolating samples). Results are given for a variety of maximum return periods T_R for training samples and as a percentage of the respective test set size (Table 5.2). Ten seeds are drawn for each T_R. Red line is the median, blue box edges are the 1st and 3rd quartiles, black lines are the min and max values, excluding outliers that are marked with a red cross.

5.1.2 Sensitivity of ANN accuracy

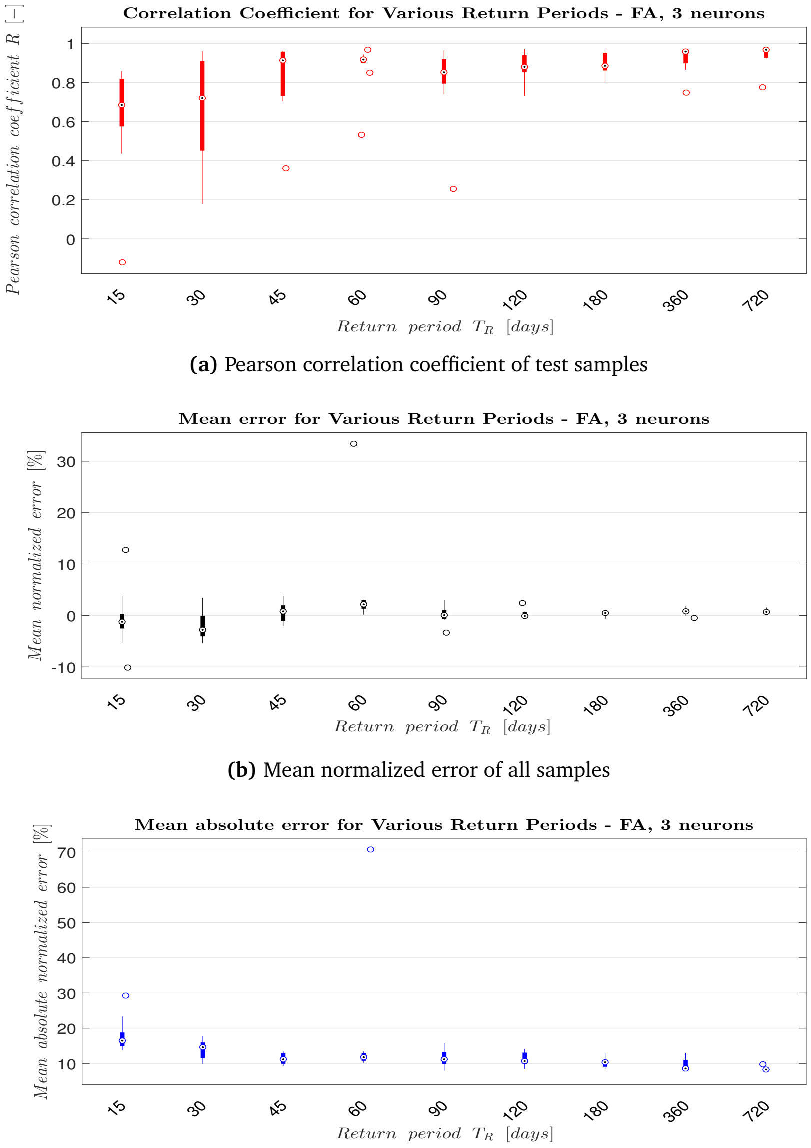
Next, ten randomly created data sets are constructed for each of the nine return periods; the obtained training sets are subsequently used for training ANNs for both the FA and SS directions. Input quantities are nacelle and TB accelerations and all statistical parameters are used for each one. The network inputs and architecture is shown in Table 5.3. Small network architectures are chosen with 3 hidden neurons in a single hidden layer. The performance is estimated on the test set and the metrics introduced in Equations 4.1, 4.2 and 4.3 are used to assess the accuracy. The results are shown in the box plots of Figures 5.2a to 5.3c. Statistics of the metrics for the 10 seeds are illustrated. As was shown in Figure 5.1, for some test samples the networks are extrapolating, but this is tolerated here in order to see the effect on the error.

The median, first and third quantile, minimum, maximum and outliers are visualized with a circle with a dot, thick line edges, thin line edges and unfilled circles

Table 5.3: Parameters of the ANNs and the data set used in the training set sensitivity study for the normal operation case.

Number of hidden layers	Number of neurons	Activation function	Train/Validation/Test [%]
1	3	$tanh$	Table 5.2
Input combination	**Statistics**	**Initialisations per $\mathbf{T_{Rmax}}$**	
4,6 (Table 2.3)	All (Table 2.4) no λ_-	10 – statistics shown in Figures 5.2a - 5.3c	

respectively. In FA direction the performance stabilises for T_{Rmax} equal and higher than 45 days, with some further improvement at the highest return periods. The correlation coefficient R converges to approximately 92%, the mean error $\epsilon_{mu,all}$ to 1.5% and the mean absolute error $\epsilon_{mu,abs,test}$ to 10%. Outliers show performance which deviates significantly with respect to the majority and are caused by the LM algorithm finding an unsuitable local optimum. In SS direction the correlation coefficient R converges to $98 - 99\%$ even from the lower T_{Rmax}, $\epsilon_{mu,all}$ converges to -2% and $\epsilon_{mu,abs,test}$ converges to 7%. The performance is as expected substantially worse in the case of an incomplete training set. Smaller training sets cover a small input range and do not properly reflect the variation in the test set; therefore predictions made with ANNs which are trained with a limited amount of input-output data are prone to extrapolation error. This is why ANNs that are trained on the smaller training sets exhibit low accuracy. This sensitivity analysis showcases the importance of good training sets for the ANNs that fully encapsulate the variety of environmental conditions that the turbine will encounter during its lifetime. To obtain such a proper training set long measurement periods are required for the target loads, in this case the mudline mending moments of the monopile. Such measurements are costly and difficult to obtain in practice; for this reason training sets created by BHawC simulations are considered as a viable alternative. The range of BHawC simulations can be extended on demand by to create larger training sets.

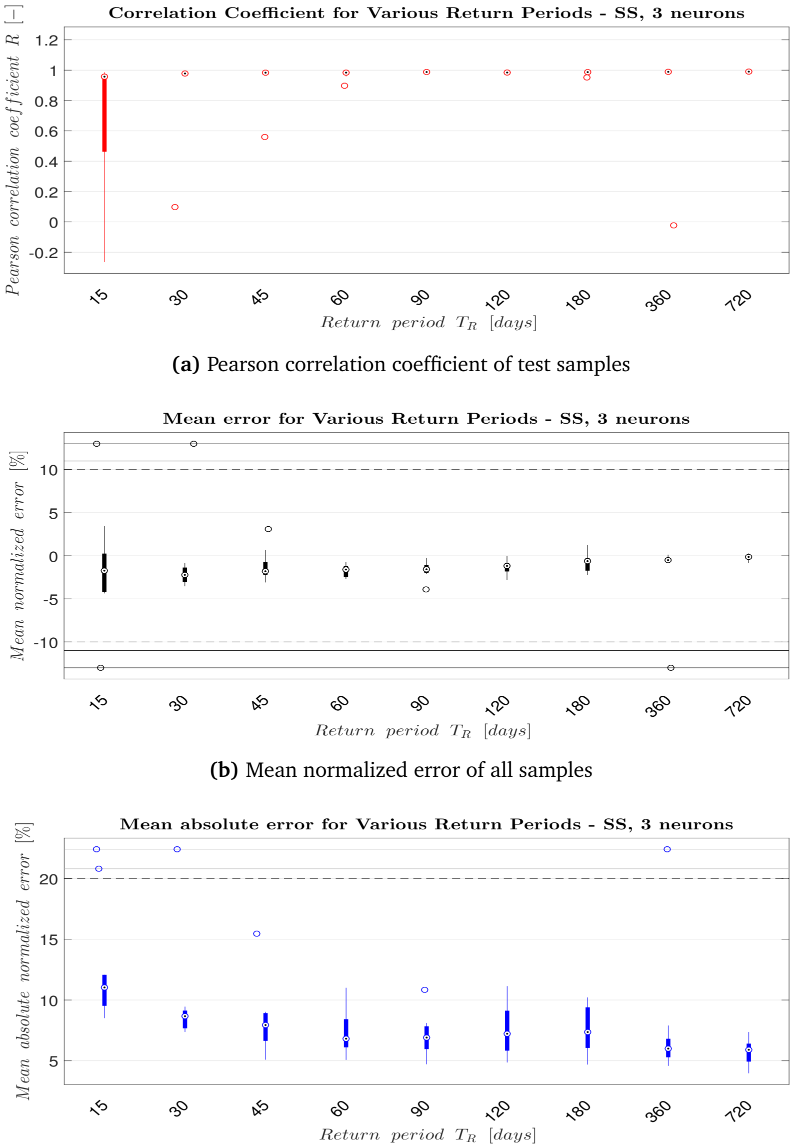


(a) Pearson correlation coefficient of test samples

(b) Mean normalized error of all samples

(c) Mean absolute normalized error of test samples

Figure 5.2: Statistics of the three performance metrics in FA direction for different values of maximum return period T_R. Ten different seeds are drawn and ten different ANNs are trained for each value of T_R. Circles with a dot are medians, thick line edges are first and third quartiles, thin line edges are min and max, circles are outliers.



(a) Pearson correlation coefficient of test samples

(b) Mean normalized error of all samples

(c) Mean absolute normalized error of test samples

Figure 5.3: Statistics of the three performance metrics in SS direction for different values of maximum return period T_R. Ten different seeds are drawn and ten different ANNs are trained for each value of T_R. Circles with a dot are medians, thick line edges are first and third quartiles, thin line edges are min and max, circles are outliers. Dotted lines indicate a change in scale in order to visualise outliers.

5.2 Turbine eigenfrequency sensitivity

A ML model that was trained on input-output observations generated from an underlying distribution should be used to make estimations for new (input) observations coming from the same distribution. For the wind turbine case, an ANN that has been trained on BHawC simulated data should be used for turbines that correspond to the specific BHawC model. Within an OWF usually the same turbine type is used and additionally turbines are clustered in several groups that share geometric characteristics of tower and foundation. As a result, it might be possible to efficiently monitor the loading of more than one turbines of a cluster using a single trained ANN. This assumes that all the specific turbines are almost identical. In practice however, soil conditions might be different from point to point within the cluster, altering the individual turbine characteristics, such as natural frequencies. For this reason, a sensitivity study with respect to the actual first turbine eigenfequency is carried out in this Section.

The normal operation load case simulations are run for five different turbine models in BHawC. The turbine type and support structure dimensions are the same in all five models and a tuning is introduced to alter the first eigenfrequency of the four models with respect to the base case f_0. This frequency shift can be correspond in practice to stiffer or softer soil conditions in a adjacent locations with the same water depth. The base case as well as the four divergent eigenfrequencies are shown in Table 5.4. The base case frequency f_0 is used to train the ANNs and its test performance is assessed. Then the trained ANNs are also tested for the other four frequencies. The datasets for all four divergent eigenfrequencies consist of input and target values for 11232 simulations. The input combinations are inserted to the ANNs trained on the base case f_0 and the outputs of the networks are compared with the correct target values that are obtained by rainflow counting. Three different ANN configurations are examined for this study, which are shown in Table 5.5. The accuracy results for the base case and the four different eigenfrequencies are illustrated in Figures 5.4a – 5.5b. Results for the three different network configurations are visualised with different markers. The study includes both FA and SS directions. The regression plots for all cases are shown in Figures B.1 – B.6 of Appendix B.

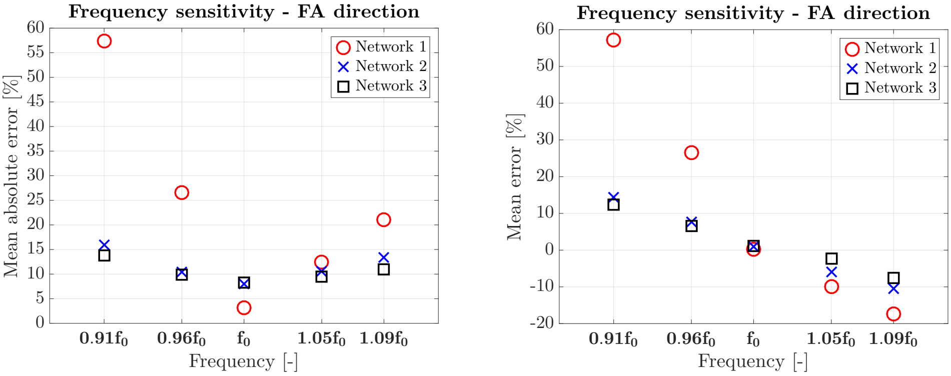
Table 5.4: Examined first turbine eigenfrequencies for the eigenfrequency sensitivity study.

