UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Physics and Technology

Short-term wind power prediction models in complex terrain based on statistical time series analysis

Andreas Aarhuus Fossem

EOM-3901 Master‘s Thesis in Energy, Climate and Environment
desember 2019

Abstract

One of the largest challenges with the utilization of wind as a renewable resource, is its natural variability and intermittent nature. To achieve a sustainable integration of wind power into the power grid, a precise and reliable prediction method is therefore required.

In this study, several short-term wind power prediction models based on statistical time series analysis were developed and tested, focusing on five wind power parks in northern Norway. All prediction models were applied to each of the five complex terrain sites, Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda wind park. The models apply meteorological forecast data, provided by the Norwegian Meteorological Institute, and the measured hourly total power output, for the time period 1. January 2017 – 31. December 2017, for each wind park. Five Markov chain models have been trained and tested using different sets of input parameters, such as wind speed, wind direction, temperature, surface air pressure and power output. Additionally, a meteorological data-customized power curve function by polynomial regression was developed and tested, using the on-site power output and forecasted wind speed and direction. The performances of all models were measured in terms of the NRMSE, and compared with that of a persistent model, by an improvement parameter. All Markov chain models were found to have lower NRMSE than the persistent model, for all five wind parks. The best performing Markov chain model at each wind park, in terms of improvement with reference to the persistent model, was found to be 6.17%, 4.86%, 9.31%, 9.48% and 12.01%, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively. A linear combination of the meteorological data-customized power curve function model and the persistent model, was found to outperform all Markov chain models at all five sites. A turbine-wise prediction for 15 turbines at Havøygavlen wind park, by the use of Markov chains, was found to attain an improvement parameter value of 8.07%. This suggested a substantial improvement gain by the turbine-wise approach, compared to the 1.98% improvement of using the same Markov chain model for the whole park. Furthermore, the wind regimes and seasonal variations at all sites are investigated by an analysis of the statistical properties of the applied wind data.

Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my supervisor, Yngve Birkelund, for providing me with the means to fulfil this thesis, and for the guidance, encouragement and feedback along the way. Thanks also to Richard Wasell at Finnmark Kraft AS, for providing power output data at Havøygavlen wind park.

I am also very grateful to my family for their support throughout my studies. And a special thanks to my mom and dad, Torunn and Rune, for proofreading my thesis and for always being there for me.

Lastly, a big thanks to my friends and classmates, for making my years in Tromsø the best time of my life. Thank you for all the fun adventures and memorable experiences.

Andreas Aarhuus Fossem
Tromsø, December 2019

Table of contents

Abbreviations

HAWT Horizontal-axis wind turbine

LOWESS Locally Weighted Scatterplot Smoothing

MCM Markov chain model

MEPS MetCoOp Ensemble Prediction System

NRMSE Normalized root mean square error

PCF Power curve function

PM Persistent model

PWPP Potential for wind power production

RIX Ruggedness index

List of Tables

List of Figures

1 Introduction

1.1 Short-term wind power prediction models

The annually rising energy consumption in the world, along with the growing environmental concern and rapidly depleting reserves of fossils fuels, makes reliable and sustainable energy alternatives a necessity for minimizing the effects of global warming and meeting the energy needs in the future. The energy transition away from fossil fuels will be enabled by technological innovation, and large-scale deployment of renewable energy sources has increased significantly since the turn of the century (Bremnes & Giebel, 2017) (Shokrzadeh, Jozani, & Bibeau, 2014). Among the renewable energy technologies, wind power has been one of the fastest growing sources of electricity generation. According to the International Energy Agency (IEA), 15% to 18% of the global electricity production, is expected to be generated by wind power by the year 2050 (IEA, 2013). However, one of the largest challenges with the utilization of wind as a renewable resource, is its natural variability and intermittent nature (Ren, Liu, Wan, Guo, & Yu, 2017). Wind power is produced instantaneously as the wind blows, and due to this fluctuating and uncertain nature, additional challenges emerges for grid operators and energy producers. Therefore, accurate forecasts of the expected wind power output is crucial for an effective integration of wind power into the power grid, planning and decision making, and economic efficiency (Hodge & Milligan, 2011).

As storage of harvested wind energy is not yet achievable on a large scale, the production of electricity has to match the consumption at all times. Therefore, the electricity balance normally requires a mix of energy resources to maintain stability in the electricity system (Yoder, Hering, Navidi, & Larson, 2013). To meet the electricity demands, utilities employ both slow and fast-starting dispatchable generation units along with the non-dispatchable, variable renewable

resource units, such as wind power. The dispatchable power sources can be used on demand at the request of power grid operators, to maintain stability by balancing supply with demand (Yoder et al., 2013). This results in an energy marked where suppliers sell potential future power production at both long and short time scale. The intermittent power generation from the wind power industry leads to uncertainty in the power grid and therefore higher economical risk (Birkelund, Alessandrini, Byrkjedal, & Monache, 2018) (Früh, 2012). Regarding market integration, the forecast lead times typically vary between 30 minutes and 12 hours (Fowler, 2012). Substantial deviations from the estimated power production due to inaccurate forecasting, may result in companies being dependent on operating reserves to maintain stability, or expensive last minute power transactions which would lead to higher operating costs (Yoder et al., 2013).

Short-term prediction of power production is recognized as one of the primary contributors for reliable large-scale wind power integration. Hence, improving the performance of prediction models is extensively identified in energy research as a priority of great importance (Madsen, Pinson, Kariniotakis, Nielsen, & Nielsen, 2005). Accurate short-term predictions of wind power may lower the economic impact of wind power systems integrated into electricity grids, by allowing to schedule dispatchable generation and reduce uncertainty by being able to take day-ahead decisions in the electricity market (Yoder et al., 2013) (Hodge & Milligan, 2011).

Short time prediction of wind power is traditionally divided into two main approaches: A physical approach using numerical weather forecasting models, and a statistical approach where a vast amount of historical and current data is analysed and used as input to a mathematical model. Most operational and commercial models today use hybrid models, combining physical models and statistical time series analysis (Giebel, Brownsword, Kariniotakis, Denhard, & Draxl, 2011).

A physical model attempts to formulate the wind field of a given site, by using physical information regarding the area of interest, such as local topography, air pressure, obstacles and roughness. Whereas a statistical model attempts to identify dependencies between explanatory variables, traditionally wind speed and wind direction forecasts, and the produced power output (Holttinen, Miettinen, & Sillanpää, 2013) (Carpinone, Giorgio, Langella, & Testa, 2015). Generally, these models are based on machine learning algorithms, deriving functional dependencies directly from the observations. Recent comparisons demonstrate that machine learning prediction models are well suited for short-term wind power prediction with forecast horizons up to a few hours (Heinermann & Kramer, 2016).

Many natural processes are considered to be stochastic and memoryless, satisfying the properties of Markov processes (Hocaoglu, Gerek, & Kurban, 2008). When modelling wind data, Markov chains are intuitively appealing due to its ability to calculate the probability of going from one state to another (Brokish & Kirtley, 2009). Each state could for instance represent a given wind speed, wind direction, temperature or wind power. And from any given state, there is some probability distribution function of what the next wind speed/direction/temperature/power will be. The Markovian wind models have proven to be far superior to the simple Monte Carlo approach with no temporal correlation (Brokish & Kirtley, 2009). The simplicity of the Markov chain model, and the advantage of it providing probabilistic forecasts and not only point predictions, allows for beneficial modifications to the algorithm regarding the inclusion of different models and how the probabilistic forecasts are applied. This allows for an optimization of the model output.

1.2 Purpose of the study

Norway possesses a vast potential for wind power production. The main reason for this is that the latitudinal location of Norway often coincides with the polar front, in which the warm air from the south meets the colder air from the north. Strong weather systems occur along this front, as a result of the big difference in air pressure, along with the far-reaching coastline towards the open sea, allowing strong winds to reach the shore unaffected (The Norwegian Water Resources and Energy Directorate, 2019). This study will investigate the use of statistical time series analysis for short-term wind power prediction, focusing on five wind power parks in Northern Norway. The five wind parks, Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, are located in a cold climate region characterized by its unique environment conditions, which stands in considerable contrast to anywhere else in the world (Bilal, 2016). Northern Norway holds the rare concurrent features of being a cold climate region and having complex terrain, with large mountains, fjords and valleys. This emphasizes the significance and value of improving the performance of short-term wind power prediction models, in order to reduce both the technical and financial risk related to the uncertainty of wind power production, for all electricity market participants.

The primary purpose of this study is to develop and evaluate several short-term wind power prediction models with 2-hour prediction horizon, using different methods of time series analysis. By applying the models to five different complex terrain sites, their reliability and overall performance will be comprehensively tested. The study will also provide a turbine-wise wind power prediction at Havøygavlen wind park, solely using the concept of Markov chains. Several models will be developed and tested using various input parameters, such as the on-site power production data and meteorological forecast data obtained from the open and operational weather forecast provided by the Norwegian Meteorology Institute. This includes wind speed, wind direction, temperature and surface pressure.

The performance of the proposed models will be measured in terms of Normalized Root Mean Square Error (NRMSE), and compared to those of the naive Persistent Model (PM) and the WindPRO park modelled power curve function for the specific wind park, using the same input data. Considering the topography and complex terrain of the sites investigated, this study will also provide an analysis of the statistical properties of the applied wind data, by illustrating the wind regimes of the sites through wind roses.

1.3 Structure of the study

This study is divided into 6 chapters and organized as follows. In chapter two the relevant wind theory and statistical theory are presented. Chapter 3 provides descriptions of the different wind parks, the data used in the study, and the prediction models. Results are presented and discussed in chapter 4. And in chapter 5 the study is summarized and concluded, along with suggestions for possible future research.

2 Theory

2.1 Wind

2.1.1 Origin of wind

Wind energy is the kinetic energy of the compensatory movement of air mass due to an imbalance in temperature and pressure on Earth. This imbalance is a result of solar radiation, which is primarily absorbed by the land and the sea, heating up the surrounding air. Because of the disparity of absorption for different materials, the temperature of the adjacent air will differ, causing the convection and pressure changes resulting in wind (Andrews & Jelley, 2013). On a global scale, the shape of the Earth causes the intensity of the incident solar radiation to be higher at the equator than at the poles, due to the angle of attack. The non-uniformity causes warm air to rise at the equator and colder air to flow from the poles. Wind will vary both with time and location (Andrews & Jelley, 2013). Close to the sea or ocean shore there are breeze winds which blow from the sea to the land during the day, and vice versa during the night (Shpilrain, 2009). Relatively steady winds with high velocities exists in mountainous areas, at ridges, and in passes or gorges. However, wind is variable depending on local conditions in general, such as topography, surface character and the vertical wind profile for each location.

