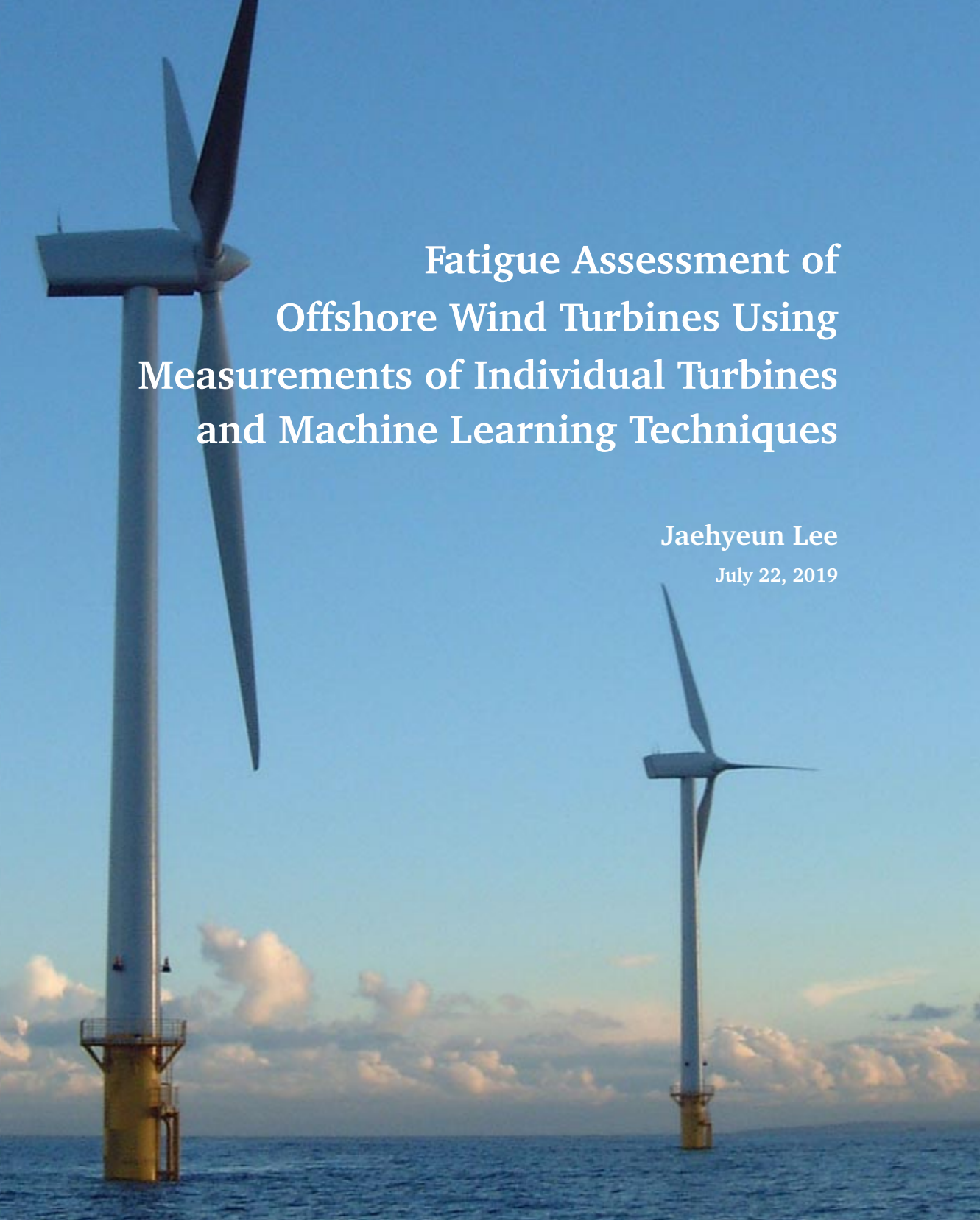


Fatigue Assessment of Offshore Wind Turbines Using Measurements of Individual Turbines and Machine Learning Techniques

Jaehyeun Lee

July 22, 2019

TU Delft Delft University of Technology

NTNU

SIEMENS Gamesa RENEWABLE ENERGY

Fatigue Assessment of Offshore Wind Turbines using Measurements of Individual Turbines and Machine Learning Techniques

Master of Science Thesis

Author:	Jaehyeun Lee	
Committee:	Prof. Andrei Metrikine	TU Delft - Chairman
	Prof. Michael Muskulus	NTNU
	Dr. Eliz-Mari Lourens	TU Delft
	Dr. Alexandros Iliopoulos	Siemens Gamesa Renewable Energy
	Ir. Hendrik Kramers	Siemens Gamesa Renewable Energy

July 22, 2019

TU Delft Delft University of Technology NTNU SIEMENS Gamesa RENEWABLE ENERGY

- CONFIDENTIAL -

Summary

Fatigue is often a governing design factor for offshore wind turbines. Since the design of offshore wind turbines includes conservatism, the actual accumulated fatigue damage can be lower than what the turbine is designed for. In this case, the operator can make a decision on life time extension of existing wind turbines. Therefore, it is important to estimate the actual accumulated fatigue damage to support decision making on life time extension, and for optimization of support structure design. However, fatigue critical locations are located near mudline where it is unfeasible to install strain gauges to measure the accumulated fatigue damage.

The first purpose of this thesis is to investigate if data-driven approaches (linear regression and feed-forward neural network) can be applied to estimate the accumulated fatigue damage both in individual turbines and farm-wide levels. The second purpose is to determine the minimum number of sensors and quantity of data required for accurate estimation.

Towards this goal, real measurement data of two offshore turbines in the same wind farm have been used. Specifically, the data-driven approaches have been applied with real measurement data from the SCADA system, measurements at the top and bottom of the tower, and data from a wave measurement system. This data was used to estimate the accumulated fatigue damage at multiple locations (tower bottom, transition piece and two levels on the monopile) in the form of damage equivalent loads. Throughout the study, 10 min. statistical properties of the measurement data have been used as input to the learning algorithms. One remark is that the estimation has not been performed for the fatigue critical location near mudline itself, but it is expected that estimation with these approaches can be expanded to the fatigue critical location if accurate response estimation at multiple locations on the support structure is possible.

The results of this thesis show that the data-driven approaches can give accurate estimates damage equivalent loads on individual turbine level at multiple locations on the support structure when moment or inclination signals at tower bottom is used.

For farm-wide level load estimation as well, it has been proven that the data-driven approaches can give quite accurate estimates the damage equivalent load. However, it should be noted that the turbines used in this study have similar dynamic properties. Therefore, the farm-wide level load estimation with the data-driven approaches should be further investigated in the future.

Acknowledgements

I would like to express my gratitude to my supervisors Dr. Alexandros Iliopoulos and Ir. Hendrik Kramers for their patient guidance and all the daily cares throughout the thesis project. Furthermore, I would like to thank Dr. Eliz-Mari Lourens for introducing this thesis project. Additionally, I am grateful for her trust, advice, and assistance in keeping my progress on schedule. I would like to express my appreciation to Prof. Michael Muskulus for his suggestions and useful critiques. Especially, he gave me opportunities to apply different machine learning techniques during the thesis project. It makes me have profound insight into the machine learning techniques. Also, I am profoundly grateful to Prof. Andrei Metrikine for the suggestions and comments I have never thought of. It makes me improve myself.

I would also like to extend my thanks to my friends. During the two years of my study, many friends are helpful to color my life. Especially, I am grateful to all my EWEM friends for making my fantastic student life.

Special thanks are given to my family. My parents always support me. Most importantly, I wish to thank my loving and supportive wife, Dawon, and my son, Juho, who always inspire and make me break my limitation. I am really appreciated for all the efforts my wife has been making, and the trust she showed me.

Lastly, I thank my God for letting me through all the difficulties. I have always felt Your presence, and it makes me encouraged, not to be disappointed in any situation. I am in Your grace. Thank you, Lord Jesus Christ.

"I can do all this through him who gives me strength" (Philippians 4:13)

Contents

List of Figures

List of Tables

Nomenclature

Latin symbols

D_{eq}	Equivalent damage	-
$\hat{\boldsymbol{y}}$	Target vector	-
$\boldsymbol{C}$	Damping matrix	-
$\boldsymbol{E}$	Error vector	-
$\boldsymbol{e}$	Error vector	-
$\boldsymbol{g}$	Gradient vector	-
$\boldsymbol{H}$	Hessian matrix	-
$\boldsymbol{I}$	Identity matrix	-
$\boldsymbol{J}$	Jacobian matrix	-
$\boldsymbol{K}$	Stiffness matrix	-
$\boldsymbol{M}$	Mass matrix	-
$\boldsymbol{w}$	Vector weights	-
$\boldsymbol{y}$	Output vector	-
$\Delta \boldsymbol{w}$	Vector of changes in weights	-
Δw	Change in weight	-
F	Nonlinear function	-
g	Gradient	-
$\tilde{\boldsymbol{c}}$	New candidate	-
$\boldsymbol{c}$	Cell state	-
$\boldsymbol{f}$	Forget rate	-
$\boldsymbol{h}$	Hidden state	-
$\boldsymbol{i}$	Input rate	-

$\boldsymbol{o}$	Output rate	-
$\hat{X}$	Dataset	-
$\hat{Y}$	Dataset	-
A	Area of a section	m^2
a	Activation function	-
Acc	Acceleration	m/s^2
$Acc_{FA,Btm}$	Acceleration in fore-aft direction at tower bottom	m/s^2
$Acc_{FA,Top}$	Acceleration in fore-aft direction at tower top	m/s^2
$Acc_{SS,Btm}$	Acceleration in side-by-side direction at tower bottom	m/s^2
$Acc_{SS,Top}$	Acceleration in side-by-side direction at tower top	m/s^2
D	Damage	-
E	Young's modulus	GPa
f_i	i th frequency	Hz
H_i	i th hidden neuron	-
H_s	Significant wave height	m
I	Second moment of inertia of a section	m^4
I_i	i th input neuron	-
J_{TR}	Training error	-
J_{VAL}	Cross validation error	-
M	Moment	kNm
m	Slope of S-N curve	-
M_X	Moment along with X direction	kNm
M_Y	Moment along with Y direction	kNm
M_{eq}	Damage equivalent moment	kNm
$M_{FA,Btm}$	Moment in fore-aft directional at tower bottom	kNm
$M_{FA,Top}$	Moment in fore-aft directional at tower top	kNm
M_i	i th moment	kNm
M_{SN}	Moment along with south-north direction	kNm
$M_{SS,Btm}$	Moment in side-by-side directional at tower bottom	kNm
$M_{SS,Top}$	Moment in side-by-side directional at tower top	kNm
M_{WE}	Moment along with west-east direction	kNm
$M_{X,MP1}$	Moment along with X direction at upper monopile	kNm

$M_{X,MP2}$	Moment along with X direction at lower monopile	kNm
$M_{X,TP}$	Moment along with X direction at transition piece	kNm
M_x	Moment along with x measurement coordinate	kNm
$M_{Y,MP1}$	Moment along with Y direction at upper monopile	kNm
$M_{Y,MP2}$	Moment along with Y direction at lower monopile	kNm
$M_{Y,TP}$	Moment along with Y direction at transition piece	kNm
M_y	Moment along with y measurement coordinate	kNm
N	Total number	-
N_i	Total number of cycles of i th stress range $\Delta\sigma_i$	-
n_i	Actual number of cycles of i th stress range $\Delta\sigma_i$	-
N_{eq}	Equivalent number of cycles	-
N_{ref}	Reference number of cycles	-
O_i	i th output neuron	-
P_{act}	Active power	kW
P_{axial}	Axial force	kN
R	Pearson's correlation coefficient	-
S	Power spectral density	$Unit^2/Hz$
T_p	Peak period	s
U_w	Wind speed	m/s
w	Weight	-
w_{ij}^k	Weight factor of i th neuron in $(k-1)$ th layer toward j th neuron in k th layer	-
X	Global coordinate cooresponding to west-east direction	-
x_i	i th data	-
x_u	Distance from neutral axis to strain gauge at East & West side	m
Y	Global coordinate cooresponding to south-north direction	-
y	Output signal	-
y_i^j	Output signal from i th neuron in j th layer	-
y_u	Distance from neutral axis to strain gauge at South & North side	m
z	Input signal	-

Greek symbols

μ	Mean	-

$\Delta\sigma$	Stress range	kPa
$\Delta\sigma_i$	i th stress range	kPa
$\Delta\sigma_{eq}$	Equivalent stress range	kPa
$\Delta\sigma_{ref}$	Reference stress range	kPa
ϵ	Strain	-
ϵ_E	Strain at east	-
ϵ_N	Strain at north	-
ϵ_S	Strain at south	-
ϵ_W	Strain at west	-
λ_n	n th spectral moment	-
$\mu_{\hat{X}}$	Mean of dataset $\hat{X}$	-
$\mu_{\hat{Y}}$	Mean of dataset $\hat{Y}$	-
μ_c	Combination coefficient	-
Ω_{gen}	Rotational speed	rad/s
$\phi_{FA,Btm}$	Inclination in fore-aft directional at tower bottom	$degree$
$\phi_{SS,Btm}$	Inclination in side-by-side directional at tower bottom	$degree$
$\sigma_{\hat{X}}$	Standard deviation of dataset $\hat{X}$	-
$\sigma_{\hat{Y}}$	Standard deviation of dataset $\hat{Y}$	-
σ_{std}	Standard deviation	-
θ_b^p	Blade pitch angle	$degree$
θ_i^j	i th bias of layer j	-
θ_{offset}	Offset angle	$degree$
θ_{yaw}	Yaw angle	$degree$

Abbreviations

ANN	Artificial neural network
BhawC	Bonus Horizontal Axis Wind turbine Code
CS	Case Study
DEL	Damage Equivalent Load
FA	Fore-aft direction
FLS	Fatigue Limit State
LSTM	Long short-term memory

MAPE	Mean Apsolute Percentage Error
MPE	Mean Percentage Error
OWT	Offshore Wind Turbine
PCA	Principal component analysis
PSD	Power Spectral Density
RNA	Rotor Nacelle Assembly
RNN	Recurrent neural network
SCADA	Supervisory Control And Data Acquisition system
SGRE	Siemens Gamesa Renewable Energy
SS	Side-by-side direction
T1A2	Training with turbine 1, Applied to turbine 2

Chapter 1

Introduction

This chapter gives a general introduction to the research performed in this study. In Section 1.1, the research motivation is described. In Section 1.2, results of literature study is explained. The research question and methodology are given in Section 1.3 and 1.4 respectively. Lastly, the software used for this study and the outline of this thesis are presented in section 1.5 and 1.6 respectively.

1.1 Motivation

The wind energy industry is growing quickly around the world. Especially, offshore wind energy industry is growing at a remarkable rate. In 2018, Europe solely installed offshore wind farms with overall capacity of 2.6 GW and the next 4 years, as shown in Figure 1.1, foresees over 2 GW of newly installed offshore wind farms [9, 10]. The two most widely used support structure types are monopile and jacket. Especially, when the water depth is shallow, monopile is a preferred option due to the fact that it is simpler and has fewer weld joints compared to jacket. As a result, of all offshore wind turbines installed in 2018 in Europe, 66% has monopile support structure while 33% uses jacket type support structure [10].

In addition, offshore wind turbines are exposed to various sources of loads including hydrodynamic loads, aerodynamic loads, load from rotating blades and sometimes ice loads. Moreover, all the sources have interaction with the structural dynamics of the turbines. As a result, high dynamic excitation forces are applied to the turbines. Consequently, fatigue is one of the most governing design factors in offshore wind turbines. Especially, for the monopile support structure, high level

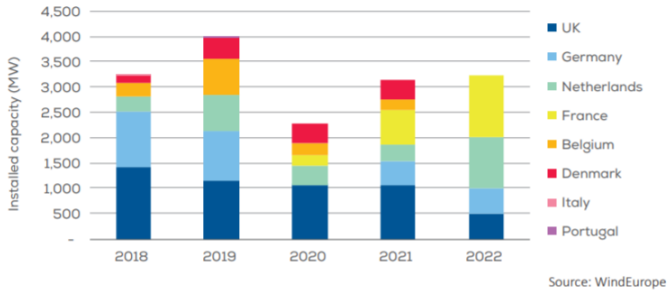


Figure 1.1: Offshore wind outlook

of bending moment is acting at a section near mudline location and this location is where the fatigue damage is accumulated the most rapidly.

In general, the lifetime of offshore wind farm is 20 years. At the end of its lifetime, turbines can either be decommissioned or be used for a longer period. Obviously, it is economically beneficial to extend the lifetime of a wind turbine. However, from a technical point of view, to justify the decision making on a lifetime extension of wind turbines, all of the aspects shown in Table 1.1 should be assessed [8]. Especially, assessment of accumulated fatigue damage at fatigue critical location is crucial for foundation assessment to support decision making on the lifetime extension. In addition to that, by estimating accumulated damage at the end of lifetime, applied design safety factor for fatigue limit state (FLS) can be assessed by comparing actual accumulated damage with design value. In other words, support structure design can be optimized.

To be able to estimate the accumulated damage at fatigue critical location, strain gauges can be installed. However, as explained above, fatigue critical location is near mudline location where it is unfeasible to install strain gauges. Therefore, it is desirable to use remote sensing technique. The remote sensing technique uses measurements at some locations where it is feasible to install sensors and estimate response at other locations.

1.2 Literature study

To estimate response with the remote sensing techniques, either the physics-based approach and data-based approach can be applied. The physics-based approach uses the laws of physics to describe and solve a problem based on it. On the other hand,

Table 1.1: Assessment scopes for lifetime extension

Component	System Parts
Rotor Blade	Blade
Machinery components	Hub
	Main shaft
	Torque arm
	...
	Yaw system
Tower	Tower segments
	Tower connections
	Door opening
Foundation	Anchor bolt connection
	...
	Jacket structure (offshore)
	Monopile structure (offshore)
Control and protection system	Sensors
	Braking systems
	Control software
Electrical Equipment	Generator
	Lightning protection

the data-based approach uses input and target data and solves a problem by finding a relationship between the input and output data. For remote sensing, Kalman filter and modal decomposition and expansion can be regarded as the physics-based approach. For wind turbines, studies have been performed and remote sensing can capture well the response of wind turbine [13, 35]. In addition to that, the data-based approach has also been performed and it shows the data-based approach can also be used for remote sensing [36].

Especially, a neural network has been applied for some studies and it shows that the feed-forward neural network can be used for remote sensing.

Specifically, in the previous study from Cosack [7], regression, neural networks, and physics-based models have been applied to estimate damage equivalent load at different locations with standard signals (from Supervisory Control And Data Acqui-

sition system (SCADA)) for onshore wind turbines.

In detail, onshore variable-speed pitch-controlled wind turbines with a rated power of 1.5 MW has been considered to validate if the feed-forward neural network can be used for the estimation. For the turbine, simulation data have been generated with Flex5 for varying turbulence intensity and mean wind speeds. With the simulation data, damage equivalent load (hereinafter, DEL) of rotor torque has been estimated with the feed-forward neural network with one hidden layer. As input data, statistical properties such as mean, standard deviation and one to fourth spectral moments of standard signals have been used. The standard signals include generator rotational speed, electrical power, tower top acceleration and so on. As a result, it has been proven that the feed-forward neural network with one hidden layer can give an accurate estimation with mean percentage error (hereinafter, MPE) of less than 4%.

Then, the feed-forward neural network with a single hidden layer has been applied with real measurement data of onshore wind turbines including one Multibrid M5000 turbine and two Nordex N80 turbines. For both types of turbines, the feed-forward neural network has been trained on around 120 hours of data and tested with around 30 hours of data in which the turbines are in normal operating condition. For individual turbine level load estimation, it has been found that the feed-forward neural network gives an accurate estimation of DEL of tower bottom moment, blade root bending moment, rotor torque and thrust force for both types of turbines with MPE of less than 4%.

For two Nordex N80 turbines, farm-wide estimation has been performed. Specifically, the feed-forward neural network trained on one turbine (T8) has been applied to another turbine (T6). The estimation accuracy was not as accurate as individual turbine level load estimation with MPE of around 8% even though two turbines are only 600 m away from each other, and both turbines (T6 and T8) have the same design. However, no clear explanation has been made for the inaccuracy in farm-wide level load estimation.

In addition to that, the study from M.Souliotis [28] found that the feed-forward neural network can be applied to offshore wind turbines as well.

Concretely, total 11232 simulation data for normal operating condition and 2304 simulation data for idling condition have been generated with a simulation software Bonus Horizontal Axis Wind turbine Code (BhawC). Each simulation data includes 10 min. averaged statistical properties of signals, such as mean, standard deviation, equivalent value and negative second to positive second spectral moments. In the study, not only the standard signals (from SCADA), but also acceleration at tower bottom have been used as input data. As a target value, moment DEL at mudline

has been used.

As a result, it has been found the feed-forward neural network with a single hidden layer can accurately estimate DEL at mudline of offshore wind turbines with mean absolute percentage error (hereinafer, MAPE) of less than 4%.

In addition to that, linear regression has been applied to estimate moment DEL at mudline with inclination at tower bottom. The result shows that linear regression can also give accurate estimation with MAPE of less than 3% if the inclination at the tower bottom has been used as input data. The estimation accuracy of the linear regression was comparable to that of the feed-forward neural network.

In summary, for both onshore and offshore wind turbines, estimation of moment DEL with the feed-forward neural network has been proven to be accurate. For offshore wind turbines, estimation with the feed-forward neural network can be accurate and it has been validated with simulation data. For onshore wind turbines, the feed-forward neural network has been validated both with simulation and real measurement data and the estimation was accurate for individual turbine level load estimation. However, for farm-wide level load estimation, it has been found that the feed-forward neural network gives relatively inaccurate result compared to individual turbine level.

As a result of the literature study, it has been found that the following remaining researches need to be done:

- Validation of feed forward neural network for offshore wind turbine with real measurement data
- Validation of farm-wide level load estimation

In this study, studies have been made for those items. In addition, not only the linear regression and feed-forward neural network, but the recurrent neural network has also been applied since it has not been researched.

1.3 Research questions

As explained in the previous section, neural networks have not been proved for offshore wind turbines with real measurement data. Therefore, in this study, neural networks have been applied with real measurement data with the first research question of **'Can the neural network technique be applied to accurately estimate**

moment DEL at already known fatigue critical locations for offshore wind turbines with real measurement data?'. The research question addresses first the potential for load estimation on an individual turbine level. That is, the input (measurements at tower and RNA) and target (moment DEL at already known fatigue critical location - near mudline) both are from the same turbine and are all taken from real measurement data. In addition, it also needs to be answered **'What is the required number of data to train a neural network?**.

Expanded to the first and second research question, the following third research question can be asked: **'Can the machine learning technique (neural network) be expanded to farm-wide fatigue assessment?'**.

In addition, for both individual turbine and farm-wide level fatigue assessment, the following question needs to be answered: **'What is the minimum number of sensors that gives accurate DEL estimation?'**.

1.4 Methodology

In this study, real measurement data from two offshore wind turbines have been used. Specifically, measurement data from SCADA and distributed sensors installed at the tower have been used as input data, and the strain signals installed at bottom of the tower, transition piece and monopile have been used as target data for neural networks. For strain signal at monopile, it should be noted that there are no strain gauges at mudline location but the strain gauges are installed at two levels above mudline. If neural networks can estimate DEL at those two locations at monopile, it is expected that the technique can also be used to estimate response at mudline. The detailed description of measurement data is shown in Chapter 2.

1.4.1 Individual turbine load estimation

Firstly, linear regression has been applied to estimate moment DEL for specific operating conditions where the number of data is not enough for training of neural networks. Secondly, the feed-forward neural network has been applied. In this approach, different neural network structures have been compared in terms of accuracy.

Thirdly, the recurrent neural network has been applied. Specifically, of all different types of recurrent neural networks, long short-term memory (LSTM) has been applied. Then, accuracy has been compared between feed-forward neural network and recurrent neural network.

Then, with the feed-forward neural network, case studies have been made to answer the research question of ***'what is the minimum number of sensors that give accurate DEL estimation?***. In addition, different approaches have been examined to figure out if any improvement can be achieved.

Finally, the size of training data has been varied and the accuracy level has been evaluated to answer the research question of ***'what is the required number of data to train a neural network?***.

1.4.2 Farm-wide load estimation

To answer the research question ***'Can the neural network be expanded to obtain a farm-wide fatigue assessment?***, the feed-forward neural network has been trained on one turbine, and applied to the other turbine.

First of all, linear regression has also been applied in farm-wide level.

Secondly, the same feed-forward neural network used in individual turbine level has been applied with different input combinations to answer the research question of ***'what is the minimum number of sensors that give accurate DEL estimation?***.

Thirdly, to include different dynamic properties and wake effect between two turbines, case studies have been performed to figure out if any improvement can be achieved.

1.5 Software

Throughout the study, two main software have been used.

- MATLAB R2018b
- Python

Specifically, MATLAB has been used to pre-process the data and build up neural networks (both feed-forward neural network and recurrent neural network) with Deep Learning Tooldbox.

Python has been used to apply random forest regression with pandas and scikit-learn library.

1.6 Outline

Chapter 1 presents general introduction of this study. Firstly, the motivation for this study is explained. Then, research questions and objectives are described. Lastly, the

methodology used to answer the research questions is presented.

Chapter 2 introduces sensor configurations and data pre-processing procedures taken in this study. Firstly, a general description of the turbines is given. Then, the sensor configuration and detailed description of each sensor are given. In addition, a summary of all the pre-processed data is shown. Then, correlation between different features and moment DEL is presented. Lastly, plots of pre-processed data are presented.

In Chapter 3, theories of feed-forward neural network and recurrent neural network are described. Firstly, a basic description of the feed-forward neural network is given. Then, the gradient descent and learning process of the neural network (forward and backward propagation) are given. In addition, a general description of the recurrent neural network is presented focusing on long short-term memory (LSTM).

In Chapter 4, fatigue assessment on individual turbine level is shown. Firstly, the base case is described. Then, results with linear regression, feed-forward neural network, and recurrent neural network are shown and comparison of their accuracy comparison is made. Lastly, the results of various case studies are shown.

In Chapter 5, fatigue assessment on farm-wide level is shown. Firstly, result with linear regression is explained. Then, the result of the feed-forward neural network application on the base case is described followed by the result of case studies with different input combinations. Lastly, results are shown for explicit inclusion of dynamic property differences and wake effect.

In Chapter 6, conclusion is given together with recommended future work.

Chapter 2

Data overview and preprocessing

In this chapter, the measurement data that have been used in this study is described. Firstly, the installed sensor configurations and the form of measurement data from SCADA and the vibration sensors are introduced in Section 2.1. Then, a summary of processed measurement data is shown in Section 2.2. In Section 2.3, correlation between different features and moment DEL at tower bottom is presented. Lastly, important plots of pre-processed measurement signals are presented in Section 2.4. Especially, all the measurement data shown in this chapter have been pre-processed into statistical properties. The detailed description is not shown in this chapter, but it can be found in Appendix A.

2.1 Sensor configuration and format of data

The two offshore wind turbines examined in this study have the same configuration of sensors. There are two accelerometers at the top and bottom of the tower, two calibrated moment signals converted from strain gauges at the top and bottom of the tower, one inclinometer at bottom of the tower. In addition, raw strain signals at one transition piece level and two different monopile levels are available. At each level, the strain gauges are installed equidistantly along the circumference as shown in Figure 2.1. The subscript E, N, W and S in the figure stand for East, North, West, and South respectively. In addition, the strain gauges measure the axial strain (along

with local z-axis of the tower in Figure 2.1). The overall sensor configurations are shown in Figure 2.2.

In addition to the sensors described above, there is SCADA in Rotor-Nacelle Assembly (RNA). Since SCADA is a default system that is installed in all turbines, data from SCADA will be denoted as 'Standard signal' while data from the other sensors will be denoted as 'Non-standard signal' since the sensors attached at the tower, transition piece and monopile are not installed on every turbine by default.

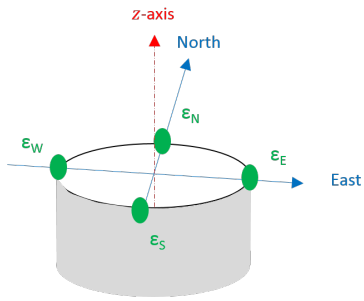


Figure 2.1: Configuration of strain gauges

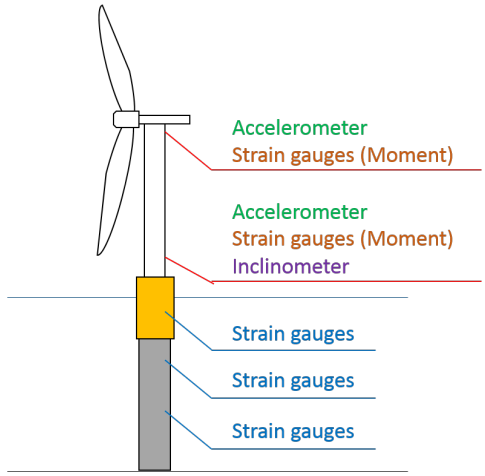


Figure 2.2: Overall configuration of sensors

2.1.1 Data from SCADA system (Standard signals)

SCADA system records high-frequent data. Firstly, the data provided for this study are based on **10 min. statistical properties** of the signals by default. Specifically, all the 10 min. statistical properties shown in Table 2.1 of all the data shown in Table 2.2 are recorded in SCADA with time stamp. However, it should be noted that time series for all the SCADA data except acceleration are also available with the sampling frequency of 0.04 sec. (25 Hz) in addition to the statistical data. For acceleration signal, time series has been recorded with varying sampling frequencies (event-driven sampling frequency).

Table 2.1: Standard statistical properties recorded in SCADA

Number	Statistical Properties	Symbol
1	Mean	μ
2	Standard deviation	σ_{std}
3	Max.	-
4	Min.	-

Table 2.2: Data from SCADA

Number	Signal	Symbol
1	Acceleration	Acc_{FA}, Acc_{SS}
2	Blade Pitch angle	θ_b^p
3	Wind Speed	U_w
4	Active Power	P_{act}
5	Rotational Speed	Ω_{gen}
6	Yaw Direction	θ_{yaw}
7	Operating Condition	-

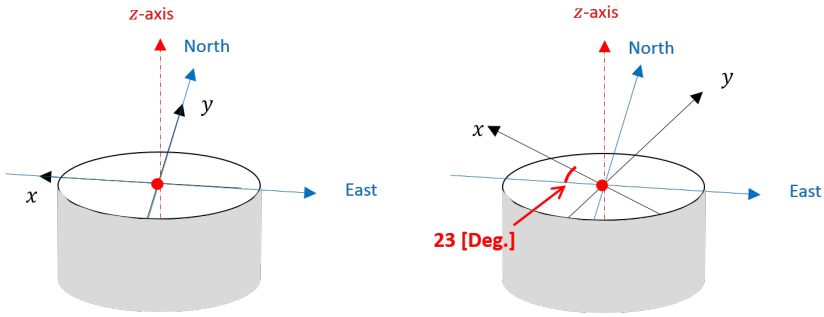
Operating condition recorded in SCADA system consists of multiple different conditions which are divided according to wind speed, pitch angle, active power and so on. Four different operating conditions are found within the one-month data. Specifically, three operating conditions and parked condition are in the dataset.

2.1.2 Data from Sensors (Non-standard signal)

All the data from sensors are recorded as 0.04 sec. **time series**. The sampling frequency is fixed as 25 Hz in all sensors and the one-month time series are available.

As depicted in Figure 2.2, a number of sensors are installed at different levels. The strain gauges at transition piece and monopile are installed with the configuration shown in Figure 2.1.

For top and bottom of the tower, moment signals are available. The moment signals had been pre-processed from strain signals by SGRE. Both the moment signals at top and bottom of the tower consist of M_x and M_y which stand for the moment based on measurement coordinate system (lower case x and y). For turbine 1, the axis x and y are located along with East-West and South-North directions respectively as shown in Figure 2.3a. However, for turbine 2, the strain gauges at top and bottom of the tower are installed with the offset of 23 deg. as shown in Figure 2.3b. As a result, the moment M_x and M_y of turbine 1 and 2 are based on the different coordinate systems.



(a) Moment coord. (Turbine 1) **(b)** Moment coord. (Turbine 2)

Figure 2.3: Moment measurement coordinate system

At transition piece and monopile, strain signals are available. Therefore, it has been converted into moment so that it can be further processed into moment DEL. The detailed description is shown in Appendix A. One remark here is that the strain gauges are not calibrated and the information about the calibration is not available. Accordingly, M_{WE} and M_{SN} (hereinafter, M_X and M_Y respectively) have been used instead of M_{FA} and M_{SS} for those locations.

For accelerometer, the sensor records x and y directional acceleration Acc_x and Acc_y, and its coordinate system is shown in Figure 2.4. For acceleration, the same

measurement coordinate system is applied for both turbine 1 and 2. The summary of non-standard signals is tabulated in Table 2.3.

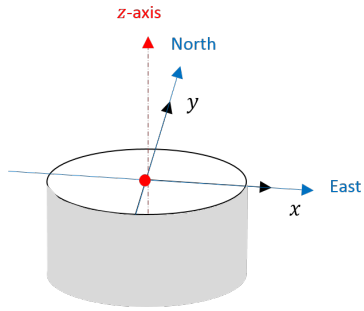


Figure 2.4: Acceleration coord. (Both turbines)

Table 2.3: Data from Sensors

Number	Signal	Symbol
1	Acceleration at Tower Top	$Acc_{FA,Top}$, $Acc_{SS,Top}$
2	Moment at Tower Top	$M_{FA,Top}$, $M_{SS,Top}$
3	Acceleration at Tower Bottom	$Acc_{FA,Btm}$, $Acc_{SS,Btm}$
4	Moment at Tower Bottom	$M_{FA,Btm}$, $M_{SS,Btm}$
5	Inclination at Tower Bottom	$\phi_{FA,Btm}$, $\phi_{SS,Btm}$
6	Moment at Transition Piece	$M_{X,TP}$, $M_{Y,TP}$
7	Moment at Monopile (Upper level)	$M_{X,MP1}$, $M_{Y,MP1}$
8	Moment at Monopile (Lower level)	$M_{Y,MP2}$, $M_{X,MP2}$

2.1.3 Wave data

In addition to the standard and non-standard signals, wave data have also been recorded. Wave has been measured at one location in the wind farm and it has been assumed that the same wave data is applicable to all the wind turbines.

The wave data includes multiple wave statistics such as H_s, T_p, Wave direction and spectral moments.

2.2 Summary of measurement data

For two turbines, all the data have been pre-processed into 10 min. statistical data as explained in previous sections. The total number of 10 min. time window within 1 month is 4644. However, the actual number of data available is less than that since measurements stop recording in some cases. Therefore, a total of 3791 and 1868 data points are available for turbine 1 and 2 respectively. The plot of mean wind speed and rotational speed for different operating conditions in both turbines is shown with the number of data points in Figure 2.6. The total number of data points according to its operating condition is tabulated in Table 2.4. One remark here is that sometimes the power generation is limited below its rated power for turbine 2 as shown in Figure 2.5. In Figure 2.5a, normalized value of 1.0 is corresponding to rated power and -1.0 is corresponding to 0. In Figure 2.5b, the normalized value of 1.0 and -1.0 is corresponding to rated power and 0 respectively, and the normalized value of around -0.7 is corresponding to the case when it generates below-rated power.

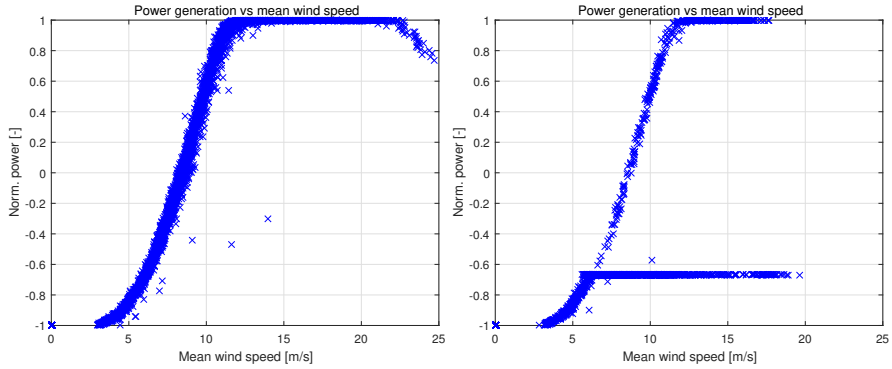
Table 2.4: Summary of data

Turbine 1	Operating			Parked	Total
	Cond. 1	Cond.2	Cond.3		
No. of data	3523	0	117	151	3791
Portion	92.9%	0.0%	3.1%	4.0%	100.0%
Turbine 2	Operating			Parked	Total
	Cond. 1	Cond.2	Cond.3		
No. of data	1529	1	64	274	1868
Portion	81.8%	0.1%	3.4%	14.7%	100.0%

In addition, total 6 standard signals (Number 1-6 in Table 2.2) have been pre-processed and total 8 non-standard signals (Number 1-8 in Table 2.3) have been pre-processed.

2.3 Correlation between different features and moment DEL at tower bottom

In this section, linear correlation between different statistical properties and moment DEL at tower bottom are investigated. By comparing correlation factors between



(a) Turbine 1 **(b)** Turbine 2

Figure 2.5: Mean power generation vs mean wind speed

different features and moment DEL, features that have a large effect on moment DEL can be recognized. In addition, only the correlation in between different features and moment DEL at tower bottom is shown in this section.

Especially, the Pearson´s correlation coefficient shown in Equation (2.1) has been used to find the correlation.

$$R = \frac{\text{cov}(\hat{X}, \hat{Y})}{\sigma_{\hat{X}}\sigma_{\hat{Y}}} = \frac{\text{E}[(\hat{X} - \mu_{\hat{X}})(\hat{Y} - \mu_{\hat{Y}})]}{\sigma_{\hat{X}}\sigma_{\hat{Y}}} \tag{2.1}$$

Where R is Pearson´s correlation coefficient, $\hat{X}$ and $\hat{Y}$ are different datasets, $\mu_{\hat{X}}$ and $\mu_{\hat{Y}}$ are mean of $\hat{X}$ and $\hat{Y}$ respectively, $\sigma_{\hat{X}}$ and $\sigma_{\hat{Y}}$ are standard deviation of $\hat{X}$ and $\hat{X}$ respectively. Pearson´s correlation coefficient R has a value between -1 and 1. The more the absolute value of R is close to 1, the larger the correlation is in between $\hat{X}$ and $\hat{Y}$.

For all the pre-processed statistical properties, the Pearson´s correlation coefficient R has been calculated. However, Pearson´s correlation coefficient R has not been calculated for moment signal at tower bottom itself since it is obvious that it will have the largest correlation and it is meaningless to investigate. As a result, the top 15 most correlated features are tabulated in Table 2.5 to 2.7 with Pearson´s correlation coefficient R.