Eigenfrequency	f_1	f_2	f_0	f_3	f_4
Percentage of f_0 [%]	91	96	0	105	109
Used for	Test	Test	Training & Test	Test	Test

The load estimation on turbines with diverging eigenfrequencies from the one used for training exhibits higher errors than the load estimation in the original turbine. Both the mean absolute error and the mean error are higher for the four eigen-

Table 5.5: The three different network configurations that are examined for the eigenfrequency sensitivity study.

Network	**1**	**2**	**3**
Hidden layers	1	1	1
Hidden neurons	8	8	4
Input signals (Table 2.3)	$1-5$	$1-5$	$1-5$
Statistics (Table 2.4)	All	All for $1-3,5$, only σ for 4	All for $1-3,5$, only σ for 4



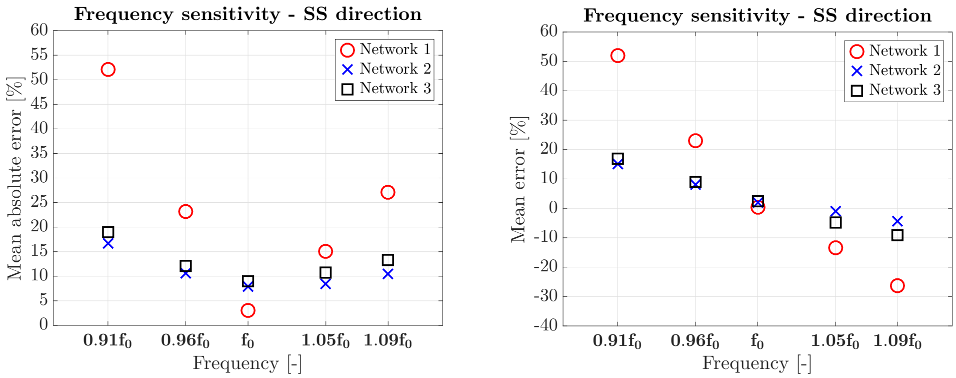
(a) Mean absolute error

(b) Mean (residual) error

Figure 5.4: Performance metrics for the eigenfrequency sensitivity analysis in the FA direction. Three different network configurations and five eigenfrequencies are examined. Errors are computed between output and target values.

frequencies than for f_0. The higher the deviation from f_0, the higher are the errors. By examining the mean (residual) error it is observed that DELs are consistently overestimated by the networks for lower eigenfrequencies, while for higher eigenfrequencies the DELs are underestimated. Additionally, estimation for lower frequencies shows lower accuracy than for higher frequencies in all cases.

As far as the performance of the individual networks is concerned, Network 1 has the best estimation performance for the f_0 case but its performance is by far the lowest for the other four frequencies. Network 1 includes as inputs the spectral moments of nacelle accelerations, which are highly correlated to the target loads and for this reason it has the best performance for the base case. Networks 2 and 3 do not include these inputs, thus they are less accurate. However spectral moments are very sensitive to changes in the turbine's natural frequency, because in Equation A.4 for the computation of spectral moments frequencies are weighted exponentially. This is what causes the high error of Network 1 compared to Networks 2 and 3 that do



(a) Mean absolute error

(b) Mean (residual) error

Figure 5.5: Performance metrics for the eigenfrequency sensitivity analysis in the SS direction. Three different network configurations and five eigenfrequencies are examined. Errors are computed between output and target values.

not use spectral moments. Thus, there is a trade–off between accuracy, when the eigenfrequencies that correspond to the training coincide with the eigenfrequencies of estimation and high errors when these frequencies actually deviate. When deviation in the actual eigenfrequencies is expected, it is preferable to use Network 2 or 3. The difference between Networks 2 and 3 is the number of hidden neurons, which is smaller for Network 3. Network 2 is performing slightly better in the FA direction and Network 3 slightly better in the SS direction, so no clear distinction between the two is made in terms of performance.

5.3 Input signal noise sensitivity

This Section examines the influence of noise on the accuracy of ANNs when training, validation and test samples come from noisy signals. This includes understanding how noise affects the various 10-minute statistics of the time signals and assessing the reduction in estimation accuracy. In practice different wind turbine measured signals contain different noise levels. For simplicity only acceleration signals of the normal operation case (DLC 1.2) have been considered for this study, both nacelle and tower bottom. The acceleration signals have been contaminated with artificial, Gaussian, white noise with zero mean and standard deviation $\sigma_{noise} = 0.01 m/s^2$. The value of σ_{noise} is kept constant for all 10-minute simulations and corresponds to a noise–to–signal ratio between 3% and 20%. Smaller ratios correspond to stronger acceleration signals. Figure 5.6 shows how the different statistics of the FA nacelle acceleration are affected by the noise contamination.

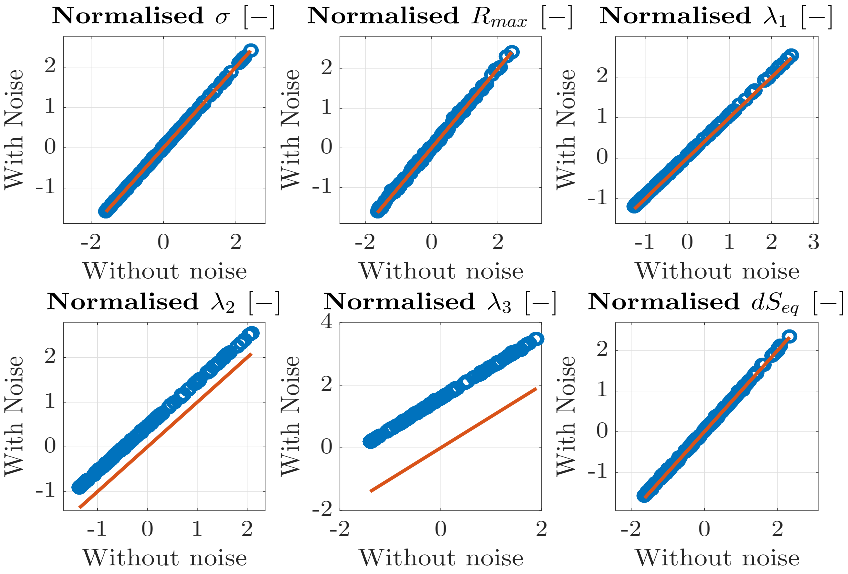


Figure 5.6: Influence of artificial Gaussian white noise on statistics of FA nacelle acceleration. Noise standard deviation is $\sigma_{noise} = 0.01m/s^2$. Blue points are the noisy statistics and the orange line shows the noiseless statistics.

Standard deviations are shifted slightly towards higher values in the noisy cases. The magnitude of the shift depends on the noise–to–signal ratio and is generally minimal for the specific ratio. Maximum range R_{max} and equivalent range dS_{eq} are also shifted slightly either upwards or downwards. The most sensitive statistics to noise are spectral moments λ_n. Higher order moments are affected more, as can be seen in the case of the λ_3. This indicates that if an ANN is trained on noiseless data using these noise–sensitive statistics and then noisy data are used for the estimation, inaccuracies can be expected. A measure to mitigate this phenomenon can be to also corrupt with similar noise the training set. The effectiveness of this method is tested here. Two ANNs are trained on the same input signals and statistics and the same training, validation and test sets are used. In the first case the signals are noiseless, whereas in the second case the signals are corrupted with the noise format described previously. The network properties are described in Table 5.6 and the accuracy results are shown in Table 5.7.

Judging from the results, the accuracy of the estimation does not deteriorate significantly for the particular noise format and noise–to–signal ratios. Statistics that are shifted in a consistent manner by signal noise, such as standard deviation and spectral

Table 5.6: Parameters of the ANNs, data set and noise used for the input noise sensitivity study for the normal operation case.

Number of hidden layers	Number of neurons	Activation function	Train/Validation/Test [%]
1	8	*tanh*	70/15/15
Input combination	**Statistics**	**Noise properties**	
4,6 (Table 2.3)	All except for μ and λ_{-} (Table 2.4)	Gaussian, white, noise–to–signal ratio:3% – 20%	

Table 5.7: Performance metrics of the load estimation for the input noise sensitivity study. Comparison between ANNs trained and tested on noiseless and noisy inputs.

Input signals	FA			SS		
	$R[-]$	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$	R[-]	$\epsilon_{mu,abs,test}[\%]$	$\epsilon_{mu}[\%]$
Noiseless	0.986	5.4	0.4	0.997	3.5	0.3
Noisy	0.985	5.6	0.7	0.997	3.9	0.3

moments are not expected to drastically decrease the network accuracy. When noisy data are used for training, the network learns the correlation between shifted inputs and targets. If the inputs of the estimation phase are shifted in the same way as in the training phase the network estimates can be accurate.

5.4 Conclusions

The importance of a proper training set for ANNs is showcased in Section 5.1. Good training sets for the should reflect the entire variation of environmental conditions that the turbine will encounter during its lifetime. Small training sets can lead to a big number of novel iputs during the estimation phase, for which estimations should not be taken into account because they would introduce significant errors due to extrapolation. Since it is difficult to obtain a proper training set by actual measurements, using BHawC simulations to create the training set seems like a promising alternative.

If the turbine whose loads are monitored has different natural frequency from the one corresponding to the training set, significant errors can be introduced, depending on the magnitude of the deviation. The decrease in accuracy also depends on the inputs used for the ANNs; spectral moments are statistics highly sensitive to shifts in the natural frequency and should be ommitted from the estimation procedure if there is high uncertainty about the natural frequency. Furthermore, there is a consistent trend of load overestimation for lower frequencies and load underestimation for higher frequencies. When statistics of acceleration signals contaminated with Gaussian, white noise of noise–to–signal ratio between 3% and 20% are used to train and test an ANN, the performance does not deteriorate significantly with respect to the noiseless case. This conclusion can be used to contaminate BHawC simulation results with noise according to the properties of the turbine sensors in order to create a training set that has similar noise properties with the measured signals. An alternative to this approach can be to denoise the measured signals instead.

Chapter 6

Linear Regression Models

Section 6.1 presents a brief overview of the linear regression mathematical model. In Section 6.2 a training set corresponding to a virtual measurement campaign of limited duration of two weeks is created. The remaining samples form the test set, which is 3 times bigger than the training set and contains many samples whose inputs are outside of the training range. A LR model is made, with inputs the standard deviations of nacelle accelerations in FA and SS directions and its accuracy of estimating mudline moment damage equivalent loads is assessed. In Section 6.3 again LR models are used with the same inputs, but in this case the samples are a priori divided in three bins based on their 10-minute mean wind speed: below, around and above rated. Consequently, six LR models were used, one for each bin and direction. Section 6.4 introduces a new signal as potential input to the estimation: TB rotation angle. First, its high correlation with the target signal (monopile bending moment) is shown. Then the 10-minute equivalent range of the rotation signals is used as input to the piecewise linear regression scheme, in addition to standard deviations of nacelle accelerations, resulting in high estimation accuracy for this difficult data set.

6.1 Linear regression theory

According to the definition in [7], regression is described as the process of infering the relationship between some independent and one or more dependent variables. Considering $x \in \Re^d$ the input with dimension d, each of the d components represents one independent variable (also named in literature covariate, feature etc.). The output $y \in \Re$ is a variable dependent on x and is generally called the ”response”. A *regression function* is a function $f : \Re^d \rightarrow \Re$ for the data pair (x, y), such that $y \approx f(x; w)$. The free parameters of the function are the weights w and contain the model information. The regression model is called linear if the prediction function f is a linear function of the unknown weights w. The basic form of the linear model used in this work has the form:

$$y_i \approx f(x_i; w) = w_0 + \sum_{j=1}^{d} x_{ij} w_j, \quad i = 1, ..., n \tag{6.1}$$

The term w_0 is called bias, offset or intercept and w_j can be interpreted as the slope of the function at the respective dimension j. The total number of samples is equal to n. The training dataset (x, y) will be used to find a weight vector w such that $y_i \approx f(x_i; w)$. For that reason the least squares (LS) objective function L is introduced in order to find the weight vector w that minimises the sum of squared errors (SSE).

$$L = \sum_{i=1}^{n} (y_i - f(x_i, w))^2 \tag{6.2}$$

The LS solution w_{LS} for the linear regression problem is defined as [31]:

$$w_{LS} = arg \min_{w} L = arg \min_{w} \sum_{i=1}^{n} (y_i - f(x_i, w))^2 \tag{6.3}$$

Using matrix notation, the input matrix X with dimensions $(n \times d+1)$, the output column array y with dimensions $(n \times 1)$ and the weight column array w with dimensions $(d \times 1)$ have the following form.