2.1.2 Global wind patterns

The locations of high and low wind occurrence are distinctly determined by the effect of Earth's rotation. Without rotation, the latitudinal imbalance of intensity of solar radiation would set up a simple north-south convective flow of air mass (Andrews & Jelley, 2013). However, the rotation of the earth leads to a phenomenon near the surface of the Earth called the Coriolis effect. Due to this effect, the straight north-south movement of air mass from the high pressure areas to the low pressure areas is diverted, giving rise to the wind patterns known as the

trade winds, westerlies and polar easterlies. In the northern hemisphere, wind tends to rotate clockwise, whereas in the southern hemisphere the motion of the wind is anti-clockwise.

As the wind moves away from the equator, its eastward component of velocity increases. But by latitude 30° the air flow meets air moving from the poles, cools down and starts sinking due to an increase of density. At these latitudes the sinking air compresses and gets warmer. As a result, some of the air is forced back to the low pressure belt at the equator, creating a cycle of air circulation between latitudes 0° and 30°. This is known as a Hadley cell, and is illustrated in Figure 2.1. The remaining sinking air moves toward the poles, and creates another low pressure belt in the encounter with cold air from the poles. The resulting winds are called the westerlies, and are located between latitudes 40° and 60°. Once more, a second convection cell is created at mid-latitudes, due to some of the air returning to 30° latitude. As the remaining air continues to approach the poles, it is cooled down by the extremely cold air moving toward the equator, resulting in a third and last convection cell, the Polar cell (Jacobsen, 2014).

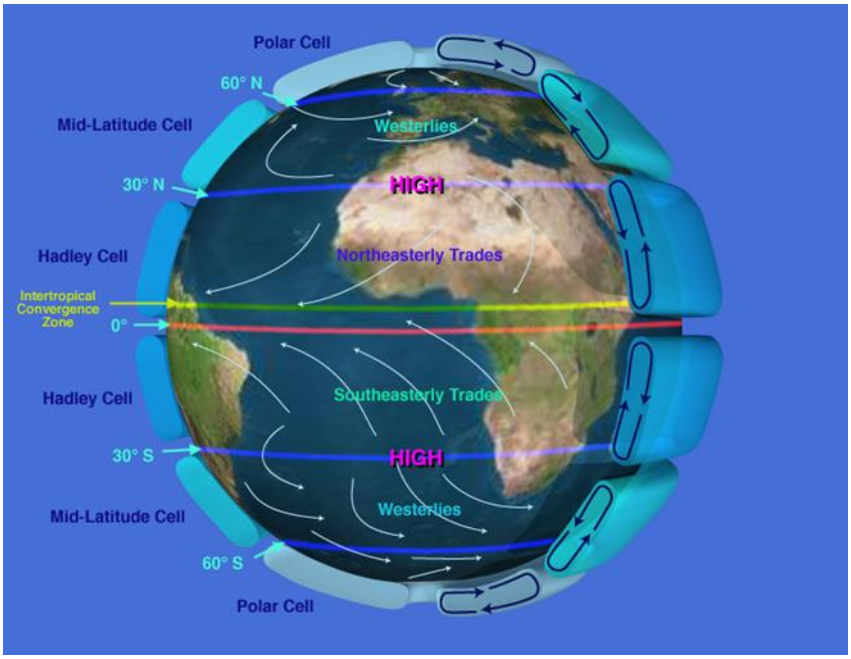

Figure 2.1: Illustration of the global wind circulation (NASA).

2.1.3 Wind Energy

The moving molecules in the wind contains kinetic energy. As stated by Andrews and Jelley (2013), for a wind speed u and air density ρ, the energy per unit volume E is given by

$$E = \frac{1}{2}\rho u^2 \tag{2.1}$$

The volume of air moving through a cross-sectional area A per second, where A is normal to the wind direction, is uA (Andrews & Jelley, 2013). Therefore, the kinetic energy per second of the volume of air flowing through area A is given by $P = EuA$, which we can rewrite as

$$P = \frac{1}{2}A\rho u^3 \tag{2.2}$$

2.2 Wind power production

2.2.1 Wind turbines

Wind turbines act as energy converters, by transforming kinetic energy into mechanical rotational energy, which then is converted into electrical power. The majority of modern wind turbines are horizontal-axis wind turbines (HAWTs), including the ones installed at all five wind parks assessed in this study (Andrews & Jelley, 2013). The turbines consists of a tower with a nacelle mounted on top of it, which encloses the other important components of the wind turbine. The gearbox, generator, brake and control unit are all located inside the nacelle, as well as the bearings for the turbine shaft which is mounted to the rotor blades. An illustration of the different components of a modern HAWT is shown in Figure 2.2.

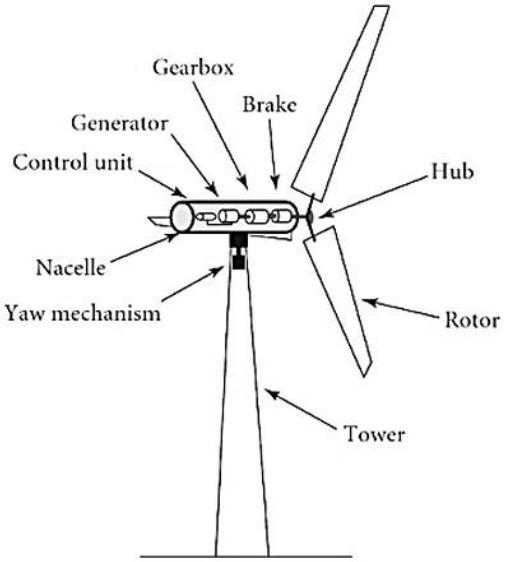

Figure 2.2: Components of a horizontal-axis wind turbine (Andrews & Jelley, 2013).

The rotor is generally oriented so that the plane of rotation is perpendicular to the direction of the wind, called an upwind design, with the nacelle behind (Andrews & Jelley, 2013). The rotor blades of the wind turbine, typically two or three, is attached to the hub. The blades are shaped like aerofoils providing lift perpendicular to the flow direction when the wind moves over them. The lift force generates a driving torque causing rotation. The rotor hub is placed at a height so that the tips of the blades are clear of any turbulent layer of air near the ground, as this can cause reduced energy capture since the wind speed generally increases with height, as well as additional fatigue loading (Andrews & Jelley, 2013).

To optimize its power output, the wind turbine is controlled by the yaw control mechanism, orienting the turbine into the wind. In addition, in the generation of electrical energy from the wind, the power output of most wind turbines is controlled by pitch or active stall by turning the blades of the rotor (Andrews & Jelley, 2013). This allows for an efficient energy harvest at all times, with constantly varying wind conditions. Maximum output power can therefore be generated at wind speeds exceeding the minimum required to operate the turbine, while keeping loads on the turbine within safe limits (Jacobsen, 2014).

The wind turbine can be stopped by applying the shaft brake, whose main function is to provide supplementary breaking to the breaking effect obtained by pitching, but is also used when conducting maintenance or during downtime due to technical or mechanical failures (Jacobsen, 2014).

A wind turbine cannot extract all of the power in the wind. In order to maintain flow, some of the kinetic energy is carried downstream of the turbine, as illustrated by Figure 2.3. Wind turbines work by slowing down the passing wind in order to extract energy (Andrews & Jelley, 2013). This introduces a theoretical maximum efficiency for the extraction of power from the wind, known as *Betz limit*, stating that only 59,3% of the kinetic energy can be extracted (Andrews & Jelley, 2013). Moving upstream, the wind has a velocity of u_0 passing through a cross-sectional area A_0. When it reaches the turbine, the velocity of the wind has decreased to u_1 moving through the swept area of the blades A_1. As the wind flows through the turbine, it is slowed down as some of the kinetic energy is extracted. Moving downstream of the turbine, through an increased cross-sectional area A_2, the velocity of the wind has consequently decreased to u_2. This process is shown in Figure 2.3.

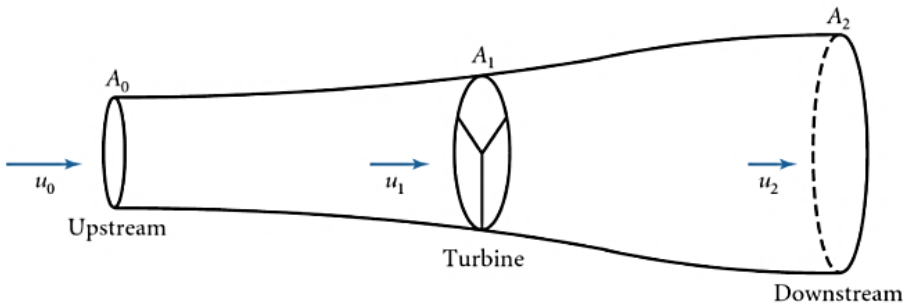

Figure 2.3: Wind flow through a turbine (Andrews & Jelley, 2013).

Maximum power generation is obtained when the velocity of the wind at the turbine u_1 is two thirds of the upstream velocity u_0, and downstream velocity u_2 is one third, or

$$u_1 = \frac{2}{3}u_0 \quad \text{and} \quad u_2 = \frac{1}{3}u_0 \tag{2.3}$$

According to the derivation of Equation (2.1) and by mass continuity, the power extracted under the conditions in Equation (2.3) is given by

$$P = \frac{1}{2}\rho A_1 \left(\frac{16}{27}\right) u_0^3 \tag{2.4}$$

and by Equation (2.2), the power in the wind passing through the upstream cross-sectional area A_1 is given by

$$P_w = \frac{1}{2}\rho A_1 u_0^3 \tag{2.5}$$

Hence, the power coefficient C_P, defined as the fraction of power extracted by the turbine, is given by

$$C_P = P \Big/ \left(\frac{1}{2}\rho A_1 u_0^3\right) \tag{2.6}$$

which under the conditions of Equation (2.3) would be equal to *Betz limit*.

