ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

BlockLab
Blockchain for Energy and Logistics

Master of Science in Data Science & Marketing Analytics

Erasmus School of Economics

Erasmus University, The Netherlands

Graduation Thesis

IS THERE A WINDOW OF OPPORTUNITY IN THE WEATHER? WIND SPEED & WAVE HEIGHT FORECASTS FOR OPERATION AND MAINTENANCE OF GEMINI WIND FARM – THE CASE OF JACK-UP VESSEL

Author: Phạm Anh Tuấn
Student ID: 532458tp

Supervisor: Prof. Dr. Ir. Rommert Dekker
Second Assessor: Ass. Prof. NM Almeida Camacho

30th April 2022

Table of Contents

List of tables

Table of figures

Abbreviation

Abbreviation	Meaning
AICc	Corrected Akaike's Information Criteria
ANN	Artificial Neural Network
ARIMA	Auto-Regressive Integrated Moving Average
ARMA	Auto-Regressive Moving Average
CNN	Convolutional Neural Network
CRPS	Continuous Ranked Probability Score
CV	Cross Validation
DBN	Deep Belief Network
ENS-ELM	Ensemble Extreme Learning Machine
ETS	Exponential Smoothing methods with Trends and Seasonality
F-ARMA	Fourier - Auto Regressive Moving Average
GAMs	Generalized Additive Models
GW	Gigawatt
ICEEMDAN	Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
IR	Innovation Residuals
kW	Kilowatt
LSTM	Long Short-term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Errors
MSE	Mean Squared Error
MW	Megawatt
NAR	Non-linear Auto-Regressive Neural Network without exogenous inputs
NARX	Non-linear Auto-Regressive Neural Network with exogenous inputs
O&M	Operation and Maintenance
RMSE	Root Mean Squared Errors
RNN	Recurrent Neural Network
RQ	Research question
SGB	Stochastic Gradient Boosting
STL	Seasonal and Trend decomposition using Loess
SVM	Support Vector Machine
UC	Unobserved Components

Foreword

This thesis has been submitted as the graduation requirement for the Master of Science in Data Science and Marketing Analytics at The Erasmus University of Rotterdam and the internship at Block Fieldlab B.V. First of all, I would like to express my sincerest appreciation to Professor. Rommert Dekker and Mr. Aljosja Beije for their idea and guidance throughout the project. Your words of wisdom have paved the way to contribute my thesis to a promising and sustainable solution for humanity. Secondly, I would like to thank my family, my classmates Elina, Ayan, Aigerim, Michele, Emilia, Nathalia, Joy, Filippo, Jenny, Sofia, Isidora, Neal, Kenita, Alizera, and Markus, my colleagues, friends in Vietnam and Rotterdam who raised me up and supported me through my depression. Thirdly, I would like to give my credit to Mr. Wout Frijters - my supervisor at Blocklab, Professor Rob J Hyndman - Head of the Department of Econometrics & Business Statistics at Monash University for his instruction in building prediction model. In addition, I appreciate Mr. Maurice Vogel, Mr. Leon Maughan – Product Owner at DTN, Sitthichat Sukpholtham student in Ocean Acoustic at the University of Hawaii in Manoa, and Mr. Stefano Mattia – Earth Observation Ground Segment System Engineer at Rhea System S.A providing me with more insight into climatology and weather data. Last but by no means least, thanks to Professor. NM Almeida Camacho for his time and effort in reading and evaluating this thesis.

Executive summary

Given the commitment to renewable energy in The Netherlands, offshore wind energy is a viable source of energy supply. The further on the sea the offshore wind farm, the more energy generated, however, the larger cost of Operation and Maintenance. This cost might take up to approximately 30% of the capital expenses of offshore wind(Rodrigues *et al.*, 2015). In addition, the further away from the shore, the more dangerous the maintenance service is due to more volatile oceanic weather conditions. In addition, a jack-up vessel is necessary to conduct maintenance and replacement of large turbine components such as the gear box. However, this vessel is quite expensive and requires a thorough and long process to implement the maintenance services. Therefore, we are interested in the following research problem:

"How offshore wind farm operators can apply the weather window of wave height and wind speed forecasting in planning the jack-up vessel in an efficient way while ensuring some safety standards?"

In the first part of this research, we focus on data collection, cleaning, and analysis to discover the nature of the weather and wave time series. These data were first collected at an hourly level. We then applied random forest imputation to the wind speed time series to regenerate the wind speed pattern after more than 2 months of missing values. The random forest model using 50 trees and 7 variables to impute the data provided the lowest imputation error in general, however, it also took the longest computation time. After that, the wind speed time series were aggregated into average daily mean wind speed to conduct a long-term wind speed prediction of one year. The hourly mean wind speed, wind power, and significant wave height were used for short-term wind speed prediction of up to one day. After that, these data were decomposed by the STL decomposition method to discover the multi-seasonality feature and trend during the period under the study. There was a decreasing trend in both the wind speed and wind power output which required a more efficient jack-up vessel chartering plan and execution to optimize the O&M service of the offshore wind farm for more profit gain. These data were also non-stationary and not normally distributed; therefore, a Box-Cox transformation was applied to make their distribution more symmetric. In addition, there was a strong yearly – seasonal V-shape pattern in both wind speed and wind power and we can divide the wind climate into two seasons. The winter wind season is from October to March, and the summer wind season is from April to September. There is not an obvious pattern in the significant wave height time series, however, we can also divide the wave season the same way as the wind season.

In the second part of this research, we implemented a long-term prediction model for the wind speed using the daily wind speed time series with Fourier Autoregressive Moving Average (F-ARMA) method. We then conducted the short-term wind speed, wind power, and significant wave height prediction at the hourly level by using the Non-linear Autoregressive Neural Network model without exogenous outputs (NAR). The F-ARMA model outperformed a seasonal naïve model in prediction accuracy, therefore, we used its forecast to select the weather window for O&M activities. Based on our long-term wind speed prediction for 2020, we took advantage of the decreasing trend of the wind speed at the beginning of the summer and compared the weather condition with the safety standards of the jack-up vessel to select the local minimum in March, April, and May 2020. Specifically, we forecasted that in 2020, the local minimum of the average daily mean wind speed in March was the 30th of March and the local minimum of average daily mean wind speed was the 28th of April. After that, we recommend building a medium-term prediction to verify these local minima of each weather window before implementing short-term predictions.

Due to the time scope, we implemented short-term wind speed, wind power, and significant wave height prediction one day before and after these two selected days to verify the long-term prediction results and established the trade-off between the revenue gained by the higher wind speed and the demurrage rate of the jack-up vessel. The performance of the NAR model for wind speed was much better on the 29th and 30th of March 2020 with the MAPE of less than 15%, however, we could not conduct maintenance service on these days due to unsafe wind speed and wave height conditions.

Regards the forecasted weather window in April 2020, the wind speed and wave height were predicted to be more favorable to implement maintenance although the prediction error was a bit larger than the forecasted weather window in March 2020. The wind power prediction was unfortunately not accurate due to the wind energy data collection could not capture the correct nominal wind energy produced by the Gemini wind farm.

In the third part of the research, we calculate the trade-off between the revenue gained by the wind power when we delay the maintenance service and the demurrage rate of the jack-up vessel. We found that it is not cost-effective to offset the demurrage rate of the jack-up vessel. However, we can apply long-term and short-term predictions of the weather data every year to establish the schedule of chartering the same jack-up vessel between different wind farms. Thus, enabling the vessel sharing approach of the jack-up vessel to optimize the vessel use, and reduce the chartering cost, documentation time, and contract negotiation. In this way, we could reduce the lead time of jack-up vessel deployment and the downtime cost of the broken turbine. We also suggest investigating the weather window from October 31st to November 1st since the delay at that time may be more profitable

1. INTRODUCTION

1.1. Research problem and motivation

The Netherlands government is targeting 70% of total energy coming from renewable sources in 2030 (Government of The Netherlands, 2020). More than half of which will be generated from offshore windfarm thanks to the favorable conditions of The North Sea's shallow water, seasonal wind climate, and the proximity of the wind farm to the nearest port (Government of The Netherlands, 2020). Specifically, in 2019, offshore wind turbines generated around 1 gigawatt (GW) of power in total (Government of The Netherlands, 2020), and nearly 1.5 GW of new offshore wind was installed in 2020, making the Netherlands the second-largest market followed by Belgium (706 MW) ((GWEC), 2021). In addition, continuous effort is being invested to raise the capacity of offshore wind electricity consumption to 40% (Accenture, 2021) of total electricity consumption in the country by the end of 2030. The larger the wind farm is, the more operation expenditure is spent since operation and maintenance (O&M) costs account for roughly 25% to 30% (Röckmann, Lagerveld, and Stavenuiter., 2017) of the total expenditure. Therefore, the optimization of operation and maintenance services for the offshore wind farm would make the wind electricity more affordable to the citizen and support the Netherlands to realize its commitment.

Operation and Maintenance of the wind farm consist of three main activities which are the transportation of maintenance personnel, shipment of larger equipment and components, and lifting activities. One of the main challenges in designing a cost-effective offshore maintenance service is to predict the weather window including the wind speed and wave height prediction so that the wind farm operator can plan the O&M activities. This will not only result in the minimum downtime cost but also ensure the safest and fastest accessibility of the maintenance crew, shipments of large replacement components, and lifting activities to the wind farm (Sperstad *et al.*, 2017). Therefore, this thesis aims to solve the following research problem:

"How offshore wind farm operators can apply the weather window of wave height and wind speed forecasting in planning the jack-up vessel in an efficient way while ensuring some safety standards?"

Regards of economic benefit, this thesis determines the downtime cost of the interruption in the energy output based on the wind speed prediction and the trade-off between flexible chartering and pre-fixed chartering of the jack-up vessel (Faulstich, Hahn and Tavner, 2011). In terms of safety standards, wind speed and wave height forecasts are used to evaluate the accessibility of the maintenance crew to the wind farm and the weather restriction for lifting activities, maintenance, and repair activities. Thus, this thesis first constructs a long-term wind speed prediction model to identify wind speed seasonality and select the period for O&M activities. After that, short–term wind speed and wave height forecasting predictions are executed to determine and forecast the specific economic benefit of O&M activities in the selected period, loss or gain of the expected downtime due to the delay in maintenance, and the trade-off of chartering the jack-up vessel. Then, the wind farm operator can schedule the vessel and prepare the contract for collaborations between different wind farms.

In order to apply the thesis's model to the real-life case study, we analyzed the case of Gemini Wind Park which is located 85 km away from the shore of The Dutch North Sea at the depth of 30 meters under the seawater level. There are 150 Siemens SWT- 4.0 - 130 wind turbines installed in the wind park with a total nominal capacity of 600 Megawatt per year. This wind park was fully operational in 2017 with the operation and maintenance service implemented by Siemens Gamesa until 2036[1].

1.2. How does the thesis contribute to current literature?

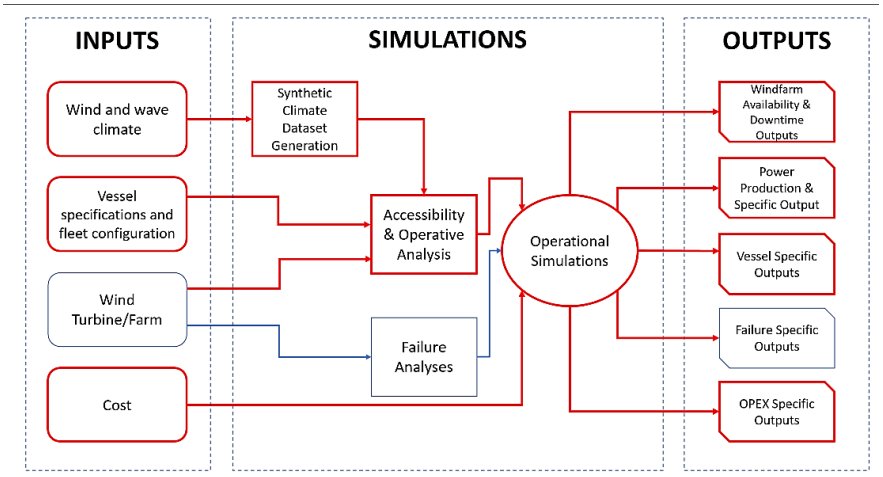
This thesis aims to optimize the schedule for O&M activities of the offshore wind farm based on the predicted weather window with wind speed and wave height prediction along with the cost data. Yalcin Dalgic mapped out different methodologies in optimizing Operation and Maintenance activities by diving them based on the input data, simulations, and expected outputs for the research as in Figure

[1] https://www.geminiwindpark.nl/now-at-gemini-wind-park.html#n6

1 (Yalcin Dalgic, 2015). Based on this figure, the thesis contributes to recent literature in the following ways.

First of all, it contributes to the (Schouten, 2019) research of wind farm O&M activities and failure analysis by predicting the difference in energy production and downtime outputs for the certain delay. Secondly, it provides Synthetic Climate Data Prediction of both the wind speed and the wave height for the accessibility and operative analysis. Thirdly, the thesis contributes to the planning of jack-up vessels for the wind farm in the long term. In addition, by applying short-term wind speed and wave height prediction, we can assess the accessibility of the wind farm for jacking, lifting, and maintenance activities of the jack-up vessel. Last but not least, the thesis applies the most updated data on the downtime cost of the Dutch wind energy market in estimation and modeling for the case of Gemini Wind Farm.

Figure 1: Developed O&M Methodologies of the offshore wind farm (Source: Dalgic, Lazakis, Dinwoodie, et al., 2015).



1.3. The purpose of the thesis and the research questions

The thesis aims to create a weather window of long-term wind speed forecast, short-term wind speed, and wave height prediction which empowers the decision-making process to solve the research problem in maintenance planning and vessel schedule for the safety of the maintenance crew with the least downtime cost.

Regarding the prediction methods, machine learning algorithms, nowadays, pose promising prediction results with higher accuracy compared to traditional statistical models. Therefore, this thesis also aims to apply not only the traditional auto-regressive model but also other machine learning algorithms including shallow and deep learning methods to recommend the most suitable model which best addresses the objectives and the research questions.

Thus, the thesis targets to address the following main research questions (RQ):

- RQ1: How can the offshore wind farm operator forecast the wind speed over a short-term and long-term period?
- RQ2: How can the offshore wind farm operator forecast the wave height over a short-term period?
- RQ3: How should the wind farm operator charter the jack-up vessel, in which season of the year, on which day, and for how long given certain maintenance needs? (Figure 2)

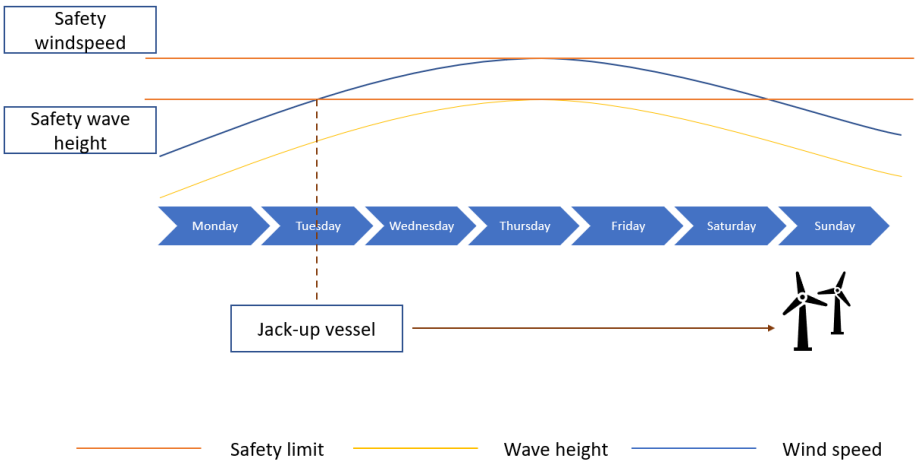


Figure 2: Illustration of Research Question 3

- RQ4: How can the wind farm operator determine the trade-off between the revenue gained by 1 day/1 week/1 month delay in maintenance for more energy generation thanks to better weather? (Figure 3)

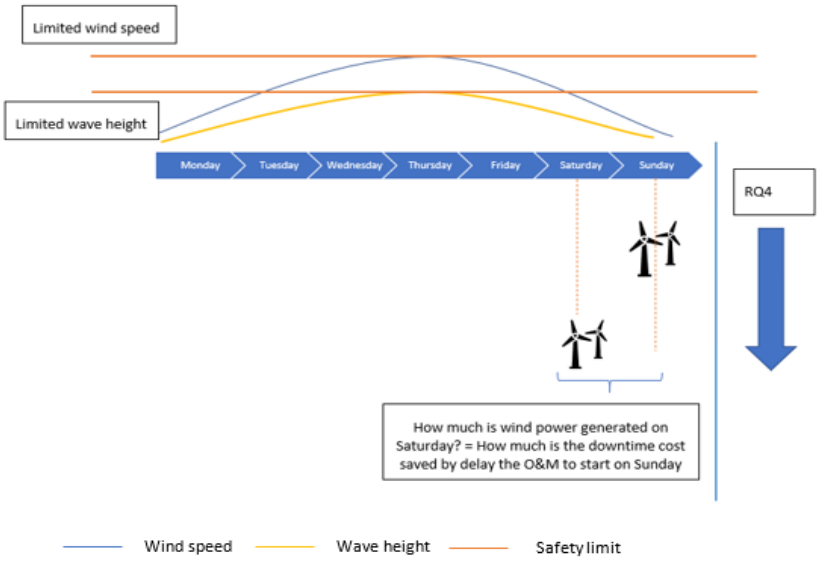


Figure 3: Illustration of Research Question 4

- RQ5: How can the wind farm operator organize the jack-up vessel to conduct maintenance on the wind farm by establishing the trade-off between revenue of the wind farm gained in the delay and the demurrage rate of the jack-up vessel? (Figure 4)

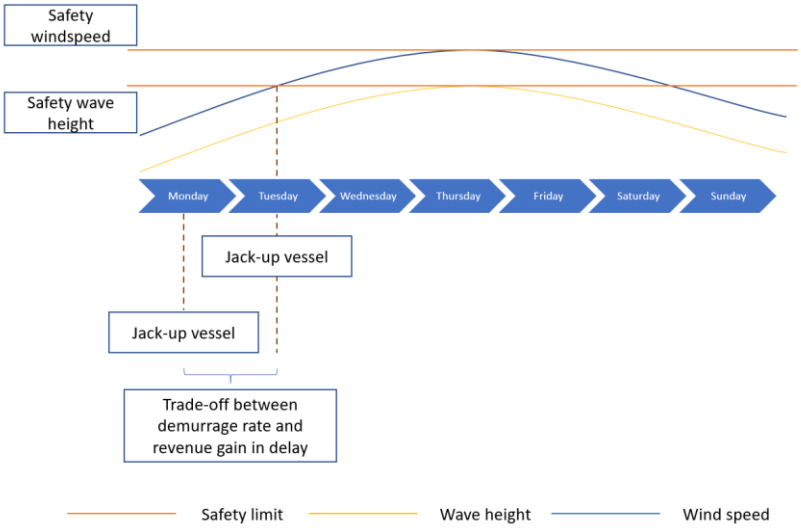


Figure 4: Illustration of Research Question 5

1.4. Thesis outline

We present the general overview of research about offshore windfarm O&M optimization and related methodology in wind speed and wave height prediction in Chapter 2. In Chapter 3, we collected the data, cleaned, and analyzed the weather data time series. In Chapter 4, we implemented the long-term wind speed prediction to answer RQ1 and sketched out the weather window for 1-year planning and scheduling of the jack-up vessel to answer RQ3. After that, we selected the weather window with the day with the lowest wind speed to conduct short-term wind speed, wind power, and wave height prediction to answer RQ1 and RQ2. After that, we conducted an economic analysis to evaluate the trade-off between the downtime cost and demurrage rate, and how the wind farm operator could use these models as a tool to assist their planning and scheduling of jack-up vessels.

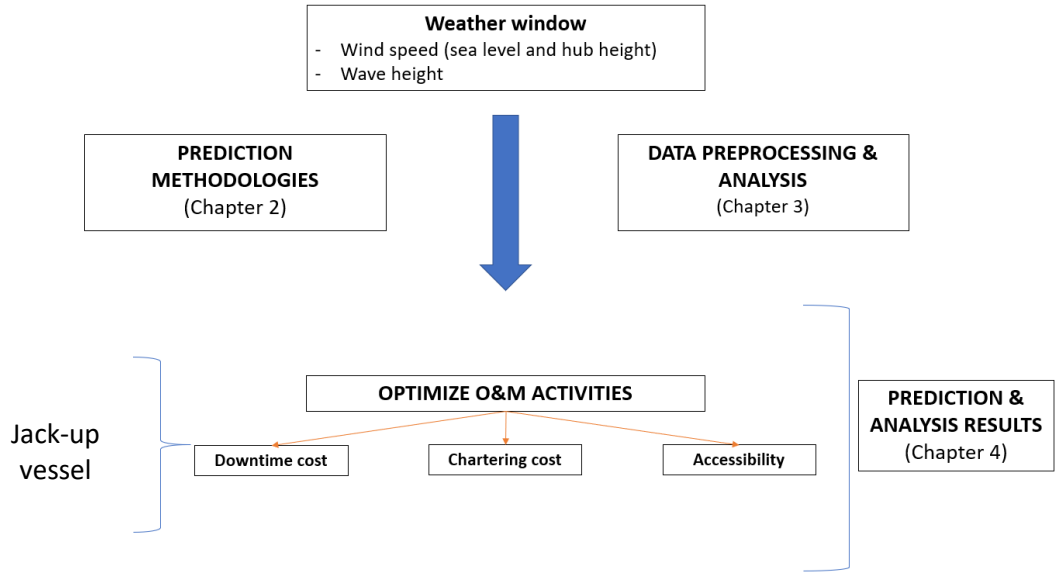


Figure 5: Layout of the thesis

2. LITERATURE REVIEW AND METHODOLOGY

In this chapter, section 2.1 consists of the various literature sources about current state-of-the-art offshore wind farm O&M strategies involving short-term and long-term wind speed, wave height prediction, and forecasting methods and section 2.2 presents selected methodologies for the wind speed and wave height forecasting techniques. To do so, we search for papers with key terminologies like “off-shore wind farm maintenance,” “off-shore wind farm accessibility,” “long-term wind speed prediction,” “short-term wind speed prediction,” and “short-term wave height prediction,” “jack-up vessel planning”.

2.1. Literature review

2.1.1. Offshore wind farm operation and maintenance literature review.

2.1.1.1. Some regulations, guidelines, and safety standards concerned in the thesis

Regulations are defined as administrative legislation at the national or state level which governs the rights and responsibilities (Sirnivas et al., 2012). Standards are documents established from best practice, research, and know-how and are applied by the agreement of the stakeholders. A standard illustrates how a product is to be designed, constructed, examined and operated (Sirnivas et al., 2012). In addition, guidelines are suggested practice documents that are not subjected to a formal protocol or vote of constituencies. They are developed by classification societies and only depend on the internal quality process and peer review of the originating society. Guidelines consist of proposed, nonmandatory administration that assists standard or act as references when no applicable standard is in place. They are strongly recommended in some cases (Sirnivas et al., 2012).

In this thesis, we concentrate on the regulations and safety standards of the weather conditions (wind speed and wave height) which ensure the normal and safe operation of the wind turbine, jack-up vessel, and the maintenance crew. Therefore, we based on the safety standards of the turbine and jack-up vessel’s technical design and the Guidelines for the Safe Management and Operation of Offshore Support Vessels to make the decision on vessel planning and scheduling in this thesis (UK offshore Operators Association/ Chamber of Shipping, 2002)

2.1.1.2. Optimal operation & maintenance policies for offshore wind farm

Schouten (2019) implemented long-term prediction for maintenance planning however, they did not measure in detail the option to delay the maintenance for better weather days. If the postponement of the maintenance help to generate more energy and offset the cost of this delay, the O&M operator can opt for this option. In order to do so, a short-term prediction model of wind speed is also necessary. Byon, Ntaimo, and Ding (2010) introduced stochastic weather limits for preventive maintenance. Besnard et al., (2009) applied wind speed forecasting and wave height correlation in determining the optimal schedule to conduct preventive and corrective maintenance. However, they did not use wave height prediction to assess the accessibility of the wind farm.

2.1.1.3. Jack-up vessel planning and optimization

Regards jack-up vessel planning, researchers build models of wind speed and wave height weather window to assess the downtime cost and accessibility separately or combine them without considering the restriction of the jack-up vessel. Broek et al (2019) created a simulation model with stochastic processes of the weather pattern and component failures to evaluate the resource sharing strategy for offshore wind farm maintenance. However, in their model, wind speed is only estimated to calculate the downtime cost of the wind turbine and not for scheduling the jack-up vessel. In addition, wave height prediction is not used to support vessel planning. Zhang et al (2019) programmed stochastically a dynamic opportunistic maintenance model related to the wind speed. These authors did not consider the weather restriction of lifting activity for the jack-up vessel.

Dalgic et al., (2015) optimized the jack-up vessel chartering strategy for offshore wind farm O&M activities. They also included wave height and wind speed prediction into their simulation to calculate wind power, weather windows, and repair time, however, they did not describe explicitly their prediction model and the trade-off between the flexibility of the vessel chartering period and the demurrage rate within a short-term period of 1 year (52 weeks).

2.1.1.4. Accessibility of offshore wind farm

The accessibility to the offshore wind farm, in reality, depends on a combination of factors such as the characteristics of the vessel such as the motion compensating capabilities of the access system, the size and hull design of the vessel on which it is mounted, the location of the vessel, and environmental conditions (Hu and Yung (2020). In the offshore industry, waiting time due to weather constraints approximately accounts for 89.4% of the downtime cost. Unfortunately, there is not yet a standard method to define the accessibility that allows comparison between different access systems; thus, this thesis uses figures based on data provided by the manufacturers or designers. In addition, this thesis also assumes the stylized condition in which all accessibility conditions of the offshore wind farm are satisfied except for the restriction of wind speed and the significant wave height per type of vessel.

Regarding the wave height, it affects the lifting of the turbine's main component, transporting of the maintenance crew, and other vessels–related costs accounting for approximately 45% of the total O&M cost (Sperstad *et al.*, 2017). As a consequence, a weather window which combines these two predictions could strongly support the optimal planning of the maintenance service and scheduling of the vessels.

After reviewing scientific articles about offshore wind farms' O&M optimization strategy, research about the combination of wind speed and wave height forecast to optimize O&M activities of offshore wind farms is being conducted with a variety of approaches and methods but mostly either to find the optimal vessel fleet composition which is based on the cost of the vessel type or evaluate the accessibility of the wind farm under certain weather conditions or minimizing downtime cost due to the interruption of wind power generation. Lazakis and Khan (2021) took into account climate and weather restrictions in their optimization framework for scheduling and routing the vessels but the wind speed is used for safety conditions but not for calculating the downtime cost. Therefore, the combination of short-term wind speed prediction and wave height prediction results in a more holistic cost model based on the power output with realistic weather restrictions for the safety of the maintenance crew and vessel type.

The goal of this thesis is to contribute to the recent literature on this field in the following ways. Firstly, it contributes to the (Schouten, 2019) research of wind farm O&M optimization strategy with short-term wind speed prediction and related downtime cost. Secondly, it provides more holistic scheduling with both wind speed and wave height for the lifting activities and mobilizing of the jack-up vessel. Thirdly, the thesis contributes to the planning of a jack-up vessel for the offshore wind farm by evaluating its accessibility. Last but not least, the thesis applies the most updated data on the downtime cost of the Dutch wind energy market in estimation and modeling.

2.1.2. Weather forecasting methodologies

In this section, for our problem, we wish to forecast the wind speed both in the long and short-term and the wave height prediction in the short term. Based on Lawan et al., (2014), short-term and long-term wind speed prediction is defined by the forecasting horizon. The short-term forecasting horizon is from 30 minutes to 6 hours ahead. The medium-term forecasting horizon is from 6 hours to 1 day ahead, and the long-term forecasting horizon is 1 day to 1 week ahead. However, with current developments in wind speed and wind power forecasting techniques, the forecasting horizon is enlarged per each type of forecasting. In this thesis, we defined long-term forecasting as the forecasting of average daily mean wind speed and average wind power one year ahead, and the short-term prediction horizon was defined as around one hour to 24 hours ahead.

Based on Foley et al., (2012), regarding the difference between long-term and short-term prediction, long-term wind forecasting is implemented for the major strategic decision-making process by describing the seasonality, trends, and repeated pattern in wind speed variation. In the light of these trends and patterns, the wind farm operator can outline the overall O&M plan and operation. Therefore, for long–term wind speed prediction, we are more interested in the distribution of the average daily mean wind speed and average daily wind power. However, the further the future, the more vague and less accurate the long-term prediction is, therefore, the drawback of long-term

prediction is vague with wider fluctuation thus requiring short-term prediction for more detailed and tactical execution. Short-term forecasting applies more current and past data to predict the near future with more certainty and less volatility. As the result of more detailed predictions, the wind farm operator can prepare for more scenarios while implementing the planning and scheduling of O&M activities.

In terms of forecasting methods, there are two approaches: prediction methods and optimizing methods.

In terms of prediction methods, they are physical models, statistical models, spatial correlation models, artificial intelligence, and deep learning models. Physical models apply numerical physical parameters like temperatures and atmospheric pressure as predictors and are normally used for long-term prediction, however, they require a large amount of time and data with computation intensity (Wang *et al.*, 2015; Wang and Li, 2016; Allen *et al.*, 2017; Galanis, Papageorgiou and Liakatas, 2017; González-Aparicio *et al.*, 2017). Spatial correlation models struggled to implement perfect wind speed prediction due to the enormous amount of information (Tascikaraoglu *et al.*, 2016). Therefore, these two types of methods are not used in the thesis due to the limitation of time, data, and computation resources.

Statistical models try to characterize the inherent uncertainty of the weather data in which the Gaussian random process is a solid foundation. The Gaussian random process provides marginal modeling in terms of mean and covariance functions, thorough research on time series data, and scalable for large data (Heaton *et al.*, 2019). However, this type of method suffers from the non-linear trend, seasonality, and high noise inherently in weather data like wind speed and wave height. The artificial intelligence and deep learning methods successfully capture the non-linear relationship of the weather data for accurate short-term prediction. These two types of forecasting methods are more relevant to our program so the thesis will apply these types of predictions to predict the wind speed and wave height for O&M planning of the wind farm.

The second approach is parameter optimization methods which include many technologies to optimize the hyperparameter for a single prediction model or the weight parameters of the multiple prediction models. These methods can increase the accuracy, stability, and computation speed of the prediction models; however, this is not the focus of our thesis, and they are often more complex, which require extensive programming expertise and time. Therefore, in this thesis, optimization methods are not considered.

2.1.2.1. Long-term wind speed prediction.

For long-term wind speed prediction, the thesis aims to find the method which can capture the multi-seasonality of the average daily mean wind speed time series to find the favorable weather windows to schedule vessels. This time series is not normally distributed and not stationary.

Jamaludin et al., 2016 compared between conventional Auto-Regressive Moving Average (ARMA) model with the Fourier-ARMA model in modeling and forecasting the wind speed data in Malaysia. It was found that the Fourier – ARMA model was the best-fitted model to forecast one year in time. This method is called Dynamic Harmonic Regression which uses Fourier terms to capture the seasonality of the data (Rob J Hyndman, 2021). Since the thesis focuses on the average daily mean wind speed prediction of Gemini wind farm only and the annual seasonal pattern of this time series is stable over time, we can assume that the seasonal pattern does not change over time to use the Fourier – Autoregressive Integrated Moving Average (ARIMA) model to forecast the daily wind speed with the multi-seasonality and the large frequency of the data is 365.25 as in Young, Pedregal and Tych, 1999 and Ludlow and Enders, 2000. In addition, this forecasting method can be implemented by the R package fpp3 with function forecast() and command xreg = fourier() (Rob Hyndman, 2021).

On the other hand, to benchmark the prediction accuracy, computational time, and the ability to characterize the multi-seasonality of the wind speed, we will use the seasonal Naïve model as a simple benchmark (Rob J Hyndman, 2021). In this method, each forecast is the last observation of the same season in the past.

2.1.2.2. Short-term wave height prediction.

Short-term wave height prediction is implemented in this thesis to ensure the planning of O&M activities meets the safety standard and sea-keeping specifications of the vessels. The short-term prediction horizon was defined as around one hour to 48 hours ahead in this thesis.

Currently, many machine learning methods are being developed to predict significant wave height and most of them involve a kind of neural network model. Huang and Dong (2021) improved the complete ensemble empirical mode decomposition algorithm and applied recurrence quantification analysis to separate the original time series into deterministic and stochastic components. Then they predicted each decomposed by the long short-term memory network. Finally, they integrated the deterministic and stochastic predictions to arrive at the final significant wave height prediction. This hybrid model was better than the two separated models. It can describe both non-stationarity and anisotropy with advantageous computational properties, and easily interpretable parameters. Miky et al., (2021) proposed a Recurrent – Cascade - Neural network – Nonlinear Autoregressive with exogenous inputs to forecast the wave height based on the wave characteristics. This method is robust to hourly and daily significant wave height time series. Kumar, Savitha, and Al Mamun (2018) experienced an ensemble of Extreme Learning Machine (ENS-ELM) to predict the daily wave height and proved that its performance was better than Support Vector Regression SVR (Support Vector Regression)). Berbić et al (2017) applied neural network (ANN) and support vector machine (SVM) in real-time wave height prediction. The prediction accuracy decreases as the period for the forecast increases, however, ANN and SVM were both accurate in some cases and ANN is slightly more accurate for the next 0.5 to 5.5 hours. Prahlada and Deka (2015) implemented a wavelet decomposed neural network model to forecast wave height time series for lead time up to 48 hours. This hybrid model was proved to perform better than the single machine learning model. Therefore, methods involving artificial neural networks seem to demonstrate the efficiency of neural networks in predicting the short-term significant wave height, however, they are complicated with the lack of interpretability.

Given the scope of the thesis and the complex seasonality of the wave height time series data, Zubier (2020) included wind speed, wind shear velocity, and differences between wind and wave direction as exogenous variables for the neural net autoregressive model. These exogenous variables help to increase the accuracy of the prediction model. Among artificial neural network types, the feed-forward neural network autoregressive model could bridge the gap between deep learning and traditional time-series modeling with its fast computing and interpretability (Triebe, Laptev, and Rajagopal, 2019). Therefore, we applied a feed-forward neural net autoregression model to our significant wave height time series data based on the instruction in Forecasting Principle and Practice (Hyndman, 2021) with the support of the fpp3 package in R (Rob Hyndman, 2021).