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1d} \\ 1 & x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix}$$

$$y = [y_1, y_2, \dots, y_n]^T$$

$$w = [w_0, w_1, \dots, w_d]^T$$

The objective function L of Equation 6.2 then takes the following form:

$$L = \sum_{i=1}^{n} (y_i - f(x_i, w))^2 = \|y - Xw\|^2 = (y - Xw)(y - Xw)^T \tag{6.4}$$

Setting the gradient of the objective function L with respect to the weight vector w equal to zero, the w_{LS} that minimizes L is obtained:

$$\nabla_w L = 2X^T X w - 2X^T y = 0 \Rightarrow \quad w_{LS} = \left(X^T X\right)^{-1} X^T y \tag{6.5}$$

In order to calculate w_{LS} according to Equation 6.5 the inverse matrix $\left(X^T X\right)^{-1}$ must exist. The matrix $\left(X^T X\right)^{-1}$ exists when the matrix $X^T X$ is of full rank, which means that the matrix X with dimensions $n \times d+1$ has *at least d+1 independent rows*.

If $n < (d+1)$ then LS regression is not possible, since if $\left(X^T X\right)^{-1}$ does not exist there are infinite possible solutions for the objective function L minimisation problem.

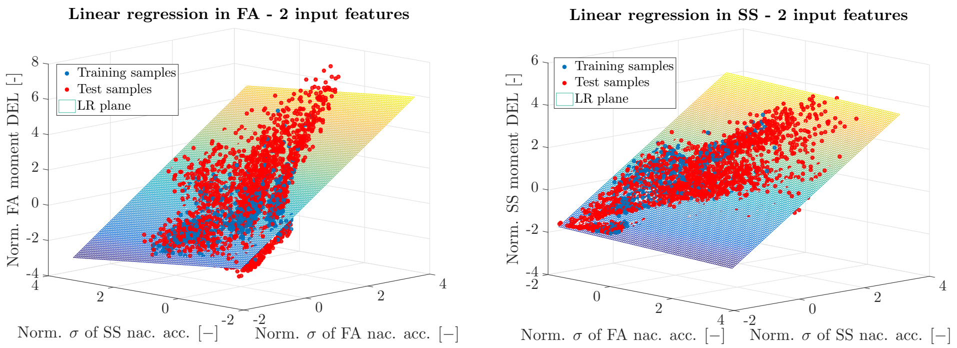
6.2 Estimating loads with linear regression

In Figures 4.2a and 4.2b the linear correlation between the DEL of the monopile MBM and the standard deviation of the nacelle acceleration in the respective direction is exhibited. This linear correlation does potentially allow for a linear regression scheme to be used for estimating DELs. Probably this LR method will result in lower accuracy than an ANN, but it is a significantly simpler model and the weights $\mathbf{w}$ that constitute the model can be directly associated with the respective input, such as the acceleration std. This allows for useful conclusions to be extracted regarding the input-output correlation. This is much harder to achieve in ANNs, since the network weights do not have a direct physical interpretation, especially in networks with multiple hidden layers.

6.2.1 Data set for Linear Regression

In order to examine the effectiveness of the LR method, the simulation set for Westermost Rough that was described in Section 5.1 is used. It consists of 4896 simulations of DLC 1.2. Each of those simulations is associated with a occurence weight, which denotes how often it occurs over the OWT lifetime. To make things more challenging and assess how well the LR model extrapolates, the training set is created based on a virtual measurement campaign of just 2 weeks duration. These 2 weeks correspond to 2160 10-minute intervals. As a result, 2160 samples are drawn out of the 4986 simulations according to the occurence weights and with replacement of the drawn simulations (they can be drawn again). In practice this method is drawing samples according to the specific distribution that generated the occurence weights. Seasonal dependencies and correlations between subsequent 10-minutes are not taken into account.

Of the 2160 drawn samples, duplicates are removed in order to have only unique samples in the training set. If the duplicates were not removed they would induce a bias for better estimation of these duplicate smples. The removal of duplicates results in a training set of 1264 samples, while the remaining 3632 samples constitute the test set, since no validation set is required for simple LR. The training set is 3 times smaller than the test set, which is used to assess the performance and generalization of the LR model when limited data is available and extrapolation is necessary. In other words, it is assessed whether the linear correlations that are obtained by the LR model are valid over the whole range of the operational conditions.



(a) FA direction

(b) SS direction

Figure 6.1: Linear regression in FA and SS directions for the vrtual measurement campaign data set. Inputs are standard deviation σ of $\ddot{y}_{na}$ and $\ddot{x}_{na}$. Blue points stand for training set samples, red datapoints stand for test set samples. The linear regression model is illustrated by the coloured linear plane.

6.2.2 Linear Regression results

A linear regression model can be fit with an arbitrary number of input variables. In Figures 6.1a and 6.1b two input variables have been used for the linear regression: the standard deviation σ of nacelle accelerations $\ddot{y}_{na}$ and $\ddot{x}_{na}$. These two statistics show the highest correlation with the moment DELs at the mudline. From the two Figures it can be seen that a trained LR model is represented by a linear plane in the three dimensional space. These linear planes can be compared with the nonlinear planes of the ANN models that are shown in Figures 4.3 and 4.4. The LR model is evaluated on the data set described in Subsection 6.2.1. It is trained on the training set samples that are depicted by blue points and then is tested on the test samples, depicted by red points in the Figures. The coloured plane shows the LR model output for the specific domain of the two inputs, which is the plane that minimizes the sum of the squared prediction errors for the training data set. The details of this LR model are given in Table 6.1. The performance of the LR model is summarised with the use of the performance metrics in Table 6.2:

Table 6.1: Information on the input signals and statistics for the single linear regression model and the data set used (normal operation case).

Input combination	Statistics	Train/Test [samples]
4 (Table 2.3)	σ (Table 2.4)	1264/3632

From the results it can be deduced that the LR model provides for the FA and SS

Table 6.2: Accuracy metrics of the linear regression model for the virtual measurement campaign data set. Inputs are standard deviations σ of $\ddot{y}_{na}$ and $\ddot{x}_{na}$. Results are shown for both training and test sets.

Direction	Training			Test		
	$R[-]$	$\epsilon_{mu,abs}[\%]$	$\epsilon_{mu}[\%]$	$R[-]$	$\epsilon_{mu,abs}[\%]$	$\epsilon_{mu}[\%]$
FA	0.913	11.2	1.6	0.941	14.1	4.5
SS	0.987	13.4	1.0	0.986	13.4	0.1

directions mean absolute errors $\epsilon_{mu,abs} > 10\%$ for the specific training and test data sets. A fair comparison cannot be made with the performance of an ANN because the two methods were used on different data sets. The errors are higher than in the case of an ANN, for which the full data set of 11232 simulations were used and the training set was properly distributed over the whole range of the simulations. However, considering the fact that for the present LR model the training set is significantly smaller and the model is extrapolating for a big number of test samples, the method is considered promising. It will be further examined and improved by considering a piecewise linear regression scheme in Sections 6.3 and 6.4.

6.3 Piecewise linear regression – binning samples based on mean wind speed

A piecewise linear regression scheme [15] is examined in this Section. Piecewise regression is a method in regression where independent vaiables are clustered in different groups and a separate regression model is fit in each cluster. A simple example is the approximation of a nonlinear curve with several linear ones. Here the samples will be divided in three clusters according to the 10-minute mean wind speed. Using the a priori knowledge of the control system for the blade pitch, which is distinct for wind speeds below and above the rated wind speed, the simulations can be divided in three bins. Each bin contains simulations with 10-minute mean wind speed below, around and above the turbine rated wind speed, according to Figure 6.2. This is a deterministic division of simulations and the term binning will be used, as opposed to clustering, where the various clusters are determined based on a statistical technique. This binning on mean wind speed affects both training and test sets and three different LR models will be fit to the training data of each bin and evaluated on the respective test sets. The reason for this is to make a separate model for each operational range that exhibits distinct behavior, than one model for all.

The same input features as in the previous global LR model have been considered and are given in Table 6.3. The results are illustrated with the use of regression plots

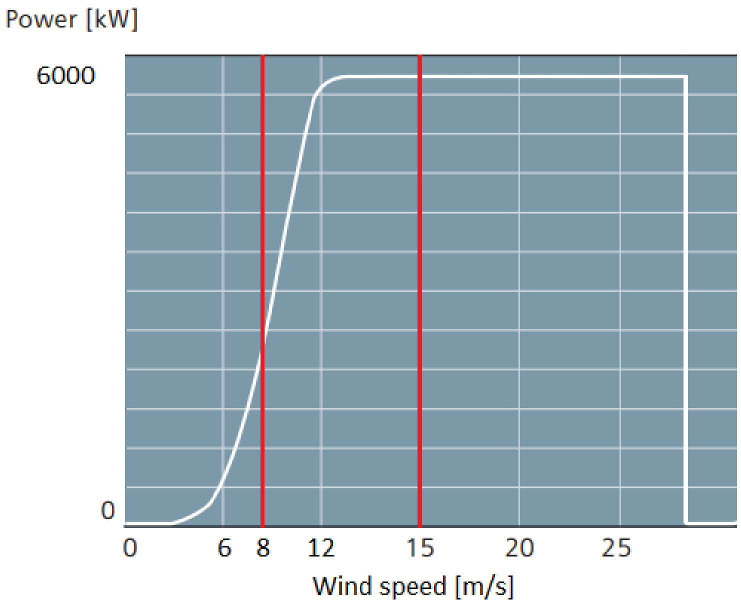


Figure 6.2: Illustration of the below, around and above rated operation domains on the turbine power curve. Below rated includes wind speeds between 3 and 8 m/s, around rated $9 - 15$ m/s and above rated $16 - 28$ m/s.

in Figures 6.3a to 6.3f and 6.4a to 6.4f for the FA and SS directions. The performance metrics are summarised in Table 6.4. The mean absolute errors for the binned case are generally lower than for the case of a single LR model, whose results are shown in Table 6.2. The only exception is the below rated domain in FA direction, which has a high mean absolute error of 15% for the test set and also a high mean error of 10%. Estimation in SS direction is consistently more accurate than in the FA direction. It is interesting that the above rated domain, where the training set is smaller, has low residual (mean) errors and the peformance is better compared to the below rated domain. In Section 6.4 the same approach of piecewise linear regression will be followed, but with statistics of the tower bottom rotation signal as inputs to the LR models.

Table 6.3: Information on the input signal, statistics and binning for the piecewise linear regression model and the data set used for the normal operation case.

Input combination	Statistics	Train/Test [samples]	Number of bins
4 (Table 2.3)	σ (Table 2.4)	1264/3632	3, according to Figure 6.2

Table 6.4: Accuracy of the piecewise linear regression model with simulations binned into three intervals based on the 10-minute mean wind speed. Input quantities are standard deviations σ of $\ddot{y}_{na}$ and $\ddot{x}_{na}$. The virtual measurement campaign data set is used.

Direction	Operation	Training			Test		
	wrt. rated	$R[-]$	$\epsilon_{mu,abs}[\%]$	$\epsilon_{mu}[\%]$	$R[-]$	$\epsilon_{mu,abs}[\%]$	$\epsilon_{mu}[\%]$
FA	Below	0.921	10.6	1.8	0.895	15.3	9.4
	Around	0.794	10.7	1.5	0.843	9.9	1.5
	Above	0.820	11.6	1.7	0.848	12.6	2.1
SS	Below	0.992	9.5	2.3	0.992	11.7	2.2
	Around	0.990	9.6	3.0	0.988	9.6	2.9
	Above	0.983	8.3	2.0	0.980	8.5	-0.3