2.2.2 Power curve

The power curve of a wind turbine illustrates the expected output power from a turbine as a function of wind speed, and is a key concept for understanding their efficiency. Figure 2.4 shows a typical power curve, where cut-in, cut-out, and rated wind speed correspond to a Nordex N80/2500 wind turbine, same as 15 of the 16 turbines operating at Havøygavlen wind park. The wind speed at which the turbine first starts to rotate and generate power, 4 m/s, is called the *cut-in wind speed*. When wind speeds reach 15 m/s, the power output reaches the limit of which the electrical generator is capable of. This limit is called *rated wind speed*. At higher wind speeds, the turbine is designed to limit the power to this maximum output power, either by altering the pitch of the blades or altering the generator torque by changing the electrical load. At 25 m/s, a braking system is employed in order to avoid excessive loads, fatigue and damage to the turbine. This is called the *cut-out wind speed*.

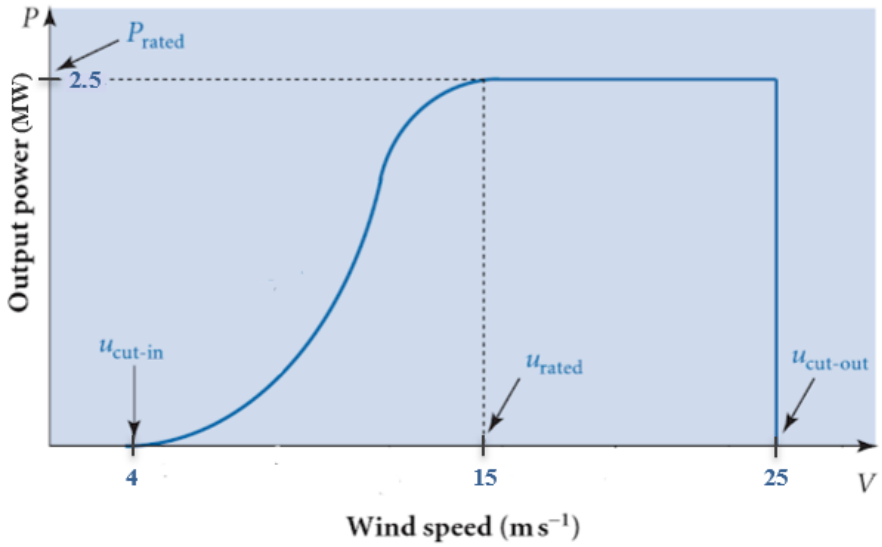

Figure 2.4: Power curve illustrating the output power versus wind speed, with assigned wind speeds corresponding to that of a Nordex N80/2500 turbine.

2.2.3 Effect of topography

The effects of the topographic features on site and surrounding area of a wind park, such as hills, valleys and cliffs, are very complex. Wind directions and velocities vary greatly around such features (Kondo, Tsuchiya, & Sanada, 2002). Therefore, when assessing a possible wind park site and the potential type and positioning of the turbines, it is crucial to estimate the wind flow field around the site. Some wind flow patterns may be favourable for wind power production, whereas other patterns should be avoided since they can create considerable turbulence (Ragheb, 2016). Complex topography may cause turbulence, which can cause varying loads on wind turbines, reducing their lifespan and affecting the power production. Obstacles in close proximity, such as hills, ridges and cliffs affect the wind velocity profile, by decreasing the wind speed and creating turbulence. Figure 2.5 illustrates how the incident wind on an obstacle (black box) is affected, flowing around the obstacle and creating turbulence zones. The turbulence occurs in a lesser extent in front of the obstacle, and largely behind it (Ragheb, 2016).

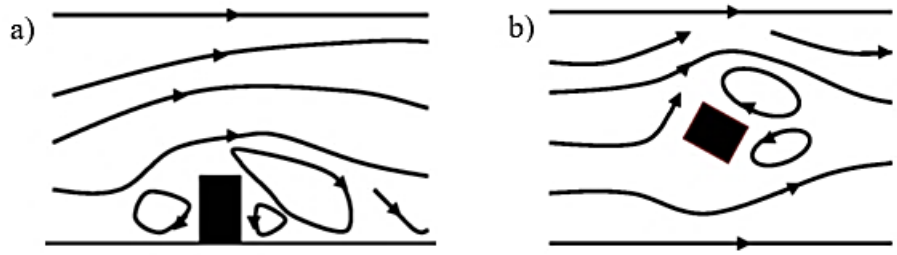

Figure 2.5: Elevation view of the wind flow around an obstacle a), and top view of the wind flow around an obstacle (Ragheb, 2016).

The ruggedness index (RIX) is an objective measure of the complexity in the terrain, by describing the extent of steep slopes in a given radius (The Norwegian Water Resources and Energy Directorate, 2019). The interpretation of RIX values, should be done with caution. A legitimate interpretation is that the model uncertainty in simplified models, that do not explicitly calculate the turbulent atmospheric boundary layer, will be higher in areas with large RIX values.

Mesoscale models such as WRF falls into this category. The level of turbulence, on the other hand, also depends on how the wind rose is distributed in relation to local topography for the given site. Therefore, high RIX value does not categorically mean large occurrences of turbulence (Byrkjedal & Åkervik, 2009).

2.2.4 Wake effects

When a wind turbine extracts energy from the mass of air that flows through it, some of its kinetic energy will be converted to electrical energy. The implications of this process is that the flow is reduced, while the turbulence intensity increases at a region behind the turbine rotor. The region in which these changes occur, is known as the wake of the wind turbine, and the impacts exerted by these changes on the overall performance of the wind park is known as wake effects (Yussuff, 2017). At wind park level, the reduced wind speed and turbulence downstream of a turbine, will negatively affect the energy production and increase wake-induced fatigue of downwind turbines (Manwell, McGowan, & Rogers, 2009).
In wind park scale, typical wake losses lie in the range of 4-15% (Barthelmie, 2007). Both for optimizing the power production, and to accurately predict the power production at a wind park, it is very important to assess the positioning of the turbines, with the subsequent turbine wake effects on wind speeds and turbulence reaching downstream turbines.

2.3 Time series analysis

Time series analysis involve the application of statistical methods for analysing and modelling an ordered sequence of observations. Henrik Madsen (2008) defines a time series as an observed or measured realization of a stochastic process. Analysing the time series data using different statistical methods, can provide useful characteristics and forecasts based on formerly observed data.

2.3.1 Markov chains

A Markov chain is a stochastic model developed by the Russian mathematician Andrey Andreyevich Markov. The theory was originally released in 1907 in his paper *Extension of the Limit Theorems of Probability Theory of a Sum of Variables Connected in a Chain* (Markov, 1907).
A Markov chain is defined by Sheldon M. Ross (2014) as a stochastic process that takes on infinite or countable number of possible values, in which the probability of attaining each value depends only on the previous value. A stochastic process $\{X(t), t \in T\}$ is a collection of random variables, so that for each $t \in T$, $X(t)$ is a random variable. Commonly, t is interpreted as time, and consequently we often refer to $X(t)$ as the state of the process at time t (Ross, 2014).

Let $\{X_n, n = 0, 1, 2, \dots, \}$ be a discrete-time stochastic process indexed by nonnegative integers and that takes on finite or countable number of possible values. If $X_n = i$, then the process is said to be in state i at time n.
We suppose that there is a fixed probability that the process will next be in state j whenever in state i, defined as P_{ij} (Carpinone et al., 2015). That is, for a first-order Markov chain, we suppose that

$$\begin{aligned} &P\{X_{n+1} = j | X_n = i\ , X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} \\ &= \ P\{X_{n+1} = j | X_n = i\} \ = \ P_{ij} \end{aligned} \tag{2.7}$$

for all states $i_0, i_1, \dots, i_{n-1}, i, j$ and all $n \geq 0$. For a Markov chain, Equation (2.7) can be interpreted as the conditional distribution of any future state being independent of the past states, and depends only on the present state (Ross, 2014).

The first-order transition matrix P for k states includes all transition probabilities between said states. Shamshad et al. (2005) defines this transition matrix as

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,k} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k,1} & p_{k,2} & \cdots & p_{k,k} \end{bmatrix} \tag{2.8}$$

Since probabilities are nonnegative and since the process must take a transition into some state, we have the following conditions regarding the transition probabilities above

$$p_{ij} \geq 0, \forall\, i, j \geq 0; \qquad \sum_{j=1}^{k} p_{ij} = 1, \;\; i = 1, 2, \ldots, k \tag{2.9}$$

For a second-order Markov chain, the process depends both on the current state and the immediately preceding state. The same pattern applies for higher order Markov chains. The transition matrix P for a second-order Markov chain, as stated by Shamshad et al. (2005), assumes the following form

$$P = \begin{bmatrix} p_{1,1,1} & p_{1,1,2} & \cdots & p_{1,1,k} \\ p_{1,2,1} & p_{1,2,2} & \cdots & p_{1,2,k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1,k,1} & p_{1,k,2} & \cdots & p_{1,k,k} \\ p_{2,1,1} & p_{2,1,2} & \cdots & p_{2,1,k} \\ p_{2,2,1} & p_{2,2,2} & \cdots & p_{2,2,k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k,k,1} & p_{k,k,2} & \cdots & p_{k,k,k} \end{bmatrix} \tag{2.10}$$

In this case, p_{ijl} represents the transition probability of going to state l, given that the current state is j and the previous state was i.

Similarly as for the first-order Markov chain, we have the following conditions regarding the transition probabilities of the second-order Markov chain

$$\begin{aligned} &p_{ijl} \geq 0, \forall\, i,j,l \geq 0; \\ &\sum_{l=1}^{k} p_{ijl} = 1, \quad i = 1,2,\dots,k \ \text{ and } \ j = 1,2,\dots,k \end{aligned} \tag{2.11}$$

Using maximum likelihood, it is possible to estimate the transition probabilities of first- and second-order Markov chains. As stated by Sheldon M. Ross (2014), for any pair of states i and j, or trio of states i, j and l, we have

$$p_{ij} = \frac{q_{ij}}{\sum_{j=1}^{k} q_{ij}} \tag{2.12}$$

$$p_{ijl} = \frac{q_{ijl}}{\sum_{l=1}^{k} q_{ijl}} \tag{2.13}$$

where the quantities q_{ij} and q_{ijl} are called the instantaneous transition rates. Meaning, when in state i, it is the rate at which the process makes a transition into state j, or to state j and then state l, respectively. The instantaneous transition rates can be obtained by the observed number times that a specific sequence of states occurs in a process.

2.3.2 Polynomial regression

Time series is an outcome of a stochastic process, or an observation of a dynamical phenomenon. However, methods which are generally related to the analysis and modelling of static phenomena, such as regression analysis, can prove beneficial in regard to time series forecasting (Madsen, 2008).