Table 2.5: Pearson's correlation coefficient, w.r.t. Tower bottom moment DEL, Operating Cond. 1

Rank	Turbine 1			
	FA (Tower Bottom)		SS (Tower Bottom)	
	R value	Features	R value	Features
1	0.924	Tower Top Mom FA Eq. value	0.972	RNA Accel SS σ_{std}
2	0.918	RNA Accel FA σ_{std}	0.970	RNA Accel SS λ_{-1}
3	0.916	RNA Accel FA Eq. value	0.945	Tower Top Accel SS λ_{-1}
4	0.914	Tower Top Accel FA σ_{std}	0.944	Tower Top Mom SS Eq. value
5	0.910	Tower Top Mom FA Range	0.940	RNA Accel SS λ_0
6	0.900	RNA Accel FA Range	0.903	Tower Top Mom SS λ_1
7	0.895	RNA Accel FA λ_0	0.903	RNA Accel SS Eq. value
8	0.889	Tower Top Accel FA λ_0	0.893	RNA Accel SS Range
9	0.887	Tower Top Accel FA Range	0.880	RNA Accel SS Max
10	0.887	RNA Accel FA Max	0.798	Tower Top Accel SS σ_{std}
11	0.886	Tower Top Accel FA Eq. value	0.785	RNA Accel SS λ_1
12	0.882	RNA Accel FA λ_{-1}	0.769	Tower Top Accel SS λ_0
13	0.868	RNA Accel FA λ_1	0.765	Tower Top Mom SS Range
14	0.855	Tower Top Accel FA Max	0.693	Tower Top Accel SS Range
15	0.854	Tower Top Mom FA λ_1	0.686	RNA Accel SS λ_{-2}
Rank	Turbine 2			
	FA (Tower Bottom)		SS (Tower Bottom)	
	R value	Features	R value	Features
1	0.881	RNA Accel FA σ_{std}	0.982	Tower Top Mom SS Eq. value
2	0.854	RNA Accel FA λ_0	0.974	RNA Accel SS σ_{std}
3	0.850	RNA Accel FA Eq. value	0.966	RNA Accel SS λ_{-1}
4	0.844	RNA Accel FA Range	0.960	Tower Top Mom SS λ_1
5	0.843	RNA Accel FA λ_{-1}	0.958	RNA Accel SS λ_0
6	0.833	RNA Accel FA Max	0.890	RNA Accel SS Eq. value
7	0.815	Tower Top Mom FA Range	0.880	RNA Accel SS Range
8	0.798	Tower Top Accel FA σ_{std}	0.867	RNA Accel SS Max
9	0.776	Tower Top Mom FA σ_{std}	0.832	Tower Top Mom SS Range
10	0.770	RNA Accel FA λ_1	0.789	Tower Top Mom SS σ_{std}
11	0.765	Tower Top Mom FA Eq. value	0.789	Tower Top Accel SS σ_{std}
12	0.763	Tower Top Accel FA λ_0	0.782	Tower Top Accel SS λ_0
13	0.721	Tower Top Accel FA Range	0.773	RNA Accel SS λ_1
14	0.708	Tower Top Accel FA Eq. value	0.687	RNA Accel SS λ_{-2}
15	0.703	Tower Top Mom FA λ_0	0.678	Tower Top Mom SS λ_2

Table 2.6: Pearson's correlation coefficient, w.r.t. Tower bottom moment DEL, Operating Cond. 3

Rank	Turbine 1			
	FA (Tower Bottom)		SS (Tower Bottom)	
	R value	Features	R value	Features
1	0.999	RNA Accel FA Eq. value	0.999	Tower Top Mom SS Eq. value
2	0.999	Tower Top Mom FA Eq. value	0.998	RNA Accel SS σ_{std}
3	0.998	RNA Accel FA σ_{std}	0.997	Tower Top Mom SS σ_{std}
4	0.994	Tower Top Mom FA σ_{std}	0.990	RNA Accel SS Eq. value
5	0.991	Tower Top Accel FA σ_{std}	0.975	Tower Top Mom SS Range
6	0.980	Tower Bottom Incli. FA Eq. value	0.974	Tower Top Accel SS σ_{std}
7	0.980	RNA Accel FA Range	0.964	Tower Top Mom SS λ_1
8	0.977	RNA Accel FA Max	0.963	RNA Accel SS λ_0
9	0.964	Tower Top Mom FA Range	0.963	RNA Accel SS λ_{-1}
10	0.947	RNA Accel FA λ_{-1}	0.960	RNA Accel SS λ_1
11	0.946	RNA Accel FA λ_0	0.959	Tower Top Mom SS λ_2
12	0.943	RNA Accel FA λ_1	0.942	Tower Top Accel SS λ_0
13	0.939	Tower Top Accel FA λ_0	0.940	Tower Top Mom SS λ_0
14	0.936	Tower Top Mom FA λ_1	0.925	RNA Accel SS Range
15	0.931	Tower Top Mom FA λ_0	0.923	RNA Accel SS Max
Rank	Turbine 2			
	FA (Tower Bottom)		SS (Tower Bottom)	
	R value	Features	R value	Features
1	0.998	RNA Accel FA Eq. value	0.998	Tower Top Mom SS Eq. value
2	0.995	RNA Accel FA σ_{std}	0.998	RNA Accel SS σ_{std}
3	0.981	Tower Top Mom FA σ_{std}	0.985	Tower Top Mom SS σ_{std}
4	0.977	RNA Accel FA λ_0	0.983	RNA Accel SS Eq. value
5	0.977	RNA Accel FA λ_{-1}	0.983	RNA Accel SS λ_0
6	0.976	RNA Accel FA λ_1	0.982	RNA Accel SS λ_{-1}
7	0.968	RNA Accel FA Max	0.980	RNA Accel SS λ_1
8	0.967	Tower Top Mom FA Top λ_0	0.975	Tower Top Mom SS λ_1
9	0.966	RNA Accel FA Range	0.972	Tower Top Accel SS σ_{std}
10	0.942	Tower Top Mom FA Eq. value	0.966	Tower Top Mom SS Range
11	0.931	Tower Top Mom FA Range	0.958	Tower Top Accel SS λ_0
12	0.929	Tower Top Accel FA σ_{std}	0.938	Tower Top Mom SS λ_0
13	0.923	Tower Top Accel FA Var	0.903	Tower Top Mom SS Max
14	0.923	Tower Top Accel FA λ_0	0.899	Tower Bottom Accel SS σ_{std}
15	0.922	RNA Accel FA λ_2	0.885	Tower Top Accel FA σ_{std}

Table 2.7: Pearson's correlation coefficient, w.r.t. Tower bottom moment DEL, Parked Condition

Rank	Turbine 1			
	FA (Tower Bottom)		SS (Tower Bottom)	
	R value	Features	R value	Features
1	0.999	RNA Accel FA Eq. value	0.999	Tower Top Mom SS Eq. value
2	0.999	Tower Top Mom FA Eq. value	0.997	RNA Accel SS Eq. value
3	0.997	RNA Accel FA σ_{std}	0.989	Tower Top Accel SS σ_{std}
4	0.996	Tower Top Accel FA σ_{std}	0.988	RNA Accel SS σ_{std}
5	0.996	Tower Top Mom FA σ_{std}	0.987	Tower Top Accel SS Eq. value
6	0.988	Tower Top Accel FA Eq. value	0.985	Tower Top Mom SS σ_{std}
7	0.971	Tower Top Accel FA λ_0	0.971	Tower Top Accel SS λ_0
8	0.971	Tower Top Mom FA λ_1	0.969	Tower Top Mom SS λ_1
9	0.971	RNA Accel FA Range	0.968	RNA Accel SS λ_1
10	0.971	Tower Top Mom FA λ_0	0.968	RNA Accel SS λ_0
11	0.971	RNA Accel FA λ_{-1}	0.968	RNA Accel SS λ_{-1}
12	0.970	RNA Accel FA λ_1	0.967	Tower Top Mom SS λ_0
13	0.970	RNA Accel FA Max	0.949	Tower Top Mom SS λ_2
14	0.970	RNA Accel FA λ_0	0.937	Tower Top Mom SS Range
15	0.970	Tower Top Mom FA Range	0.935	RNA Accel SS λ_2
Rank	Turbine 2			
	FA (Tower Bottom)		SS (Tower Bottom)	
	R value	Features	R value	Features
1	0.999	Tower Top Mom FA Eq. value	0.998	RNA Accel SS Eq. value
2	0.999	RNA Accel FA Eq. value	0.995	Tower Top Accel SS σ_{std}
3	0.996	Tower Top Accel FA Eq. value	0.995	RNA Accel SS σ_{std}
4	0.996	Tower Top Accel FA σ_{std}	0.992	Tower Top Mom SS σ_{std}
5	0.996	Tower Top Mom FA σ_{std}	0.989	Tower Top Mom SS Eq. value
6	0.995	RNA Accel FA σ_{std}	0.978	Tower Top Accel SS λ_0
7	0.987	Tower Bottom IncliFA σ_{std}	0.978	RNA Accel SS λ_1
8	0.981	Tower Bottom Accel FA σ_{std}	0.978	RNA Accel SS λ_{-1}
9	0.973	Tower Top Mom FA λ_2	0.977	RNA Accel SS λ_0
10	0.969	RNA Accel FA Range	0.977	Tower Top Accel SS λ_{-1}
11	0.968	RNA Accel FA Max	0.975	Tower Top Accel SS Eq. value
12	0.966	Tower Top Mom FA λ_1	0.973	Tower Top Mom SS λ_0
13	0.965	Tower Top Mom FA λ_0	0.971	Tower Top Mom SS λ_1
14	0.963	RNA Accel FA λ_1	0.964	RNA Accel SS λ_2
15	0.963	Tower Top Accel FA λ_0	0.952	Tower Bottom Accel SS σ_{std}

As shown in the tables, for all conditions, acceleration at RNA, acceleration and moment at top of the tower have the highest correlation with moment DEL at bottom of the tower. Especially, standard deviation, range, and equivalent value are the most relevant statistical properties.

In addition, for operating condition 1, the highest R values are smaller compared to operating condition 3 and parked condition. As shown in Figure 2.6, operating condition 1 is corresponding to the condition where the turbine is rotating in its rated rotational speed while rotational speed is low in operating condition 3 and parked condition. Therefore, the more complicated relationship exists in operating condition 1 than other conditions. Accordingly, the relationship between different features and moment DEL at the tower bottom is more complex in operating condition 1.

In summary, it is expected that moment DEL at tower bottom can be estimated even with linear regression with standard signal (RNA acceleration) in operating condition 3 and parked condition, but it may not be applicable in operating condition 1. The result of linear regression is shown in Section 4.3.

For transition piece and monopile locations, the same analysis has been done and the results are shown in Appendix E. For transition piece and monopile, moment DEL is not expected to be estimated with linear regression with the standard signal for all conditions. However, the moment and inclination signals at the tower bottom have large R values.

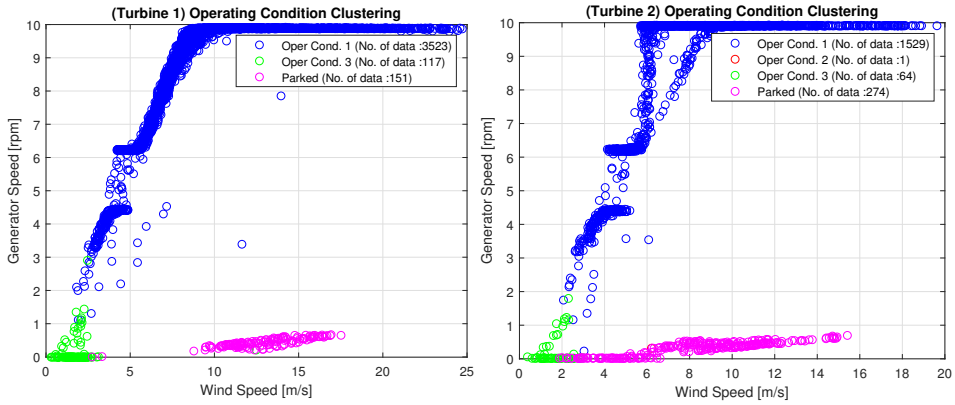
Lastly, similar studies have been performed to find the most important features to estimate moment DEL. For this study, random forest with decision tree algorithm and principal component analysis (hereinafter, PCA) have been performed. The studies are shown in Appendix F.

2.4 Plots of the measured signals

As described in previous sections, total 14 signals (6 standard signals and 8 non-standard signals) are measured and each signal has its statistical properties as shown in Table A.1. Accordingly, it is impossible to plot all the data into one 2D or 3D plot. Therefore, in this section, only some of the data are selected and put into a 2D plot to visualize the relationship between them.

2.4.1 Mean wind speed vs rotational speed

The mean rotational speed varies according to the turbine's operating conditions. Especially, when the rotational speed is plotted with mean wind speed, it can be used to give an overview of the distribution of dataset in term of operating condition. The mean wind speed versus rotational speed is shown in Figure 2.6.



(a) Turbine 1 **(b)** Turbine 2

Figure 2.6: Mean wind speed versus rotational speed

As shown in Figure 2.6, in operating condition 1 (Blue in the figure), most of the data points have high wind speed and rotational speed which indicates it is corresponding to normal operating condition. For operating condition 3 (Green in the figure), it has low wind speed and low rotational speed. For parked condition (Pink in the figure), obviously, the rotational speed is at its lowest value. For operating condition 2, it only has one data point and it has not been investigated.

One remark here is that there are two different trends in turbine 2 in operating

condition 1 in Figure 2.6. This is induced by different generations. As mentioned above, turbine 2 sometimes generates power below its rated power in which the rotational speed is different from its normal operating case.

2.4.2 Standard deviation of acceleration vs moment DEL at tower bottom

As described in Section 2.3, one of the most linearly correlated features with moment DEL at bottom of the tower is standard deviation of acceleration at top of the tower. In this section, only FA direction is shown. In Appendix D, moment DEL at the tower bottom and standard deviation of acceleration in SS direction are shown.

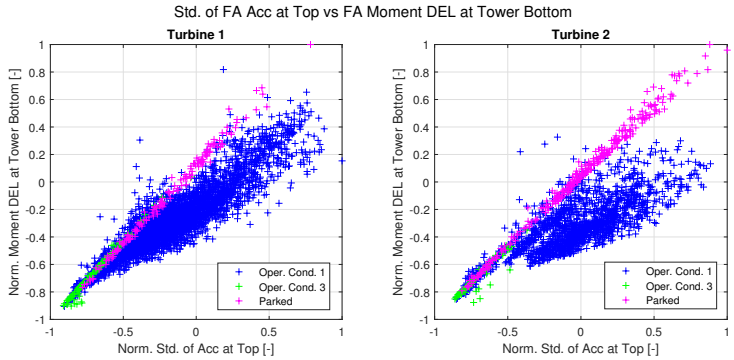


Figure 2.7: Std. of Acc at tower top vs Moment DEL at tower bottom (FA Direction)

As shown in Figure 2.7, standard deviation of acceleration at tower top is highly correlated with moment DEL at tower bottom in operating 3 and parked condition.

2.4.3 Moment DEL at tower top vs moment DEL at tower bottom

Moment DEL at tower top is obviously highly correlated with moment DEL at the tower bottom. In this section, only FA direction is shown in Figure 2.8. SS direction is shown in Appendix D.

Similar to the standard deviation of acceleration at the tower top, moment DEL at tower top has linear relationship with moment DEL at tower bottom in operating condition 3 and parked condition. The correlation is smaller in operating condition 1.

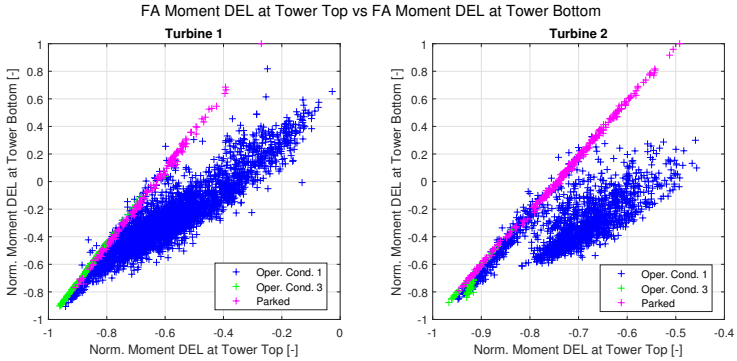


Figure 2.8: Moment DEL at tower top vs Moment DEL at tower bottom (FA Direction)

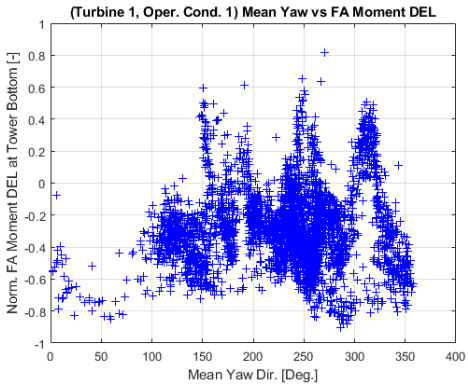
2.4.4 Mean yaw angle vs moment DEL at tower bottom

In a wind farm, there are multiple wind turbines. Therefore, larger wake turbulence is induced in downstream turbines as a result of wake effect from upstream turbines. Consequently, downstream wind turbines have higher fatigue loads (which is equivalent to moment DEL) than upstream wind turbines [4, 18]. Therefore, mean yaw angle versus moment DEL at the tower bottom plot can give an indication of wake effect from other turbines. Specifically, it is expected that larger moment DEL is induced when wind direction is corresponding to the direction where the effect from other turbines exists.

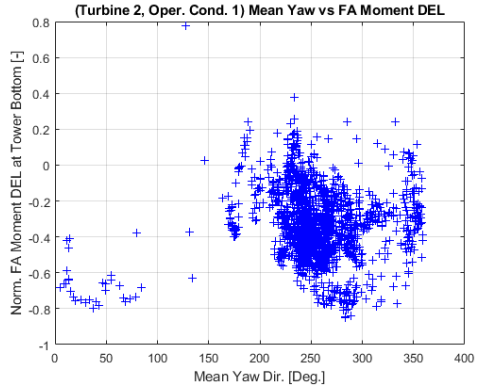
Here, it should be noted that the strength of the wake turbulence is affected by different sources such as distance in between turbines, operating conditions of the upstream turbines, turbulence intensity of atmosphere and so on. Therefore, it is possible that wake direction cannot be recognized by simply comparing mean yaw direction and moment DEL.

In this section, only operating condition 1 is considered since there is not enough data exists for other conditions. The plot is shown in Figure 2.9 and 2.10.

From Figure 2.9 and 2.10, it is clearly seen that near 150, 180, 250 and 320 deg., turbine 1 is in wake. For turbine 2, though it is not as clear as turbine 1, turbine 2 is in wake effect when yaw direction is 180, 240 and 350 deg..

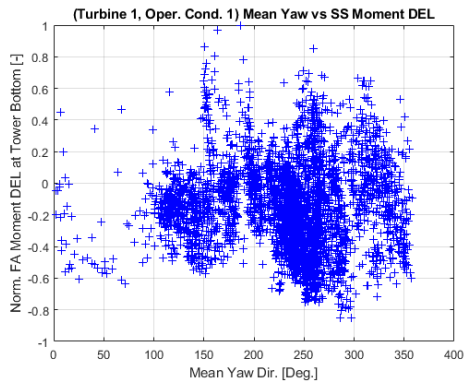


(a) Turbine 1

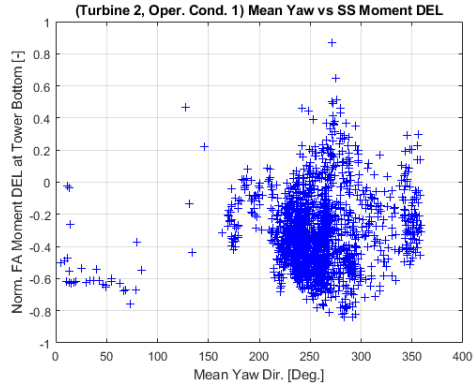


(b) Turbine 2

Figure 2.9: Mean yaw angle versus moment DEL at tower bottom (FA Direction)



(a) Turbine 1



(b) Turbine 2

Figure 2.10: Mean yaw angle versus moment DEL at tower bottom (SS Direction)

2.4.5 Comparison of moment DEL at tower bottom

In this section, moment DEL at the tower bottom is compared to figure out the wake directions. As shown in Section 2.4.4, at a specific yaw angle, the turbines are in the wake effect where the moment DEL is measured higher than free stream condition. To specify wake directions, Figure 2.9 and 2.10 have been used. As a result, the four

regions shown in Figure 2.11 and Table 2.8 are specified as directions where there is wake effect. In addition to the moment DEL comparison, the wake directions have been found from layout information. As a result, the same wake directions have been recognized except for one case. The one case is that turbine 2 is also in wake when the yaw direction is in 300∼320 deg..

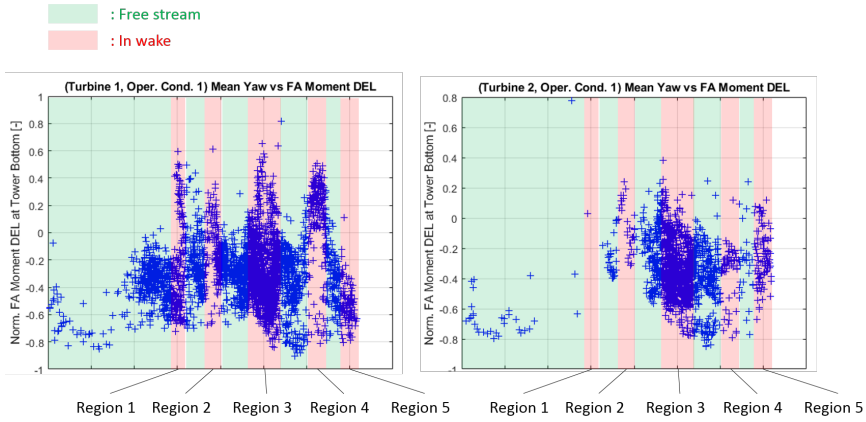


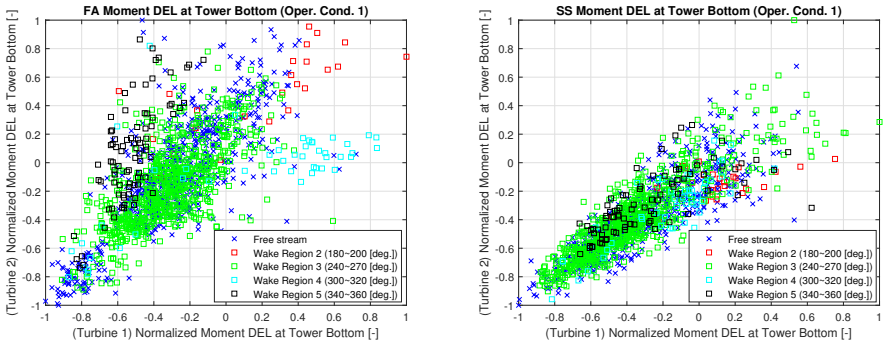
Figure 2.11: Definition of wake regions

Table 2.8: Definition of wake regions

Region	Yaw direction [deg.]	Turbines in wake
1	140 ∼160	Turbine 1
2	180 ∼200	Turbine 1 & 2
3	240 ∼270	Turbine 1 & 2
4	300 ∼320	Turbine 1 & 2
5	340 ∼360	Turbine 2

In addition, only when both turbines are in operating condition 1 has been considered in comparison since there are not enough number of data in which both turbines are in operating condition 3 or parked condition. As a result, a comparison of moment DEL at the tower bottom can be made as shown in Figure 2.12. It should be noted that there is no data point in wake region 1 since turbine 2 has only one data point in this region, and turbine 1 is not in operating condition 1 for that case.

From Figure 2.12, it is clearly visualized that normalized moment DEL is different

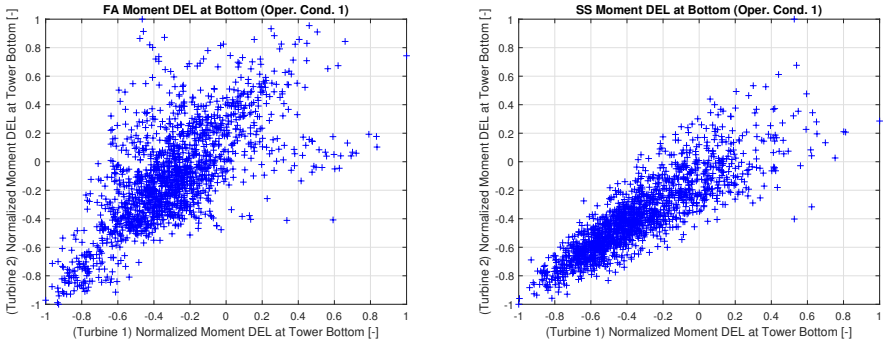


(a) FA Direction **(b)** SS Direction

Figure 2.12: Tower bottom moment DEL comparison

when only one of the two turbines is in wake (corresponding to wake region 5). In addition, even when both turbines are in wake (corresponding to wake regions 2, 3 and 4), moment DEL is quite different. The reason for this is that the distances from upstream turbines are different, and operating conditions of upstream turbines also affect the wake. Lastly, when both turbines are in free stream, both turbines have similar normalized moment DEL compared to the other cases.

When both turbines are in free stream, the comparison presented in Figure 2.13 is obtained. Obviously, moment DEL is correlated, but they are scattered a lot. Accordingly, pearson's correlation coefficient R has been calculated as 0.7939 and 0.8366 for FA and SS direction respectively.



(a) FA Direction **(b)** SS Direction

Figure 2.13: Tower bottom moment DEL comparison (only free stream condition)

Chapter 3

Neural Network

In this chapter, theories behind the feed-forward neural network and recurrent neural network are described. Firstly, a description of the feed-forward neural network are given in Section 3.1. Description of the recurrent neural network is shown in Section 3.2.

3.1 Feed forward neural network

One of the most widely used machine learning techniques is neural network. As the name 'neural network' suggests, it mimics neural network in the human brain. The similarity of biological neural network and neural network in machine learning is not discussed here but it can be found in many literature [1, 12].

The feed-forward neural network is a neural network with layer-wise architecture as shown in Figure 3.1. As the name suggests, data flow into forward direction from the input layer toward the output layer through all the hidden layers. In this section, the basic architecture and training process of the feed-forward neural network will be discussed.

3.1.1 Basic architecture & concept

The basic architecture of a neural network is shown in Figure 3.1.

First of all, a neural network consists of different *layers*. Basically, there are three categories in layers: input layer, hidden layer, and output layer. As shown in Figure

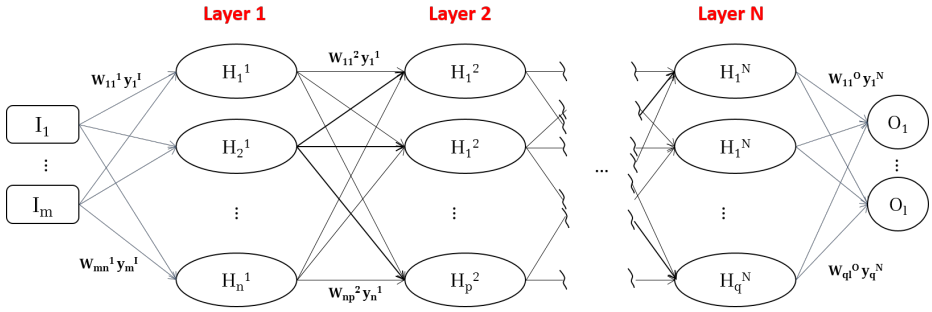


Figure 3.1: Basic structure of neural network

3.1, the number of input and output layers is 1 while there are multiple hidden layers.

Secondly, each layer consists of a number of *neurons*. In Figure 3.1, each circle where the arrows are connected is a neuron. Each neuron in hidden layers (denoted with letter 'H') has input and output signals while neurons in the input layer (denoted with letter 'I') only have output signals and neurons in the output layer (denoted with letter 'O') only have input signals. Neurons in hidden layers and output layer process the input signals to generate output signals. In addition, the role of neurons in the input layer just transfers the input values into neurons in the first hidden layer.

In addition, the notation needs to be clarified. In each layer, there are different number of neurons. In Figure 3.1, the number of neurons are m in input layer, n, p and q in hidden layer 1,2 and N respectively, and l in output layer. For neurons, the subscript stands for neuron's number and superscript stands for layer's number. For example, H_2^1 means 2nd neuron in 1st hidden layer. In addition, y and w stand for output signal and weight factor respectively. Specifically, y_i^j stands for output signal from i'th neuron in j'th layer, and w_{ij}^k stands for weight factor for output signal from i'th neuron in $(k-1)'$th layer toward j'th neuron in k'th layer.

The fundamental idea of neural network is that the relationship between input and output can be expressed as a combination of basic signals between a number of neurons. The basic signal is a signal processed with the basic function called *activation function*. The *activation function* is described in Section 3.1.2.

In addition, the goal of neural network is to find optimized weights w. To achieve the goal, *training* of neural network is required in which the optimized weights w

can be found. Specifically, during the training, weights w that give minimum error between output values (from neural network) and target values (known values) should be found. This process is described in Section 3.1.4.

3.1.2 Activation function

As explained, the activation function is used to process the input signal. The most frequently used activation functions are sigmoid, hyperbolic tangent, rectified linear unit (RELU) and pure linear functions. Each activation functions are shown in Equations (3.1) to (3.4).

Sigmoid function

$$a_{sigmoid}(z) = \frac{1}{1 + \exp(-z)} \tag{3.1}$$

hyperbolic tangent function

$$a_{hyper-tan}(z) = \tanh z \tag{3.2}$$

Rectified linear unit (RELU)

$$a_{RELU}(z) = \max(0, z) \tag{3.3}$$

Linear function

$$a_{linear}(z) = z \tag{3.4}$$

Where z is input signal which will be processed with the activation function. The plot of activation function is shown in Figure 3.2.

3.1.3 Forward propagation

Forward propagation is a process to calculate the final outputs at the output layer. As shown in Figure 3.1, output signals from the previous layer are processed with activation function and generate a new output signal.

Specifically, when the output value at a $i'th$ neuron of layer j is written as $y_i^j = a(z_i^j)$ (where $a(z)$ is a result of the pre-determined activation function), z_i^j can be calculated as shown in Equation (3.5).

$$z_i^j = \sum_{k=1}^{n} (w_{ki}^j y_k^{j-1}) - \theta_i^j \tag{3.5}$$

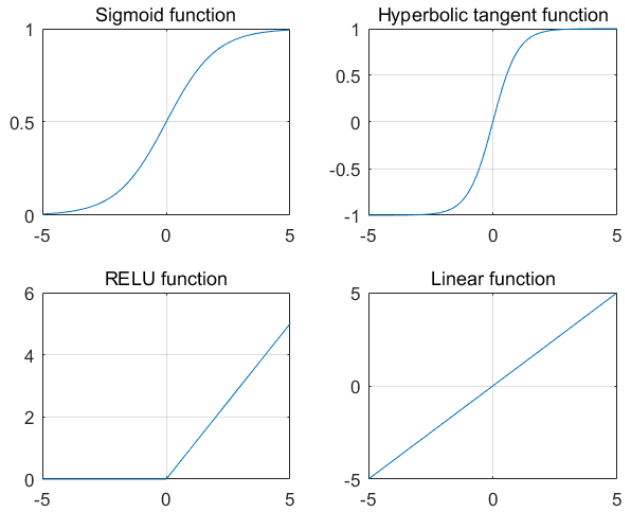


Figure 3.2: Activation functions

Where j is a layer in consideration, i is index of a neuron of layer j, k is index of a neuron of layer $j-1$ and θ_i^j is $i'th$ bias term of layer j. By using Equation (3.5), all the output y_i^j for all layers can be calculated. Eventually, the output signals at the output layer can be calculated.

As shown in Equation 3.5, weights should be known to perform the *forward propagation*. However, the first *forward propagation* stage, all the weights are unknown and therefore output y_i^j cannot be calculated. Accordingly, random initialization is used in which all the weights are randomly initialized.

3.1.4 Backward propagation

As described in Section 3.1.1, weights that give the minimum error in between output values and target values should be found by training a neural network. Backward propagation is a process to optimize the weights so that the error in between output signals at the output layer and target values can be minimized.

There are many different algorithms for training including gradient descent, gradient descent with momentum, conjugate gradient, Levenberg-Marquardt and so on. Of all the algorithms, gradient descent can be said to be the basic algorithm. In addition, the Levenberg-Marquardt algorithm is one of the most efficient training

algorithms and it has been applied in this study. Accordingly, gradient descent and Levenberg-Marquardt algorithms [22, 34] are explained.

- Gradient descent back propagation

When a vector containing target values is denoted as $\hat{\boldsymbol{y}}$ and the output vector from neural network is denoted as $\boldsymbol{y}$, the error vector $\boldsymbol{E}$ can be expressed as written in Equation (3.6).

$$\boldsymbol{E} = \frac{1}{2} \cdot (\boldsymbol{y} - \hat{\boldsymbol{y}})^2 \tag{3.6}$$

It should be noted that the element-wise square should be taken. Then, the error vector $\boldsymbol{E}$ is containing $\boldsymbol{E} = (E_1, E_2, \ldots, E_l)$. Now, let's consider the output layer. To find the weights w_{ml}^O that makes error $\boldsymbol{E}$ minimum, gradient of error vector $\boldsymbol{E}$ with respect to w_{ml}^O should be calculated. In here, chain rule can be applied as follows:

$$\begin{aligned} \frac{\partial E_l}{\partial w_{ml}^O} &= \frac{\partial E_l}{\partial y_l^O} \frac{\partial y_l^O}{\partial z_l^O} \frac{\partial z_l^O}{\partial w_{ml}^O} \\ &= (t_1 - y_l^O) \cdot a'(z_l^O) \cdot y_m^H \end{aligned} \tag{3.7}$$

where $a'(z)$ is derivative of activation function.

Then, the updated weight w_{ml}^O from gradient descent can be expressed as follows:

$$\begin{aligned} w_{ml}^O &= w_{ml}^O - \alpha^O \frac{\partial E_l}{\partial w_{ml}^O} \\ &= w_{ml}^O - \alpha^O \cdot (t_1 - y_l^O) \cdot y_m^H a'(z_l^O) \\ &= w_{ml}^O - \alpha^O \cdot \delta_l \cdot y_m^H \end{aligned} \tag{3.8}$$

where α^O is learning late at output layer (layer O) and δ_l is:

$$\delta_l = (t_1 - y_l^O) \cdot a'(z_l^O) \tag{3.9}$$

Similarly, in the hidden layer, the weight can be updated as follows:

$$\begin{aligned} w_{nm}^H &= w_{nm}^H - \alpha^H \frac{\partial E_l}{\partial w_{nm}^H} \\ &= w_{nm}^H - \alpha^H \cdot \delta_m \cdot y_n^I \end{aligned} \tag{3.10}$$

where α^H is learning late at hidden layer (layer H) and δ_m is:

$$\delta_m = \sum_{k=1}^{l} (w_{mk}^O \delta_k) \cdot a'(z_m^H) \tag{3.11}$$

In short, the weights of all layers can be updated by using gradient descent. And the gradient can be calculated by using the chain rule.

- Levenberg-Marquardt back propagation

Levenberg-Marquardt algorithm is one of the second-order backpropagation algorithms. To explain Levenberg-Marquardt algorithm, Newton's method should be discussed first.

In Newton's method, for simplicity, the weights are expressed with a single index such that $w_1, w_2, ..., w_N$ where N is the total number of weights. In Levenberg-Marquardt algorithm, gradient components $g_1, g_2, ..., g_N$ are assumed to be the functions of weights as shown in Equation (3.12).

$$\begin{aligned} g_1 &= F_1(w_1, w_2, ..., w_N) \\ g_2 &= F_2(w_1, w_2, ..., w_N) \\ &... \\ g_N &= F_N(w_1, w_2, ..., w_N) \end{aligned} \tag{3.12}$$

where $F_1, F_2, ..., F_N$ are nonlinear functions connecting weights and gradients.

Equation (3.12) can be expanded with first order Taylor expansion as shown in Equation (3.13).

$$\begin{aligned} g_1 &\approx g_{1,0} + \frac{\partial g_1}{\partial w_1}\Delta w_1 + \frac{\partial g_1}{\partial w_2}\Delta w_2 + ... + \frac{\partial g_1}{\partial w_N}\Delta w_N \\ g_2 &\approx g_{2,0} + \frac{\partial g_2}{\partial w_1}\Delta w_1 + \frac{\partial g_2}{\partial w_2}\Delta w_2 + ... + \frac{\partial g_2}{\partial w_N}\Delta w_N \\ &... \\ g_N &\approx g_{N,0} + \frac{\partial g_N}{\partial w_1}\Delta w_1 + \frac{\partial g_N}{\partial w_2}\Delta w_2 + ... + \frac{\partial g_N}{\partial w_N}\Delta w_N \end{aligned} \tag{3.13}$$

In addition, by using the definition of gradient g, second-order term in Equation (3.13) can be written as follows:

$$\frac{\partial g_i}{\partial w_j} = \frac{\partial(\frac{\partial \boldsymbol{E}}{\partial w_i})}{\partial w_j} = \frac{\partial^2 \boldsymbol{E}}{\partial w_i \partial w_j} \tag{3.14}$$

By using (3.14), Equation (3.13) can be rewritten as follows:

$$\begin{aligned}
g_1 &\approx g_{1,0} + \frac{\partial^2 \boldsymbol{E}}{\partial w_1^2}\Delta w_1 + \frac{\partial^2 \boldsymbol{E}}{\partial w_1 \partial w_2}\Delta w_2 + ... + \frac{\partial^2 \boldsymbol{E}}{\partial w_1 \partial w_N}\Delta w_N \\
g_2 &\approx g_{2,0} + \frac{\partial \boldsymbol{E}}{\partial w_2 \partial w_1}\Delta w_1 + \frac{\partial^2 \boldsymbol{E}}{\partial^2 w_2}\Delta w_2 + ... + \frac{\partial^2 \boldsymbol{E}}{\partial w_2 \partial w_N}\Delta w_N \\
&\dots \\
g_N &\approx g_{N,0} + \frac{\partial^2 \boldsymbol{E}}{\partial w_N \partial w_1}\Delta w_1 + \frac{\partial^2 \boldsymbol{E}}{\partial w_N \partial w_2}\Delta w_2 + ... + \frac{\partial^2 \boldsymbol{E}}{\partial w_N}\Delta w_N
\end{aligned} \tag{3.15}$$

When the error is minimum, the gradients should be zero. Therefore, when the error is minimum, the relationship in between gradient vector $\boldsymbol{g}$ and $\Delta \boldsymbol{w}$ can be written as shown in Equation (3.16) by setting $g_1, g_2, ...g_N = 0$ from Equation (3.15).

$$\begin{aligned}
-\boldsymbol{g} &= \boldsymbol{H}\Delta \boldsymbol{w} \\
\rightarrow \Delta \boldsymbol{w} &= -\boldsymbol{H}^{-1}\boldsymbol{g}
\end{aligned} \tag{3.16}$$

where $\boldsymbol{H}$ is Hessian matrix shown in Equation (3.17).

$$\boldsymbol{H} = \begin{bmatrix}
\frac{\partial^2 \boldsymbol{E}}{\partial w_1^2} & \frac{\partial^2 \boldsymbol{E}}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 \boldsymbol{E}}{\partial w_1 \partial w_N} \\
\frac{\partial^2 \boldsymbol{E}}{\partial w_2 \partial w_1} & \frac{\partial^2 \boldsymbol{E}}{\partial^2 w_2} & \cdots & \frac{\partial^2 \boldsymbol{E}}{\partial w_2 \partial w_N} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 \boldsymbol{E}}{\partial w_N \partial w_1} & \frac{\partial^2 \boldsymbol{E}}{\partial w_N \partial w_2} & \cdots & \frac{\partial^2 \boldsymbol{E}}{\partial^2 w_N}
\end{bmatrix} \tag{3.17}$$

As a result, weight should be updated with the amount of $\Delta \boldsymbol{w}$ and it can be expressed as follows:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - \boldsymbol{H}^{-1}\boldsymbol{g} \tag{3.18}$$

In Newton's method, to calculate $\boldsymbol{H}$, second-order derivative of total error should be calculated but it is very complicated. Accordingly, Jacobian matrix $\boldsymbol{J}$ is introduced

in Gauss-Newton Algorithm. The Jacobian matrix $\boldsymbol{J}$ is shown in Equation (3.19).

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial e_1}{\partial w_1} & \frac{\partial e_1}{\partial w_2} & \cdots & \frac{\partial e_1}{\partial w_N} \\ \frac{\partial e_2}{\partial w_1} & \frac{\partial e_2}{\partial w_2} & \cdots & \frac{\partial e_2}{\partial w_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_l}{\partial w_1} & \frac{\partial e_l}{\partial w_2} & \cdots & \frac{\partial e_l}{\partial w_N} \end{bmatrix} \tag{3.19}$$

where l is the number of outputs. Then, the gradient vector $\boldsymbol{g}$ can be expressed as follows:

$$\boldsymbol{g} = \boldsymbol{J}\boldsymbol{e} \tag{3.20}$$

where $\boldsymbol{e}$ is error vector shown in Equation (3.21).

$$\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_l \end{bmatrix} \tag{3.21}$$

In addition, the Hessian matrix H and Jacobian matrix J have the following relationship:

$$\boldsymbol{H} \approx \boldsymbol{J}^T\boldsymbol{J} \tag{3.22}$$

As a result, Equation (3.18) can be expressed with Jacobian matrix $\boldsymbol{J}$ as follows:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - (\boldsymbol{J}_k^T\boldsymbol{J}_k)^{-1}\boldsymbol{J}_k\boldsymbol{e}_k \tag{3.23}$$

In Equation (3.23), matrix $\boldsymbol{J}^T\boldsymbol{J}$ can be not invertible. In that case, weights cannot be updated. To solve the problem, In Levenberg-Marquardt algorithm, Hessian matrix is assumed as shown in Equation (3.24).

$$\boldsymbol{H} \approx \boldsymbol{J}^T\boldsymbol{J} + \mu_c\boldsymbol{I} \tag{3.24}$$

where μ_c is combination coefficient which is always positive, $\boldsymbol{I}$ is identity matrix. With this approach, the approximated Hessian matrix $\boldsymbol{H}$ is always invertible.

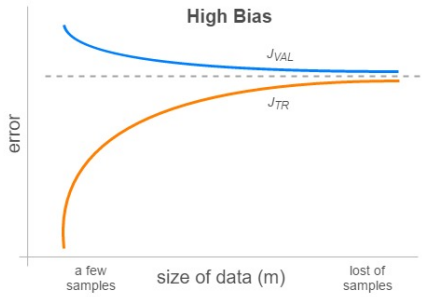
Eventually, weight update rule in Levenberg-Marquardt algorithm can be expressed as follows:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - (\boldsymbol{J}_k^T \boldsymbol{J}_k + \mu \boldsymbol{I})^{-1} \boldsymbol{J}_k \boldsymbol{e}_k \tag{3.25}$$

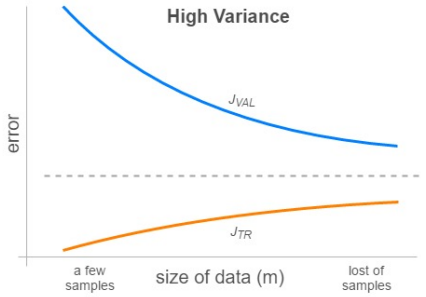
The advantage of Levenberg-Marquardt algorithm is that it is stable since the approximated Hessian matrix $\boldsymbol{H}$ is always invertible. In addition, it converges very fast compared to other algorithms [34].

3.1.5 Cross validation

After train the neural network, the trained network should be validated with cross-validation data set to see if it is under/over fitted. During cross-validation, so-called *Learning curve* is used. The *learning curve* is a graph showing a relationship between the size of the training set m and resultant training and cross-validation error. The learning curve for those two cases are shown as follows [3]:



(a) Learning curve - underfitting



(b) Learning curve - overfitting

The term 'high bias' and 'high variance' are corresponding to underfitting and overfitting respectively. In Figure above, J_{VAL} is an error with the cross-validation set and J_{TR} is an error with the training set.

In the case of underfitting, both J_{VAL} and J_{TR} are large even though the size of the training set m is large. It is corresponding to the case where the trained neural network gives inaccurate results. In this case, the number of hidden layers or neurons should be increased so that the neural network can capture a more complicated relationship between input and output.

In the case of overfitting, J_{TR} is quite low while J_{VAL} is large when the size of the training set m is large. In other words, the result can represent the training set

very well while it cannot be used for generalized data set. In this case, the number of layers or neurons can be decreased. In this case, an increase in the size of the training set m can solve the overfitting problem.

In summary, to apply the trained neural network for general data with which the neural network is not trained on, the neural network should be validated with cross-validation data set and both J_{TR} and J_{VAL} should be reasonably small.

After all these cross-validation is done, the test data set is used to calculate the accuracy of the estimation. It should be noted that the test set is a different data set from the cross-validation data set. One important remark is that the training, cross-validation and test data set is randomly decided. Therefore, accuracy can be different for different divisions of the training, cross-validation, and test data set. As a result, multiple divisions of training, cross-validation and test set should be applied for appropriate accuracy calculation.

3.1.6 Vanishing and exploding of gradient

As explained above, weights are updated during the training process with gradient. When the neural network is deep, during the weight update process, these gradients are multiplied multiple times. It causes a problem called *vanishing* or *exploding* of gradient.

Specifically, when the gradient is smaller than 1, the multiple multiplications of gradient causes *vanishing* of gradients. On the other hand, when the gradient is larger than 1, the gradient is *exploding*. In either case, the updated weights will not converge into optimized points where the error is minimum.

When vanishing and exploding occur, the neural network would give very inaccurate results. Especially, the error would not converge in the process of training. Therefore, it is easily be captured.

To solve the vanishing and exploding, many different approaches can be taken [11, p. 60]. First of all, one can restructure a neural network to have shallower architecture. Secondly, for a recurrent neural network, long short-term memory (LSTM) architecture can be used. In addition, dropout can be used to prevent overfitting [29]. Lastly, gradient clipping can be applied to prevent exploding. With gradient clipping, gradients are limited during the training process if the gradients exceed a threshold value.