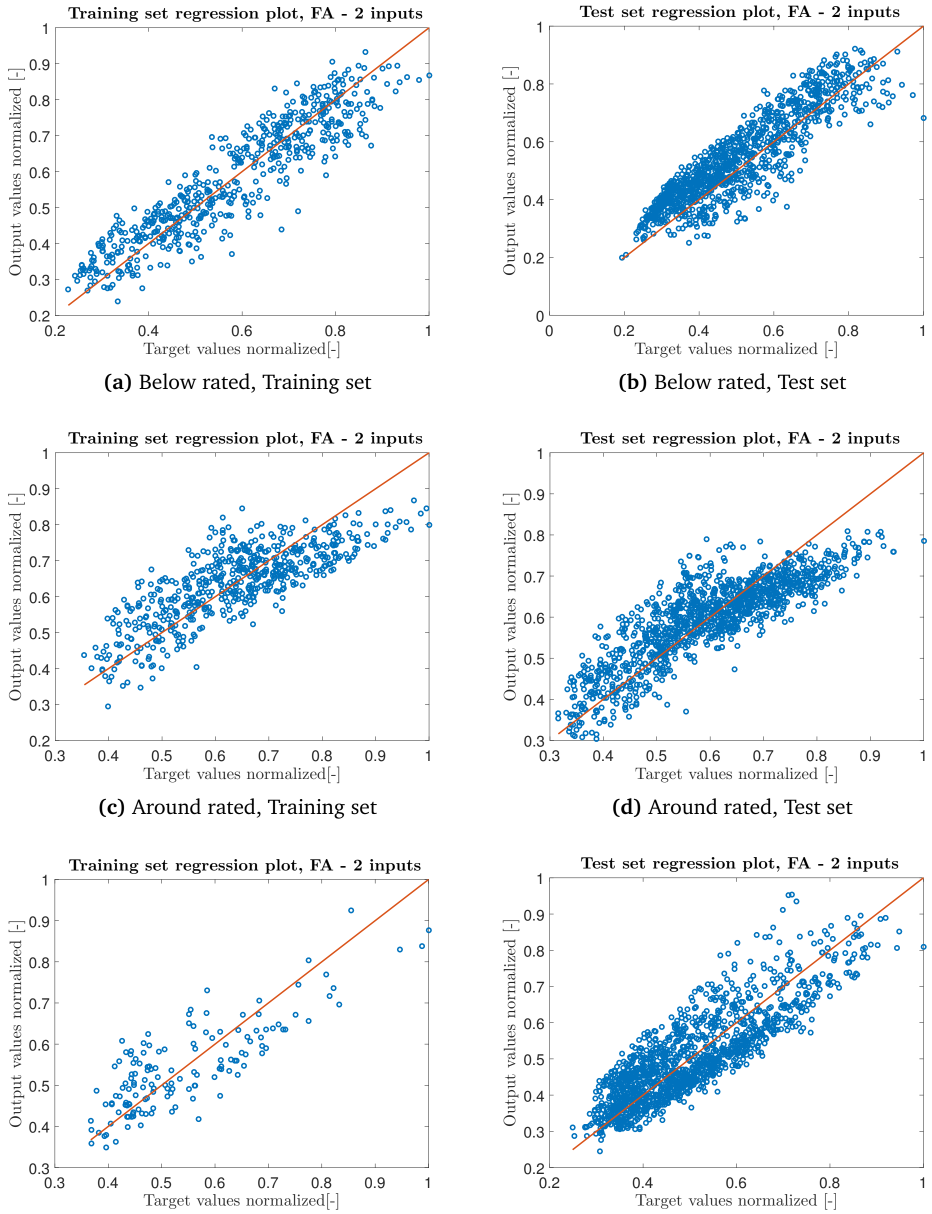
6.4 Rotation angle: an interesting signal for load monitoring

Accelerometers are maybe the most commonly used sensors for monitoring wind turbine structural response. They measure the (translational or rotational) acceleration of the structure at the location where they are installed. Acceleration signals provide valuable information for Operational Modal Analysis among others. However for load monitoring purposes, when comparing response time signals to find correlations it might be interesting to consider signals like displacements and rotations in addition to acceleration signals.

6.4.1 Correlation of input – target time signals

In the case examined in this work, the target quantity for the load monitoring is the bending moment of the monopile at the mudline level (MBM). This terget signal is highly correlated with displacements and rotations at different levels of the monopile and tower. The following Figures 6.5a - 6.5b, 6.6a - 6.6b and 6.7a - 6.7b show the correlation of the rotations, displacements and displacement accelerations at TB respectively with the MBM. Two perpendicular directions are considered in all cases (substructure fixed coordinates according to BHawC notation).

From the presented figures it can be concluded that there is a strong correlation between rotations and displacements at the TB and MBM. Each of the former two signals are very similar with the moment signal when normalized. This correlation stands for the whole range of the operational conditions as it is described by the 4896 10-minute simulations of DLC1.2 for the Westermost Rough OWF. This means



(a) Below rated, Training set

(b) Below rated, Test set

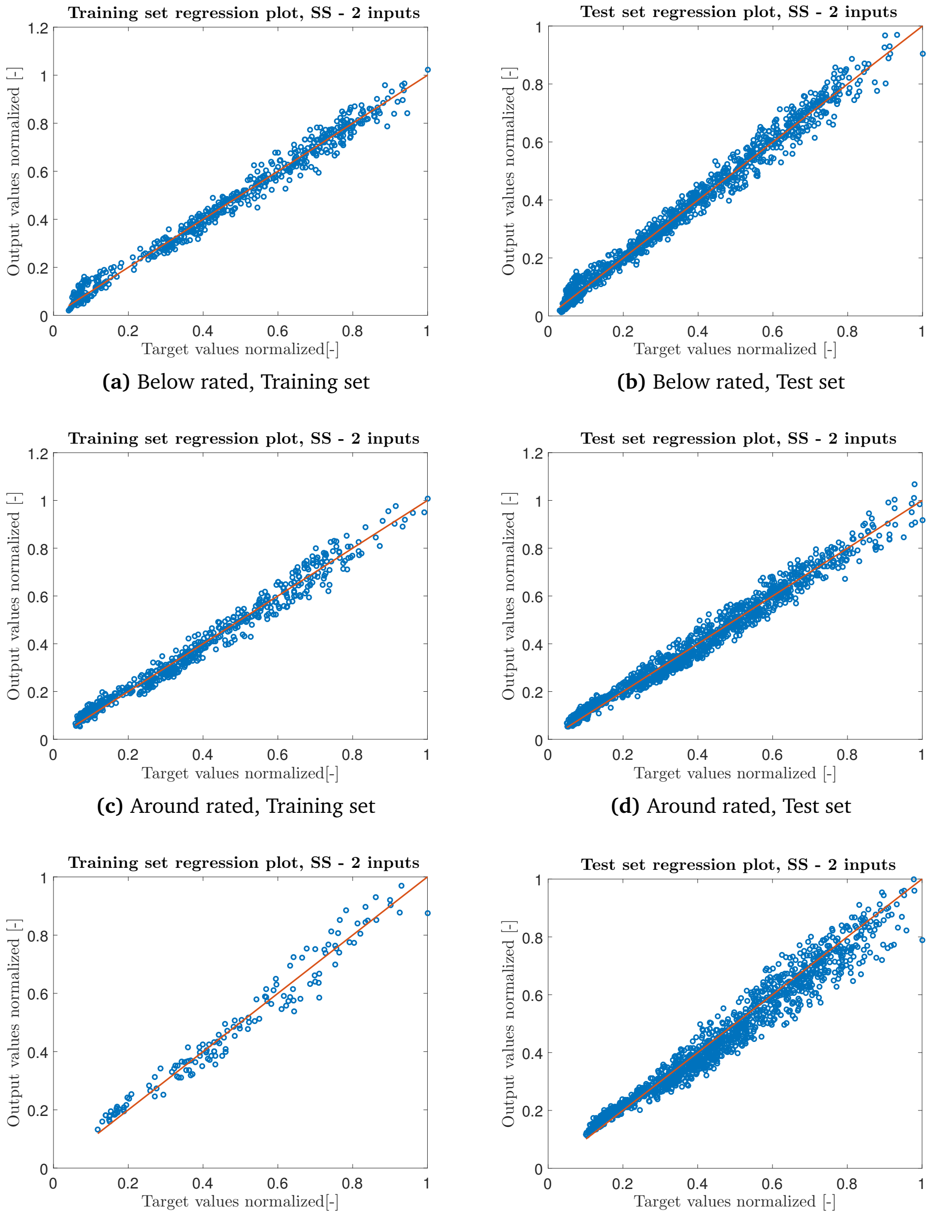
(c) Around rated, Training set

(d) Around rated, Test set

(e) Above rated, Training set

(f) Above rated, Test set

Figure 6.3: Linear regression in FA direction using σ of $\ddot{y}_{na}$ and $\ddot{x}_{na}$ as inputs. Operation is divided in below, around and above rated wind speed. Results are shown for both training and test sets.



(a) Below rated, Training set

(b) Below rated, Test set

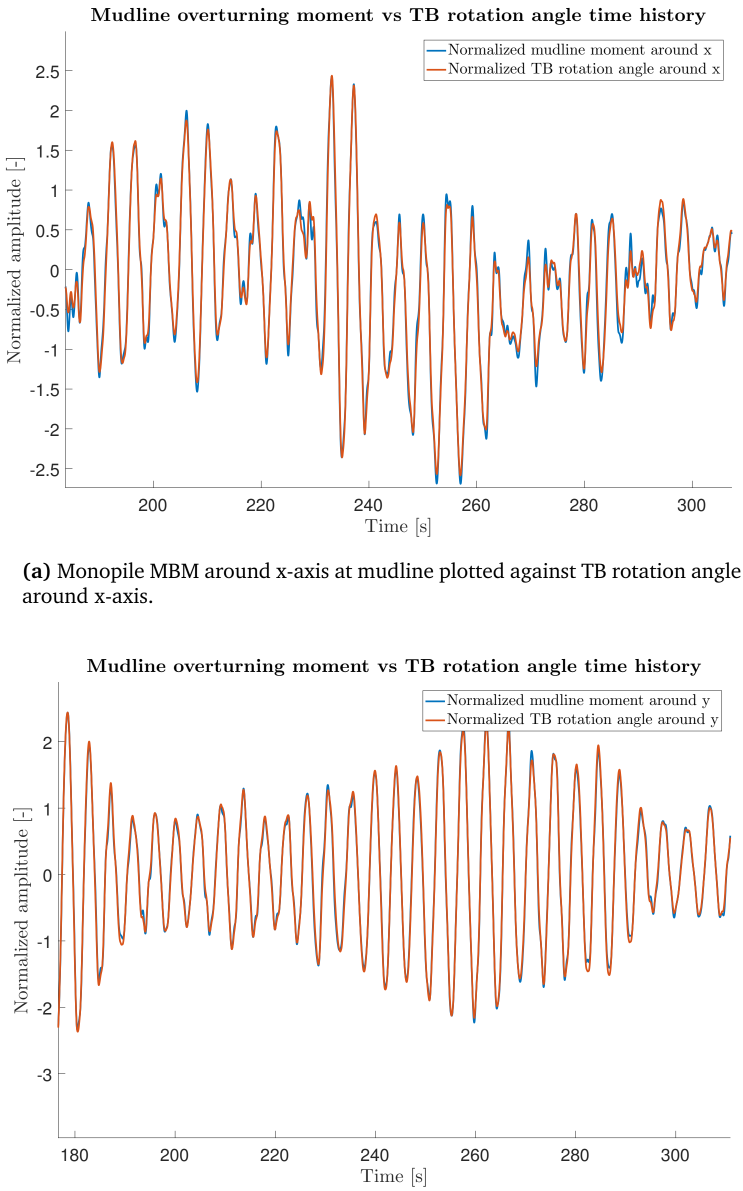
(c) Around rated, Training set

(d) Around rated, Test set

(e) Above rated, Training set

(f) Above rated, Test set

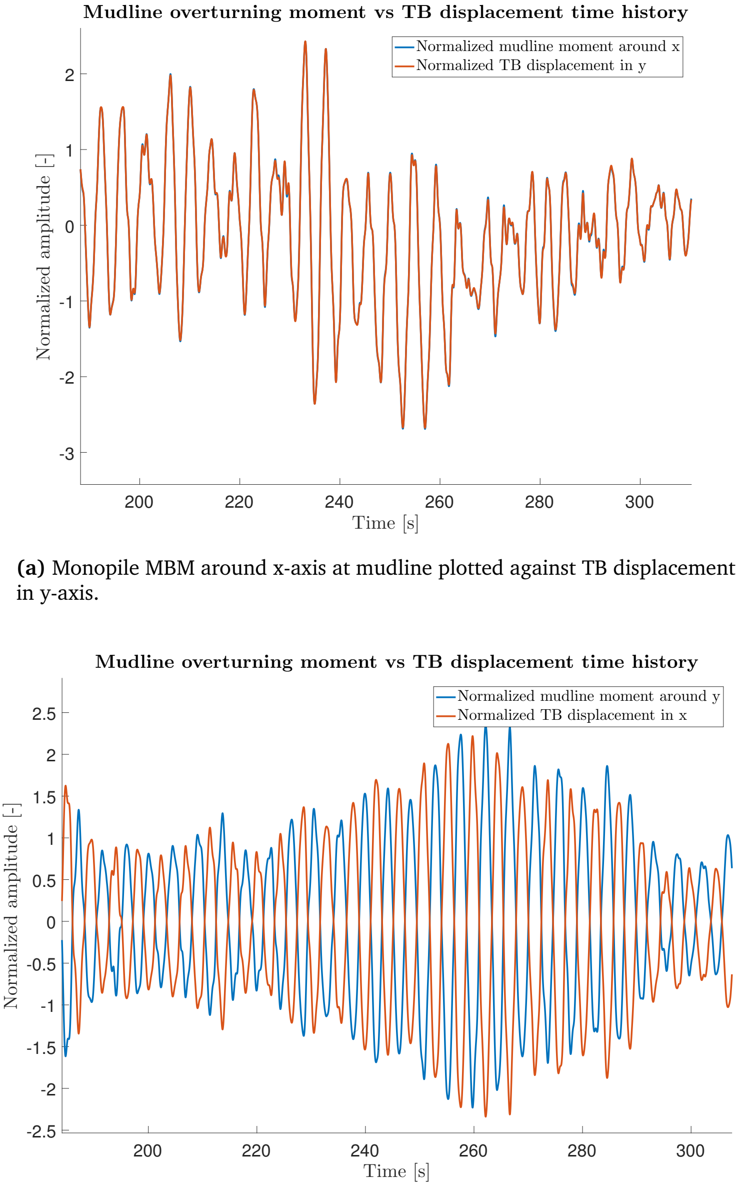
Figure 6.4: Linear regression in SS direction using σ of $\ddot{y}_{na}$ and $\ddot{x}_{na}$ as inputs. Operation is divided in below, around and above rated wind speed. Results are shown for both training and test sets.