Madsen (2008) defines the classical regression model as a way of describing the static relation between a dependent variable p_i, and the $k + 1$ independent

variables $\beta_0, \beta_1, \dots, \beta_k$. Polynomial regression is extensively featured in literature concerning the estimation of power curves of wind turbines (Shokrzadeh, Jozani, & Bibeau, 2014). The model can be expressed as the standard extension of the linear regression $p_i = \beta_0 + \beta_1 x_i + \epsilon_i$, with the polynomial function

$$p_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \epsilon_i \tag{2.14}$$

As defined by Shokrzadeh et al. (2014), Equation (2.14) can be written as

$$\boldsymbol{P} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.15}$$

where $\boldsymbol{P} = (p_1, p_2, \dots, p_n)^{\mathrm{T}}$, $\boldsymbol{X}$ is a matrix of which the ith row is defined as $\boldsymbol{X}_i = \left(\mathbf{1}, \boldsymbol{x_i}, \boldsymbol{x_i^2}, \dots, \boldsymbol{x_i^k}\right)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^{\mathrm{T}}$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^{\mathrm{T}}$.

We estimate the coefficients in $\boldsymbol{\beta}$ by minimizing the *residual* sum of squares *S*, using the method of *least squares*. With the residuals being defined as the difference between the observed value, and the fitted response value provided by the model (Madsen, 2008). As defined by Shokrzadeh et al. (2014), we have

$$S(\boldsymbol{\beta}) = (\boldsymbol{P} - \boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{P} - \boldsymbol{X}\boldsymbol{\beta}) \tag{2.16}$$

Differentiating Equation (2.16) with respect to $\boldsymbol{\beta}$, we can solve

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{P} - \boldsymbol{X}\boldsymbol{\beta}) = 0 \tag{2.17}$$

and thus we obtain the following estimation for $\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{P} \tag{2.18}$$

2.3.3 Wind roses

A wind rose is a circular histogram used to graphically present wind conditions, specifically wind speed and direction, at a specific site. The bars of a regular histogram are replaced by segments around a circle. Collected wind data is divided into bins and sorted by wind direction, so that the radius of each segment represents the number of observed wind speed measurements for each given directional segment. Thus indicating the wind behaviour at a site. An example of a wind rose is shown in Figure 2.6. The circle is divided into 16 directional segments, with each representing a 22.5° interval. The segments represents the four cardinal directions, with the four respective intercardinal directions, and eight secondary intercardinal directions. Each of these segments are further divided into bins representing 12 different wind speed intervals, indicated by a colour scale.

Figure 2.6 tells us that the most common wind directions in this case are originating from between the secondary intercardinal directions of west-southwest (WSW) and south-southwest (SSW). It is also from this directional region where we primarily find the highest occurrence of high wind speeds.

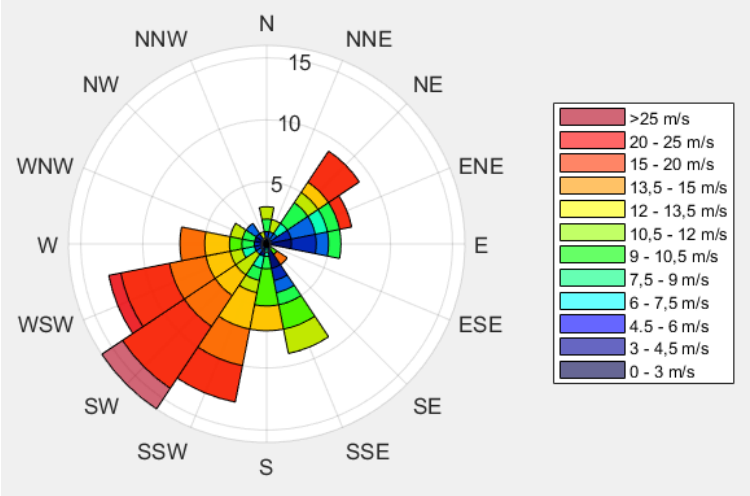

Figure 2.6: Example of a wind rose.

2.4 General statistics

2.4.1 Linear combination

For a vector w in a vector space V, w is said to be to be a *linear combination* of the vectors $v_1, v_2, \dots, v_r$ in V if w can be expressed in the form of

$$w = k_1 v_1 + k_2 v_2 + \cdots + k_r v_r \tag{2.19}$$

where $k_1, k_2, \dots, k_r$ are scalars, called *coefficients* of the linear combination (Anton & Rorres, 2015).

2.4.2 Error measures

To evaluate the performance of short-term wind power prediction models, an error measure is needed. As stated by Madsen et al. (2005), the prediction error is defined as the difference between the measured value and the predicted value

$$e(t+k|t) = P(t+k) - \hat{P}(t+k|t) \tag{2.20}$$

Where P and $\hat{P}$ are the true and predicted power output respectively, and $t+k$ is the lead time.

For the purpose of comparison, it is often convenient to introduce the normalized prediction error ϵ, which is obtained by dividing the prediction error e on the installed capacity P_{inst}

$$\epsilon(t+k|t) = \frac{1}{P_{inst}}\Big(P(t+k) - \hat{P}(t+k|t)\Big) \tag{2.21}$$

Any prediction error can be decomposed into a systematic error μ_e, or *bias*, and a random error ξ_e (Madsen, Pinson, Kariniotakis, Nielsen, & Nielsen, 2005). Thus, we can write

$$e = \mu_e + \xi_e \tag{2.22}$$

where μ_e is a constant and ξ_e is a zero mean random variable.

2.4.2.1 Normalized root mean square error

The root mean square error (RMSE) is a standard statistical metric to measure model performance, and commonly employed in model evaluation studies (Chai & Draxler, 2014) (Birkelund, et al., 2018). The RMSE is a measure of the difference between the values actually observed and the values predicted by the model. Particularly, the square root of the variance of the residuals, indicating the absolute fit of the model to the observed data. The RMSE is useful when comparing different models on the same particular set of data, but not across different data sets, as it is scale-dependent (Hyndman & Koehler, 2006).

To allow for a direct comparison between different wind parks, one can normalize the RMSE to the range of the observed data, meaning the installed capacity of the wind parks. The Root Mean Square Error (RMSE) and the Normalized Root Mean Square Error (NRMSE) can be determined by the following equations

$$\mathrm{RMSE(k)} = \left(\frac{1}{\mathrm{N}}\sum_{\mathrm{t=1}}^{\mathrm{N}} \mathrm{e}^2(\mathrm{t+k|t})\right)^{\frac{1}{2}} \tag{2.23}$$

$$\mathrm{NRMSE(k)} = \left(\frac{1}{\mathrm{N}}\sum_{\mathrm{t=1}}^{\mathrm{N}} \epsilon^2(\mathrm{t+k|t})\right)^{\frac{1}{2}} \tag{2.24}$$

Where N is the number of predictions made in the given time period that is being evaluated. Both the systematic error μ_e and the random error ξ_e contribute to the criterion of these error measures (Madsen, el al.,2005).

2.4.2.2 Comparison of models

To attain a constructive evaluation and quantification of the gain of using a new model compared to a reference model, we introduce the improvement parameter I. For a given lead time, this parameter is defined by Madsen et al. (2005) as

$$I_{ref,EC} = 100 \cdot \frac{EC_{ref}(k) - EC(k)}{EC_{ref}(k)}\% \tag{2.25}$$

Where EC is the considered Evaluation Criterion, which can be for instance the RMSE or the NRMSE of the reference model and the new model.

3 Methods

3.1 Site and time

The short-term power prediction models developed in this study, will be focusing on five wind power parks in Northern Norway. The locations of the five wind parks are shown in Figure 3.1. The figure shows the terrain elevation in Northern Norway in colour from dark blue to red, from 0 to 1500 meters above sea level, respectively, with the ocean in white. The cold climate region in which all five wind parks are located, represents good wind resources in general, but is challenging with its complex terrain (Byrkjedal & Åkervik, 2009) (Birkelund, Alessandrini, Byrkjedal, & Monache, 2018). A description overview for the five wind park sites is presented in Table 3.1, serving as a summarized comparison in terms of installed capacity, location, site ruggedness (RIX) (Byrkjedal & Åkervik, 2009), and potential for wind power production (PWPP).

The ruggedness index (RIX) is an objective measure of the complexity in the terrain, by describing the extent of steep slopes in a given radius (The Norwegian Water Resources and Energy Directorate, 2019). Wind speed is the major factor determining the potential for wind power generation for a given area. Areas with high mean wind and even distribution of wind at different speeds have the best production potential. In addition, the area should have little extreme winds, icing and turbulence. The Norwegian Resources and Energy Directorate has analysed 43 areas in Norway in the process of identifying the most suitable areas for wind power production, inlcuding the areas in which the five wind parks in this study are located (The Norwegian Water Resources and Energy Directorate, 2019). Each analyzed area is given a score of 1 to 10, representing the potential for wind power production. The associated potential for wind power production (PWPP) for all five wind park site areas, are given in Table 3.1.

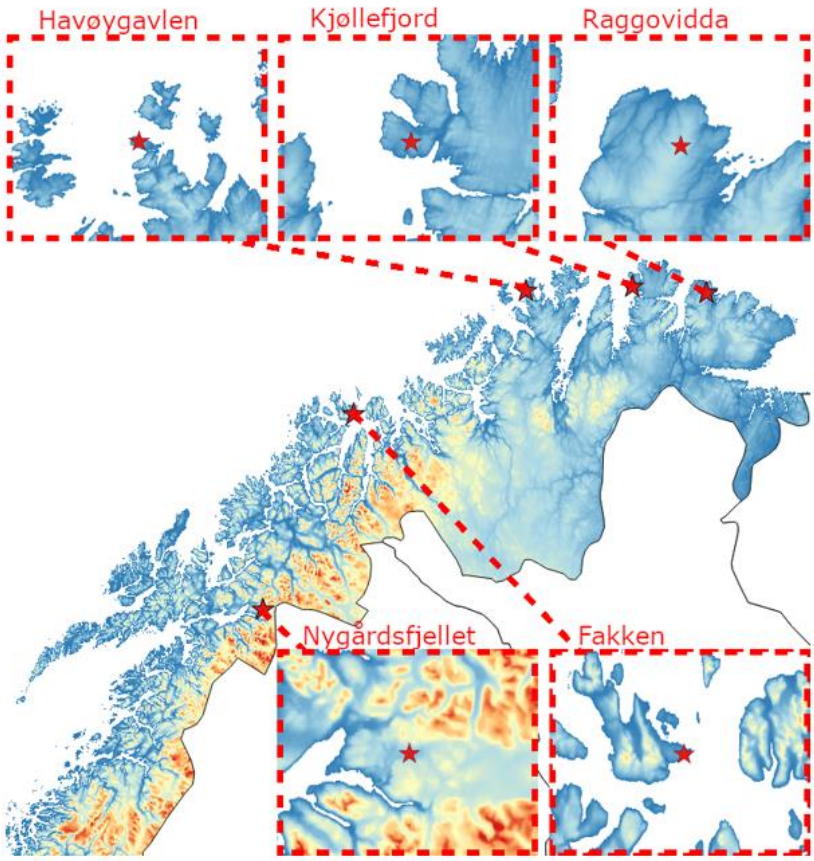

Figure 3.1: Northern Norway, with the ocean shown in white. Terrain elevation shown in colour from dark blue to red, from 0 to 1500 meters, respectively. The wind park locations are marked with stars (Birkelund, Alessandrini, Byrkjedal, & Monache, 2018).