2.1.2.3. Short-term wind speed and wind energy prediction

In this thesis, short-term wind speed prediction is conducted to calculate the detailed wind power generated and downtime cost caused by the O&M activities, thus providing the optimal scheduling for the vessels. Currently, under the development of data science, machine learning methods for weather forecasts prove to provide accurate hourly and daily predictions. Specifically, deep learning and hybrid models in wind speed forecasting are being researched intensively for the application of time-series data. With regards to the deep learning approach, they have been widely applied to classification and regression thanks to their robustness to non-linearity of real-world data and ability to automatically extract hidden data. Current deep learning methods are Convolutional Network (CNN) (Lecun, Bengio, and Hinton, 2015), Deep Belief Network (DBN) (Hinton and Osindero, no date), generalized denoising auto-encoders (Bengio *et al.*, 2013), and Long-short Term Memory (LSTM) (Hu and Chen, 2018).

In the field of wind speed prediction, hybrid models based on deep learning approaches are gaining momentum since they can deal with the non-linear features of the wind speed time series by applying different data preprocessing before feeding into the forecasting model. The most current techniques in this approach are the following: Neshat et al (2021) suggested a hybrid model of a bidirectional long short-term memory neural network, a hierarchical evolutionary decomposition

technique, and an improved generalized normal distribution optimization algorithm for hyper-parameter tuning. This hybrid model can derive the non-linearity features of wind speed data. The parallel bi-directional long-term memory model outperformed the recurrent model in both whole and detailed sub-series prediction. The combination of a generalized normal distribution algorithm, a downhill local search, and a further mutation is robust and effective in adjusting the hyper-parameters of recurrent deep learning models. In addition, categorizing the wind speed time series into four seasons enhanced the performance of their models sometimes. Duan et al (2021) developed a hybrid prediction system which consists of Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), Recurrent Neural Network (RNN), ARIMA (Auto-Regressive Integrated Moving Average), and error correction method. First of all, the ICEEMDAN decomposed the nonlinear and non-stationary wind speed data into a series of more simple subseries. Secondly, the RNN model is used to forecast each subseries. Then, the predicted wind speed and error are retrieved. Next, the ICEEMDAN decomposed the error and the ARIMA predicted the error subsequence to obtain the prediction error. The final wind speed prediction was achieved by aggregating the previously predicted wind speed and the current predicted error. Daniel et al (2020) compared artificial neural networks (ANN) trained with Bayesian regularization, decision trees based on stochastic gradient boosting (SGB), and generalized additive models (GAMs). The results of this research proposed that ANN outperformed other methods based on root mean square error (RMSE). In contrast, SGB was better in terms of mean average error (MAE) and the related mean average percentage error (MAPE). However, these methods are too complicated and often take many hours to train the model. In their research, Spyros and colleagues pointed out that LSTM might over-parametrize by multiple magnitudes, which made it meaningless to compare with simpler statistical models (Spyros Makridakis, 2018). In addition, there are similar works that raised the need for proper and understandable deep–learning methods for time series modeling (Deepak A. Kaji, 2019), (Spyros Makridakis, 2018). Therefore, we did not apply them to our thesis.

Given the multivariate wind speed and wind energy time series data set at the hourly level, Blanchard and Samanta (2020) constructed two nonlinear autoregressive neural network models with and without exogenous inputs (NARX and NAR respectively). They showed that both models were superior to the baseline persistence model for 5 hours ahead forecasting, however, NARX was slightly more accurate with fewer data required than NAR. NARX with temperature as exogenous inputs was recommended if you had more weather indices at hand, NAR was more suitable when you only had the wind speed time series. NARX better performed with less than 28-hour lag. In addition, the higher the number of neurons is, the worse the model is due to the optimization process of the local optima in the network parameters. They suggested the number of hidden layer neurons varied between 2 and 3 (NAR) to 3 and 6 (NARX). Noman et al., 2021 also proposed a multistep wind speed prediction by using a Non-linear Autoregressive Neural network with Exogenous inputs thanks to its best accuracy despite limited prior knowledge of the wind speed history. Therefore, we applied Non-linear Autoregressive Neural Network with and without Exogenous inputs to predict the short-term wind speed under the instruction of Forecasting Principle and Practice, 3rd edition (Hyndman, 2021).

Regarding wind energy prediction based on wind speed, Kumar et al., (2017) proposed a 3-layer artificial neural network to extract the dependency of the variable in the training process by showing a non-linear complex relationship of the time series. They used air density, temperature, and wind direction to forecast the wind speed before applying predicted wind speed to forecast the wind energy. In addition, a feed-forward neural network with a hidden layer of 15 was used to forecast wind daily average wind energy (Cristian-Dragos Dumitru & Adrian Grigor, 2017). This model could capture the evolution of wind energy considering its history and the average wind speed of the area under the study. Therefore, inspired by these studies, we also applied Non-linear Autoregressive Neural networks with and without Exogenous inputs to predict wind energy in short term under the instruction of Forecasting Principle and Practice, 3rd edition (Hyndman, 2021).

2.2. Methodology

In this section, we introduce the prediction methods which are applied to forecast the wind speed and wave height for O&M activities. Firstly, a general description of forecasting methods is provided, after that, each method applied in the thesis is discussed in more detail.

2.2.1. Long-term wind speed prediction

In terms of long-term wind speed prediction, we applied the average daily mean wind speed time series from 2014 to 2019. This time series is slightly decreasing with multi-seasonality features whose annual seasonality is stable. In addition, this data set is not stationary with a non-zero mean and volatile variance.

2.2.1.1. Dynamic Harmonic Regression

The Dynamic Harmonic Regression of Unobserved Components (UC) type takes advantage of the Fourier terms to capture the seasonal pattern, trend, cyclical and irregular components of the data while applying the non–seasonal Autoregression Integrate Moving Average (ARIMA) (Fourier – ARIMA) to model the error terms (Young, Pedregal, and Tych, 1999). It applies Kalman Filter and Fixed Interval Smoothing algorithms to approximate recursively the unobserved components while optimizing the hyperparameters by the cost function. This cost function is not based on Maximum Likelihood but on the difference between the logarithmic autoregressive spectrum of the data and the logarithmic pseudo-spectrum of the model. Here, the Fourier terms are the pairs of sine and cosine defined as follows to capture the seasonality in the data:

$$S_k(t) = sine(\frac{2\pi kt}{m}) \text{ and } C_k(t) = cosine(\frac{2\pi kt}{m}) \quad (1)$$

The forecasting function of the weather data is written as follows:

$$y_t = \beta_0 + \sum_{k=1}^{K}[\alpha_k S_k(t) + \gamma_k C_k(t)] + \varepsilon_t \quad (2)$$

Here the m is the seasonal period length, β_0 is the constant, $S_k(t)$ and $C_k(t)$ is the k^{th} Fourier terms at time t, while K is the upper limit of how many pairs of Fourier terms should be included in the model, and y_t is the weather data time series. The Fourier terms' frequency increases when the number of pairs K increases. α_k and γ_k are the regression coefficients of the k^{th} Fourier terms. Since the seasonality of the time series is modeled by these Fourier terms, e_t which is the error terms of time t, is modeled by the non-seasonal Autoregressive Integrated Moving Average model.

This Fourier – ARIMA model assumes that the seasonal pattern does not change over time which is suitable for the average daily mean wind speed since its yearly seasonal components have the same pattern over time as in Figure 22. It is also the strongest seasonal component in the time series. The only concern is the weekly seasonal components fluctuate more wildly as in Figure 22, which may affect the accuracy of the model.

One more advantage of Fourier – ARIMA over the Seasonal ARIMA and Exponential Smoothing methods with Trends and Seasonality (ETS) is that Fourier – ARIMA can be constructed with any seasonal period length, especially for large period time series as daily time series (Rob J Hyndman, 2021).

Thirdly, since the average daily mean wind speed time series has both weekly and yearly seasonal components, Fourier – ARIMA can develop different Fourier terms for different seasonal periods separately. This enables the model to handle the multi-seasonality nature of our data set. In addition, the smaller the K is, the smoother the seasonal pattern. In addition, the value of K is not larger than half of the seasonal period length m. Therefore, we can choose K based on the given selection criteria which will be discussed in section 2.2.

Finally, although Seasonal ARIMA allows the seasonality of the time series to change over the period under the study, in this case, the annual seasonal pattern is substantially constant. In addition, because the annual and weekly seasonal patterns are not smooth, it does not make sense to make the seasonal differencing of one year, or in another word, compare the current value with the observation one year ago.

2.2.1.2. Seasonal Naïve prediction

Seasonal Naïve prediction is a simple benchmarking forecast which takes the last observation of the same season in the past as the forecast. For example, the prediction of all January is equal to the last value observed in January (Rob J Hyndman, 2021).

2.2.2. Non-linear Autoregressive Neural Network with and without Exogenous input (NARX)

Regarding the short-term prediction method for wind speed, wind power, and sea state, we applied a Non-linear Auto-regressive Neural network with Exogenous inputs (NARX) to capture the non-linear relationship between the time series of the predictors and response variable without prior assumptions. There are, in general, many advantages of artificial neural network methods (Alberto Pliego Marugán, 2018). Firstly, they are adaptive learning via the training process and construct their structure to represent the data. Secondly, they are robust to missing values or damaged structure in the data. Thirdly, they can be computed in parallel and online to accelerate the speed of the model. Finally, there are special chips designed for neural networks to integrate into the system.

The Non-linear Auto-regressive Neural Network with Exogenous inputs is the neural network structure that resembles the brain structure and nodes are organized in different layers. The top layer is the output of the prediction, and the bottom layer is the set of input values which are the time lag of the response and independent variables. In between these two layers are intermediate layers containing "hidden nodes" or "hidden neurons" (Rob J Hyndman, 2021). In R, we used the fable package (Mitchell O'Hara-Wild, 2021) which applies the multilayer feed-forward network with 1 layer of hidden nodes. In this neural network, lagged values of the response and exogenous variables are used as predictors in the input layer. In addition, the previous layer of nodes creates input for the next layer of nodes.

In this neural network structure, the "weights" are the coefficients attached to these predictors. The weights are selected in the neural network framework using a "learning algorithm" that minimizes a "cost function" such as the Mean Squared Error (MSE) or Mean Absolute Percentage Error(MAPE). In the hidden layer, we apply a weighted linear combination to the inputs before transferring them to each hidden node. For example, the weighted linear combination is (Rob J Hyndman, 2021):

$$z_j = b_j + \sum_{i=1}^{n} w_{i,j} x_i \tag{3}$$

We then adjust the results by non-linear function like the sigmoid function below before being the output or being transferred to the next layer (Rob J Hyndman, 2021):

$$s(z) = \frac{1}{1+e^{-z}} \tag{4}$$

This process tends to minimize the impact of extreme input values, thus making neural networks robust to outliers. Here, z_j is the result of layer j, n is the number of nodes in layer j. b_j and $w_{i,j}$ (weights) are parameters that are estimated in the data. We restrict the weights from being too large by the "decay parameter". We take a random value of the weights to start the process and then update them using the data. This also creates randomness in the data; therefore, we train the network several times using different random starting points and then average out the results (Rob J Hyndman, 2021). The Neural Non-linear Auto-regressive Neural Network without Exogenous inputs shares the same structure, however, we do not feed the exogeneous inputs in the model.

In this thesis, we denote the Non-linear Auto-Regressive Neural network with Exogenous inputs as $NARX(p, P, k)_m$ and Non-linear Auto-Regressive Neural network as $NAR(p, P, k)_m$ to indicate p lagged inputs, k neurons in the hidden layer, P is the number of seasonal lagged inputs, m is the seasonal period. The fable package automatically chooses the value of p, P, k with the default value of P is 1, p is selected from the optimal linear model fitted to the seasonally adjusted data, and $k = \frac{p+P+1}{2}$ (rounded to the nearest integer). For multi-step-ahead forecasting, we use the one-step ahead and historical data as inputs and continue this process until we complete the forecast horizon (Mitchell O'Hara-Wild, 2021).

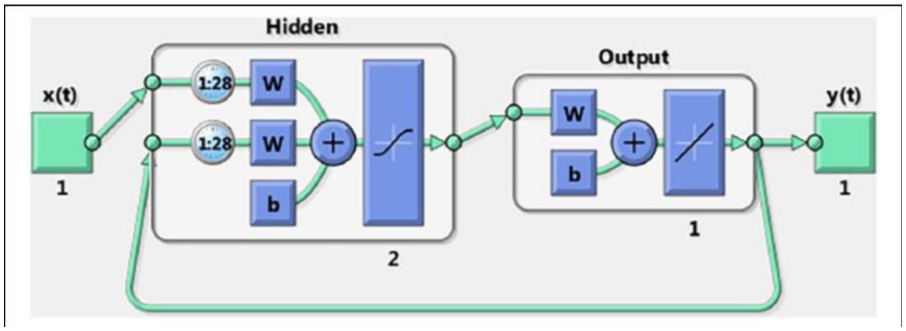


Figure 6: Neural network structure for a non-linear autoregressive model with exogeneous input for the multi-step-ahead forecast. Source: Blanchard and Samanta (2020)

Since a neural network is not created from the well-defined stochastic model, we could not derive the prediction interval from the forecasted results. However, we can apply simulation of multiple future sample paths based on bootstrapped residuals to calculate the prediction intervals. Here, we assume that the errors are homoskedasticity and normally distributed, then, for each value of errors randomly drawn out from historical values, we generate one possible future sample-path drawn from the forecast distribution. By repeating this process, we construct the information of all the future values based on the fitted neural network (Hyndman, 2021).

Since the distributions of the hourly wind speed and wind power time series are not normal and there were zero values in them, we could not apply Box-Cox transformation to these data sets. In addition, the fable package does not integrate the extension of the Box-Cox transformation for non-positive values, therefore, we scale the input data by subtracting the column means and dividing by their respective standard deviations.

The limitation of the neural network forecasting method is its lack of interpretability and the forecasting horizon. A neural network is a type of black-box model which requires further analysis algorithm to explain why the model produces such results. Unfortunately, there is no package of black box interpretation available for the NARX model run by the fable package in R. Secondly, since the forecasting horizon of short-term NARX might be less than 48 step-ahead, therefore, we applied NARX to mean wind speed and average power at hourly prediction with the forecast horizon of 24 hours only. Moreover, the results of NARX and NAR are not stable so we have to set the same sample before running the model.

2.2.3. Model validation and selection criteria

2.2.3.1. Ljung – Box test for residual diagnostics

Fitted values after Box-Cox transformation at time t: $\widehat{yb_t}$ is a forecasted value at time t using all previous observations of the Box-Cox transformed time series.

Innovation residual at time t is the difference between the real observation of the time series after the Box-Cox transformation (yb_t) and its predicted value at time t (Rob J Hyndman, 2021):

$$IR_t = yb_t - \widehat{yb_t} \quad (5)$$

There are two essential criteria of a good forecasting model in terms of innovation residuals (Rob J Hyndman, 2021). Firstly, the innovation residuals should not be correlated with each other or there should not be any pattern in the residual graph. Because uncorrelation between residuals shows that the model captures all the information of the data set.

Secondly, the innovation residual should have zero mean, or else the forecast is biased.

These two criteria are important to assess the ability to capture all information of the original data set for prediction. If they are not satisfied, then the forecasting models can still be improved. However, these conditions are not meant for selecting the best forecasting method since there might be many prediction models satisfying them at the same time. In addition, these qualified models can also be improved.

One more thing to note is that there are two more properties to support the evaluation of the prediction intervals. Firstly, are the innovation residuals homoskedastic, or do they have a constant variance? Secondly, is the distribution of the innovation residuals normal? (Rob J Hyndman, 2021).

To formally evaluate the autocorrelation function between residuals, Ljung – The box test is carried out (G. M. Ljung, G. E. P. Box, 1978) based on:

$$Q^* = B(B+2)\sum_{k=1}^{l}(B-k)^{-1}AR_k^2 \quad (6)$$

Here, B is the number of observations in the time series, AR_k is the autocorrelation for lag k, and l is the first l autocorrelation that we use to test. This Q^* statistic follows a χ^2 distribution with $(l-K)$ degree of freedom where K is the number of parameters in the model. Then we conduct a hypothesis testing with the null hypothesis *(*H_0*)* and the alternative hypothesis (H_a) as follow:

$$\begin{cases} H_0: & The\ residuals\ are\ white\ noise\ series\ or\ random \\ H_a: & The\ residuals\ are\ not\ white\ noise\ series\ or\ not\ random \end{cases} \quad (7)$$

If the result of this test is statistically significant (p-value < 0.05), there is enough evidence to reject H_0 and the residuals are correlated. In contrast, if this test is not statistically significant (p-value $\geq$ 0.05), then there is not enough evidence to reject H_0 or the residuals are not different from a white noise series or random series.

2.2.3.2. Corrected Akaike's Information Criterion (AICc)

Prediction accuracy is an important criterion to select a forecasting model, however, different models with a variable number of predictors or parameters can achieve the same desired accuracy. Therefore, we need to find a parsimonious prediction model which achieved a targeted accuracy with the least predictors. Therefore, the Corrected Akaike's Information Criterion (AICc) (Hurvich and Tsai, 1989) will also be used to evaluate different models constructed in this thesis. It is defined as:

$$\text{AICc} = T log\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + 2 \times (k+2) + \frac{2\times(k+2)\times(k+3)}{T-k-3} \quad (8)$$

Here, T is the observation number used to construct the model, k is the predictor number in the model. In this way, the AICc penalizes the sum of squared error with the predictor number T. Therefore, the smaller the AICc value is, the more parsimonious the model is.

2.2.3.3. Evaluation of the distributional forecast accuracy

2.2.3.3.1. Winkler score

To measure the accuracy of the prediction interval for the wind speed, we apply the Winkler score to check how far one actual observation lies within or outside of the prediction interval (Winkler, 1972). If the $100(1-\alpha)\%$ prediction interval at time t is defined by $[l_{\alpha,t}; h_{\alpha,t}]$, the Winkler score is the interval length with the penalty for the observation which is outside the interval:

$$\begin{cases} W_{\alpha,t} = (h_{\alpha,t} - l_{\alpha,t}) + \frac{2}{\alpha}(l_{\alpha,t} - y_t) & if\ y_t < l_{\alpha,t} \\ W_{\alpha,t} = (h_{\alpha,t} - l_{\alpha,t}) & if\ l_{\alpha,t} \leq y_t \leq h_{\alpha,t} \\ W_{\alpha,t} = (h_{\alpha,t} - l_{\alpha,t}) + \frac{2}{\alpha}(y_t - h_{\alpha,t}) & if\ y_t > l_{\alpha,t} \end{cases} \quad (9)$$

We can see that if the observation lies in between the prediction interval, the Winkler score is the interval length, and it is small. Whereas, if the observation falls outside the interval, the penalty will be proportional to the distance between the observation and the interval, which makes the Winkler score larger.

2.2.3.3.2. Continuous Rank Probability Score (CRPS)

To establish a weather window of the average daily wind speed, we need to evaluate the accuracy of the whole forecasting distribution. Therefore, we can apply the Continuous Ranked Probability Score (CRPS) proposed by Gneiting and Katzfuss, 2014.

First, we calculate the quantile forecast $f_{p,t}$ of probability p at time t. If we denote y_t as an observation at time t, then with the 20th percentile, we have $f_{0.2,t}$ and y_t is expected to be less than $f_{0.2,t}$ with the probability of 20%. Then the Quantile Score $Q_{p,t}$ is defined as:

$$Q_{p,t} = \begin{cases} 2 \times (1-p)(f_{p,t} - y_t) \; if \; y_t < f_{p,t} \\ 2 \times p(y_t - f_{p,t}) \qquad if \; y_t \geq f_{p,t} \end{cases} \tag{10}$$

The Quantile score $Q_{p,t}$ can be considered as an absolute error if p = 0.5. If p > 0.5 then $Q_{p,t}$ will place a heavier penalty to $y_t > f_{p,t}$ than when $y_t < f_{p,t}$ and vice versa.

The CRPS is an average of $Q_{p,t}$ over all values of p therefore, it takes the average for each of the predictive densities and the corresponding observation over the period under the study. The CRPS and Mean Absolute Error (MAE) (for deterministic forecasting) are directedly comparable since CRPS is the generalization of MAE with a given lead time and nominal/normalized scale. Given the same lead time, nominal scale, and data set, if a probabilistic forecast has a CRPS value of 7%, it is better than the deterministic forecast with the MAE value of 8%. The lower the CRPS, the better the prediction distribution.

2.2.3.4. Evaluation of the point forecast accuracy

2.2.3.4.1. Root Mean Squared Error (RMSE)

Since the prediction methods used in this paper will automatically transform the result back to its original scale and within each method, we also tune the different parameters to find the most optimal model, a scale-dependent error measurement is necessary. In addition, the same time series is used across different models and different types of prediction methods. Therefore, we use Root Mean Squared Error (RMSE) to measure the accuracy of the forecasting model (Rob J Hyndman, 2021).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_t - \widehat{y_t})^2} \tag{11}$$

2.2.3.4.2. Mean Absolute Percentage Error (MAPE)

To give a better perspective of how large an error is relative to the real observation value, we used Mean Absolute Percentage Error (MAPE). This measurement emphasizes the errors which account for large percentages of the actual observations. MAPE can also be applied to evaluate how well the model fits with different data sets:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|100 \times \frac{y_t - \widehat{y_t}}{y_t}\right| \tag{12}$$

% Accuracy = 100% - MAPE

The % Accuracy of over 70% is desirable. MAPE is scale-independent, however, there are two constraints of MAPE. Firstly, if y_t is 0, MAPE will be undefined and if y_t is close to 0, MAPE will be extremely large for any t period. Secondly, MAPE requires a meaningful error for the unit of measurements (Hyndman and Koehler, 2006). The less MAPE is, the better the model is. In this thesis, there is no day without wind so the average daily mean wind speed can use MAPE to measure the point forecast accuracy. However, for the hourly mean wind speed and hourly significant wave height, there will be many observations of zero, therefore, MAPE will not be applied for them.

2.2.3.4.3. Forecast evaluation on a rolling origin – Time-series Cross-Validation

Forecast evaluation on a rolling origin or Time series cross-validation is a procedure in which the training set consists of all observations that occurred before a series of test sets, while the test sets are formed by only one observation in the time series. In this way, the origin of the forecast rolls forward in time (Rob J Hyndman, 2021).

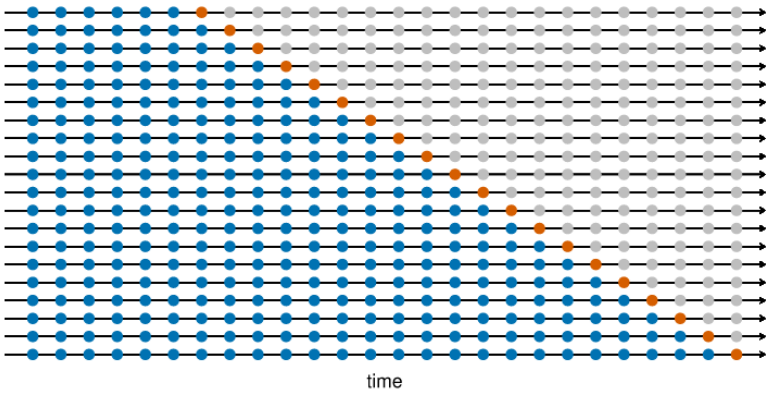


Figure 7: Illustration of Timeseries CV (Blue dots are training set, the orange dot is test set) *(Rob J Hyndman, 2021)*

The forecast accuracy is calculated by taking the average of error measurements like RMSE or MAPE over the test sets. Therefore, the best model will have the lowest error. This procedure can be modified to allow multi-step ahead prediction. This procedure is used to compare different models with the same time series.

2.2.4. Data decomposition method – Seasonal and Trend decomposition using Loess (STL)

After several reviews of wind speed and wave height prediction methods, we saw that the weather indices time series were reported with multiple seasonality, non – stationary, and non-linear trends in data Heaton et al., 2019 Huang and Dong, 2021 for example. In addition, specifying the type of seasonality, trend, and their impact on the existing time series is crucial to capture the variance of the data in the statistical modeling beside visual diagnosis. Therefore, in current research, for more accurate quantitative analysis to discover the trend and seasonality of the chosen weather time series, "Seasonal and Trend decomposition using Loess,"(Cleveland *et al.*, 1990) or STL for short was applied to the time series before constructing prediction model. Because STL is a filtering process that can evaluate and decompose the time series into trend, the multi-seasonality, and the remaining components robustly with different granularities and missing values, it was applied to wind speed time series as in Laib et al., 2018, Guignard et al., 2019. For significant wave height time series, Yang et al., (2021) also applied STL decomposition to the wave height times series before using a convolutional neural network for forecasting. This decomposition method was also applied to significant wave height data in Ramachandra, (2019) and Rodolfo Piscopia, (2015).

Since the λ values of the Box-Cox transformation for weather data in the Table 10 are between 0 and 1 and the time series illustrate a stable magnitude of seasonality in Figures 9, 10, 19, we can assume that the data has an additive decomposition where X_t is the original wind speed time series, $Season_t$ are the seasonal components, $Trend_t$ is the trend components, and $Remainder_t$ is the remainder components (Rob Hyndman, 2021). The weather indices time series can be rewritten as follow:

$$X_t = Season_t + Trend_t + Remainder_t \quad (13)$$

Seasonally adjusted data is the time series after eliminating the seasonal components, which are:

$$X_t - Season_t \quad (14)$$

Detrended data is the time series after eradicating the trend components, which is:

$$X_t - Trend_t \quad (15)$$

STL decomposition is also useful in measuring the strength of trend and seasonality of the data (Xiaozhe Wang, 2006). Trend strength is calculated as:

$$Strength\_trend = max(0, 1 - \frac{Var(Remainder_t)}{Var(Remainder_t + Trend_t)}) \quad (16)$$

since if the time series is strongly trended, the variance of seasonally adjusted data is much stronger than the variance of the remainder component itself. This makes $\frac{Var(Remainder_t)}{Var(Remainder_t+Trend_t)}$ smaller and closer to 0, which turns $Strength\_trend$ closer to 1. In contrast, if the trend components are not that strong, the two variances should be approximately similar. Therefore, $Strength\_trend$ will be closer to 0.

Season strength is defined as:

$$Strength\_season = max(0,1 - \frac{Var(Remainder_t)}{Var(Remainder_t+Season_t)}) \quad (17)$$

since if the time series has a strong seasonal component, the variance of detrended data is much stronger than the variance of the remainder component itself. This makes $\frac{Var(Remainder_t)}{Var(Remainder_t+Season_t)}$ closer to 0, which turns $Strength\_season$ closer to 1 and vice versa.

STL decomposition can be conducted by using function model() with STL command in fpp3 package in R (Rob Hyndman, 2021). The formulas of component strength for the decomposed time series with more than two seasonal components are described in the appendix.

The advantage of STL decomposition is that it can automatically capture and decompose the time series without prior assumptions of trend and seasonality. Therefore, we can use it to recognize the multi-seasonality in data without any bias. Thus, it supports the Fourier prediction in defining the multi-seasonality of the time series and selecting the number of Fourier terms. However, the components might not have the same unit of measurement as the original time series.

3. DATA

This section provides more information about the data sources which are used for the thesis, their reliability, and the transformations required to perform the analysis. Because the thesis focuses on the case of Gemini Wind Park (no. 5 in Appendix 6.1), the wind climate data, wave height, wind turbine specification, and other related data will be gathered within or near the location of Gemini Wind Park with the coordination of 54°02'10"North 5°57'47"East where the red cross in the below picture locates.

3.1. Data overview

In this thesis, among different weather indices provided by the KNMI, we applied hourly mean wind speed, hourly significant wave height, and hourly wind power output for the short-term prediction and the average daily mean wind speed, average daily temperature for the long-term prediction. The average daily mean wind speed and the average daily temperature were aggregated by taking the average of the hourly mean wind speed and hourly temperature respectively. We selected the temperature as the exogenous input to predict wind speed in the long term since their correlation efficiency is statistically correlated. In addition, in practice, the temperature is the easiest weather index that can be measured and converted into different altitudes. Whereas, max wind gust is almost linearly correlated with mean wind speed. Besides, the air pressure, wind direction, and relative humidity are more difficult to measure in practice or hard to convert into different altitudes to feed into the prediction model.

In addition, since the safety wind speed is defined at 10-meter heigh, the wind speed which can rotate the turbine blades are at hub height level (89.5 m), while the temperature is at sea level. Therefore, before feeding these data into any prediction models, we have to convert them into the weather indices of the same height corresponding to the purpose of that model.

3.1.1. Hourly wind speed data in the Dutch North Sea

The data was collected from The Royal Netherlands Meteorological Institute (KNMI)[2] from 2011 to 2020 from station AWG-1 located at 53° 30' North 05° 57' East which is closest to the Gemini Wind farm (marked as a red cross in the figure above). The data set consists of hourly records of the weather data as in the following table with 87672 observations. In this data set, the mean wind speed is the hourly mean wind speed recorded at 10 m above sea level. From this point onward, if nothing else is mentioned, the wind speed presented in the thesis is the hourly mean wind speed recorded at 10 m height. The full data set detail can be found in Appendix 6.2.

Table 1: Wind speed data set and variable definitions (Source: KNMI)

Variable name	Definition
YYYYMMDD	Date (YYYY=year, MM=month, DD=day)
HH (Time)	Time (time (hour, Amsterdam time)
DD (Wind_direction)	Mean wind direction (in degrees) during the 10 minutes preceding the time of observation
T (Temperature)	The temperature at 1.50 m at the time of observation

As can be seen in Table 2 below, for Wind_direction, the value of 990 degrees means that the wind blew in many directions and causes unnecessary outliers to the time series. There are 324 observations with this value and the associated weather indices are in the normal values so we will first turn them into missing values and impute them later with the imputation algorithm. This leads to

[2] https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens_Noordzee

3332 missing observations in Wind_direction at the hourly level. For the full data set statistics explanation, please see Appendix 6.2.

For short-term wind speed and wind power prediction, hourly data was applied to construct the forecasting model. In this data set, the number of consecutive missing values in each weather indices is particularly large for Wind_direction, Mean_wind_speed, and Max_wind_gust. More than 2400 hourly consecutive observations at the beginning of the data set were not recorded, which means there are missing data for more than 100 consecutive days of the weather data at the beginning of the time series. The values in the data set were missing due to the failure of the weather station at that time. This issue creates a loss of information in one season in the data set. Table 3 below summarizes the total missing values of each weather index and the largest number of consecutive missing values in each one of them. For long-term wind speed forecast, the thesis aims to establish the annual pattern of the wind speed to schedule the vessel by a statistical model for daily data, therefore, missing values should first be treated. For outliers which are higher than the third quartile plus 1.5 times the interquartile range, they are always present in the data set due to natural extreme weather conditions and they are not errors. Therefore, we kept them in the data set to ensure the integrity of the data. We aggregated the average daily mean wind speed by taking the average of the hourly mean wind speed every 24 hours.

Table 2: Hourly climate data set and variable descriptions excluding missing values (Source: KNMI).

Variable name	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
YYYYMMDD	-	-	-	-	-	-
HH (Time)	-	-	-	-	-	-
DD (Wind_direction)	(360=north, 90=east, 180=south, 270=west, 0=calm 990=variable)	-	-	-	-	-
FH (Mean_wind_speed)	m/s	6.99	7	3.18	25	0
T (Temperature)	degrees Celsius	10.76	10.40	5.65	34.10	-9.10

Table 3: Summary of missing values of weather indices

Variable name	Missing values	The largest number of consecutive missing values
DD (Wind_direction)	3008	2423
FH (Mean_wind_speed)	3030	2424
T (Temperature)	983	312

Regards the distribution of the hourly mean wind speed in the data set, it is skewed to the right as in Figure 8.

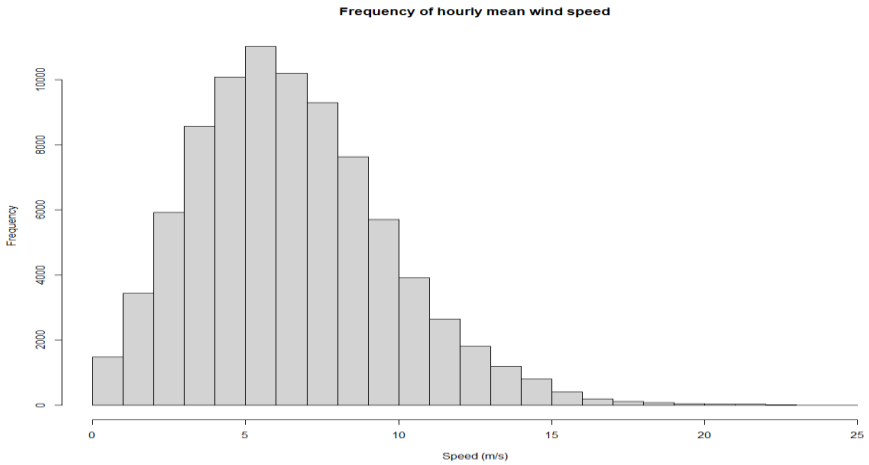


Figure 8: Hourly mean wind speed distribution before imputation

The boxplot of the hourly mean wind speed per month after imputation in Figure 9 shows that the wind season in the North Sea can be divided into two seasons. The extended winter from October to the next March with the median wind speed of around 8 m/s with the range from 0 to 25 m/s. Whereas the extended summer season with a lower median wind speed of just around 5 m/s and a smaller range from March to the end of September. As in Figure 10, the hourly mean wind speed time series has no constant mean, and the variance also evolves. However, we can see a clear seasonal pattern in the time series in which the wind speed was much higher from October to March then lower down and hit the bottom in the middle of the year.