3.2 Recurrent neural network (LSTM)

When the data contain sequential dependency, the recurrent neural network can be used to include this sequential dependency. There are many different recurrent neural network structures including fully-recurrent neural network, echo-state neural network, gated recurrent unit (GRU) and long short-term memory (LSTM).

The biggest challenge to train the recurrent neural network is *vanishing and exploding gradient problem*. It is briefly explained in Section 3.1.6. Especially when the recurrent neural network deals with long-term memory, it can be regarded as deep neural network as shown in Figure 3.4. Specifically, $\mathbf{Input_1}$ and $\mathbf{Output_t}$ in Figure 3.4 are connected through a lot of neural network structures, and it can be regarded as deep neural network. In this process, *vanishing and exploding gradient problem* is frequently taken place.

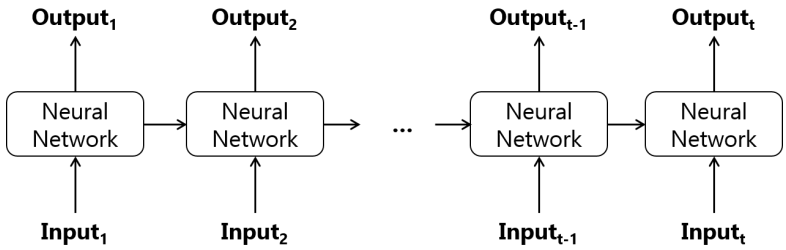


Figure 3.4: Expanded basic RNN

To solve the problem, different kinds of recurrent neural networks have been developed. Of all the structures, LSTM has been applied in this study and it will be described in this section.

3.2.1 Basic architecture & concept

First of all, The main idea of LSTM is to prevent the *vanishing* problem by allowing long-term memory to flow without being changed. To achieve it, LSTM uses cell state c. The basic structure of LSTM is shown in Figure 3.5 [37].

In Figure 3.5, '$\times$' and '$+$' in green circle is multiplication and summation process respectively. In addition, 'σ' and 'tanh' in gray rectangular is sigmoid and hyperbolic tangent activation function shown in Section 3.1.2 respectively. Added to that, $\boldsymbol{c}$ is cell state applied to prevent *vanishing* problem. In addition, $\boldsymbol{h}_t$ is hidden state and

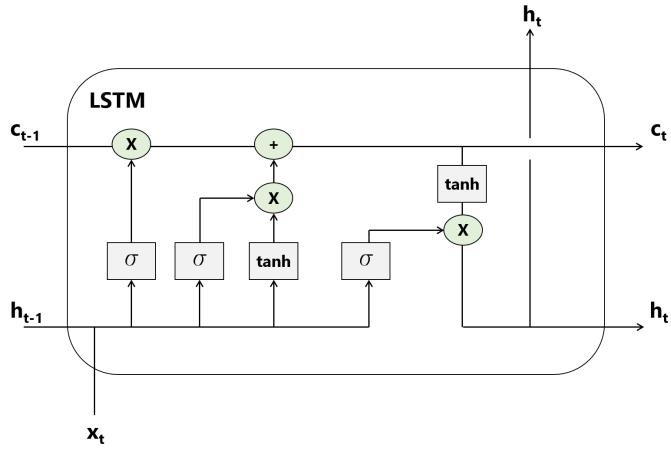


Figure 3.5: LSTM architecture

x_t is input. After calculating hidden state $\boldsymbol{h}_t$, output $\boldsymbol{y}_t$ can be calculated by applying weight into the hidden state as shown in Equation 3.26.

$$\boldsymbol{y}_t = \boldsymbol{w}_{hy}\boldsymbol{h}_t \tag{3.26}$$

As described, the main idea of LSTM is to use cell state c to prevent *vanishing* problem. As shown in Figure 3.5, this cell state c is flowing through the whole layer with one multiplication and summation process. It is also flowing from the start of the sequence to the end. In this process, LSTM can add or remove information to cell state. To decide what to be added or removed, LSTM uses three gates; forget gate, input gate and output gate and it is explained in Section 3.2.2.

3.2.2 Gates

In LSTM, there are three gates: forget gate, input gate, and output gate. The role of these gates is to decide what to add, what to remove and what to transfer to cell state.

First of all, forget gate is shown in Figure 3.6.

As its name suggests, forget gate decides how much information it forgets from cell states. As a result of sigmoid function denoted as 'σ' in Figure 3.6, f_t has a value

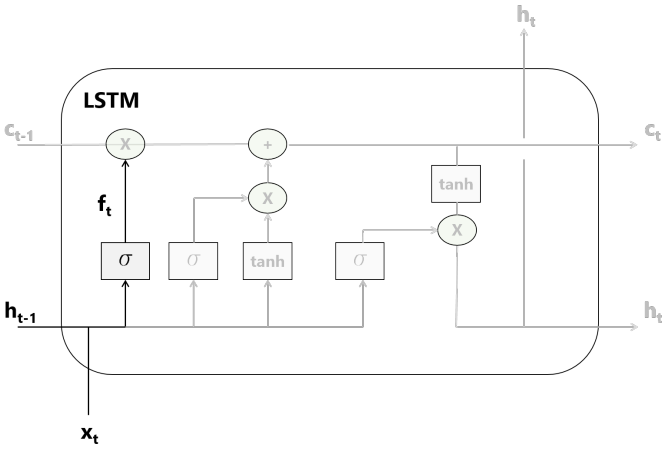


Figure 3.6: Forget gate

in between 0 to 1. If it is 1, that means it will keep the full data while it forgets full data if it is 0. The process can be expressed as shown in Equation (3.27).

$$\boldsymbol{f}_t = \sigma(\boldsymbol{w}_f \boldsymbol{x}_t + \boldsymbol{u}_f \boldsymbol{h}_{t-1} + \boldsymbol{\theta}_f) \tag{3.27}$$

where $\boldsymbol{x}_t$ is input at time step t, $\boldsymbol{h}_{t-1}$ is hidden state at time step $t-1$, $\boldsymbol{w}_f$ is forget weight matrix for input $\boldsymbol{x}_t$, $\boldsymbol{u}_f$ is forget weight matrix for hidden state $\boldsymbol{h}_{t-1}$ and $\boldsymbol{\theta}_f$ is forget bias.

Secondly, input gate is shown in Figure 3.7.

Input gate decides what will be added to cell state. In consists of sigmoid activation function and hyperbolic tangent activation function. Through the hyperbolic tangent activation function, new candidate $\tilde{\boldsymbol{c}}_t$ is calculated. Then, through the sigmoid activation function, it decides how much of the new candidate $\tilde{\boldsymbol{c}}_t$ will be added to cell state with the portion of $\boldsymbol{i}_t$. It can be expressed as shown in Equation (3.28) and (3.29).

$$\boldsymbol{i}_t = \sigma(\boldsymbol{w}_i \boldsymbol{x}_t + \boldsymbol{u}_i \boldsymbol{h}_{t-1} + \boldsymbol{\theta}_i) \tag{3.28}$$

$$\tilde{\boldsymbol{c}}_t = \tanh(\boldsymbol{w}_c \boldsymbol{x}_t + \boldsymbol{u}_c \boldsymbol{h}_{t-1} + \boldsymbol{\theta}_c) \tag{3.29}$$

where subscript i is for $\boldsymbol{i}_t$ calculation through sigmoid activation function and

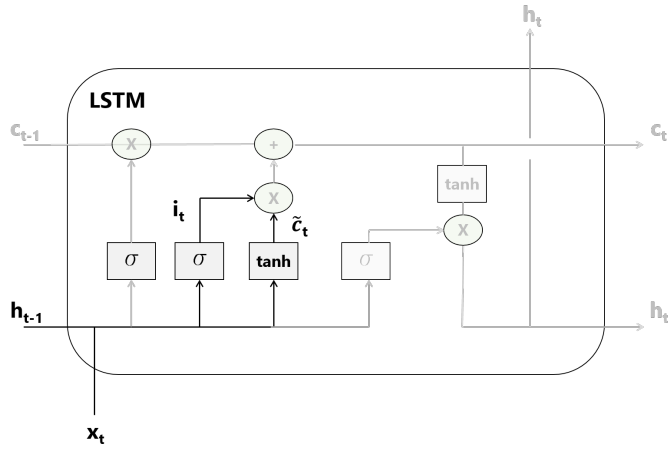


Figure 3.7: Input gate

subscript c for new candidate calculation through hyperbolic tangent activation function.

With the values calculated from forget gate and input gate, new cell state $\boldsymbol{c}_t$ can be calculated. This process is shown in Figure 3.8.

With the calculated forget rate $\boldsymbol{f}_t$, input rate $\boldsymbol{i}_t$ and new candidate $\tilde{\boldsymbol{c}}_t$, new cell state $\boldsymbol{c}_t$ can be calculated as shown in Equation (3.30).

$$\boldsymbol{c}_t = \boldsymbol{f}_t \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \tilde{\boldsymbol{c}}_t \tag{3.30}$$

The first term in Equation (3.30) is a process to forget the data by multiplying forget ratio $\boldsymbol{f}_t$ with $\boldsymbol{c}_{t-1}$. The second term is a process to add new candidate $\tilde{\boldsymbol{c}}_t$ with ratio of $\boldsymbol{i}_t$.

The last gate of LSTM is output gate and it is shown in Figure 3.9.

Output gate decides how much data will be transferred to the next sequence. It also consists of sigmoid activation function and hyperbolic tangent activation function. Similar to input gate, sigmoid activation function gives an output ratio $\boldsymbol{o}_t$ and hyperbolic tangent activation function processes the updated cell state $\boldsymbol{c}_t$ to be in a range of -1 to 1. As a result, the hidden state $\boldsymbol{h}_t$ can be calculated as shown in Equation (3.31).

$$\boldsymbol{h}_t = \boldsymbol{o}_t \tanh(\boldsymbol{c}_t) \tag{3.31}$$

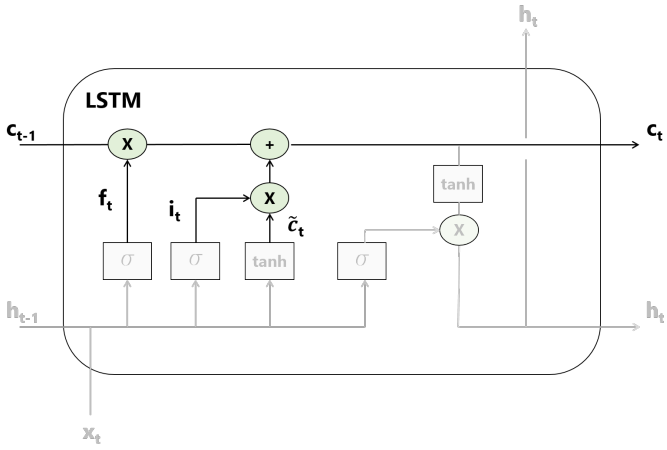


Figure 3.8: Update cell state

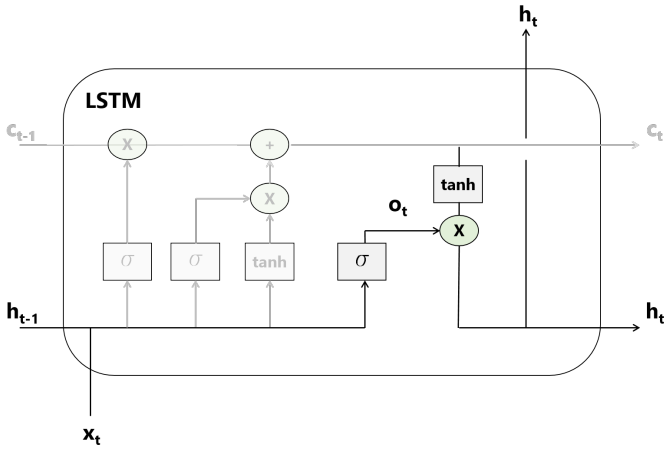


Figure 3.9: Output gate

In summary, through forget, input and output gates, cell state and hidden state can be updated. In this process, all the weights in all gates are variables required to be optimized by training.

3.2.3 Forward & Backward propagation

Training (Forward & backward propagation) of the recurrent neural network is the same as the feed-forward neural network. The only difference is in the backpropagation algorithm. As explained in Section 1.5, Deep Learning Toolbox in MATLAB has been used to build both the feed-forward neural network and recurrent neural network. In Deep Learning Toolbox in MATLAB, only three optimization algorithms are available for the recurrent neural network which are stochastic gradient descent with momentum (SGDM), RMSProp and ADAM. Of the three algorithms, ADAM optimizer has been applied in this study.

Chapter 4

Individual turbine level load estimation

In this chapter, individual turbine level load estimation is described. First of all, the introduction to the dataset used in individual turbine load estimation and description of the base case are shown in Section 4.1. The accuracy measurement method is introduced in Section 4.2. The linear regression analysis for operating condition 3 and parked condition is shown in Section 4.3. Then, the result with the feed-forward neural network is described in Section 4.4. The result with the recurrent neural network is shown in Section 4.5. In addition, various case studies have been performed to answer one of the research questions: ***'what is the minimum number of sensors for accurate estimation?'***. The case study results are presented in Section 4.6. Additionally, it has been investigated if any improvement can be achieved with the division of operating condition 1 and PCA. these applications are explained in Section 4.7. In addition, to answer the research question of ***'what is the required number of data for accurate estimation?'***, accuracy level has been compared with respect to the size of training data. It is described in Section 4.8. Lastly, conclusion is given in Section 4.9.

4.1 Applied dataset

For individual turbine load estimation, the base case has been applied to decide hyperparameters of the neural network including the number of hidden layers and

neurons. In Section 4.1.1, input and target data are described. Then, the base case is described in Section 4.1.2.

4.1.1 Input and target signals

For individual turbine level load estimation, the pre-processed data described in Chapter 2 have been applied. It is shown in Table 4.1.

Table 4.1: Base case

Input			
Category	Number	Signal	Statistical properties
Standard	1	Acceleration	All (1-5 in Table A.1)
	2	Blade Pitch angle	All (1-5 in Table A.1)
	3	Wind Speed	All (1-5 in Table A.1)
	4	Active Power	All (1-5 in Table A.1)
	5	Rotational Speed	All (1-5 in Table A.1)
	6	Yaw Direction	All (1-5 in Table A.1)
Non-standard	1	Acceleration at Tower Top	All (1-5 in Table A.1)
	2	Moment at Tower Top	All (1-5 in Table A.1)
	3	Acceleration at Tower Bottom	All (1-5 in Table A.1)
	5	Inclination at Tower Bottom	All (1-5 in Table A.1)
Target			
Category	Number	Signal	Statistical properties
Non-standard	4	Moment at Tower Bottom	DEL (4 in Table A.1)
	6	Moment at Transition Piece	DEL (4 in Table A.1)
	7	Moment at Monopile (Upper level)	DEL (4 in Table A.1)
	8	Moment at Monopile (Lower level)	DEL (4 in Table A.1)

4.1.2 Base case

As a base case, all the input and target shown in Table 4.1 have been used. Basically, the base case uses all the measurements as input. Therefore, it is expected that the results would be quite accurate with the base case.

4.2 Accuracy measurement

To evaluate the accuracy of estimation from neural networks, two methods have been used in this study.

Firstly, Pearson's correlation coefficient has been applied. As described in Section 2.3, Pearson's correlation coefficient R measures the linear correlation between two data.

In addition, the mean absolute percentage error (MAPE) has been used. MAPE can be calculated as shown in Equation (4.1).

$$\text{MAPE} = 100\% \times \frac{1}{n} \sum_{i} \frac{y_i - \hat{y}_i}{\hat{y}_i} \tag{4.1}$$

4.3 Linear regression (for subsets of operating conditions)

As explained in Section 2.2, for parked condition and operating condition 3 where the turbine is operating with low rotational speed and low wind speed, Pearson's correlation coefficient in between moment DEL at **tower bottom** and some standard signals were found to be almost 1. It means moment DEL at the tower bottom can be accurately estimated with linear regression. However, for **transition piece** and **monopile**, Pearson's correlation coefficient is quite off from 1 with standard signals. However, Pearson's coefficient is close to 1 with some non-standard signals such as moment DEL and a standard deviation of inclination at the tower bottom. Accordingly, to investigate if linear regression can be applied for operating condition 3 and parked condition, linear regression has been applied to estimate moment DEL with features with Pearson's correlation coefficient of almost 1.

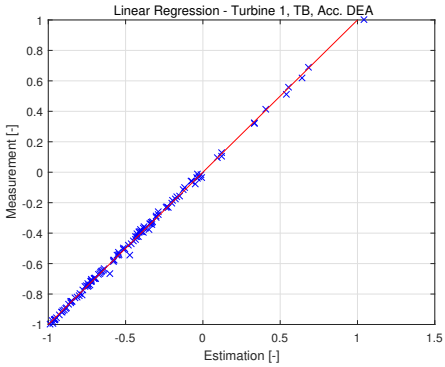
From Table 2.6, 2.7 and Tables in Appendix E, the features which have Pearson's coefficient close to 1.0 are extracted and tabulated in Table 4.2.

Table 4.2: Linearly correlated features with moment DEL at tower bottom

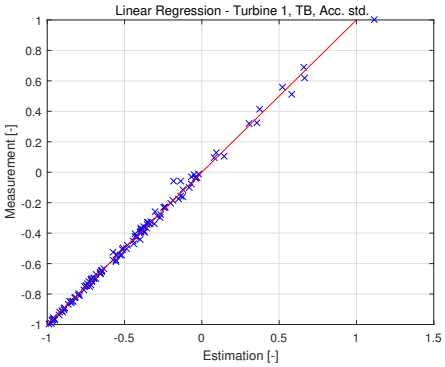
Category	At tower bottom	At transition piece and monopile
Standard	RNA Accel Eq. value	-
	RNA Accel σ_{std}	-
Non-standard	Tower Top Mom σ_{std}	Tower Bottom Mom Eq. value
	-	Tower Bottom Inclination σ_{std}

For tower bottom, the results of linear regression are shown in Table 4.3 and Table 4.4 for FA direction. In addition, comparison between estimation and measurement is shown in Figure 4.1 and 4.2 for operating condition 3 and parked condition respectively.

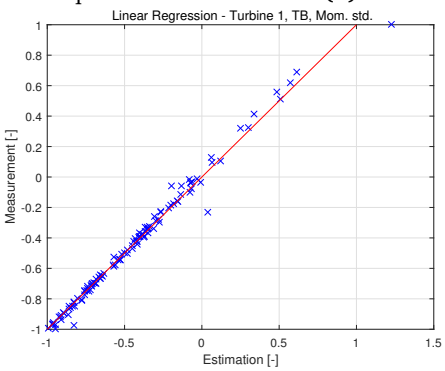
For transition piece and monopile, the results are shown in Table 4.5 and Table 4.6 for X direction.



(a) w.r.t. RNA Acc. Eq. value



(b) w.r.t. RNA Acc. std.



(c) w.r.t. Tower Top Mom. std.

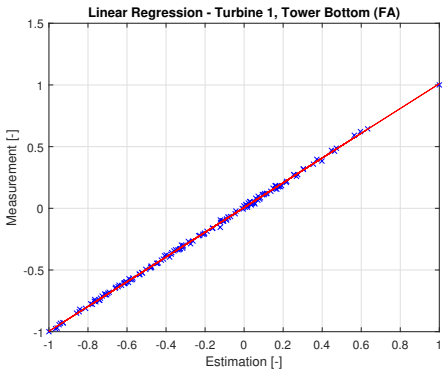
Figure 4.1: Results for linear regression in between FA directional tower bottom moment DEL and different features, Operating condition 3, Turbine 1

Table 4.3: Results for linear regression in between FA directional tower bottom moment DEL and different features, Operating condition 3

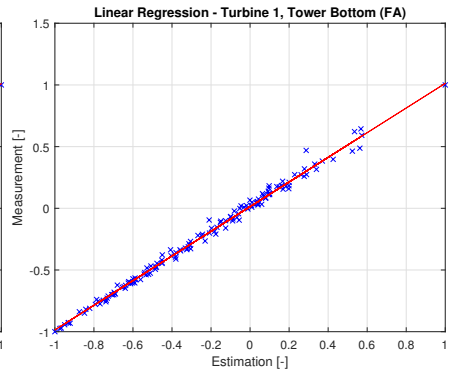
Location	Feature	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
Tower Bottom	RNA Accel Eq. value	0.999	1.86%	0.999	1.71%
	RNA Accel σ_{std}	0.998	2.64%	0.996	3.33%
	Tower Top Mom σ_{std}	0.998	2.67%	0.997	3.83%

Table 4.4: Results for linear regression in between FA directional tower bottom moment DEL and different features, Parked condition

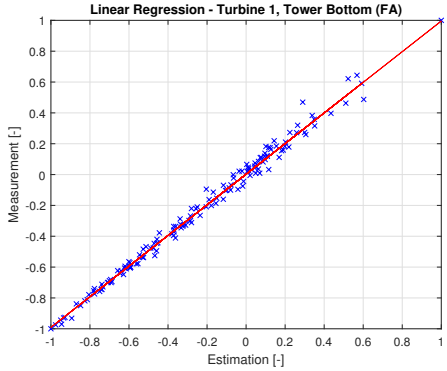
Location	Feature	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
Tower Bottom	RNA Accel Eq. value	0.999	1.39%	0.999	2.60%
	RNA Accel σ_{std}	0.997	2.21%	0.995	3.55%
	Tower Top Mom σ_{std}	0.996	2.38%	0.996	2.39%



(a) w.r.t. RNA Acc. Eq. value



(b) w.r.t. RNA Acc. std.



(c) w.r.t. Tower Top Mom. std.

Figure 4.2: Results for linear regression in between FA directional tower bottom moment DEL and different features, Parked condition, Turbine 1

Table 4.5: Results for linear regression in between X directional transition piece and monopile moment DEL and different features, Operating condition 3

Location	Feature	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
Transition piece	Tower Bottom Eq. value	0.998	1.13%	0.987	4.89%
	Tower Bottom Mom Incli. σ_{std}	0.995	3.45%	0.971	5.35%
Upper Monopile	Tower Bottom Eq. value	0.995	4.18%	0.979	5.01%
	Tower Bottom Mom Incli. σ_{std}	0.996	2.73%	0.972	6.97%
Lower Monopile	Tower Bottom Eq. value	0.977	6.12%	0.973	5.69%
	Tower Bottom Mom Incli. σ_{std}	0.968	9.19%	0.964	8.99%

Table 4.6: Results for linear regression in between X directional transition piece and monopile moment DEL and different features, Parked condition

Location	Feature	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
Transition piece	Tower Bottom Eq. value	0.997	0.91%	0.999	0.90%
	Tower Bottom Mom Incli. σ_{std}	0.946	6.82%	0.988	4.57%
Upper Monopile	Tower Bottom Eq. value	0.993	2.86%	0.998	2.57%
	Tower Bottom Mom Incli. σ_{std}	0.953	6.96%	0.988	4.64%
Lower Monopile	Tower Bottom Eq. value	0.979	4.55%	0.993	5.02%
	Tower Bottom Mom Incli. σ_{std}	0.958	6.50%	0.986	6.01%

As shown in the results, linear regression gives quite accurate estimation at the tower bottom for both operating condition 3 and parked condition. The same result has been found for SS direction.Specifically, only with standard signals, the estimation is accurate with MAPE of less than 4%. Especially, when the equivalent value of RNA acceleration has been used, MAPE of less than 3% can be achieved.

For transition piece and monopile, linear regression gives less accurate results. Especially, when the standard deviation of inclination has been used, estimation was not so accurate. However, when moment DEL at the tower bottom has been used, estimation was quite accurate. Specifically, linear regression gives MAPE of less than about 5% for operating condition 3 and 3% for parked condition at transition piece and upper monopile level. However, at lower monopile level, the accuracy level is lower than that with MAPE of around 6% for operating condition 3 and around 5% for parked condition. In addition, the higher error is expected at the fatigue critical location near mudline.

4.4 Feed forward artificial neural network

First of all, as described in Section 1.5, Deep learning toolbox in MATLAB 2018b has been used for the feed-forward neural network application. Throughout this study, Levenberg-Marquardt backpropagation algorithm explained in Section 3.1.4 has been used. Its related parameters are tabulated in Table 4.7.

Table 4.7: Parameters used for feed forward neural network training

Parameter	Value
Max. number of epoch	1000
Min. gradient	10^{-7}
Max. validation check	10

In addition, all the features and targets have been normalized in a range of -1∼1 with min-max normalization. About the activation function, the tangent-sigmoid activation function has been applied for the hidden layer and output layer.

Lastly, 70% of the total data set has been used for training and 15% has been used as cross-validation and test data set respectively.

For all cases, ten (10) different divisions of training, cross-validation and test set have been made. Then, the averages of the accuracy measurements have been taken as the representative accuracy measurements.

In addition to that, for each division of training, cross-validation and test set, ten (10) different random weights initialization have been applied. Then, the accuracy measurements of averaged estimation have been calculated and used as the representative value. It has been done to prevent the feed-forward neural network being optimized for local minimum, instead of the global minimum.

4.4.1 Decision of hyper parameters

With the base case described in Section 4.1.2, hyperparameters including the number of hidden layers and neurons of the feed-forward neural network have been decided. To decide the hyperparameters, only the moment DEL at the tower bottom has been used.

Specifically, five different number of neurons (4,8,12,16 and 20) and two different number of hidden layers (1 and 2) are compared. The comparison is based on Pearson's correlation coefficient R and MAPE of the test data set.

As a result, the following Pearson's correlation coefficient R and MAPE have been calculated for Turbine 1.

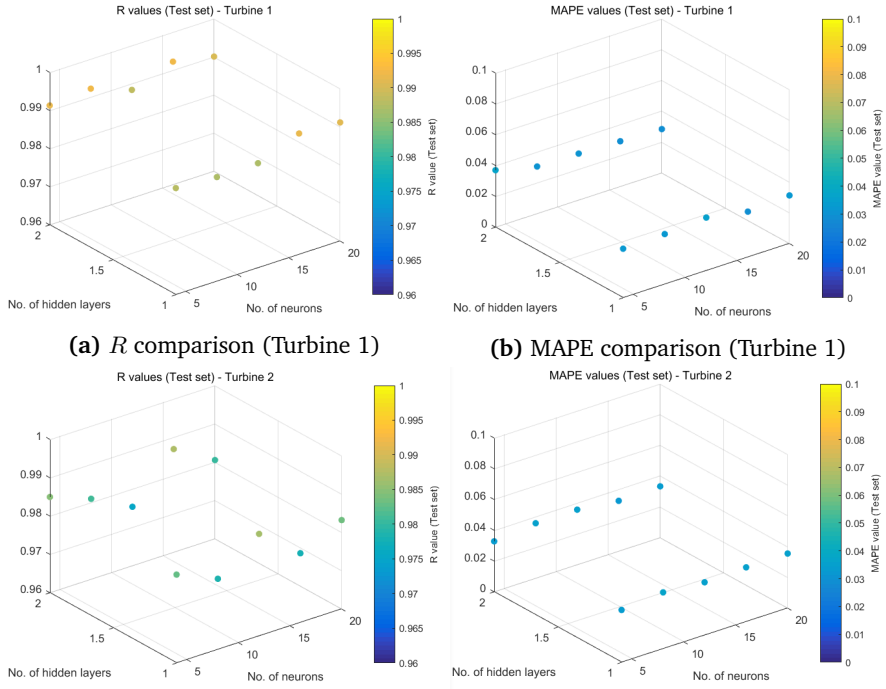
As shown in Figure 4.3, the accuracy level has no significant difference. In detail, the accuracy measurements for the simplest (smallest number of neurons and number of layers) and the most complex architecture (largest number of neurons and number of layers) have almost the same values as tabulated in Table 4.8. As a result, the simplest feed-forward neural network architecture with 4 neurons in one hidden layer has been used for further studies.

Table 4.8: Comparison of accuracy between simplest and the most complex structures

	Turbine 1		Turbine 2	
	R value	MAPE	R value	MAPE
1 Layer, 4 neurons	0.988	3.19%	0.983	3.40%
2 Layers, 20 neurons	0.990	2.85%	0.980	3.34%

4.4.2 Estimation result for base case

For the base case, moment DEL estimations for all locations have been made. The purpose of the base case application is to investigate if the feed-forward neural network can give accurate estimation, and the achievable level of estimation accuracy. As briefly mentioned in Section 4.1.2, the base case application is expected to give quite accurate estimation since all the measurement data have been used.



(a) R comparison (Turbine 1)

(b) MAPE comparison (Turbine 1)

(c) R comparison (Turbine 2)

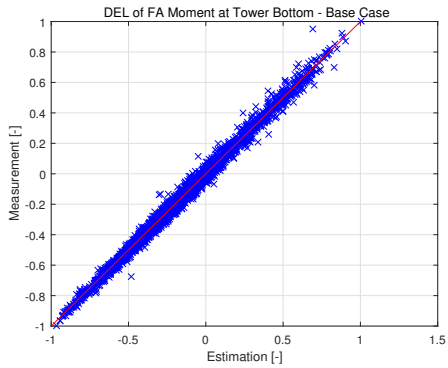
(d) MAPE comparison (Turbine 2)

Figure 4.3: Accuracy comparison for different number of neurons and layers

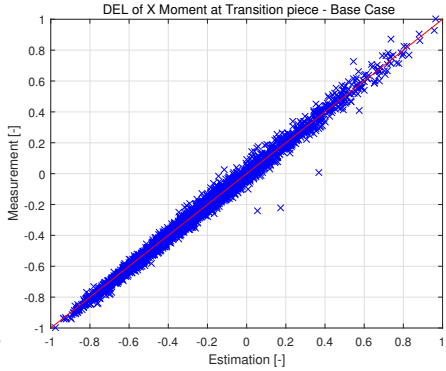
In Figure 4.4, FA directional moment DEL at tower bottom and X directional moment DEL at transition piece and monopile for operating condition 1 of turbine 1 are shown. In Figure 4.5, the same results for operating condition 3 are shown. It should be noted that the results for all the different sets including training/cross validation and test sets are shown in both figures. However, to explain the overfitting problem, distinction in between three sets has been made on Figure 4.5.

In Table 4.9, 4.10, 4.11 and 4.12, all the accuracy measurements are shown for tower bottom, transition piece, monopile (upper level) and monopile (lower level) estimations respectively.

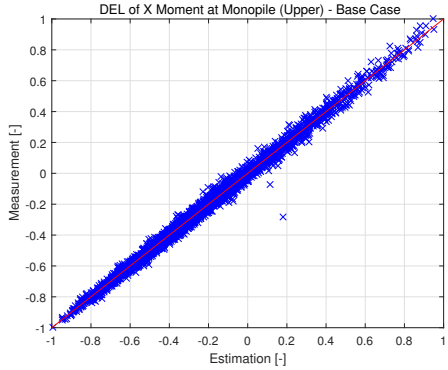
From the results, it is found that the estimation is accurate for operating condition 1 in all locations. Specifically, MAPE is less than 4% for all locations except lower monopile location where MAPE is about 5%. In addition, as shown in Figure



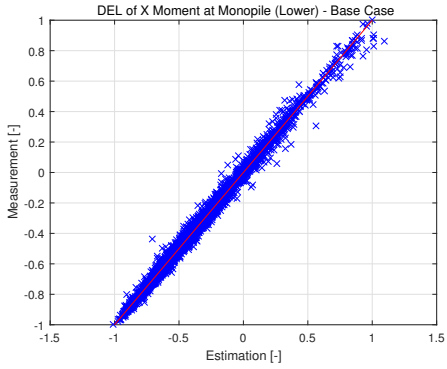
(a) Tower bottom, FA moment DEL



(b) Transition piece, X moment DEL



(c) Monopile (upper), X moment DEL

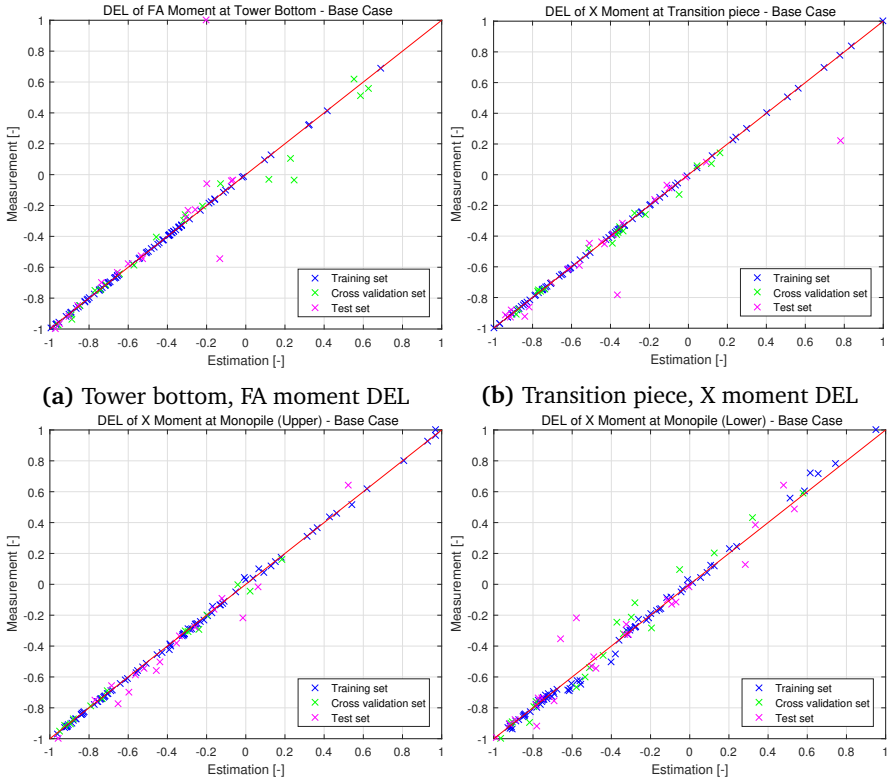


(d) Monopile (lower), X moment DEL

Figure 4.4: Results of individual turbine level load estimation, FA and X directions, Operating Condition 0, Turbine 1, Base Case, Feed forward neural network

Table 4.9: Results of individual turbine level load estimation, Tower bottom, Base Case, Feed forward neural network

	Turbine 1 (Tower bottom)				Turbine 2 (Tower bottom)			
	FA		SS		FA		SS	
	R value	MAPE	R value	MAPE	R value	MAPE	R value	MAPE
Oper. Cond. 1	0.988	3.19%	0.997	1.52%	0.983	3.40%	0.995	1.55%
Oper. Cond. 3	0.946	7.80%	0.972	7.64%	0.933	11.65%	0.751	15.82%
Parked Cond.	0.945	7.19%	0.977	3.39%	0.984	2.69%	0.991	1.48%



(a) Tower bottom, FA moment DEL

(b) Transition piece, X moment DEL

(c) Monopile (upper), X moment DEL

(d) Monopile (lower), X moment DEL

Figure 4.5: Results of individual turbine level load estimation, X directions, Operating Condition 3, Turbine 1, Base Case, Feed forward neural network

Table 4.10: Results of individual turbine level load estimation, Transition piece, Base Case, Feed forward neural network

	Turbine 1 (Transition piece)				Turbine 2 (Transition piece)			
	X		Y		X		Y	
	R value	MAPE	R value	MAPE	R value	MAPE	R value	MAPE
Oper. Cond. 1	0.990	3.24%	0.988	3.02%	0.982	3.45%	0.982	3.68%
Oper. Cond. 3	0.957	8.23%	0.903	9.50%	0.923	16.20%	0.951	11.87%
Parked Cond.	0.972	6.81%	0.986	4.89%	0.990	3.30%	0.990	3.23%

Table 4.11: Results of individual turbine level load estimation, Monopile (upper), Base Case, Feed forward neural network

	Turbine 1 (Monopile (upper))				Turbine 2 (Monopile (upper))			
	X		Y		X		Y	
	R value	MAPE	R value	MAPE	R value	MAPE	R value	MAPE
Oper. Cond. 1	0.992	3.37%	0.989	3.00%	0.993	2.88%	0.988	3.19%
Oper. Cond. 3	0.881	8.17%	0.823	11.77%	0.903	16.88%	0.953	11.09%
Parked Cond.	0.975	5.74%	0.972	5.43%	0.975	4.40%	0.990	3.55%

Table 4.12: Results of individual turbine level load estimation, Monopile (lower), Base Case, Feed forward neural network

	Turbine 1 (Monopile (lower))				Turbine 2 (Monopile (lower))			
	X		Y		X		Y	
	R value	MAPE	R value	MAPE	R value	MAPE	R value	MAPE
Oper. Cond. 1	0.984	5.22%	0.987	4.48%	0.985	3.96%	0.985	4.90%
Oper. Cond. 3	0.916	7.72%	0.924	14.68%	0.869	14.88%	0.913	19.11%
Parked Cond.	0.981	5.37%	0.950	6.93%	0.980	3.74%	0.990	3.39%

4.4, results are not biased. At the lower monopile location, moment is induced by wind as well as wave, and the wave-induced moment is larger at the lower monopile location. As a result, higher inaccuracy is expected at the lower monopile location since the base case does not include any information about wave as its input.

In addition, accuracy is low in operating condition 3 with MAPE of above 8% for turbine 1 and MAPE of above 11% for turbine 2. This inaccuracy is induced by the small number of the training set. Concretely, overfitting problem caused by the small number of training set induces the inaccuracy. As shown in Table 2.4, only 117 and 64 data points exist for turbine 1 and turbine2 respectively for operating condition 3. Among them, only 70% (82 and 45 for turbine 1 and turbine 2 respectively) have been used for training and it is not enough to train the neural network. Therefore, as shown in Figure 4.5, results are well matched only for the training set, but for cross-validation and test set, the result is not well matched since the neural network has overfitting problem.

A similar trend is found in parked condition. For parked condition, MAPE has been calculated as above 5% for turbine 1 and about 4% for turbine 2. As shown in Table 2.4, the number of data points is 151 and 274 and 70% of those data points (106 and 192 respectively) have been used for training. Since the number of data points is larger than operating condition 3, the results are more accurate for parked condition. Especially, for turbine 2 with 192 data points, the result is quite accurate

with MAPE of about 4%.

4.5 Recurrent neural network (LSTM)

In addition to the feed-forward artificial neural network, the recurrent neural network has also been applied. The idea behind applying recurrent neural network is to investigate if more accurate estimation can be achieved. Specifically, the estimation accuracy would be increased if the moment DELs from previous time stamps have a relationship with moment DEL at time stamp in question since recurrent neural network gives accurate estimation when the data has a sequential dependency.

Of all the recurrent neural network structures, LSTM has been used for this study. By trial and error, the neural network architecture that gives minimum MAPE has been decided. The recurrent neural network architecture is shown in Figure 4.6.

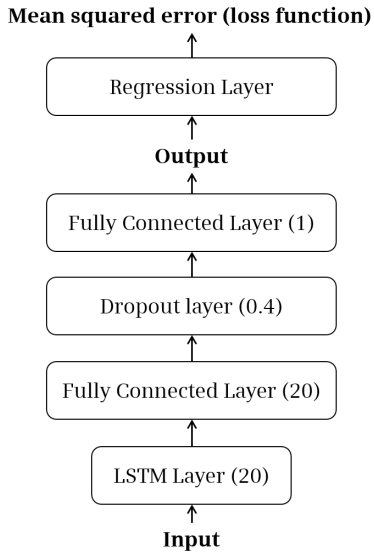


Figure 4.6: RNN architecture

The number in the bracket in 'LSTM layer' box indicates the number of hidden units in LSTM layer. In other words, 20 LSTM units shown in Figure 3.5 are stacked up inside the LSTM layer. After the LSTM layer, a fully connected layer and dropout

layer are connected. The fully connected layer is the same as one hidden layer of the feed-forward neural network. The number in the bracket of 'Fully connected layer' box indicates the number of neurons in the layer. Dropout layer is added to prevent overfitting problem and the number in the bracket of 'Dropout layer' box indicates dropout rate. At the end of the architecture, the fully connected layer and regression layer are connected. The fully connected layer is added to estimate output data (In this study, moment DEL). The regression layer is added to calculate mean squared error for training.

For LSTM application, only the operating condition 1 has been considered. Especially, data points where the operating condition is constantly the same have been extracted since sequential dependency is expected to disappear when operating condition is changed. As a result, the number of data points which consecutively are corresponding to operating condition 1 has been obtained as shown in Table 4.13.

Table 4.13: Number of data points used for RNN

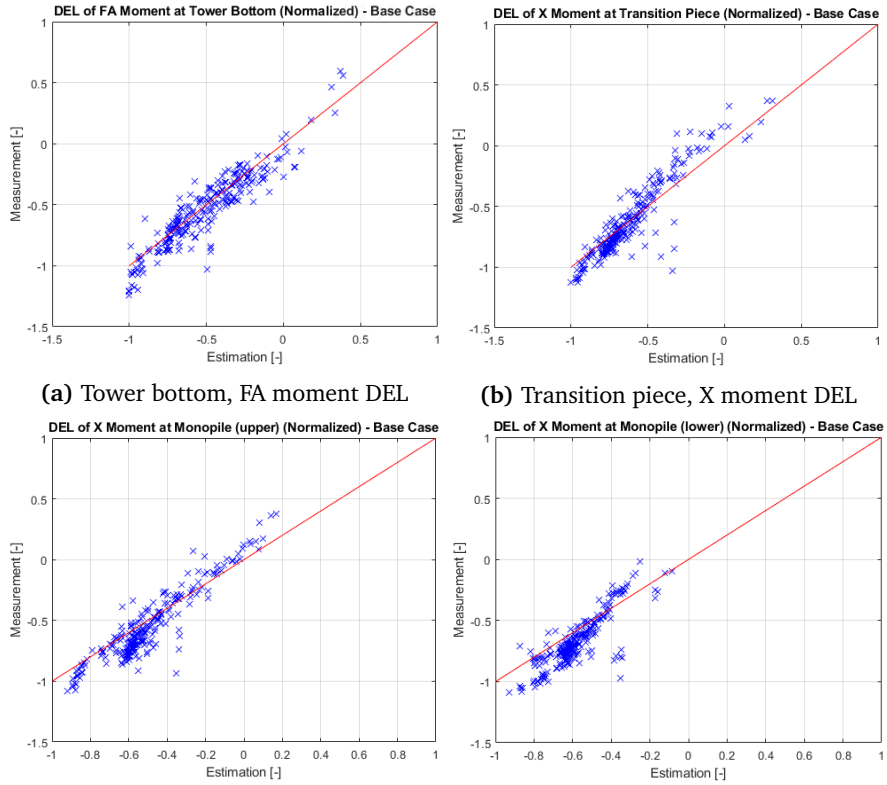
	Turbine 1	Turbine 2
No. of Data	1665	439

4.5.1 Estimation result for base case

Moment DEL estimation for all locations has been made with the recurrent neural network. In Figure 4.7, FA and X directional estimation of turbine 1 is shown only for test set. In Table 4.14, accuracy measurements for all the results are tabulated.