Table 3.1: Description overview of the wind park sites.

Wind park	MW	Location °N / °E	RIX	PWPP
Havøygavlen	40.5	71.012 / 24.589	5-10	10
Kjøllefjord	39.1	70.922 / 27.268	10-20	10
Nygårdsfjellet	32.2	68.504 / 17.879	0-5	3
Fakken	54.0	70.098 / 20.081	5-10	4
Raggovidda	45.0	70.769 / 29.094	0-5	9

Separate descriptions of the sites, their surrounding topography and turbine information for all five wind parks, will be presented in sections 3.1.1 – 3.1.5. Information regarding turbine type, hub height and rotor diameter of the operative wind turbines of each site, are collected from the publicly available database of developed wind parks in Norway (The Norwegian Water Resources and Energy Directorate, 2019).

Three types of data have been provided and used in this study. Measured hourly total power output for each site listed in Table 3.1, provided by the Norwegian Water Resources and Energy Directorate (NVE), active power for each turbine at Havøygavlen wind park, provided by Richard Wasell, operations Manager at Finnmark Kraft AS, and finally meteorological data for all five sites from the open and operational weather forecast provided by the Norwegian Meteorological Institute (MET Norway). All data, for all five wind park sites, are chosen so that all prediction lead times are within the time period of 1. January 2017 – 31. December 2017.

3.1.1 Havøygavlen

Havøygavlen wind park is located on the island Havøya, northwest in Finnmark county. The island is partially exposed to the Norwegian Sea, with Rolvsøya and Ingøya to the west, Hjelmsøya to the north and Måsøya to the east, shown in Figure 3.1. The topography of the site and surrounding area is flat, with little vegetation due to the polar forest boundary going all the way down to sea level. The wind park is located at an altitude of about 200 meters, and there are no large mountains nearby. However, there are steep cliffs down to the ocean enclosing the site, strongly affecting the ruggedness for Havøygavlen (Birkelund, et al., 2018). The cliffs, which can be seen in Figure 3.2, affect the wind velocity profile, and makes the site prone to turbulence (Ragheb, 2016).

The wind park consists of 16 wind turbines, 15 Nordex N80/2500 and 1 Siemens SWT-3.0-101, resulting in a total installed capacity of 40.5 MW. The hub height of the turbines is 80 meters above the ground, with rotor diameters of 80 meters for the Nordex turbines and 101 meters for the Siemens turbine. This study will also provide a turbine-wise wind power prediction at Havøygavlen wind park. A list of all turbines with their respective name, type and location, features in Table 3.2. The 16 wind turbines are placed in two parallel lines, with a distance of approximately 400 meters between each wind turbine. The location of each turbine is given in Figure 3.2.

Figure 3.2: Satellite photo of the site and surrounding area of Havøygavlen wind park, with all 16 wind turbines pointed out (The Norwegian Water Resources and Energy Directorate, 2019).

Table 3.2: List of the wind turbines at Havøygavlen, with their name, type and location.

Name	Turbine type	Latitude	Longitude
HAVWTG001	Nordex N80/2500	71.01095	24.600052
HAVWTG002	Nordex N80/2500	71.00951	24.605536
HAVWTG003	Nordex N80/2500	71.007798	24.6121
HAVWTG004	Siemens SWT-3.0-101	71.003761	24.608
HAVWTG005	Nordex N80/2500	71.010915	24.582402
HAVWTG006	Nordex N80/2500	71.009153	24.588599
HAVWTG007	Nordex N80/2500	71.007691	24.594766
HAVWTG008	Nordex N80/2500	71.006407	24.600891
HAVWTG009	Nordex N80/2500	71.01271	24.592858
HAVWTG010	Nordex N80/2500	71.015691	24.580131
HAVWTG011	Nordex N80/2500	71.01723	24.572046
HAVWTG012	Nordex N80/2500	71.018187	24.56587
HAVWTG013	Nordex N80/2500	71.014554	24.58694
HAVWTG014	Nordex N80/2500	71.012686	24.573184
HAVWTG015	Nordex N80/2500	71.014022	24.566608
HAVWTG016	Nordex N80/2500	71. 01509	24.559993

3.1.2 Kjøllefjord

Kjøllefjord wind park is located on Dyfjord peninsula in northern Finnmark, at a low flat mountain area ranging from 260 meters to 300 meters above sea level (Birkelund, et al., 2018) (Østreng, 2014). A map showing its location and surroundings is given by Figure 3.3. To the west there is a large north-south going fjord, Laksefjorden, running along the whole peninsula and emptying out into the Barents Sea in the north. Two smaller east-west going fjords, Kjøllefjorden and Eidsfjorden, are located in the north and south respectively. In the northwest of the peninsula, there are several mountains close to the coastline. Whereas in the south and southwest, the terrain is gradually lower towards the ocean.

The wind park consists of 17 Siemens SWT-2.3-82VS wind turbines with rated power 2.3 MW, yielding a total installed capacity of 39.1 MW. The hub height of the turbines is 70 meters above the ground, and the rotor diameters are 80 meters. The 17 wind turbines are placed in two parallel lines, perpendicular to the southwestern intercardinal direction, shown in Figure 3.3.

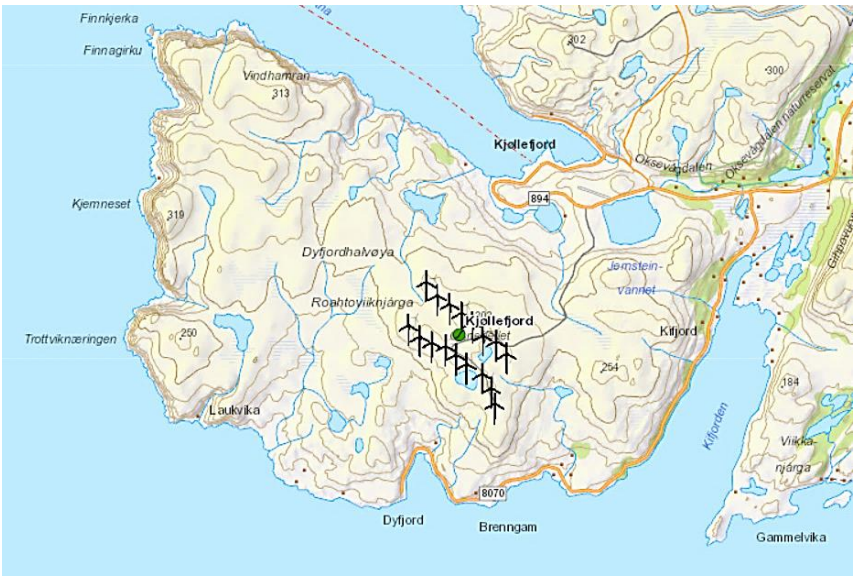

Figure 3.3: Map showing the location and surrounding area of Kjøllefjord wind park (The Norwegian Water Resources and Energy Directorate, 2019).

3.1.3 Nygårdsfjellet

Nygårdsfjellet wind park is located in a complex terrain northeast of Narvik in Nordland county, near the Norwegian-Swedish border. The wind park is placed in a west-east going valley at an altitude of about 400 meters above sea level, with large mountains in the north and south, shown in Figure 3.1 and Figure 3.4. Lake Jernvatnet partially surrounds the wind park in the south, and the turbines are also situated around a smaller lake, given in Figure 3.4. Across the border in the east, a mountainous terrain with lower gradient goes towards Torneträsk, a 332 square km lake serving as a natural wind channel for the wind park (Bilal, 2016). Due to its topography and elevation, the wind park is prone to icing conditions (Jin, 2017) (The Norwegian Water Resources and Energy Directorate, 2019).

The wind park consists of 14 Siemens SWT-2.3-93 wind turbines with rated power 2.3 MW, yielding a total installed capacity of 32.2 MW. The hub height of the turbines is 80 meters above the ground, and the rotor diameters are 93 meters. The 14 wind turbines are placed in three north-south going non-parallel lines, shown in Figure 3.4.

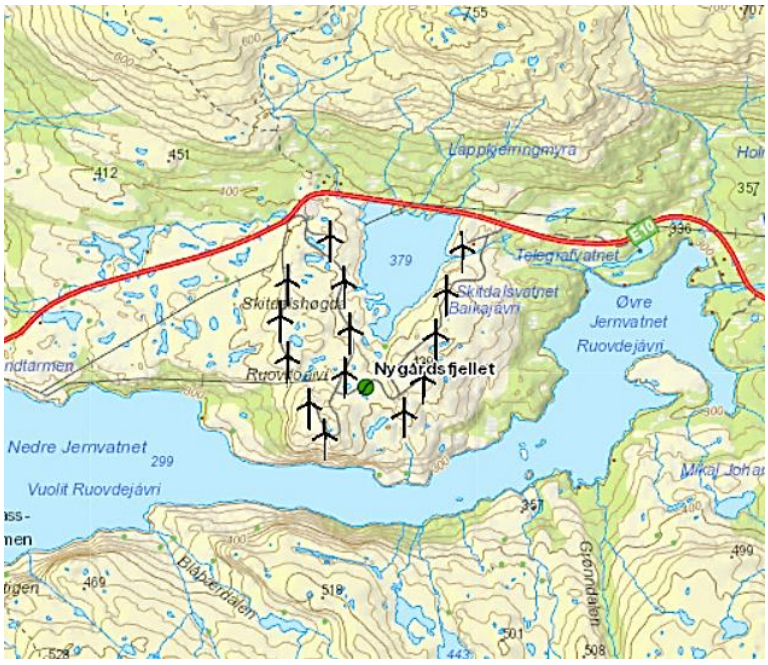

Figure 3.4: Map showing the location and surrounding area of Nygårdsfjellet wind park (The Norwegian Water Resources and Energy Directorate, 2019).