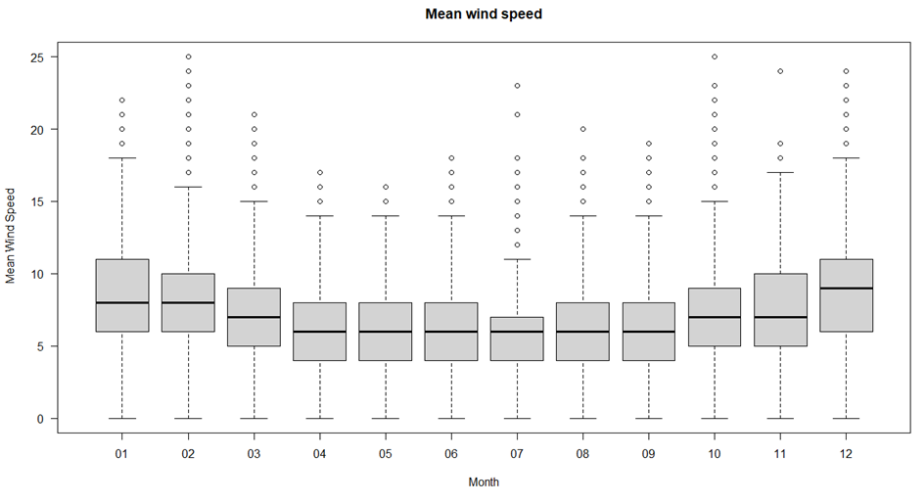


Figure 9: Hourly mean wind speed per month

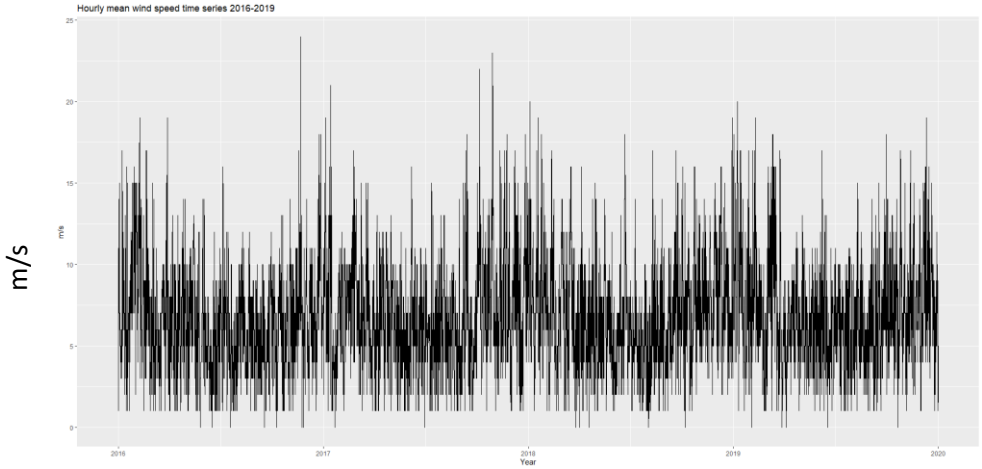


Figure 10: Hourly mean wind speed 2016-2019

In general, the hourly mean wind speed data is not stationary with a strong seasonal pattern. In addition, it is not normally distributed but skewed to the right and there are many missing values within the data. Its distribution is not normal but skews to the right.

3.1.2. Average daily mean wind speed data in the Dutch North Sea

Firstly, for long-term wind speed prediction, the average daily mean wind speed from 2014 to 2019 was used to build the forecasting model for 2020 since we are interested in the average daily downtime energy output of the wind farm. Secondly, the time series from 2014 to 2019 does not have many missing values to impute which makes the data aggregation by the daily average value more precise. Thirdly, if there is any seasonality in the data, the forecasted seasonal pattern shall be mostly influenced by the shape of the seasonality near the end of the time series but not so much by the pattern at the beginning of the time series. Moreover, to reduce the computation time, we decided to use the data from 2014 to 2019 with 2191 observations. The summary of the average daily climate data after being aggregated from the selected imputed hourly data set is provided in below Table 4. This data set is calculated based on the already imputed hourly data set. According to it, there is no day with zero average daily mean wind speed. For the full data set statistics, please see Appendix 6.2.

As can be seen in Figure 11 below, from 2014 to 2019, there was not a visible trend in average daily mean wind speed. This time series is not stationary since there is no constant mean and the variance increases at the end of the time series. This time series also illustrates a strong seasonal pattern which increases in size as the levels of the series increase.

Zooming into the most recent three years from 2017 to 2019 in Figure 12, there is an annual seasonal pattern of the average daily mean wind speed. As can be seen, the average daily mean wind speed often increased and peaked at the end of the year (from October to next March) and hit the bottom in the middle of the year (from March to September). The red arc connects two observations on the 1st of November while the green arc connects two observations on the 1st of May in 2018 and 2019.

Table 4: Average daily climate data from 2014 – 2019.

Variable name	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Index	Year – Month – Day	-	-	-	-	-
Mean_wind_speed	m/s	6.94	6.63	2.61	17.92	1.29
Temperature	degrees Celsius	11.05	10.63	5.49	26.53	-5.79

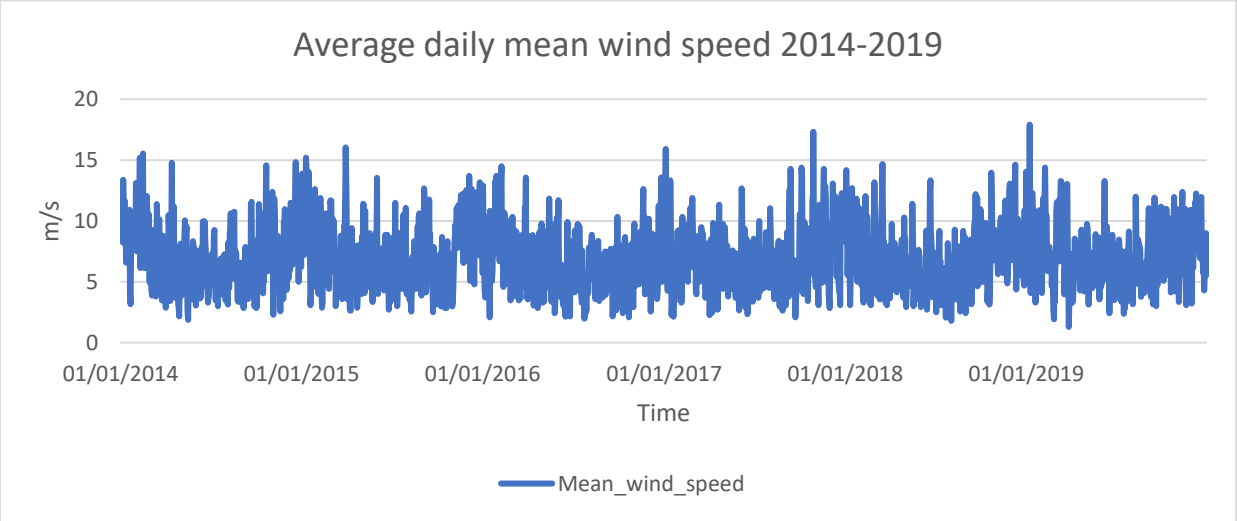


Figure 11: Average daily mean wind speed after imputation 2014-2019

This annual seasonal pattern is more obvious for data aggregated at the monthly level as in Figure 13. The average monthly wind speed started to increase from October to next March and lower from March to September. Except for the sudden dips in October 2015 and 2016 due to El Nino (Becker, 2016) Therefore, we can divide the annual wind climate into two seasons, the strong wind season is the extended winter from last October to March and the calm season is the extended summer from March to September.

Based on Figure 14, the average daily mean wind speed from 2014 to 2019 is skewed to the right. In addition, its autocorrelation function (ACF) plot illustrates a sinusoidal pattern whose autocorrelations are larger for seasonal lags of one year. In addition, in Figure 15, we can see that the autocorrelations decrease when the lags go from the beginning of the year to midyear before recovering when the lags go toward the end of the year. This shows a stable annual seasonality in the data (Hyndman, 2021). In general, this average daily mean wind speed possesses a strong and stable annual seasonality, and it is not stationary. Besides, its distribution is not normal.

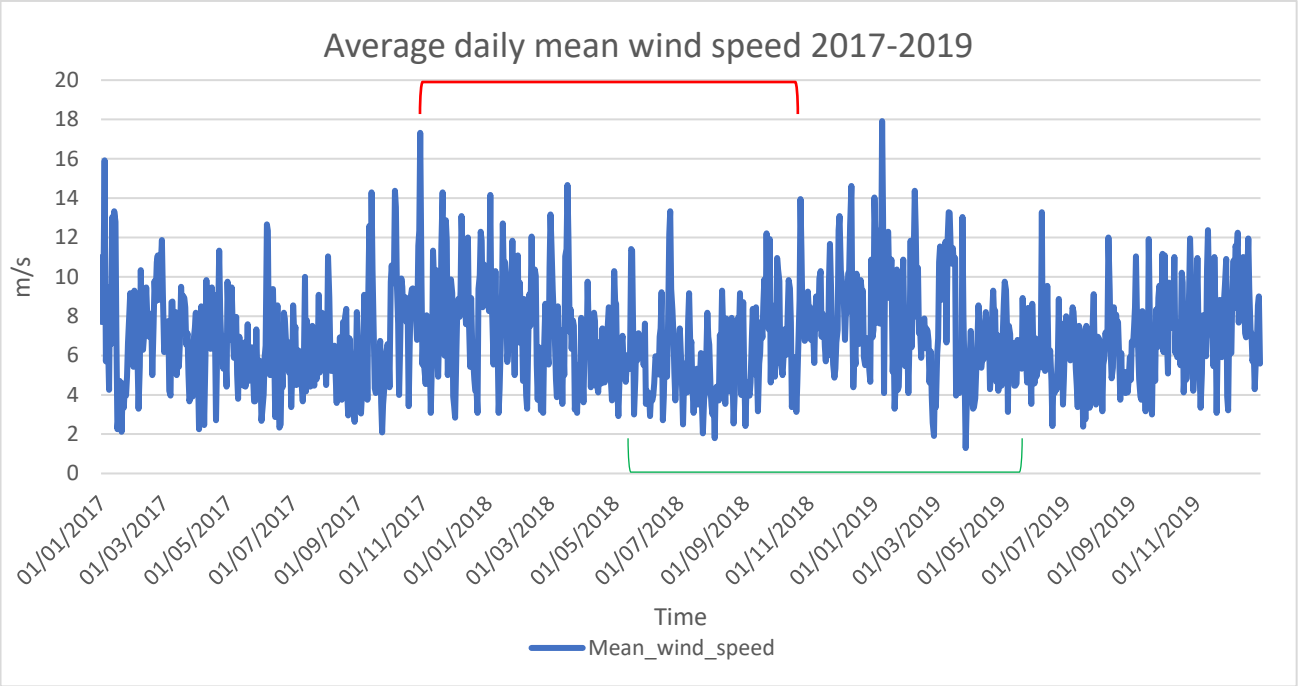


Figure 12: Average daily mean wind speed 2017-2019

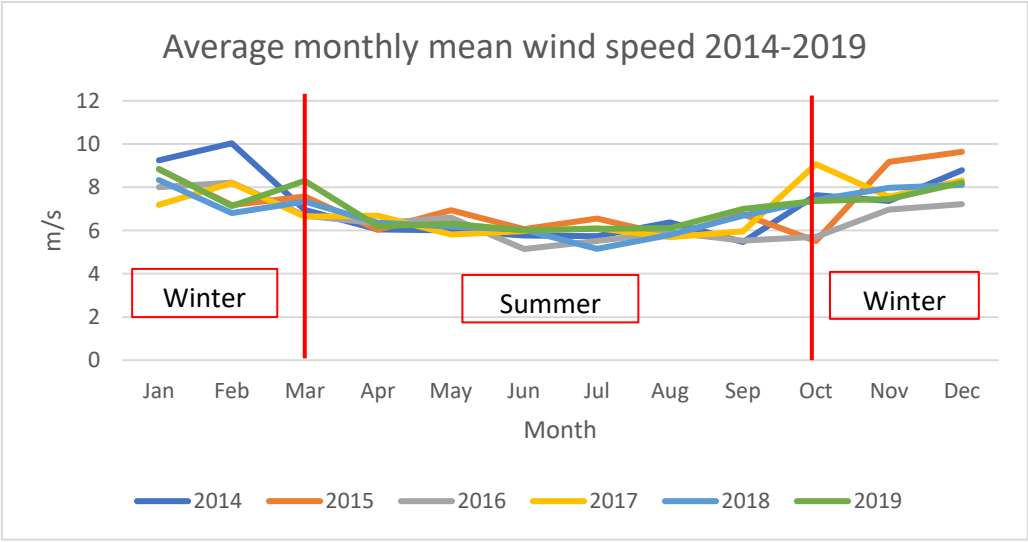


Figure 13: Monthly average daily wind speed 2014-2019

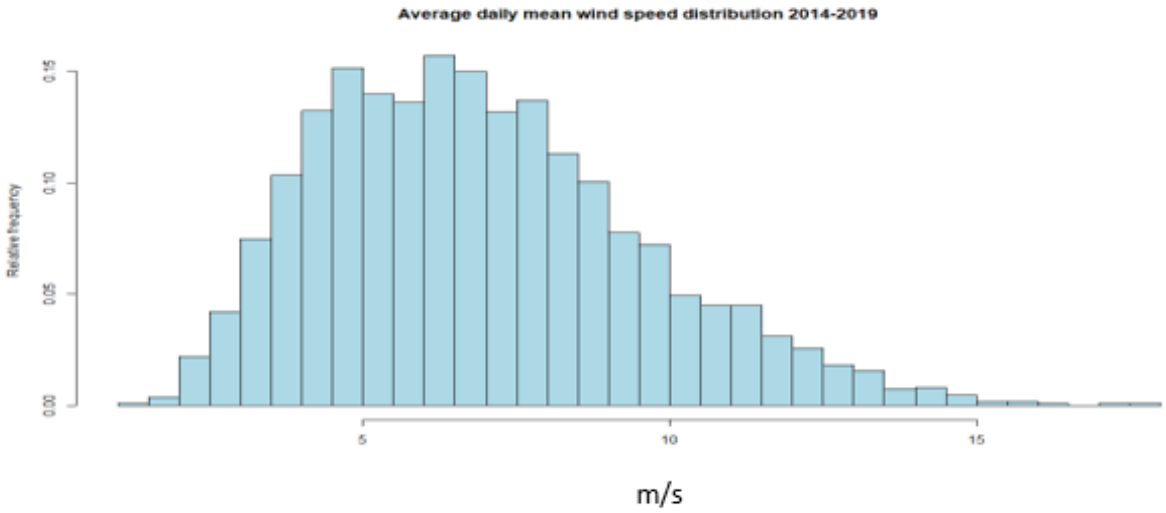


Figure 14: Distribution of the average daily mean wind speed 2014-2019

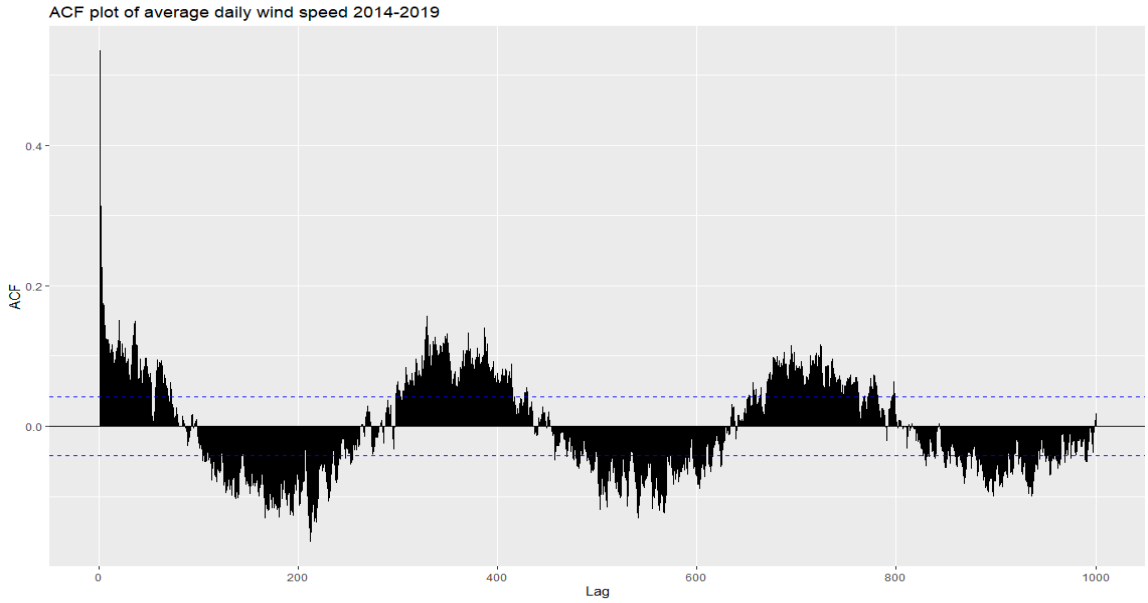


Figure 15: Autocorrelation function plot of the average daily mean wind speed 2014-2019

3.1.3. Hourly wave height data in the Dutch North Sea

The wave height data of the Dutch North Sea was collected from the Copernicus Marine Service[3] which consists of hourly significant wave height in meters at the Gemini wind farm location. For vessel design, the limiting wave states of vessels have traditionally been described only with significant wave height (H_S) which is defined as the average value of one-third highest waves (Bai and Jin, 2016).

The data set obtained only includes the wave height measurement in meters and the time it was recorded. Therefore, the seasonality and nonlinearity of the data will be analyzed by STL decomposition.

There are 34381 hourly observations of the significant wave height from 2018 to 11th Jan 2021 in the data set. To assist the analysis, time variables such as Hours, days, Weekdays, Months, Year were included in the data set. Fortunately, there is no missing value in this data set. The table below explains the variables used in this data set. Since the analysis is in The Netherlands, the date and time were converted to Amsterdam time (UTC+1).

[3] https://view-cmems.mercator-ocean.fr/ARCTIC_ANALYSIS_FORECAST_WAV_002_014

The boxplot of hourly significant wave height per month from 2018 to 2021 (Figure 16) also illustrates that hourly significant wave heights from March to October are lower with the median around 0.8 m and their range is between around 0.3 and 3.2 m with outliers up to 4.5 m. Significant wave height in the winter has a median of around 1.5m and ranges from about 0.3 to 4.5 m with outliers up to more than 6m.

Table 5: Variable definition and description of wave height data set 2018-2021

Variable name	Definition	Minimum	Maximum	Mean	Median	Standard Deviation
Time	Date (YYYY=year, MM=month, DD=day) and Time (HH=hour, MM=minute, SS=second)	-	-	-	-	-
VHM0	Significant wave height (m)	0.15	6.35	1.40	1.20	0.84

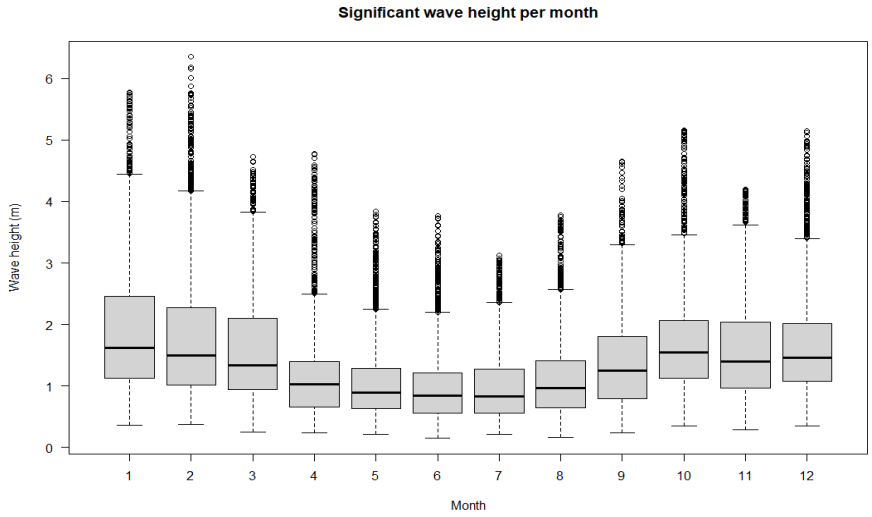


Figure 16: Boxplot of significant wave height per month 2018-2021

As can be seen in Figure 17, the hourly significant wave height time series does not have a constant mean and its variance evolves. In addition, the distribution of the wave height is skewed to the right as in Figure 18.

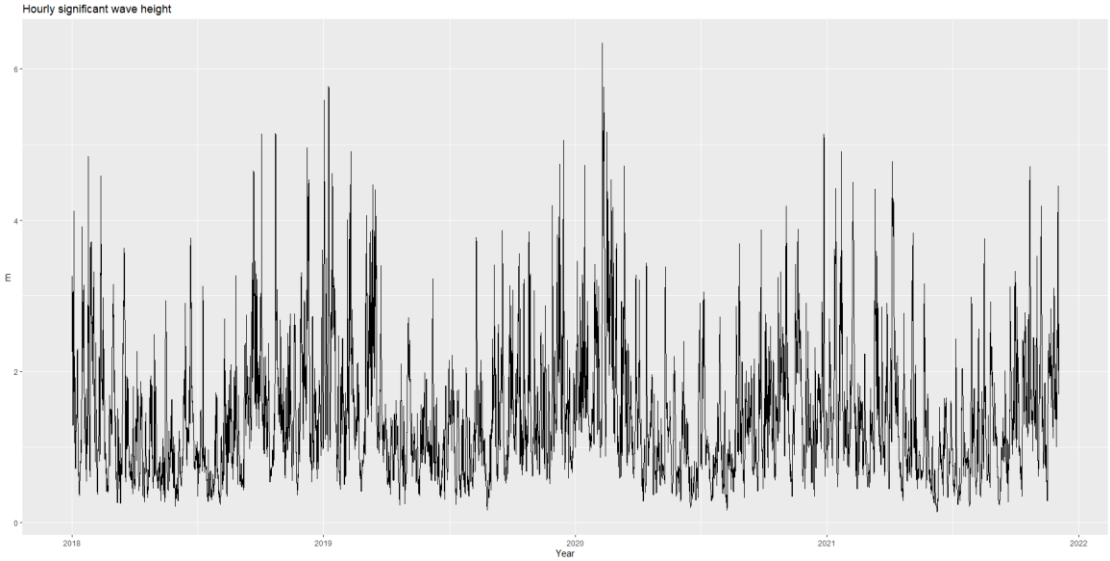


Figure 17: Hourly significant wave height time series

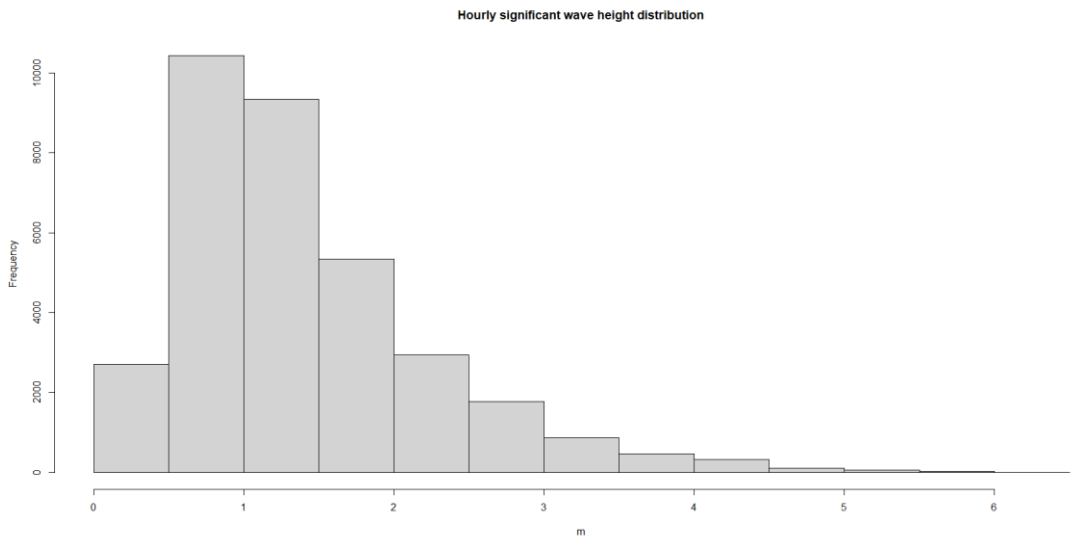


Figure 18: Hourly significant wave height distribution 2018-2021

3.1.4. Wave height restriction of transportation vessel

Besides, if the maintenance service requires the replacement of the turbine's components like blades or generator, transportation vessels that assist the lifting process such as jack-up vessels have to be chartered.

A jack-up vessel is a self-elevating unit with a buoyant hull equipped with about 3-6 legs. These legs can station themselves on the seafloor to raise their hull above the sea surface, which enables crane operation in a stable environment despite harsh weather conditions (Dalgic, Lazakis, Turan, *et al.*, 2015a). Table 6 below provides the weather restriction of the jack-up vessel. As can be seen from the table, conditions for the operation of jack-up vessels are stricter than jacking conditions. However, wind speed at the hub level is an additional condition that enables lifting operations to take place.

During the mobilization of the jack-up vessel, the survivability wave height and wind speed allow the vessel to sail without the elevated risk of sinking and capsizing. Since the author assumed the major repairs could not be interrupted after the maintenance activities start, the jack-up vessel can only start the O&M activity if there is no expected storm during the maintenance period. During the operation of the jack–up vessel, it can only be jacked up and down within the wave height and wind speed at sea level jacking limit. In addition, the lifting activity of the jack-up vessel can only be implemented if the wind speed at the hub is less than the wind speed at the hub level limit because the main O&M activities involve heavy equipment lifting.

In the case of the Gemini Wind Park, Van Oord's Aeolus vessel was chartered, its specifications are presented in the table below (Crol, 2015). The picture of this vessel can be found in Appendix 6.11.

Table 6: Weather restriction of Van Oord's Aeolus vessel. Source: (Crol, 2015)

	Jacking	Operability	
		Jacking	Lifting
Max wave height (m)	1.80	1.80	3.60
Max wind speed at sea level (m/s)	40	15	Not Applicable
Max wind speed at hub level (m/s)	Not Applicable	Not Applicable	20

3.1.5. Cost data

3.1.5.1. Energy production of Gemini wind farm

Gemini wind farm features 150 Siemens SWT-4.0-130 wind turbines[4] with a total capacity of 600 MW. This turbine model has a rated power of 4000 kW with a cut-in wind speed of 5.0 m/s, rated wind speed of 12 m/s, and cut–out wind speed of 25 m/s. 85% of the electricity generated from Gemini will be purchased by Delta and the rest is purchased by HVC. Since their purchased prices are confidential, in the thesis, the author will use an electricity price of 17 cents/kWh as reported in De Ingenieur (www.deingenieur.nl, 2017)

The actual Dutch offshore wind energy production time series was retrieved from Entsoe – Transparency Platform (transparency.entsoe.eu, 2022) from 2017 to 2020. The energy data is aggregated over all Dutch offshore wind farms every 15 minutes. Therefore, to extract the average energy production of Gemini per hour per turbine from the total offshore wind energy, we followed this process of calculation. First, we calculate the actual average total offshore wind energy per hour in The Netherlands by summing every 4 observations to get the aggregated power per 15 minutes. After that, we divided the results by 4 to get the average offshore wind power generated in The Netherlands per hour in MWh.

Table 7: Operational wind farms and their respective nominal capacity in the Netherlands. Source: ((RVO), 2021)

Wind farm	Capacity wind farm (MW)	Proportion (%)	In use since
Borssele V (Innovation site)	19	0.77	2021
Borssele I & II	752	30.57	2020
Borssele III & IV	731.5	29.74	2020
Gemini	600	24.39	2016
Luchterduinen	129	5.24	2015
Prinses Amalia	120	4.88	2008
Egmond aan Zee (OWEZ)	108	4.41	2007
Total	**2459.5**		

The second step is to extract the average energy production of Gemini per hour out of the whole country's production by estimating its proportion of energy contribution. As can be seen in Table 7 the proportion of nominal energy contributed by the Gemini wind farm is approximately 24.39% in 2021. However, since the time series under study was from 2017 to 2020, therefore, we have to recalculate the proportion of Gemini wind farms in 2020 when the Borssele I, II, III, and IV wind farms were installed. From 2017 to March 2020, the proportion of energy contributed by Gemini to the Dutch national gridline was 62.69%. From the 13th of April[5] to the 27th of November[6], there were 94 turbines installed in Borssele I and II with a nominal capacity of 8MW each turbine. Therefore, we assumed that each month, there were 8 turbines installed and connected to the national gridline in Borssele I and II. From the 7th of August[7] to the 26th of November[8], there were 77 turbines installed and connected in

[4] https://www.geminiwindpark.nl/turbines--nacelles--rotorblades.html#t1

[5] https://www.deme-group.com/news/turbine-installation-borssele-1-2-successfully-completed-deploying-deme-offshores-unique

[6] https://www.offshorewind.biz/2020/11/27/borssele-1-2-fully-commissioned/

[7] https://www.blauwwind.nl/en/news/2020/7/first-power-borssele-iii-iv

[8] https://www.blauwwind.nl/en/news/2020/11/installation-of-the-final-turbine-for-borssele-iii-iv

Borssele III and IV with a nominal capacity of 9.5 MW for each turbine. Therefore, we assumed that each month, there were 19 turbines installed and connected to the national gridlines in Borssele III and IV. Table 9 below shows the corresponding proportion of energy generated by Gemini each month in 2020 while the four wind farms above were being installed in 2020.

The third step is to calculate the wind energy generated per turbine per hour in the Gemini wind farm. Each observation was divided by 150 (since there are 150 turbines in the Gemini wind farm) and then multiplied by 1000 to convert from MWh to kWh.

After cleaning and aggregating the data, the summary statistics of the wind power per turbine in Gemini from 2017 to 2020 are described in the below table. Wind electricity generation is measured at the hourly level. In this thesis, the term "wind electricity generation", "wind energy", and "wind power" are used interchangeably with the same unit of measurement. Since the nominal of the Siemens SWT 4.0 – 130 turbine model in Gemini has the nominal power generating capacity of 4000 kWh, the maximum value of the real power of 3814.4 makes sense. This indirect power output calculation does not consider the possible energy lost due to the daily maintenance. Therefore, this issue might affect the accuracy of the prediction model. The detail of the proportion estimation can be found in appendix 6.10.

Table 8: Wind power per turbine in Gemini offshore windfarm statistics from 2017-2020

Index	Wind electricity generation (kWh)
Min	0
Median	1516
Mean	1677.80
Max	3814.4

Another way to collect the Gemini energy output directly is using the proportion of Gemini wind farm's average actual annual output compared to the total actual annual offshore wind energy production[9]. From 2017 to 2020, the proportion of the Gemini's actual annual output is around 67.2% which was only 5% higher than the nominal output but it is harder to calculate the wind energy output each month. Therefore, we will proceed with the proportion of the nominal output in this thesis.

3.1.5.2. Chartering cost of Jack-up vessel

Since the chartering and related cost of The Van Oord's Aeolus jack-up vessel is confidential, the thesis will apply the cost data from (Dalgic, Lazakis, Dinwoodie, *et al.*, 2015b). It is assumed that the Aeolus jack-up vessel is chartered on a fix-on-fail basis (2-8 weeks), not being purchased for the whole lifetime of the wind farm. It is available during the charter period. After that, if there are more incomplete tasks, the daily demurrage rate will be charged until all repairs are done and the vessel will sail back to shore if the agreed charter period is not long enough to fix all issues. Because we do not have the exact cost of The Aeolus, we use the chartering rate of the jack-up vessel in Dalgic, Lazakis, Turan, et al., 2015b in the table below. As you can see, the daily charter rate is seasonally-based and the winter charter rate is cheaper than the summer charter rate. The whole cost of a jack-up vessel is given in Appendix 6.7.

Table 9: Jack-up short-term charter parameters. (Source:Dalgic, Lazakis, Turan, et al., 2015b)

Parameter	Value
Daily charter rate	172500 Pound/day (April – September)
	116250 Pound/day (October – March)

[9] https://nl.wikipedia.org/wiki/Windturbines_in_Nederland

Parameter	Value
Demurrage	+30% daily rate
Jack-up/down period	3 hours
Hub removal time	8 hours

3.2. Data preprocessing and analysis methods

3.2.1. Random forest imputation for missing values

To estimate the missing values, methods like the last observation carried forward, the next observation carried backward, and linear interpolation between time stamps, are not applicable since they cannot generate the seasonality in the original data. Since there are 2424 consecutive missing values in wind speed data, the last observation carried forward only uses the value of the last observation to fill the whole range of missing values and turns the weather indices in this period into straight lines. The next observation carried backward shares the same mechanism and results. The linear interpolation between time stamps method assumes a linear relationship between the two known values at the beginning and the end of the missing period, therefore, the imputed values also create a straight line in the missing period of 101 days or around 4 months. Therefore, it fails to regenerate the pattern of the missing parts in the data set and the imputed data loses the seasonal variance inhibited in the weather indices. In addition, if we copy the whole data of the same period next year to the missing part, we will lose the correlation between these two years.

In contrast, the imputation of missing values based on the random forest algorithm of (Leo Breiman and Adele Cutler, 2011) is a non-parametric imputation method which is robust to both categorical and numeric variables, non-linear relations, multi-collinearity, seasonality, and high dimensionality (Stekhoven, 2011). It requires no assumption or transformation of the data before imputing and it preserves the seasonal variance of the weather indices better. Therefore, the missForest package (Daniel J. Stekhoven & Peter Bühlmann, 2011) was applied to impute the hourly missing values of the data set. It applies a random forest to the observed parts of each variable and then predicts the missing parts of that variable. It repeats these two steps continuously and runs iteratively to update the imputed matrix variable-wise. Next, it assesses its performance between iterations by considering the difference(s) between the results of the previous and new imputation. The algorithm will stop (for both one-type and mixed-type variables) in case this difference increases or a specified maximum iteration made by the user is reached. This package allows at max 10 iterations, however, due to large computational time, 5 iterations (maxiter = 5) were used. Table 13 below summarizes the computational time and errors of the different tuning parameters with 5 iterations.

The fundamental concept of random forest is a decision tree learner which uses the tree-like structure to form the relationship between features and outcomes. A random forest is constructed by a number of trees (ntree) on bootstrapped training samples to minimize the variance of outcome and the correlation between these decision trees. In each training sample, only a subset of variables in the data set (mtry) is selected randomly to grow a full tree and find the results.