(a) Monopile MBM around x-axis at mudline plotted against TB rotation angle around x-axis.

(b) Monopile MBM around y-axis at mudline plotted against TB rotation angle around y-axis.

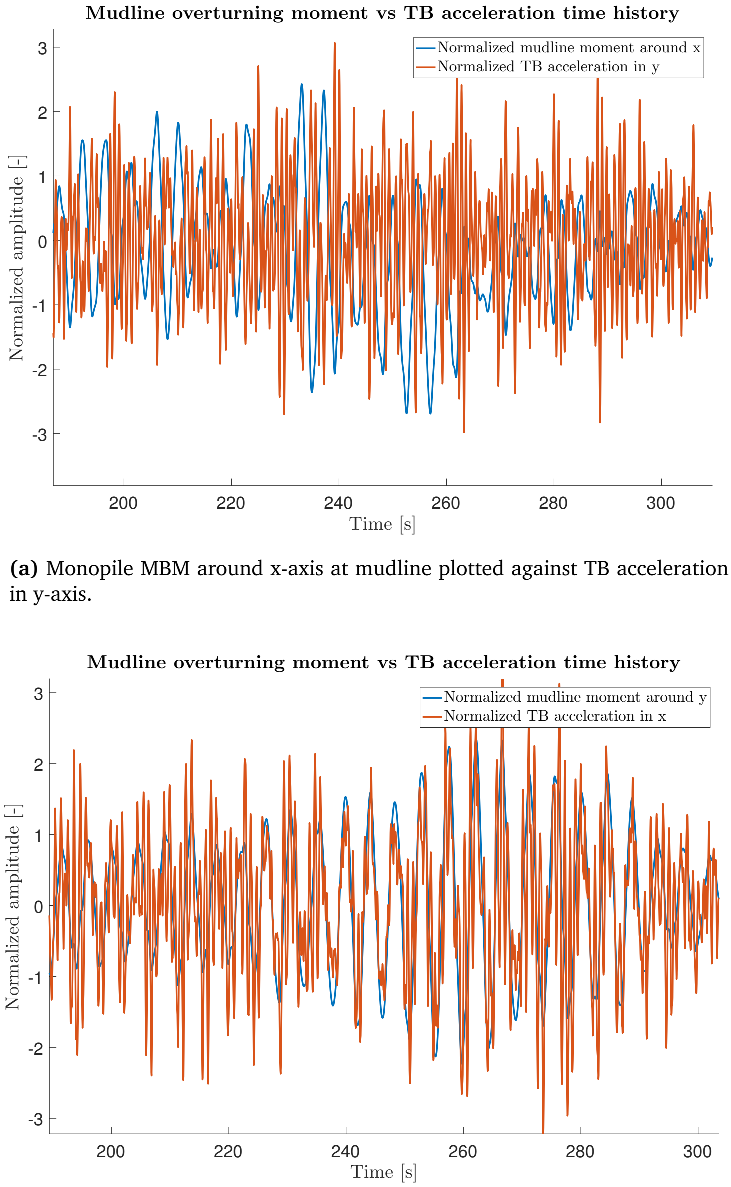
Figure 6.5: Comparison of normalised MBM and TB rotation angle time signals in two perpendicular directions.



(a) Monopile MBM around x-axis at mudline plotted against TB displacement in y-axis.

(b) Monopile MBM around y-axis at mudline plotted against TB displacement in x-axis.

Figure 6.6: Comparison of normalised MBM and TB displacement time signals in two perpendicular directions.



(a) Monopile MBM around x-axis at mudline plotted against TB acceleration in y-axis.

(b) Monopile MBM around y-axis plotted against TB acceleration in x-axis.

Figure 6.7: Comparison of normalised MBM and TB translational acceleration time signals in two perpendicular directions.

that an acquisition of those signals by means of a suitable sensor can provide accurate information on the MBM, which is a highly inconvenient spot for deploying and maintaining sensors. On the other hand, acceleration signals are not clearly correlated with moment signals. Higher frequencies are significantly more dominant in acceleration signals than in moments.

Methods for recording displacements are generally too expensive and are not preferred in the wind industry. The option to record accelerations and perform double integration in order to obtain the respective displacements is also problematic in practice, since the discrete integration scheme and the noise of the measured accelerations usually lead to significant errors, such as the phenomenon called ”drifting”. On the other hand, rotation angle is a signal that can be captured with standard sensors such as inclinometers. However in practice the important characteristics of these sensors, such as precision and durability have to be considered and compared to alternatives. If the sensor precision is low it can induce significant errors in the estimation.

An alternative signal that can be considered for providing information on the MBM is the TB moment. It can be measured with the use of strain gauges, which measure strain at that level, which is transformed to normal stress and subsequently moment of the monopile section. This method is proposed in Ziegler et al [33], where MBM DELs are estimated by measuring moments at TB, taking their DEL and using a nearest neighbor scheme to make the estimation. The time series of normalized MBMs and TB moments are plotted in Figure 6.8. Figures 6.9a and 6.9b show the cross-correlation of the normalised MBM time signal with the normalised TB rotation and the TB bending moment time signals respectively [28]. The cross-correlation of MBM with TB rotation is 5.3% higher than with TB moment for the zero lag case. As a result, if the two sensor types are assumed similar in terms of cost, performance and durability, it might be preferable to measure rotations than strains.

6.4.2 Piecewise linear regression including tower bottom rotation statistics as inputs

In this Subsection a piecewise linear regression approach is followed with the same binning method based on mean wind speed, as in Section 6.3. The rotation signals and specifically the 10-minute equivalent rotation ranges ΔS_{eq}^{m} in the FA and SS directions are inserted as inputs in the estimation procedure, in addition to the standard deviation σ of FA and SS nacelle accelerations, as is shown in Table 6.5. The performance metrics for the piewise LR model are summarised in Table 6.6. Additionally, the results are illustrated in the regression plots in Figures 6.10a to 6.10f and 6.11a to 6.11f for the FA and SS directions.

The mean absolute error is $\epsilon_{mu,abs} < 2.5\%$ in all operation domains and both

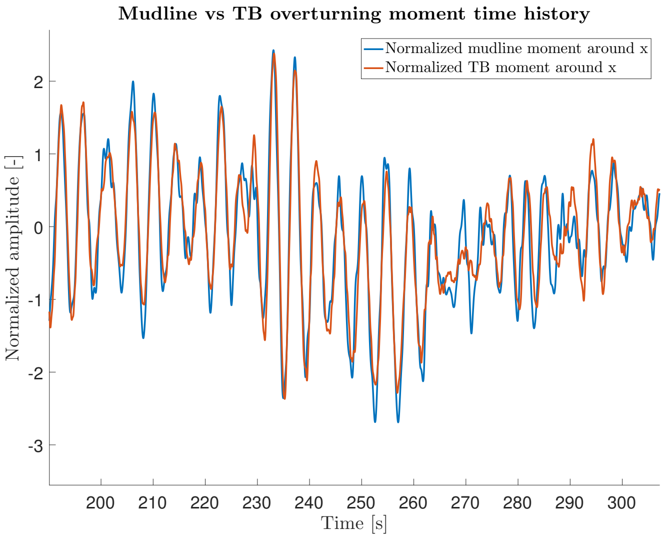
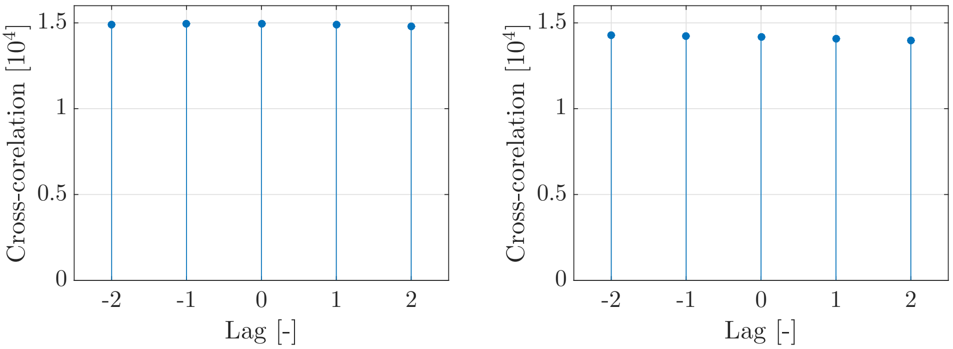


Figure 6.8: Comparison of normalised MBM and TB moment time signals around x-axis.



(a) Cross-correlation of normalised MBM with normalised TB rotation.

(b) Cross-correlation of normalised MBM with normalised TB moment.

Figure 6.9: Cross-correlation of normalised MBM around time signal with normalised TB rotation and moment time signals. 5 lag values considered. All signals are around x-axis in the monopile coordinate system.

directions. Additionally, the residual errors have small values. This performance is better than the one achieved with ANNs in Section 4.3, although a smaller training set and a simpler LR model is used. This fact indicates the importance of input features

Table 6.5: Input signals, statistics and binning for the piecewise linear regression model when statistics of rotation signals at tower bottom are included as inputs. The size of the virtual measurement campaign data set is also shown.

Input combination	Statistics	Train/Test [samples]	Number of bins
4 (Table 2.3) and TB rotations	σ (Table 2.4) for 4, ΔS_{eq}^{m} for rotations	1264/3632	3, according to Figure 6.2

Table 6.6: Accuracy of the piecewise linear regression model with simulations binned into three intervals based on the 10-minute mean wind speed. Input quantities are equivalent ranges ΔS_{eq}^{m} of TB rotations and standard deviations σ of $\ddot{y}_{na}$ and $\ddot{x}_{na}$. The virtual measurement campaign data set is used.

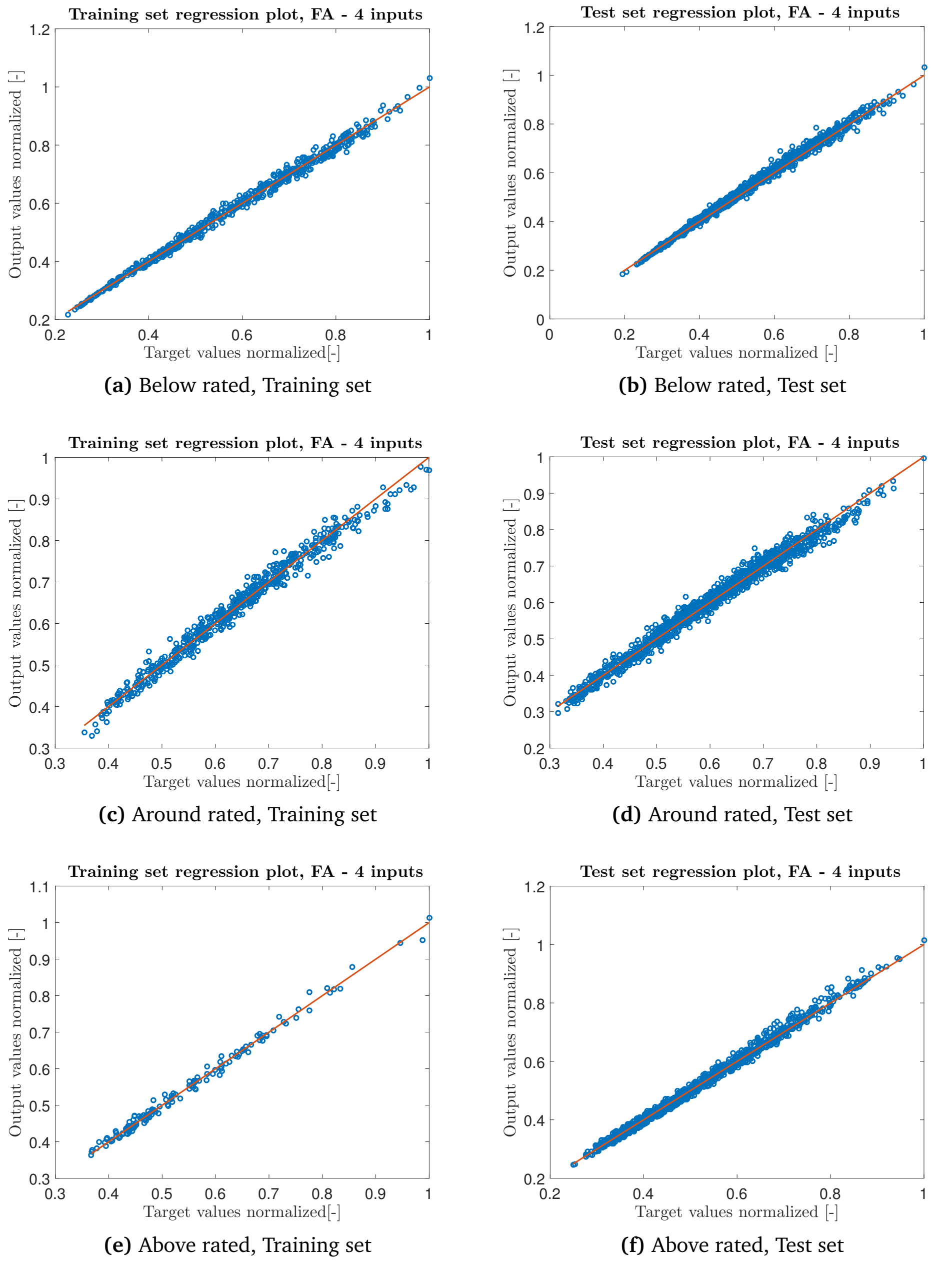
Direction	Operation	Training			Test		
	wrt. rated	$R[-]$	$\epsilon_{mu,abs}[\%]$	$\epsilon_{mu}[\%]$	$R[-]$	$\epsilon_{mu,abs}[\%]$	$\epsilon_{mu}[\%]$
FA	Below	0.997	1.8	0.0	0.997	2.1	1.2
	Around	0.991	2.3	0.0	0.991	2.3	0.1
	Above	0.997	1.6	0.0	0.997	1.7	0.1
SS	Below	1.000	1.4	0.0	1.000	1.8	0.1
	Around	0.999	2.0	0.2	0.999	2.0	0.1
	Above	0.999	1.6	0.2	0.998	2.1	-0.4