Table 4.14: Results of individual turbine level load estimation, Operating condition 1, All locations, Base Case, RNN

Case	Location	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
Base Case	Tower Bottom (FA)	0.933	15.80%	0.916	15.73%
	Tower Bottom (SS)	0.890	13.01%	0.785	38.23%
	Transition Piece (X)	0.938	12.85%	0.909	30.45%
	Transition Piece (Y)	0.942	11.33%	0.981	6.70%
	Monopile (upper) (X)	0.929	13.00%	0.758	37.44%
	Monopile (upper) (Y)	0.942	11.76%	0.988	15.92%
	Monopile (lower) (X)	0.849	19.24%	0.401	47.66%
	Monopile (lower) (Y)	0.893	21.26%	0.973	13.80%



(a) Tower bottom, FA moment DEL

(b) Transition piece, X moment DEL

(c) Monopile (upper), X moment DEL

(d) Monopile (lower), X moment DEL

Figure 4.7: Results of individual turbine level load estimation, FA and X directions, Operating Condition 0, Turbine 1, Base Case, RNN

As shown in the results, the recurrent neural network gives poor estimation. Specifically, MAPE has been calculated above 10% for all cases except Y directional moment DEL estimation at transition piece. This might be due to the fact that moment DELs have no sequential dependency, or the number of training data is not enough for training but it is difficult to make a conclusion of what causes the inaccuracy. However, it can be concluded that the feed-forward neural network fits better than the recurrent neural network for this study with one-month data.

4.6 Case study with different input combination

To answer the research question of ***'what is the minimum number of sensors that gives accurate DEL estimation?***, case studies have been performed. Since it has been found from Section 4.4 and 4.5 that the feed-forward neural network performs better than recurrent neural network for this study, the feed-forward neural network has been applied for case studies. In addition, for all case studies, only operating condition 1 has been considered since the number of data points is not enough in operating condition 3 and parked condition as shown in Section 4.4.

In Section 4.6.1, accuracy measurement comparison for different input combination is shown. In Section 4.6.2, estimation result with only the standard signals as input is presented. Then, effect of wave data is explained in Section 4.6.3. In Section 4.6.4, estimation with non-standard signals is described. Summary of estimation results for different input combination is shown in Section 4.6.5. Lastly, weights comparison is described in Section 4.6.6.

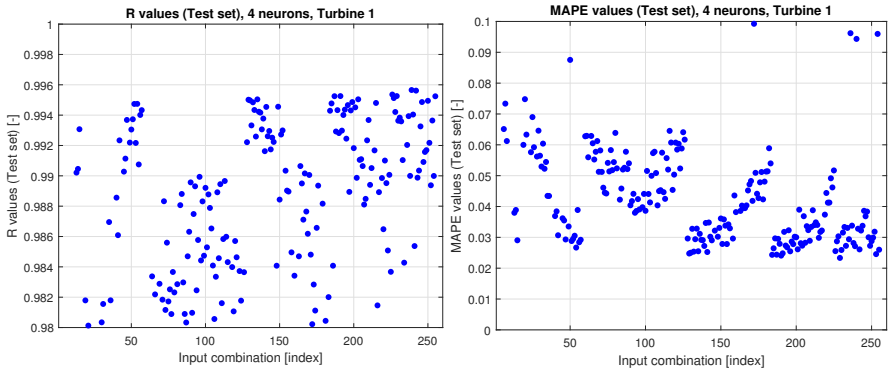
Two remarks here are that, firstly all the comparisons between measurement and estimation are shown for all the different sets including training/cross validation and test sets. Secondly, the name of each case study is abbreviated to 'CS' with case number at the end.

4.6.1 Accuracy measurements for different input combinations

To figure out the minimum number of sensors that gives a reasonably accurate estimation, the accuracy measurements have been compared for different input combinations. The same feed-forward neural network used in Section 4.4 (with 1 hidden layer and 4 neurons) has been applied. For this study, different combinations of input features shown in Table 4.1 have been made. As a result, a total of 255 different input combinations have been generated. For all the features, all the statistical properties have been applied as input. In addition, FA moment DEL at the tower bottom of turbine 1 has been used as the target value for this study. The resultant accuracy measurements are shown in Figure 4.8.

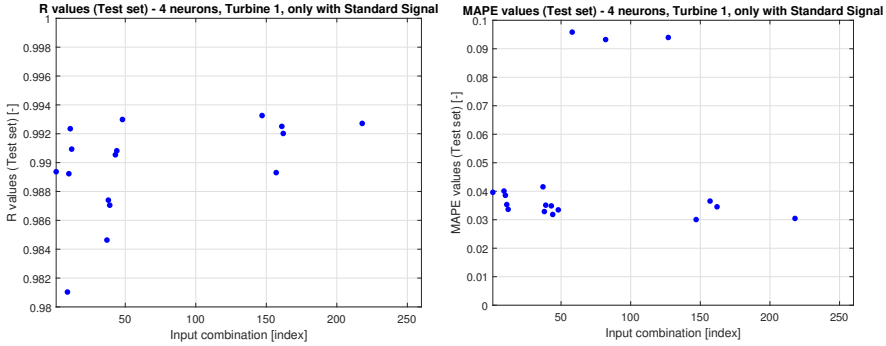
About index number of input combination in Figure 4.8, smaller index number indicates the input combination is simpler and containing a small number of input features while it contains a larger number of input features when index number is larger. As a result, it has been found that the input combination with the smaller number of input features gives almost the same level of accuracy with the input combination with the larger number of input features.

The same accuracy measurement comparison has been made for input combinations which include only the standard input signals as shown in Figure 4.9.



(a) R comparison **(b)** MAPE comparison

Figure 4.8: Accuracy comparison for different input combinations, all input combinations



(a) R comparison **(b)** MAPE comparison

Figure 4.9: Accuracy comparison for different input combinations, input combinations only with standard signals

As shown in Figure 4.9, it can be concluded that the estimation can be accurate only with the standard signals with R value of almost 0.994 and MAPE of around 4%.

4.6.2 Estimation with standard signals

As shown in Section 4.6.1, FA directional moment DEL estimation at tower bottom of turbine 1 can be accurately estimated with the feed-forward neural network only with standard signals as its input.

Therefore, the feed-forward neural network only with standard signals has been applied to estimate moment DEL at other locations (transition piece and monopile) to investigate if only the standard signals are enough for accurate estimation.

The input signals and statistical properties used for this study are tabulated in Table 4.15 and the results for FA and X directional moment DEL estimation of turbine 1 are shown in Figure 4.10. The accuracy measurements are presented in Table 4.16.

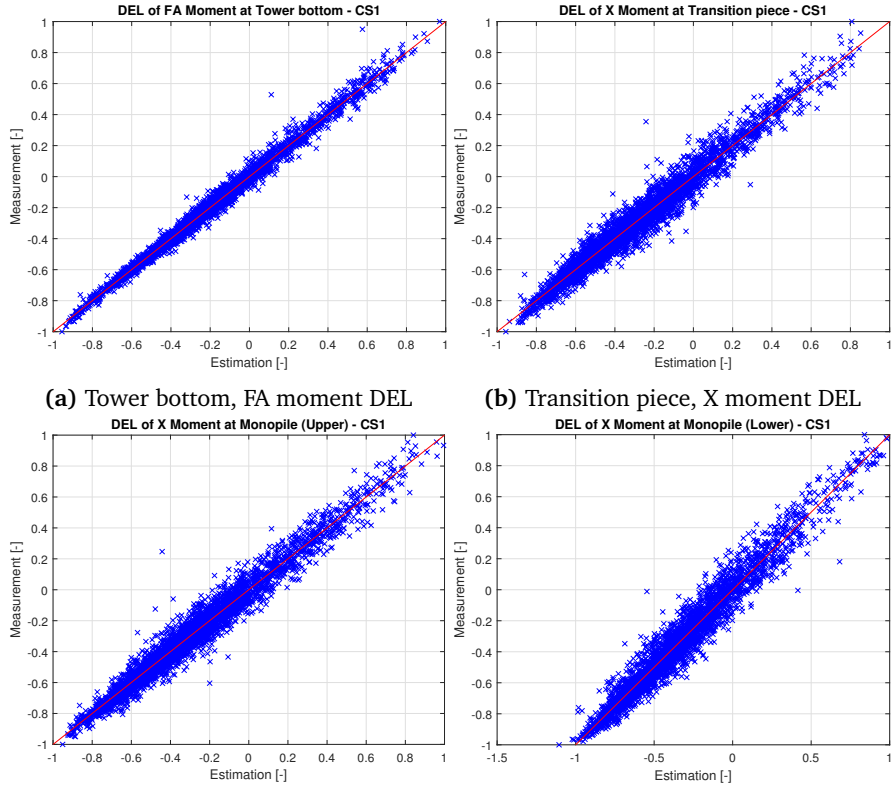
Table 4.15: Input data, only with standard signals with all statistical properties, CS1

Input (CS1)			
Category	Number	Signal	Statistical properties
Standard	1	Acceleration	All (1-5 in Table A.1)
	2	Blade Pitch angle	All (1-5 in Table A.1)
	3	Wind Speed	All (1-5 in Table A.1)
	4	Active Power	All (1-5 in Table A.1)
	5	Rotational Speed	All (1-5 in Table A.1)
	6	Yaw Direction	All (1-5 in Table A.1)

Table 4.16: Results of individual turbine level load estimation, Operating condition 1, All locations, Only with standard signals, All statistical properties, Feed forward neural network, CS1

Case	Location	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
CS1	Tower Bottom (FA)	0.965	3.17%	0.976	3.62%
	Tower Bottom (SS)	0.996	1.77%	0.994	1.92%
	Transition Piece (X)	0.972	6.14%	0.957	5.95%
	Transition Piece (Y)	0.963	5.61%	0.943	5.56%
	Monopile (upper) (X)	0.977	5.79%	0.934	5.67%
	Monopile (upper) (Y)	0.967	5.23%	0.962	5.45%
	Monopile (lower) (X)	**0.963**	**7.91%**	**0.957**	**6.68%**
	Monopile (lower) (Y)	**0.953**	**6.66%**	**0.959**	**6.26%**

From the results, it is concluded that the moment DEL at the tower bottom can



(a) Tower bottom, FA moment DEL

(b) Transition piece, X moment DEL

(c) Monopile (upper), X moment DEL

(d) Monopile (lower), X moment DEL

Figure 4.10: Results of individual turbine level load estimation, FA and X directions, Operating Condition 0, Turbine 1, Only with standard signals, All statistical properties, Feed forward neural network, CS1

be accurately estimated only with standard signals with MAPE of less than 4%. However, for transition piece and monopile, MAPE of higher than 5% has been obtained. Especially, at the lower monopile level, the result is the most inaccurate with MAPE of higher than 6%. Added to that, from Figure 4.10, it has been found that the estimations are not biased for all locations.

One remark is that there is a larger error in FA directional estimation than SS directional at tower bottom. This is due to the larger range of input features and

target moment DEL in FA direction compared to that of SS direction. Specifically, the thrust force is larger than other forces in offshore wind turbines and it is acting in FA direction. Accordingly, ranges of input features and target moment DEL in FA direction is larger than SS direction. In other words, the feed-forward neural network should be trained with more data in FA direction to cover all these larger ranges of input features and target moment DEL compared to SS direction. However, the one-month data seems not enough to train the feed-forward neural network thoroughly.

In addition to that, all the standard signals only with standard statistical properties have been applied. Specifically, among all the statistical properties, some of the statistical properties can only be calculated when time series is available. Such statistical properties are damage equivalent value and spectral moments.

However, default SCADA system does not provide those statistical properties. In other words, to calculate those statistical properties, additional effort should be put. Therefore, it has been investigated if estimation without those statistical properties can give accurate estimation.

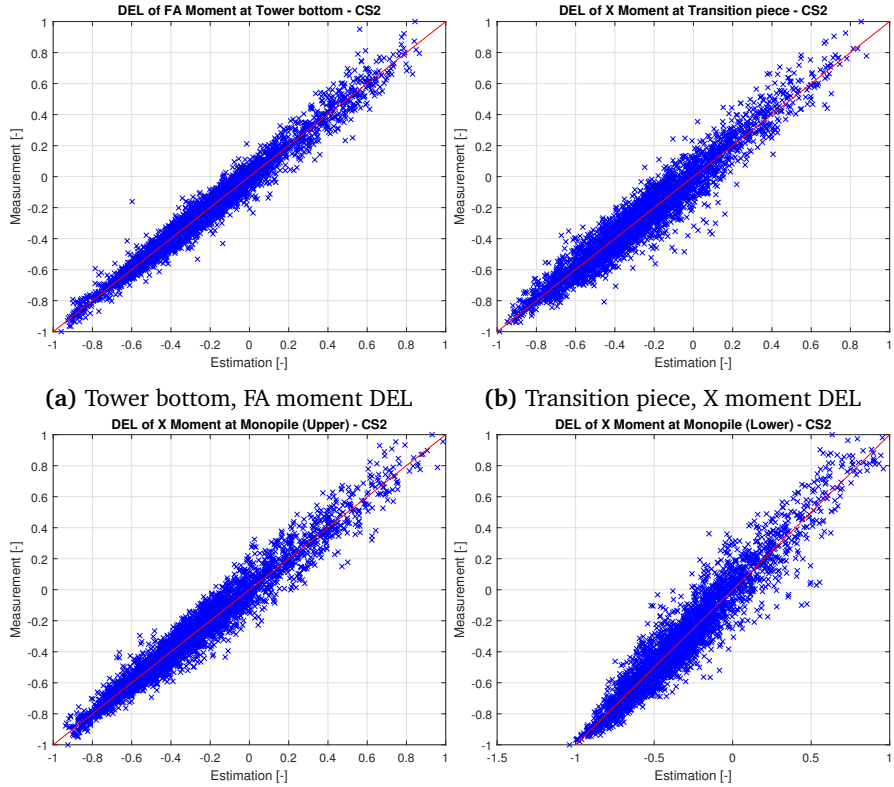
The input data used for this study is shown in Table 4.17. The results for FA and X directional moment DEL estimation of turbine 1 are shown in Figure 4.11. The accuracy measurements are presented in Table 4.18.

Table 4.17: Input data, only with standard signals, with standard statistical properties, CS2

Input (CS2)			
Category	Number	Signal	Statistical properties
Standard	1	Acceleration	1-3 in Table A.1
	2	Blade Pitch angle	1-3 in Table A.1
	3	Wind Speed	1-3 in Table A.1
	4	Active Power	1-3 in Table A.1
	5	Rotational Speed	1-3 in Table A.1
	6	Yaw Direction	1-3 in Table A.1

As shown in the results, overall accuracy is decreased compared to the results when all statistical properties have been used. Specifically, for all locations, MAPE of higher than 5% has been obtained except SS directional moment DEL estimation at tower bottom.

In summary, it has been found that standard signals can give accurate estimation for tower bottom location with MAPE of less than 5%. However, estimation has been



(a) Tower bottom, FA moment DEL

(b) Transition piece, X moment DEL

(c) Monopile (upper), X moment DEL

(d) Monopile (lower), X moment DEL

Figure 4.11: Results of individual turbine level load estimation, FA and X directions, Operating Condition 0, Turbine 1, Only with standard signals, with standard statistical properties, Feed forward neural network, CS2

relatively inaccurate at transition piece and monopile with MAPE of less than 10%. With only the standard statistical properties, the results have been even worse.

Furthermore, given that the fatigue critical location is located near mudline, higher inaccuracy is expected in estimation at the fatigue critical location.

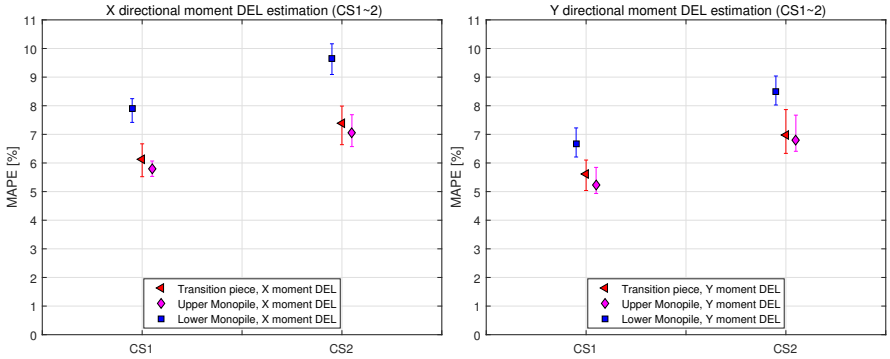
The possible explanation of the inaccuracy in estimation at transition piece and monopile is that standard signals do not include any information about waves. The wave induced moment is getting larger along with the depth. Accordingly, it is ex-

Table 4.18: Results of individual turbine level load estimation, Operating condition 1, All locations, Only with standard signals, with standard statistical properties, Feed forward neural network, CS2

Case	Location	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
CS2	Tower Bottom (FA)	0.981	4.68%	0.976	4.88%
	Tower Bottom (SS)	0.993	2.70%	0.993	3.01%
	Transition Piece (X)	0.959	7.38%	0.954	7.62%
	Transition Piece (Y)	0.952	6.99%	0.946	6.39%
	Monopile (upper) (X)	0.967	7.06%	0.963	6.79%
	Monopile (upper) (Y)	0.959	6.79%	0.957	6.31%
	Monopile (lower) (X)	**0.942**	**9.64%**	**0.956**	**7.81%**
	Monopile (lower) (Y)	**0.956**	**8.51%**	**0.957**	**6.92%**

pected that wave effect is higher at lower level.

Lastly, for turbine 1, MAPE of CS1 and CS2 are shown together with its variance in Figure 4.12 for clear comparison in between CS1 and CS2.



(a) X directional **(b)** Y directional

Figure 4.12: MAPE Comparison in between CS1 and CS2, Turbine 1

4.6.3 Estimation with standard signals and wave data

In Section 4.6.2, it has been found that the feed-forward neural network only with standard signals is expected to give inaccurate estimation at the fatigue critical location. In this section, wave effect has been included to investigate standard signals and wave data are enough for moment DEL estimation.

As shown in Section A.1, wave data has also been pre-processed into 10 min. basis. For this study, H_s, T_p and spectral moments (from -2 to 2) of wave have been applied.

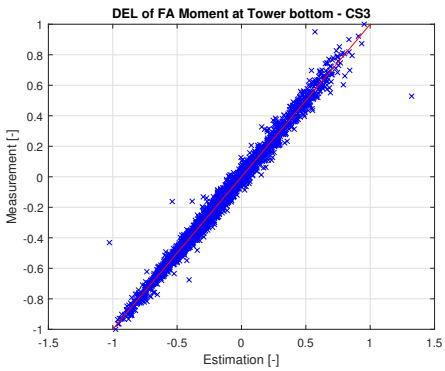
First of all, standard signals with all the statistical properties have been applied with the wave statistics. The input signals and statistical properties are tabulated in Table 4.19 and the results for FA and X directional moment DEL estimation of turbine 1 are shown in Figure 4.13. The accuracy measurements are presented in Table 4.20.

Table 4.19: Input data, only with standard signals, with all statistical properties, with wave data, CS3

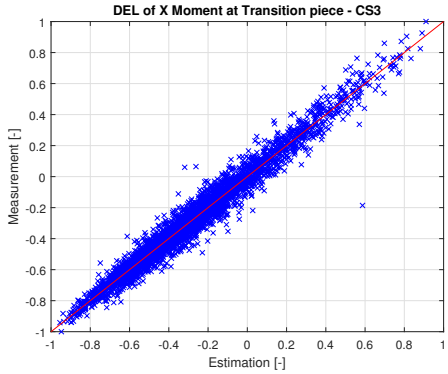
Input (CS3)			
Category	Number	Signal	Statistical properties
Standard	1	Acceleration	All (1-5 in Table A.1)
	2	Blade Pitch angle	All (1-5 in Table A.1)
	3	Wind Speed	All (1-5 in Table A.1)
	4	Active Power	All (1-5 in Table A.1)
	5	Rotational Speed	All (1-5 in Table A.1)
	6	Yaw Direction	All (1-5 in Table A.1)
Wave	-	Wave data	H_s
			T_p
			$\lambda_{-2\sim2}$

From the results, it is clearly seen that the wave data improve accuracy. Especially, there is a relatively larger improvement in accuracy at the lower monopile level where the wave-induced moment is higher compared to other locations. As a result, even at the lower monopile level, MAPE of about 6% can be achieved. However, the improvement is not so significant at the tower bottom and transition piece.

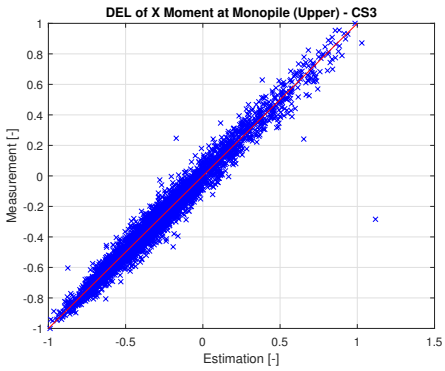
The same study has been done for standard signals only with standard statistical properties to investigate if additional process with time series can be disregarded (that is, additional effort to calculate equivalent value and spectral moments). The input signals and statistical properties applied for this case study are tabulated in



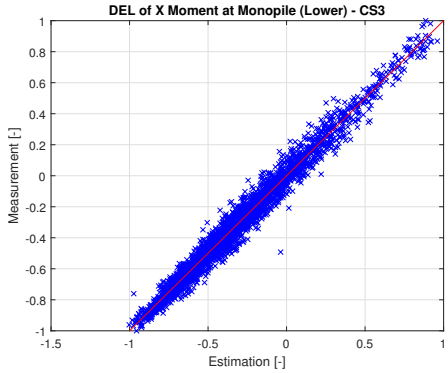
(a) Tower bottom, FA moment DEL



(b) Transition piece, X moment DEL



(c) Monopile (upper), X moment DEL



(d) Monopile (lower), X moment DEL

Figure 4.13: Results of individual turbine level load estimation, FA and X directions, Operating Condition 0, Turbine 1, Only with standard signals, All statistical properties, With wave data, Feed forward neural network, CS3

Table 4.21 and the results for FA and X directional moment DEL estimation of turbine 1 are shown in Figure 4.14. The accuracy measurements are presented in Table 4.22.

From the results, a similar conclusion can be made. Specifically, the inclusion of wave improves the accuracy of the estimation. Especially, at monopile locations, the accuracy is improved the most and MAPE of around 7% can be achieved even only with standard statistical properties. Compared to the case where all the statistical properties of standard signals have been used, the results are less accurate for all

Table 4.20: Results of individual turbine level load estimation, Operating condition 1, All locations, Only with standard signals, All statistical properties, With wave data, Feed forward neural network, CS3

Case	Location	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
CS3	Tower Bottom (FA)	0.991	3.17%	0.949	3.69%
	Tower Bottom (SS)	0.997	1.73%	0.988	2.04%
	Transition Piece (X)	0.974	6.15%	0.946	5.80%
	Transition Piece (Y)	0.959	5.43%	0.950	5.43%
	Monopile (upper) (X)	0.967	5.40%	0.966	5.30%
	Monopile (upper) (Y)	0.978	4.98%	0.969	5.12%
	Monopile (lower) (X)	**0.982**	**5.79%**	**0.970**	**5.73%**
	Monopile (lower) (Y)	**0.967**	**5.73%**	**0.964**	**5.68%**

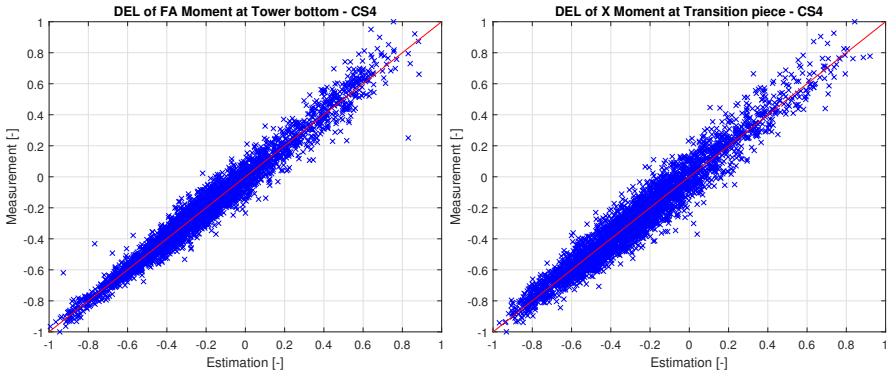
Table 4.21: Input data, only with standard signals, with standard statistical properties, with wave data, CS4

Input (CS4)			
Category	Number	Signal	Statistical properties
Standard	1	Acceleration	1-3 in Table A.1
	2	Blade Pitch angle	1-3 in Table A.1
	3	Wind Speed	1-3 in Table A.1
	4	Active Power	1-3 in Table A.1
	5	Rotational Speed	1-3 in Table A.1
	6	Yaw Direction	1-3 in Table A.1
Wave	-	Wave data	H_s
			T_p
			$\lambda_{-2\sim2}$

locations.

In summary, it has been found that the inclusion of wave improves accuracy. However, the estimation at the lower monopile level (closest to the fatigue critical location) is still not so accurate with MAPE of higher than 5%.

Lastly, for turbine 1, MAPE of CS3 and CS4 are shown together with its variance in Figure 4.15.



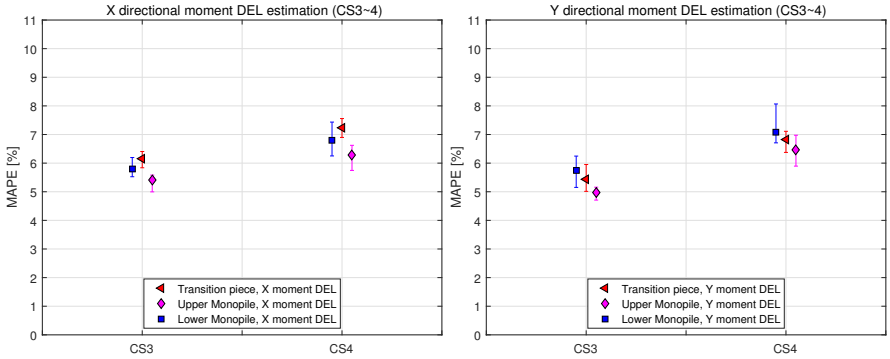
(a) Tower bottom, FA moment DEL **(b)** Transition piece, X moment DEL

(c) Monopile (upper), X moment DEL **(d)** Monopile (lower), X moment DEL

Figure 4.14: Results of individual turbine level load estimation, FA and X directions, Operating Condition 0, Turbine 1, Only with standard signals, Standard statistical properties, With wave data, Feed forward neural network, CS4

Table 4.22: Results of individual turbine level load estimation, Operating condition 1, All locations, Only with standard signals, Standard statistical properties, With wave data, Feed forward neural network, CS4

Case	Location	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
CS4	Tower Bottom (FA)	0.983	4.73%	0.956	5.00%
	Tower Bottom (SS)	0.993	2.62%	0.993	2.97%
	Transition Piece (X)	0.963	7.24%	0.952	7.72%
	Transition Piece (Y)	0.956	6.81%	0.944	6.21%
	Monopile (upper) (X)	0.975	6.29%	0.967	6.27%
	Monopile (upper) (Y)	0.960	6.46%	0.961	5.67%
	Monopile (lower) (X)	**0.977**	**6.79%**	**0.970**	**6.52%**
	Monopile (lower) (Y)	**0.964**	**7.09%**	**0.966**	**5.86%**



(a) X directional **(b)** Y directional

Figure 4.15: MAPE Comparison in between CS3 and CS4, Turbine 1

4.6.4 Estimation with non-standard signals

From previous sections, only with the standard signals and wave data, it has been found that estimation accuracy at transition piece and monopile levels are not so accurate with MAPE of higher than 5%. Therefore, not only the standard signals and wave data, some of the non-standard signals have been applied in this section.

As shown in Appendix E, moment DEL at monopile levels are closely correlated with moment and inclination at tower bottom. In addition, inclinometer and strain gauges can relatively easily be installed at tower bottom. Accordingly, among all the non-standard signals, inclination and moment at tower bottom have been applied to increase estimation accuracy at transition piece and monopile.

In this section, only the accuracy measurements of the lower monopile level are compared in between different cases since lower monopile level is where the closest to the fatigue critical location near mudline, and the lowest accuracy has been found from the previous case studies. In the end, accuracy measurements of all locations are shown for the best case together with plots showing X directional moment DEL comparison between measurement and estimation.

For this study, a total of eight (8) cases have been made as shown in Table 4.23. From CS5 to CS10, it includes both standard and non-standard signals while CS11 and CS12 only include non-standard signals. Since the mean of the non-standard signals is expected to have no influence on moment DEL estimation, it has not been used.

As a result, accuracy measurements have been obtained as shown in Table 4.24. For a clear comparison, MAPE of all case studies are shown in Figure 4.16 together with its variance.

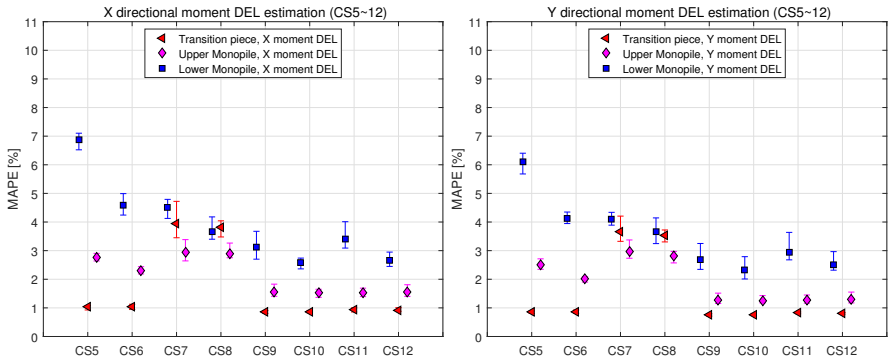
From the results, it is clearly seen that the accuracy increased even at the lower monopile level. Specifically, in some cases, MAPE is as low as around 3%.

In detail, when CS5 (Standard signal with standard statistical properties as well as moment signal at tower bottom) and CS7 (Standard signal with standard statistical properties as well as inclination signal at tower bottom) are compared, the inclusion of inclination signal give the larger increase in accuracy. Concretely, the inclusion of the inclination signal gives MAPE of around 4.5%.

From CS6 (Standard signal with standard statistical properties as well as moment signal at tower bottom with wave data) and CS8 (Standard signal with standard statistical properties as well as inclination signal at tower bottom with wave data), it is shown that the accuracy level is increased in both cases. In detail, when wave data is included in addition to moment or inclination signal at tower bottom, MAPE of less than 5% and 4% can be achieved respectively.

Table 4.23: Input data, CS5∼CS12

Input (CS5∼CS12)				
Case	Category	Number	Signal	Statistical properties
CS5∼10	Standard	1	Acceleration	1-3 in Table A.1
		2	Blade Pitch angle	1-3 in Table A.1
		3	Wind Speed	1-3 in Table A.1
		4	Active Power	1-3 in Table A.1
		5	Rotational Speed	1-3 in Table A.1
		6	Yaw Direction	1-3 in Table A.1
CS5	Non-standard	6	Moment at Tower Bottom	2-5 in Table A.1
CS6	Non-standard	6	Moment at Tower Bottom	2-5 in Table A.1
	Wave	-	Wave data	H_s, T_p and $\lambda_{-3\sim3}$
CS7	Non-standard	5	Inclination at Tower Bottom	2-5 in Table A.1
CS8	Non-standard	5	Inclination at Tower Bottom	2-5 in Table A.1
	Wave	-	Wave data	H_s, T_p and $\lambda_{-3\sim3}$
CS9, CS11	Non-standard	5	Inclination at Tower Bottom	2-5 in Table A.1
		6	Moment at Tower Bottom	2-5 in Table A.1
CS10, CS12	Non-standard	5	Inclination at Tower Bottom	2-5 in Table A.1
		6	Moment at Tower Bottom	2-5 in Table A.1
	Wave	-	Wave data	H_s, T_p and $\lambda_{-3\sim3}$



(a) X directional **(b)** Y directional

Figure 4.16: MAPE Comparison in between CS5∼CS12, Turbine 1

Table 4.24: Results of individual turbine level load estimation, Operating condition 1, At lower monopile, Feed forward neural network, CS5∼CS12

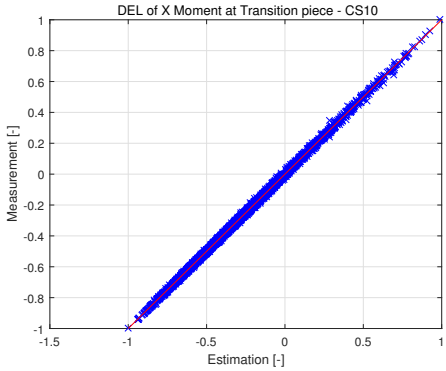
Case	Direction	Turbine 1		Turbine 2	
		R	MAPE	R	MAPE
CS5	X	0.971	6.86%	0.975	5.37%
	Y	0.975	6.11%	0.979	4.70%
CS6	X	0.987	4.59%	0.984	3.86%
	Y	0.990	4.12%	0.986	3.82%
CS7	X	0.980	4.50%	0.986	3.93%
	Y	0.988	4.10%	0.986	3.70%
CS8	X	0.993	3.66%	0.990	3.43%
	Y	0.992	3.66%	0.986	3.55%
CS9	X	0.993	3.13%	0.990	3.09%
	Y	0.995	2.68%	0.988	2.72%
CS10	**X**	**0.996**	**2.57%**	**0.990**	**2.66%**
	Y	**0.996**	**2.32%**	**0.992**	**2.41%**
CS11	X	0.992	3.41%	0.993	2.72%
	Y	0.995	2.95%	0.992	2.74%
CS12	X	0.996	2.65%	0.985	2.67%
	Y	0.995	2.50%	0.995	2.47%

CS9 (Standard signal with standard statistical properties as well as moment and inclination signals at tower bottom) and CS10 (Standard signal with standard statistical properties as well as moment and inclination signals at tower bottom with wave data) include the standard signals as well as both moment and inclination signals at tower bottom. Without wave data (CS9), it gives an accurate estimation with MAPE of around 3% which outperforms all the previous cases (CS5∼CS8). When the wave data is added (CS10), the accuracy level increases further with MAPE of less than 3%.

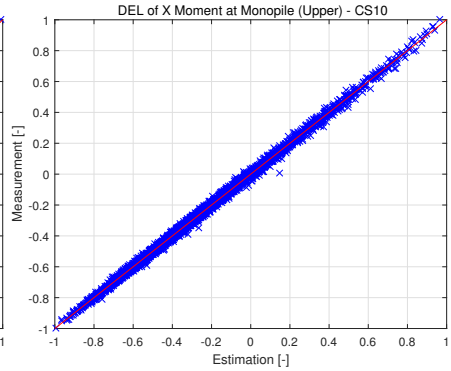
Lastly, without standard signals, only the moment and inclination signals have been applied with or without wave data for CS11 (Moment and inclination signals at tower bottom) and CS12 (Moment and inclination signals at tower bottom with wave data). In both cases, the accuracy level is almost the same as that of CS9 and CS10 respectively.

From all the observations, the best accuracy can be obtained at CS10 and CS12 with MAPE of less than 3% in both cases. Accordingly, the estimation of all the

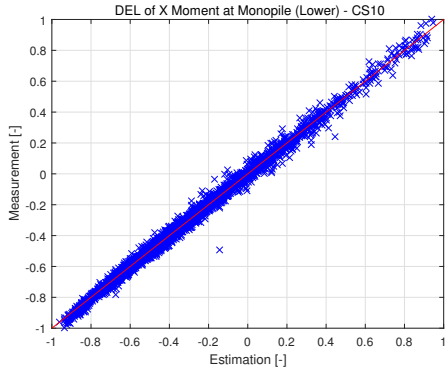
locations has been made for CS10. The results are shown in Table 4.25 and Figure 4.17.



(a) Transition piece, X moment DEL



(b) Monopile (upper), X moment DEL



(c) Monopile (lower), X moment DEL

Figure 4.17: Results of individual turbine level load estimation, Operating condition 1, All locations, With standard and non-standard (moment and inclination at tower bottom) signals, Standard statistical properties, With wave data, Feed forward neural network, CS10

As expected, estimation on transition piece and upper monopile level locations are more accurate than lower monopile location. With CS10, MAPE of less than 3% can be obtained for all locations.

In summary, the number of sensors required for accurate estimation of moment DEL is dependent on the level of accuracy. If MAPE of 5% needs to be achieved at

Table 4.25: Results of individual turbine level load estimation, Operating condition 1, All locations, With standard and non-standard (moment and inclination at tower bottom) signals, Standard statistical properties, With wave data, Feed forward neural network, CS10

Case	Location	Turbine 1		Turbine 2	
		R value	MAPE	R value	MAPE
CS10	Transition Piece (X)	0.999	0.87%	0.999	0.85%
	Transition Piece (Y)	0.999	0.76%	0.999	0.81%
	Monopile (upper) (X)	0.998	1.53%	0.997	1.39%
	Monopile (upper) (Y)	0.998	1.26%	0.998	1.30%
	Monopile (lower) (X)	**0.992**	**2.57%**	**0.990**	**2.66%**
	Monopile (lower) (Y)	**0.996**	**2.32%**	**0.992**	**2.41%**

lower monopile location, strain gauges or inclinometers should be installed at tower bottom in addition to the standard signals from SCADA and wave data (corresponding to CS6 and CS8). Especially, when wave measurement data is available, the inclusion of wave measurement data gives an increase in accuracy. However, it should be noted that the accuracy level calculated in this section is for lower monopile level which is located upper than the fatigue critical location near mudline. Therefore, higher MAPE is expected at the fatigue critical location located near mudline.

4.6.5 Summary of the results

The estimation results shown in previous sections are summarized to clearly distinguish the accuracy level according to different input combinations. In this section, a summary of the estimation results at the lower monopile level is shown.

The summarized results are shown in Table 4.26.

4.6.6 Combined weights for CS10, Oper. cond. 1

To investigate which features have a high effect on estimation, the combined weights have been compared. The combined weights can be calculated as explained in Appendix G.

In this study, for both turbines in operating condition 1, the feed-forward neural network trained with input features of CS10 estimating X directional moment DEL at lower monopile level has been considered since CS10 gives the highest accuracy. It should be noted that the trained weights can be different according to different

Table 4.26: Summary of results

Standard signals?	Additional features	Case number	MAPE
O	None	CS1	<8%
O	Wave	CS3	<6%
O	Moment at tower bottom	CS5	<7%
O	Inclination at tower bottom	CS7	<4.5%
O	Moment and Inclination at tower bottom	CS9	<3.5%
O	Moment at tower bottom + Wave	CS6	<5%
O	Inclination at tower bottom + Wave	CS8	<4%
O	Moment and Inclination at tower bottom + Wave	CS10	<3%
X	Moment and inclination at tower bottom	CS11	<3.5%
X	Moment and Inclination at tower bottom + Wave	CS12	<3%

divisions of training, cross-validation and test set.

As a result, the features shown in Table 4.27 have been found as the top ten the most influencing features.

Table 4.27: Combined weight comparison for both turbines

Rank	Turbine 1	Turbine 2
1	Tower Incli. Bottom Back λ_0	Tower Incli. Bottom Back λ_{-1}
2	Tower Incli. Bottom Back λ_{-1}	Tower Incli. Bottom Back λ_0
3	Tower Incli. Bottom Back λ_{-2}	Tower Incli. Bottom Back λ_{-2}
4	Tower Incli. Bottom Back λ_{-3}	Tower Incli. Bottom Back λ_{-3}
5	Tower Incli. Bottom Back λ_2	Tower Mom Bottom λ_2
6	Tower Mom Bottom λ_{-1}	Tower Incli. Bottom Back λ_3
7	Tower Mom Bottom λ_2	Tower Mom Bottom λ_0
8	Tower Incli. Bottom λ_3	Wave spectrum λ_1
9	Tower Incli. Bottom Back λ_3	Tower Mom Bottom DEL
10	Tower Mom Bottom DEL	Tower Incli. Bottom λ_1

From the result, it is clearly seen that the important features are somewhat different from the results obtained from Pearson's correlation coefficient comparison in

Section 2.3. Given that the feed-forward neural network finds a nonlinear relationship in between input features and targets, while Pearson's correlation coefficient finds a linear relationship in between one feature and target, the important features can be obtained differently.

In addition, spectral moments of inclination, moment and wave spectrum have been found to be the important features. The result can be interpreted that varying dynamic properties have a high effect on moment DEL estimation for operating condition 1. Specifically, the dynamic properties of wind turbines are highly affected by aerodynamic damping (especially, FA directional dynamic property). The aerodynamic damping is changed according to wind speed, pitch angle and rotational speed of wind turbine. Accordingly, the response frequency is changed and moment DEL would be affected since the number of cycles would be affected by the changed response frequency (i.e. the number of cycles would be higher when response frequency gets higher and vice versa).

Secondly, it is shown that 0 th and negative spectral moments of inclination signals have the highest combined weights for both turbines. Given that negative spectral moments are related to low-frequency components in PSD, it can be interpreted that low-frequency components such as wind, blade passing and lowest natural frequency have a higher effect on estimation compared to higher frequency components.

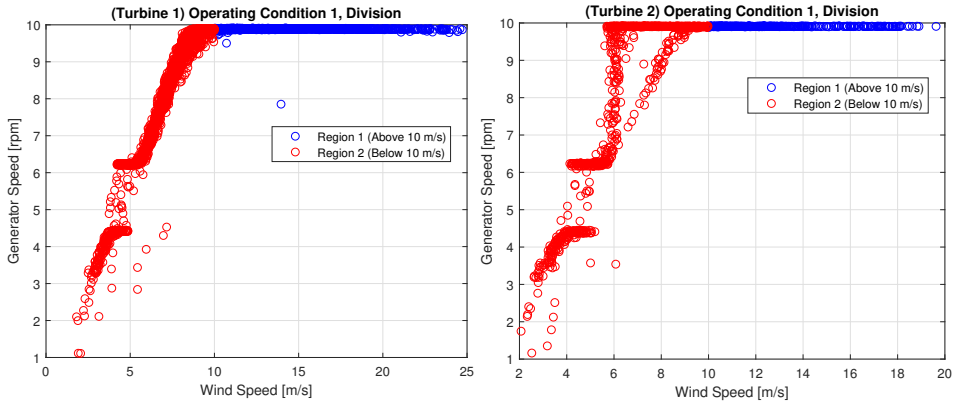
4.7 Case study for improvement

In addition to the different input combinations, case studies have been performed to investigate if any improvement can be achieved. In Section 4.7.1, division of operating condition 1 is explained. In Section 4.7.2, principal component analysis is described.