To impute the missing data in the training sample, the missing value is first filled by the median of the variable if it is a continuous variable, then we refine this guessing by finding similar observations to the one with the missing value. This refining process is conducted by feeding all the data to each tree in the random forest, the similar observations are those which stay in the same node. After that, the missing value is calculated as the weighted average of these observations based on the proximity. We iterate this refining process around 5 times until the missing values converge. In this way, each tree in the random forest is not correlated with other trees and the algorithm removes potential correlation or multi-correlation between variables in the data set since only a subset of variables were applied to build each tree. These imputations were assessed by the Out-Of-Bag errors (Mean Squared Error – MSE) per variable (Daniel J. Stekhoven & Peter Bühlmann, 2011) and implemented by an Intel i7 chip with an 8-core CPU.

In addition, to provide the imputation algorithm with the perspective of time, three variables of the day, month, and weekdays were added to the data set, thus there are 6 weather index variables (Wind_direction, Mean_wind_speed, Max_wind_gust, Temperature, Air_pressure, Relative_humidity) and 4 variables indicating time (Time, Month, Day, Weekdays) used to feed the missForest imputation algorithm. We tuned the number of trees in each forest (ntree) from 2,10,25 to 50. For the number of variables randomly sampled per split (mtry), we used the default value of the square root of the number of variables in the data set and 7. We will impute the data set from 2011 to 2019 for building the forecasting models and keep the data in 2020 as-is for testing purposes.

3.2.2. Box-Cox Transformation

Since the distributions of the average daily mean wind speed, the hourly mean wind speed, and hourly wave height are all skewed to the right, we need to normalize the distribution of these time series to ensure the assumptions of the statistical forecast models. Unfortunately, log transformation could not successfully make these data sets normally distributed as in the below pictures. The log-transformed weather data sets are skewed to the left. Therefore, Box-Cox transformation was applied to these data to better normalize the distribution of the data (G. E. P. Box & D. R.Cox , 1964). This transformation applies both a logarithmic and power transformation based on the parameter λ as follows:

$$yb_t = \begin{cases} \log(y_t) & if\ \lambda\ =\ 0 \\ \frac{{y_t}^{\lambda} - 1}{\lambda} & if\ \lambda\ \neq\ 0 \end{cases} \tag{18}$$

yb_t is the Box-Cox transformed value of the weather index y_t at time t, λ can take any values, and $\log(y_t)$ is the natural logarithm. With function coef() and powerTransform() in the Companion to Applied Regression (CAR) package, we can find the best value of lambda to normalize the distribution of the data (John Fox and Sanford Weisberg, 2019). After implementing these above functions, the weather indices are more normally distributed as in Figure 19.

The back-transformation of the Box-Cox transformation is:

$$y_t = \begin{cases} \exp(yb_t) & if\ \lambda\ =\ 0 \\ \exp\left(\frac{log(1+\lambda yb_t)}{\lambda}\right) & if\ \lambda\ \neq\ 0 \end{cases} \tag{19}$$

One more thing to note is that the point forecast after the Box-Cox transformation of the prediction is the median, not the mean of the forecast distribution, therefore when we use the mean of the forecast distribution, we call it a bias-adjusted point forecast which is calculated by the following formula (Rob J Hyndman, 2021):

$$\widehat{y_{t+h|h}} = \begin{cases} \exp\left(\widehat{yb_{t+h|h}}\right)\left[1 + \frac{\sigma_h^2}{2}\right] & if\ \lambda\ =\ 0 \\ (\lambda \widehat{yb_{t+h|h}} + 1)^{\frac{1}{\lambda}}\left[1 + \frac{\sigma_h^2(1-\lambda)}{2(\lambda \widehat{yb_{t+h|h}}+1)^2}\right] & if\ \lambda\ \neq\ 0 \end{cases} \tag{20}$$

$\widehat{y_{t+h|h}}$ is the back-transformed mean of the h-step ahead forecast distribution from time t, $\widehat{yb_{t+h|h}}$ is the h-step ahead forecast mean and σ_h^2 is the variance of the h-step forecast in the Box-Cox transformed scale. As can be seen, the further the forecast away, the larger the forecast variance. This back-transformation will be automatically transformed by the fable package in R(Mitchell O'Hara-Wild et al., 2021). One thing to note is that not all the weather data types are non-negative. For climatological data like temperature, the data may contain many negative observations, therefore, the Box-Cox transformation family allowing negative values was applied (D. M. Hawkins, S. Weisberg, 2017). It is the Box-Cox transformation of z statistics where:

$$z\ =\ 0.5\ \times (y_t + \sqrt{y_t^2 + \gamma^2}) \tag{21}$$

Data set	Lambda (λ)
Hourly mean wind speed	0.55
Average daily mean wind speed	0.33
Hourly significant wave height	0.08

Table 10: Lambda value of the Box-Cox transformation of weather data

3.2.3. Weather data transformation at different height

3.2.3.1. Wind speed at different height

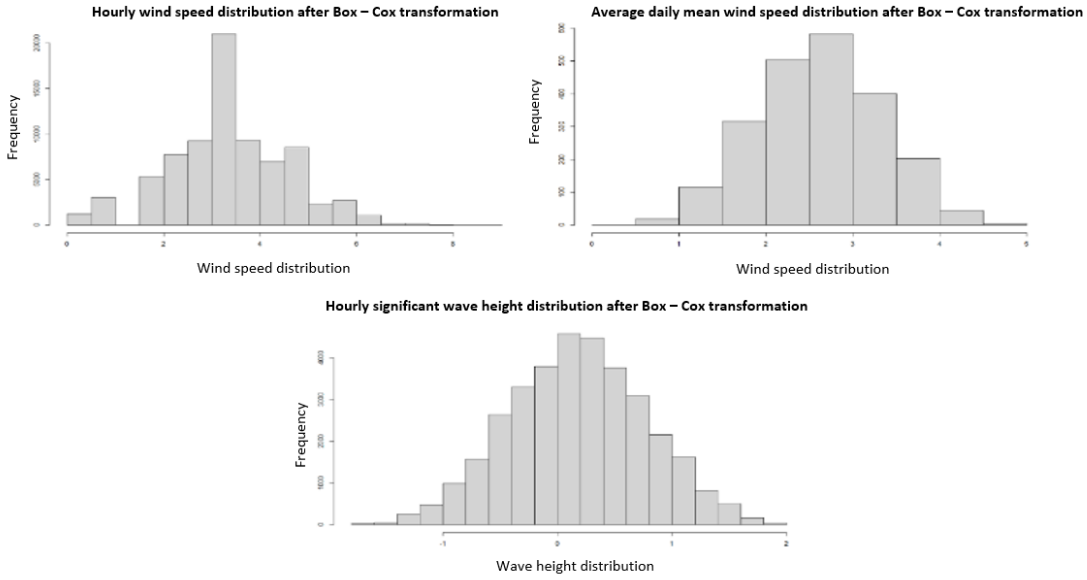


Figure 19: Distribution of weather data after Box-Cox Transformation

Since the wind turbine tower is 89.5 meters high, the wind speed was also converted to the corresponding height for calculation instead of the 10 meters height as was used in the original data set. Therefore, the thesis applies the hourly wind speed transformation at 10m above the sea level to the hub height of 89.5m based on the formula from Kubik, Coker, and Hunt (2011):

$$\frac{u(h_2)}{u(h_1)} = \frac{log(h_2) - log(h_0)}{log(h_1) - log(h_0)} \quad (22)$$

Where $u(h_1)$ is the wind speed u at h_1 = 10 m, $u(h_2)$ is the wind speed at h_2 = 89.5 m, h_0 = 0.0001 m. Therefore, the wind speed at 89.5-meter height is:

$$u(89.5) = 1.19\,u(10) \quad (23)$$

3.2.3.2. The temperature at different height

Since the temperature data of KNMI was recorded at 1.5 meters height, we converted it to the temperature at 89.5 meters height based on the US Standard Atmosphere 1976 (NOAA, NASA, & The United States Airforce, 1976):

$$Temp(h_2) = Temp(h_1) + L_b \times (h_2 - h_1) \quad (24)$$

where $Temp(h_2), Temp(h_1)$ are the temperature in Kevin at h_2 and $h_1$respectively, L_b is the base temperature lapse rate per kilometer of geopotential altitude. L_b = -6.5 K/km = - 0.0065 K/m. Therefore, the temperature at 89.5 meter height in Celsius:

$$Temp(89.5) + 273.15 = Temp(1.5) - 0.0065 \times (89.5 - 1.5) + 273.15$$

$$\leftrightarrow Temp(89.5) = Temp(1.5) - 0.0065 \times (89.5 - 1.5)$$

$$\leftrightarrow Temp(89.5) = Temp(1.5) - 0.572 \quad (25)$$

Because we have used temperature as an exogenous input to predict the wind speed, the temperature has to be at the same altitude as the wind speed. In addition, we selected the temperature as the only exogenous input since other weather indices like wind direction, atmospheric pressure, and relative humidity are hard to measure and convert into different altitudes, while maximum wind gust is almost linearly correlated with the mean wind speed.

3.3. Data preprocessing and analysis results

3.3.1. Hourly mean wind speed data imputation results

As can be seen in Table 11 below, given the same number of variables randomly sampled per split (mtry), increasing the number of trees in the random forest for each variable reduces the overall imputation errors but increases the computation time. If we also increase the value of mtry along with the number of trees, we trade off between the accuracy of the imputation algorithm and the computation time. Here, when we doubled ntree and took mtry = 7 instead of $\sqrt{10}$ the computation time almost tripled. The more trees and mtry, the less imputation error and the longer time we have to wait for completing the data because there might be observations which are missing completely without any weather indices recorded. Therefore, if mtry = $\sqrt{10}$ there will be the case that the randomly selected variables for the subset are all missing and harder to find similar observations to refine the first guess, whereas, with mtry = 7, the day, month, and weekdays will have a higher chance of being included and provide more data to each tree to find the similar observations and increase the accuracy of the imputation. The higher the number of trees and the number of randomly selected variables, the more accurate the imputation is. If we keep increasing ntree and mtry, the imputed wind speed pattern will more and more resemble the real wind speed pattern, however, in this case, we have to trade off the long computation time for accuracy since it already took nearly 6 hours for imputing the data set with mtry = 7 and ntree = 50.

Table 11: OOB errors per variable of each imputation model

		ntree = 2 and mtry = $\sqrt{10}$	ntree = 10 and mtry = $\sqrt{10}$	ntree = 25 and mtry = $\sqrt{10}$	ntree = 50 and mtry = 7
OOB error (MSE)	Mean_wind_speed	0.79	0.58	0.52	**0.52**
	Temperature	1.49	1.10	0.88	**0.49**
Computation time (minutes)		11.23	44	154.87	**352.51**

As can be seen from Figure 20 below, the most missing records of wind speed were between February and May 2011. The red squares mark the same period from February to May in 2011 and 2012 when there is much less missing value for comparison. Random forest imputation imputed the missing data with a volatile pattern of the mean wind speed instead of a straight line therefore, it preserves the variance of the wind speed time series. However, in comparison with the same period in 2012, the volatility and the variance of the imputed mean wind speed were smaller. This is negligible since the missing values only account for around 3% of the data set. Besides, missing value imputation does not increase the information of the data set but prevents the loss of information when applying data analysis methods that require complete data. In addition, with mtry = 7 and ntree = 50, the imputed data resemble the wind speed pattern of the same period the most, therefore, we will use this imputed data set in our further research.

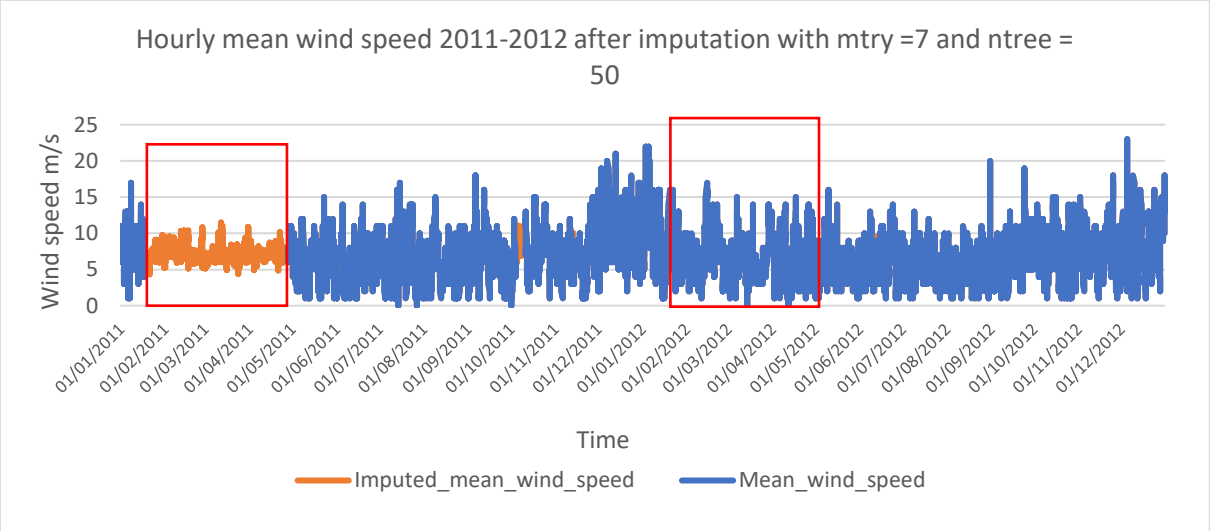


Figure 20: Hourly mean wind speed 2011-2012 after imputed with mtry = 7 and ntree = 50 (red squares indicate the same period of the two given years)

3.3.2. Hourly mean wind speed data analysis

After implementing STL on the hourly mean wind speed at 10 meters height by the fpp3 package in R, the data can be decomposed as follows with three seasonal components:

$$V_t = Season\_day_t + Season\_week_t + Season\_year_t + Trend_t + Remainder_t \quad (26)$$

where V_t is a wind speed time series, $Season\_year_t$ is a yearly seasonal component, $Season\_week_t$ is a weekly seasonal component, $Season\_day_t$ is a daily seasonal component, and $Trend_t$ is a trend-cycle component, and $Remainder_t$ is a remainder component. The Mean wind speed is the original average daily mean wind speed, the Trend column is the trend following the movement of the time series ignoring the daily, weakly, yearly seasonal fluctuations, and random fluctuations.

In Figure 21, the trend component slightly declines over the period from 2015 to 2019 with a deep plunge at the end of 2016. This plunge was due to the El Nino and La Nina effects in 2015 and 2016. The yearly seasonal component remained almost the same with the V-shape pattern ranging from -2.5 to 7.5. The wind rose from the middle of the year to the end of the year before decreasing at the beginning of the next year. The weekly seasonal component fluctuated around 0 with a larger variance at the beginning and the end of the year compared to that in the middle of the year. It ranged from -4 to 6. The daily seasonal component also fluctuated around 0 with a more constant variance.

One more thing to note is the vertical scale of each component. The longer the vertical bar on the left is, the smaller the range of the data is and the weaker the corresponding component is. Here in Figure 21, the daily and weekly seasonal components are both weaker than the yearly seasonal component.

The strength of each component shows that the yearly seasonal component (0.291) is the strongest in the time series following the weekly seasonal component (0.141) and daily seasonal component (0.073). In addition, the yearly seasonal component peaks in December and hit the bottom in June, whereas the trough and peak of the weekly and daily seasonal components are in December and January respectively. The formulas of each component's strength with three seasonal components are given in Appendix 6.3.

In conclusion, the hourly mean wind speed time series from 2014 to 2019 is not stationary with a multi-seasonality of annual, weakly, and daily seasonal components. This hourly time series has a slightly decreasing trend. The annual seasonality of this time series does not seem to evolve much over time. Since the hourly wind speed at 89.5-meter height is the direct linear transformation of hourly wind speed at 10-meter height, the STL decomposition structure and components of hourly wind speed at 89.5-meter height remain the same as that of hourly wind speed at 10-meter high.

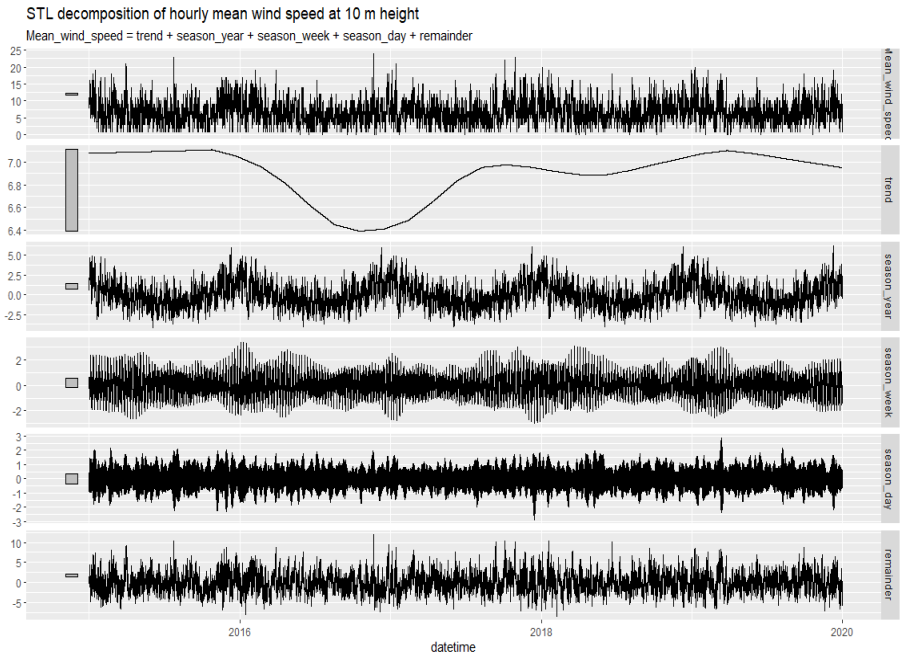


Figure 21: Hourly mean wind speed 2015 - 2019 (top) and its three additive components obtained from a robust STL decomposition

3.3.3. Average daily mean wind speed time series analysis

After implementing STL on the average daily wind speed at 10 meters in height, the data can be decomposed as followed with two seasonal components:

$$V_t = Season\_week_t + Season\_year_t + Trend_t + Remainder_t \quad (27)$$

where V_t is a wind speed time series, $Season\_year_t$ is a yearly seasonal component, $Season\_week_t$ is a weekly seasonal component, $Trend_t$ is a trend-cycle component, and $Remainder_t$ is a remainder component. The Mean wind speed is the original average daily mean wind speed, the Trend column is the trend following the movement of the time series ignoring the weakly, yearly seasonal fluctuations and random fluctuations

As can be seen, the yearly seasonal component is the strongest and influence the pattern shape of the series, followed by the weekly seasonal component and trend component (0.32, 0.15, and 0.02 respectively)

As can be seen in Figure 22 below, the trend component of average daily mean wind speed is approximately similar to that of hourly mean wind speed. It slightly decreases over the period under the study with the deepest plunge at the end of 2016. The weekly seasonal components fluctuate around zero mean with no specific pattern. However, we can notice that these components fluctuate with a wider range at the beginning and the end of the year while they remain close to zero with a much smaller range in the middle of the year. In this figure, both the largest and smallest weekly seasonal components are in December.

One more thing to note is that since this time series is the average daily mean wind speed, the possible cycles are 7 days per week and then 365 days per year only. This might be the reason why there is no daily seasonal component in the average daily mean wind speed.

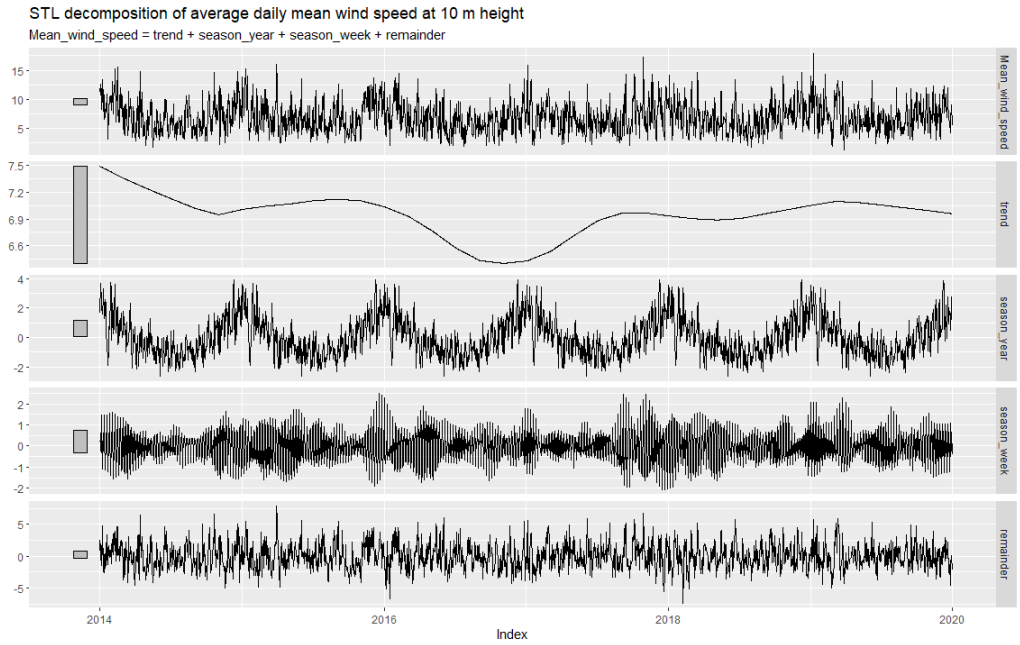


Figure 22: Average daily mean wind speed 2014-2019 (top) and its three additive components obtained from a robust STL decomposition

In Figure 22, the yearly seasonal components clearly illustrate a repetitive V-shape pattern with a fluctuation range from around -3 to 5 every year. The largest yearly seasonal component is in December and the smallest yearly seasonal component is in June. However, for the weekly seasonal component, the largest and smallest value were both in December. The remaining components in the below figure fluctuate around zero, however, with no trend and seasonal pattern.

One more thing to note is the vertical scale of each component. The longer the vertical bar on the left is, the smaller the range of the data is and the weaker the corresponding component is. Here, we can see that $Season\_week_t$ is weaker than season_year. In addition, at different granularities (hourly and daily), the yearly seasonal component of wind speed has the same pattern of V-shape.

In conclusion, the average daily mean wind speed time series from 2014 to 2019 is not stationary with a multi-seasonality of annual and weakly seasonal components, and a slightly decreasing trend. The annual seasonality of this time series does not seem to evolve much over time. Since the daily wind speed at 89.5-meter height is the direct linear transformation of daily wind speed at 10-meter height, the STL decomposition structure and components of daily wind speed at 89.5-meter height remain the same as that of daily wind speed at 10-meter height.

The hourly wind power distribution is not normally distributed, along with the multi-seasonality features would require a robust prediction method that can handle these characteristics of the time series.

3.3.4. Hourly wind power time series analysis

After implementing STL on the hourly mean wind speed at 10 meters height, the data can be decomposed as follows with three seasonal components with P_t is the average power generated at time t (Figure 23):

$$P_t = Season\_day_t + Season\_week_t + Season\_year_t + Trend_t + Remainder_t \quad (28)$$

As can be seen from the figure below, the wind power was quite stable from 2017 to 2019 before gradually decreasing in 2020, this trend is also confirmed by the declining trend of the wind speed at the Gemini site in the hourly wind speed time series analysis above. This trend is the weakest component of the hourly average wind power time series with a strength value of around 0.01.

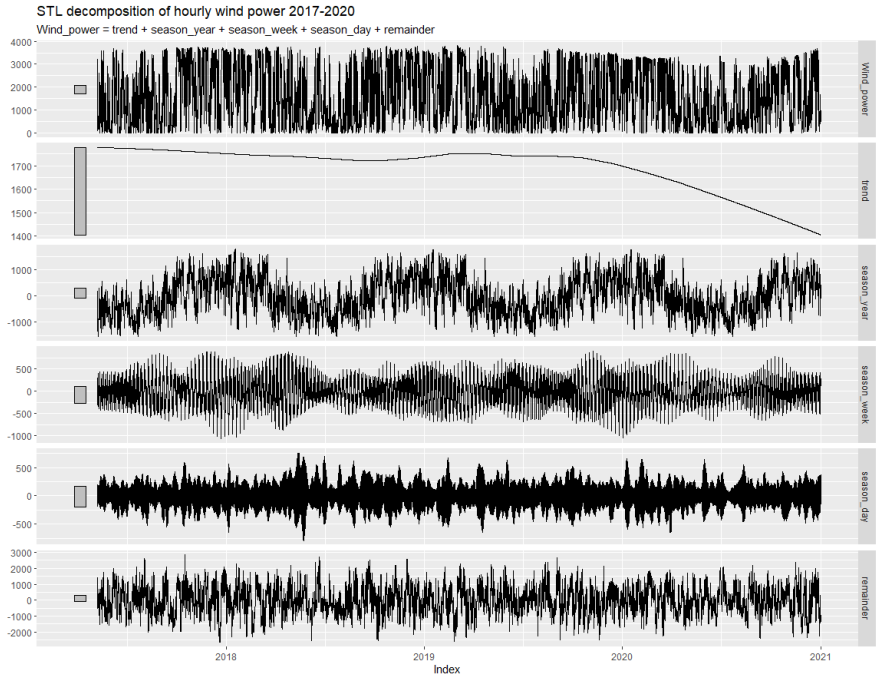


Figure 23: STL decomposition of the hourly average wind power generated by each turbine in Gemini wind farm from 2017 – to 2020

The $Season\_year_t$ the component also followed a V-shape pattern which increased and peaked from October to March before decreasing toward the middle of the year (ranging from about 2000 to -1250). This is also the strongest component with a strength of 0.39. The $Season\_day_t$ and $Season\_week_t$, however, do not have a specific pattern and fluctuated wildly. For the $Season\_week_t$ component, the fluctuation range was the largest around the winter months and started to shrink toward the middle of the year (ranging from around 1000 to -1000 and 250 to -250 respectively). The $Season\_week_t$ component of the hourly average wind power is weaker than the $Season\_day_t$ component with the strength (0.14 and 0.05 respectively). In addition, the distribution of this time series is not normal and symmetric (Figure 24).

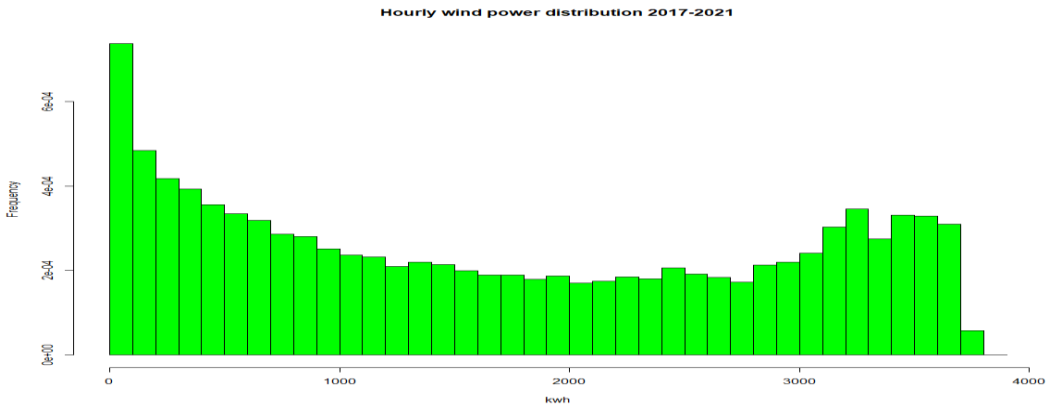


Figure 24: Hourly average wind power distribution 2017 - 2020

3.3.5. Hourly wave height time series analysis

As can be seen from Figure 25 below, there are two seasonal components in the hourly significant wave height time series but there is not a visible trend in significant wave height. In addition, unlike other time series at the hourly level in this study, hourly significant wave height only has weekly and daily seasonality:

$$Wave_t = Season\_day_t + Season\_week_t + Trend_t + Remainder_t \quad (29)$$

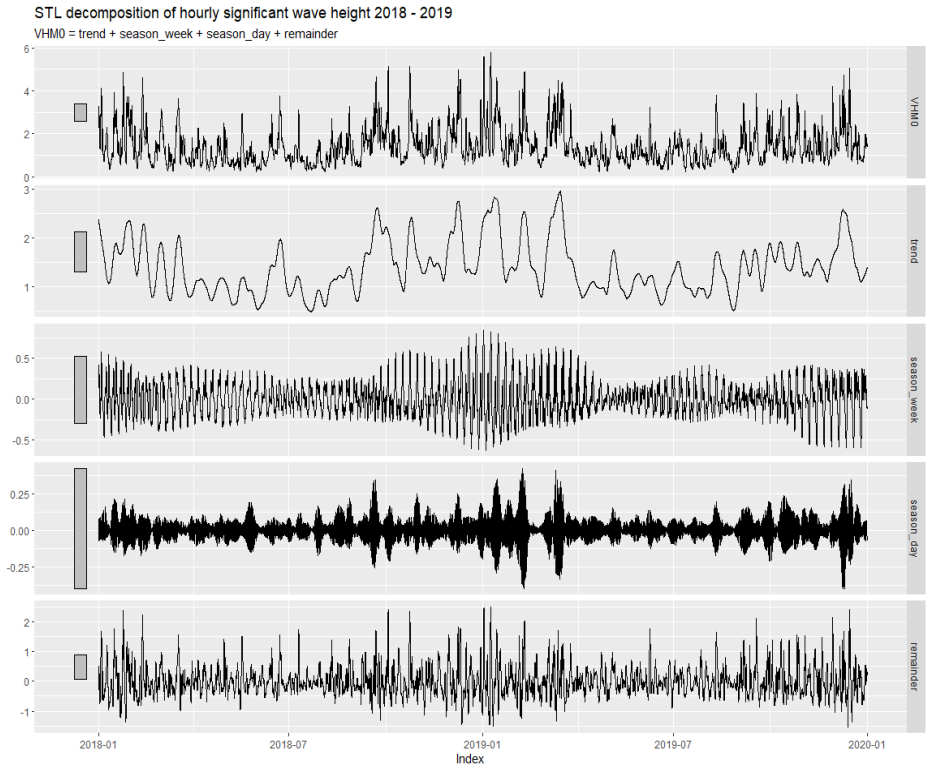


Figure 25: STL decomposition of hourly significant wave height 2019 - 2019

For the $Trend_t$ component, there is no specific trend from 2018 to 2019, however, within each year, there was a decrease from the beginning to the middle of the year before recovering from July to December. The $Trend_t$ component is also the strongest component with a strength of 0.58. The $Season\_day_t$ and $Season\_week_t$ both fluctuated wildly from October to March before shrinking their range in the middle of the year. For $Season\_week_t$ component, the fluctuate from around 1 and -0.75, whereas, in the summer, its range was only from -0.25 to 0.25 with a strength of 0.19. For $Season\_day_t$ component, in the winter, it fluctuated around -0.5 and 0.5 during the winter months and from -0.125 to 0125 during the summer months with a strength of 0.03.

In conclusion, the wind speed, wind power, and significant wave height time series are all non-stationary with multi-seasonality. While the wind speed at both daily and hourly lever and the hourly average wind power have a significantly strong yearly seasonal component with a V-shape pattern, the significant wave height time series did not have a yearly seasonal component after STL decomposition. Since there were only 2 years under study, there might not be enough data for the algorithm to capture the yearly component of the significant wave height time series. In addition, the decreasing trend in both wind speed and wind power in the Gemini site raises requires a better prediction and O&M scheduling to reduce the downtime cost and vessel chartering fee.

4. PREDICTION AND ANALYSIS RESULTS

In this chapter, we present a long-term prediction of 1 year ahead from the 1st of January 2020 to the 31st of December 2020 with average daily mean wind speed to sketch the weather window of one year and then select the month and date which are forecasted with the least wind speed for maintenance. Since the least wind speed will produce the least energy, we will schedule the jack-up vessel to conduct maintenance or replacement of the wind turbine accordingly. Since the location of the Gemini wind farm is further away from other Dutch offshore wind farms, the wind climate might be different. Therefore, they might have different days with the least wind to conduct O&M service. In this scenario, different wind farms can share the same jack-up vessel to reduce the chartering cost and utilize the vessel. In this chapter, sections 4.1 and 4.3 provide the answer for RQ 1 and RQ2, while section 4.2 is dedicated to RQ3.

After selecting the weather window by long-term prediction in 2020, we implemented short-term prediction on the hourly mean wind speed, hourly significant wave height, and hourly wind energy generated to assess the accessibility and safety of the wind farm in each forecasted weather window. Given the forecasted weather condition and power generated, the wind farm operator could manage the O&M activities accordingly. We suggest the wind farm operator apply this process of long-term wind speed prediction every year to prepare for the scheduling and sharing of the jack-up vessel while applying short-term weather window and power prediction to make a final adjustment of the maintenance execution to the actual weather.

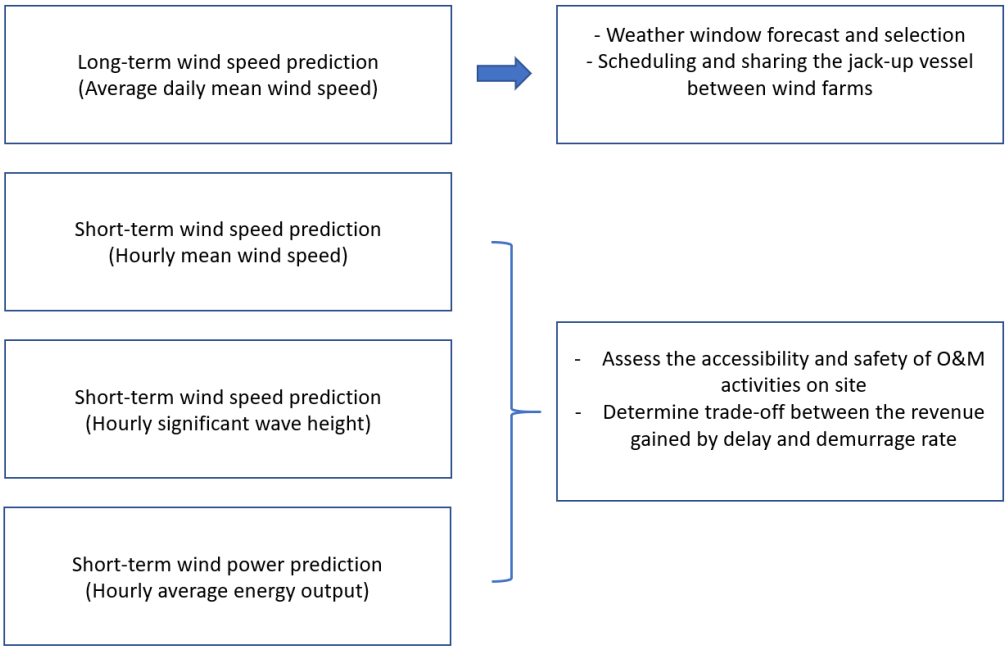


Figure 26: Process of using long-term and short-term forecast

4.1. Long-term wind speed prediction

Regarding the long-term daily wind speed prediction, we aim to find the best forecast model in terms of both accuracy and parsimony. In section 3.3.3 we established that there are two seasonal periods in the average daily wind speed time series, namely the weekly and yearly components. This multi-seasonality could not be captured by a seasonal ARIMA model. In addition, the strongest seasonal component in this time series has a repetitive V-shape pattern every year, therefore, we can assume a constant seasonal pattern over time to apply Dynamic Harmonic Regression. In the Fourier forecasting methods, the number of the pairs of Fourier terms for the weekly seasonality can be 3 at max and the number of the pairs of Fourier terms for the yearly seasonality can reach 182. Thus, we first selected the model with the least AICc score of 3928.50 whose number of weekly Fourier pairs is 3 and 2 pairs of annual Fourier pairs with first-order autoregression model - AR(1,0,0) for the error terms. We denote it as F-ARMA(3,2)x(1,0,0) model. This selection process is implemented by our algorithm in R and it returned the final results with the least AICc score.