that are highly correlated with the target values. The 10-minute equivalent range ΔS_{eq}^{m} of TB rotations is highly correlated with the MBM DELs and furthermore the correlation is linear. This leads to accurate estimation using a simple LR model.

6.5 Conclusions

LR models are simple and more interpretable than ANNs, since each weight w_i is directly associated with a respective input, as is shown in Section 6.1. When the input-output correlation is (close to) linear LR is a good model choice and it is commonly used. The normal operation load case DLC 1.2 is used to test LR models and a $25\%/75\%$ train/test division without duplicate samples is used to create a demanding data set. When a single model is used for the whole data set using nacelle acceleration standard deviations as inputs, errors of around 14% for the test set are observed for both FA and SS moment DELs.

When samples are divided in three bins based on 10-minute mean wind speed and separate LR models are used with same inputs as before, errors generally decrease with respect to the global model, except for the Below Rated, FA direction case, which shows a higher error. Next, the rotation angle signal at the TB is exam-



(a) Below rated, Training set

(b) Below rated, Test set

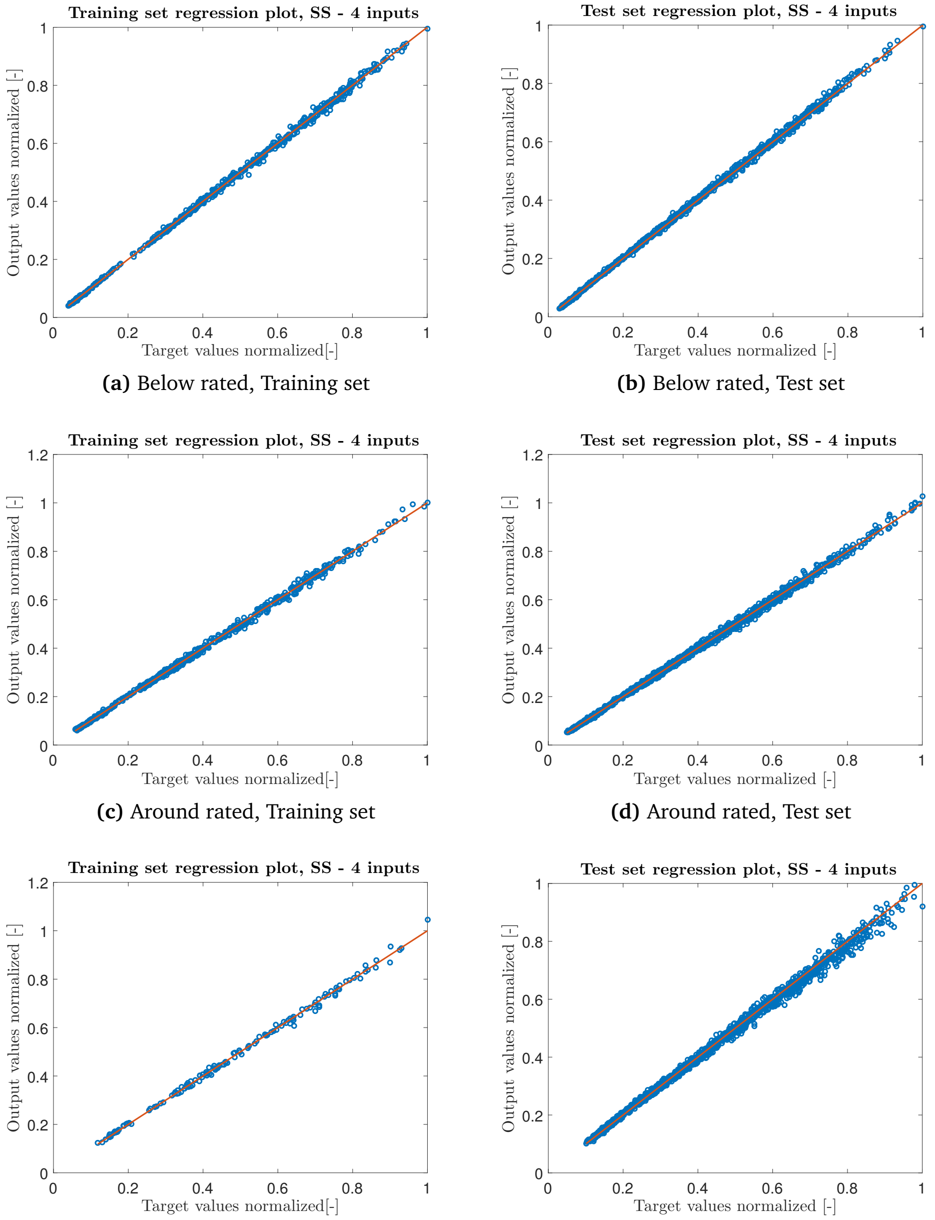
(c) Around rated, Training set

(d) Around rated, Test set

(e) Above rated, Training set

(f) Above rated, Test set

Figure 6.10: Linear regression in FA using σ of $\ddot{y}_{na}$ and $\ddot{x}_{na}$ and ΔS_{eq}^{m} of TB rotations. Operation divided in below, around and above rated wind speed. Results for training and test sets.



(a) Below rated, Training set

(b) Below rated, Test set

(c) Around rated, Training set

(d) Around rated, Test set

(e) Above rated, Training set

(f) Above rated, Test set

Figure 6.11: Linear regression in SS using σ of $\ddot{y}_{na}$ and $\ddot{x}_{na}$ and ΔS_{eq}^{m} of TB rotations. Operation divided in below, around and above rated wind speed. Results for training and test sets.

ined as potential input to LR, after it has been identified as highly correlated with the mudline moment signal. When the 10-minute equivalent range of those rotations are used, errors lower than 2.5% are achieved in all three bins and two directions. Thus, it is shown that if those rotation signals are available – for instance by deploying a dual-axis inclinometer at TB – a highly accurate estimation of the monopile moments is possible using linear regression. However, the precision of such sensors should be additionally considered, since measurement errors can affect the accuracy of the proposed method.

Chapter 7

Conclusions

7.1 Conclusions

The results of this study indicate that it is possible to accurately estimate the loading history of the monopile in the form of bending moment damage equivalent loads (DELs) using feedforward artificial neural networks (ANNs) with a single hidden layer. The mean absolute error is $< 3.5\%$ for both the fore-aft (FA) and side-side (SS) directions for normal operation load cases and the residual error is $< 0.5\%$ for both directions when the inputs to the ANNs are suitable statistics of the standard signals (Table 2.3) that are measured in wind turbines. It is shown that negative spectral moments of the input signals significantly increase accuracy when included in the estimation. These results are achieved using data from Bonus Horizontal axis wind turbine Code (BHawC) simulations that naturally do not include noise that is expected to be present in actual measurements. Also a training data set that properly encapsulates the variety of environmental conditions is used. Of the used input signals, nacelle accelerations are identified as the most important for the estimation accuracy. They exhibit the best accuracy among all cases where a single input is used, with 4.1% mean absolute error in the FA and 3.0% in the SS direction. Including more standard signals as inputs generally leads to a decrease of estimation error. If additionally tower bottom (TB) acceleration signals are available, errors are decreased to $< 3.0\%$ in both FA and SS directions.

In the absence of nacelle acceleration signals, if a combination of the remaining standard signals is used, the error would rise to approximately 15% in both directions. This fact showcases the importance of nacelle acceleration signals for load monitoring of offshore wind turbines (OWTs). Although decent results have been achieved in the case of onshore wind turbines without using nacelle accelerations in Cosack [5], the same levels of accuracy cannot be achieved in the offshore environment, where wave-induced loading is significant. The reason for this is that while the remaining signals contain important information about the wind-induced loading, they contain little information about the wave-induced loading. The combined hydrodynamic and aerodynamic loading that affects mostly the FA direction leads to a more complex behavior than in the SS direction. In the case of idling load cases the importance of nacelle accelerations is exhibited once more, since it is the only standard signal that leads to accurate predictions of moment DELs. A mean absolute estimation error of less than 3.5% in both directions is achieved using only nacelle accelerations.

A proper training set for ANNs is crucial for an accurate model, since extrapolation is not possible. Measurement campaigns in situ are costly and especially for remote underwater locations, such as the seabed level, they are quite difficult. Additionally, measurements of small duration are less likely to make a good training set, because they do not represent the whole variation of loading conditions that the turbine will encounter during its lifetime. In this work it is shown that small training sets lead to high estimation errors. For this reason it might be more realistic to train ANNs on simulated data, since new simulations can be run on demand when it is necessary to expand the training set. Accuracy is also sensitive to changes in the underlying model's properties. When an ANN is trained on data from a specific turbine and is then tested on a similar turbine that has a first eigenfrequency that deviates from the original, error is introduced in the estimation. The acceleration spectral moments are identified as frequency-sensitive input statistics and the error is decreases if they are excluded from the inputs.

Linear regression (LR) models are simpler and more interpretable than ANNs, since each weight w_i is directly associated with a respective input. When the input-output relationship is (close to) linear LR is a good model choice and it is commonly used. The operational load case DLC 1.2 is used to test LR models and a $25\%/75\%$ train/test division is used to create a demanding data set. When a single LR model is used for the whole data set using nacelle acceleration standard deviations as inputs, errors of around 14% (test set) are observed for both FA and SS moment DELs. When samples are binned based on 10-minute mean wind speed and separate LR models are used with same inputs as before in a piecewise LR scheme, errors generally decrease with respect to the global model, except for the below rated, FA direction case.

When 10-minute equivalent ranges of rotation angles at tower bottom are additionally used as inputs in piecewise LR scheme, errors lower than 2.5% are achieved in all three wind speed bins and two directions. Thus, it is shown that if those rotation signals are available – for instance by deploying a dual-axis inclinometer at TB – a highly accurate estimation of the monopile moments is possible using simple linear regression. However, the precision of such sensors has to be considered.

7.2 Recommendations for practical implementation

The methods presented in this thesis can be implemented in a number of different ways. The first step in any case is to validate the findings of this work on one or more offshore wind turbines, using real data, see also Section 7.3. ANNs and LR models can be used on their own or combined, mainly depending on the type and quality of the available training data set. Both methods require past observations of the examined (underlying) system in order to model it in a data-driven way. As a result, the properties of these observations/samples that will be used to train and test the model will determine the optimal implementation scenario for each case. The observations that comprise the available data set can be categorised based on two important characteristics.

1. The way the data set of observations has been acquired. This can be either from:

- Measurements collected from the Supervisory Control And Data Acquisition system (SCADA), Programmable Logic Controller (PLC) and sensors of the actual turbine.
- Results coming from simulations with a computational model of the structure, such as BHawC. This computational model should be preferably validated or updated with measurements from the actual turbine.