3.1.4 Fakken

Fakken wind park is located on the small island Vannøya in Troms county. The wind park is sited on a small hill at the southwestern edge of the island, at an altitude of 40 to 200 meters above sea level (Birkelund, et al., 2018). A north-south going mountain range is located to the west, shown in Figure 3.5. The open sea to the north, several other mountainous islands in close proximity and two large north-south going fjords in the south, leads to a complex wind regime (Jacobsen, 2014).

The wind park consists of 18 Vestas V90-3.0 wind turbines with rated power 3.0 MW, yielding a total installed capacity of 54 MW. The hub height of the turbines is 80 meters above the ground, and the rotor diameters are 90 meters. The 18 wind turbines are placed in two roughly parallel lines, perpendicular to the southeastern intercardinal direction, shown in Figure 3.5.

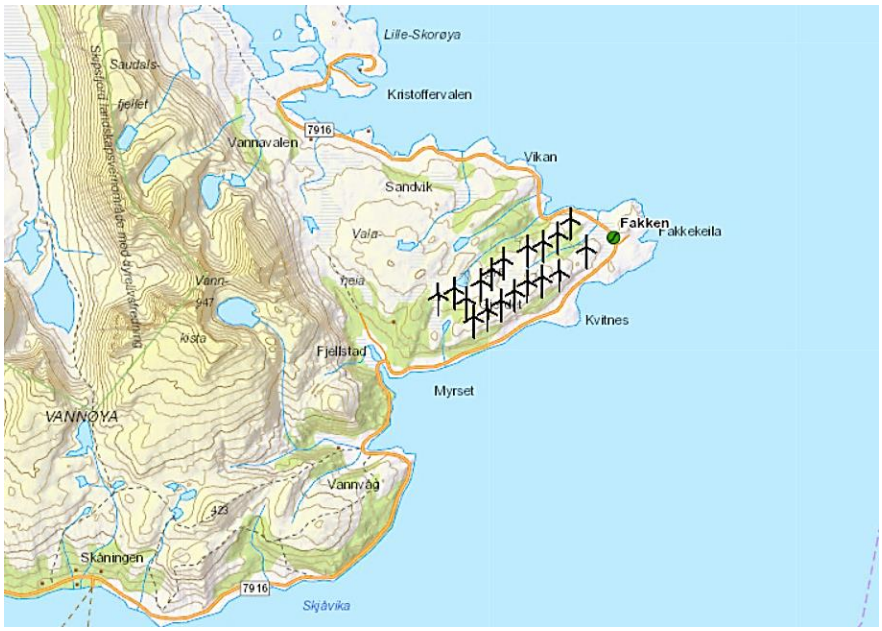

Figure 3.5: Map showing the location and surrounding area of Fakken wind park (The Norwegian Water Resources and Energy Directorate, 2019).

3.1.5 Raggovidda

Raggovidda wind park is located on the Rakkocearro plateau, about 10 km south of Berlevåg in Finnmark county. A map showing its location and surroundings is given by Figure 3.6. The wind park is situated on a large inland mountain area, dominated by blockfield and without vegetation, between 380 and 440 meters above sea level. The area has very good wind resources, and a wind regime characterized by the open sea in the north, the mountains close by in the north and south, and by north-south going fjords in the east and the west.

The wind park consists of 15 Siemens SWT-3.0-101 wind turbines with rated power 3.0 MW, yielding a total installed capacity of 45 MW. The hub height of the turbines is 80 meters above the ground, and the rotor diameters are 101 meters. The 15 wind turbines are placed in four parallel lines, perpendicular to the southwestern intercardinal direction, shown in Figure 3.6.

Figure 3.6: Map showing the location and surrounding area of Raggovidda wind park (The Norwegian Water Resources and Energy Directorate, 2019).

3.2 Data

3.2.1 MetCoOp Ensemble Prediction System (MEPS)

The hourly-point meteorological forecast data used in this paper are obtained from the state-of-the-art numerical weather prediction system MEPS (MetCoOp Ensemble Prediction System). This is a convection-permitting atmosphere ensemble model which is run in operational routine in cooperation between The Norwegian Meteorological Institute (MET Norway), The Swedish Meteorological and Hydrological Institute (SMHI) and Finland Meteorological Institute (FMI) (Køltzow, 2017). MEPS has a horizontal resolution of 2.5 km, using a horizontal grid of 739x949 points centred at 63.5°N and 15°E, and has 65 vertical levels and 10 ensemble members (Birkelund, et al., 2018). MEPS is run four times daily at 00, 06, 12, 18UTC, with member 0, 1 and 2 running up to 66 hours lead time, and the rest up to 54 hours.

In this study, the control run member (member 0) of the model is used. This ensemble member has unperturbed initial and lateral boundary conditions, producing four forecasts per day, each with a forecast horizon of 66 hours. The model data includes predictions of zonal and meridional wind components, u and v, temperature two meters above the ground, and air pressure. Each data point value represents the predicted average over the last hour, interpolated from a point located near the centre of the given wind park. For the forecasted zonal and meridional wind components, the vertical levels chosen to attain the data are done so to correlate with the average hub height of the given wind farm, described in sections 3.1.1 - 3.1.5. The model temperature estimation is based at two meters above ground, while the air pressure is estimated at the surface. From each initialisation, at the four daily term times, it takes 1 hour and 15 minutes to receive and process all available observations, and approximately 50 minutes to run the MEPS system on a high performance computer (Køltzow, 2017).

The MEPS data set had some missing data for the time period being evaluated. In these cases, all model variables are assigned the numeric data type value NaN, representing an undefined value.

The zonal and meridional wind components are converted into two new parameters, wind speed (U) and wind direction (D). The wind speed is found using the Pythagorean theorem, whereas the meteorological wind direction, the direction from which the wind originates from, is found using the four-quadrant inverse tangent of the zonal and meridional components, converting the result from radians to degrees, and adding 180 degrees. We have

$$U = \sqrt{u^2 + v^2} \tag{3.1}$$

$$D = \frac{180}{\pi} \left(\text{atan2}(u, v)\right) + 180 \tag{3.2}$$

where u is the zonal velocity, the component of the horizontal wind towards the east, and v is the meridional velocity, the component of the horizontal wind towards the north.

3.2.2 On-site production data

In addition to the MEPS model data sets, the measured hourly total power output provided by The Norwegian Water Resources and Energy Directorate (NVE), is used for both developing the prediction models and for evaluating their performance. These are aggregated production values, recorded by Statnett at the point of entry to the power grid. The power production time series are limited to the installed capacity of each wind park, listed in Table 3.1, and adapted to correlate and fit with the meteorological forecast data.

The power production data, as well as the meteorological forecast data used in this study, have been collected and processed by Yngve Birkelund of UiT the Arctic University of Norway.

3.2.3 Turbine-wise production measurements at Havøygavlen

In this study, there will also be developed a turbine-wise prediction model, by the use of Markov chains. For this purpose, the on-site measured power output of each of the 16 wind turbines at Havøygavlen wind park, has been provided by Richard Wasell, operations manager at Finnmark Kraft AS. The power data consists of 16 separate power production time series, with 10-minute average values, named according to Table 3.2. The Siemens SWT101 3.0 MW wind turbine, or HAVWTG004, has been excluded in the prediction due to incomplete and faulty power data. The total rated power of the remaining 15 Nordex N80 2.5 MW wind turbines is 37.5 MW.

All of the 10-minute average power data are loaded from csv-format into Matlab, and converted to hourly arithmetic average data. This is carried out in the manner so that each data point is derived from the six 10-minute average data points leading up to it.

Meaning, the measured hourly average power timestamped 06.00, will be the arithmetic average of the 10-minute average values, timestamped 05.00 through 05.50. The resulting power production time series are limited to the rated power of the turbine, and adapted to correlate and fit with the meteorological forecast data.

3.3 Prediction models

3.3.1 Notations

When describing and carrying out the calculations of the subsequent power prediction models, the following notations are used.

P: Measured power provided by NVE and Finnmark Kraft AS [MW]

$\hat{P}$: Forecasted power [MW]

$\widehat{U}$: Forecasted wind speed using MEPS [m/s]

$\widehat{D}$: Forecasted wind direction using MEPS [degrees]

$\hat{T}$: Forecasted temperature at 2m using MEPS [Kelvin]

$\hat{p}$: Forecasted surface air pressure using MEPS [kPa]

t_0: Current time [hours]

t_m: Time at m hours from current time [hours]

3.3.2 Persistent model

One of the most commonly used reference models regarding performance comparison of a forecasting model, is the simplistic and naive persistence method (Giebel, Brownsword, Kariniotakis, Denhard, & Draxl, 2011). The model implies that future values of the time series will be the same as at the current time. It is given by (Madsen, et al., 2005) as

$$\hat{P}(t_{h+m} \mid t_h) = P(t_h) \tag{3.3}$$

for $m = 1, 2, 3, \ldots$ and where $\hat{P}(t_{h+m} \mid t_h)$ is the forecasted output power for future time t_{h+m}, done at time t_h. And $P(t_h)$ is the measured power over the last hour prior to the time of forecasting t_h. Due to the slow scale of changes in the atmosphere, the performance of this model is rather good at short-time prediction horizons (e.g. 4-6 hours) (Madsen, et al., 2005). For that reason, any prediction model should first be measured by the extent it can improve on predictions made by the persistent model.

3.3.3 Power curve function

The second type of reference forecasting model included in this study, will be a simple "black box model", using direction-specific power curve functions obtained with a wind park model from WindPRO software, developed by EMD International A/S. A WindPRO project has been developed by Kjeller Vindteknikk for each wind park listed in Table 3.1, to calculate the site-specific power curves for the different wind parks (Weir, 2014). These are based on synthetic time series of air density-dependent wind speed and wind direction, developed by the mesoscale numerical weather forecasting model WRF (NCAR). The WindPRO Park Power Verification Model being applied in these projects, considers the turbine positions, the associated power curves of the given turbines, height, wake effects between the turbines, and local topography for each site. Thus producing a set of representative power curves, assuming a uniform wind direction distribution, for 12 sectors at 30° intervals. Accordingly, the power output of the wind park is given by a directional-specific power curve function $pcf[\cdot]$ from the park specifications, which is discretized into 30 power outputs corresponding to 30 wind speed intervals. The functions cover a wind range of 0 to 30 m/s, at 1 m/s intervals (Weir, 2014). An illustration of the power curve function for Raggovidda wind park, is shown in Figure 3.7.