4.7.1 Division of operating condition

As shown in Figure 2.6, operating condition 1 can be divided into two parts: above and below the wind speed where the generator speed keeps constant, which is 10 m/s. The reason for that is since the aerodynamic damping is expected to be different for these two regions [26]. The divided regions of operating condition 1 are shown in Figure 4.18. The number of data points for each region is shown in Table 4.28.

As shown in Table 4.28, Turbine 2 has only 435 data points with wind speed higher than 10 m/s which may not be enough for training.



(a) Turbine 1 **(b)** Turbine 2

Figure 4.18: Mean wind speed versus rotational speed, Division of operating condition 1

Table 4.28: Summary of data

Turbine 1	Operating 1		Total
	Above 10 m/s	Below 10 m/s	
No. of data	1597	1926	3523
Portion	45.3%	54.7%	100%
Turbine 2	**Operating 1**		**Total**
	Above 10 m/s	Below 10 m/s	
No. of data	435	1094	1529
Portion	28.4%	71.6%	100%

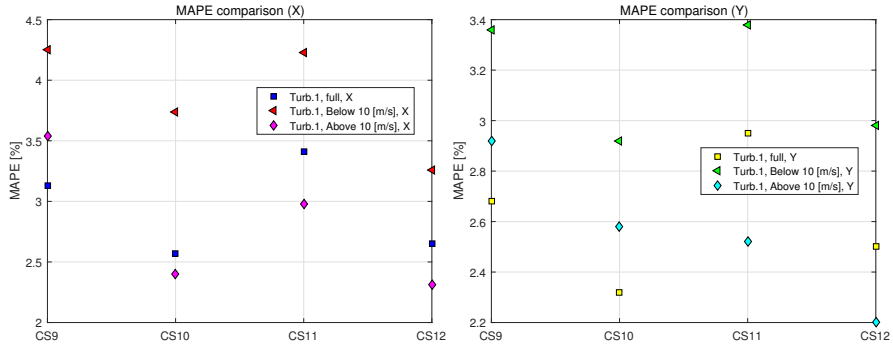
For this study, the feed-forward neural network has been trained separately for Region 1 and Region 2 in Figure 4.18a. In addition, only the accuracy level at the lower monopile level has been compared for CS9∼CS12. The results are shown in Table 4.29.

From the results, it is shown that the estimation is more accurate when the wind speed is above 10 m/s than the case where wind speed is below 10 m/s. For turbine 1, the result is compared with the results shown in Table 4.24 where the operating condition has not been divided as shown in Figure 4.19.

Specifically, when the wind speed is above 10 m/s, the estimation accuracy is

Table 4.29: Results of individual turbine level load estimation, Operating condition 1, Divided based on wind speed, At lower monopile, Feed forward neural network, CS9∼CS12

Case	Direction	Turbine 1				Turbine 2			
		Above 10 m/s		Below 10 m/s		Above 10 m/s		Below 10 m/s	
		R	MAPE	R	MAPE	R	MAPE	R	MAPE
CS9	X	0.963	3.54%	0.985	4.25%	0.974	3.61%	0.990	3.46%
	Y	0.982	2.92%	0.986	3.36%	0.968	2.77%	0.991	3.11%
CS10	X	0.989	2.40%	0.990	3.74%	0.977	2.44%	0.992	2.90%
	Y	0.988	2.58%	0.984	2.92%	0.983	2.25%	0.987	2.87%
CS11	X	0.991	2.98%	0.982	4.23%	0.984	2.66%	0.992	3.46%
	Y	0.992	2.52%	0.986	3.38%	0.985	2.36%	0.988	3.02%
CS12	X	0.994	2.31%	0.992	3.26%	0.991	2.20%	0.988	2.97%
	Y	0.994	2.20%	0.990	2.98%	0.978	2.12%	0.994	2.72%



(a) X direction **(b)** Y direction

Figure 4.19: Accuracy comparison in between estimation with full operating condition 1 and divided operating condition 1

sometimes higher than that of the case where full operating condition 1 is considered. On the other hand, when the wind speed is below 10 m/s, the result is always less accurate.

In addition to that, the combined MAPE can be calculated as shown in Equation

(4.2). The combined MAPE has been tabulated in Table 4.30.

$$\text{MAPE}_{combined} = \frac{\text{MAPE}_1 \cdot n_1 + \text{MAPE}_2 \cdot n_2}{n_1 + n_2} \tag{4.2}$$

Table 4.30: Combined MAPE

Case	Direction	Turbine 1	Turbine 2
CS9	X	3.93%	3.50%
	Y	3.16%	3.01%
CS10	X	3.13%	2.77%
	Y	2.77%	2.69%
CS11	X	3.66%	3.23%
	Y	2.99%	2.83%
CS12	X	2.83%	2.75%
	Y	2.63%	2.55%

The comparison of accuracy level in between estimations with full operating condition 1 and divided operating condition 1 is shown in Figure 4.20.

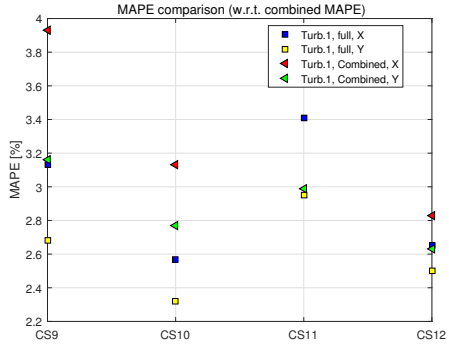


Figure 4.20: Accuracy comparison in between estimation with full operating condition 1 and divided operating condition 1, Combined MAPE

From the Figure, it is found that the accuracy level is always lower when the operating condition 1 is divided into two parts. In detail, for CS9 and CS10, the accuracy level decreases more than CS11 and CS12. The reason for that is since the CS9 and CS10 have included SCADA data as its input (i.e. wind speed, rotational speed, pitch angle and so on). Specifically, since the feed-forward neural network

trained for the different wind speed as well, the neural network can capture the different relationship between when the wind speed is below and above 10 m/s. However, for CS11 and CS12 where wind speed has not been included as its input, different behavior below and above the specific wind speed cannot be captured. That is why there is a smaller decrease in accuracy in CS11 and CS12.

In summary, with one-month data, division of operating condition 1 gives the lower accuracy level compared to the case when full operating condition 1 has been used.

Theoretically, when operating condition is divided into two parts, it is expected that the accuracy level would be almost the same in CS9 and CS10 since wind speed is included as input, and would be higher in CS11 and CS12 since wind speed is not used as input. However, the results show that the separation of operating condition 1 gives the lower accuracy level when operating condition 1 is divided. The lower accuracy level might be caused by the small number of data for training since two separate training has been done for around 1000∼2000 data points while 3520 data points have been used when operating condition 1 is not divided.

4.7.2 Estimation with principal components

All the studies explained previously based on the original dataset, and it is possible that some of the features are redundant. When there are these irrelevant features, the result can be inaccurate [2, 25]. Accordingly, it is necessary to select important features so that the irrelevant features can be removed from training. However, it is difficult to know beforehand which features are irrelevant features. Therefore, instead of removing the irrelevant features, Principal Component Analysis (PCA) can be applied. The main idea of PCA is to find low dimensional axes that can represent the original data accurately. Specifically, it uses an orthogonal transformation to convert the original data into the low dimensional axes [16], and the converted data is called principal component. As a result of PCA, if there are irrelevant features, its contribution factor in the conversion process would be negligible and the converted principal component would have small influence from the irrelevant features.

However, it is not true that PCA gives better results compared to the cases where all the original features are applied. From the various studies, it has been found that PCA can give either better or worse results [14, 17, 31] and it is difficult to know beforehand if PCA would give a better or worse result. In addition, it has also been discussed that principal components with small variance can also be important in the regression problem [21, pp. 244–246], or only the principal components with high variance are important [23, pp. 397–398]. Therefore, in this section, it has been investigated if PCA gives improvement for moment DEL estimation of offshore

wind turbines, and the effect of principal components with small variance has also been studied.

For this study, CS10 has been investigated since it includes the largest number of input features which indicates there is the highest possibility of including irrelevant features. In addition, X directional moment DEL at the lower monopile location of turbine 1 has been studied.

The total of three different cases shown in Table 4.31 have been compared. The percentage variance limit shown in Table 4.31 represents how well the original data points can be expressed with the low dimensional axes as a result of PCA. PCA Case 1 has a percentage variance limit of 99% and it only extracts the first few principal axes so that the accumulated percentage variance is above 99%. In short, PCA Case 1 will have a smaller number of features compared to CS10. Similarly, PCA Case 2 will have the same number of converted features with CS10 since the percentage variance limit has been set to be 100%.

Table 4.31: Different cases applied in PCA study

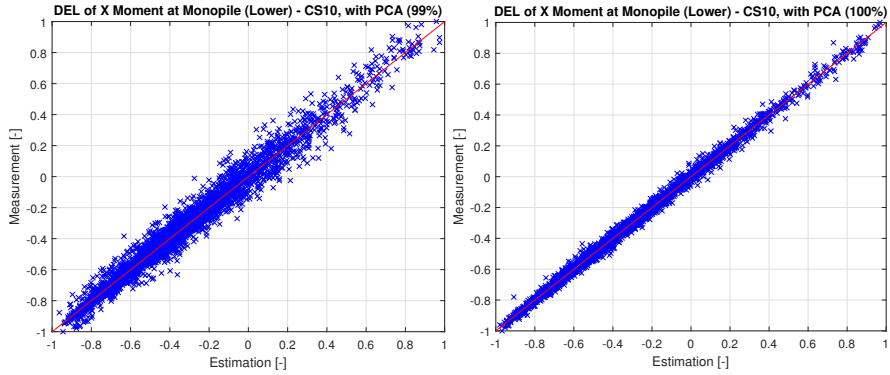
Case	Percentage variance limit	No. of features
CS10 (with original features)	-	88
PCA Case 1	99%	25
PCA Case 2	100%	88

The accuracy comparison is shown in Table 4.32. The estimation and measurement comparison for PCA Case 1 and 2 are shown in Figure 4.21.

Table 4.32: Accuracy comparison in between original features, PCA Case 1 (99% of percentage variance) and PCA Case 2 (100% of percentage variance)

Case	R value	MAPE
CS10 (with original features)	0.994	2.57%
PCA Case 1	0.983	5.28%
PCA Case 2	0.996	2.80%

From the result, it is found that PCA with 99% percentage variance limit gives less accurate result compared to the original case (CS10) while almost the same level of accuracy can be achieved when PCA with 100% percentage variance limit applied. In addition, from Figure 4.21, it is found that the results are not biased for both cases.



(a) PCA Case 1 **(b)** PCA Case 2

Figure 4.21: Results of individual turbine level load estimation, Turbine 1, Operating condition 1, Lower monopile, Feed forward neural network, PCA Case 1 (99% of percentage variance) and PCA Case 2 (100% of percentage variance)

In addition, it has also been found that estimation with a few principal components (PCA Case 1) can give less accurate estimation compared to the original case (CS10). In other words, even though principal components have small percentage variance (with percentage variance smaller than 1%), it cannot be disregarded in moment DEL estimation.

4.8 Required number of data points for accurate estimation

In this section, the number of data points required for accurate estimation is described. For this study, for operating condition 1 and turbine 1, X directional moment has been estimated at lower monopile level with input data of CS10 since this case has been proven to be the most accurate. Specifically, the number of training set has been varied from 5% to 100% with 5% steps of total data. During the study, the portion in between training, cross-validation and test set has been constant as 70%, 15%, and 15% respectively. Since the total number of data is 3523 as shown in Table 2.2, 100% of the total number of training data is 2466 (70% of the total data). Lastly, MAPE has been plotted with respect to the number of training data to find the required number of training data. The result is shown in Figure 4.22.

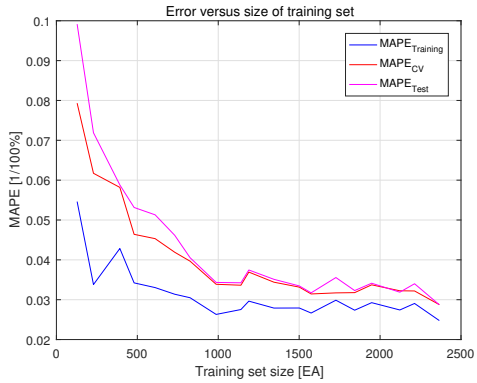


Figure 4.22: MAPE versus size of training set, Operating condition 1, Lower monopile, Feed forward neural network, CS10

First of all, as shown in Figure 4.22, both cross-validation and training errors are in the similar level and both are low when all the data is used, which indicates the feed-forward neural network trained on the full data has no overfitting/underfitting problem.

Secondly, the error is expected to be lower when there is more data available given that the error is not converged completely. In addition, the required number of data is dependent on the target level of accuracy. For CS10, if the target level of accuracy is 5%, the required number of data points is about 700 where the test set error (colored with magenta) is below 5%. The 700 data points are corresponding to 7000 min. since each data point stands for a statistical property based on 10 min. time series. Therefore, with input data of CS10 for operating condition 1, at least 5 days ($\sim$116.67 hours) of measurement data is required to train the neural network with less than 5% MAPE. However, it should be noted that these 5 days of measurement data should cover all the different ranges of features (i.e. different wind speed, different pitch angles, different standard deviation of acceleration and so on).

In addition to that, the result has been linearly extrapolated to investigate how much data is required to achieve MAPE of less than 1%. It is shown in Figure 4.23.

It has been found that at least around 7300 data points are required. It is corresponding to 51 days ($\sim$1216.67 hours). However, it should be noted that the MAPE can converge at some point. In other words, the lowest MAPE that the feed-forward neural network is able to achieve can be higher than 1%.

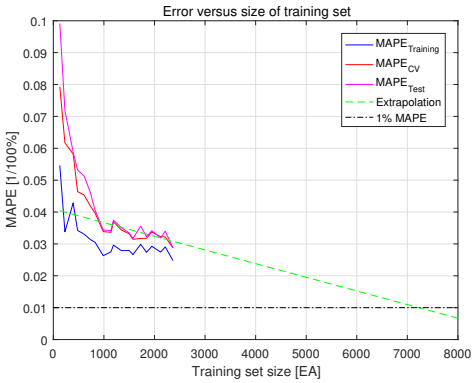


Figure 4.23: MAPE versus size of training set with extrapolation, Operating condition 1, Lower monopile, Feed forward neural network, CS10

4.9 Conclusion

In this chapter, individual turbine load estimation of moment DEL at four different levels has been made: tower bottom, transition piece and two different levels at monopile.

Firstly, with the base case including all the standard and non-standard signals, both the feed-forward neural network and recurrent neural network have been applied. For moment DEL estimation, it has been found that the feed-forward neural network outperforms recurrent neural network. The inaccuracy in recurrent neural network might be caused by small sequential dependency of moment DEL, or the small number of training data. Accordingly, the feed-forward neural network has been used for case studies.

Secondly, different case studies have been performed with the feed-forward neural network to find out the minimum required number of sensors for accurate estimation. Specifically, only the standard signals with and without non-standard statistical properties (damage equivalent values and spectral moments) have been applied. Only with the standard signals with all the statistical properties, estimation has been proved to be accurate with MAPE of around 3.5% at tower bottom, and less accurate with MAPE of around 7% at lower monopile level. When only the standard statistical properties have been applied, accuracy level decreases.

In addition, wave data has also been included as input. When the wave data is included, improvement in accuracy can be achieved for monopile locations. Especially at monopile locations, MAPE of around 6% could be achieved with standard signals. However, for the tower bottom and transition piece, the inclusion of wave data gives no significant improvement.

Added to that, non-standard signals at the tower bottom have been used as inputs. Specifically, moment and inclination signals at the tower bottom have been used. When all the standard signals and non-standard signals have been applied together with wave data, estimation is accurate with MAPE of less than 3%. When only the moment signal at the tower bottom is used together with standard signals and wave measurement data, MAPE of less than 5% can be achieved. Lastly, only the inclination signal at the tower bottom is used together with standard signals and wave measurement data, MAPE of less than 4% can be achieved.

As a result, the answer to one of the research question of the minimum required number of sensors has been found: it depends on target accuracy level. For example, to achieve MAPE of 5% for all locations, in addition to the standard signal from SCADA and wave measurement data, either strain gauges or inclinometers needs to be installed at tower bottom. If MAPE of less than 3% is target accuracy level, both strain gauges and inclinometers need to be installed at tower bottom.

In addition to the case studies, two more studies have been done to figure out if more improvement can be achieved.

First of all, operating condition 1 has been divided into two parts based on the mean wind speed of 10 m/s. For the divided operating condition 1, the accuracy level is increased when the mean wind speed is above 10 m/s while it decreases when the mean wind speed is below 10 m/s. When that estimation errors were combined, it has been found the accuracy level is worse than the case where full operating condition 1 has been used. That inaccuracy might be caused by the small number of training data since there were only around 1500 data points for training when operating condition 1 was divided into two parts, while there were around 3500 data points when full operating condition is considered.

Secondly, PCA has been applied to remove a possible negative effect from irrelevant features. As a result, it has been found that PCA with 99% percentage variance gave less accurate results. However, when PCA with 100% percentage variance has been applied, almost the same level of accuracy can be achieved with the case where original features are used.

Lastly, the required number of training data has been investigated with CS10 for the estimation of moment DEL at the lower monopile level. First of all, the required

number of data was dependent on the level of accuracy. If the target level of accuracy level is 5%, the required number of training data has been found as 700, which is corresponding to 5 days of measurements. However, it should be noted that the 5 days of measurement should include all the variation of input features (wind speed, rotational speed and so on). Therefore, the actual measurement period could be longer than 5 days.

Chapter 5

Farm-wide level load estimation

In this chapter, farm-wide level load estimation is described. First of all, the methodology applied for this study is described in Section 5.1. The applied dataset is explained in Section 5.2 followed by linear regression analysis for operating condition 3 and parked condition in Section 5.3. In Section 5.4 and 5.5, base case application and case studies with different input combinations are explained. Then, case studies with explicit inclusion of dynamic properties and wake effect are explained in Section 5.6. Lastly, conclusions are shown in Section 5.7.

5.1 Methodology

For farm-wide level load estimation, the same data used in individual turbine level load estimation have been applied. Since it has been proved that the feed-forward neural network gives better results than the recurrent neural network, the feed-forward artificial neural network has been used for farm-wide level load estimation as well.

Specifically, the feed-forward neural network has been trained for one turbine. Then, it has been applied for the other turbine and estimation accuracy has been calculated. If it can accurately estimate the response of the other turbine, it is expected that the trained neural network can be applied to all the turbines in the same wind farm. It is depicted in Figure 5.1.

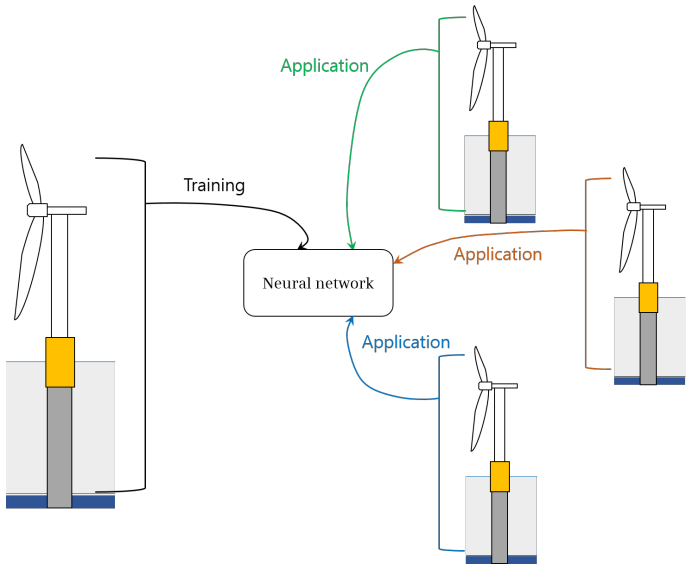


Figure 5.1: Farm-wide estimation, neural network application scheme

For farm-wide level load estimation, the following two additional aspects should be considered; different dynamic property and different wake directions.

In a wind farm, all the turbines are installed far from each other to minimize wake effect. Accordingly, water depth and soil condition can be different from turbine to turbine even though they are in the same wind farm. As a result, dynamic properties such as natural frequency and mode shapes can be different from turbine to turbine.

Specifically, natural frequency difference generates a difference in structural response frequency that causes a difference in fatigue damage accumulation at the end. Therefore, it is important to include this aspect in farm-wide level load estimation.

The difference in response frequency is expected to be included in spectral moments, and the spectral moments have already been included as input data for a neural network. Specifically, in PSD, peak points appear not only at the frequencies of excitation forces (i.e. wind, wave, 1P, 3P frequencies and so on), but also at structural response frequency. One example is shown in Figure 5.2 for moment signal. In the figure, the biggest peak is corresponding to the first damped natural frequency,

and the other frequencies are corresponding to frequencies of excitation forces.

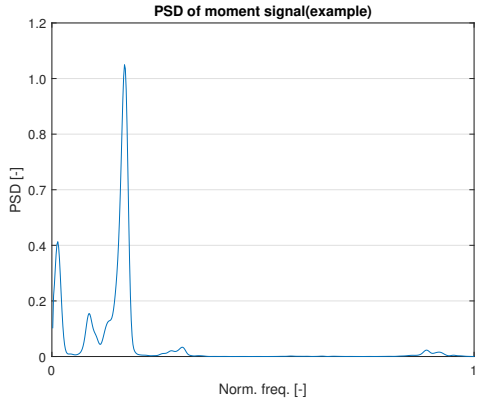


Figure 5.2: Example of PSD

As explained in Equation (A.7), a spectral moment is a function of frequency and PSD. Therefore, it is expected that the difference in dynamic properties is included in spectral moments. Especially, as shown in Equation (A.7), negative spectral moments can be thought to represent the effect of lower frequency components and positive spectral moments represent that of higher frequency components.

In addition to the spectral moments, different case studies have been performed to explicitly include the response frequency differences. In this chapter, it will be discussed.

In addition, according to a layout of a wind farm, wake direction and wake turbulence can be different from turbine to turbine. When a turbine is in wake, moment DEL tends to be higher compared to the case when the turbine is in free-stream condition [32, 33]. Therefore, the different wake turbulence effect should be included for farm-wide level load estimation.

In addition, It is expected that wake turbulence affects not only moment DEL, but also input features. Concretely, when there is higher wake turbulence, it is expected that there is higher fluctuation of moment, as well as acceleration. Therefore, a neural network is expected to be able to learn those different relationships in-between moment DEL and input features even though the wake direction and strength of turbulence are not explicitly put into the neural network.

Accordingly, multiple case studies have been performed with and without explicit inclusion of the wake effect.

Lastly, all the studies shown in this chapter is based on training with turbine 1 and application to turbine 2 (which will be denoted as T1A2: Training with turbine 1, Applied to turbine 2). The first reason for this is the number of data. As shown in Table 2.4, the number of data is not enough to train the feed-forward neural network when data of turbine 2 is used. Secondly, the range of input features and target moment DEL is larger in the turbine 1. Therefore, if a neural network is trained on turbine 1, the error which can be caused by the extrapolation problem shown in Section 5.2.1 can be minimized.

5.2 Applied dataset

For farm-wide level load estimation, the same one-month data used for individual turbine level load estimation has been used. The summary of data points is shown in Table 2.4. As already discussed in the individual turbine level load estimation, for operating condition 1,3 and parked condition, the number of data points is not enough for neural network training. For these cases, linear regression has been applied and the results are shown in Section 5.3. Only for operating condition 1, the feed-forward neural network has been applied. For the feed-forward neural network application, data points are sub-divided to differentiate extrapolation. It is described in Section 5.2.1. In addition, change of response frequency near structural natural frequency is explained in Section 5.2.2

5.2.1 Extrapolation with neural network

A neural network can give inaccurate result for extrapolation problem. In other words, a neural network is expected to be inaccurate for the data points in which the neural network is not trained on [30]. For individual turbine level load estimation, the possibility of extrapolation problem is low. However, In farm-wide level load estimation, all the turbines can have different ranges of input and target values (i.e. acceleration, wind speed, wind direction, moment DEL and so on). Therefore, there is a high chance of the extrapolation problem in farm-wide level load estimation. The two turbines used in this study also have different ranges of input and target values.

To investigate the effect of extrapolation, the feed-forward neural network has been trained on turbine 1 and applied to turbine 2 (T1A2). Input data for CS10 in Table 4.23 has been applied and X directional moment DEL at lower monopile location has been estimated. The result is shown in Figure 5.3.

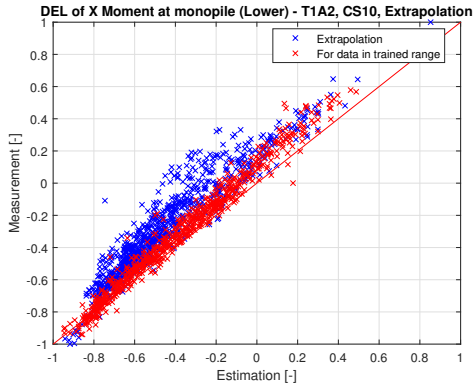


Figure 5.3: Result of Farm-wide level load estimation with effect of extrapolation, Operating condition 1, Lower monopile, With standard and non-standard (moment and inclination at tower bottom)signals, Standard statistical properties, With wave data, Feed forward neural network, T1A2, CS10

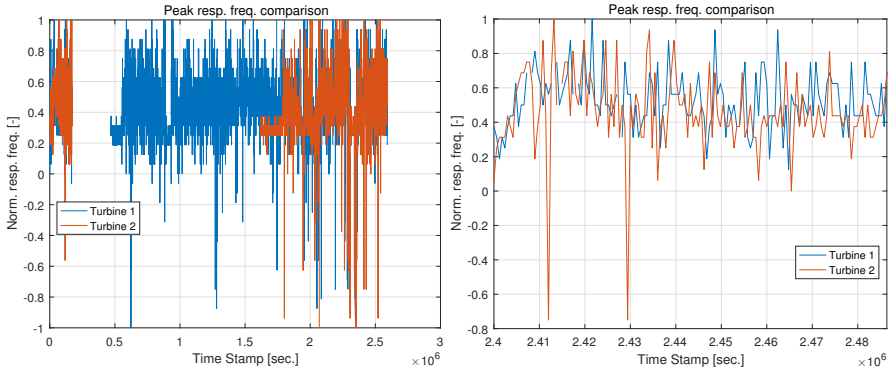
From Figure 5.3, it is found that the estimation for data located within the training range (colored with red) is closely gathered along with the linear line which means estimation is quite accurate. However, when the data points are located outside the training range (extrapolation, colored with blue), it is shown that the estimation is scattered and off from the linear line which indicates the estimation is inaccurate.

As a result, data points located outside the training range have been separately marked so that the error can be separately evaluated in extrapolation region. In addition, it should be noted that the number of data points located outside the training range can be different case by case since different combinations of input and target has been used for training.

5.2.2 Response frequency changes

During the one month, it has been found that the first damped natural frequency kept changing. The reason for this is that the aerodynamic damping is changed according to wind speed and rotational speed, and the effect of aerodynamic damping is large in offshore wind turbine [26]. From PSD of FA directional acceleration at RNA (from SCADA system), the first damped natural frequency has been extracted and compared in between two turbines. The result is shown in figure 5.4. The figure

on the left-hand side shows the difference for the full one month, and the figure on the right-hand side shows the difference for one day to clearly see the difference.



(a) Comparison for a month **(b)** Comparison for a day

Figure 5.4: Peak response frequency comparisons

As shown in the figure, two turbines have similar ranges of the first damped natural frequency changes.

5.3 Linear regression (for subset of operating conditions)

For individual turbine level load estimation, it has been found that linear regression can give the accurate estimation of moment DEL at the tower bottom, transition piece and upper monopile level for operating condition 3 and parked condition with MAPE of less than 5%. Additionally, It is expected that the linear regression for one turbine can be applied to other turbines in a wind farm if the design is the same. Therefore, in this section, linear regression has been applied for farm-wide estimation.

For this study, linear regression analysis has been performed on turbine 1, and it has been applied to the turbine 2. For linear regression, the equivalent value of acceleration at RNA has been used to estimate moment DEL at tower bottom, and moment DEL at tower bottom has been used for estimation at transition piece and monopile since those features have given the best results as shown in Section 4.3.

The results of X directional moment DEL estimation are shown in Table 5.1.

Table 5.1: Results for linear regression for farm-wide estimation, X directional moment DEL estimation, T1A2, Operating condition 3 and parked condition

Location	Feature	Oper. Cond. 3		Parked Cond.	
		R value	MAPE	R value	MAPE
Tower Bottom	RNA Acc Eq. value	0.996	1.84%	0.999	1.76%
Transition Piece	Tower Bottom Mom Eq. value	0.997	3.77%	0.993	1.30%
Monopile (upper)	Tower Bottom Mom Eq. value	0.996	3.60%	0.997	2.84%
Monopile (lower)	Tower Bottom Mom Eq. value	0.966	7.96%	0.950	3.93%

As shown in the results, linear regression gives almost the same accuracy in estimation compared to individual turbine level load estimation. For parked condition, linear regression gives less than 4% MAPE for all locations. For operating condition 3, it gives MAPE of less than 4% except for the lower monopile level where there is a relatively large MAPE of around 8%.

As a result, given that almost the same level of accuracy can be achieved with individual turbine level load estimation, it can be concluded that almost the same linear relationship exists between two turbines.

5.4 Feed forward neural network for base case

For farm-wide level load estimation, the same feed-forward neural network architecture used in individual turbine level load estimation has been applied. As a first study, the same base case used for individual turbine level load estimation has been applied to figure out if the feed-forward neural network can be used for farm-wide level load estimation. The base case is shown in Table 4.1. For this study, only operating condition 1 has been considered since the number of data is not enough for training as explained in Section 4.4.2.

As explained in Section 5.1, the difference in dynamic properties and wake effect should be included in farm-wide level load estimation. For the base case application, it should be noted that those effects have not been included explicitly.

5.4.1 Estimation result for base case

As a result of the base case application, the accuracy measurement comparison shown in Table 5.2 has been obtained. In addition, estimation of FA directional

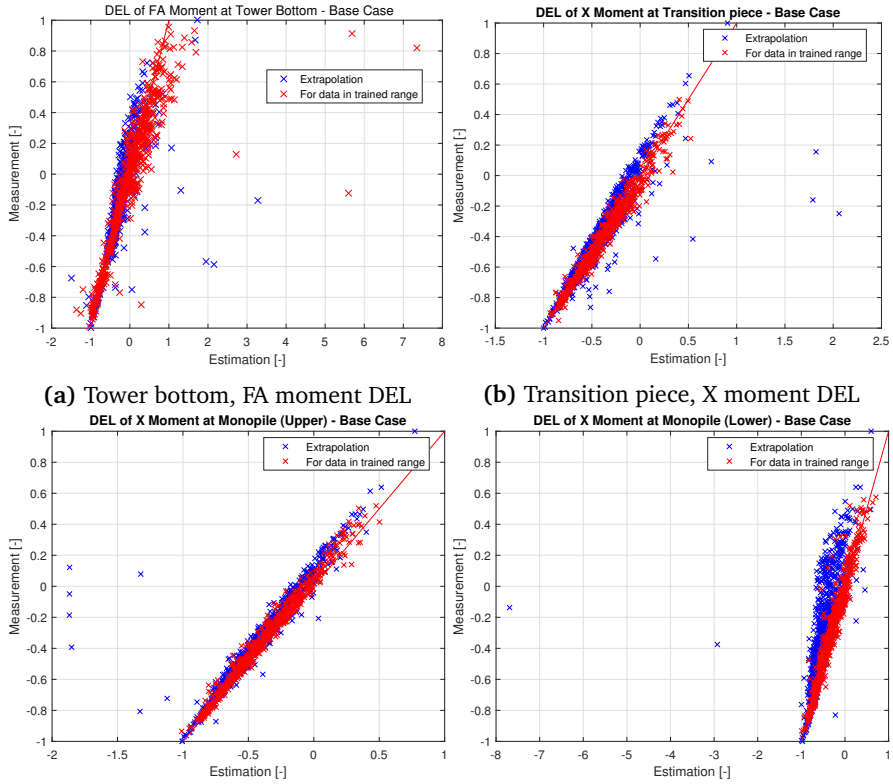
moment DEL at tower bottom and X directional moment DEL at transition piece and monopile are shown in Figure 5.5.

Table 5.2: Results of farm-wide estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, Base Case

Location	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Tower Bottom (FA)	0.899	10.97%	0.793	21.93%	0.841	17.47%
Tower Bottom (SS)	0.991	5.34%	0.972	6.85%	0.977	6.23%
Transition Piece (X)	0.983	5.26%	0.904	11.19%	0.932	8.79%
Transition Piece (Y)	0.960	7.46%	0.843	11.30%	0.896	9.75%
Monopile (upper) (X)	0.978	6.80%	0.893	16.09%	0.919	12.34%
Monopile (upper) (Y)	0.975	5.76%	0.876	8.99%	0.915	7.69%
Monopile (lower) (X)	**0.954**	**8.72%**	**0.749**	**21.63%**	**0.802**	**16.41%**
Monopile (lower) (Y)	**0.957**	**12.23%**	**0.877**	**17.93%**	**0.903**	**15.63%**

First of all, it has been found that the feed-forward neural network can be used for farm-wide level load estimation. In detail, from Table 5.2, it is shown that MAPE is above 5% for all locations. When the MAPE is compared between the results from estimation with data within the training range and outside the training range (extrapolation), it is obvious that there is a higher error in extrapolation. From Figure 5.5, it can be seen that extrapolation (blue-colored) has higher scatter. Added to that, when the error at tower bottom is compared, it is seen that there is a higher error in FA direction. In addition, similar to individual turbine level load estimation, the error tends to increase as going down deeper.

From Figure 5.5, it is shown that there is higher scatter compared to individual turbine level load estimation. Especially, for FA directional moment DEL estimation at tower bottom shown in Figure 5.5a, when moment DEL is higher, there is bias and estimation is higher than measurement. On the other hand, At upper and lower monopile in Figure 5.5c and 5.5d, the results tend to underestimate moment DEL regardless of the magnitude of moment DEL. At the transition piece, the estimation is not biased. At all locations, even when the data within the training range is applied (red-colored), there are some outliers that indicate there is a large difference in between estimation and measurement. This large error is caused by the different operating conditions in two turbines. Specifically, as mentioned in Section 2.2, the power generation of turbine 2 is sometimes limited to below its rated power even when the wind speed is high enough. However, for turbine 1, it always generates power along with its power curve. In other words, the neural network trained on



(a) Tower bottom, FA moment DEL

(b) Transition piece, X moment DEL

(c) Monopile (upper), X moment DEL

(d) Monopile (lower), X moment DEL

Figure 5.5: Results of farm-wide level load estimation, FA and X directions, Operating Condition 1, T1A2, Feed forward neural network, Base Case

turbine 1 cannot fully capture the different behavior when the turbine generates below its rated power when wind speed is high enough.

In Section 5.4.2, estimation result without these data points from turbine 2 is shown.

5.4.2 Estimation without power generation below rated

After removing all the data points corresponding to the case where power generation is below its rated power for turbine 2, the total number of data has been found to be 647. The power generation versus wind speed with removed data points are shown in Figure 5.6.

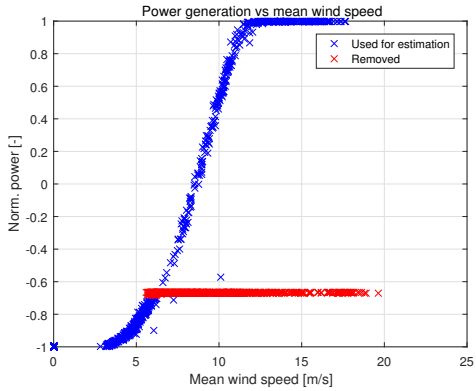


Figure 5.6: Mean power generation vs mean wind speed (turbine 2), with removed data points

With the base case, the results shown in Table 5.3 and Figure 5.7 have been obtained.

First of all, from Table 5.3, it is found that the accuracy level is increased when data points which the neural network is trained on have been applied. However, for extrapolation, it gives worse results. Overall, the accuracy level is increased. Especially at tower bottom level, the difference between FA and SS directional estimation is reduced. It can be interpreted that the different operating condition affects FA directional features more than SS directional features. For example, if the turbine 2 is operating with different pitch angle and rotational speed compared to turbine 1, thrust force would be affected largely, which causes the large difference in input features such as acceleration and moments. However, there are still differences in estimation accuracy between FA and SS direction. It can be explained with the larger range of input features and target moment DEL in FA direction as described in Section 4.6.2.

From Figure 5.7, it is found that the estimation is less scattered compared to the case before. However, there is still under or overestimation of moment DEL.

Table 5.3: Results of farm-wide level load estimation, moment DEL estimation for all locations, Operating condition 1, T1A2, Feed forward neural network, Base Case without some data points below rated power

Location	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Tower Bottom (FA)	0.971	6.82%	0.886	26.28%	0.946	10.64%
Tower Bottom (SS)	0.994	4.59%	0.913	6.58%	0.980	4.83%
Transition Piece (X)	0.986	4.90%	0.934	11.20%	0.976	5.87%
Transition Piece (Y)	0.968	7.60%	0.929	18.43%	0.955	9.27%
Monopile (upper) (X)	0.990	5.92%	0.979	8.69%	0.989	6.35%
Monopile (upper) (Y)	0.984	5.07%	0.950	12.79%	0.977	6.27%
Monopile (lower) (X)	**0.972**	**7.18%**	**0.891**	**15.31%**	**0.959**	**8.43%**
Monopile (lower) (Y)	**0.966**	**12.50%**	**0.912**	**30.51%**	**0.956**	**15.28%**

From the results, again, it has been found that neural network should be applied on dataset which the neural network is trained on. Therefore, for further studies, the data points corresponding to the case where turbine 2 is generating less than its rated power have been disregarded.

5.5 Case study for different input combinations

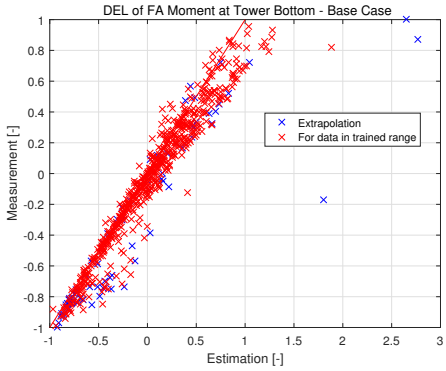
5.5.1 Estimation results

For farm-wide level load estimation, multiple case studies have been performed.

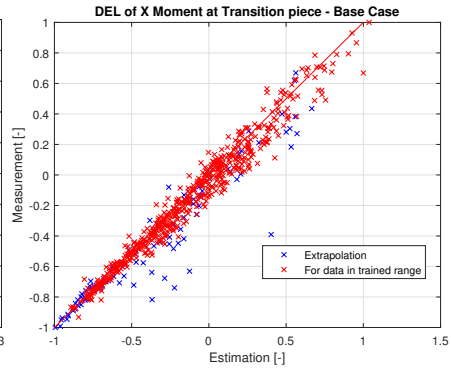
Specifically, case studies with different combinations of input features have been performed to answer the research question of **'What is the minimum number of sensors that gives accurate DEL estimation?'**. For this study, CS1∼CS12 in Section 4.6 have been considered excluding the mean yaw angle. The reason for that is since the mean yaw angle can cause an error in estimation. For instance, if a neural network is trained on mean yaw angle as well, it is expected that the neural network would give the wrong estimation at the specific yaw angles where turbine 1 (with which the neural network is trained on) is in wake, but turbine 2 is not in wake since the neural network would make estimation as if turbine 2 is also in wake.

The result is shown only for the lower monopile level. Then, for the best case, all the results are shown.

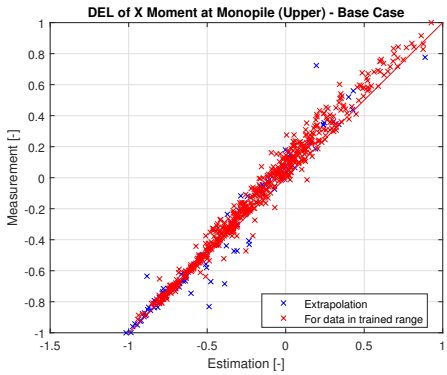
The accuracy measurement comparison is shown in Table 5.4 and Figure 5.8.