2. Representativeness of the available data set.

- In the case of measurements collected from the actual turbine it is likely that the data set corresponds to a limited measurement period and potentially does not capture the whole range of conditions that the turbine will encounter during its lifetime.
- In the case where simulation results are used to create the data set it is easier to extend the range of simulations carried out in order to include conditions met in situ that where not originally accounted for. It is thus more likely for an artificially created data set to be complete and this method shows the biggest potential.

The most appealing scenario is to train ANNs on data from simulations with a validated/updated BHawC model. Then 10-minute statistics from turbine sensors can be input to the ANNs in real time or in batches to estimate the respective DELs. This method matches well with the resources and assets of Siemens (BHawC, signal measurements). If measured data are available for the target quantities, for example from strain gauges on the monopile, it is also possible to use them for ANN training. They can be augmented with simulated data for loading conditions that are missing.

In the case that historic data are recorded from wind turbines they can be potentially used for an a posteriori estimation of the load history in the form of 10-minute DELs. An ANN can be trained on simulation results from a – preferably validated – BHawC model, using exclusively inputs that are among the recorded data. An indication for the accuracy that the available input combinations can offer can be looked up in Tables 4.6 and 4.10 of this work for both normal operation and idling conditions. An overview of the proposed method is given in Figure 7.1. Future work with real data is necessary to validate or update the accuracy levels found in this work.

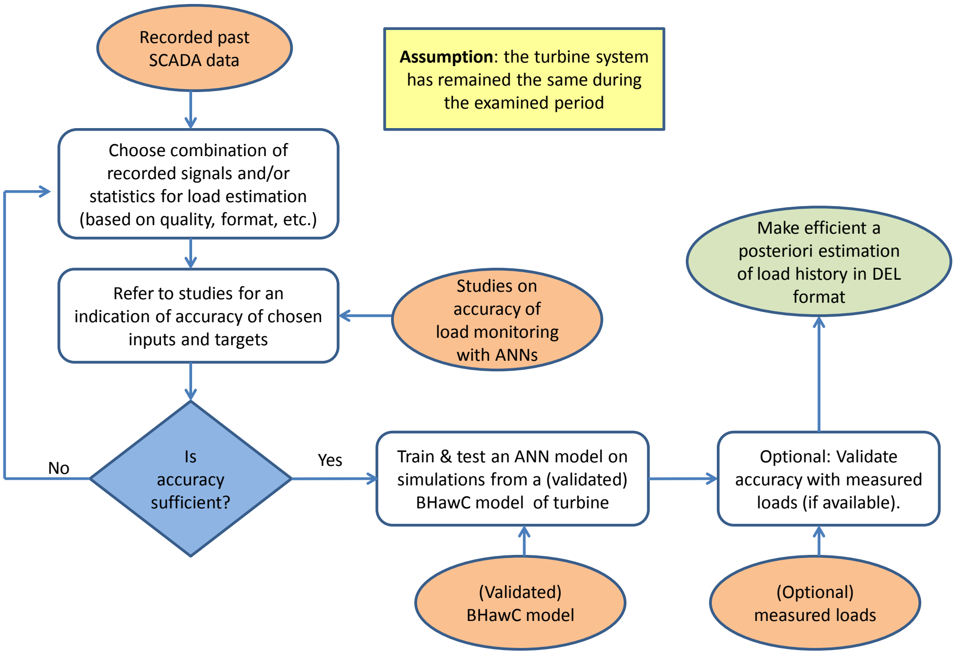

Recorded past SCADA data
Assumption: the turbine system has remained the same during the examined period
Choose combination of recorded signals and/or statistics for load estimation (based on quality, format, etc.)
Refer to studies for an indication of accuracy of chosen inputs and targets
Studies on accuracy of load monitoring with ANNs
Make efficient a posteriori estimation of load history in DEL format
No
Is accuracy sufficient?
Yes
Train & test an ANN model on simulations from a (validated) BHawC model of turbine
Optional: Validate accuracy with measured loads (if available).
(Validated) BHawC model
(Optional) measured loads

Figure 7.1: Flowchart for the process of a posteriori estimation of load histories in DEL format using ANNs.

If additional sensors are considered for load monitoring, a dual-axis inclinometer at TB can be used for straighforward and accurate estimation of DELs with a simple LR model. Again the training data can be either measured or simulated.

An implementation aspect that can increase the efficiency of load monitoring is the application of a single trained model for multiple similar wind turbines, for instance within a cluster. Even within a cluster, where turbines generally share geometrical characteristics it is possible to have deviations for instance due to local soil conditions. If these deviations manifest themselves in the turbine's eigenfrequencies, the impact on the estimation accuracy can be assessed using an approach as the one followed in Section 5.2. If the introduced error is estimated to be high, techniques like model updating [16] are proposed to update the BHawC model and create a new updated data set for training.

7.3 Recommendations for future work

A promising implementation scenario presented in this work includes using data obtained from BHawC simulations to train the ANN model, while data from actual turbines is then fed to the model in order to make estimations for the load quantities (DELs). In this case there are several factors that may decrease the method accuracy and need to be considered. Firstly, it is possible that there is a degree of divergence between the BHawC model and the actual wind turbine. For instance, the actual natural frequencies may be different than the designed. In this work it is shown that deviations in the natural frequencies can introduce significant error in the estimation. Using model updating methods could potentially be used to address this problem, and research in this topic is suggested. In these methods a chosen group of the finite element (FE) model parameters is updated to match the measured vibration characteristics (e.g. natural frequencies and eigenvectors) by minimising a loss function. A survey in the field of model updating in structural dynamics is given in Mottershead and Friswell [16].

Another issue is the influence of noise. Data from FE simulations carried out within Siemens with the BHawC aeroelastic code contain no noise or disturbance in any of the signals that are used. When these models are actually applied, they will use real mesurements from the SCADA systems and potentially sensors as inputs. These measurements inherently contain a level of noise, which depends on the type of signal. This noise can possibly deteriorate the estimation accuracy. A sensitivity study assuming Gaussian, white noise for the input signals is carried out in this thesis, but more work with different noise types is advised. The methods proposed in this thesis should be validated with real data from existing OWTs. A follow up work could thus train an ANN model on BHawC data and then use data from a geared OWT to assess the model's accuracy.

Different signals and/or statistics can be examined as inputs to the data-driven models and their value can be assessed. They can be provided by existing SCADA and PLC systems or by retrofitted sensors. Also, other critical loads can be examined, such as the tower bottom and blade root bending moments. Depending on the specific target load different inputs are expected to provide the highest accuracy. Furthermore, in addition to the normal operation and idling cases that are investigated in this work, other load cases like transients should be investigated.

Transferability of trained models to similar wind turbines within the farm (turbine cluster) is a field that could be examined. In the case that a model has been trained on a specific wind turbine and is performing well, it could be possible to be used for load monitoring of turbines that are similar (in terms of geometrical characteristics, material properties, etc.) and are expected to exhibit a similar behavior. A first case study in this direction is performed in this thesis, where the model applicability is tested on OWTs with slightly different natural frequency than the one it is trained on. This possibility of using one model per cluster of OWTs should be examined/validated with the use of measured data from an actual wind farm. The efficiency of the load monitoring procedure could be significantly improved in the case that a single model is proven sufficient for a cluster of OWTs.

In the case where measured data is used to train the models, generally data coresponding to limited duration are available. It is interesting to examine how well different subsets of the recorded data reflect the variation of the entire data set. For instance, in the case where 1 year of measurements is available, ANNs can be trained on all possible 2-month data sets and used to estimate the remaining 10 months. The same can be done with different lengths for the training set. This approach has been presented in Seifert et al. [25] for the case where flapwise blade root bending moment is the target load and measurements come from Baltic 1 OWF, which consists of 21 Siemens 2.3-93 wind turbines. A similar study for different target loads, such as the monopile mudline section or tower bottom bending moments might be interesting.

Finally, different architectures, activation functions and training algorithms of ANNs can be considered and compared in terms of performance. In addition to that, different machine learning models can be investigated as potential transfer functions between measured signals and fatigue loads.

Appendix A

POWER DENSITY SPECTRUM AND SPECTRAL MOMENTS

The power spectrum $S_{XX}(f)$ of a time series $X(t)$ provides a description of how power is distributed into the various frequency components composing that signal, Stoica et al [28]. According to Fourier analysis any physical signal can be decomposed into a superposition of a number of discrete frequencies, or a spectrum of frequencies over a continuous range. The power density spectrum of an even-length signal sampled at a frequency F_s and with number of samples N can be calculated using the Fourier transform of the signal as follows. The discrete Fourier transform of the time series $X(t)$ is calculated as:

$$Y(k) = \sum_{j=1}^{n} X(j) W_n^{(j-1)(k-1)} \tag{A.1}$$

where $W_n = \exp^{(-2\pi)/N}$. Y ia a complex-valued array with length N. Then the power spectrum values are calculated as:

$$S_{XX}(f) = \frac{1}{F_s N} Y conj(Y) \tag{A.2}$$

Because the signal is real-valued, power estimates for either the positive or negative frequencies is needed (one-sided spectrum). For that reason only the first half values of the symmetric S_{XX} are retained. In order to conserve the total power of the original spectrum, all frequencies that occur in both sets – positive and negative – are multiplied by a factor of 2. Zero frequency and the Nyquist frequency, which is equal to half the sampling frequency ($\frac{F_s}{2}$) do not occur twice. The respective vector of frequencies to which the $S_{XX}(f)$ values correspond is found as:

$$f = \frac{F_s}{N} [0, 1, \ldots, N] \tag{A.3}$$

Spectral moments λ_n of order n can be used to describe the power spectrum. They have a meaning analogous to the physical concepts of center of mass and various moments of area.

$$\lambda_n = \int_0^\infty f^n S_{XX}(f) df \tag{A.4}$$

where f is the frequency and $S_{XX}(f)$ is the power spectral density of that frequency. In the case of the discrete PSD, the n-th order spectral moment is calculated as:

$$\lambda_n = \sum_{i=1}^{m} f_i^n S_{XX}(f_i) \tag{A.5}$$

where m is the number of points in the PSD.

Appendix B

REGRESSION PLOTS FOR NATURAL FREQUENCY SENSITIVITY STUDY

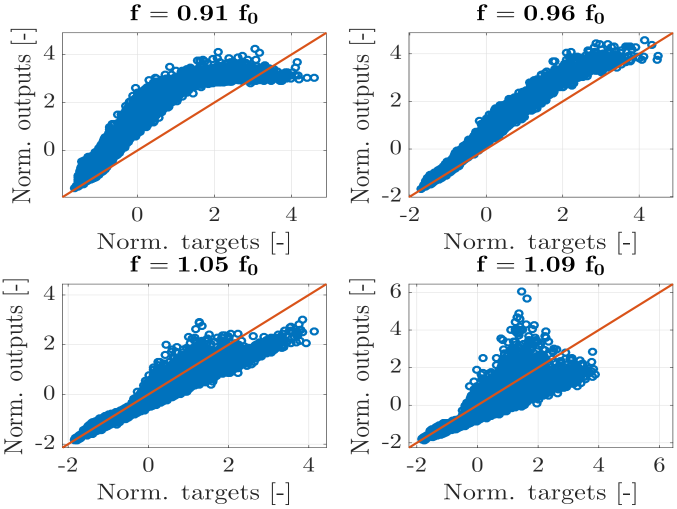


Figure B.1: DEL estimation with ANNs for turbine eigenfrequency different than the one corresponding to the training set. Source turbine eigenfrequency is f_0. Inputs are all standard signals and statistics. Eight neurons in the hidden layer. FA direction is examined.

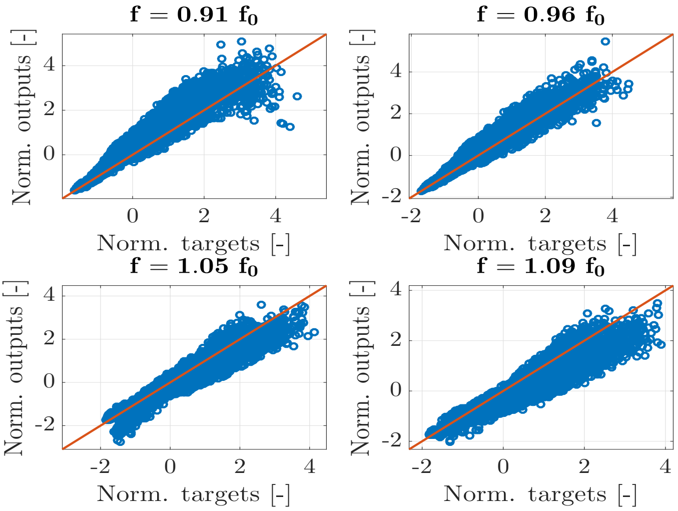


Figure B.2: DEL estimation with ANNs for turbine eigenfrequency different than the one corresponding to the training set. Source turbine eigenfrequency is f_0. Inputs are all standard signals and statistics, only σ of nacelle accelerations. Eight neurons in the hidden layer. FA direction is examined.