The predicted power output of the wind park, using the site-specific power curve function, is given by

$$\hat{P} = pcf[\hat{U}, \hat{D}] \tag{3.4}$$

Using the same input meteorological data as all the other models developed and tested in this study, allows for a fair comparison as an idealized power output for the whole verification period at all sites evaluated.

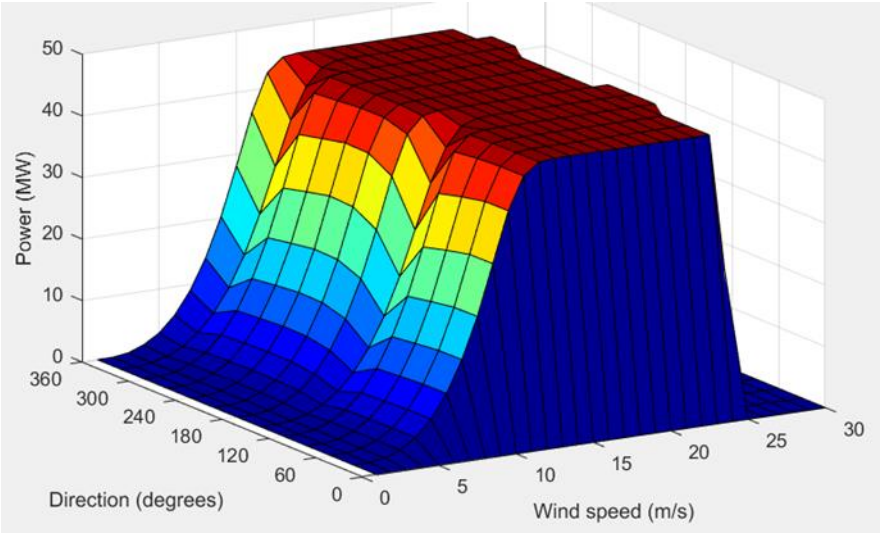

Figure 3.7: Directional-specific power curve for Raggovidda wind park.

3.3.4 Training and verification data

To determine the performance of the prediction models, the meteorological forecast data and power production data is split into a training set and a verification set, with no overlap. This allows the models to be trained using the training set, and their prediction performances to be measured based on new and independent data. In this study, the prediction models will be validated for a whole year of verification data, by the use of leave-one-out cross validation. The leave-one-out cross validation method is appealing in regard to the prediction models developed in this study, as the size of the training set is maximized. However, excessive training may lead to an over-parameterization of the models, overfitting the input training data, and leaving them ineffective in their actual applications. Choosing an optimal training period depends on the geographic location and its seasonal cycle of wind generation patterns and meteorological parameters. Higher wind speeds in the winter, as cold fronts move in from the polar regions, will for instance require a sufficiently large training period, so that the adjacent training periods of the verification period are able to capture the same

seasonal characteristics. The accuracy of a model might therefore suffer by choosing too short a training period, leaving out important information. The ideal training period would pick up the important drivers and patterns for different times of the year (Milligan, Schwartz, & Wan, 2003).

In this study, the data for 2017 are sectioned into 12 separate months of forecast data and power production data. Each month is then used to calculate the prediction performance, with the remaining other 11 months serving as training data for the models. Thus producing a predicted power production time series for the whole year, forming the basis of the error measures. When the models are tested on new independent data, through all seasonal cycles of the year, we quantify the predictive ability of the prediction models and attain a measure of the quality of the prediction models in practice (Madsen, et al., 2005) (Cheng, Garrick, & Fernando, 2017).

The available forecast data and power production data for the time period of 1. January 2017 – 31. December 2017, consisting of 1452 time series each of 66 lead times, has been split into 12 equally sized periods of 121 time series. Hence, the error measures will be based on each of the 12 verification periods, with the remaining 1331 time series serving as training data.

3.3.5 Markov chain model

All input data are loaded into Matlab in 67x1452 sized arrays, one for each parameter, for each of the five wind parks listed in Table 3.1. Thus producing five equally sized arrays for wind speed, wind direction, temperature, surface pressure, and output power, for the respective sites. The indices of all missing measurement values in the meteorological forecast data, assigned NaN, are located and removed from all arrays by discarding the indexed time series for all data sets. Consequently leaving all arrays consisting of the same amount of data and synchronized, so that all meteorological forecast data and power data have the same relative time.

For the meteorological input data parameters, wind speed U [m/s], wind direction D [degrees], temperature T [Kelvin] and surface pressure *p* [kPa], the number of state intervals and spacing between these states are chosen according to Table 3.3. The reasoning behind these choices was to capture the distribution of the data for each parameter in an optimal manner. Hocaoglu et al. (2008) found that by increasing the dimension of state space from 0.5 m/s to 1 m/s, more accurate modelling results could be obtained. The wind speed states in this study are chosen so that the first state includes all wind speeds between 0 m/s and the average cut in speed of the turbines for all five wind parks. Further, the size of the state intervals are 1.5 m/s, until surpassing the average rated wind speed, in which the size of the intervals are 5 m/s. Finally, the last state includes all wind speeds exceeding the average cut-off wind speeds of the turbines for all five wind parks.

The input data containing the forecasted wind directions are assigned 12 equally spaced states, with intervals of 30°. Whereas the states for the temperature and surface pressure data, are chosen to be 12 equally sized intervals of 4 K and 1 kPa respectively, covering all values between the forecasted minimum and maximum of the respective arrays.

Table 3.3: Spacing of state intervals for all meteorological input parameters.

State	Wind speed boundaries (m/s)	Wind direction boundaries (Degrees)	Temperature boundaries (Kelvin)	Surface air pressure boundaries (kPa)
1	0 – 3	0° - 30°	< 254	< 93
2	3 – 4.5	30° - 60°	254 – 258	93 – 94
3	4.5 – 6	60° - 90°	258 – 262	94 – 95
4	6 – 7.5	90° - 120°	262 – 266	95 – 96
5	7.5 – 9	120° - 150°	266 – 270	96 – 97
6	9 – 10.5	150° - 180°	270 – 274	97 – 98
7	10.5 – 12	180° - 210°	274 – 278	98 – 99
8	12 – 13.5	210° - 240°	278 – 282	99 – 100
9	13.5 – 15	240° - 270°	282 – 286	100 – 101
10	15 – 20	270° - 300°	286 – 290	101 – 102
11	20 – 25	300° - 330°	290 – 294	102 – 103
12	> 25	330° - 360°	294 – 298	103 – 104

For the output power states of all five wind parks, the number of state intervals and spacing between these states, are chosen according to Table 3.4. The number of states are different for all wind parks, except for Havøygavlen and Kjøllefjord, which have the exact same state intervals. Considering the power output of a wind park very often equals zero or nominal wind farm power, these values represents the state boundaries, with equally spaced states of 2 MW between the respective boundaries. For the turbine-wise prediction at Havøygavlen wind park, the Nordex N80 turbine are chosen to have state intervals of size 0.1 MW, enclosed by zero production and the rated power 2.5 MW.

Table 3.4: Spacing of output power state intervals for all wind parks, and the Nordex N80 2.5MW wind turbine.

State	Havøygavlen (MW)	Nordex N80 (MW)	Kjøllefjord (MW)	Nygårdsfjellet (MW)	Fakken (MW)	Raggovidda (MW)
1	= 0	= 0	= 0	= 0	= 0	= 0
2	0 – 2 (>0)	0 – 0.1 (>0)	0 – 2 (>0)	0 – 2 (>0)	0 – 2 (>0)	0 – 2 (>0)
3	2 – 4	0.1 – 0.2	2 – 4	2 – 4	2 – 4	2 – 4
4	4 – 6	0.2 – 0.3	4 – 6	4 – 6	4 – 6	4 – 6
5	6 – 8	0.3 – 0.4	6 – 8	6 – 8	6 – 8	6 – 8
6	8 – 10	0.4 – 0.5	8 – 10	8 – 10	8 – 10	8 – 10
7	10 – 12	0.5 – 0.6	10 – 12	10 – 12	10 – 12	10 – 12
8	12 – 14	0.6 – 0.7	12 – 14	12 – 14	12 – 14	12 – 14
9	14 – 16	0.7 – 0.8	14 – 16	14 – 16	14 – 16	14 – 16
10	16 – 18	0.8 – 0.9	16 – 18	16 – 18	16 – 18	16 – 18
11	18 – 20	0.9 – 1.0	18 – 20	18 – 20	18 – 20	18 – 20
12	20 – 22	1.0 – 1.1	20 – 22	20 – 22	20 – 22	20 – 22
13	22 – 24	1.1 – 1.2	22 – 24	22 – 24	22 – 24	22 – 24
14	24 – 26	1.2 – 1.3	24 – 26	24 – 26	24 – 26	24 – 26
15	26 – 28	1.3 – 1.4	26 – 28	26 – 28	26 – 28	26 – 28
16	28 – 30	1.4 – 1.5	28 – 30	28 – 30	28 – 30	28 – 30
17	30 – 32	1.5 – 1.6	30 – 32	30 – 32	30 – 32	30 – 32
18	32 – 34	1.6 – 1.7	32 – 34	32 – 32,2	32 – 34	32 – 34
19	34 – 36	1.7 – 1.8	34 – 36		34 – 36	34 – 36
20	36 – 38	1.8 – 1.9	36 – 38		36 – 38	36 – 38
21	38 – 40	1.9 – 2.0	38 – 40		38 – 40	38 – 40
22		2.0 – 2.1			40 – 42	40 – 42
23		2.1 – 2.2			42 – 44	42 – 44
24		2.2 – 2.3			44 – 46	44 – 45
25		2.3 – 2.4			46 – 48	
26		2.4 – 2.5			48 – 50	
27					50 – 52	
28					52 – 54	

Each value in all parameter arrays containing the forecasted meteorological data and power production data, are converted to their corresponding state value. Meaning for instance a wind direction value of 45° would be denoted as state 2, whereas a wind direction value of 175° would be denoted as state 6 out of the total 12 states. This is done for the sake of practicality when implementing the

algorithm, allowing a more effective data processing and calculation, as well as making it easier to handle multiple input parameters model and creating transition matrices (Jacobsen, 2014).

Using different sets of parameters as input, several Markov chain models can be established. For instance, this could include the forecasted wind speed in two hours $\widehat{U}_{t_2}$ combined with the 2 hour point forecasted wind direction $\widehat{D}_{t_2}$, or any other combination of parameters from the available data sets. The output state parameter is chosen to be the predicted power for all models.