As can be seen in Figure 27, the mean point forecast and the 95% prediction interval of this model are quite smooth and could not capture the variance of the true observations well. For the long-term wind speed prediction, we need an accurate prediction interval to sketch out the schedule of the O&M activities. Although F-ARMA(3,2)x(1,0,0) satisfied all the requirements of the residuals assumption which were further explained in Appendix 6.4, we need to improve it to produce better forecast distribution and capture the dynamic of wind speed 1 year ahead of time.

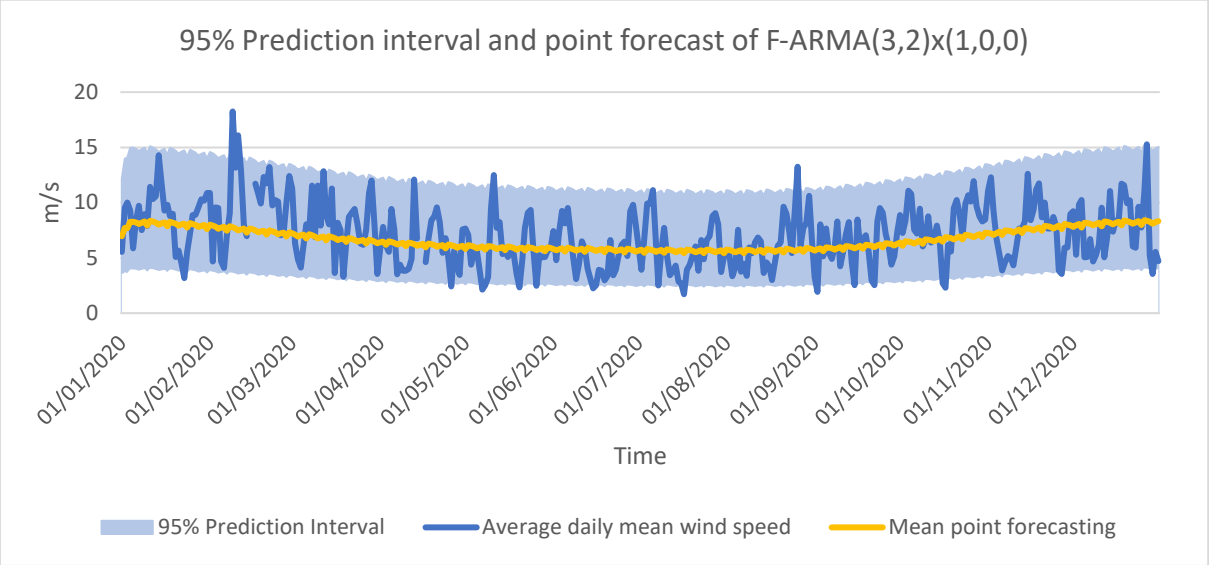


Figure 27: Mean point forecasting and 95% Prediction interval of F-ARMA(3,2)x(1,0,0) in 2020

We improved this model in two directions. The first direction is increasing the number of pairs of Fourier terms for yearly seasonality until the innovation residual is correlated with each other based on the Ljung-Box test while keeping the number of pairs of Fourier terms for weekly seasonality at 3. The second direction is including the Temperature at 10-meter height as external regressors of the model and finds the model with the lowest AICc.

Regarding the first approach, the number of pairs of Fourier terms can reach 20 pairs without making the innovation residuals correlated with each other. We denote it as F-ARMA(3,20)x(1,0,0). In Figure 28, we can see that the forecast distribution with a 95% prediction interval is more volatile and follows the pattern of the true observations better than that of F-ARMA(3,2)x(1,0,0). F-ARMA(3,20)x(1,0,0)'s point forecast is also more wiggly to capture more variance of wind speed dynamic. The only problem with F-ARMA(3,20)x(1,0,0) is a large number of estimators, which may not be optimal in terms of model parsimony.

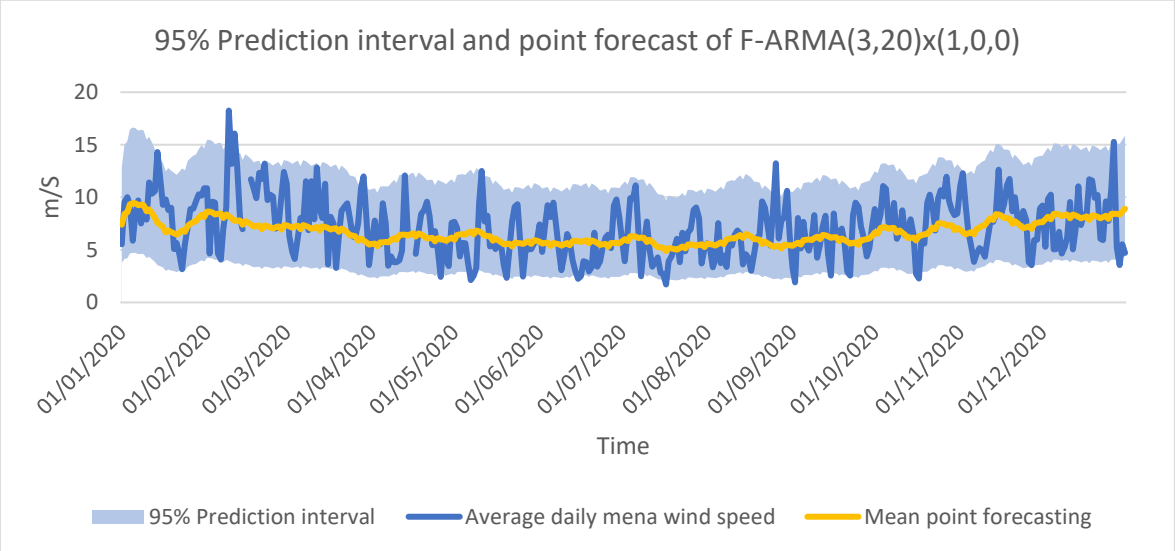


Figure 28: 95% Prediction interval and point forecast of the F-ARMA(3,20)x(1,0,0) in 2020

Regarding the second approach, we aim to select an external regressor that might affect the wind speed to include in the model. We selected temperature at 10-meter height as the only external regressors since the correlation coefficient (-0.23) is statistically significant. We can see that the pattern of the daily temperature time series repeated over time, especially the yearly seasonal components remained almost the same from 2017 to 2020 in Figure 29. Therefore, we could apply Fourier ARMA to predict the temperature and use this forecasted temperature as the exogenous inputs for our wind speed prediction model. We denote this model as F-ARMA(3,2)x(1,0,0)xTemp.

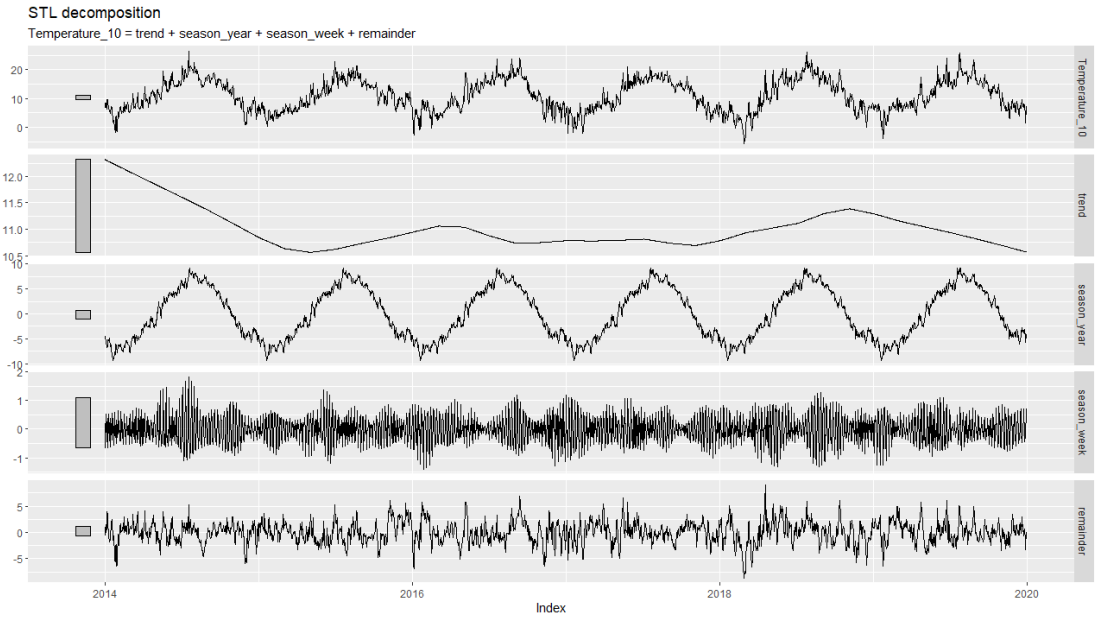


Figure 29: STL decomposition of Temperature at 10-meter height

The innovation residuals of F-ARMA(3,2)x(1,0,0)xTemp are uncorrelated based on the Ljung-Box test result (Table 14). Therefore, it could also capture all information of the data and the mean point forecast is unbiased. Besides, its innovation residuals are normally distributed with constant variance, therefore, its prediction interval and forecast distribution are credible which is further explained in Appendix 6.4. In Figure 30, we could see that the F-ARMA(3,2)x(1,0,0)xTemp's forecast distribution captures the dynamic of the wind speed better than F-ARMA(3,2)x(1,0,0) since its mean point forecast follows the wind speed pattern, for example, on the 1st of February 2020, both the real observation and mean point forecasting hit the bottom before recovering afterward.

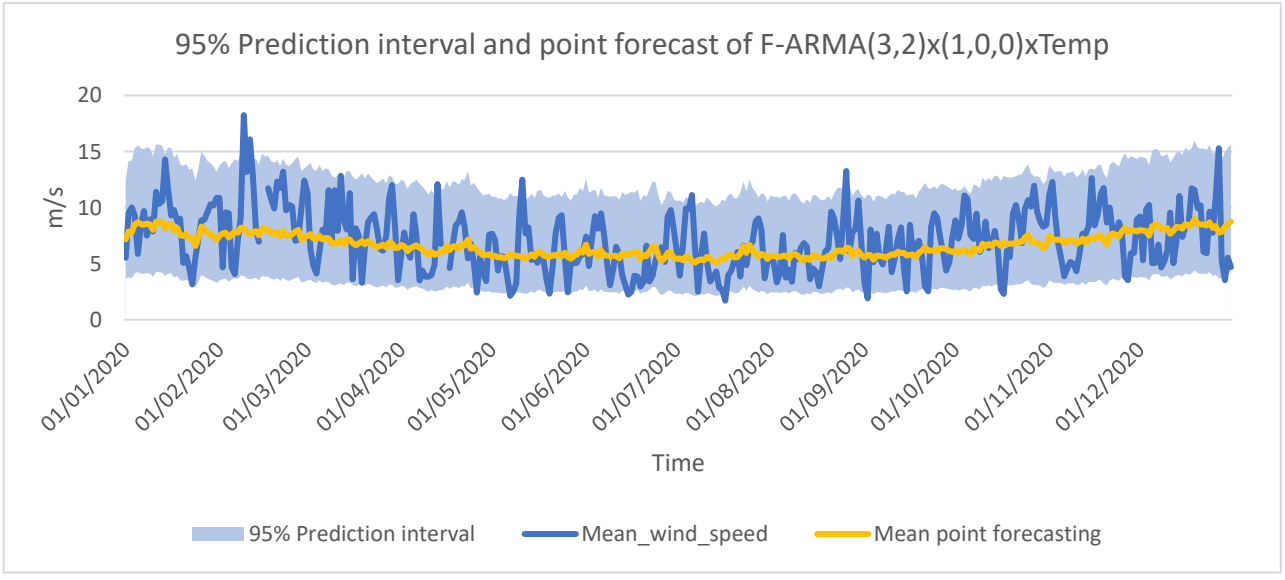


Figure 30: 95% Prediction interval and point forecast of F-ARMA(3,2)x(1,0,0)xTemp in 2020

In terms of prediction accuracy, F-ARMA(3,2)x(1,0,0)xTemp has better RMSE, MAPE than F-ARMA(3,2)x(1,0,0) and F-ARMA(3,20)x(1,0,0) with much less parameters compared to F-ARMA(3,20)x(1,0,0). Besides, the CRPS of F-ARMA(3,2)x(1,0,0)xTemp is also smaller than those of F-ARMA(3,2)x(1,0,0) and F-ARMA(3,20)x(1,0,0), therefore, its forecast distribution is more accurate. In addition, it has the lowest AICc score of 3905.86 among the three models, which makes it the most parsimonious model. Besides, the mean point forecast could capture the pattern of real observations in a way that its peaks and troughs occur close to the day of the real peaks and troughs of the wind speed. Therefore, F-ARMA(3,2)x(1,0,0)xTemp is selected to benchmark with seasonal naïve prediction. This is also proved by its CRPS score which is 5 times higher than that of F-ARMA(3,2)x(1,0,0)xTemp. Therefore, the prediction interval of this model might not be reliable. In addition, the RMSE of the time series cross-validation for the training set of F-ARMA(3,2)x(1,0,0)xTemp is 1 unit lower than that of seasonal naïve prediction (2.60 and 3.55 respectively).

Therefore, to answer **Research Question 1,** we selected a Fourier model with 3 pairs of Fourier terms for weekly seasonality, 2 pairs of Fourier terms for yearly seasonality, the average daily temperature as an external regressor, and a first-order autoregressive model for error terms to predict long-term wind speed. The F-ARMA(3,2)x(1,0,0)xTemp is written as follows where yb_t is the wind speed at time t after Box-Cox transformation, $\varphi_t = 0.465\varphi_t + \varepsilon_t$ is the first order autoregression of the error term ε_t with t = 0,1,2…:

$$\begin{aligned} yb_t = {} & 2.213 + 0.010sin\left(\frac{2\pi t}{7}\right) - 0.013cos\left(\frac{2\pi t}{7}\right) - 0.030sin\left(\frac{4\pi t}{7}\right) + 0.015cos\left(\frac{4\pi t}{7}\right) \\ & - 0.017sin\left(\frac{6\pi t}{7}\right) + 0.022cos\left(\frac{6\pi t}{7}\right) + 0.137sin\left(\frac{2\pi t}{365.25}\right) + 0.570cos\left(\frac{2\pi t}{365.25}\right) \\ & - 0.039sin\left(\frac{4\pi t}{365.25}\right) + 0.068cos\left(\frac{4\pi t}{365.25}\right) + 0.036Temp_t + \varphi_t \end{aligned}$$

$$\text{Where } \varphi_t = 0.465\varphi_t + \varepsilon_t$$

$$\varepsilon_t \sim NID(0, 0.345) \qquad (30)$$

Table 12: Accuracy measurements of wind speed forecasting models

	AICc	RMSE[10]		MAPE[11]		Winkler Score[12]	CRPS[13]	tsCV RMSE	Ljung-Box test P-value	Conclusion
		Train	Test	Train	Test					
F-ARMA(3,2)x(1,0,0)	3928.50	2.15	2.60	27.71	34.72	11.1	1.46	-	0.61	Uncorrelated residuals
F-ARMA(3,20)x(1,0,0)	3972.56	2.13	2.58	27.68	35.6	10.9	1.45	-	0.08	Uncorrelated residuals
F-ARMA(3,2)x(1,0,0)xTemperature	3905.86	2.13	2.60	27.38	34.0	11.20	1.36	2.60	0.71	Uncorrelated residuals
Seasonal Naïve model	Not available	Not available	3.58	Not available	51.53	142	5.08	3.55	0.00	Correlated residuals

With t = 0,1,2…, the pairs of Fourier terms represent the weekly and yearly seasonality of the average daily mean wind speed after Box-Cox transformation in the corresponding timestep. In

[10] Error calculated by forecast package which use indexing from t = 1

[11,12, 13] Error calculated by forecast package which use indexing from t = 0

addition, After Box-Cox transformation with t = 0,1,2…, if other factors stay the same, Temp(t) increases by 1 unit will make the wind speed ybt increase by 0.036 unit.

Although this type of model has a MAPE of 35.5 which means on average, its error of point forecasting is 35% as large as the real wind speed value, the purpose of long–term wind speed forecasting is to establish the wind speed window which focuses more on the forecasting distribution. Therefore, the smallest CRPS of 1.43 is acceptable.

4.2. Weather window prediction results

At Gemini, SOV and helicopters are used for daily maintenance. For helicopters, they are only occasionally used in emergencies and their wind speed limit is around 20 m/s which is the same as the wind limit of the jack-up vessel's lifting activity (Hu and Yung, 2020a). Therefore, we did not consider helicopters in this research. Based on industry knowledge from Blocklab.nl, the Windea La Cour is not able to carry out major maintenance and replacement of big wind turbine components such as turbine blades, gearboxes, and generators. In addition, the Gemini wind farm has been operating for more than 4 years which increases the failure rates of these main components. In this case, the Gemini wind farm operator must schedule a jack-up vessel to conduct maintenance and replacement for these major components. Therefore, we selected Van Oord's Aeolus jack-up vessel as a research object for this thesis since it was also used to install monopiles for the Gemini wind farm during the construction phase (Gemini Wind Park, 2018).

In this thesis, we assumed that the jack-up vessel only carries the gearbox for a major replacement to simplify the calculation. This might not be true in real life since to fully optimize the capacity of the jack-up vessel, they often load many components onto one vessel at once. However, since this thesis concentrates on optimizing the logistics and supply chain of offshore wind farm maintenance, simulating the whole operation with one major component of the wind turbine is the first step for the financial model. In addition, the jack-up vessel can only lift one component to repair one turbine at a time. Therefore, the chartering duration of the vessel will add up concerning the number of major components it carries and the repair time of the respective components.

Based on Carroll et al; (2015), the gearbox has the highest failure rate of 0.154 failures per turbine per year with a considerable average replacement time of 231 hours (excluding travel times, lead time, and waiting time due to inaccessibility). We used the data from this research since the turbine model used in the Gemini wind farm has similar characteristics to the wind turbines of this research sample which were offshore wind turbines from 3 to 10 years old with a capacity between 2 and 4 MW. The failure rate in this research was small since this failure rate was defined as the "visit to a turbine, outside of a scheduled operation, in which material is consumed" which did not consider the total stoppage of the turbine during its operation.

After that, to create the weather window of wind speed based on the long-term prediction for 2020, the mean point forecast, upper bound, and lower bound of the 95% prediction interval in F-ARMA(3,2)x(1,0,0)xTemp is converted from 10-meter height to 89.5-meter height to compare with the cut-in and cut-out wind speed value (5 m/s and 25 m/s respectively) of the turbine model. This means that below 5 m/s and above 25 m/s, the wind turbine stops producing energy. With this wind speed forecasting distribution, the wind farm operator could calculate the average daily wind energy based on the power coefficient of the wind turbine to estimate the largest and smallest energy generated per day, per week, and month in 2020. However, since this power coefficient is confidential, we could not estimate the wind power generated throughout 2020. Therefore, we assume that the day with the lowest average daily wind speed will produce the least wind energy and vice versa.

From Figure 31, based on our forecasts in 2020, at hub height level, the wind speed was able to produce energy almost all year round since the upper limit of 95% prediction interval was lower than both the cut–out wind speed (25 m/s) and lifting wind speed limit (20 m/s). The cut-in wind speed lies within the forecast distribution, therefore, on each day, there would be a time that the wind speed was either lower than 5 m/s or higher than 5 m/s.

For the jacking operation of the jack-up vessel, the wind speed must be lower than the jacking wind speed limit (15 m/s). Based on our calculation and Figure 32, we could not conduct a jacking operation in January and December since the upper limit of the 95% prediction interval might be higher than 15 m/s. Specifically, from the 4th to the 9th, 11th to 13th, 15th to 16th of January 2020, and from the 5th to 7th, 16th to 24th, 26th to 31st of December 2020.

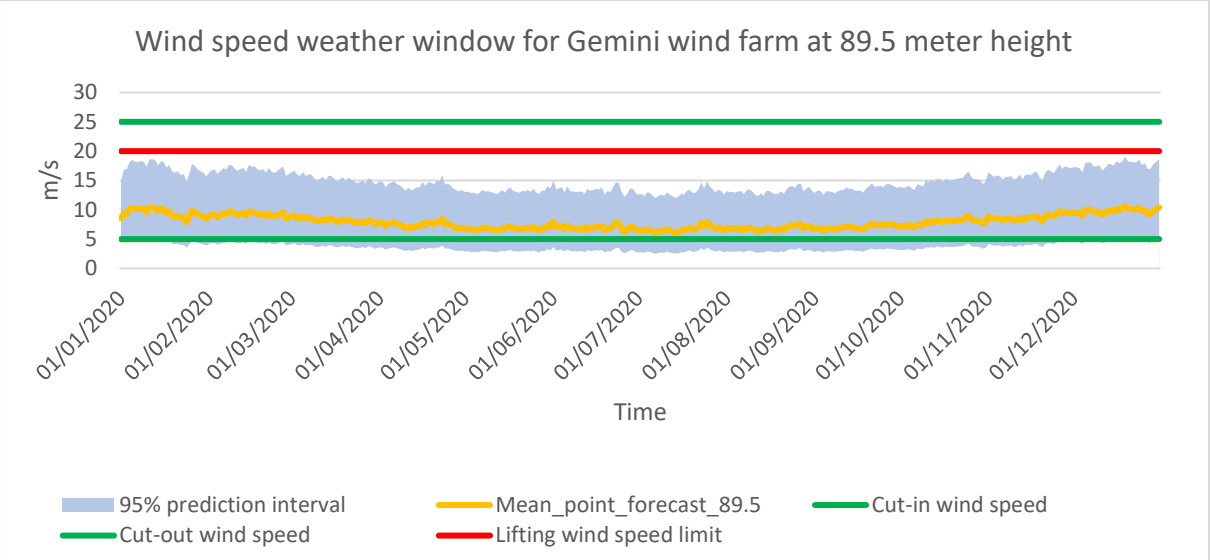


Figure 31:Wind speed weather window for Gemini wind farm at 89.5-meter height in2020

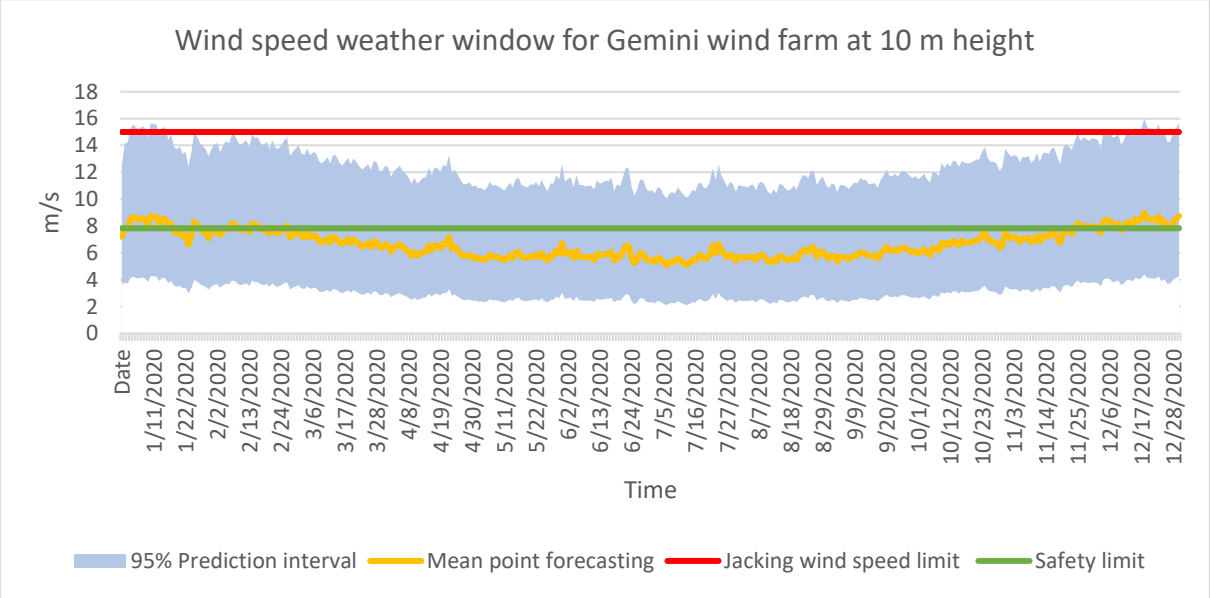


Figure 32: Wind speed weather window for Gemini wind farm at 10-meter height in 2020

Based on Adverse Weather Sailing and Adverse Weather Working Guidelines (UK offshore Operators Association/ Chamber of Shipping, 2002), when the mean wind speed at a 10-meter height is from 20 to 25 knots (approximately 7.83 to 9.78 m/s) and/or the wave height is from 3 to 4 m/s, the shipmaster must conduct a risk assessment, before beginning or continuing any operation alongside an installation, and crane driver. Therefore, we used the wind limit of 7.83 m/s as the reference for human safety to work outside of the vessel and compared it with the mean point forecast at 10-meter height. Because on average, it takes around 10 days to repair one gearbox (Carroll et al; 2015), the jack-up vessel could not operate in January, February, and December 2020 due to insufficient weather windows. Specifically, we forecasted that the weather window in 2020 could not allow the jack-up vessel to operate from the 2nd to 18th and 26th to 27th of January, from the 6th to 17th and 22nd to 27th of February, from 25th of November to 31st of December.

Since the summer is the peak season for chartering a jack-up vessel to conduct maintenance and replacement service, the chartering rate is the most expensive in June, July, and August. In addition, the wind speed decreased from March to the middle of the year before recovering until the end of the year based on section 3.1.2. Therefore, the downtime cost might increase from September to November and lead us to forecast weather windows in March, April, and May in 2020. Then within each month, we picked the day with the lowest forecasted average daily mean wind speed as the starting point to predict the weather window of 1 day before and 1 day after that for maintenance since we assume the lowest average daily wind speed will produce the least energy and revenue. After that, we can determine the hourly mean wind speed, hourly significant wave height, and hourly wind power of the weather window around that day for downtime cost calculation and accessibility of the wind farm. This process is illustrated in Figure 33 below. In March, April, and May of 2020, the day with the lowest forecasted average daily mean wind speed point forecast is 30th March, 28th April, and 27th May 2020 respectively.

Therefore, to answer **Research Question 3,** each year, we suggest making a one-year ahead wind speed prediction to find the favorable weather window with the least wind speed each month. Then, the wind farm operator could start to schedule the jack-up vessel and share the vessel with other wind farms if possible. Based on the long-term forecast of 2020 for the Gemini wind farm**,** we recommend conducting maintenance with the jack-up vessel from March to May. More specifically, the jacking and lifting activities of jack-up vessels might be around the 30th of March, 28th of April, and 27th of May 2020 respectively for the least forecasted wind in the selected weather window. We recommend building another medium-term prediction of 3 months to verify the local minimum of each selected weather window, however, due to time scope, we skip this step in our thesis. The chartering period of the jack-up vessel depends on the number of turbines which need maintenance and replacement, however, for major maintenance and replacement of the gearbox, 1 day per turbine on average is recommended (Carroll, Mcdonald, and Mcmillan, 2015). For other wind farms, each year, they should update the weather and power data to construct a new forecast for the next year before planning and chartering the vessel.

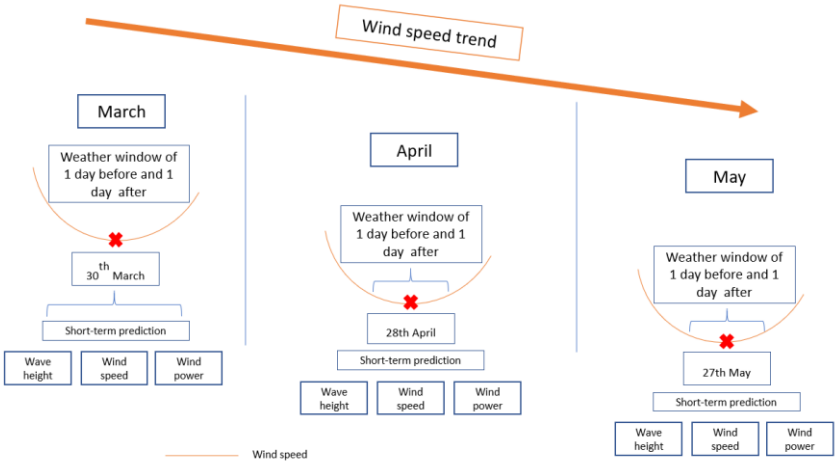


Figure 33: Simplified process of weather window selection

4.3. Short-term wind speed and wave height prediction

Suppose the jack-up vessel was booked to implement maintenance and replacement on the 30th of March or the 28th of April 2020 based on the long-term wind speed prediction, for each scenario, delays might occur due to inevitable external factors on-site or the wind farm operator wants to gain more revenue from high wind speed on that day. In this thesis, we want to find out whether it is cost-effective to delay the maintenance service for more revenue gained from the wind farm or not. Based on The Crown Estate, (2014), there can be a 1- or 2-days delay in the execution of a jack-up vessel on-site due to various reasons. Therefore, for each selected day in the chosen weather window, we will conduct short-term predictions of the wind speed, wind power, and wave height to measure the

downtime cost of the preventive maintenance given these delays. As a wind farm operator or the captain of the jack-up vessel, we have to consider the chance of weaker or stronger wind speed given the current weather conditions. For example, if today, the wind speed is already strong, there will be a higher chance that on the next day the wind speed will be weaker, therefore, with short-term prediction, we could make the decision-makers more informed about their next step. This is even more important since based on the wind speed analysis in chapter 3, there was a decreasing trend in hourly mean wind speed from 2014 to 2019.

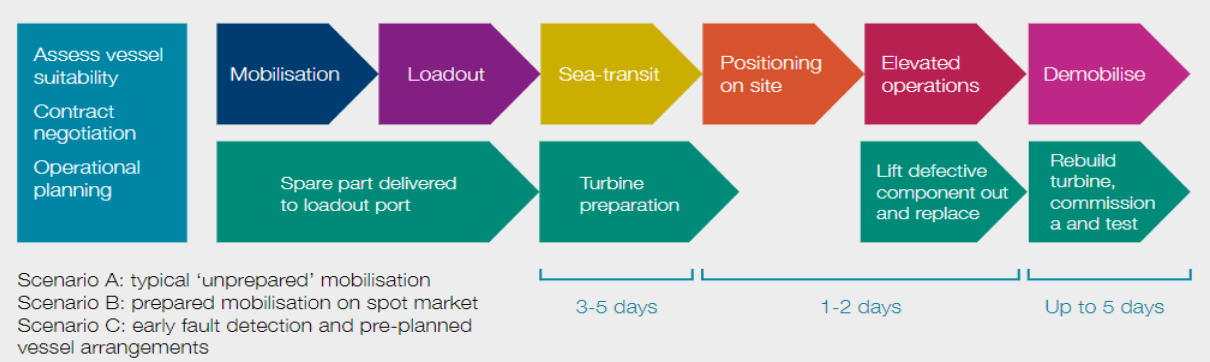


Figure 34: Overview of planning requirements and potential for delays depending on repair scenario utilized. Source: The Crown Estate, (2014)

Due to the limitation of the Non-linear Autoregressive Neural Network with and without exogenous inputs (NARX and NAR), we can only predict the power 24 hours ahead of time. Therefore, with the scenario that the jack-up vessel is mobilizing to the wind farm on the 27th of March 2020, we will forecast the wind speed, wind power, and wave height on the 28th of March 2020 based on the real observations until the present value of wind speed on the 27th of March 2020. In another word, we have to decide whether to continue the maintenance service or delay this operation to the 28th of March 2020 since the real wind speed on the 27th of March 2020 was quite strong, there would be a higher chance that the wind speed of the next day might be lower. The forecast for the 29th of March will be based on the real values until the 28th of March, and so on for the 31st of March. The same process will be applied to the scenario that the jack-up vessel is sailing to the wind farm on the 27th, 28th, and 29th of April 2020. However, in this case, the real wind speed on the 27th of April 2020 was quite weak, therefore, the chance of a weaker wind speed on the next day might be lower. For the time scope of this thesis, we will not consider the downtime cost of the 27th of May and its delays.

Regards the short-term prediction, we focused on the mean point forecasting of the wind speed and wind power one day ahead of time. Therefore, only RMSE and MAPE were considered to evaluate the model accuracy. Firstly, we compared NARX with wind speed as exogenous inputs and NAR for wind power prediction. Although the NARX(44,1,23)$_{24}$ performed slightly better than NAR(44,1,22)$_{24}$ with the MAPE of 24.4% and 26.5% respectively for the prediction of hourly mean wind speed on the 28th of March 2020, we found that NARX(44,1,23)$_{24}$ and NAR(44,1,22)$_{24}$ both had normally distributed and homoscedastic residuals. However, the residuals of NARX(44,1,23)$_{24}$ were more or less correlated based on the autocorrelation function plot in Figure 35 while the residuals of NAR(44,1,22)$_{24}$ were not. For both NARX and NAR models, we could not conduct the Ljung-Box test since R could not find appropriate degrees of freedom for these models. Therefore, we decided to use a Neural Network Non-linear Autoregression model without Exogenous inputs for short-term prediction only since this method could capture all the information in the data series (Figures 35 and 36).