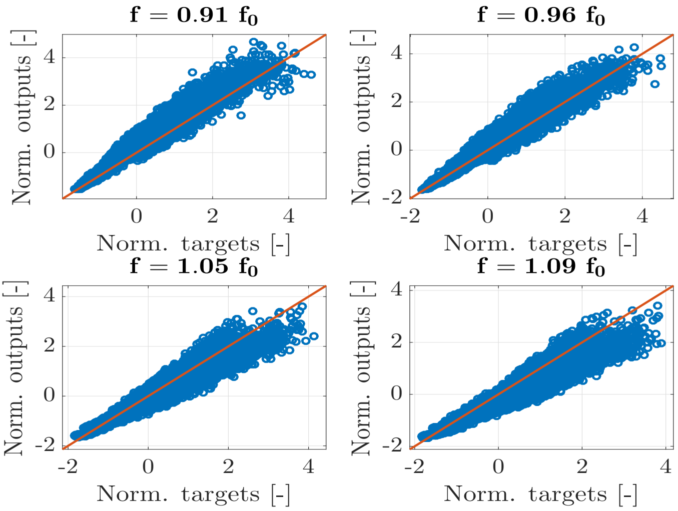


Figure B.3: DEL estimation with ANNs for turbine eigenfrequency different than the one corresponding to the training set. Source turbine eigenfrequency is f_0. Inputs are all standard signals and statistics, only σ of nacelle accelerations. Four neurons in the hidden layer. FA direction is examined.

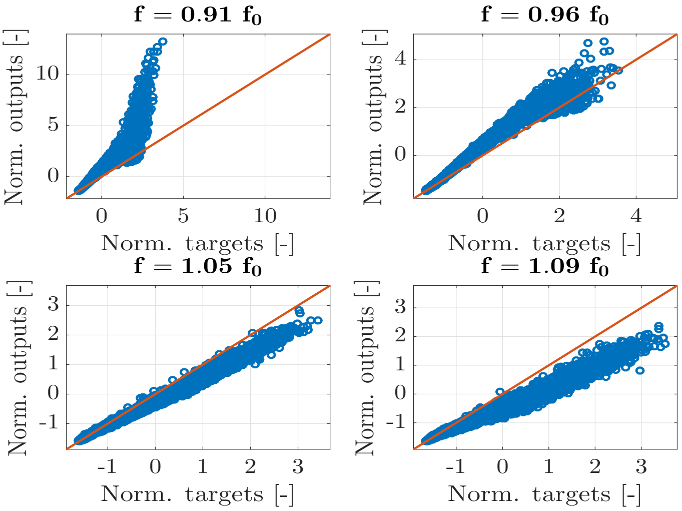


Figure B.4: DEL estimation with ANNs for turbine eigenfrequency different than the one corresponding to the training set. Source turbine eigenfrequency is f_0. Inputs are all standard signals and statistics. Eight neurons in the hidden layer. SS direction is examined.

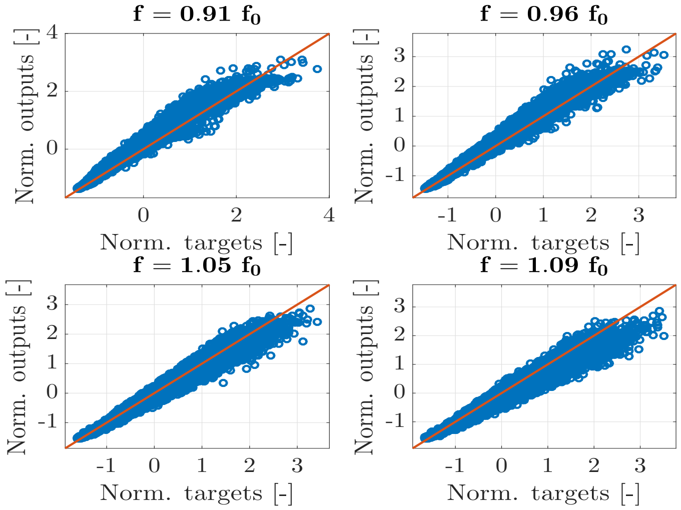


Figure B.5: DEL estimation with ANNs for turbine eigenfrequency different than the one corresponding to the training set. Source turbine eigenfrequency is f_0. Inputs are all standard signals and statistics, only σ of nacelle accelerations. Eight neurons in the hidden layer. SS direction is examined.

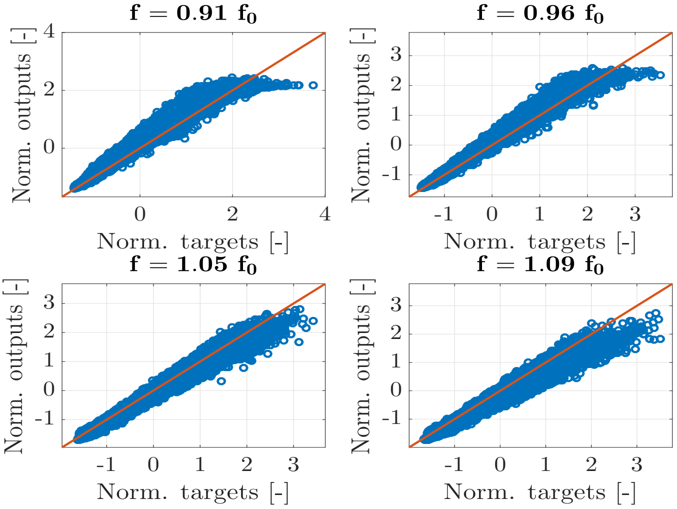


Figure B.6: DEL estimation with ANNs for turbine eigenfrequency different than the one corresponding to the training set. Source turbine eigenfrequency is f_0. Inputs are all standard signals and statistics, only σ of nacelle accelerations. Four neurons in the hidden layer. SS direction is examined.

Bibliography

[1] Iec 61400 wind turbines international standard part 3: Design requirements for offshore wind turbines. Edition 1.0 2009-02. Cited on P. 34

[2] AskABiologist. Neuron anatomy. `https://askabiologist.asu.edu/neuron-anatomy`. [Online; accessed May 22, 2017]. Cited on P. ix, 24

[3] R. Babuska. Lecture notes: Computational intelligence in modeling and control. Delft University of Technology, 2009, Chapter 7. Cited on P. 27

[4] T. Chen and H. Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995. Cited on P. 26

[5] N. Cosack. *Fatigue load monitoring with standard wind turbine signals*. PhD thesis, University of Stuttgart, 2011. Cited on P. 3, 12, 52, 85

[6] S.D. Downing and D.F. Socie. Simple rainflow counting algorithms. *International journal of fatigue*, 4(1):31–40, 1982. Cited on P. 15

[7] D. H Freedman. *Statistical Models: Theory and Practice*. 01 2005. Cited on P. 67

[8] M. Hudson Beale, M.T. Hagan, and H.B. Demuth. Neural network toolbox, user's guide. Matlab 2016a, Mathworks. Cited on P. 21, 35

[9] A. Kusiak, Z. Zhang, and M. Li. Optimization of wind turbine performance with data-driven models. *IEEE TRANSACTIONS ON SUSTAINABLE ENERGY*, VOL(1):66–76, 2010. Cited on P. 1

[10] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, 6(6):861–867, 1993. Cited on P. 25, 31, 39

[11] Ch.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa. Handwritten digit recognition: benchmarking of state-of-the-art techniques. *Pattern recognition*, 36(10):2271–2285, 2003. Cited on P. 20

[12] M. Lourakis. A brief description of the levenberg-marquardt algorithm implemented by levmar. `http://www.ics.forth.gr/~lourakis/levmar/levmar.pdf`, 01 2005. Cited on P. 27

[13] K. Maes, A. Iliopoulos, W. Weijtjens, C. Devriendt, and G. Lombaert. Dynamic strain estimation for fatigue assessment of an offshore monopile wind turbine using filtering and modal expansion algorithms. *Mechanical Systems and Signal Processing*, 76:592 – 611, 2016. Cited on P. 2

[14] R. May, G. Dandy, and H. Maier. Review of input variable selection methods for artificial neural networks. In *Artificial neural networks-methodological advances and biomedical applications*. InTech, 2011. Cited on P. 37

[15] V.E. McZgee and W.T. Carleton. Piecewise regression. *Journal of the American Statistical Association*, 65(331):1109–1124, 1970. Cited on P. 71

[16] J.E. Mottershead and M.I. Friswell. Model updating in structural dynamics: A survey. *Journal of Sound and Vibration*, 167(2):347 – 375, 1993. `http://www.sciencedirect.com/science/article/pii/S0022460X83713404`. Cited on P. 89

[17] V. Nair and G.E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010. Cited on P. 25

[18] T. Obdam, L. Rademakers, and H. Braam. Flight leader concept for wind farm load counting: First offshore implementation. OWEMES conference, Brindisi, Italy, 5 2009. Cited on P. 4

[19] J. Paisley. Lecture notes: Machine learning. Columbia University, course offered through EdX, January 2017. Cited on P. ix, 21

[20] D. Rahul. Why do we need validation? `https://am207.github.io/2017/wiki/validation.html`. [Online; accessed June 15, 2017]. Cited on P. ix, 23

[21] S. Reid. 10 misconceptions about neural networks. `www.turingfinance.com/misconceptions-about-neural-networks/`, May 8, 2014. [Online; accessed May 22, 2017]. Cited on P. ix, 26

[22] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958. Cited on P. 25

[23] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Learning Internal Representations by Error Propagation, pages 318–362. MIT Press, Cambridge, MA, USA, 1986. Cited on P. 27

[24] P.A. Rutecki. Neuronal excitability: voltage-dependent currents and synaptic transmission. *Journal of Clinical Neurophysiology*, 9(2):195211, 1992. Cited on P. 24

[25] J. Seifert, L. Vera-Tudela, and M. Kühn. Training requirements of a neural network used for fatigue load estimation of offshore wind turbines. Cited on P. 55, 90

[26] U. Smolka and P.W. Cheng. On the design of measurement campaigns for fatigue life monitoring of offshore wind turbines. In *Proceedings of the Twenty-third (2013) International Offshore and Polar Engineering Anchorage, Alaska, USA*, June 30 July 5, 2013. Cited on P. 5

[27] U. Smolka, D. Kaufer, and P. W. Cheng. Are sea state measurements required for fatigue load monitoring of offshore wind turbines? *Journal of Physics: Conference Series*, 555(1):012095, 2014. Available at `http://stacks.iop.org/1742-6596/555/i=1/a=012095`. Cited on P. 5

[28] P. Stoica, R. L. Moses, et al. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005. Pages 1-4. Cited on P. 79, 91

[29] Grimsby Telegraph. Robust results for dong energy as wind powers on. `http://humberbusiness.com/news/robust-results-for-dong-energy-as/story-3501-detail/story`, November 8, 2016. [Online; accessed November 22, 2017]. Cited on P. ix, 12

[30] A. Vuckovic, D. Popovi, and V. Radivojevic. Artificial neural network for detecting drowsiness from eeg recordings, 10 2002. Cited on P. ix, 27

[31] E. Weisstein. Least squares fitting. `http://mathworld.wolfram.com/LeastSquaresFitting.html`. From MathWorld–A Wolfram Web Resource. Cited on P. 68

[32] P.J. Werbos. Beyond regression : new tools for prediction and analysis in the behavioral sciences /. 01 1974. Cited on P. 27

[33] L. Ziegler, U. Smolka, N. Cosack, and M. Muskulus. Structural monitoring for lifetime extension of offshore wind monopiles: can strain measurements at one level tell us everything? 2:469–471, 09 2017. Cited on P. 79

Definitions

Machine learning: A subfield of computer science whose objective is to explore the study and construction of algorithms that can learn from and make predictions on data, rather than being explicitly programmed.

Supervised learning: The process of inferring an underlying function from labeled training data. The training data comprises samples of the input vectors along with their corresponding target vectors. Depending on the task supervised learning is divided in regression and classification.

Regression: A statistical process for estimating the relationships between a dependent variable and one or more independent variables (or ”predictors”). In the field of supervised learning, the term regression is used for the prediction of continuous dependent variables (as opposed to discrete or categorical variables).

Unsupervised learning: The process of exploring unlabeled data and discovering structure in it such as hidden patterns or groupings. Unlabeled data samples means they consist only of input and no corresponding label.

Overfitting: A phenomenon that occurs when a too complex model is used for a specific problem. A sign of overfitting is when the training dataset is very well or even perfectly fit, but when the complex model is evaluated on new data (testing), it performs poorly. In other words, the model does not generalize well. For example using higher order polynomials for fitting a linear (first order) relationship can lead to overfitting.

TU Delft Delft University of Technology

SIEMENS Gamesa RENEWABLE ENERGY