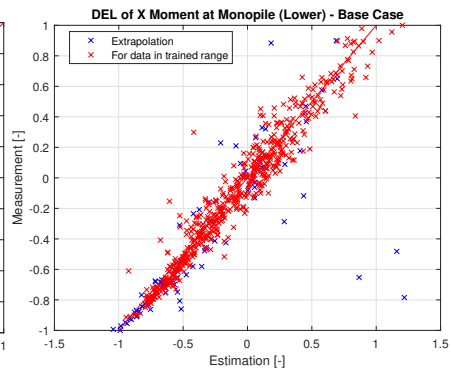
(a) Tower bottom, FA moment DEL



(b) Transition piece, X moment DEL

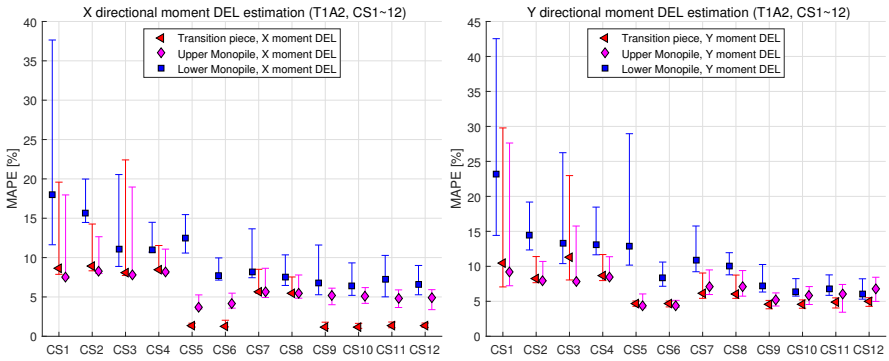


(c) Monopile (upper), X moment DEL



(d) Monopile (lower), X moment DEL

Figure 5.7: Results of farm-wide level load estimation, FA and X directions, Operating Condition 1, T1A2, Feed forward neural network, Base Case without some data points below rated power



(a) X directional **(b)** Y directional

Figure 5.8: MAPE Comparison in between CS1∼CS12, T1A2

Table 5.4: Results of farm-wide estimation, Lower monopile, Operating condition 1, T1A2, Feed forward neural network, CS1∼CS12

	Direction	within Train Range		Extrapolation		Full	
		R value	MAPE	R value	MAPE	R value	MAPE
CS1	X	0.908	18.01%	0.810	33.18%	0.895	19.20%
	Y	0.875	23.14%	0.774	42.85%	0.815	24.69%
CS2	X	0.911	15.70%	0.849	21.14%	0.905	15.97%
	Y	0.933	14.50%	0.915	26.99%	0.932	15.14%
CS3	X	0.930	11.04%	0.730	27.94%	0.894	12.37%
	Y	0.952	13.30%	0.872	19.81%	0.935	13.81%
CS4	X	0.952	10.97%	0.935	17.46%	0.950	11.29%
	Y	0.944	13.08%	0.856	27.47%	0.936	13.82%
CS5	X	0.915	12.45%	0.943	12.42%	0.954	12.44%
	Y	0.912	12.85%	0.957	27.03%	0.956	17.08%
CS6	X	0.955	7.67%	0.962	8.80%	0.975	8.00%
	Y	0.966	8.33%	0.978	18.98%	0.982	11.50%
CS7	X	0.970	8.18%	0.981	6.76%	0.971	8.09%
	Y	0.946	10.86%	0.976	15.65%	0.950	11.17%
CS8	X	0.978	7.53%	0.981	12.72%	0.979	7.87%
	Y	0.958	10.01%	0.972	16.10%	0.960	10.41%
CS9	X	0.958	6.82%	0.975	6.88%	0.977	6.84%
	Y	0.949	7.20%	0.982	14.38%	0.976	9.34%
CS10	**X**	**0.962**	**6.52%**	**0.978**	**6.92%**	**0.980**	**6.67%**
	Y	**0.961**	**6.35%**	**0.984**	**13.41%**	**0.981**	**8.46%**
CS11	X	0.968	7.27%	0.975	6.65%	0.984	7.11%
	Y	0.950	6.76%	0.979	14.10%	0.977	8.71%
CS12	X	0.968	6.55%	0.968	6.62%	0.984	6.57%
	Y	0.956	7.00%	0.972	14.21%	0.979	8.58%

In general, the results are similar to that of individual turbine level load estimation. However, the overall accuracy level is decreased. Specifically, when only the standard signals have been used (CS1∼CS4), MAPE of higher than 10% has been obtained. When either moment or inclination signal has been used (CS5∼CS8), the accuracy level is higher than 7%, and the best accuracy level has been obtained when moment signal at tower bottom location is used together with wave data in addition to standard signals (CS6).

When both moment and inclination signals have been used, the accuracy level increased with MAPE of around 7%. The best accuracy level has been found when both moment and inclination signals have been used together with wave data in addition to standard signals (CS10).

Since the best case is CS10, accuracy measurements are calculated for all locations as shown in Table 5.5. The comparisons in between estimation and measurement are shown not only for both X, but also for Y direction since it needs to be discussed. It is shown in Figure 5.10.

Table 5.5: Results of farm-wide estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10

	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Transition Piece (X)	0.998	1.21%	0.999	1.11%	0.999	1.18%
Transition Piece (Y)	0.998	4.60%	0.999	4.18%	0.999	4.48%
Monopile (upper) (X)	0.996	5.10%	0.997	3.97%	0.998	4.76%
Monopile (upper) (Y)	0.984	5.84%	0.994	4.12%	0.992	5.32%
Monopile (lower) (X)	**0.962**	**6.42%**	**0.978**	**6.92%**	**0.980**	**6.57%**
Monopile (lower) (Y)	**0.961**	**6.35%**	**0.984**	**13.41%**	**0.981**	**8.46%**

From table 5.5, it can be found that the accuracy level is the worst at the lower monopile level and gets better as the level is higher. At the transition piece level, estimation is accurate with MAPE of less than 5%. Especially, Y directional estimation is less accurate than X directional estimation.

As shown in figure 5.10, X directional moment DEL estimation at transition piece level is not biased, while Y directional estimation is biased and it overestimates the response. At the lower and upper monopile levels, both X and Y directional moment DEL estimation are biased. In opposite to that, estimation at the upper monopile level tends to underestimate the response.

In this study, since there is not enough data for strain gauges at transition piece and monopile, the root cause of the different trends could not be found. Furthermore, since X and Y directional moment have been used instead of FA and SS directional for transition piece and monopile, it is difficult to interpret the result. If FA and SS directional moment had been used, it is expected that FA directional estimation would not be biased, but SS directional estimation would be biased. The reason for this is that aerodynamic damping is similar in both turbines, while soil condition and water depth are different. For FA direction, aerodynamic damping would be dominant compared to the other effects such as soil damping and drag damping. However, for SS direction, soil damping and drag damping are the governing factors and the different soil condition and water depth would affect the response. Therefore, if structural dynamic properties are similar in both turbines, a similar dynamic response is expected in FA direction while there would be a difference in SS direction.

To explain the different trends with the given data, some hypotheses have been made.

Firstly, it is possible that strain gauges are not installed equidistantly. Since the moment calculated from the strain assumes equidistant installation of strain gauges, the actual moment can be different. Especially at transition piece, the overestimation only in Y directional moment DEL estimation can be explained in this way.

Secondly, it is possible that the strain gauges are installed at different elevations. In other words, the 'transition piece', 'upper monopile' and 'lower monopile' level can be different in two turbines. If that is the case, the different bias in between upper monopile level and lower monopile & transition piece can be explained. For example, if the 'upper monopile' level of turbine 1 is higher than the turbine 2 and the feed-forward neural network is trained on turbine 1, moment DEL can be underestimated since moment is lower at a higher level.

Thirdly, different mode shapes in higher mode could affect the estimation. Specifically, turbine 1 has deeper water depth than turbine 2. When the typical second mode shape of fixed type wind turbines is considered, it is possible that the upper monopile level is located near the node for turbine 1, while it is located far from the node. It is shown in Figure 5.9. In that case, moment induced by the second mode would be higher in turbine 2, which indicates the feed-forward neural network trained on turbine 1 would underestimate moment from the second mode. In this way, the underestimation in the upper monopile level can be explained.

Lastly, the answer to the research question of the required number of sensors has been found. Similar to the individual turbine level load estimation, the required

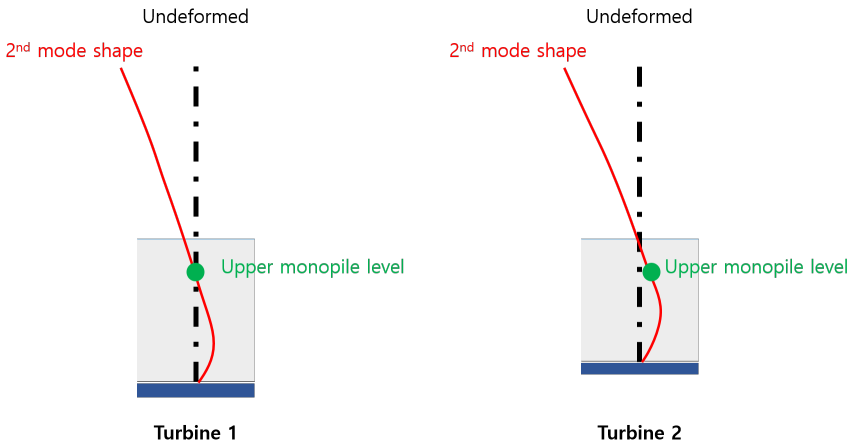
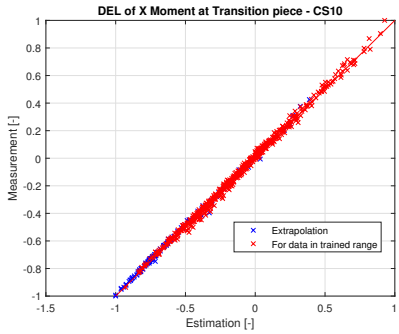
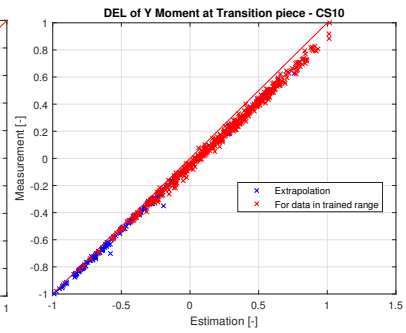


Figure 5.9: Different second mode shapes in two turbines

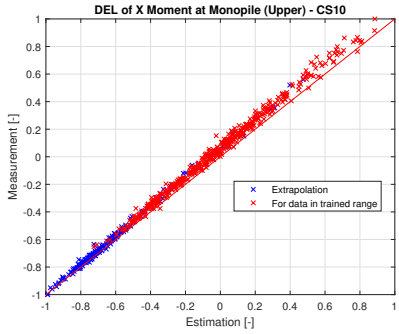
number of sensors for accurate estimation is dependent on the level of accuracy. For example, to achieve MAPE of less than 7%, it can be concluded that both inclination and moment signals should be applied in addition to wave measurement and standard signals. However, it should be noted that estimation accuracy can change. Specifically, two turbines used in this study have a very similar range of first damped natural frequencies as shown in Section 5.2.2. Therefore, if the feed-forward neural network is used for a turbine which has a significantly different range of damped natural frequency, the different estimation accuracy can be obtained.



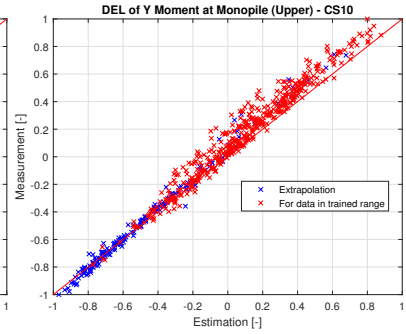
(a) Transition piece, X moment DEL



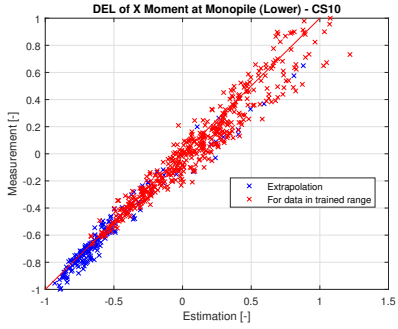
(b) Transition piece, Y moment DEL



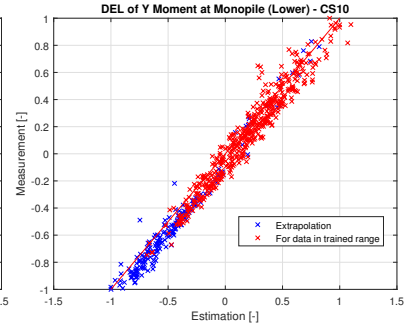
(c) Monopile(upper), X moment DEL



(d) Monopile(upper), Y moment DEL



(e) Monopile(lower), X moment DEL



(f) Monopile(lower), Y moment DEL

Figure 5.10: Results of farm-wide estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10, X direction

5.5.2 Summary of the results

For a clear distinction, the estimation results are summarized. Similar to the individual turbine level, a summary of the estimation results at the lower monopile level are shown in this section.

The summarized results are shown in Table 5.6.

Table 5.6: Summary of results

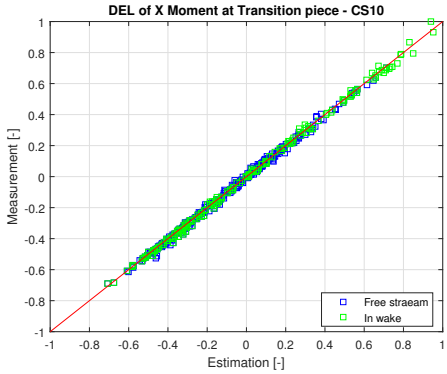
Standard signals?	Additional features	Case number	MAPE
O	None	CS1	<23.5%
O	Wave	CS3	<13.5%
O	Moment at tower bottom	CS5	<13%
O	Inclination at tower bottom	CS7	<11%
O	Moment and Inclination at tower bottom	CS9	<7.5%
O	Moment at tower bottom + Wave	CS6	<8.5%
O	Inclination at tower bottom + Wave	CS8	<10.5%
O	Moment and Inclination at tower bottom + Wave	CS10	<7%
X	Moment and inclination at tower bottom	CS11	<7.5%
X	Moment and Inclination at tower bottom + Wave	CS12	<7%

5.6 Case study for improvement

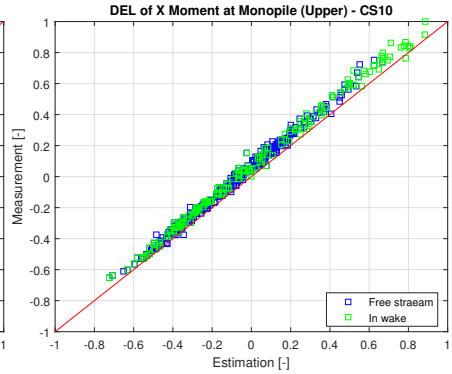
First of all, to figure out if the different wake in both turbines affects the results, resultant estimation shown in Figure 5.10 has been labeled with 'free stream' and 'in wake' by comparing mean yaw direction with the wake directions shown in table 2.8. As shown in the figure, the level of accuracy is almost the same in both 'free stream' and 'in wake' conditions, which indicates wake effect can be captured by the feed-forward neural network.

However, both dynamic properties and wake effect have been included explicitly in the feed-forward neural network to investigate if any improvement can be achieved.

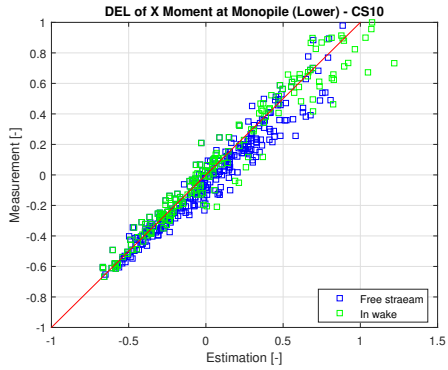
In Section 5.6.1 and Section 5.6.2, explicit inclusion of dynamic properties and



(a) Transition piece, X moment DEL



(b) Monopile (upper), X moment DEL



(c) Monopile (lower), X moment DEL

Figure 5.11: Results of farm-wide level load estimation with wake direction, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10

its results are explained. In Section 5.6.3, normalization of signals and its results are described. Lastly, in Section 5.6.4, wake effect inclusion is shown.

5.6.1 Inclusion of peak frequencies and values

To include dynamic property as input, peak frequencies and its PSD values are directly used as input. Since two turbines have similar changes in the first damped natural frequency, it is expected that the feed-forward neural network trained for turbine 1 can differentiate the different behavior according to its first damped natu-

ral frequency change, and can be applied to turbine 2.

For this study, two cases have been investigated:

- Case 1: 1 peak near structural natural frequency is included with corresponding frequency
- Case 2: Highest 4 peaks and its frequencies are included

The first case has been performed to include the first damped natural frequency, and the second case has been performed to include peak wind frequency, 1P, and 3P frequencies as well. In Figure 5.12, an example of moment PSD with four peaks is shown.

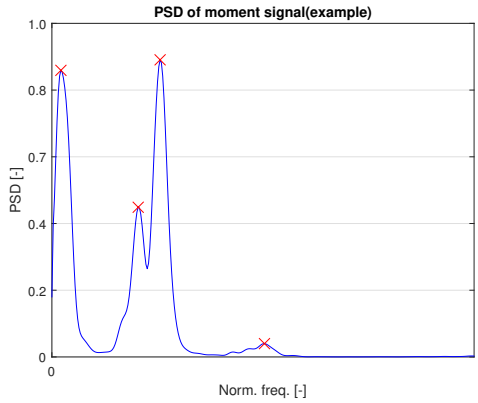


Figure 5.12: Example of PSD with 4 peaks

As a result, accuracy measurements have been calculated as shown in Table 5.7 and 5.8. It has been fount the result is almost the same and sometimes worse. Furthermore, the estimation has still been biased as shown in Figure 5.13 and 5.14.

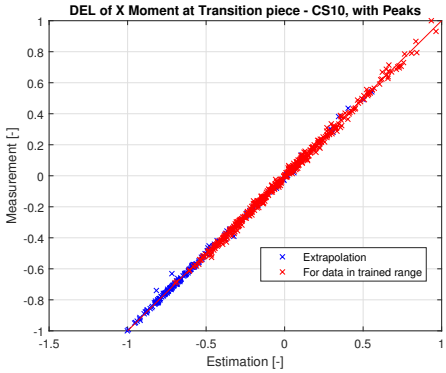
The result can be interpreted with spectral moments. As already explained in Section 5.1, spectral moments already include dynamic properties. From the result, it can be concluded that the explicit inclusion of PSD peaks and its corresponding frequencies give no additional information to the feed-forward neural network.

Table 5.7: Results of farm-wide level load estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with 4 peaks from PSD

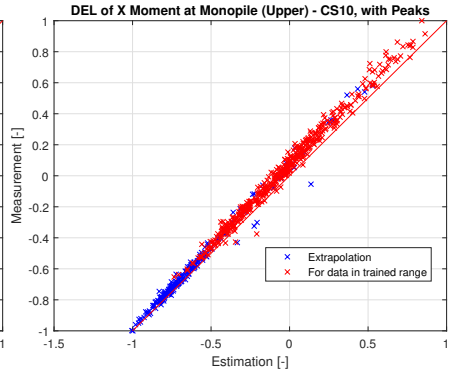
	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Transition Piece (X)	0.999	1.21%	0.999	1.18%	0.999	1.20%
Transition Piece (Y)	0.998	4.78%	0.999	4.29%	0.999	4.63%
Monopile (upper) (X)	0.996	5.06%	0.997	3.71%	0.998	4.65%
Monopile (upper) (Y)	0.980	5.30%	0.994	4.16%	0.991	4.95%
Monopile (lower) (X)	0.963	6.38%	0.980	7.36%	0.980	6.68%
Monopile (lower) (Y)	0.962	6.93%	0.983	15.38%	0.982	9.53%

Table 5.8: Results of farm-wide level load estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with 1 peak from PSD

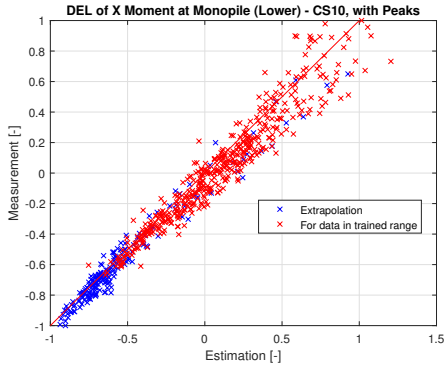
	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Transition Piece (X)	0.999	1.20%	0.998	1.11%	0.999	1.17%
Transition Piece (Y)	0.998	4.73%	0.999	4.25%	0.999	4.58%
Monopile (upper) (X)	0.997	5.10%	0.997	4.04%	0.998	4.78%
Monopile (upper) (Y)	0.984	5.77%	0.995	4.13%	0.993	5.28%
Monopile (lower) (X)	0.962	6.63%	0.980	6.91%	0.980	6.71%
Monopile (lower) (Y)	0.959	6.56%	0.982	14.08%	0.980	8.80%



(a) Transition piece, X moment DEL

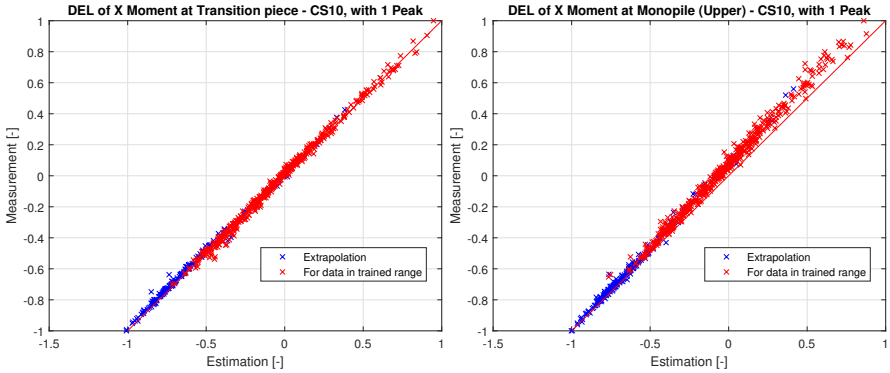


(b) Monopile (upper), X moment DEL

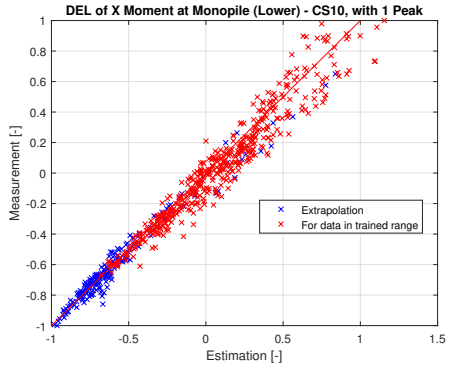


(c) Monopile (lower), X moment DEL

Figure 5.13: Results of farm-wide level load estimation with wake direction, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with 4 peaks from PSD



(a) Transition piece, X moment DEL **(b)** Monopile (upper), X moment DEL



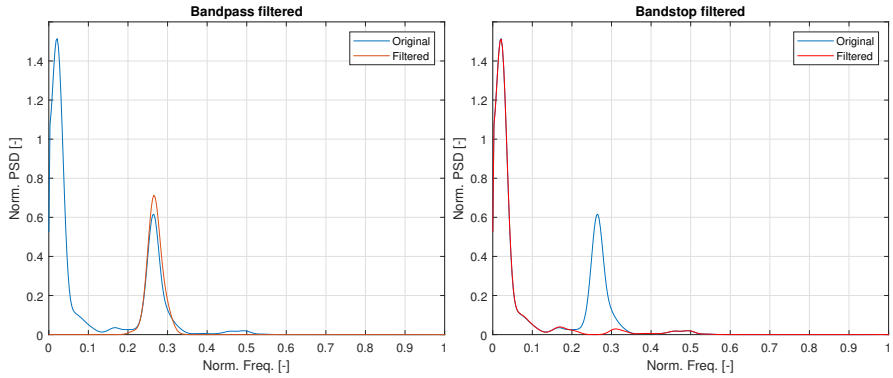
(c) Monopile (lower), X moment DEL

Figure 5.14: Results of farm-wide level load estimation with wake direction, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with 1 peak from PSD

5.6.2 Inclusion of filtered signals

To include dynamic property, original signals (i.e. acceleration, moment, inclination and so on) have been filtered. For the filtered signals, all the statistical properties have been calculated and it has been applied to the feed-forward neural network.

Specifically, original signals have been divided into two filtered signals; bandpass and bandstop filtered for its first damped natural frequency. Example PSDs of the two filtered signals are shown in Figure 5.15.



(a) Bandpass filtered signal **(b)** Bandstop filtered signal

Figure 5.15: Example PSDs of filtered signals

The idea behind this approach is that the bandpass filtered signal represents response related to the structural natural frequency and the bandstop filtered signal represents response related to the others (i.e. excitation forces). Here, the bandstop filtered signal is expected to be almost the same for both turbines since the excitation forces are the same for both turbines, while the bandpass filtered signal is expected to be different since two turbines have different dynamic properties. However, as explained in Section 5.2.2, peak frequency near structural natural frequency keeps changing during the one month. Therefore, it might be possible to capture how the different combinations of the bandstop filtered signal and the bandpass filtered signal (with the changing peak response frequency near structural natural frequency) are related to moment DEL with the feed-forward neural network. One remark is that higher structural modes are not filtered and it has been included in bandstop filtered signals. However, even though filtering is based only on the first mode, it is expected that the results would not be affected significantly by higher mode

components since the contribution of higher modes is lower than other frequency components as shown in the PSDs.

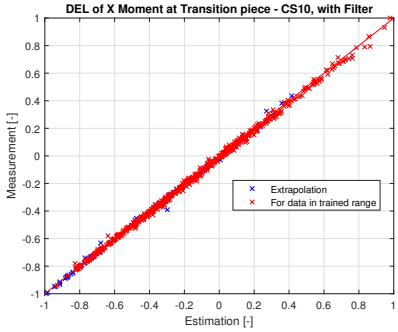
For this study, the 4th order Butterworth filter has been applied. The bandpass and bandstop filtered signals have been applied in addition to the input features of CS10. The result is shown in Table 5.9 and Figure 5.16.

Table 5.9: Results of farm-wide level load estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with filtered signals

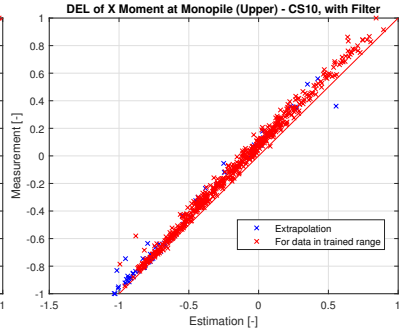
	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Transition Piece (X)	0.999	1.21%	0.939	1.72%	0.978	1.36%
Transition Piece (Y)	0.998	4.49%	0.998	3.36%	0.999	4.15%
Monopile (upper) (X)	0.994	4.59%	0.979	4.54%	0.994	4.58%
Monopile (upper) (Y)	0.987	3.96%	0.917	9.86%	0.974	5.72%
Monopile (lower) (X)	0.957	7.02%	0.919	10.44%	0.969	8.04%
Monopile (lower) (Y)	0.962	8.45%	0.894	28.56%	0.958	14.45%

As shown in the table, the estimation accuracy has been decreased. In addition, the results have been still biased as shown in the figure.

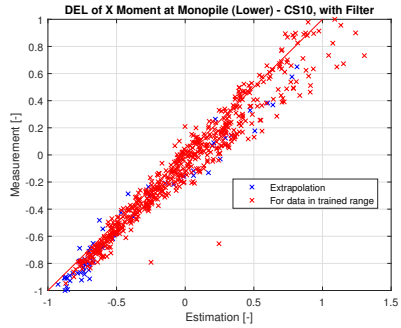
From the result, it could be concluded that the explicit inclusion of dynamic properties in a form of filtered signals is not beneficial to estimate moment DEL.



(a) Transition piece, X moment DEL



(b) Monopile(upper), X moment DEL



(c) Monopile(lower), X moment DEL

Figure 5.16: Results of farm-wide level load estimation with wake direction, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with filtered signals

5.6.3 Normalization of signals

To include dynamic property, explicit inclusions have been performed as explained in previous sections. In addition to that, normalization of signals has also been investigated and it is explained in this section.

Specifically, acceleration, moment and inclination signals have been normalized with respect to its frequency component which is corresponding to the response peak frequency near the first structural natural frequency. The idea behind this approach is that the normalized signals might not include effect from structural dynamics. Therefore, if the statistical properties of the normalized signals can be used to estimate the normalized moment DEL, it is expected that the feed-forward neural network can be used to estimate normalized moment DEL regardless of different dynamic properties of different turbines.

One remark here is that both input features and target moment DEL should be normalized to be able to remove the effect from structural dynamics. In this case, even though the feed-forward neural network can accurately estimate the normalized target moment DEL, it is difficult to convert the normalized target moment DEL back into the original target moment DEL. Even though there is a difficulty, it has been investigated since it might also be possible to apply a neural network technique to convert the normalized moment DEL into the original moment DEL.

For this normalization, the following approaches have been taken.

- Multiplication : $S_{norm} = S_{original} \times S_{filtered}$
- Division : $S_{norm} = S_{original} / S_{filtered}$

Where S_{norm} is the normalized signal in the time domain, $S_{original}$ is the original signal in the time domain, $S_{filtered}$ is bandpass filtered signal around response peak frequency near structural natural frequency and $\mu_{original}$ is mean of $S_{original}$. In this section, only the results of normalization with multiplication are shown. Results with the division are in Appendix H.

Concretely, 10 min. time series of original signal $S_{original}$ has been used for the normalization. Example PSD of $S_{original}$ and $S_{filtered}$ are shown in blue and brown line in Figure 5.15a respectively.

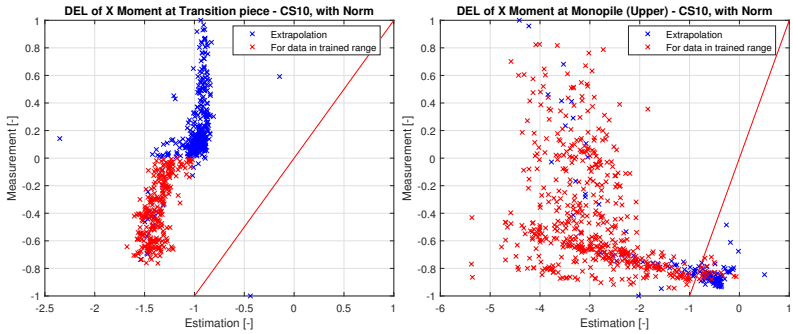
For this study, input features in CS10 have been used. Among all the input features, acceleration and moment and inclination signals have been normalized first, and then applied to the feed-forward neural network. The results for normalization with multiplication are shown in Table 5.10 and Figure 5.17.

From the results, it has been found the normalized signals give the completely wrong estimation even though the target moment DEL has also been normalized in the same way. For normalization with division, it has given similar results.

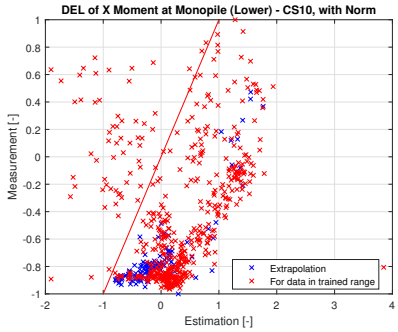
Table 5.10: Results of farm-wide level load estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with normalization (multiplication)

	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Transition Piece (X)	0.309	>100%	0.253	>100%	0.620	>100%
Transition Piece (Y)	-0.923	>100%	-0.571	>100%	-0.376	>100%
Monopile (upper) (X)	0.842	>100%	0.832	>100%	0.849	>100%
Monopile (upper) (Y)	-0.643	>100%	-0.178	>100%	-0.606	>100%
Monopile (lower) (X)	0.770	>100%	0.759	>100%	0.756	>100%
Monopile (lower) (Y)	0.161	>100%	0.796	>100%	0.659	>100%

From the results, it can be concluded that the statistical properties from the normalized signals with multiplication and division cannot be used. In other words, the hypothesis that the statistical properties of the normalized signals can be used for estimation is not correct. Given that the estimation has been scattered largely, it can be concluded that there is no clear relationship between input features and moment DEL calculated from normalized signals.



(a) Transition piece, X moment DEL **(b)** Monopile(upper), X moment DEL



(c) Monopile(lower), X moment DEL

Figure 5.17: Results of farm-wide level load estimation with wake direction, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with normalization (multiplication)

5.6.4 Inclusion of wake effect

To explicitly include wake effect, wake regions shown in Table 2.8 have been used. Specifically, two more features have been added. The first additional feature is 'wake switch'. When the mean yaw angle is corresponding to one of the wake regions shown in Table 2.8, the 'wake switch' has been set as 1, otherwise, it has been set as 0. The second additional feature is 'wake distance'. The distance has been calculated from layout information, The 'wake distance' is different for different wake regions.

The idea behind this approach is that the feed-forward neural network would clearly recognize the different behavior when the turbine is in wake. To be able to make the feed-forward neural network recognize the turbine is in wake, 'wake switch' has been used. Specifically, weights related to 'wake switch' would affect the estimation only when the 'wake switch' is set as 1 since weights would be multiplied with 0 when 'wake switch' is set as 0. In other words, according to 'wake switch', the feed-forward neural network is expected to use different relationships. In addition, the strength of wake turbulence would be affected by many factors; distance to the upstream wind turbine, operating condition of the upstream wind turbine, turbulence intensity of the region and so on. Of all the factors, only the distance data could be obtained and it has been used. Specifically, wake turbulence is stronger when the distance to the upstream turbine is shorter. Therefore, it is expected that the feed-forward neural network can capture the strength of wake turbulence by using distance to the upstream turbine.

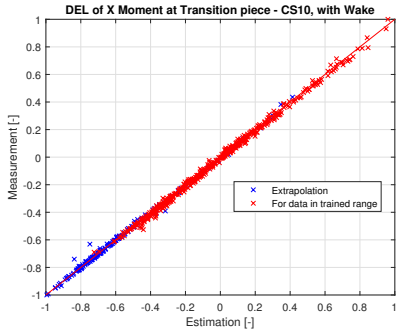
For this study, input features of CS10 have been used in addition to the two new features related to wake effect. As a result, accuracy measurements shown in Table 5.11 and Figure 5.18 has been found.

Table 5.11: Results of farm-wide level load estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with wake effect

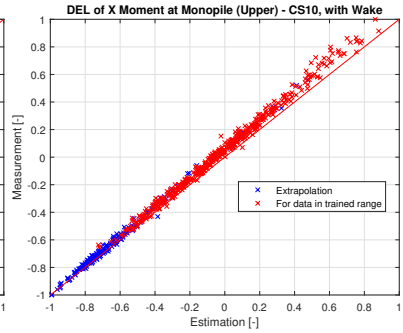
	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Transition Piece (X)	0.999	1.11%	0.999	1.24%	0.999	1.17%
Transition Piece (Y)	0.998	4.65%	0.999	4.24%	0.999	4.47%
Monopile (upper) (X)	0.996	5.06%	0.997	4.38%	0.997	4.75%
Monopile (upper) (Y)	0.983	5.52%	0.995	4.90%	0.992	5.24%
Monopile (lower) (X)	0.964	6.92%	0.986	8.13%	0.980	7.46%
Monopile (lower) (Y)	0.939	7.33%	0.978	12.31%	0.968	9.58%

As a result, a similar level of accuracy has been calculated compared to the case

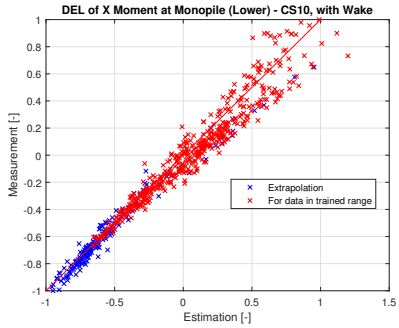
without explicit inclusion of wake effect. When the combined weights of the trained neural network have been compared, it has been found that the combined weights for the additional two features are small compared to the other features as shown in Figure 5.19.



(a) Transition piece, X moment DEL



(b) Monopile(upper), X moment DEL



(c) Monopile(lower), X moment DEL

Figure 5.18: Results of farm-wide level load estimation with wake direction, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with wake inclusion

As already explained in Section 5.1, wake not only affects target moment DEL, but also input features such as moment DEL at the tower bottom. In Section 5.6, it is shown that MAPE of 'free stream' and 'in wake' conditions have been almost the same which indicates both target moment DEL and input features are affected by wake, and the feed-forward neural network can capture it. From the study in this section, the same conclusion could be made. In addition, given that the combined weights

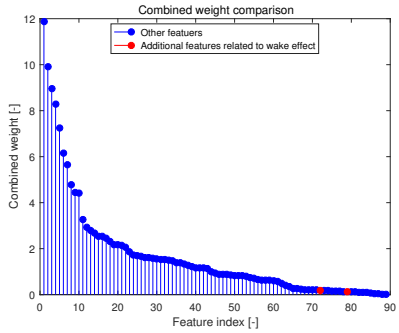


Figure 5.19: Combined weights comparison, CS10 with wake effect

of the additional features are very small compared to the other features, it has also been concluded that explicit inclusion of wake effect in a form of 'wake switch' and 'wake distance' gives no important information to the feed-forward network.

5.7 Conclusion

In this chapter, farm-wide level load estimation of moment DEL has been performed. Specifically, the feed-forward neural network has been trained on turbine 1, and applied to turbine 2 since turbine 1 has more data points.

Firstly, with the base case application, it has been found that the feed-forward neural network can be used in farm-wide level load estimation for the two turbines used in this study, but the accuracy level is lower than individual turbine level load estimation. Especially, since turbine 2 has the different operating condition in which it generates below rated power even though the wind speed is high enough. Since the feed-forward neural network trained on turbine 1 has not been trained on this operating condition, it gives inaccurate estimation.

Secondly, as a result of case studies with different input combinations, it has been found that CS10 which includes moment and inclination at tower bottom in addition to standard signals and wave measurement data gives the best estimation with MAPE of around 7%. Either moment or inclination signal at tower bottom has been used with standard signals and wave measurement data, it gives MAPE of around 10%. Moreover, not like the individual turbine level load estimation, the results have been biased. Specifically, the feed-forward neural network tends to overestimate moment DEL at transition piece (only Y-directional) and lower monopile level (both direction). However, at the upper monopile level, it tends to underestimate.

As a result of the case studies, the answer to the research question of ***'what is the minimum number of sensors that gives accurate DEL estimation?*** can be found. Basically, it depends on the target level of accuracy. For example, to achieve MAPE of less than 7%, both strain gauges and inclinometers should be installed at the tower bottom. However, the number '7%' is only valid for the specific two turbines used in this study, which have a similar first damped natural frequency. Therefore, it should be noted that different accuracy can be obtained if the feed-forward neural network is applied for other turbines which have a quite different first damped natural frequency.

Added to that, the case studies performed to explicitly include dynamic properties; inclusion of peak frequencies and filtered signals. It has been found that there is no improvement in estimation. One possible explanation of the result is that spectral moments already include dynamic property.

In addition to that, to remove the effect from the difference in the dynamic property, inclination, moment and acceleration signals have been normalized with re-

spect to the bandpass filtered signal by multiplying or dividing the filtered signals from the original signals. From the result, it has been concluded that the normalized signals give the completely wrong estimation.

Lastly, wake effect has been included in the form of wake switch and distance. It has been found the explicit inclusion of wake effect is not giving any difference in estimation. This is due to the fact that wake effect not only affects target moment DEL, but also other input features. Therefore, the feed-forward neural network could learn the relationship between input features and target moment DEL even without the explicit inclusion of wake effect.

Chapter 6

Conclusions and future work

This chapter summarizes conclusion and suggested future work. In Section 6.1, conclusions obtained from Chapter 4 and 5 are summarized for individual turbine and farm-wide level load estimation respectively. In Section 6.2, recommended implementation procedures are explained. Lastly, in Section 6.3, recommended future work is described.

6.1 Conclusions

In this study, one-month real measurement data of two turbines in the same wind farm has been used. The measurement data has been categorized into two parts; standard signals and non-standard signals. The standard signals include data from the SCADA system, and the non-standard signals include acceleration and moment signals at the tower top, acceleration, moment and inclination signals at the tower bottom and strain signals at the transition piece, upper and lower monopile levels. In addition to that, wave measurement data has also been used.

Throughout the study, 10 min. statistical properties have been used as inputs. Except for the wave measurement data, the statistical properties include mean, standard deviation, range, the equivalent value and spectral moments. Here, the equivalent value is the value which is calculated in the same way with DEL. For wave measurement data, the statistical properties include significant wave height, peak period and spectral moments.

As a target, 10 min. moment DEL at different levels (The tower bottom, transition piece and two levels at monopile) have been used.

First of all, individual turbine level load estimation has been performed. For individual turbine level load estimation, a neural network is trained on a wind turbine. Then, the trained neural network is applied to the same wind turbine it is trained on.

Firstly, it has been found that linear regression gives an accurate estimation at tower bottom when the turbine is operating in low wind speed or not operating (parked condition) with MAPE of less than 3%. Concretely, accurate estimation has been obtained for linear regression with measurements from the SCADA system and tower top. The signals include standard deviation, the equivalent value of acceleration from the SCADA system, and standard deviation of the tower top moment.

With non-standard signals at tower bottom, linear regression can give quite accurate estimations for the transition piece and upper monopile level with MAPE of less than 5%. Specifically, linear regression with moment DEL at the tower bottom or inclination can give quite accurate estimations for the transition piece and upper monopile level. However, at the lower monopile level, estimation was not so accurate with MAPE of higher than 5%. Furthermore, given that the accuracy level is decreased along with its depth, it is expected that there is higher inaccuracy in estimation at the fatigue critical location near mudline.

In addition, both the feed-forward neural network and recurrent neural network have been applied for the normal operating condition. It has been found that the feed-forward neural network gives a better estimation than the recurrent neural network.

Then, the answer to the research question of ***'Can the neural network technique be applied to accurately estimate moment DEL at already known fatigue critical locations for offshore wind turbines with real measurement data?'*** be found. It is found that the feed-forward neural network can be used and it gives an accurate estimation. Specifically, at the lower monopile level, MAPE of less than 5% can be achieved when either moment or inclination signals at the tower bottom is used together with standard signals and wave measurement data. However, when wave measurement data is not included, the accuracy level is decreased. When both moment and inclination signals are used, MAPE of less than 3% can be achieved.

In addition, the answer to the research question of ***'What is the minimum number of sensors that gives accurate DEL estimation?'*** is also found. Basically, it depends on the level of target accuracy. For example, to achieve MAPE of less than 5% at the lower monopile, either inclinometer or strain gauges need to be installed at tower bottom together with the wave measurement system. Both inclinometer and

strain gauges need to be installed with the wave measurement system to achieve MAPE of 3% at the same location.

As a result of case studies performed to see if any improvement can be achieved, two conclusions can be found. Firstly, Operating condition 1 (normal operating condition) has been divided into two parts based on a wind speed where generator speed and pitch angle start to change. However, there is no improvement in estimation. Secondly, When principal components with 100% accumulated percentage variance from PCA are applied, almost the same accuracy level can be obtained compared to the accuracy level obtained by using all the original features. However, estimation with principal components with an accumulated percentage variance of 99% lower the estimation accuracy from MAPE of around 3% to 5%. From the result, it can be concluded that the principal component with very small percentage variance cannot be ignored in the regression problem.

Then, the answer to the research question of ***'What is the required number of data to train a neural network?*** is found. The required number of training data is dependent on the level of accuracy. When moment and inclination signals at the tower bottom are used together with standard signals and wave measurement data to estimate moment DEL at the lower monopile level, at least 700 data points are required to achieve MAPE of less than 5%. Given that each data point stands for statistical properties of 10 min. time series, it is corresponding to 5 days of measurement period. However, it should be noted that those 700 data points should cover a wide range of input features such as wind speed, generator speed, pitch angle and so on to get an accurate estimation.