Regards the hourly mean wind speed, the NAR(44,1,22)$_{24}$ and NAR(44,1,40)$_{24}$ were constructed to predict the weather on the 29th and the 30th of March 2020. They provide reliable forecasts on the 29th, the 30th of March 2020, and the 27th of April 2020 with the MAPE value of less than 15%. However, the prediction of the hourly wind speed on the 31st of March 2020 and the 29th of April 2020 are not reliable with the MAPE values of 84.3% and 155% respectively. As can be seen in Figure 36, the forecasted wind speed on the 31st of March 2020 and the 29th of April 2020 captured the pattern of wind speed on the same days in 2018 and 2019. Because the Neural Network is a black-box model, we

could not explain its relationship between parameters and lagged input like a normal autoregression model. However, if we compared the forecasted values with the real values of that day and the same days in previous years, we might have some clues. Since the NAR(44,1,22)$_{24}$ learned from 44-hour lagged data, its 24-step ahead prediction could not deviate too much from the past data. Therefore, it could not capture the sudden large drop in the real observations of the wind speed on these days in 2020 well. This sudden change might be because 2020 was an abnormal year in the Netherlands in which many unpredicted and never-before-seen extreme weather events happened, namely the spring drought[12] and the warmest March in history[13]. Therefore, we can only rely on the prediction of the 29th, 30th of March 2020, and the 27th, 28th of April 2020 for wind speed analysis.

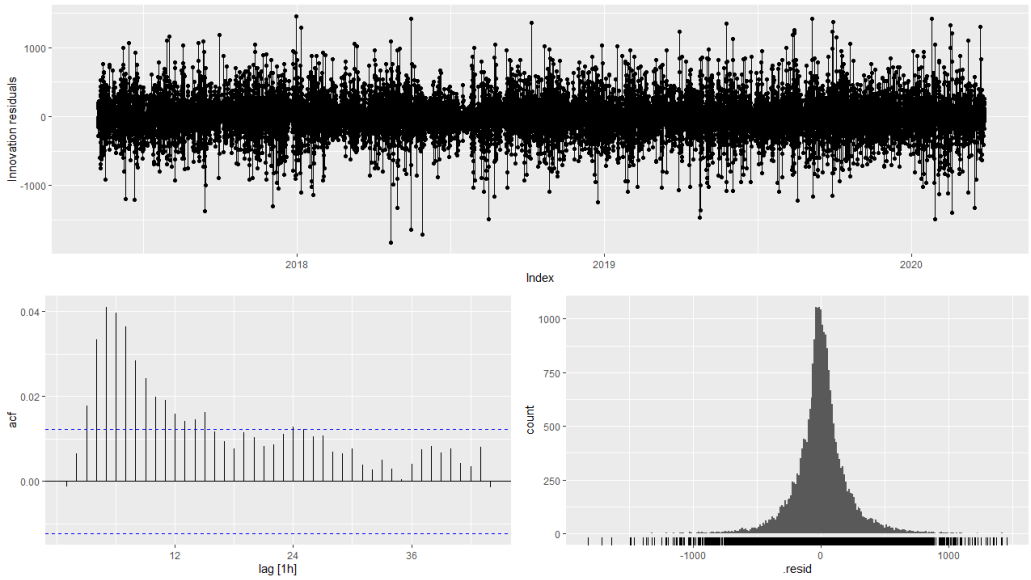


Figure 35: Residual checks of NARX(44,1,23)$_{24}$ for the 28th March 2020

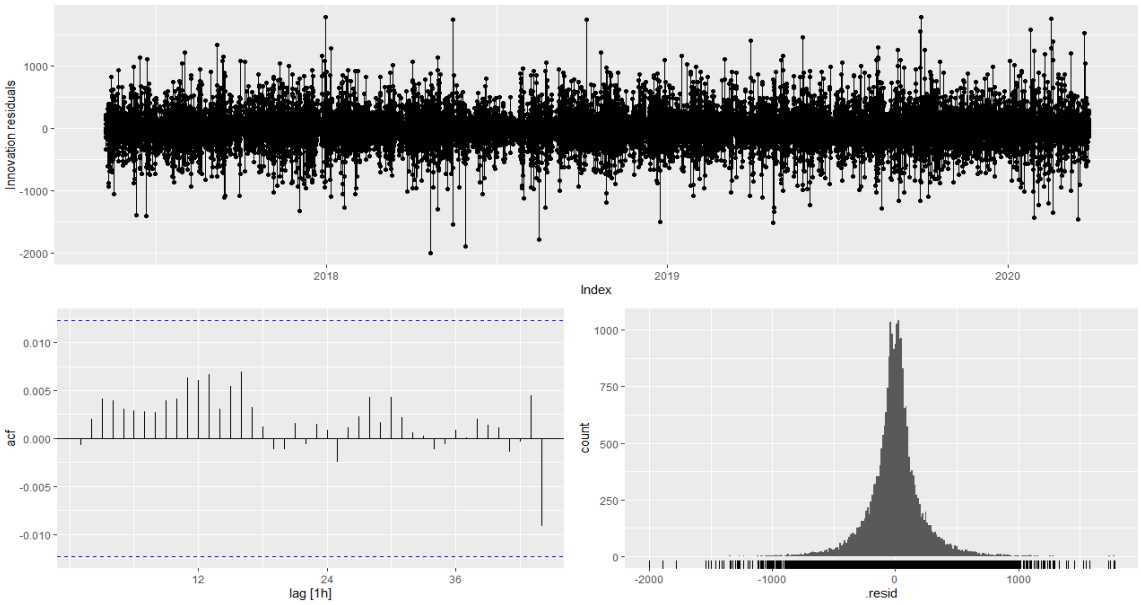


Figure 36: Residuals check of NAR(44,1,22)$_{24}$ for the 28th March 2020

[12] https://www.amsterdamiww.com/water-solutions-3-risks-and-resilience/more-drought-in-the-netherlands/

[13] https://www.iamexpat.nl/lifestyle/lifestyle-news/tuesday-officially-warmest-march-22-ever-recorded-netherlands

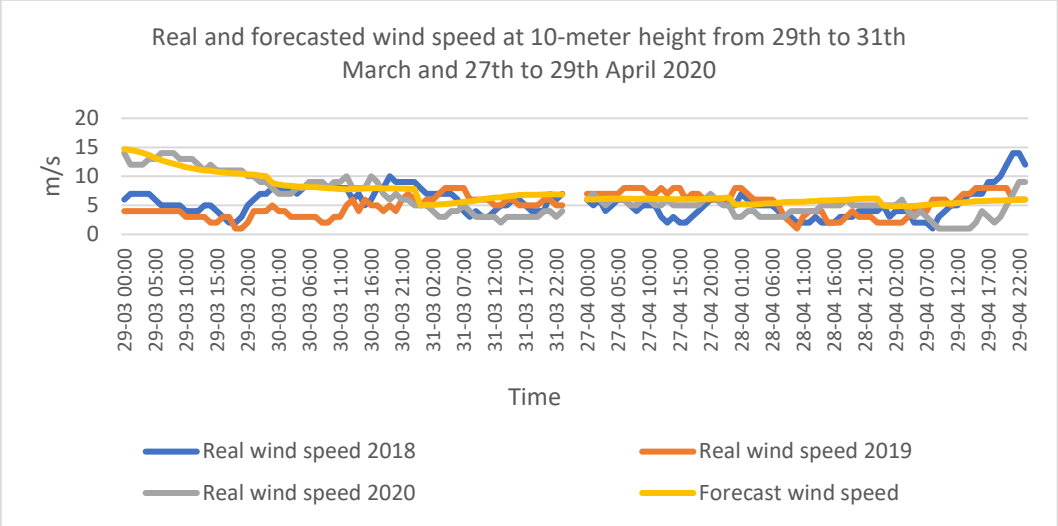


Figure 37: Real and forecasted wind speed from 29th to 31st March and 27th to 29th April 2020

Regards wind power, the NAR(44,1,22)$_{24}$, NAR(44,1,40)$_{24}$, and NAR(26,1,30)$_{24}$ model only provides an accurate forecast on the 29th, 30th of March with the MAPE value of 15.58%, and 31.8%. However, the forecast of wind power on the 31st of March 2020, 27th, 28th, and 29th of April 2020 was overestimated with quite large prediction errors (up to 2096%). These large errors might be due to the inconsistency in the energy output data of the Gemini wind farm which was aggregated from The Dutch National offshore wind energy production. We aggregated the nominal wind energy capacity of this wind farm without considering the energy lost due to daily and major maintenance during the period under the study. Therefore, with the hourly mean wind speed of 7.14 m/s at hub height level, each turbine produced 622.87 kWh on average on 27th April 2018, 2133.62 kWh on average on 27th April 2019, and 1910.34 kWh on average 30th March 2020. However, for the same model of a wind turbine, it should produce the same amount of wind energy per given wind speed. Moreover, the real power generated was close to zero in many observations (Figure 38), they also inflated the MAPE exponentially based on the formula (14). However, since the fable package in R will construct the new model again with the new data, it will provide more accurate data with the real energy output provided by the wind farm. Therefore, to make the use case and prototype of the solution for Blocklab, we will proceed with the forecasted wind energy for further analysis of downtime cost.

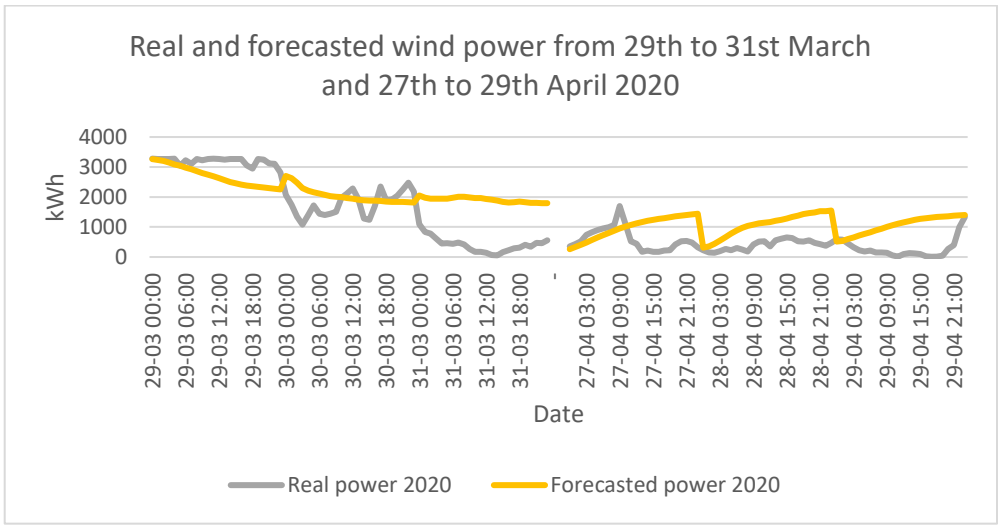


Figure 38: Real and forecasted wind power from 29th to 31st March and 27th to 29th April 2020

Regards the significant wave height forecast, the NAR(42,1,22)$_{24}$ model predicted the significant wave height with acceptable accuracy with MAPE values from 10% to 36% on the 29th, 30th of March 2020, from 27th to 29th of April 2020, except for the prediction on the 31st of March whose MAPE was around 50%. Therefore, we will use this model's results for further analysis.

Although other papers using wind speed, wind power, and wave height at different locations, at different granularity, with different forecast horizons, and with a different type of neural network, the MAPE value of less than 25% is considered as a reliable prediction in (Kadhem *et al.*, 2017; Kumar *et al.*, 2017; Dhiman and Deb, 2020; López and Arboleya, 2022). Therefore, $NAR(40,1,22)_{24}$, $NAR(44,1,22)_{24}$, $NAR(26,1,30)_{24}$, $NAR(42,1,22)_{24}$ for wind speed, wind power, and significant wave height prediction on the 29th and 30th of March is reliable.

For more detailed prediction accuracy of all prediction models, please refer to Table 19 in Appendix 6.6.

In conclusion, to answer **Research Question 1,** the fable package constructed a Non-linear Autoregressive Neural Network structure that iterates 20 networks and uses 1013 or 1841 weights, each network uses 44 lags, 1 seasonal lag of 24 hours, and 22 or 40 hidden nodes - $NAR(44,1,22)_{24}$, $NAR(44,1,40)_{24}$ to predict the wind speed in 24 hours ahead of time on the 29th, 30th March 2020, and 27th, 28th April 2020. Regards **Research Question 2, the** fable package constructed a Non-linear Autoregressive Neural Network structure that iterates 20 networks and uses 969 weights, each network uses 42 lags, 1 seasonal lag of 24 hours, and 22 hidden nodes to predict the significant wave height is 24 hours ahead of time. However, since NAR is a black-box model, we could not interpret the model easily and need further research to interpret it.

4.4. Economic analysis of offshore wind farm O&M activities based on the wind speed prediction.

After forecasting the energy output per turbine, we multiplied the energy output in kWh by its price of 17cent/kWh as in section 3.1.5.1.

From 29th to 30th March 2020, the model $NAR(44,1,22)_{24}$ and $NAR(26,1,30)_{24}$ captured the decreasing trend in power generation and downtime cost of the turbine (Figure 39). In addition, the total downtime cost error of forecasted energy output was less than 15.5% of the real energy output price on both days. We consider; therefore, this forecasting total energy output and downtime cost were reliable on these days of 2020. If we conducted the maintenance on the 29th of March 2020, we would lose 11008.35 euros per turbine based on the forecasted energy output, whereas, if we conducted the maintenance on the 30th of March 2020, we would only lose 8336.64 euros per turbine based on the forecasted energy output. Therefore, if we delayed the maintenance activity for one day, we would save 2671.71 euros per turbine. However, based on Figures 41 and 42, the 29th and 30th of March 2020 were not safe for jack-up vessel Aeolus to conduct jacking and maintenance. Thus, we recommend conducting daily and small maintenance on these days using a vessel with better sea-keeping capability such as the service operation vessel Windea La Cour (ModernPowerSystem, 2016). This short-term prediction, also verifies the result of F-ARMA(3,2)x(1,0,0)xTemp that the wind speed on the 30th March was weaker than on other days.

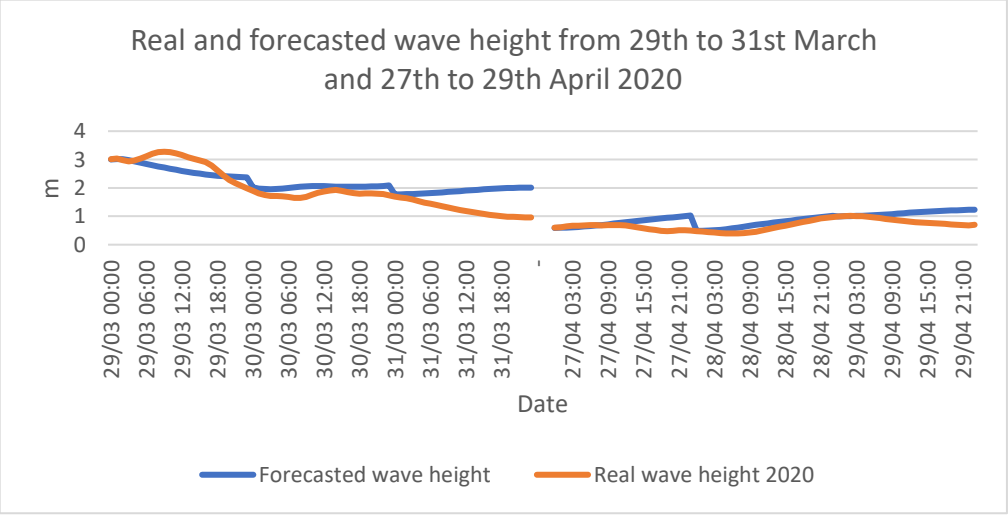


Figure 39: Real and forecasted wave height from 29th to 31st March and 27th to 29th April 2020

From 27th to 28th April 2020, the model could not capture the decreasing trend in power generation and the downtime cost of each turbine, however, from the 28th to 29th of April 2020, the prediction model could predict a slight decrease in wind power (Figure 40). In addition, the total downtime cost error of forecasting energy output was more than half of the real energy output price, therefore, we recommend using other meteorological models to predict the wind speed and then the power in case you do not have the exact energy output time series at hand. However, if we assume that these forecasts are correct, delaying the maintenance service from the 27th to the 28th of April 2020 will make us lose 365 euros per turbine, whereas, if we keep running the turbine until the 28th of April 2020 and conduct maintenance service on the 29th of April 2020, we save 107 euro. Moreover, the weather window on these days allows the Aeolus jack-up vessel to conduct maintenance.

Therefore, regards **Research Question 4,** due to the limitation of our forecasting horizon in the Non-linear Autoregressive Neural Network method and the data collection method, we could not determine the trade-off of the delay in maintenance activities for more than 1 day in terms of wind power generated and its corresponding revenue. However, in the case of Gemini offshore wind farm with Aeolus jack-up vessel, based on our forecast in 2020, if we delay the maintenance service for 1 day from the 28th to the 29th of April 2020, we will save 107 euros in downtime cost per turbine.

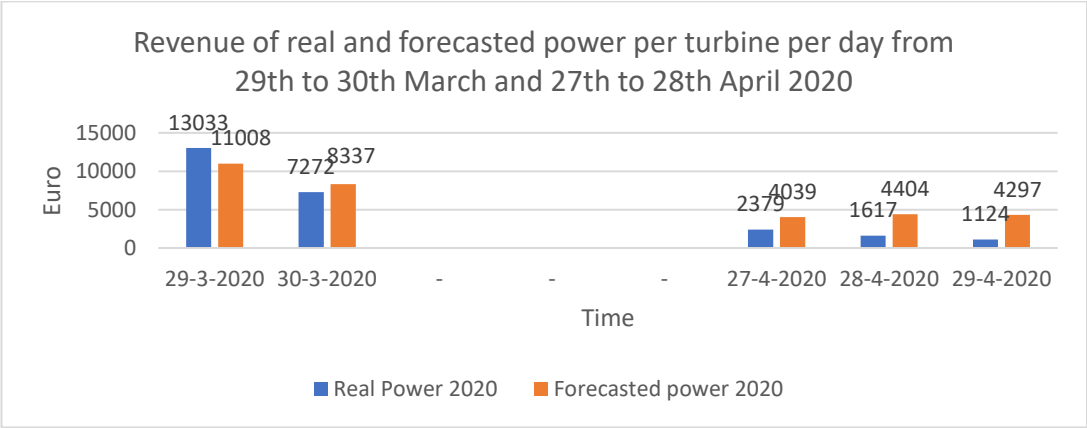


Figure 41: Downtime cost of real and forecasted power generated per turbine from 29th to 30th March and 27th to 28th April 2020

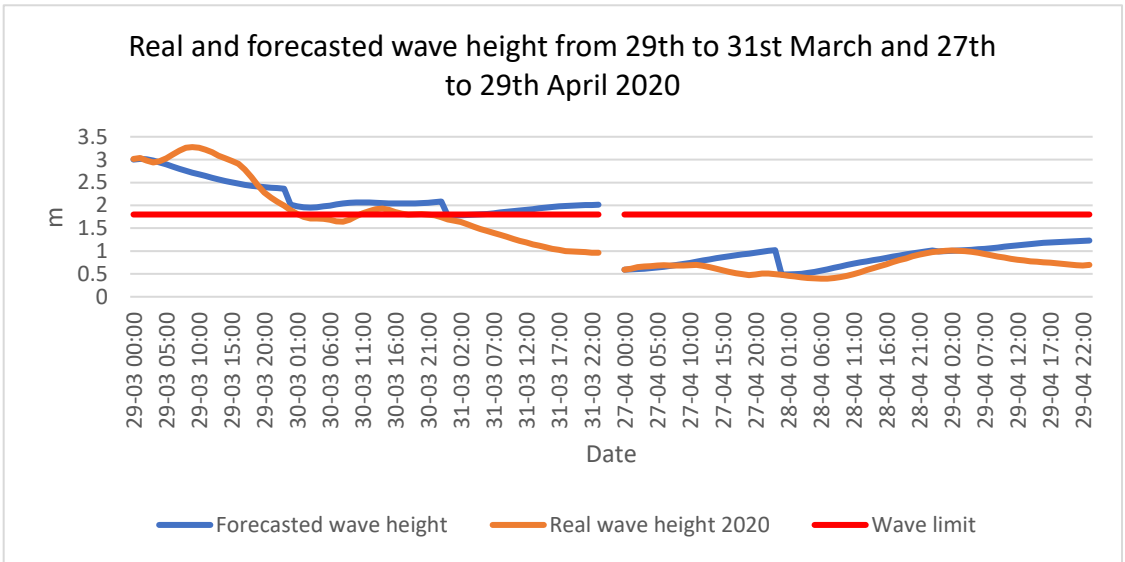


Figure 40: Forecasted wave height window for O&M activities

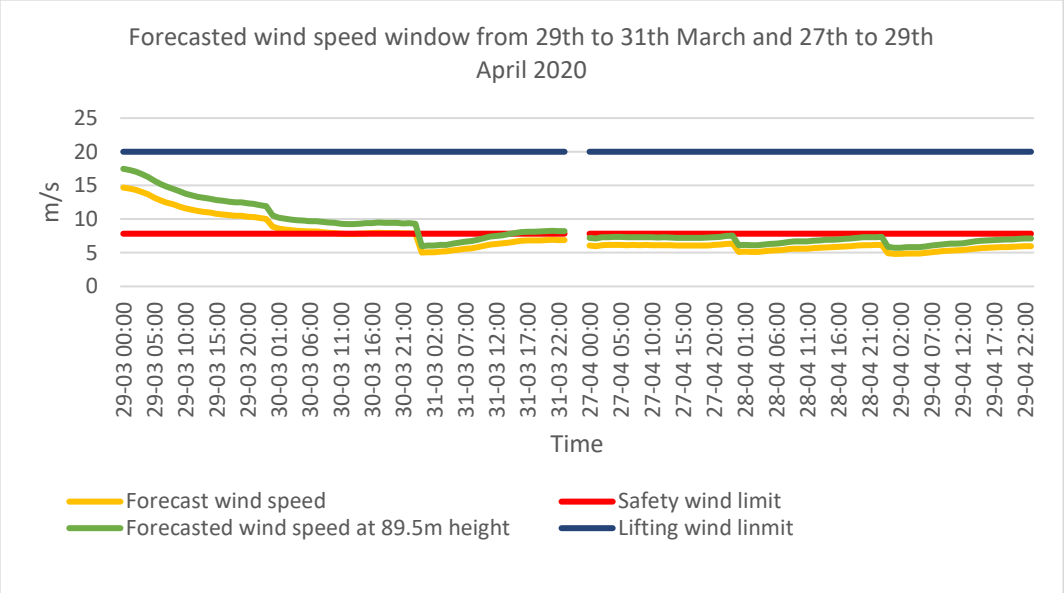


Figure 42: Forecasted weather window of wind speed at 10-meter and hub height level

Based on Table 10, the demurrage rate of a jack-up vessel (including inflation from 2015) is around 119813 euros[14] per day (summer charter rate). Therefore, if we assume on the 28th of April, all turbines generate 4404 euros per day in revenue and 4297 euros per day on the 29th of April 2020 and if we delay the maintenance service to the 29th of April, each day we need to conduct maintenance to 119813/4297 ≈ 28 turbines to offset the demurrage rate of the jack-up vessel alone. However, in practice, it is impossible to conduct maintenance for 28 turbines in a day since the jack-up/jack down and hub removal times are 11 hours on average per turbine. In addition, it takes on average 22 hours to conduct major repairs for one gearbox (Carroll, Mcdonald, and Mcmillan, 2015). This downtime will create a tremendous loss in revenue for the wind turbine which our model, unfortunately, cannot foresee plus the daily chartering rate and demurrage rate. Therefore, regards **Research Question 5,** the wind farm operator should not delay the maintenance activities of the jack-up vessel for more energy generated since the demurrage rate and other related costs of the jack-up vessel is too large compared to the revenue gained by the energy generated by the wind turbines.

4.5. Suggestion for the weather window prediction in the renewable market application.

Given the wide range of geophysical, geotechnical, and environmental differences present between different offshore wind farms in the Dutch North Sea and different turbines located on a single windfarm means that on the same day or in the same months, the wind climate and energy output can be much different. The Fourier F-ARMA(3,2)x(1,0,0)xTemp prediction model provides a weather window of up to one year to plan. Therefore, if a group of offshore wind farms first run this model to find their favorable weather window to conduct the maintenance, they can then schedule and share the jack-up vessel to repair the turbine in one wind farm after another as in Figure 43.

Besides, the short-term weather and significant wave height forecast based on a Non-linear Autoregressive Neural Network could evaluate the accessibility and safety to execute the replacement and major maintenance of huge components in the wind turbines on-site when the preventive maintenance starts.

[14] We assume 10% inflation rate and the exchange rate is on the 23rd of April 2022: 1 Pound Sterling = 1.19 EUR

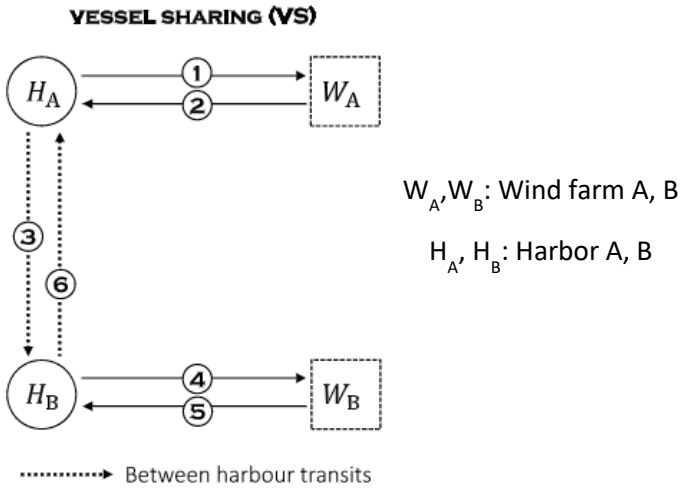


Figure 43: Vessel sharing policy illustration. Source: (Uit Het Broek et al., 2020)

By using these models to support the planning and deployment of the jack-up vessel, we also recognize that it is not possible to trade off between the revenue gained by delaying the maintenance service with the chartering and demurrage rate of the jack-up vessel if the wind farm charters it alone. However, if we develop the planning and scheduling of the jack-up vessel in a resource-sharing approach between different offshore wind farms, the revenue will increase from around 52 million to 110 million Pound Sterling in the case of the UK offshore wind farms (The Crown Estate, 2014). In the case of vessel sharing policy based on (Uit Het Broek *et al.*, 2020), some maintenance costs which depend on chartering and leasing time will partly become fixed costs. Thus, if we purchase a vessel and share it between wind farms, we could take advantage of the economy of scale and lower the fixed cost with larger collaboration. In this way, five participating wind farms with service providers deliver the best collaboration with the highest cost-benefit of 45% (6.5 million euros), the highest utilization of 89.4% compared to vessel leasing. However, to execute vessel sharing we must optimize the pre-operational stages, especially the planning process, which takes up to 6 months for one jack-up vessel provider (The Crown Estate, 2014)

Given the opportunity of vessel sharing in optimizing the O&M activities, Blocklab can integrate our long-term prediction models to harmonize the scheduling and chartering of a jack-up vessel by establishing the schedule and route of a jack-up vessel shared by the different wind farms to reduce the paperwork, less lead time and less lost in power production during this process. The short-term forecasting models, on the other hand, support ad-hoc evaluation of offshore windfarm accessibility on-site to adjust the plan and to ensure the safety of the crew. Therefore, the long-term and short-term prediction models could all be applied to optimize the project planning, project consent, contract negotiation, and technical information exchange not only within the group of offshore wind farms but also between them and the vessel providers (Figure 44).

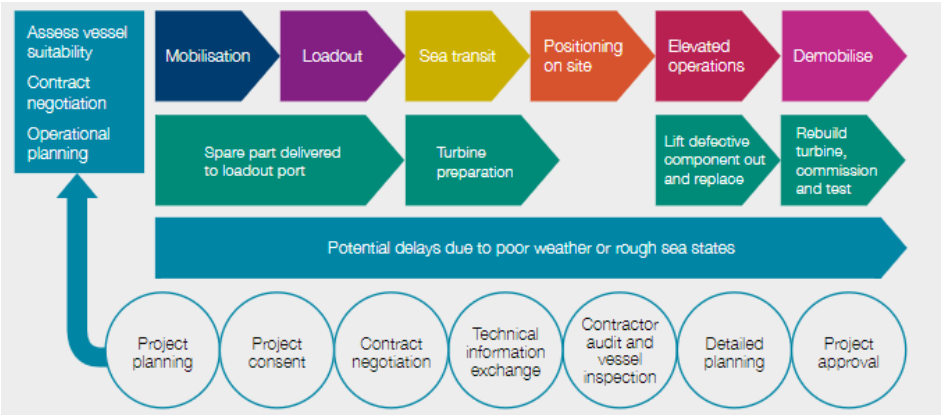


Figure 44: Interaction between planning, jack-up vessel operations, and turbine maintenance tasks

5. CONCLUSION

In this chapter, we summarize the answers to our research question and discuss the limitation of the thesis.

5.1. Conclusion

After developing a forecast for long-term wind speed, short-term wind speed, significant wave height, and wind energy output for the Gemini wind farm in 2020, we can answer the research problem:

"How offshore wind farm operators can apply the weather window of wave height and wind speed forecasting in planning the jack-up vessel in an efficient way while ensuring some safety standards?"

as follows.

First, we conducted the long-term average daily mean wind speed prediction by the Fourier-ARMA model to sketch out the wind speed forecast distribution for 2020 **(Research question 1)**. This prediction model can capture the dynamic and the variance of the real average daily wind speed time series in 2020 with its 95% prediction interval. Based on this long-term prediction of 2020, we forecasted a decreasing trend of the wind speed at the beginning of the summer and compared the weather condition with the safety standards of the jack-up vessel to select the local minimum in March, April, and May 2020. Specifically, we forecasted that in 2020, the local minimum of the average daily mean wind speed in March was the 30th and the local minimum of average daily mean wind speed in April was the 28th of April **(Research question 3)**. After that, we recommend building a medium-term prediction to verify these local minima of each weather window before fixing the maintenance moments.

Due to time restrictions, we skipped the medium-term prediction and implemented the Non-linear Autoregressive Neural Network for short-term wind speed, wind power, and significant wave height prediction one day before and after these two selected days to verify the long-term prediction results and establish the trade-off between the revenue gained by the higher wind speed and the demurrage rate of the jack-up vessel. The performance of the NAR model for wind speed was much better on the 29th and 30th of March 2020 with the MAPE of less than 15%, however, it appeared that we could not conduct maintenance service on these days due to unsafe wind speeds and wave height condition **(Research question 1)**. Regardings the forecasted weather window in April 2020, the wind speed and wave height were predicted to be more favorable to implement maintenance although the prediction error was a bit larger than the forecasted weather window in March 2020 **(Research question 2).** The wind power prediction was unfortunately not accurate due to the wind energy data collection that could not capture the correct nominal wind energy produced by the Gemini wind farm.

In the third part of the research, we calculate the trade-off between the revenue gained by the wind power when we delay the maintenance service and the demurrage rate of the jack-up vessel. We found that delaying is not cost-effective to offset the demurrage rate of the jack-up vessel **(Research question 4)**. However, we can apply long-term and short-term predictions of the weather data every year to establish the schedule of chartering the same jack-up vessel between different wind farms. Thus, enabling the vessel sharing approach of the jack-up vessel to optimize the vessel use, reduce the chartering cost, documentation time, and contract negotiation **(Research question 5)**. In this way, we could reduce the lead time of jack-up vessel deployment and the downtime cost of the broken turbine. We also suggest investigating the weather window from October 31st to November 1st since a delay at that time may be more profitable due to a lower demurrage rate.

5.2. Limitations and further research

The first limitation is the computation time of the machine learning models for missing values imputation and short-term prediction. For the missing value imputation model, we recommend stronger machine learning methods such as Multiple Imputation through XGboost (Deng and Lumley, 2021). This method is the fastest implementation of gradient boosted trees available in R. It can capture the interactions, variability, and non-linear relations while maintaining high computational

efficiency. For short-term prediction using a Non-linear Autoregressive Neural Network, we suggested running our models on cloud services to take advantage of its computation power and reduce running time compared to our personal computer.

The second limitation of this work is the forecasting horizon of the Non-linear Autoregressive Neural Network model. Since this method can only make a reliable 24-step ahead forecast for hourly data. It restricts the wind farm operators to organize the maintenance and operation activities on-site only one day ahead of time. We suggest applying the Bayesian Neural Network (Mohsin, Ramli, and Imdad, 2021) or Long Short-Term Memory Recurrent Neural Network (Akram and El, 2016) to predict the wind speed and wind energy 72 hours ahead of time. This longer forecasting horizon will be more beneficial to the orchestra of operation and maintenance service of the offshore wind farm.

The third limitation of the Non-linear Autoregressive Neural Network in this thesis is the unstable prediction accuracy and model interpretability. This method performs very well on some days for wind speed, wind power, and wave height with the MAPE values of less than around 25%, however, its errors can reach approximately 2090% on other days. Therefore, we recommend applying parameter optimization methods to improve the model prediction results such as the novel hybrid forecasting system of ensemble empirical mode decomposition, extreme learning machine, and multi-objective grey wolf optimization(Wu *et al.*, 2020) to balance the accuracy and stability. For the interpretability of the model, we need further research on this topic to improve the model or we can, instead, Fourier ARMA as the last resort if the model interpretation is necessary.

In addition, since the thesis was written to forecast the weather window in 2020, the prediction is only valid in the Gemini location in 2020, we recommend the users to update the most current weather data, wave height, and wind energy at your location to make the long-term and short-term forecasting before planning the O&M activities and chartering the jack-up vessel in the future.

Finally, the large errors in the wind energy prediction may mostly be due to the indirect calculation of Gemini wind farm's wind energy output which did not consider the energy lost due to daily and major maintenance during the period under the study instead of the actual energy output produced by the wind farm itself. Therefore, the prediction will be more accurate if we can collect the actual energy output which is confidential.