Then, farm-wide level load estimation has been performed. For farm-wide level load estimation, a neural network is trained for a wind turbine. Then, the trained neural network is applied to the other wind turbine it is not trained on. Here, it should be noted that the two turbines used in this study have similar first damped natural frequency ranges. Therefore, the results found in this study are very specific for those two turbines.

Firstly, linear regression has been performed for the same operating conditions (when the wind turbine is operating in low wind speed or in parked condition) and with the same input features used in individual turbine level load estimation. As a result, the same conclusions have been obtained. Specifically, it is found that linear regression can give accurate estimation at the tower bottom, transition piece and upper monopile level with MAPE of less than 5%. However, the linear regression gives

a quite inaccurate estimation at the lower monopile level with MAPE of around 8% (when the wind turbine is operating in low wind speed).

Secondly, the answer to the research question of ***'Can the machine learning technique (neural network) be expanded to farm-wide fatigue assessment?'*** is found; the feed-forward neural network can be used for farm-wide level load estimation for the two turbines used in this study. However, it is also found that the feed-forward neural network can be inaccurate when data that the feed-forward neural network is not trained on is fed into (Extrapolation).

Then, as a result of the multiple case studies, the answer to the research question of ***'What is the minimum number of sensors that gives accurate DEL estimation?'*** is found. Similar to individual turbine level load estimation, it depends on the level of target accuracy. For instance, either inclinometer or strain gauges need to be installed at tower bottom with the wave measurement system to achieve MAPE of less than 13% at lower monopile, and both inclinometer and strain gauges need to be installed with the wave measurement system to achieve MAPE of 7% at the same location. Again, it should be noted this conclusion is very specific for the turbines used in this study.

Lastly, multiple case studies have been performed to see if any improvement can be achieved. Specifically, two studies have been performed to explicitly include dynamic property differences. As a first study, peak information of PSD has been additionally fed into the feed-forward neural network. As a second study, signals have been filtered near its first damped natural frequency. Then, the filtered signals have been additionally fed into the feed-forward neural network. However, no improvement can be achieved. The reason for this is that dynamic properties are already included in spectral moments. In addition to that, signals have been normalized with the bandpass filtered signal. Then, the normalized signals have been used as additional features. However, it has been found that the estimation with the normalized signals is completely wrong. Lastly, one case study has been performed to explicitly include wake effect. In this study, wake switch and wake distance have been additionally used as inputs to the feed-forward neural network. However, no improvement can be achieved as well. The reason for this is that the wake effects affect not only the moment DEL but also input features.

6.2 Recommended implementation procedure

From this study, the feed-forward neural network has been proven to accurately estimate the moment DEL. However, to be able to use the feed-forward neural network in practice, some cares should be taken. Accordingly, the recommended implementation procedures are described in this section. In Section 6.2.1, data pre-processing is described. In Section 6.2.2, estimation with the feed-forward neural network is explained.

6.2.1 Data pre-processing

Overall procedure of data pre-processing is shown in Figure 6.1.

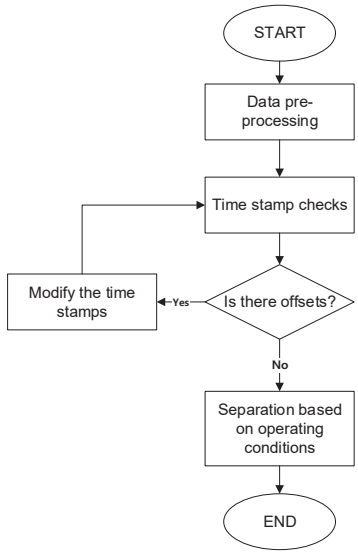


Figure 6.1: Data pre-processing procedure

First of all, it is essential to make sure that time stamps are the same for all the signals. In other words, one needs to make sure that each set of the inputs and target are from the same time stamp. Fortunately, all the measurement signals

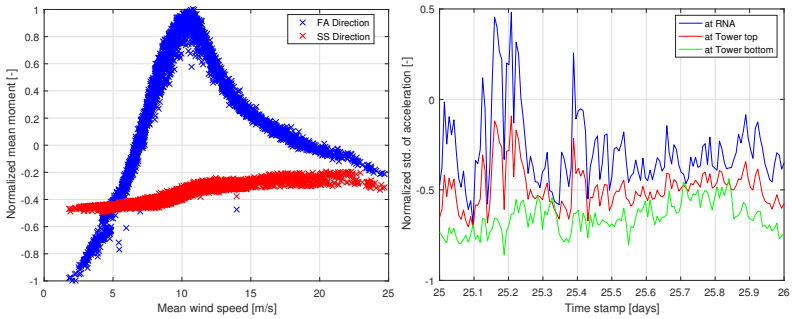
used in this study have the same time stamps. However, it is possible that the time stamps are differently recorded in some of the measurement signals. For example, time stamp 0 in a measurement signal can be corresponding to time stamp 600 in the other measurement signal. In this case, one needs to apply offset so that all the measurement signals are at the same time stamps.

To check if the time stamps are correctly recorded, manual checks should be performed. In this check, all the measurement signals are tuned to be matched with standard signals from the SCADA system. For example, the following manual checks can be performed for moment and acceleration:

Moment signal: Moment signals can be checked by plotting the mean moment versus mean wind speed graph for the normal operating condition. Specifically, FA directional mean moment should follow the thrust curve which has the maximum value at rated wind speed and getting smaller when the mean wind speed is far from the rated wind speed. On the other hand, SS directional mean moment should increase while the mean wind speed increases. The example plot is shown in Figure 6.2a. If there is an offset in time stamps, the graph will not follow the trends. In that case, time stamps in the moment signals should be modified.

Acceleration signal: Acceleration signal can be checked by comparing its standard deviation with that of acceleration at RNA (from the SCADA system) for the normal operating condition. Specifically, the first mode has the biggest contribution when the fixed offshore wind turbine is taken into account. The typical first mode shape is shown as the blue line in Figure 6.3. Accordingly, the standard deviation of acceleration is the highest at the top (RNA) and it is decreasing along with its length from the top. Therefore, the standard deviation of acceleration should be decreasing, but overall changes along with time axis should be the same as shown in Figure 6.2b. If there is offset in time stamps, there will be offset in the graph. In that case, time stamps of acceleration at tower locations should be modified.

Secondly, after all the data preprocessing is done, it is important to make sure that the different operating conditions are separated. For example, the normal operating condition can be separated from the parked condition. It can be easily done by using operating condition information recorded in the SCADA system.



(a) Mean wind speed versus mean moment

(b) Comparison of standard deviation of accelerations

Figure 6.2: Manual checks of pre-processed data

Figure 6.3: Typical first mode shape of a fixed offshore wind turbine

6.2.2 Neural network application

The overall procedure of training of the feed-forward neural network is shown in Figure 6.4.

As shown in the procedure, the feed-forward neural network can be evaluated by comparing the error level. If the error level is small enough so that the target error level is achieved, the trained neural network can be extracted. Specifically, all the

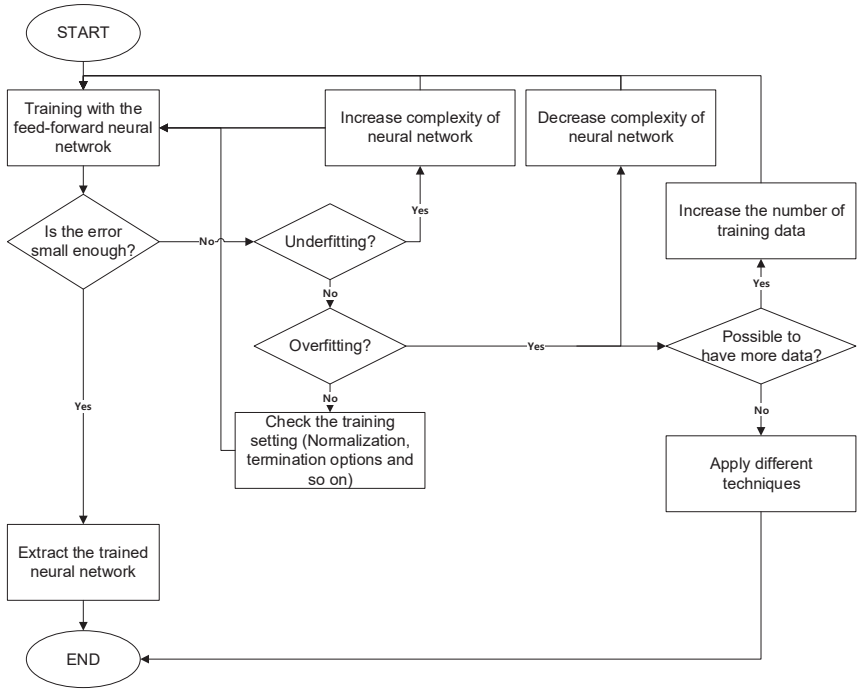


Figure 6.4: Training of the feed-forward neural network procedure

optimized weights and neural network architecture need to be extracted to apply the trained neural network for further use.

However, if the error level is higher than the target level, some actions need to be taken to increase the accuracy level.

For this task, a learning curve should be investigated. The learning curve is explained in Section 3.1.5. According to the learning curve, one can decide whether the feed-forward neural network has an underfitting or overfitting problem. If the feed-forward neural network has an underfitting problem, the complexity of the feed-forward neural network needs to be increased. Specifically, the number of hidden layers or the number of hidden neurons can be increased. If there is an overfitting problem, two actions can be taken; increase the number of training data, and

decrease the complexity of the feed-forward neural network. However, when the decreased complexity gives no improvement and it is not possible to increase the number of training data, different techniques (for example, linear regression) can be considered.

In addition, it is possible that the error is not caused by underfitting or overfitting but caused by something else. For example, if the normalization setting of the training is turned off, the large error can be caused since there is a high possibility that the gradient descent cannot be converged. Therefore, in that case, the training settings should be investigated. From the author's point of view, the following items can affect the performance of the feed-forward neural network the most:

- Normalization (min-max normalization has been used in this study)
- Activation function (Tangent-sigmoid function for hidden layer, pure-linear function for output layer have been used in this study)
- Learning method (Levenberg-marquardt algorithm has been used in this study)

When the trained feed-forward neural network is applied in practice, there are some recommendations.

First of all, extrapolation should be carefully evaluated. As explained in Section 5.2.1, extrapolation can give inaccurate estimation. Ideally, the feed-forward neural network should be trained on all the different cases so that the feed-forward neural network can avoid being applied to the extrapolation problem. However, if the feed-forward neural network is trained on a limited dataset, and it is the only option to estimate the moment DEL, the following methods can be used:

Firstly, to distinguish the extrapolation, ranges of input features and the target can be recorded during the training. Then, in the real application, the input features and the target can be compared with the recorded values. If the trained feed-forward neural network is used for extrapolation, a factor (larger than 1) can be multiplied to the estimated moment DEL to apply conservatism for extrapolation.

Furthermore, the operating condition recorded in the SCADA system is not enough to distinguish all the different conditions. Specifically, as explained in Section 2.2, even though the operating condition is recorded the same as the normal operating condition, power generation can be limited in some cases. In this case, the power curve can be used to distinguish the different conditions. Specifically, the power curve (a plot of mean wind speed versus mean grid power) can be made for the data points used in training. Then, the new data points used in the real application can be added to the power curve. If the new data points are far from the power curve,

the new data points can be recognized as extrapolation.

Secondly, one of the sensors can fail. In that case, the feed-forward neural network which is using the signal from the failed sensor can give inaccurate estimation. To deal with the situation, a back-up feed-forward neural network that is not using the signal from the failed sensor can be prepared. Specifically, two (or more) feed-forward neural networks can be used to estimate moment DEL. For example, one of the feed-forward neural networks is designed to use all the available signals including moment and inclination signals at the tower bottom, while the other (back up) is designed to use all the signals except for the inclination signal at the tower bottom. In that case, even when the inclinometer at the tower bottom is failed, the back-up feed-forward neural network can be used to estimate the moment DEL even though the accuracy level can be lower.

6.3 Future work

6.3.1 Training with larger dataset

As explained in Section 4.8, the one-month data used in this study was not enough to make test set MAPE converged. In addition, for operating condition 3 and parked condition, the number of data was not enough for training as shown in Section 4.4.2 and linear regression has been applied instead of neural network. However, at the lower monopile level in operating condition 3, linear regression has given a somewhat inaccurate estimation, but it is expected that the neural network can give accurate estimation if the number of data is enough for operating condition 3.

Therefore, it is recommended to apply the neural network with more than one-month data that covers various operating condition and broader ranges of features (i.e. wind direction, wind speed and so on). From the study, it can be investigated if the neural network gives an accurate estimation for all the different operating conditions. In addition, it can also be investigated how much data (how long the measurement period) is required for error convergence.

For farm-wide level load estimation, it has been found that the feed-forward neural network can give inaccurate estimation if it is applied to extrapolation. Specifically, two turbines had the different operating conditions in which one turbine generates limited power even when the wind speed is high enough, and the other turbine generates power along with its power curve. However, it has not been investigated if the feed-forward neural network can make an accurate estimation for this operating

condition.

Therefore, it is recommended to study further if the neural network can capture that different behavior if the feed-forward neural network is trained with large enough training data that cover all those cases.

6.3.2 Application of different machine learning techniques

Throughout this study, the feed-forward neural network and recurrent neural network (LSTM) have been applied with 10 min. statistics. As a result, it has been found that feed-forward neural network outperforms recurrent neural network for moment DEL estimation.

However, for the training of recurrent neural network (LSTM), only 1665 and 439 data points have been applied for turbine 1 and 2 respectively as shown in Section 4.5. Added to that, the recurrent neural network has more complicated architecture than the feed-forward neural network which indicates more number of data might be necessary for training.

Therefore, it is recommended to train the recurrent neural network (LSTM) with more number of data.

Additionally, it is also recommended to apply other recurrent neural networks. Specifically, gated recurrent unit (GRU) can be applied.

In addition, instead of 10 min. statistics, time series itself can be estimated with neural networks. If time series itself can be accurately estimated with neural networks, not only the fatigue load (i.e. moment DEL), but also the extreme load can be estimated.

Especially, a higher sequential dependency is expected when the time series itself is considered. Therefore, it is recommended to apply both the feed-forward neural network and recurrent neural network.

6.3.3 Training with simulation

As explained in Section 5.2.1, if the data which a neural network is not trained for is applied, the result can be inaccurate.

One way to train the neural network is to use larger dataset as explained above. However, it takes a lot of time for the measurement to cover all the different cases.

Therefore, it is recommended to train a neural network with simulation data. Specifically, multiple simulations with different water depth and different dynamic

properties can be performed. Then, the results of the simulations can be used to train the neural network.

In addition, more complicated feed-forward neural network architecture needs to be investigated. The reason for this is that a more complicated relationship is expected when all the different dynamic properties and water depths are used as input.

However, it should be noted that the real turbine can be different from the simulation models. For example, soil condition is frequently different between the simulation model and the real turbine. In addition, the actual mode shape can also be different between the simulation model and the real turbine. In that case, the neural network trained with simulation models can give inaccurate estimation.

To deal with the problem, transfer learning can be applied.

Specifically, the following procedure can be applied; Firstly, pre-train the feed-forward neural network with simulation data. It will be called pre-trained feed-forward neural network. Secondly, extract the weights of the pre-trained feed-forward neural network. Then, set the extracted weights as a starting point of further training of the feed-forward neural network. Lastly, tune the pre-trained feed-forward neural network with the real measurement data.

With the transfer learning, it is expected that the feed-forward neural network can give accurate estimation for the real turbine even though the simulation model is quite different from the real turbine since it will be tuned with real measurement data.

Furthermore, since the pre-trained feed-forward neural network will be trained to cover all the different cases with simulation data, it is expected that the feed-forward neural network will give quite accurate estimation for extrapolation problems as well. Concretely, even though the real measurement data used for the tuning cannot cover some of the conditions, the pre-trained feed-forward neural network already know how to deal with the conditions since the simulation data used in the pre-training already cover the conditions as well.

6.3.4 Farm-wide level load estimation with more than two turbines

For farm-wide level load estimation, it has been found the two turbines used in this study have a similar range of first damped natural frequency. However, it is expected that the estimation can be inaccurate if the feed-forward neural network is applied to a wind turbine which has a quite different first damped natural frequency.

Therefore, it is recommended to apply the feed-forward neural network in farm-wide level load estimation more than two turbines so that the effect of the different dynamic property can be clearly investigated.

6.3.5 Explicit wake effect inclusion in farm-wide level load estimation

In Section 5.6.4, the explicit inclusion of wake effect as a form of 'wake switch' and 'wake distance' has not given a significant increase in estimation accuracy. However, as briefly mentioned in Section 2.4.4, the strength of wake turbulence is affected by multiple sources. For example, the number of upstream turbines and operating conditions of upstream turbines can additionally affect the strength of wake turbulence. Therefore, it is recommended to perform the farm-wide level load estimation including those aspects in addition to the 'wake switch' and 'wake distance'.

Specifically, when the upstream turbine is in normal operating condition, wake turbulence would be stronger compared to the case where the upstream turbine is in parked condition. To include those effects, additional features that designate the operating status of the upstream turbine can be added. Similar to the 'wake switch', the feature can be set as 1 when the upstream turbine is in normal operation, and 0 when it is in parked condition so that the feed-forward neural network can learn the different relationships.

Appendix A

Detailed procedure of data preprocessing

First of all, time series of acceleration signal in SCADA system has varying sampling frequencies. However, it should have fixed sampling frequency to calculate spectral moments. Accordingly, the time series of acceleration signal in SCADA system has been pre-processed and it is explained in Section A.2. In addition, Wind turbine shows different behavior for fore-aft (FA) and side-by-side (SS) direction. The main cause of this difference is large thrust force which is acting on FA direction. In addition, yaw angle keeps changing according to wind direction so that the wind turbine generate the maximum energy. In other words, FA and SS directions keeps changing. Therefore, the moment and acceleration signals need to be converted into FA and SS directions. It is described in Section A.3.

In addition, strain signals from strain gauges at transition piece and monopile need to be converted into moment. The moment conversion is explained in Section A.4. As briefly explained in Section 2.1.2, the strain signals at transition piece and monopile have not been calibrated, and there is no information about the calibration. As a result the non-calibration, when the strain signals are converted into FA and SS directional moment, large error is caused. Accordingly, for transition piece and monopile, M_{WE} and M_{SN} have been used for estimation without being converted into FA and SS direction.

In Section A.5 and A.6, handling of measurement errors and damage equivalent load calculation are explained respectively.

Lastly, since all the sensors record time series, it needs to be pre-processed to

calculate statistical properties and it is described in Section A.7.

A.1 Pre-processing of wave data

As mentioned above, wave statistics have been recorded at different time lengths. Specifically, sometimes for 2 min. and sometimes for 3 min.. Therefore, the wave data should be pre-processed into 10 min. basis. To calculate the 10 min. statistics from the 2 or 3 min. data, wave data have been grouped into 10 min.. Then, the mean of all the statistical properties has been calculated as shown in Figure A.1.

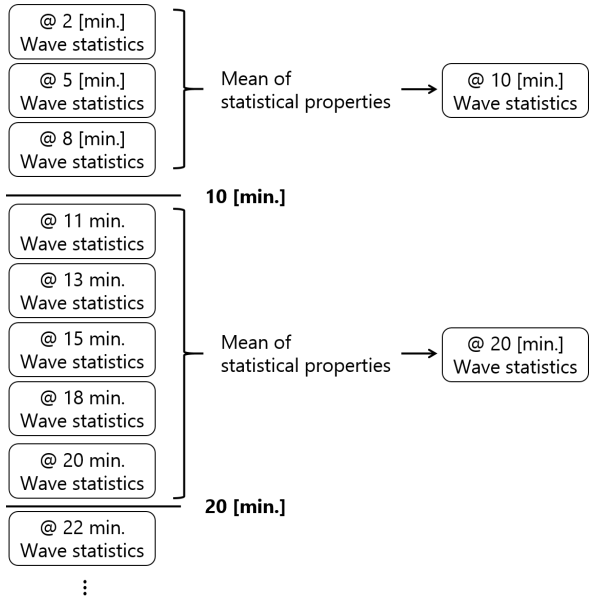


Figure A.1: Wave pre-processing

A.2 Time series of acceleration in SCADA system

The time series of acceleration signal in SCADA system has varying sampling frequencies. However, as shown in Section A.7, spectral moments have been used for

estimation. To calculate spectral moments, Fourier transform has been used and the fixed sampling frequency is required to apply the Fourier transform. Accordingly, the re-sampling of the acceleration signal in SCADA system has been performed by interpolating the original signal with the re-sampling time series. Since all the other continuous time series have fixed sampling frequency of 25 Hz (0.04 sec.), The re-sampling of the acceleration in SCADA system has also been performed based on the fixed sampling frequency of 25 Hz. A simple example of re-sampling is shown in Figure A.2 in which the red and green line is the original and re-sampled signal respectively.

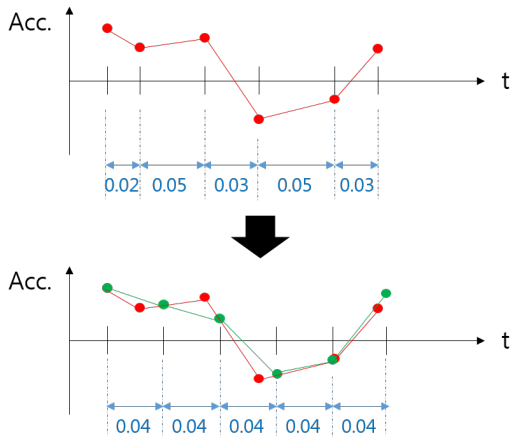


Figure A.2: Example of re-sampling

A.3 Conversion into FA & SS directional components

As explained in Section 2.1.2, moment and acceleration measurement data have different measurement coordinate system. In addition, FA and SS direction keep changing as its yaw direction changed.

Accordingly, yaw direction at each time step has been used to obtain FA & SS directional coordinate system. Then, moment and acceleration components which are corresponding to the pre-calculated FA & SS directional coordinate system have been calculated to obtain FA & SS directional moment and acceleration.

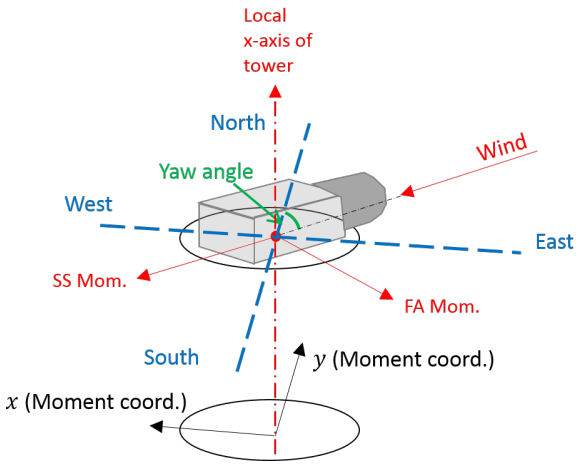
The yaw direction signal is recorded by setting 0 deg. yaw angle as when the

wind is coming from North and positive to the clock-wise direction. For example, when the wind is coming from the East, then the yaw angle is recorded as 90 deg.. By combining the yaw direction system with the moment and acceleration coordinate system shown in Figure 2.3 and 2.4, FA and SS directional moment and acceleration can easily be drawn as depicted in Figure A.3.

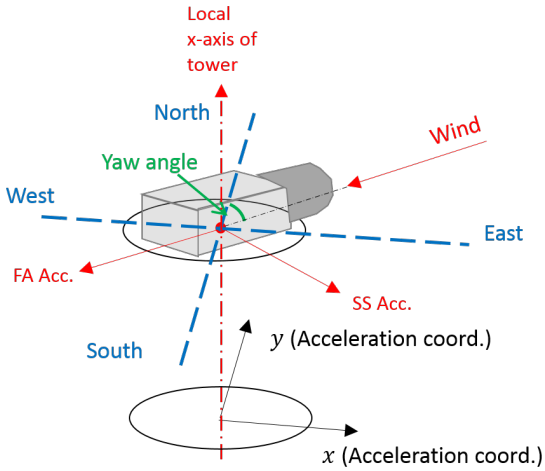
According to Figure A.3, FA & SS moment and acceleration at each time step can be calculated by Equation (A.1) and (A.2) where θ_{offset} is an offset angle which is 0 deg. and 23 deg. for moment calculation of turbine 1 and 2 respectively.

$$\begin{aligned} M_{FA} &= -M_x \cos\left(\theta_{yaw} - \theta_{offset}\right) - M_y \sin\left(\theta_{yaw} - \theta_{offset}\right) \\ M_{SS} &= M_x \sin\left(\theta_{yaw} - \theta_{offset}\right) - M_y \cos\left(\theta_{yaw} - \theta_{offset}\right) \end{aligned} \tag{A.1}$$

$$\begin{aligned} Acc_{FA} &= -Acc_x \sin\left(\theta_{yaw}\right) - Acc_y \cos\left(\theta_{yaw}\right) \\ Acc_{SS} &= Acc_x \cos\left(\theta_{yaw}\right) - Acc_y \sin\left(\theta_{yaw}\right) \end{aligned} \tag{A.2}$$



(a) Moment coord. with yaw coord.



(b) Acceleration coord. with yaw coord.

Figure A.3: Moment and acceleration coord. with yaw coord.

A.4 Conversion of strain into moment

Strain gauges are installed as shown in Figure 2.1. As described above, each strain gauge measures axial strain. As a result, to calculate moment from strain, the relationship shown in Equation (A.3) needs to be used. The relationship has been derived by using strain-forces relationship shown in Appendix B.

$$\begin{bmatrix} \epsilon_E \\ \epsilon_N \\ \epsilon_W \end{bmatrix} = \frac{1}{E} \cdot \begin{bmatrix} 1/A & -x_u/I & 0 \\ 1/A & 0 & -y_u/I \\ 1/A & x_u/I & 0 \end{bmatrix} \cdot \begin{bmatrix} P_{Axial} \\ M_{SN} \\ M_{WE} \end{bmatrix} \tag{A.3}$$

Where A, I, and E are area, second moment of inertia and young's modulus respectively. x_u, y_u are distance between neutral axis and sensors as shown in Figure A.4. Lastly, P_{Axial}, M_{SN} and M_{WE} are axial force, moment along the South-North axis and moment along the West-East axis respectively.

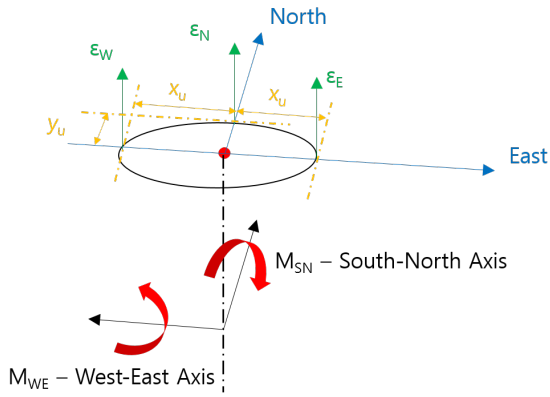
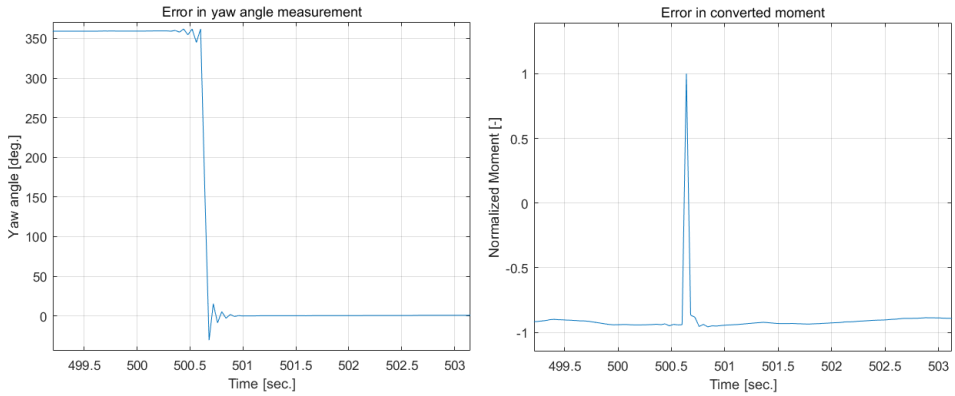


Figure A.4: Strain to moment conversion

A.5 Measurement error

In a real measurement system, there can be an unexpected error for several reasons. The causes of the measurement error include input signal error, sensor error, signal transmission error, transducer error, converter error, and computer error. It is found that the measured signals used in this study also include some errors.

Firstly, there is an error in yaw angle measurement in SCADA system. Specifically, when the yaw angle is increased from below 0 deg. (corresponding to below 360 deg.) to above 0 deg., the measured signal drops to around -10 deg. first and then stabilized afterward. In addition, the opposite case in which yaw angle changes from above 0 deg. to below 0 deg. exists as well. In that case, the yaw angle increases to around 10 deg. and then stabilized. One of the examples is shown in Figure A.5a. Accordingly, the M_{FA} and M_{SS} moment converted by Equation (A.1) has abrupt peak as shown in Figure A.5b. It causes unrealistically high moment DEL (Damage Equivalent Load) since the abrupt peak will be counted by the cycle counting method. Accordingly, the abrupt peak in the converted moment is disregarded in moment DEL calculation so that reasonable moment DEL can be obtained. Specifically, if the ratio between the absolute difference of signals in between two consecutive time step and standard deviation of the absolute differences of the whole 10 min. time series was higher than 30, the data point has been removed from the time series. One example of this modification is shown in Figure A.6 which is corresponding to the moment time series in Figure A.5b after modification.



(a) Yaw angle error **(b)** Converted moment error

Figure A.5: Example of error in yaw angle measurement and converted moment

The second measurement error found in this study is moment calibration error. At top and bottom of the tower, strain signals have been calibrated into M_x and M_y at each level. In the calibration process, an error occurred and there are some abrupt peaks in M_x and M_y signals which are similar to Figure A.5b. For this type of measurement error, the same approach described above in yaw angle error modification

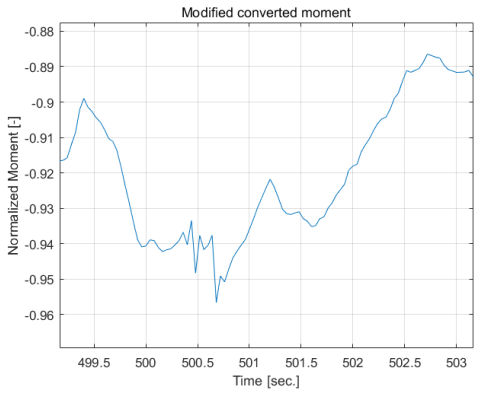


Figure A.6: Example of modified moment

has been taken to remove the abrupt peaks in M_x and M_y signals.

A.6 Damage equivalent load

The definition of damage equivalent load is that "the load that gives same fatigue damage for a given equivalent number of cycles N_{eq}". For fatigue damage calculation, the most widely used approach is to use "S-N curve". Specifically, stress range-number of cycles histogram can be made by counting the number of cycles in stress range histogram. In this study, rainflow counting has been applied to count the number of cycles. The example stress range-number of cycles histogram is shown in Figure A.7. In the figure, the total number of cycles is 1090 including 7 from residual half-cycles (meaning total 14 residual half-cycles = 7 full cycles).

With the stress range-number of cycles histogram, the damage equivalent load can be calculated as shown in Equation (A.4). (Derivation can be found in Appendix C)

$$M_{eq} = \Big[\sum_i \frac{n_i}{N_{eq}} M_i^m\Big]^{1/m} \tag{A.4}$$

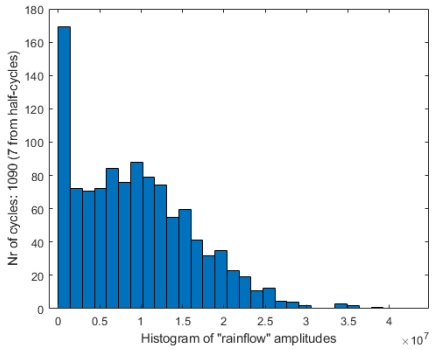


Figure A.7: Example stress range - number of cycles histogram

A.7 Statistical properties

From the time series of signals, statistical properties can be calculated. In this study, the statistical properties shown in Table A.1 have been calculated.

Table A.1: Statistical properties

Number	Statistical Properties	Symbol	Remark
1	Mean	μ	-
2	Standard deviation	σ_{std}	-
3	Max. Range	-	-
4	Equivalent value	-	Calculated in the same way of DEL
5	Spectral moment	λ_n	n = -3, -2, -1, 0, 1, 2, 3

Mean and standard deviation can be calculated by Equation (A.5) and (A.6) where N is the total number of data and x_i is i th data in a dataset.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{A.5}$$

$$\sigma_{std} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{A.6}$$

The maximum range is simply the difference between maximum and minimum values in a dataset.

The equivalent value is a value which is calculated in the same way with DEL. Specifically, count the cycles of all the signals, and obtain S_i and n_i where S_i is i th range of a signal, n_i is the corresponding cycles. Then, the equivalent value S_{eq} can be calculated by putting the S_i instead of M_i, and n_i in Equation (A.4). In this study, slope m of 3.5 and the equivalent number of cycles N_{eq} of 10^7 have been used.

Lastly, spectral moment λ has been calculated from PSD (power spectral density) by using Equation (A.7) where f_i is i th frequency and $S(f_i)$ is PSD corresponding to f_i.

$$\lambda_n = \sum_i f_i^n S(f_i) \tag{A.7}$$

Appendix B

Derivation of strain-forces relationship

For a beam shown in Figure B.1, θ_y is rotation along with y axis, w_z is displacement along with z axis, and u_0 is axial displacement along with x axis.

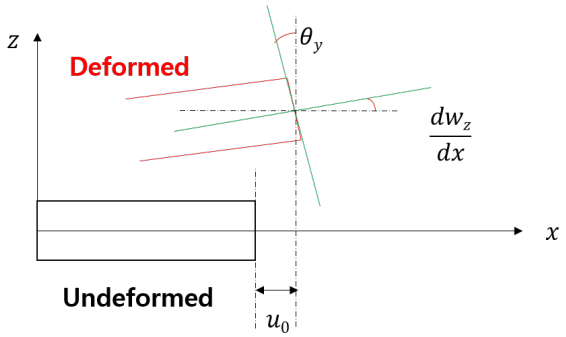


Figure B.1: Beam deformation

First of all, axial displacement in beam can be calculated by Equation (B.1).

$$u = u_0 - z\theta_y \tag{B.1}$$

Where u is total axial displacement located at z from neutral axis, z is the distance from neutral axis.

Secondly, strain can be calculated from Equation (B.1) by taking x directional derivative as shown in Equation (B.2).

$$\epsilon_{xx} = \frac{du}{dx} = \frac{du_0}{dx} - z\frac{d\theta_y}{dx} \tag{B.2}$$

Where ϵ_{xx} is total axial strain.

Lastly, by applying Hooke's law, stress-forces relationship can be obtained as shown in Equation (B.3).

$$\sigma_{xx} = E\epsilon_{xx} = E\frac{du_0}{dx} - Ez\frac{d\theta_y}{dx} \tag{B.3}$$

The first and second terms can be expressed by axial force P and moment M.

For the first term, it can easily be expressed with axial force P as shown in Equation (B.4).

$$E\frac{du_0}{dx} = E\epsilon_{xx,0} = \frac{P}{A} \tag{B.4}$$

Where E is young's modulus, $\epsilon_{xx,0}$ is strain induced by axial force, and A is cross sectional area.

Next, the second term can be expressed with respect to moment M by using moment-curvature relationship as shown in Equation (B.5).

$$M_y = EI_{yy}\frac{d\theta_y}{dx} \tag{B.5}$$

Where M_y is moment along with y axis and I_{yy} is second moment of inertia of cross section along with y axis.

Then, the second term of Equation (B.3) can be expressed as shown in Equation (B.6).

$$Ez\frac{d\theta_y}{dx} = Ez\frac{M_y}{EI_{yy}} = \frac{M_y z}{I_{yy}} \tag{B.6}$$

As a result, the following strain-forces relationship can be obtained:

$$\epsilon_{xx} = \frac{P}{EA} - \frac{M_y z}{EI_{yy}} \tag{B.7}$$

Appendix C

Derivation of damage equivalent load

By using Palmgren-Miner rule, the fatigue damage can be calculated as shown in Equation (C.1).

$$D = \sum_{i=1} \frac{n_i}{N_i} \tag{C.1}$$

Where N_i is the total number of cycles at a given stress range $\Delta\sigma_i$ which can be calculated with "S-N curve" and n_i is the actual number of cycle for a given stress range $\Delta\sigma_i$.