6. APPENDIX

6.1. Gemini wind farm and weather station location

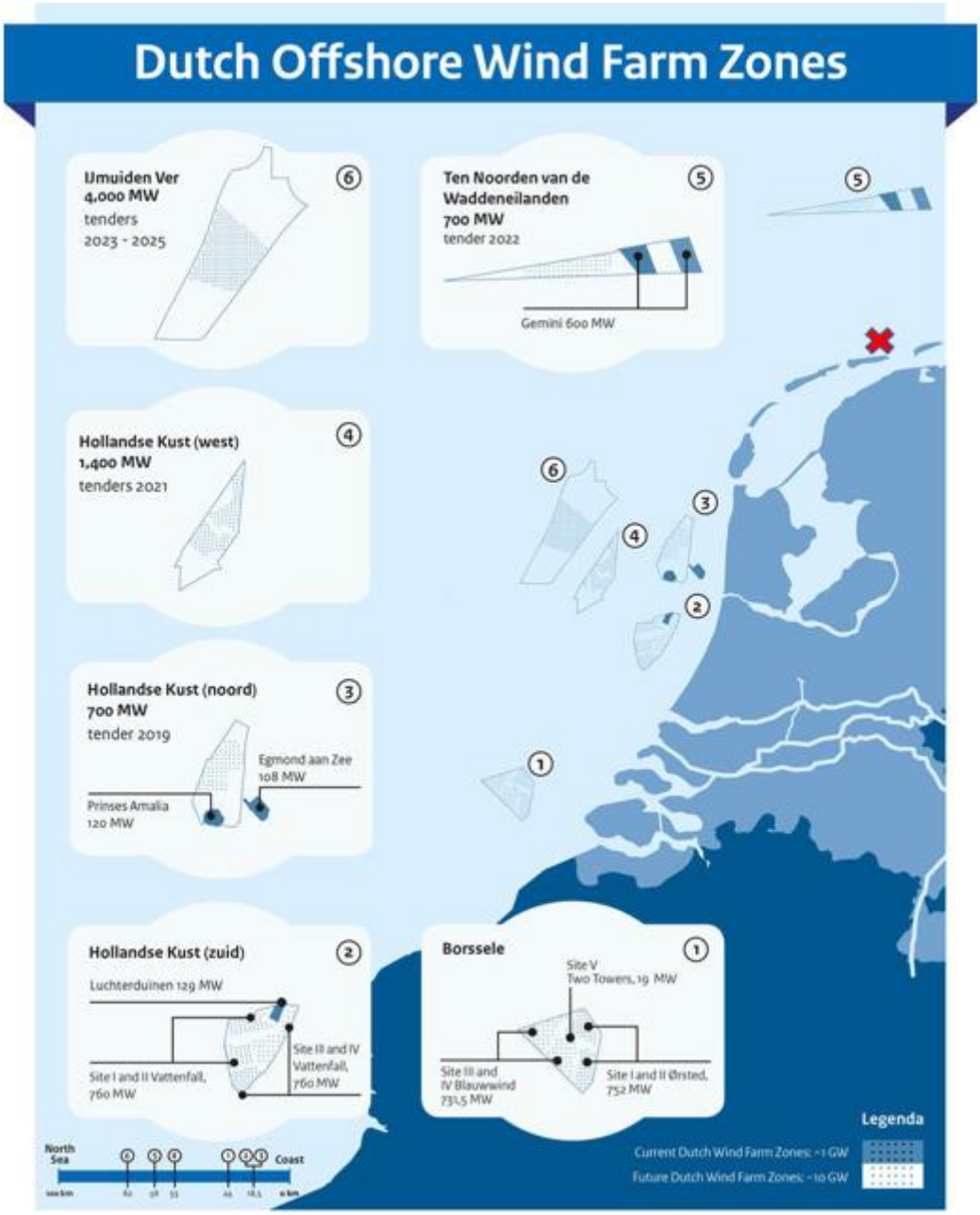


Figure 45: Offshore wind farm map planning (2021) in the Netherlands with the location of weather station marked in the red cross (Source: [The Government of The Netherlands])

6.2. Weather data set

Table 13: Wind speed data set and variable definitions (Source: KNMI)

Variable name	Definition
YYYYMMDD	Date (YYYY=year, MM=month, DD=day)
HH (Time)	Time (time (hour, Amsterdam time)
DD (Wind_direction)	Mean wind direction (in degrees) during the 10 minutes preceding the time of observation
FH (Mean_wind_speed)	Hourly mean wind speed at 10m height
FX (Max_wind_gust)	Maximum wind gust during the hourly division
T (Temperature)	The temperature at 1.50 m at the time of observation
P (Air_pressure)	Air pressure reduced to mean sea level, at the time of observation
U (Relative_humidity)	Relative atmospheric humidity at 1.50 m at the time of observation

Table 14: Hourly climate data set and variable descriptions excluding missing values (Source: KNMI).

Variable name	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
YYYYMMDD	-	-	-	-	-	-
HH (Time)	-	-	-	-	-	-
DD (Wind_direction)	(360=north, 90=east, 180=south, 270=west, 0=calm 990=variable)	-	-	-	-	-
FH (Mean_wind_speed)	m/s	6.99	7	3.18	25	0
FX (Max_wind_gust)	m/s	9.35	9	4.08	36	1
T (Temperature)	degrees Celsius	10.76	10.40	5.65	34.10	-9.10
P (Air_pressure)	kPa	101.46	101.54	1.04	104.65	97.17
U (Relative_humidity)	%	80.01	81	11.27	100	26

Table 15: Summary of missing values of weather indices

Variable name	Missing values	Largest number consecutive missing values
DD (Wind_direction)	3008	2423

FH (Mean_wind_speed)	3030	2424
FX (Max_wind_gust)	3028	2424
T (Temperature)	983	312
P (Air_pressure)	642	167
U (Relative_humidity)	1123	333

Table 16: Average daily climate data from 2014 - 2019

Variable name	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Index	Year – Month – Day	-	-	-	-	-
Wind_direction	(360=north, 90=east, 180=south, 270=west, 0=calm 990=variable)	-	-	-	-	-
Mean_wind_speed	m/s	6.94	6.63	2.61	17.92	1.29
Max_wind_gust	m/s	9.29	8.79	3.37	24.50	2.67
Temperature	degrees Celsius	11.05	10.63	5.49	26.53	-5.79
Air_pressure	kPa	101.50	101.50	1.00	104.30	97.40
Relative_humidity	%	79.89	80.29	8.74	99.83	51.67

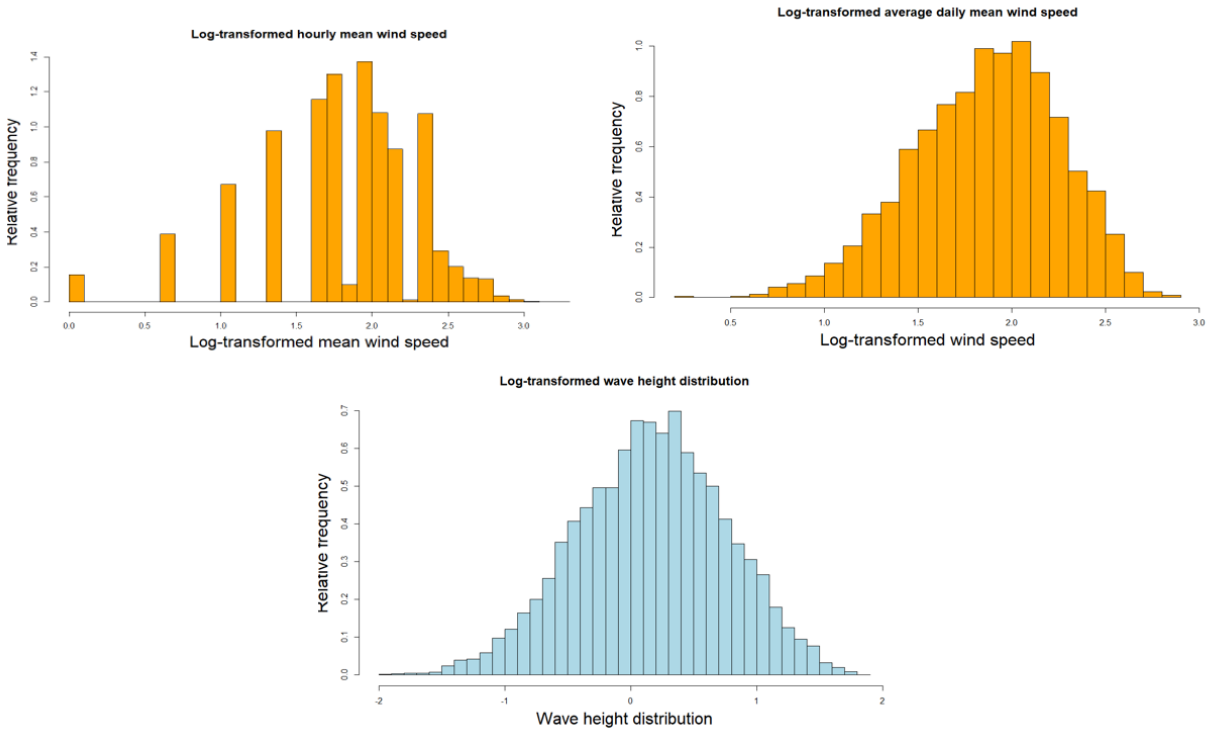


Figure 46: Distribution of log-transformed weather data

6.3. STL Decomposition

Since there are only one seasonal component in the STL decomposition of the literature review, we defined the component strength of time series with more than two seasonal components as follow:

$$Strength\_trend = max(0, 1 - \frac{Var(Remainder_t)}{Var(Trend_t + Remainder_t)})$$

$$Strength\_Season\_year_t = max(0, 1 - \frac{Var(Remainder_t)}{Var(Season\_year_t + Remainder_t)})$$

$$Strength\_Season\_week_t = max(0, 1 - \frac{Var(Remainder_t)}{Var(Season\_week_t + Remainder_t)})$$

$$Strength\_Season\_day_t = max(0, 1 - \frac{Var(Remainder_t)}{Var(Season\_day_t + Remainder_t)})$$

6.4. Residual checks of prediction models

As can be seen in Figure 47, F-ARMA(3,2)x(1,0,0) 's residuals are normally distributed and fluctuate around a zero mean, which proves that the innovation residuals are unbiased. Except for some extreme values at the beginning of 2016, 2017, and 2019, we can say that the variance of its residuals is constant. In addition, the Ljung-Box test result (Table 14) does not have enough evidence to reject the null hypothesis that the residuals are white noise or uncorrelated at a 0.05 significant level. These features ensure the homoskedasticity of the residual and the reliability of the prediction intervals. Thus,

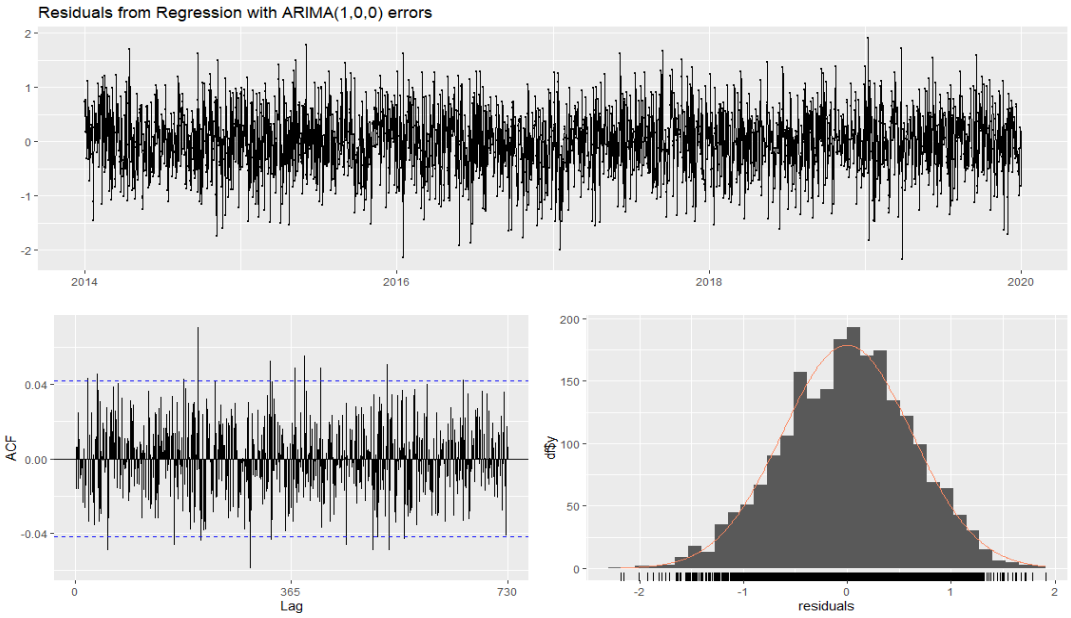


Figure 47: Residual check for Fourier(3,2)x(1,0,0)

For the F-ARMA(3,20)x(1,0,0), its innovation residuals are normally distributed with a zero mean and constant variance based on. Therefore, the prediction intervals and forecast accuracy are reliable. Also, the Ljung-Box test results showed that its residuals were not correlated.

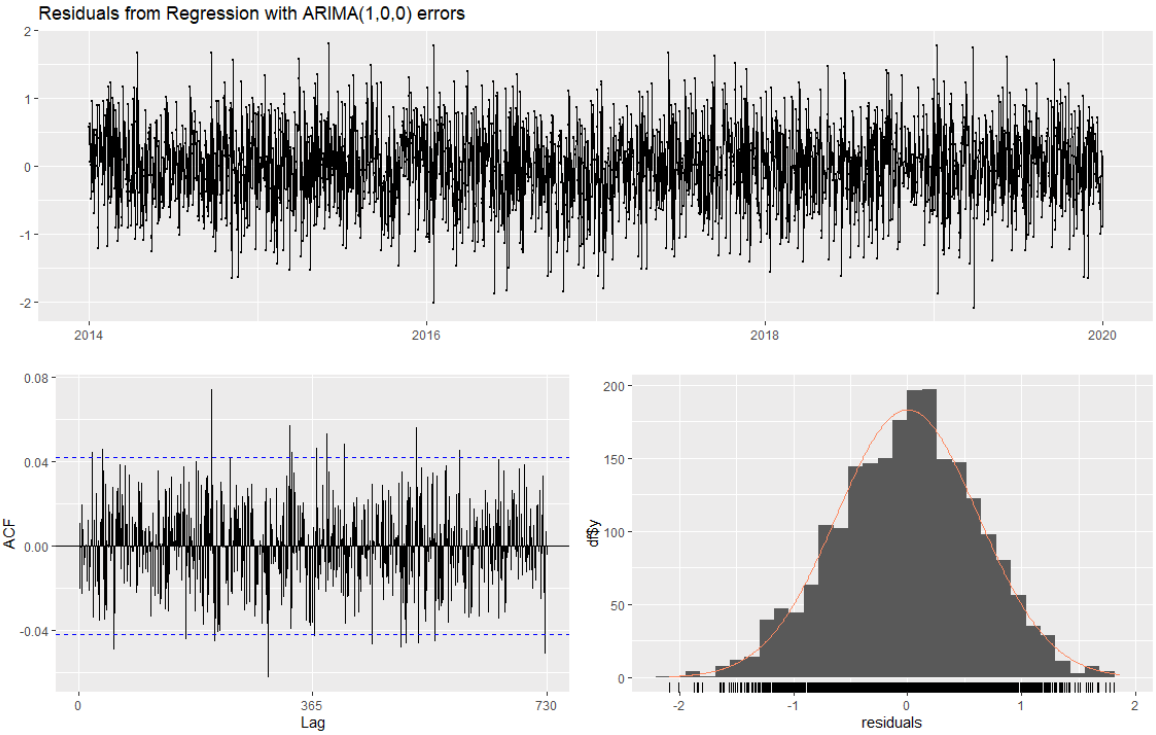


Figure 48: Residual checks of the F-ARMA(3,20)x(1,0,0)

For the F-ARMA(3,2)x(1,0,0)xTemp, its innovation residuals are normally distributed with a zero mean and constant variance based on Figure 49. Therefore, the prediction intervals and forecast accuracy are reliable. Also, the Ljung-Box test results showed that its residuals were not correlated so the model could capture all the information in the data.

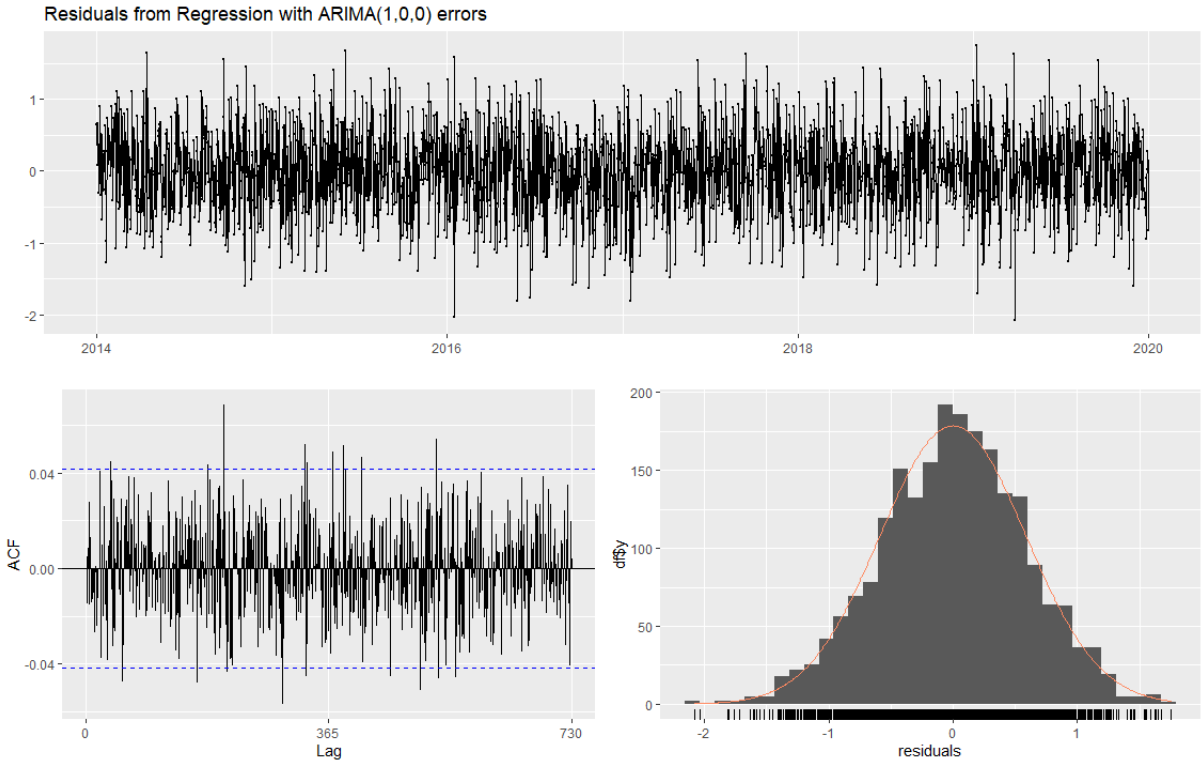


Figure 49: Residual check F-ARMA(3,2)x(1,0,0)xTemp

As can be seen in Figure 49, the innovation residuals of the seasonal naïve forecasting model have zero mean, constant variance, and normal distribution. However, they are correlated as the result of the Ljung-Box test is significant (Table 14).

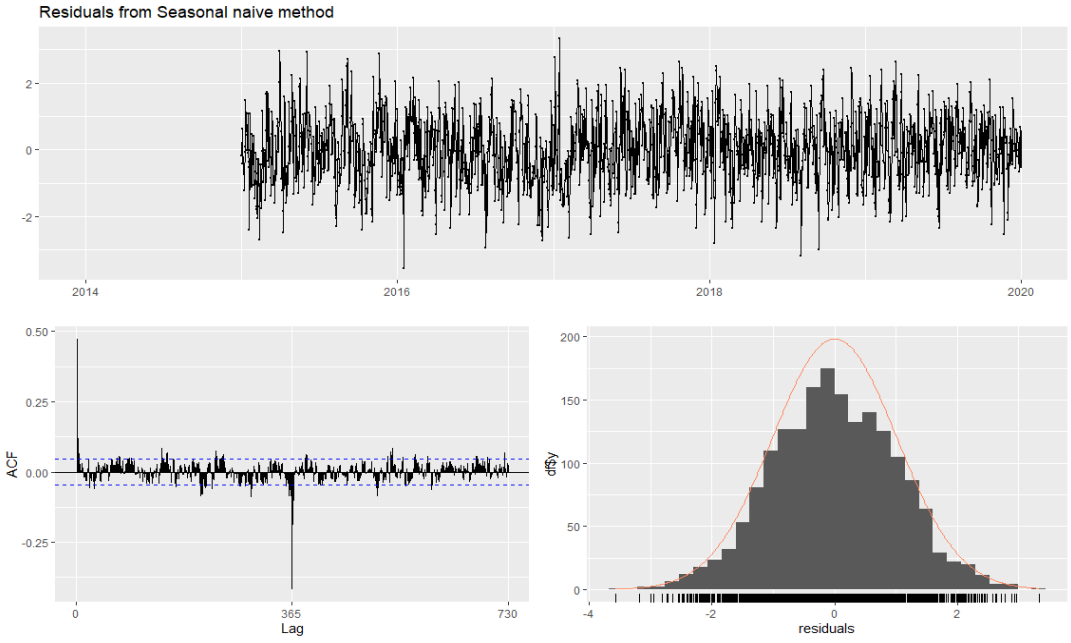


Figure 51: Residual checks of the Seasonal naive model

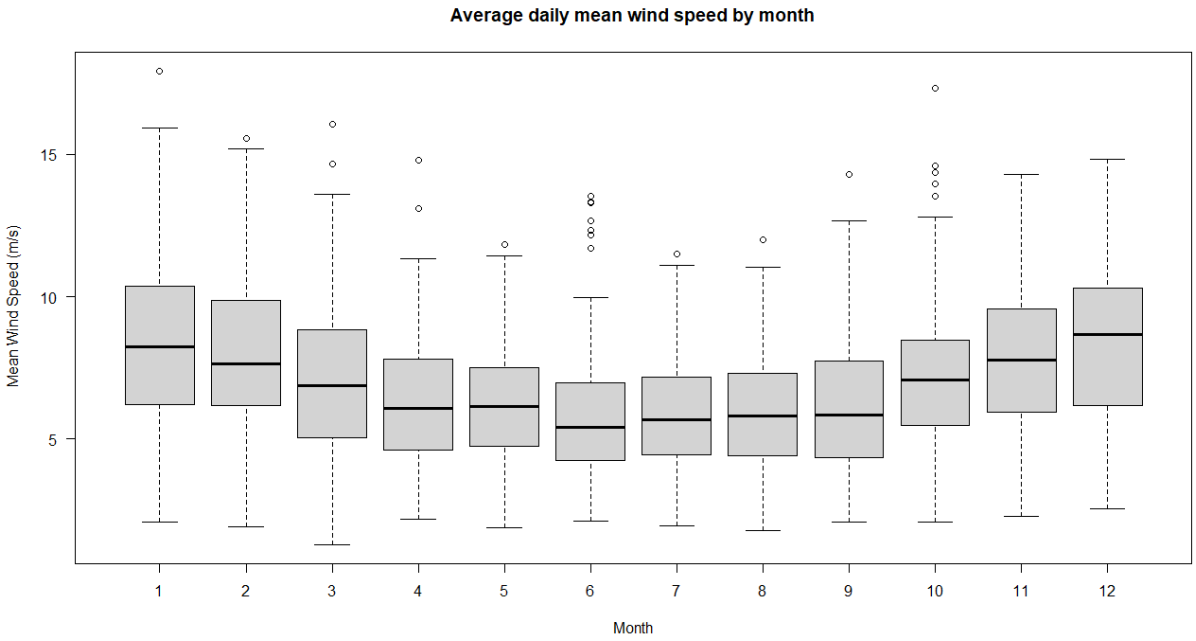


Figure 50: Average daily mean wind speed per month from 2014-2019

6.5. Missing values imputation results

Table 17: OOB errors per variable of each imputation model

		ntree = 2 and mtry = $\sqrt{10}$	ntree = 10 and mtry = $\sqrt{10}$	ntree = 25 and mtry = $\sqrt{10}$	ntree = 50 and mtry = 7
OOB error (MSE)	Wind_direction	4401.25	2695.07	2900.00	**2561.43**
	Mean_wind_speed	0.79	0.58	0.52	**0.52**
	Max_wind_gust	1.33	0.96	0.87	**0.83**
	Temperature	1.49	1.10	0.88	**0.49**
	Air_pressure	0.36	0.22	0.18	**0.13**
	Relative_humidity	44.52	31.66	23.94	**20.43**
Computation time (minutes)		11.23	44	154.87	**352.51**

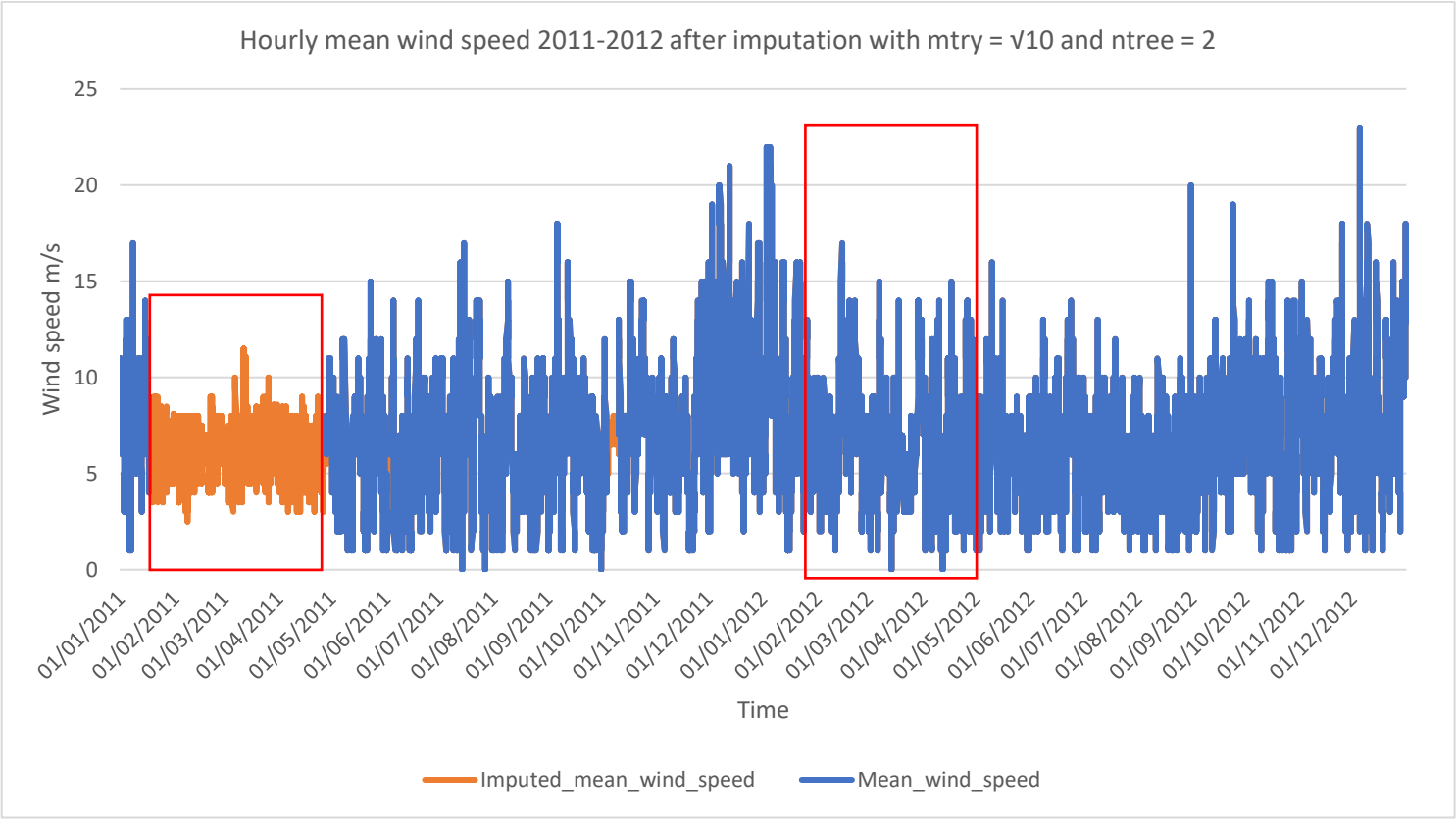


Figure 52: Hourly mean wind speed 2011-2012 after imputed with ntree = 2 and mtry = √10 (red squares indicate the same period of the two given years)

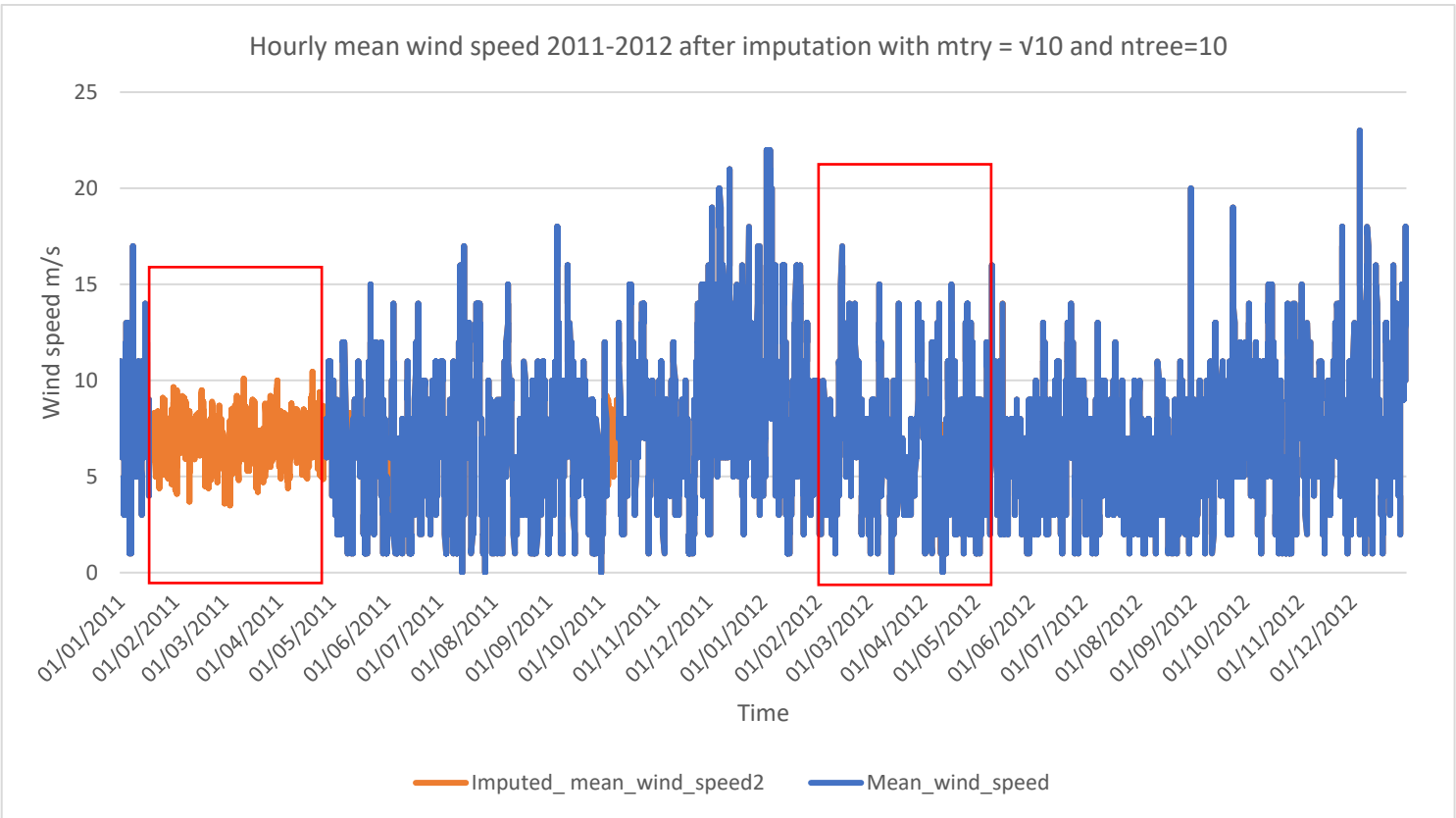


Figure 53: Hourly mean wind speed 2011-2012 after imputed with ntree = 10 and mtry = √10 (red squares indicate the same period of the two given years)

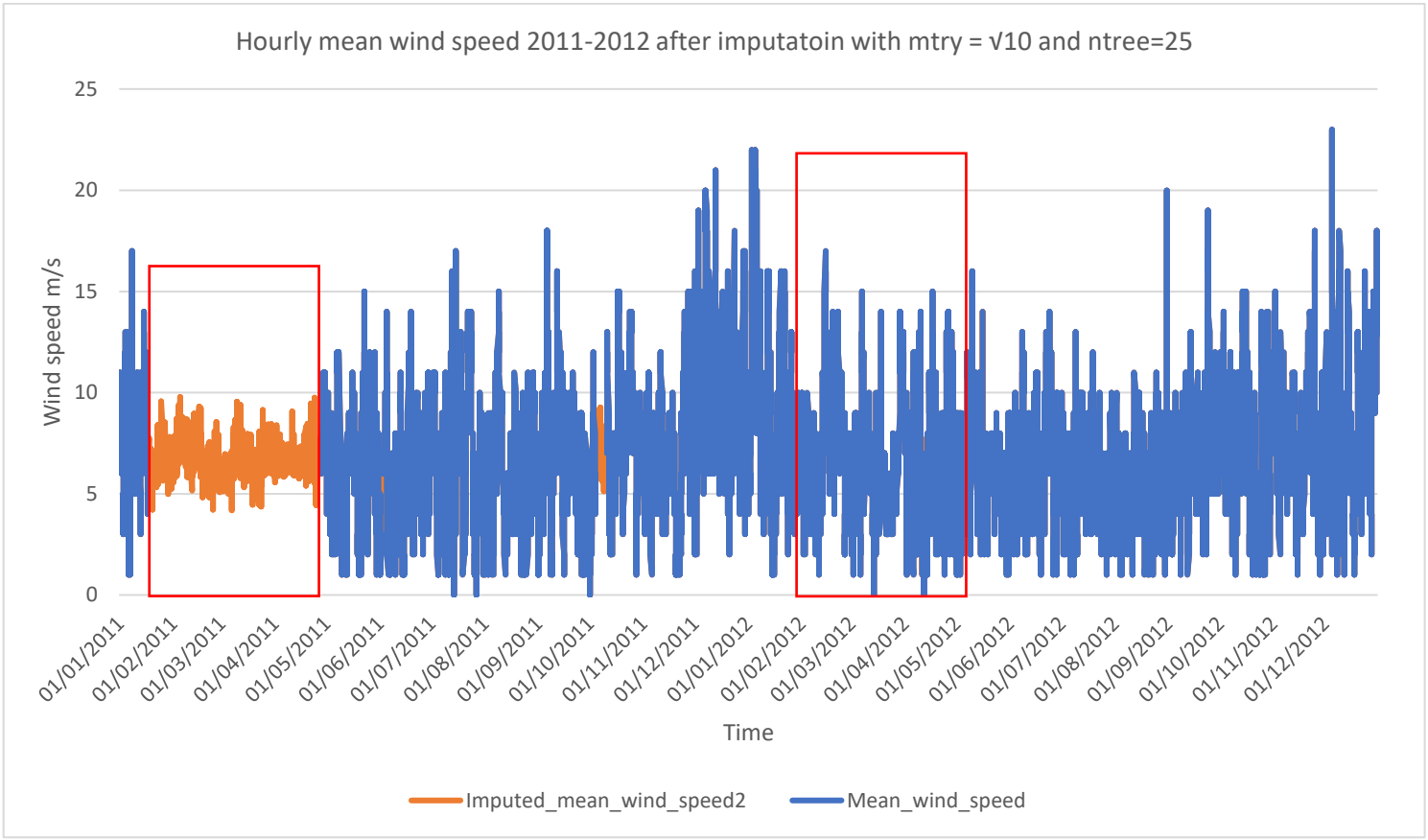


Figure 54: Hourly mean wind speed 2011-2012 after imputed with ntree = 25 and mtry = √10 (red squares indicate the same period of the two given years)