In addition to that, the relationship shown in Equation (C.2) holds in "S-N curve" approach.

$$\frac{N_i}{N_{ref}} = \left(\frac{\Delta\sigma_{ref}}{\Delta\sigma_i}\right)^m \tag{C.2}$$

Where N_{ref} and $\Delta\sigma_{ref}$ are reference number of cycles and reference stress range represents a given "S-N curve". In addition, m is the slope of a given "S-N curve". Then, by combining Equation (C.1) and (C.2), the damage in Equation (C.1) can be rewritten as shown in Equation (C.3).

$$D = \sum_i \frac{n_i}{N_{ref}} \left(\frac{\Delta\sigma_i}{\Delta\sigma_{ref}}\right)^m \tag{C.3}$$

Lastly, for a given equivalent number of cycles N_{eq}, damage can be expressed simply as a ratio between the equivalent number of cycles N_{eq} and total number of cycles at the equivalent stress range $\Delta\sigma_{eq}$ as shown in Equation (C.4).

$$D_{eq} = \frac{N_{eq}}{N(\Delta\sigma_{eq})} \tag{C.4}$$

Consequently, the damage equivalent load can be derived as shown in Equation (C.5) by equating damage D in Equation (C.3) and equivalent damage D_{eq} in Equation (C.4).

$$\Delta\sigma_{eq} = \Big[\sum_i \frac{n_i}{N_{eq}} \Delta\sigma_i^m\Big]^{1/m} \tag{C.5}$$

Especially, by using the relationship in between stress - moment using beam theory, the moment equivalent load can also simply be calculated in the same way as shown in Equation (C.6)

$$M_{eq} = \Big[\sum_i \frac{n_i}{N_{eq}} M_i^m\Big]^{1/m} \tag{C.6}$$

Appendix D

Moment DEL at tower top vs moment DEL at tower bottom (SS Direction)

Standard deviation of acceleration vs moment DEL at tower bottom, SS direction :

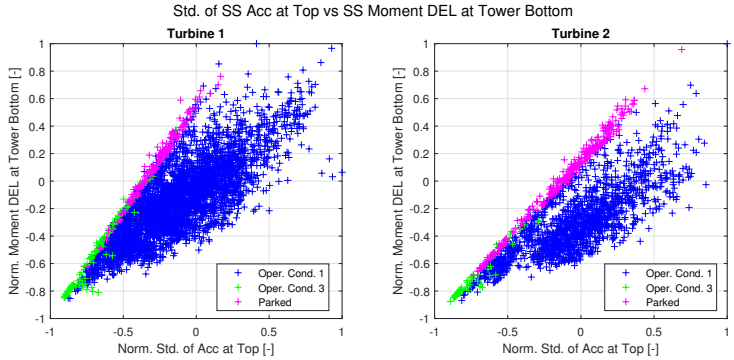


Figure D.1: Std. of Acc. at tower top vs Moment DEL at tower bottom (SS Direction)

Moment DEL at tower top vs moment DEL at tower bottom, SS direction :

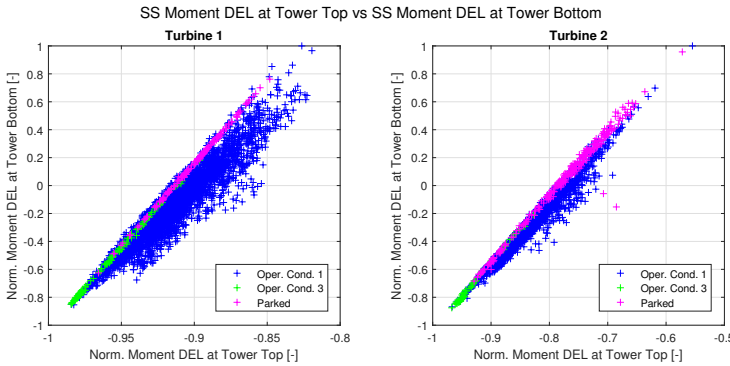


Figure D.2: Moment DEL at tower top vs Moment DEL at tower bottom (SS Direction)

Appendix E

Pearson's correlation coefficient for transition piece and monopile

Transition piece moment DEL, Operating Condition 1

Table E.1: Pearson's correlation coefficient, w.r.t. Transition piece moment DEL, Operating Condition 1

Rank	Turbine 1			
	X (Transition piece)		Y (Transition piece)	
	R value	Features	R value	Features
1	0.998	Tower Bottom Mom X Eq. value	0.998	Tower Bottom Mom Y Eq. value
2	0.953	Tower Bottom Mom X λ_1	0.940	Tower Bottom Mom Y λ_1
3	0.898	RNA Accel Y σ_{std}	0.910	RNA Accel X σ_{std}
4	0.892	Tower Top Mom X Eq. value	0.889	RNA Accel X Eq. value
5	0.882	Tower Top Accel Y σ_{std}	0.888	Tower Bottom Mom Y λ_2
6	0.877	RNA Accel Y λ_0	0.879	RNA Accel X λ_0
7	0.875	RNA Accel Y λ_{-1}	0.878	Tower Top Mom Y Eq. value
8	0.873	Tower Bottom Mom X λ_2	0.875	RNA Accel X λ_{-1}
9	0.871	RNA Accel Y Eq. value	0.869	RNA Accel X Range
10	0.869	RNA Accel Y Range	0.868	Tower Top Accel X σ_{std}
11	0.865	Tower Top Accel Y λ_0	0.856	RNA Accel X Max
12	0.865	Tower Top Accel Y Var	0.846	Tower Top Accel X λ_0
13	0.861	RNA Accel Y Max	0.819	RNA Accel X λ_1
14	0.849	Tower Bottom Mom X Range	0.811	Tower Top Mom Y λ_1
15	0.835	Tower Accel Y Top Range	0.810	Tower Top Accel X Range
Rank	Turbine 2			
	X (Transition piece)		Y (Transition piece)	
	R value	Features	R value	Features
1	0.998	Tower Bottom Mom X Eq. value	0.998	Tower Bottom Mom Y Eq. value
2	0.952	Tower Bottom Mom X λ_1	0.948	Tower Bottom Mom Y λ_1
3	0.917	RNA Accel Y σ_{std}	0.876	Tower Bottom Mom Y λ_2
4	0.913	RNA Accel Y λ_{-1}	0.847	RNA Accel X σ_{std}
5	0.905	RNA Accel Y λ_0	0.822	RNA Accel X λ_0
6	0.880	Tower Bottom Mom X λ_2	0.820	RNA Accel FA λ_0
7	0.836	RNA Accel Y Eq. value	0.819	RNA Accel FA Eq. value
8	0.822	RNA Accel Y Range	0.813	RNA Accel X λ_{-1}
9	0.813	RNA Accel Y Max	0.809	Tower Top Accel X σ_{std}
10	0.778	Tower Bottom Mom X Range	0.804	RNA Accel X Eq. value
11	0.773	Tower Top Accel Y σ_{std}	0.788	RNA Accel X Range
12	0.766	Tower Top Accel Y λ_0	0.782	Tower Top Accel X λ_0
13	0.765	RNA Accel X λ_{-1}	0.779	Tower Top Accel FA σ_{std}
14	0.757	Tower Top Mom X Eq. value	0.776	RNA Accel Y Eq. value
15	0.727	RNA Accel Y λ_1	0.776	RNA Accel X Max

Transition piece moment DEL, Operating Condition 3

Table E.2: Pearson's correlation coefficient, w.r.t. Transition piece moment DEL, Operating Condition 3

Rank	Turbine 1			
	X (Transition piece)		Y (Transition piece)	
	R value	Features	R value	Features
1	0.998	Tower Bottom Mom X Eq. value	0.999	Tower Bottom Mom Y Eq. value
2	0.995	Tower Bottom Incli. Y σ_{std}	0.988	Tower Top Accel X σ_{std}
3	0.994	Tower Bottom Incli. X Back σ_{std}	0.987	Tower Bottom Incli. Y Back σ_{std}
4	0.986	Tower Bottom Accel Y σ_{std}	0.987	Tower Bottom Incli. X σ_{std}
5	0.983	Tower Accel Y Top σ_{std}	0.953	RNA Accel X σ_{std}
6	0.968	RNA Accel Y σ_{std}	0.953	RNA Accel X Eq. value
7	0.964	RNA Accel Y Eq. value	0.952	Tower Bottom Accel X σ_{std}
8	0.960	Tower Bottom Incli. Y Eq. value	0.936	Tower Top Accel X λ_0
9	0.954	Tower Bottom Incli. Y λ_0	0.936	RNA Accel Y Eq. value
10	0.954	Tower Bottom Incli. X Back λ_0	0.936	RNA Accel Y σ_{std}
11	0.953	Tower Top Accel Y λ_0	0.929	Tower Bottom Mom Y λ_1
12	0.951	Tower Bottom Incli. X Back Eq. value	0.922	RNA Accel X λ_1
13	0.948	RNA Accel X σ_{std}	0.920	RNA Accel X λ_0
14	0.945	Tower Bottom Mom X λ_1	0.920	Tower Bottom Incli. Y Back λ_0
15	0.944	Tower Bottom Accel Y λ_0	0.920	RNA Accel X λ_{-1}
Rank	Turbine 2			
	X (Transition piece)		Y (Transition piece)	
	R value	Features	R value	Features
1	0.987	Tower Bottom Mom X Eq. value	0.994	Tower Bottom Mom Y Eq. value
2	0.975	RNA Accel X λ_1	0.992	Tower Top Accel X σ_{std}
3	0.972	RNA Accel X λ_{-1}	0.985	Tower Bottom Back Incli. Y σ_{std}
4	0.971	Tower Bottom Incli. Y λ_0	0.984	Tower Bottom Incli. X σ_{std}
5	0.971	Tower Bottom Incli. Y σ_{std}	0.984	Tower Bottom Mom Y λ_1
6	0.970	Tower Bottom Back Incli. X λ_0	0.983	RNA Accel Y σ_{std}
7	0.970	RNA Accel X λ_0	0.983	Tower Top Accel X λ_0
8	0.969	Tower Bottom Back Incli. X σ_{std}	0.977	RNA Accel Y Eq. value
9	0.968	RNA Accel X σ_{std}	0.976	RNA Accel Y λ_0
10	0.964	RNA Accel X Eq. value	0.974	RNA Accel Y λ_{-1}
11	0.937	Tower Bottom Accel Y λ_0	0.973	RNA Accel Y λ_1
12	0.932	Tower Bottom Accel Y σ_{std}	0.963	Tower Bottom Back Incli. Y λ_0
13	0.925	Tower Bottom Back Incli. Y λ_0	0.961	Tower Bottom Incli. X λ_0
14	0.925	Tower Bottom Incli. X λ_0	0.929	Tower Bottom Accel X σ_{std}
15	0.923	RNA Accel Y λ_{-1}	0.915	RNA Accel Y Max

Transition piece moment DEL, Parked Condition

Table E.3: Pearson's correlation coefficient, w.r.t. Transition piece moment DEL, Parked Condition

Rank	Turbine 1			
	X (Transition piece)		Y (Transition piece)	
	R value	Features	R value	Features
1	0.997	Tower Bottom Mom X Eq. value	0.998	Tower Bottom Mom Y Eq. value
2	0.993	Tower Accel Y Top σ_{std}	0.992	Tower Top Accel X σ_{std}
3	0.976	Tower Accel Y Top λ_0	0.989	Tower Top Accel X Eq. value
4	0.974	Tower Bottom Mom X λ_1	0.971	Tower Top Accel X λ_0
5	0.968	Tower Top Accel Y Eq. value	0.971	RNA Accel Y Eq. value
6	0.949	Tower Top Accel Y Range	0.969	Tower Bottom Mom Y λ_1
7	0.949	Tower Top Accel Y Max	0.967	RNA Accel Y σ_{std}
8	0.946	Tower Bottom Incli. Y σ_{std}	0.953	Tower Top Accel X Range
9	0.946	Tower Bottom Incli. Y λ_0	0.950	Tower Bottom Incli. X σ_{std}
10	0.945	Tower Bottom Incli. X Back λ_0	0.950	Tower Bottom Incli. Y Back σ_{std}
11	0.945	Tower Bottom Incli. X Back σ_{std}	0.946	Tower Top Accel X Max
12	0.914	Tower Bottom Mom X λ_2	0.945	Tower Top Accel X λ_1
13	0.910	Tower Bottom Accel Y λ_0	0.944	RNA Accel Y Range
14	0.909	Tower Bottom Accel Y σ_{std}	0.943	RNA Accel Y Max
15	0.899	RNA Accel Y Eq. value	0.937	RNA Accel Y λ_1

Rank	Turbine 2			
	X (Transition piece)		Y (Transition piece)	
	R value	Features	R value	Features
1	1.000	Tower Bottom Mom X Eq. value	1.000	Tower Bottom Mom Y Eq. value
2	0.996	Tower Top Mom X Eq. value	0.995	Tower Top Accel X σ_{std}
3	0.996	Tower Bottom Mom X σ_{std}	0.994	Tower Top Accel X Eq. value
4	0.995	Tower Top Accel Y σ_{std}	0.979	Tower Bottom Incli. X σ_{std}
5	0.987	Tower Bottom Incli. Y σ_{std}	0.979	Tower Bottom Back Incli. Y σ_{std}
6	0.987	Tower Bottom Back Incli. X σ_{std}	0.970	Tower Bottom Accel X σ_{std}
7	0.986	Tower Top Accel Y Eq. value	0.968	Tower Top Accel X λ_0
8	0.981	Tower Top Mom X σ_{std}	0.967	Tower Bottom Mom Y λ_1
9	0.980	Tower Bottom Accel Y σ_{std}	0.963	Tower Bottom Mom Y λ_2
10	0.976	Tower Bottom Incli. Y λ_0	0.958	Tower Top Accel X λ_1
11	0.976	Tower Bottom Back Incli. X Var	0.957	RNA Accel Y Eq. value
12	0.976	Tower Bottom Back Incli. X λ_0	0.955	Tower Top Mom Y Eq. value
13	0.974	Tower Top Mom X λ_2	0.954	Tower Bottom Back Incli. Y λ_0
14	0.973	Tower Bottom Back Incli. X λ_{-1}	0.953	Tower Bottom Incli. X λ_0
15	0.973	Tower Bottom Incli. Y λ_{-1}	0.952	RNA Accel Y σ_{std}

Monopile (Upper level) moment DEL, Operating Condition 1

Table E.4: Pearson's correlation coefficient, w.r.t. Monopile (Upper level) moment DEL, Operating Condition 1

Rank	Turbine 1			
	X (Monopile (Upper level))		Y (Monopile (Upper level))	
	R value	Features	R value	Features
1	0.979	Tower Bottom Mom X Eq. value	0.985	Tower Bottom Mom Y Eq. value
2	0.946	Tower Bottom Mom X λ_1	0.941	Tower Bottom Mom Y λ_1
3	0.881	Tower Top Accel Y σ_{std}	0.922	RNA Accel X σ_{std}
4	0.875	RNA Accel Y σ_{std}	0.905	RNA Accel X Eq. value
5	0.869	Tower Top Mom X Eq. value	0.895	Tower Top Mom Y Eq. value
6	0.866	Tower Top Accel Y λ_0	0.893	RNA Accel X λ_0
7	0.856	RNA Accel Y λ_{-1}	0.889	RNA Accel X λ_{-1}
8	0.853	RNA Accel Y λ_0	0.889	Tower Bottom Mom Y λ_2
9	0.851	RNA Accel Y Eq. value	0.883	Tower Top Accel X σ_{std}
10	0.850	Tower Bottom Mom X λ_2	0.883	RNA Accel X Range
11	0.847	RNA Accel Y Range	0.868	RNA Accel X Max
12	0.840	RNA Accel Y Max	0.863	Tower Top Accel X λ_0
13	0.840	Tower Top Accel Y Max	0.843	Tower Top Mom Y λ_1
14	0.835	Tower Top Accel Y Range	0.838	RNA Accel X λ_1
15	0.833	RNA Accel X σ_{std}	0.828	Tower Top Accel X Range
Rank	Turbine 2			
	X (Monopile (Upper level))		Y (Monopile (Upper level))	
	R value	Features	R value	Features
1	0.990	Tower Bottom Mom X Eq. value	0.985	Tower Bottom Mom Y Eq. value
2	0.940	Tower Bottom Mom X λ_1	0.957	Tower Bottom Mom Y λ_1
3	0.936	RNA Accel Y σ_{std}	0.899	Tower Bottom Mom Y λ_2
4	0.915	RNA Accel Y λ_0	0.881	RNA Accel X σ_{std}
5	0.912	RNA Accel Y λ_{-1}	0.863	RNA Accel X λ_0
6	0.884	Tower Bottom Mom X λ_2	0.861	Tower Top Accel X σ_{std}
7	0.867	RNA Accel Y Eq. value	0.852	RNA Accel X Eq. value
8	0.849	MP Mom X MP16 λ_2	0.841	Tower Top Accel X λ_0
9	0.848	RNA Accel Y Range	0.838	RNA Accel X λ_{-1}
10	0.838	RNA Accel Y Max	0.830	RNA Accel X Range
11	0.808	Tower Top Accel Y σ_{std}	0.829	RNA Accel Y Eq. value
12	0.794	Tower Top Accel Y λ_0	0.816	RNA Accel X Max
13	0.788	RNA Accel X λ_{-1}	0.812	RNA Accel Y σ_{std}
14	0.773	Tower Top Mom X Eq. value	0.792	RNA Accel Y Range
15	0.762	Tower Bottom Mom X Range	0.786	RNA Accel X λ_1

Monopile (Upper level) moment DEL, Operating Condition 3

Table E.5: Pearson's correlation coefficient, w.r.t. Monopile (Upper level) moment DEL, Operating Condition 3

Rank	Turbine 1			
	X (Monopile (Upper level))		Y (Monopile (Upper level))	
	R value	Features	R value	Features
1	0.996	Tower Bottom Incli. Y σ_{std}	0.996	Tower Bottom Mom Y Eq. value
2	0.996	Tower Bottom Back Incli. X σ_{std}	0.988	Tower Bottom Back Incli. Y σ_{std}
3	0.995	Tower Bottom Mom X Eq. value	0.987	Tower Bottom Incli. X σ_{std}
4	0.987	Tower Bottom Accel Y σ_{std}	0.984	Tower Top Accel X σ_{std}
5	0.980	Tower Top Accel Y σ_{std}	0.959	RNA Accel X σ_{std}
6	0.970	RNA Accel Y σ_{std}	0.956	RNA Accel X Eq. value
7	0.966	RNA Accel Y Eq. value	0.950	Tower Bottom Accel X σ_{std}
8	0.964	Tower Bottom IncliSS Eq. value	0.945	RNA Accel Y σ_{std}
9	0.962	Tower Bottom Incli. Y Eq. value	0.944	RNA Accel Y Eq. value
10	0.953	Tower Bottom Back Incli. X Eq. value	0.933	Tower Top Accel X λ_0
11	0.952	RNA Accel X σ_{std}	0.930	Tower Top Accel SS λ_0
12	0.950	Tower Bottom Incli. Y λ_0	0.926	RNA Accel X λ_{-1}
13	0.950	Tower Bottom Back Incli. X λ_0	0.925	RNA Accel X λ_0
14	0.944	Tower Top Accel Y λ_0	0.925	Tower Bottom Mom Y λ_1
15	0.939	Tower Bottom Accel Y λ_0	0.925	RNA Accel X λ_1
Rank	Turbine 2			
	X (Monopile (Upper level))		Y (Monopile (Upper level))	
	R value	Features	R value	Features
1	0.979	Tower Bottom Mom X Eq. value	0.992	Tower Top Accel X σ_{std}
2	0.973	RNA Accel X λ_1	0.990	Tower Bottom Mom Y Eq. value
3	0.972	RNA Accel X σ_{std}	0.986	Tower Bottom Back Incli. Y σ_{std}
4	0.972	Tower Bottom Incli. Y σ_{std}	0.985	Tower Bottom Incli. X σ_{std}
5	0.971	Tower Bottom Back Incli. X σ_{std}	0.982	Tower Top Accel X λ_0
6	0.970	RNA Accel X λ_{-1}	0.982	RNA Accel Y σ_{std}
7	0.969	RNA Accel X λ_0	0.977	Tower Bottom Mom Y λ_1
8	0.968	RNA Accel X Eq. value	0.976	RNA Accel Y Eq. value
9	0.964	Tower Bottom Back Incli. X λ_0	0.973	RNA Accel Y λ_0
10	0.964	Tower Bottom Incli. Y λ_0	0.972	RNA Accel Y λ_{-1}
11	0.929	Tower Bottom Accel Y σ_{std}	0.971	RNA Accel Y λ_1
12	0.928	Tower Bottom Accel Y λ_0	0.963	Tower Bottom Back Incli. Y λ_0
13	0.922	Tower Bottom Back Incli. Y σ_{std}	0.962	Tower Bottom Incli. X λ_0
14	0.921	Tower Bottom Incli. X σ_{std}	0.934	Tower Bottom Accel X σ_{std}
15	0.921	RNA Accel Y λ_{-1}	0.914	Tower Bottom Accel X λ_0

Monopile (Upper level) moment DEL, Parked Condition

Table E.6: Pearson's correlation coefficient, w.r.t. Monopile (Upper level) moment DEL, Parked Condition

Rank	Turbine 1			
	X (Monopile (Upper level))		Y (Monopile (Upper level))	
	R value	Features	R value	Features
1	0.993	Tower Bottom Mom X Eq. value	0.992	Tower Top Accel X σ_{std}
2	0.993	Tower Top Accel Y σ_{std}	0.992	Tower Bottom Mom Y Eq. value
3	0.973	Tower Top Accel Y λ_0	0.990	Tower Top Accel X Eq. value
4	0.971	Tower Top Accel Y Eq. value	0.976	RNA Accel Y Eq. value
5	0.968	Tower Bottom Mom X λ_1	0.974	RNA Accel Y σ_{std}
6	0.953	Tower Bottom Incli. Y σ_{std}	0.968	Tower Bottom Incli. X σ_{std}
7	0.952	Tower Bottom Incli. Y λ_0	0.968	Tower Bottom Back Incli. Y σ_{std}
8	0.952	Tower Bottom Back Incli. X σ_{std}	0.967	Tower Top Accel X λ_0
9	0.952	Tower Top Accel Y Range	0.962	Tower Bottom Mom Y λ_1
10	0.951	Tower Bottom Back Incli. X λ_0	0.952	Tower Bottom Back Incli. Y λ_0
11	0.951	Tower Top Accel Y Max	0.952	Tower Bottom Incli. X λ_0
12	0.919	Tower Bottom Accel Y λ_0	0.952	Tower Bottom Accel X σ_{std}
13	0.918	Tower Bottom Accel Y σ_{std}	0.947	Tower Top Accel X Range
14	0.910	RNA Accel Y Eq. value	0.947	RNA Accel Y Range
15	0.907	RNA Accel Y σ_{std}	0.947	Tower Top Accel X λ_1
Rank	Turbine 2			
	X (Monopile (Upper level))		Y (Monopile (Upper level))	
	R value	Features	R value	Features
1	0.998	Tower Bottom Mom X Eq. value	0.997	Tower Bottom Mom Y Eq. value
2	0.995	Tower Top Mom X Eq. value	0.995	Tower Top Accel X σ_{std}
3	0.994	Tower Top Accel Y σ_{std}	0.994	Tower Top Accel X Eq. value
4	0.994	Tower Bottom Mom X σ_{std}	0.985	Tower Bottom Incli. X σ_{std}
5	0.988	Tower Bottom Incli. Y σ_{std}	0.984	Tower Bottom Back Incli. Y σ_{std}
6	0.988	Tower Bottom Back Incli. X σ_{std}	0.975	Tower Bottom Accel X σ_{std}
7	0.986	Tower Top Accel Y Eq. value	0.968	Tower Top Accel X λ_0
8	0.980	Tower Bottom Accel Y σ_{std}	0.966	Tower Bottom Mom Y λ_1
9	0.979	Tower Top Mom X σ_{std}	0.962	RNA Accel Y Eq. value
10	0.976	Tower Bottom Incli. Y λ_0	0.962	Tower Top Accel X λ_1
11	0.975	Tower Bottom Back Incli. X λ_0	0.961	Tower Bottom Mom Y λ_2
12	0.974	Tower Bottom Incli. Y λ_{-1}	0.960	Tower Bottom Back Incli. Y λ_0
13	0.973	Tower Bottom Back Incli. X λ_{-1}	0.960	Tower Bottom Incli. X λ_0
14	0.972	Tower Bottom Accel Y λ_0	0.959	RNA Accel Y σ_{std}
15	0.970	Tower Top Mom X λ_2	0.951	Tower Bottom Accel X λ_0

Monopile (Lower level) moment DEL, Operating Condition 1

Table E.7: Pearson's correlation coefficient, w.r.t. Monopile (Lower level) moment DEL, Operating Condition 1

Rank	Turbine 1			
	X (Monopile (Lower level))		Y (Monopile (Lower level))	
	R value	Features	R value	Features
1	0.928	Tower Bottom Mom X Eq. value	0.914	Tower Bottom Mom Y Eq. value
2	0.905	Tower Bottom Mom X λ_1	0.900	Tower Bottom Mom Y λ_1
3	0.858	Tower Top Mom X Eq. value	0.889	RNA Accel X σ_{std}
4	0.848	Tower Top Accel Y σ_{std}	0.878	RNA Accel X Eq. value
5	0.834	Tower Top Accel Y λ_0	0.869	RNA Accel X λ_{-1}
6	0.823	RNA Accel Y σ_{std}	0.868	RNA Accel X λ_0
7	0.813	Tower Top Accel Y Max	0.863	Tower Top Mom Y Eq. value
8	0.811	RNA Accel X Eq. value	0.854	RNA Accel X Range
9	0.809	RNA Accel Y λ_{-1}	0.852	Tower Top Accel X σ_{std}
10	0.808	Tower Top Accel Y Range	0.841	Tower Bottom Mom Y λ_2
11	0.808	RNA Accel X σ_{std}	0.841	Tower Top Mom Y λ_1
12	0.804	RNA Accel Y Eq. value	0.841	Tower Top Accel X λ_0
13	0.804	RNA Accel Y λ_0	0.839	RNA Accel X Max
14	0.802	Tower Bottom Mom X λ_2	0.821	RNA Accel X λ_1
15	0.800	RNA Accel Y Range	0.814	Tower Top Accel Y σ_{std}
Rank	Turbine 2			
	X (Monopile (Lower level))		Y (Monopile (Lower level))	
	R value	Features	R value	Features
1	0.952	Tower Bottom Mom X Eq. value	0.937	Tower Bottom Mom Y λ_1
2	0.914	RNA Accel Y σ_{std}	0.936	Tower Bottom Mom Y Eq. value
3	0.899	Tower Bottom Mom X λ_1	0.893	Tower Bottom Mom Y λ_2
4	0.889	RNA Accel Y λ_0	0.892	RNA Accel X σ_{std}
5	0.876	RNA Accel Y λ_{-1}	0.887	Tower Top Accel X σ_{std}
6	0.856	RNA Accel Y Eq. value	0.883	RNA Accel X λ_0
7	0.847	Tower Bottom Mom X λ_2	0.877	Tower Top Accel X λ_0
8	0.832	RNA Accel Y Range	0.876	RNA Accel X Eq. value
9	0.824	RNA Accel Y Max	0.867	RNA Accel Y Eq. value
10	0.810	Tower Top Accel Y σ_{std}	0.846	RNA Accel X Range
11	0.794	Tower Top Accel Y λ_0	0.845	RNA Accel X λ_{-1}
12	0.775	Tower Top Mom X Eq. value	0.842	RNA Accel Y σ_{std}
13	0.771	RNA Accel X λ_{-1}	0.833	RNA Accel Y λ_1
14	0.758	RNA Accel Y λ_1	0.833	RNA Accel X λ_1
15	0.746	RNA Accel X σ_{std}	0.831	RNA Accel X Max

Monopile (Lower level) moment DEL, Operating Condition 3

Table E.8: Pearson's correlation coefficient, w.r.t. Monopile (Lower level) moment DEL, Operating Condition 3

Rank	Turbine 1			
	X (Monopile (Lower level))		Y (Monopile (Lower level))	
	R value	Features	R value	Features
1	0.977	Tower Bottom Incli. Y σ_{std}	0.978	Tower Bottom Mom Y Eq. value
2	0.976	Tower Bottom Back Incli. X σ_{std}	0.975	Tower Bottom Incli. X σ_{std}
3	0.968	Tower Bottom Mom X Eq. value	0.975	Tower Bottom Back Incli. Y σ_{std}
4	0.966	Tower Bottom Accel Y σ_{std}	0.967	Tower Top Accel X σ_{std}
5	0.956	Tower Top Accel Y σ_{std}	0.957	RNA Accel X σ_{std}
6	0.950	RNA Accel Y σ_{std}	0.953	RNA Accel X Eq. value
7	0.944	RNA Accel Y Eq. value	0.948	RNA Accel Y σ_{std}
8	0.932	RNA Accel X σ_{std}	0.944	RNA Accel Y Eq. value
9	0.931	Tower Bottom Incli. Y Eq. value	0.936	Tower Bottom Accel X σ_{std}
10	0.923	Tower Bottom Incli. Y λ_0	0.931	Tower Bottom Back Incli. X σ_{std}
11	0.923	Tower Bottom Back Incli. X λ_0	0.931	Tower Bottom Incli. Y σ_{std}
12	0.923	Tower Bottom Back Incli. X Eq. value	0.928	Tower Bottom Mom X Eq. value
13	0.915	RNA Accel X Eq. value	0.925	Tower Bottom Accel Y σ_{std}
14	0.912	Tower Top Accel Y λ_0	0.922	RNA Accel X λ_{-1}
15	0.910	Tower Bottom Accel Y λ_0	0.921	RNA Accel X λ_0
Rank	Turbine 2			
	X (Monopile (Lower level))		Y (Monopile (Lower level))	
	R value	Features	R value	Features
1	0.973	Tower Bottom Mom X Eq. value	0.988	Tower Top Accel X σ_{std}
2	0.970	RNA Accel X σ_{std}	0.986	Tower Bottom Back Incli. Y σ_{std}
3	0.964	Tower Bottom Incli. Y σ_{std}	0.986	Tower Bottom Mom Y Eq. value
4	0.963	RNA Accel X λ_{-1}	0.985	Tower Bottom Incli. X σ_{std}
5	0.963	RNA Accel X λ_1	0.983	Tower Top Accel X λ_0
6	0.963	Tower Bottom Back Incli. X σ_{std}	0.978	RNA Accel Y σ_{std}
7	0.963	RNA Accel X Eq. value	0.974	Tower Bottom Mom Y λ_1
8	0.961	RNA Accel X λ_0	0.974	RNA Accel Y λ_{-1}
9	0.952	Tower Bottom Incli. Y λ_0	0.974	RNA Accel Y λ_0
10	0.951	Tower Bottom Back Incli. X λ_0	0.970	RNA Accel Y Eq. value
11	0.930	Tower Bottom Back Incli. Y σ_{std}	0.970	RNA Accel Y λ_1
12	0.929	Tower Bottom Incli. X σ_{std}	0.969	Tower Bottom Back Incli. Y λ_0
13	0.925	Tower Bottom Back Incli. Y λ_0	0.967	Tower Bottom Incli. X λ_0
14	0.924	Tower Bottom Incli. X λ_0	0.937	Tower Bottom Accel X σ_{std}
15	0.922	RNA Accel Y λ_{-1}	0.924	Tower Bottom Accel X λ_0

Monopile (Lower level) moment DEL, Parked Condition

Table E.9: Pearson's correlation coefficient, w.r.t. Monopile (Lower level) moment DEL, Parked Condition

Rank	Turbine 1			
	X (Monopile (Lower level))		Y (Monopile (Lower level))	
	R value	Features	R value	Features
1	0.986	Tower Top Accel Y σ_{std}	0.980	Tower Bottom Incli. X σ_{std}
2	0.979	Tower Bottom Mom X Eq. value	0.980	Tower Bottom Back Incli. Y σ_{std}
3	0.970	Tower Top Accel Y Eq. value	0.976	Tower Top Accel X σ_{std}
4	0.961	Tower Top Accel Y λ_0	0.975	Tower Top Accel X Eq. value
5	0.958	Tower Bottom Incli. Y σ_{std}	0.970	RNA Accel Y Eq. value
6	0.957	Tower Bottom Back Incli. X σ_{std}	0.970	RNA Accel Y σ_{std}
7	0.954	Tower Bottom Incli. Y λ_0	0.968	Tower Bottom Accel X σ_{std}
8	0.953	Tower Bottom Back Incli. X λ_0	0.968	Tower Bottom Mom Y Eq. value
9	0.951	Tower Bottom Mom X λ_1	0.962	Tower Bottom Back Incli. Y λ_0
10	0.949	Tower Top Accel Y Range	0.962	Tower Bottom Incli. X λ_0
11	0.943	Tower Top Accel Y Max	0.949	Tower Bottom Accel X λ_0
12	0.925	Tower Bottom Accel Y σ_{std}	0.947	Tower Top Accel X λ_0
13	0.924	Tower Bottom Accel Y λ_0	0.942	Tower Bottom Back Incli. Y Eq. value
14	0.906	RNA Accel Y Eq. value	0.941	Tower Bottom Incli. X Eq. value
15	0.905	RNA Accel Y σ_{std}	0.940	RNA Accel Y Range
Rank	Turbine 2			
	X (Monopile (Lower level))		Y (Monopile (Lower level))	
	R value	Features	R value	Features
1	0.993	Tower Bottom Mom X Eq. value	0.989	Tower Top Accel X Eq. value
2	0.990	Tower Top Mom X Eq. value	0.989	Tower Top Accel X σ_{std}
3	0.990	Tower Top Accel Y σ_{std}	0.987	Tower Bottom Mom Y Eq. value
4	0.989	Tower Bottom Mom X σ_{std}	0.986	Tower Bottom Incli. X σ_{std}
5	0.986	Tower Bottom Incli. Y σ_{std}	0.986	Tower Bottom Back Incli. Y σ_{std}
6	0.985	Tower Bottom Back Incli. X σ_{std}	0.979	Tower Bottom Accel X σ_{std}
7	0.981	Tower Top Accel Y Eq. value	0.963	Tower Bottom Back Incli. Y λ_0
8	0.979	Tower Bottom Accel Y σ_{std}	0.962	Tower Bottom Incli. X λ_0
9	0.975	Tower Bottom Incli. Y λ_0	0.961	Tower Top Accel X λ_1
10	0.974	Tower Bottom Back Incli. X λ_0	0.960	Tower Top Accel X λ_0
11	0.974	Tower Top Mom X σ_{std}	0.957	Tower Bottom Mom Y λ_1
12	0.973	Tower Bottom Incli. Y λ_{-1}	0.956	Tower Bottom Accel X λ_0
13	0.972	Tower Bottom Back Incli. X λ_{-1}	0.954	RNA Accel Y Eq. value
14	0.971	Tower Bottom Accel Y λ_0	0.951	RNA Accel Y σ_{std}
15	0.964	Tower Top Mom X λ_2	0.950	Tower Bottom Mom Y λ_2

Appendix F

Important features to estimate moment DEL

F.1 Feature selection with random forest algorithm

To figure out the most important features, random forest regression algorithm has been applied [6, 15]. In random forest regression application, all the signals and all the statistical properties have been used to estimate FA directional moment DEL at tower bottom.

Random forest regression algorithm uses multiple decision trees. The decision tree uses successive recursive binary splits on all the features and generate multiple area which is called 'leaf'. After generating all these 'leaves', take average of target values of each leaf as a resultant output estimation of the leaf. In random forest regression, all the estimations from multiple decision trees are gathered and the final estimation can be made by taking average of all these estimations. Detailed explanation can be found in [5, 15, 19].

To calculate feature importance, binary split is taken to all the features iteratively. Then, the resultant squared error can be calculated for each split. As a result, feature importance can be calculated by comparing how the squared error be affected by each split.

As a result of the random forest regression application, feature importance has been calculated as shown in Table F.1.

Table F.1: Top 10 important features, from random forest algorithm

No.	Features
1	SCADA Accel Eq. value
2	Tower Top Mom range
3	Tower Top Accel. λ_0
4	Tower Top Accel. σ_{std}
5	Tower Top Mom Eq. value
6	RNA Accel. λ_{-1}
7	RNA Accel. σ_{std}
8	RNA Accel. λ_0
9	Tower Top Accel. λ_{-1}
10	Tower Top Mom λ_1

F.2 Feature selection with PCA

In addition to random forest, feature selection has also been performed with PCA. Feature selection with PCA has been applied to multiple different machine learning problems such as classification, object recognition and so on [20, 24, 27]. In this study, the same methodology has been applied.

Specifically, it uses principal components to select important features. As a result of PCA, percentage variance and eigen vectors (which is corresponding to principal component) can be obtained. Then, firstly, from the percentage variance, only the components corresponding to 99% of total accumulated percentage variance have been extracted. For example, if there is total m principal components exist, from the percentage variance, only p ($p < m$) components can be selected with which the summation of the p component's percentage variance is higher than 99%.

Secondly, from the selected principal components, the contribution of original features can be obtained by reading the absolute value of eigen vector components. Concretely, if an eigen vector (a principal component) consists of linear combination of total n original features, i th component of the eigen vector stands for the contribution of i th original feature. In this process, only the original features which have the contribution of higher than ϵ (ϵ is threshold value which has been set as 0.1 in this study) have been selected.

For this study, input features corresponding to CS10 in Table 4.23 has been applied. It should be noted that target data is irrelevant in feature selection with PCA.

As a result, the features shown in Table F.2 have been recognized as top 10 the

most important features.

It should be noted that the results are different from that of Pearson's correlation coefficient comparison and random forest algorithm. The reason for this is that PCA uses contribution of the input features on principal axes, while Pearson's correlation coefficient comparison and random forest algorithm find contribution of input features on moment DEL estimation.

Table F.2: Top 10 important features, CS10, from PCA

No.	Features
1	Tower Bottom Mom Eq. value
2	Tower Bottom Incli. Back Range
3	Tower Bottom Incli. Back σ_{std}
4	Tower Bottom Incli. Back Eq. value
5	Tower Bottom Incli. Eq. value
6	Tower Bottom Incli. σ_{std}
7	Tower Bottom Mom λ_2
8	Tower Bottom Incli. Back λ_{-1}
9	SCADA Accel Range
10	Tower Bottom Incli. Range

Appendix G

Combined weight calculation after training

After train the feed-forward neural network, weights can be extracted. By comparing the weights, it can be investigated which features have the highest effect on estimation. Specifically, combined weight of i th feature can be calculated by multiplying all the weights connecting i th input neuron to output neuron. It is shown in Figure G.1.

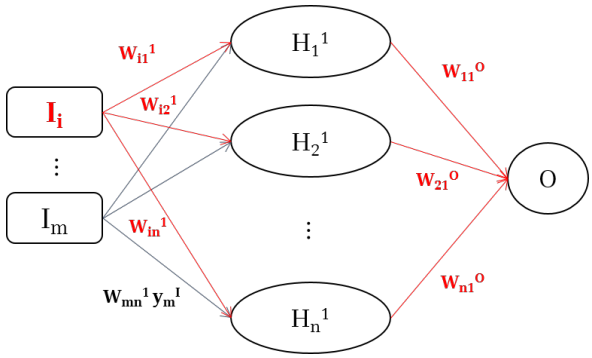


Figure G.1: Combined weight calculation

As shown in Figure G.1, the combined weight of i th feature can be calculated by Equation (G.1). It should be noted that the Equation is only valid for feed forward

neural network with one hidden layer and one output neuron.

$$w_{i,combined} = \sum_{k=1}^{n} w_{ik}^{1} w_{k1}^{O} \tag{G.1}$$

If all the features are normalized in the same scale, the higher combined weight can be interpreted as higher effect on output value.

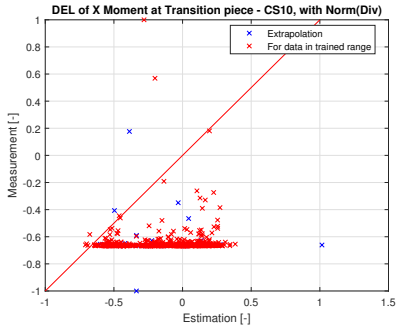
Appendix H

Farm-wide level load estimation - results of normalization

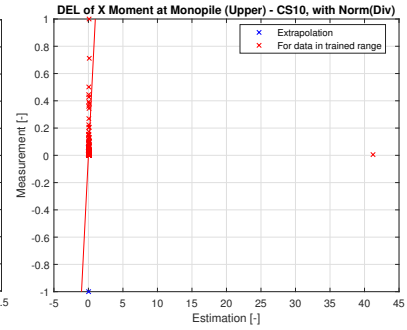
Farm-wide level load estimation results with normailzation (division) :

Table H.1: Results of farm-wide level load estimation, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with normalization (division)

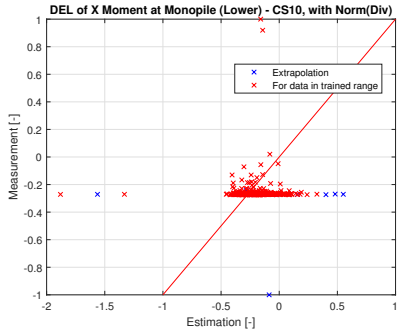
	within Train Range		Extrapolation		Full	
	R value	MAPE	R value	MAPE	R value	MAPE
Transition Piece (X)	-0.014	>100%	0.056	>100%	0.001	>100%
Transition Piece (Y)	0.013	>100%	0.190	>100%	0.031	>100%
Monopile (upper) (X)	0.007	>100%	0.092	>100%	0.016	>100%
Monopile (upper) (Y)	0.013	>100%	-0.037	>100%	0.003	>100%
Monopile (lower) (X)	0.022	>100%	-0.029	>100%	0.014	>100%
Monopile (lower) (Y)	-0.016	>100%	0.006	>100%	-0.015	>100%



(a) Transition piece, X moment DEL



(b) Monopile(upper), X moment DEL



(c) Monopile(lower), X moment DEL

Figure H.1: Results of farm-wide level load estimation with wake direction, All locations, Operating condition 1, T1A2, Feed forward neural network, CS10 with normalization (division)

Bibliography

[1] Charu C Aggarwal. *Neural networks and deep learning*. Springer. Cited at P. 27

[2] Avrim L Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2):245–271, 1997. Cited at P. 81

[3] Eduard Bonada. Diagnosing learning algorithms. `http://www.ebc.cat/2017/02/19/diagnosing-learning-algorithms/`, 2017. Accessed 02/05/19. Cited at P. 35

[4] Arno J Brand, Joachim Peinke, and Jakob Mann. Turbulence and wind turbines. *Journal of Physics: Conference Series*, 318(7):072005, dec 2011. Cited at P. 22

[5] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. Cited at P. 165

[6] Leo Breiman. *Classification and regression trees*. Routledge, 2017. Cited at P. 165

[7] Nicolai Cosack. Fatigue load monitoring with standard wind turbine signals, 2010. Cited at P. 3

[8] DNVGL. Lifetime extension of wind turbines. tech. rep., DNVGL, 2016. DNVGL-ST-0262. Cited at P. 2

[9] Wind Europe. Offshore wind in europe, key trends and statistics 2017. 2018. Cited at P. 1

[10] Wind Europe. Offshore wind in europe, key trends and statistics 2018. 2019. Cited at P. 1

[11] Yoav Goldberg. Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, 10(1):1–309, 2017. Cited at P. 36

[12] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994. Cited at P. 27

[13] Alexandros Iliopoulos, Rasoul Shirzadeh, Wout Weijtjens, Patrick Guillaume, Danny Van Hemelrijck, and Christof Devriendt. A modal decomposition and expansion approach for prediction of dynamic responses on a monopile offshore wind turbine using a limited number of vibration sensors. *Mechanical Systems and Signal Processing*, 68:84–104, 2016. Cited at P. 3

[14] IEEE Instrumentation and Measurement Society Staff Corporate Author. Feature selection for defect classification in machine condition monitoring. In *2003 IEEE Conference Instrumentation and Measurement Technology*, volume 1 of *20th IEEE Instrumentation Technology Conference (Cat. No.03CH37412)*, pages 36–41, [Place of publication not identified], 2003. I E E E. Cited at P. 81

[15] Hemant Ishwaran et al. Variable importance in binary regression trees and forests. *Electronic Journal of Statistics*, 1:519–537, 2007. Cited at P. 165

[16] Ian Jolliffe. *Principal component analysis*. Springer, 2011. Cited at P. 81

[17] Ian T Jolliffe. A note on the use of principal components in regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(3):300–303, 1982. Cited at P. 81

[18] Sang Lee, Matthew Churchfield, Patrick Moriarty, J Jonkman, and J Michalakes. Atmospheric and wake turbulence impacts on wind turbine fatigue loadings. In *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, page 540, 2012. Cited at P. 22

[19] Gilles Louppe. Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502*, 2014. Cited at P. 165

[20] Arnaz Malhi and Robert X Gao. Pca-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement*, 53(6):1517–1525, 2004. Cited at P. 166

[21] Nick Martin and Hermine Maes. *Multivariate analysis*. Academic press London, 1979. Cited at P. 81

[22] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978. Cited at P. 31

[23] Frederick Mosteller and John Wilder Tukey. Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, 1977. Cited at P. 81

[24] Nikhil Naikal, Allen Y Yang, and S Shankar Sastry. Informative feature selection for object recognition via sparse pca. In *2011 International Conference on Computer Vision*, pages 818–825. IEEE, 2011. Cited at P. 166

[25] Andrew Y Ng. *On feature selection: learning with exponentially many irreverent features as training examples*. PhD thesis, Massachusetts Institute of Technology, 1998. Cited at P. 81

[26] DJ Cerda Salzmann and J Van der Tempel. Aerodynamic damping in the design of support structures for offshore wind turbines. In *Paper of the Copenhagen offshore conference*. Citeseer, 2005. Cited at P. 77, 93

[27] Fengxi Song, Zhongwei Guo, and Dayong Mei. Feature selection using principal component analysis. In *2010 international conference on system science, engineering design and manufacturing informatization*, volume 1, pages 27–30. IEEE, 2010. Cited at P. 166

[28] Marinos Souliotis. Machine learning techniques for load monitoring of offshore wind turbines. 2017. Cited at P. 4

[29] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014. Cited at P. 36

[30] Andrew Trask, Felix Hill, Scott E Reed, Jack Rae, Chris Dyer, and Phil Blunsom. Neural arithmetic logic units. In *Advances in Neural Information Processing Systems*, pages 8035–8044, 2018. Cited at P. 92

[31] AZ Ul-Saufie, AS Yahya, and NA Ramli. Improving multiple linear regression model using principal component analysis for predicting pm10 concentration in seberang prai, pulau pinang. *International Journal of Environmental Sciences*, 2(2):415–422, 2011. Cited at P. 81

[32] W. Weijtjen, N. Noppe, T. Verbelen, C. Devriendt, and A. Iliopoulos. Fatigue life assessment of three offshore wind turbines. In *Life-Cycle of Engineering Systems: Emphasis on Sustainable Civil Infrastructure - 5th International Symposium on Life-Cycle Engineering, IALCCE 2016*, pages 742–747. CRC Press/Balkema, 2017. Cited at P. 91

[33] W WEIJTJENS, N NOPPE, A ILIOPOULOS, and C DEVRIENDT. Monitoring the consumed fatigue life on three offshore wind turbines. *Structural Health Monitoring 2015*, 2015. Cited at P. 91

[34] Hao Yu and Bogdan M Wilamowski. Levenberg-marquardt training. *Industrial electronics handbook*, 5(12):1, 2011. Cited at P. 31, 35

[35] Lisa Ziegler and Michael Muskulus. Fatigue reassessment for lifetime extension of offshore wind monopile substructures. In *Journal of Physics: Conference Series*, volume 753, page 092010. IOP Publishing, 2016. Cited at P. 3

[36] Lisa Ziegler, Ursula Smolka, Nicolai Cosack, and Michael Muskulus. Brief communication: Structural monitoring for lifetime extension of offshore wind monopiles: can strain measurements at one level tell us everything? *Wind Energy Science*, 2(2):469–476, 2017. Cited at P. 3

[37] Julian Georg Zilly, Rupesh Kumar Srivastava, Jan Koutník, and Jürgen Schmidhuber. Recurrent highway networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 4189–4198. JMLR. org, 2017. Cited at P. 37

TU Delft Delft University of Technology

NTNU

SIEMENS Gamesa RENEWABLE ENERGY