6.6. Accuracy measurements of Non-linear Autoregressive Neural Network models

Table 18: Accuracy measurements of wind speed and power short-term forecasting models

	Model description	RMSE[15]	MAPE[16]	Computation time (minutes)
Wind speed 28/03 – NAR(44,1,22)$_{24}$	20 networks and 1082 weights	4.27	27.9	20
Wind power 28/03 – NARX(44,1,23)$_{24}$	20 networks and 1082 weights	858	24.4	20
Wind power 28/03 – NAR(44,1,22)$_{24}$	20 networks and 1013 weights	931	26.5	15
Wind speed 29/03 – NAR(44,1,22)$_{24}$	20 networks and 1013 weights	1.21	8.25	15
Wind power 29/03 – NAR(44,1,22)$_{24}$	20 networks and 1013 weights	876	15.58[17]	15

[15] Error calculated by forecast package which use indexing from t = 0
[16] Error calculated by forecast package which use indexing from t = 0
[17] Since there was a zero observation in the real observation on the 29/03, MAPE could not be calculated. This might be an error of the data set, however, since the real wind speed of two observations were the same, we assume that the energy generated of the two observations were the same

Wind speed 30/03 – NAR(44,1,40)$_{24}$	20 networks and 1841 weights	1.34	15.5	20
Wind power 30/03 – NAR(26,1,30)$_{24}$	50 networks and 1841 weights	587	31.8	45
Wind speed 31/03 – NAR(44,1,40)$_{24}$	50 networks and 1841 weights	2.88	84.3	45
Wind power 31/03 – NAR(44,1,22)$_{24}$	20 networks and 1013 weights	1522	688	15
Wind speed 27/04 – NAR(44,1,22)$_{24}$	20 networks and 1013 weights	0.854	13.4	15
Wind power 27/04 – NAR(44,1,22)$_{24}$	20 networks and 1013 weights	733	205	15
Wind speed 28/04 – NAR(44,1,40)$_{24}$	40 networks and 1841 weights	1.62	40.9	45
Wind power 28/04 – NAR(44,1,22)$_{24}$	20 networks and 1013 weights	737	187	15
Wind speed 29/04 – NAR(44,1,40)$_{24}$	40 networks and 1841 weights	2.92	155	45
Wind power 29/04 – NAR(40,1,22)$_{24}$	20 networks and 1013 weights	893	2096	15
Wave height 29/03 – NAR(42,1,22)$_{24}$	20 networks and 969 weights	0.36	10	15
Wave height 30/03 – NAR(42,1,22)$_{24}$	20 networks and 969 weights	0.25	13.3	15
Wave height 31/03 – NAR(42,1,22)$_{24}$	20 networks and 969 weights	0.69	54.6	15
Wave height 27/04 – NAR(42,1,22)$_{24}$	20 networks and 969 weights	0.278	38	15
Wave height 28/04 – NAR(43,1,22)$_{24}$	20 networks and 991 weights	0.148	25.3	15
Wave height 29/04 – NAR(42,1,22)$_{24}$	20 networks and 969 weights	0.335	36	15

6.7. Chartering cost of the jack-up vessel

Table 19: Jack-up short-term charter parameters. (Source:Dalgic, Lazakis, Turan, et al., 2015b)

Parameter	Value
Fuel consumption in port	2 ton/day
Fuel consumption in operation	10 ton/day
Daily charter rate	172500 Pound/day (April – September)
	116250 Pound/day (October – March)
Demurrage	+30% daily rate
Mobilization cost and mobilization time	114000 Pound – 10 months
	229000 Pound – 8 months
	473500 Pound – 6 months
	686500 Pound – 4 months
	947500 Pound – 2 months
Low sulfur marine gas oil	550 Pound/ ton
Parameter	Value
Jack-up/down period	3 hours
Hub removal time	8 hours
Port re-supply time	24 hours

6.8. Correlation plot between weather indices

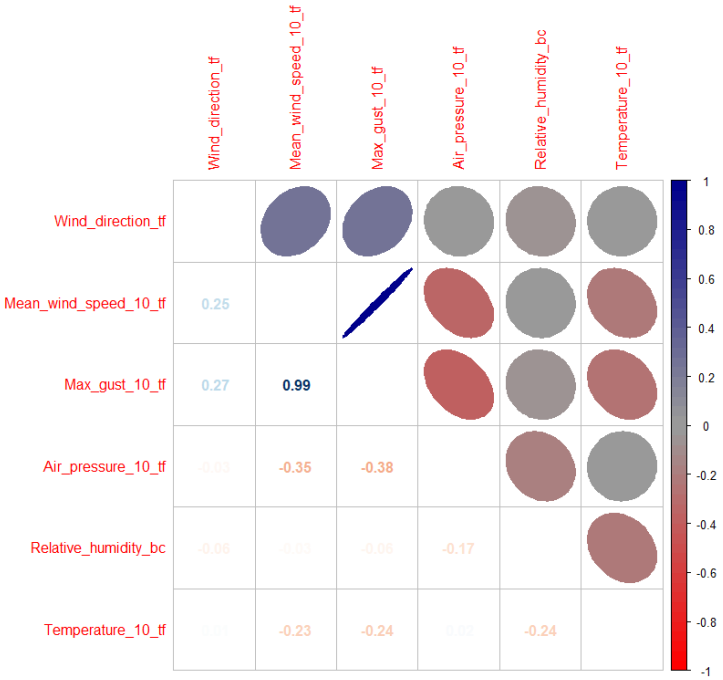


Figure 55: Correlation of weather indices at 10 meter height (Wind_direction_tf, Mean_wind_speed_10_tf, Max_gust_10_tf, Air_pressure_10_tf, Relative_humidity_bc, Temperature_tf are the wind direction, mean wind speed, max wind gust, air pressure, and relative humidity after Box-Cox transformation respectively)

Based on the correlation chart of different weather indices after Box-Cox transformation at 10-meter height in Figure 55, maximum wind gust is highly correlated with mean wind speed, whereas relative humidity is almost not linearly correlated with mean wind speed. Wind direction is harder to predict and could not be assumed to be the same in the next year. The air pressure at 10-meter height is directly correlated with the temperature at 10 m height due to the conversion formula based on (NOAA, NASA, & The United States Airforce, 1976) and it is also not easy to predict.

6.9. Prediction intervals from bootstrapped residuals

In forecasting, we express the uncertainty in our forecasts using a probability distribution. Then, we make an assumption of normal distribution for the possible future values distribution. The point forecast is the mean of this distribution.

As the result, the **prediction interval** gives an interval of where the real future values lie with a certain probability. In this thesis, we use 80% and 95% prediction intervals (Rob J Hyndman, 2021). In addition, since we use Box-Cox transformation for the data, the prediction intervals for the h-step forecast are calculated in the transformed scale, then the endpoints of the interval are transformed back to the original scale. Thus, preserving the probability coverage of the intervals, however, they will not be symmetrical around the point forecast. The prediction interval illustrates the uncertainty of the forecast.

In case, the assumption of normally distributed residuals is violated, we can use the **prediction intervals of bootstrapped residuals** or **bootstrapped prediction intervals** instead (Rob J Hyndman, 2021). This kind of prediction interval only relies on the assumption of uncorrelated residuals with constant variance.

Firstly, on the Box-Cox transformed scale, we simulate different sample paths of the future forecast by sampling the forecasting errors (eb_{t+1}) from the pool of previously observed errors (or residuals) from the past. In this way, we assume the future errors are similar to the past errors $eb_t, \widehat{yb_t}$ is the forecasted value or a fitted value

$$yb_{t+1|t} = \widehat{yb_t} + eb_{t+1} \tag{31}$$

We can continue to simulate the next future observations by repeating this process with e_{t+2} sampling from the previous forecasting errors:

$$yb_{t+2|t+1} = \widehat{yb_{t+1}} + eb_{t+2} \quad (32)$$

Thus, we can simulate the whole set of possible predictions. After that, we calculate the bootstrapped prediction intervals by taking the percentiles of the future sample paths per forecast horizon. This process, therefore, measures the uncertainty of the predicting by only historical data. However, without the assumption of normal residual distribution, the bootstrapped prediction interval is asymmetric, and the point forecast is the mean of the bootstrap samples. The number of sample paths is arbitrary so we will take a bootstrap of 100 samples.

6.10. Gemini energy output proportion calculation in 2020

Table 20: Gemini energy output proportion calculation in 2020

Month	Gemini capacity (MW)	Borssele I & II capacity (MW)	Borssele III & IV capacity (MW)	Gemini energy output proportion in The Netherlands (%)
January - March	600	0	0	62.69
April	600	12 x 8	0	57
May	600	24 x 8	0	52
June	600	36 x 8	0	48
July	600	48 x 8	0	44
August	600	60 x 8	19 x 9.5	37
September	600	72 x 8	38 x 9.5	31
October	600	94 x 8	57 x 9.5	26
Nov to Dec	600	94 x 8	77 x 9.5	24.60

6.11. Van Oord Aeolus jack-up vessel

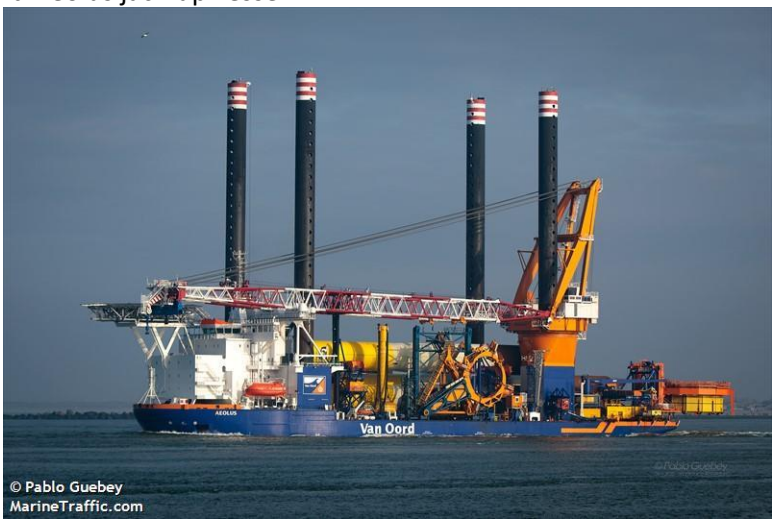


Figure 56: Van Oord Aeolus jack-up vessel. Source: https://www.vanoord.com/

7. REFERENCE

Akram, M. and El, C. (2016) “Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks,” *International Journal of Computer Applications*, 143(11), pp. 7–11. doi:10.5120/ijca2016910497.

Allen, D.J. *et al.* (2017) “A boundary layer scaling technique for estimating near-surface wind energy using numerical weather prediction and wind map data,” *Applied Energy*, 208, pp. 1246–1257. doi:10.1016/J.APENERGY.2017.09.029.

Bai, Y. and Jin, W.-L. (2016) “Wave Loads for Ship Design and Classification,” *Marine Structural Design*, pp. 73–93. doi:10.1016/B978-0-08-099997-5.00005-8.

Bengio, Y. *et al.* (2013) “Generalized Denoising Auto-Encoders as Generative Models.” Available at: http://arxiv.org/abs/1305.6663.

Berbić, J. *et al.* (2017) “Application of neural networks and support vector machine for significant wave height prediction,” *Oceanologia*, 59(3), pp. 331–349. doi:10.1016/J.OCEANO.2017.03.007.

Besnard, F. *et al.* (2009) “An optimization framework for opportunistic maintenance of offshore wind power system,” in *2009 IEEE Bucharest PowerTech*, pp. 1–7. doi:10.1109/PTC.2009.5281868.

Blanchard, T. and Samanta, B. (2020) “Wind speed forecasting using neural networks,” *Wind Engineering*, 44(1), pp. 33–48. doi:10.1177/0309524X19849846.

Byon, E., Ntaimo, L. and Ding, Y. (2010) “Optimal Maintenance Strategies for Wind Turbine Systems Under Stochastic Weather Conditions,” *IEEE Transactions on Reliability*, 59(2), pp. 393–404. doi:10.1109/TR.2010.2046804.

Carroll, J., Mcdonald, A. and Mcmillan, D. (2015) *Failure Rate, Repair Time and Unscheduled O&M Cost Analysis of Offshore Wind Turbines*.

Cleveland, R.B. *et al.* (1990) “STL- A seasonal trend decomposition procedure based on Loess,” *Journal of Official Statistics*, 6(Statistics Sweden), pp. 3–73.

Crol, J.B. (2015) *Upending of a Monopile for an Offshore Wind Turbine Foundation Graduation thesis MSc Offshore and Dredging Engineering*.

Dalgic, Y., Lazakis, I., Dinwoodie, I., *et al.* (2015a) “Advanced logistics planning for offshore wind farm operation and maintenance activities,” *Ocean Engineering*, 101, pp. 211–226. doi:10.1016/J.OCEANENG.2015.04.040.

Dalgic, Y., Lazakis, I., Dinwoodie, I., *et al.* (2015b) “Advanced logistics planning for offshore wind farm operation and maintenance activities,” *Ocean Engineering*, 101, pp. 211–226. doi:10.1016/J.OCEANENG.2015.04.040.

Dalgic, Y., Lazakis, I., Turan, O., *et al.* (2015a) “Investigation of optimum jack-up vessel chartering strategy for offshore wind farm O&M activities,” *Ocean Engineering*, 95, pp. 106–115. doi:10.1016/J.OCEANENG.2014.12.011.

Dalgic, Y., Lazakis, I., Turan, O., *et al.* (2015b) “Investigation of optimum jack-up vessel chartering strategy for offshore wind farm O&M activities,” *Ocean Engineering*, 95, pp. 106–115. doi:10.1016/j.oceaneng.2014.12.011.

Daniel, L.O. *et al.* (2020) “Short-term wind speed forecasting using statistical and machine learning methods,” *Algorithms*, 13(6). doi:10.3390/A13060132.

Deng, Y. and Lumley, T. (2021) “Multiple Imputation Through XGBoost.” Available at: http://arxiv.org/abs/2106.01574.

Dhiman, H.S. and Deb, D. (2020) “A Review of Wind Speed and Wind Power Forecasting Techniques.” Available at: http://arxiv.org/abs/2009.02279.

Duan, Jikai *et al.* (2021) “Short-term wind speed forecasting using recurrent neural networks with error correction,” *Energy*, 217, p. 119397. doi:10.1016/J.ENERGY.2020.119397.

Faulstich, S., Hahn, B. and Tavner, P.J. (2011) “Wind turbine downtime and its importance for offshore deployment,” *Wind Energy*, 14(3), pp. 327–337. doi:10.1002/we.421.

Foley, A.M. *et al.* (2012) “Current methods and advances in forecasting of wind power generation,” *Renewable Energy*, 37(1), pp. 1–8. doi:10.1016/J.RENENE.2011.05.033.

Galanis, G., Papageorgiou, E. and Liakatas, A. (2017) “A hybrid Bayesian Kalman filter and applications to numerical wind speed modeling,” *Journal of Wind Engineering and Industrial Aerodynamics*, 167, pp. 1–22. doi:10.1016/J.JWEIA.2017.04.007.

Gemini Wind Park (2018) *Living on Wind Gemini Wind Park*. Available at: http://www.studioplatz.nl/assets/gemini_boek_binnen_def_compleet.pdf (Accessed: March 23, 2022).

Gneiting, T. and Katzfuss, M. (2014) “Probabilistic forecasting,” *Annual Review of Statistics and Its Application*, 1, pp. 125–151. doi:10.1146/annurev-statistics-062713-085831.

González-Aparicio, I. *et al.* (2017) “Simulating European wind power generation applying statistical downscaling to reanalysis data,” *Applied Energy*, 199, pp. 155–168. doi:10.1016/J.APENERGY.2017.04.066.

Guignard, F. *et al.* (2019) “Investigating the time dynamics of wind speed in complex terrains by using the Fisher–Shannon method,” *Physica A: Statistical Mechanics and its Applications*, 523, pp. 611–621. doi:10.1016/J.PHYSA.2019.02.048.

Heaton, M.J. *et al.* (2019) “A Case Study Competition Among Methods for Analyzing Large Spatial Data,” *Journal of Agricultural, Biological, and Environmental Statistics*, 24(3), pp. 398–425. doi:10.1007/s13253-018-00348-w.

Hinton, G.E. and Osindero, S. (no date) *A Fast Learning Algorithm for Deep Belief Nets Yee-Whye Teh*. Available at: http://direct.mit.edu/neco/article-pdf/18/7/1527/816558/neco.2006.18.7.1527.pdf.

Hu, B. and Yung, C. (2020a) *Offshore Wind Access Report 2020*, *1.1.1.1.1.1.1*. Available at: www.tno.nl.

Hu, B. and Yung, C. (2020b) *OFFSHORE WIND ACCESS REPORT OFFSHORE WIND ACCESS REPORT 2020 Offshore Wind Access Report 2020 Date*, *1.1.1.1.1.1.1*.

Hu, Y.L. and Chen, L. (2018) “A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm,” *Energy Conversion and Management*, 173, pp. 123–142. doi:10.1016/J.ENCONMAN.2018.07.070.

Huang, W. and Dong, S. (2021) “Improved short-term prediction of significant wave height by decomposing deterministic and stochastic components,” *Renewable Energy*, 177, pp. 743–758. doi:10.1016/J.RENENE.2021.06.008.

Hurvich, C.M. and Tsai, C.-L. (1989) *Regression and Time Series Model Selection in Small Samples*. Available at: https://www.jstor.org/stable/2336663.

Hyndman, R.J. and Koehler, A.B. (2006) “Another look at measures of forecast accuracy,” *International Journal of Forecasting*, 22(4), pp. 679–688. doi:10.1016/J.IJFORECAST.2006.03.001.

Jamaludin, A.R. *et al.* (2016) “A comparative study between conventional ARMA and Fourier ARMA in modeling and forecasting wind speed data,” in *AIP Conference Proceedings*. American Institute of Physics Inc. doi:10.1063/1.4954627.

Kadhem, A.A. *et al.* (2017) “Advanced wind speed prediction model based on a combination of Weibull distribution and an artificial neural network,” *Energies*, 10(11). doi:10.3390/en10111744.

Kubik, M.L., Coker, P.J. and Hunt, C. (2011) *Using meteorological wind data to estimate turbine generation output: a sensitivity analysis*.

Kumar, N.K., Savitha, R. and Al Mamun, A. (2018) “Ocean wave height prediction using ensemble of Extreme Learning Machine,” *Neurocomputing*, 277, pp. 12–20. doi:10.1016/J.NEUCOM.2017.03.092.

Kumar, V. *et al.* (2017) “Wind Speed & Power Forecasting using Artificial Neural Network (NARX) for new York Wind Energy Farm,” *Journal for Research* [Preprint]. Available at: www.journal4research.org.

Laib, M. *et al.* (2018) “Multifractal analysis of the time series of daily means of wind speed in complex regions,” *Chaos, Solitons & Fractals*, 109, pp. 118–127. doi:10.1016/J.CHAOS.2018.02.024.

Lawan, S.M. *et al.* (2014) “Different Models of Wind Speed Prediction; A Comprehensive Review,” *International Journal of Scientific & Engineering Research*, 5(1). Available at: http://www.ijser.org.

Lazakis, I. and Khan, S. (2021) “An optimization framework for daily route planning and scheduling of maintenance vessel activities in offshore wind farms,” *Ocean Engineering*, 225, p. 108752. doi:10.1016/J.OCEANENG.2021.108752.

Lecun, Y., Bengio, Y. and Hinton, G. (2015) “Deep learning,” *Nature*. Nature Publishing Group, pp. 436–444. doi:10.1038/nature14539.

López, G. and Arboleya, P. (2022) “Short-term wind speed forecasting over complex terrain using linear regression models and multivariable LSTM and NARX networks in the Andes Mountains, Ecuador,” *Renewable Energy*, 183, pp. 351–368. doi:10.1016/j.renene.2021.10.070.

Ludlow, J. and Enders, W. (2000) “Estimating non-linear ARMA models using Fourier coefficients,” *International Journal of Forecasting*, 16(3), pp. 333–347. doi:10.1016/S0169-2070(00)00048-0.

Miky, Y. *et al.* (2021) “A Recurrent-Cascade-Neural network- nonlinear autoregressive networks with exogenous inputs (NARX) approach for long-term time-series prediction of wave height based on wave characteristics measurements,” *Ocean Engineering*, 240, p. 109958. doi:10.1016/J.OCEANENG.2021.109958.

Mitchell O’Hara-Wild *et al.* (2021) *Package “fable” Title Forecasting Models for Tidy Time Series*.

Mohsin, S., Ramli, S.N. and Imdad, M. (2021) “Medium-Term Wind Speed Prediction using Bayesian Neural Network (BNN),” *International Journal of Systematic Innovation*, 6(5), pp. 11–20. doi:10.6977/IJoSI.202109_6(5).0002.

Neshat, M. *et al.* (2021) “A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the Lillgrund offshore wind farm,” *Energy Conversion and Management*, 236, p. 114002. doi:10.1016/J.ENCONMAN.2021.114002.

Noman, F. *et al.* (2021) “Multistep short-term wind speed prediction using nonlinear auto-regressive neural network with exogenous variable selection,” *Alexandria Engineering Journal*, 60(1), pp. 1221–1229. doi:10.1016/J.AEJ.2020.10.045.

Prahlada, R. and Deka, P.C. (2015) “Forecasting of Time Series Significant Wave Height Using Wavelet Decomposed Neural Network,” *Aquatic Procedia*, 4, pp. 540–547. doi:10.1016/J.AQPRO.2015.02.070.

Ramachandra, V. (2019) *Weather event severity prediction using buoy data and machine learning*.

Röckmann, C., Lagerveld, S. and Stavenuiter, J. (2017) “Operation and Maintenance Costs of Offshore Wind Farms and Potential Multi-use Platforms in the Dutch North Sea,” in Buck, B.H. and Langan, R. (eds) *Aquaculture Perspective of Multi-Use Sites in the Open Ocean: The Untapped Potential for Marine Resources in the Anthropocene*. Cham: Springer International Publishing, pp. 97–113. doi:10.1007/978-3-319-51159-7_4.

Rodolfo Piscopia (2015) “An Expert System for Modelling Wave - Height Time - Series,” *International Journal of Engineering Research and*, V4(09). doi:10.17577/IJERTV4IS090656.

Rodrigues, S. *et al.* (2015) “Trends of offshore wind projects,” *Renewable and Sustainable Energy Reviews*. Elsevier Ltd, pp. 1114–1135. doi:10.1016/j.rser.2015.04.092.

Schouten, T.N. (2019) “Optimal maintenance policies for wind turbines under time-varying costs,” p. 94.

Sirnivas, S. *et al.* (2012) *Assessment of Offshore Wind System Design, Safety, and Operation Standards*. Available at: www.nrel.gov/publications.

Sperstad, I.B. *et al.* (2017) "Testing the robustness of optimal access vessel fleet selection for operation and maintenance of offshore wind farms," *Ocean Engineering*, 145, pp. 334–343. doi:10.1016/J.OCEANENG.2017.09.009.

Tascikaraoglu, A. *et al.* (2016) "Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform," *Applied Energy*, 165, pp. 735–747. doi:10.1016/J.APENERGY.2015.12.082.

The Crown Estate (2014) *Jack-up vessel optimisation: Improving offshore wind performance through better use of jack-up vessels in the operations and maintenance phase*. Available at: www.thecrownestate.co.uk.

Triebe, O., Laptev, N. and Rajagopal, R. (2019) "AR-Net: A simple Auto-Regressive Neural Network for time-series." Available at: http://arxiv.org/abs/1911.12436.

uit het Broek, M.A.J. *et al.* (2019) "Evaluating resource sharing for offshore wind farm maintenance: The case of jack-up vessels," *Renewable and Sustainable Energy Reviews*, 109, pp. 619–632. doi:10.1016/j.rser.2019.03.055.

Uit Het Broek, M.A.J. *et al.* (2020) "Evaluating resource sharing for offshore wind farm maintenance: The case of jack-up vessels." doi:10.1016/j.rser.2019.03.055.

UK offshore Operators Association/ Chamber of Shipping (2002) *Guidelines for the Safe Management and Operation of Offshore Support Vessels*. London. Available at: www.oilandgas.org.uk.

Wang, J. *et al.* (2015) "A self-adaptive hybrid approach for wind speed forecasting," *Renewable Energy*, 78, pp. 374–385. doi:10.1016/J.RENENE.2014.12.074.

Wang, L. and Li, J. (2016) "Estimation of extreme wind speed in SCS and NWP by a non-stationary model," *Theoretical and Applied Mechanics Letters*, 6(3), pp. 131–138. doi:10.1016/J.TAML.2016.04.001.

Winkler, R.L. (1972) "A decision-theoretic approach to interval estimation," *Journal of the American Statistical Association*, 67(337), pp. 187–191. doi:10.1080/01621459.1972.10481224.

Wu, C. *et al.* (2020) "A novel hybrid system based on multi-objective optimization for wind speed forecasting," *Renewable Energy*, 146, pp. 149–165. doi:10.1016/j.renene.2019.04.157.

Yang, S. *et al.* (2021) "A novel hybrid model based on STL decomposition and one-dimensional convolutional neural networks with positional encoding for significant wave height forecast," *Renewable Energy*, 173, pp. 531–543. doi:10.1016/j.renene.2021.04.010.

Young, P.C., Pedregal, D.J. and Tych, W. (1999) *Dynamic Harmonic Regression*. Available at: https://onlinelibrary.wiley.com/doi/epdf/10.1002/%28SICI%291099-131X%28199911%2918%3A6%3C369%3A%3AAID-FOR748%3E3.0.CO%3B2-K (Accessed: January 19, 2022).

Zhang, C. *et al.* (2019) "Opportunistic maintenance strategy for wind turbines considering weather conditions and spare parts inventory management," *Renewable Energy*, 133, pp. 703–711. doi:10.1016/J.RENENE.2018.10.076.

Zubier, K.M. (2020) *Using an Artificial Neural Network for Wave Height Forecasting in the Red Sea*, *Indian Journal of Geo Marine Sciences*.

(GWEC), G. W. E. C., 2021. [Online] Available at: https://gwec.net/global-wind-report-2021/

Aaishah Radziah Jamaludin, F. Y. I. L. K. S. M. N., 2016. *A comparative study between conventional ARMA and Fourier ARMA in modeling and forecasting wind speed data.* s.l., AIP Publishing.

Accenture, 2021. *HOW DIGITAL TECHNOLOGY CAN IMPROVE WIND FARM PERFORMANCE IN THE NETHERLANDS.* [Online]
Available at: https://acn-marketing-blog.accenture.com/wp-content/uploads/2021/03/Accenture-Digital-technologies-wind-farm-Netherlands-POV-v1.pdf

Alberto Pliego Marugán, F. P. G. M. J. M. P. P. D. R.-H., 2018. A survey of artificial neural network in wind energy systems. *Applied Energy,* Volume 228, pp. 1822-1836.

Becker, E., 2016. *Climate.org.* [Online]
Available at: https://www.climate.gov/news-features/blogs/enso/october-2015-el-ni%C3%B1o-update-pumpkin-spice
[Accessed 07 01 2022].

Brownlee, J., 2018. *Machine Learning Mastery.* [Online]
Available at: https://machinelearningmastery.com/how-to-develop-lstm-models-for-multi-step-time-series-forecasting-of-household-power-consumption/
[Accessed 21 September 2021].

Brown, Z. G., 2017. *Wind Speed Vs. Air Pressure.* [Online]
Available at: https://sciencing.com/wind-speed-vs-air-pressure-5950623.html
[Accessed 27 10 2021].

Cristian-Dragos Dumitru & Adrian Grigor, 2017. Daily average wind energy forecasting using Artificial Neural Networks. *Procedia Engineering,* Volume 181, pp. 829-836.

D. M. Hawkins, S. Weisberg, 2017. Combining the box-cox power and generalised log transformations to accommodate nonpositive responses in linear and mixed-effects linear models. *South African Statistical Journal,* 51(2), pp. 317-328.

Daniel J. Stekhoven & Peter Bühlmann, 2011. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics,,* 28(1), p. 112–118.

Deepak A. Kaji, J. R. Z. S. K. K. C. S. D. B. C. K. O., 2019. An attention based deep learning model of clinical events in the intensive care unit.. *PLOS ONE,* 14(2), pp. 1-17.

Energy, N., 2020. *nsenergybusiness.com.* [Online]
Available at: https://www.nsenergybusiness.com/projects/borssele-i-and-ii-offshore-wind-power-project/
[Accessed 21 September 2021].

G. E. P. Box & D. R.Cox , 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological),* 26(2), pp. 211-252.

G. M. Ljung, G. E. P. Box, 1978. On a measure of lack of fit in time series models. *Biometrika,* 65(2), p. 297–303.

Government of The Netherlands, 2020. *Government of The Netherlands.* [Online]
Available at: https://www.government.nl/latest/news/2020/12/04/north-sea-energy-outlook-establishes-framework-conditions-for-future-growth-of-offshore-wind-energy
[Accessed 23 09 2021].

Government of The Netherlands, 2020. *Government of The Netherlands.* [Online]
Available at: https://www.government.nl/topics/renewable-energy/offshore-wind-energy
[Accessed 23 09 2021].

Government of The Netherlands, 2020. *Government of The Netherlands.* [Online]
Available at: https://www.government.nl/topics/renewable-energy/offshore-wind-energy
[Accessed 23 09 2021].

Hawkins, D. a. W. S., 2019. Combining the box-cox power and generalised log transformations to accommodate nonpositive responses in linear and mixed-effects linear models.. *South African Statistical Journal,* 51(2), pp. 317-328..

Hyndman, R. &. A. G., 2021. *Forecasting: principles and practice, 3rd edition.* [Online] Available at: https://otexts.com/fpp3/acf.html [Accessed 11 01 2022].

John Fox and Sanford Weisberg, 2019. *An R Companion to Applied Regression.* 3rd ed. Thousand Oaks CA: Sage.

Leo Breiman and Adele Cutler, 2011. *Random Forests.* [Online] Available at: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#prox [Accessed 22 11 2011].

ModernPowerSystem, 2016. *Service Operation Vessels – one year on.* [Online] Available at: https://www.modernpowersystems.com/features/featureservice-operation-vessel-one-year-on-5694099/

NOAA, NASA, & The United States Airforce, 1976. *U.S. Standard Atmosphere 1976.* [Online] Available at: https://ntrs.nasa.gov/citations/19770009539 [Accessed 1994 03 04].

Rob Hyndman, G. A. ,. M. O.-W. R., 2021. *fpp3: Data for "Forecasting: Principles and Practice" (3rd Edition).* [Online] Available at: https://cran.r-project.org/web/packages/fpp3/fpp3.pdf [Accessed 18 01 2022].

Rob J Hyndman, G. A., 2021. *Forecasting: principles and practice, 3rd.* [Online] Available at: https://otexts.com/fpp3/ [Accessed 19 01 2022].

Spyros Makridakis, E. S. a. V. A., 2018. Statistical and machine learning forecast-ing methods: Concerns and ways forward. *PLOS ONE,* 13(3).

Stekhoven, D. J., 2011. *https://stat.ethz.ch/.* [Online] Available at: https://stat.ethz.ch/education/semesters/ss2012/ams/paper/missForest_1.2.pdf [Accessed 22 11 2021].

Tennekes, H., 1973. The logarithmic wind profile. *Journal of the Atmospheric Sciences,* pp. 234-238.

TENNET, 2021. *Technical publication.* [Online] Available at: https://www.tennet.eu/fileadmin/user_upload/Company/Publications/Technical_Publications/Annual_Market_Update_2020.pdf

transparency.entsoe.eu, 2022. *transparency.entsoe.eu.* [Online] Available at: https://transparency.entsoe.eu/generation/r2/actualGenerationPerProductionType/show# [Accessed 24 02 2022].

Ulstein.com, 2018. *HOW WALK-TO-WORK HAS TRANSFORMED OFFSHORE WIND SERVICES.* [Online] Available at: https://ulstein.com/blog/how-walk-to-work-has-transformed-offshore-wind-services

www.deingenieur.nl, 2017. *www.deingenieur.nl.* [Online] Available at: https://www.deingenieur.nl/artikel/gemini-windpark-in-gebruik-genomen [Accessed 24 02 2022].

Xiaozhe Wang, K. S. &. R. H., 2006. Characteristic-Based Clustering for Time Series Data. *Data Mining and Knowledge Discovery,* Volume 13, pp. 335-364.

Yalcin Dalgic, I. L. I. D. D. M. M. R., 2015. Advanced logistics planning for offshore wind farm operationand maintenance activities. *Ocean Engineering,* Volume 101, pp. 211-226.