With each value representing the state according to their corresponding parameter state interval presented in Table 3.3 and Table 3.4, the algorithm for the different Markov chain models loops through the 12 equally sized sections of the respective input parameters for the model, as described in Section 3.3.4. Using the training data for each section, a transition matrix is made. Using the example stated above, with input parameters being the forecasted wind speed in two hours and the forecasted wind direction in two hours, and the predicted power output in two hours as output, we would obtain the following transition matrix P

$$\mathrm{P} = \begin{bmatrix} \mathrm{p}_{1,1,1} & \mathrm{p}_{1,1,2} & \cdots & \mathrm{p}_{1,1,\hat{\mathrm{p}}} \\ \mathrm{p}_{1,2,1} & \mathrm{p}_{1,2,2} & \cdots & \mathrm{p}_{1,2,\hat{\mathrm{p}}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{p}_{1,\hat{\mathrm{u}},1} & \mathrm{p}_{1,\hat{\mathrm{u}},2} & \cdots & \mathrm{p}_{1,\hat{\mathrm{u}},\hat{\mathrm{p}}} \\ \mathrm{p}_{2,1,1} & \mathrm{p}_{2,1,2} & \cdots & \mathrm{p}_{2,1,\hat{\mathrm{p}}} \\ \mathrm{p}_{2,2,1} & \mathrm{p}_{2,2,2} & \cdots & \mathrm{p}_{2,2,\hat{\mathrm{p}}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{p}_{\hat{\mathrm{d}},\hat{\mathrm{u}},1} & \mathrm{p}_{\hat{\mathrm{d}},\hat{\mathrm{u}},2} & \cdots & \mathrm{p}_{\hat{\mathrm{d}},\hat{\mathrm{u}},\hat{\mathrm{p}}} \end{bmatrix} \tag{3.5}$$

The probability $p_{i,j,l}$ in this matrix expresses the probability of going to the power output state *l*, given that the forecasted wind speed in two hours is in state *j* and the forecasted wind direction in two hours is in state *i* (Shamshad, Bawadi, Hussin, Majid, & Sanusi, 2005). Given the number of states the parameters included in this example retain, shown in Table 3.3, the number of rows in the

transition matrix would be 144, while the number of columns would be the number of power output states for the given prediction subject. This is because the number of states for both wind speed and wind direction is 12, which to the second power is 144. And each column represents a power output state for the given wind park or turbine.

These transition matrices become the basis of future likelihood of output power for the given wind park, where the probabilities for state transitions are calculated from the frequency of transitions between states in the training data. For the verification data, each combination of the input parameters in question indicates a given row in the transition matrix. From this point, the transition matrix provides a probabilistic power output forecast, enabling the model to yield the most probable power output state in this row. The algorithm of the Markov chain models creates 12 different transition matrices based on the 12 different sets of training data, and applies the respective verification data to the transition matrix. Figure 3.8 illustrates the transition matrix of Markov chain model 1 (MCM1) for Havøygavlen, trained with data from January to November, which is then used to predict the power output for December. This model applies the predicted wind speed state in two hours and the output power state over the last hour as input parameters. And the colour bar indicates the probability from 0 to 1 of any output power state, for any given combination of input states. The first 12 rows indicate the 12 wind speed states as the output power state over the last hour was 1. The next 12 rows indicate the 12 wind speed states as the output power state over the last hour was 2, and so on.

When the predicted power output states for all 12 months are found, they are assembled and converted back to actual power output values, creating predicted power production time series for the whole year. Converting the predicted power output states, state 1 is set to be 0 MW. While the remaining states are set to be in the midpoint of the respective interval according to Table 3.4.

The NRMSE are calculated based on the actual power output time series and the predicted power time series produced, for all model compositions of input parameters.

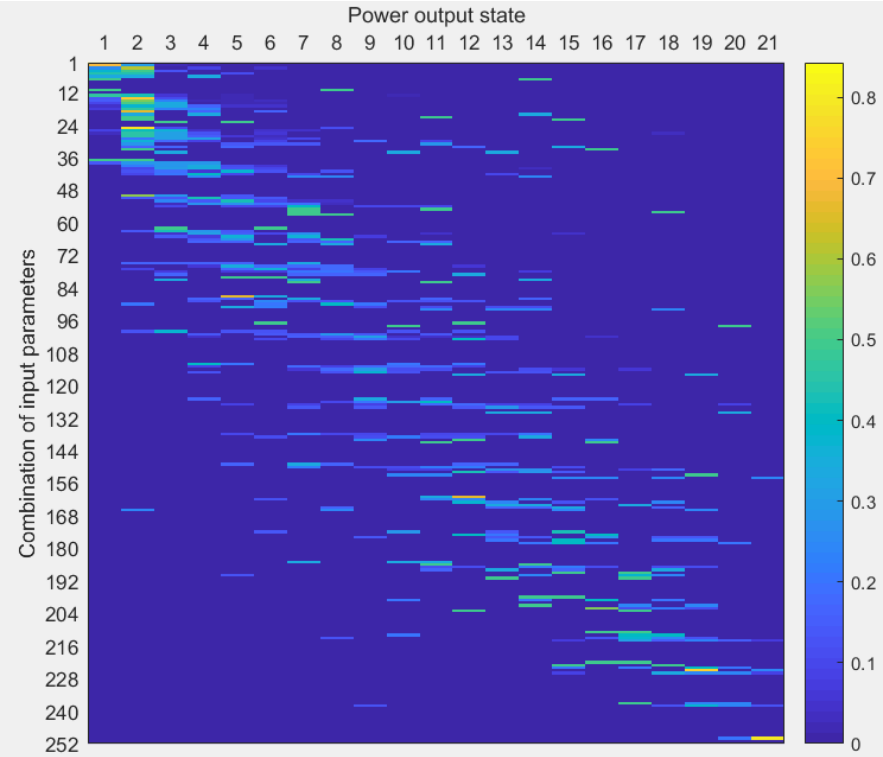

Figure 3.8: Transition matrix of MCM1 for Havøygavlen trained with data from January to November.

When generating each transition matrix, if a specific combination of parameters occurs only once in the training period, the event is too badly supported, and the row will remain all zeros. This is to avoid unnecessary bias to the prediction model. If at any time the current combination of states of the input parameters in the verification period are not found in the transition matrix, meaning no such event occurred in the training period, the model will apply the persistent model with $m = 2$, as forecast for this specific event. The frequency of the Markov chain model applying the persistent model increases with the size of the transition matrix, and therefore depends on the number of input parameters.

The biggest weakness of the persistent model is the obvious large errors when there is a sudden change in energy production, such as at the beginning or the end of a downtime period. Common for all Markov chain models is the input parameters $\widehat{U}_{t_2}$ and P_{t_0}, predicted wind speed in two hours and the output power over the last hour. Using these two parameters, we can moderate the error of the persistent model by altering it, based on the specific combination of the two parameters and what they imply. If the power output state over the last hour is small, but the predicted wind speed state in two hours is large, the predicted power output state will be larger than the power output state over the last hour, and the other way around. The modification of the persistent model will be according to a modification matrix, and is applied solely within the Markov chain models, not the independent persistent model used as reference. The entries of the modification matrices for all wind parks and the Nordex N80 turbine, are the closest integer produced by the following formula

$$MM\left(\widehat{U}_{t_2}, P_{t_0}\right) = \frac{\left(\left(\frac{P_{max}}{2} - P_{t_0}\right) \cdot \left(\frac{U_{max}}{P_{max}}\right)\right) - \left(\frac{U_{max}}{2} - \widehat{U}_{t_2}\right)}{2} \qquad (3.6)$$

where U_{max} is the maximum wind speed state, P_{max} is the maximum power output state, P_{t_0} is the power output state over the last hour, and $\widehat{U}_{t_2}$ is the predicted wind speed state in two hours. Figure 3.9 shows the modification matrix of Havøygavlen and Kjøllefjord.

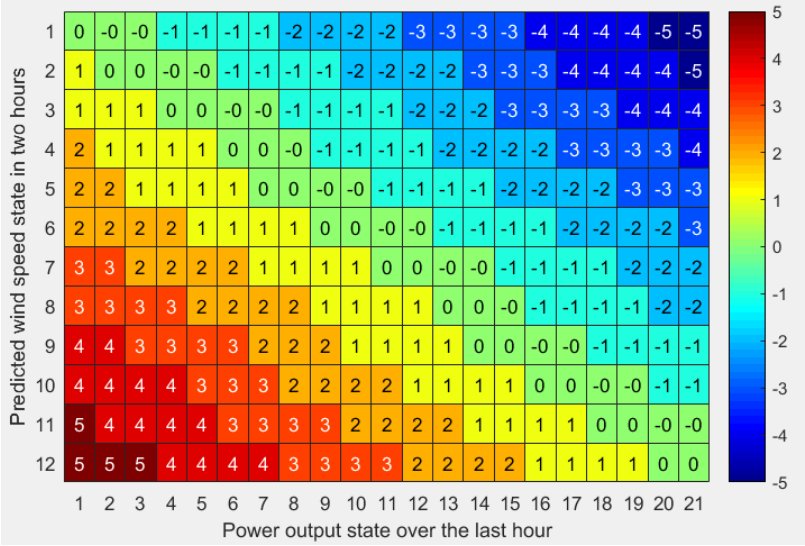

Figure 3.9: Modification matrix for Havøygavlen and Kjøllefjord

Each Markov chain model's corresponding transition matrix, for all 12 iterations of the algorithm, provides a probabilistic power output forecast, yielding the most probable power output. However, if at any time the current combination of states of the input parameters in the verification period, corresponds to a row in the transition matrix containing multiple equal maximum probabilities, the forecasted power output state closest to the persistent model state is used.

By inspecting the biggest errors along the predicted power production time series, and analysing the transition matrix for the given combination of parameters producing these errors, there is an evident pattern that the entries are sparse, and the maximum probability is below 50%. Therefore, if the maximum probability in the row indicated by a set of input parameters is smaller than 50%, the produced predicted power state is the closest integer of the sum of all occurred training states multiplied by their respective probability for the given combination of parameters.

This is illustrated in Figure 3.10, showing row 66 of the transition matrix of MCM1 for Kjøllefjord, trained with data from January to November. Here, the largest probability, 0.25, is located at power output state 4. However, this does not represent the whole probabilistic forecast. To counter this issue, the algorithm uses the closest integer to the mean row probability, in this case 8, as power output state.

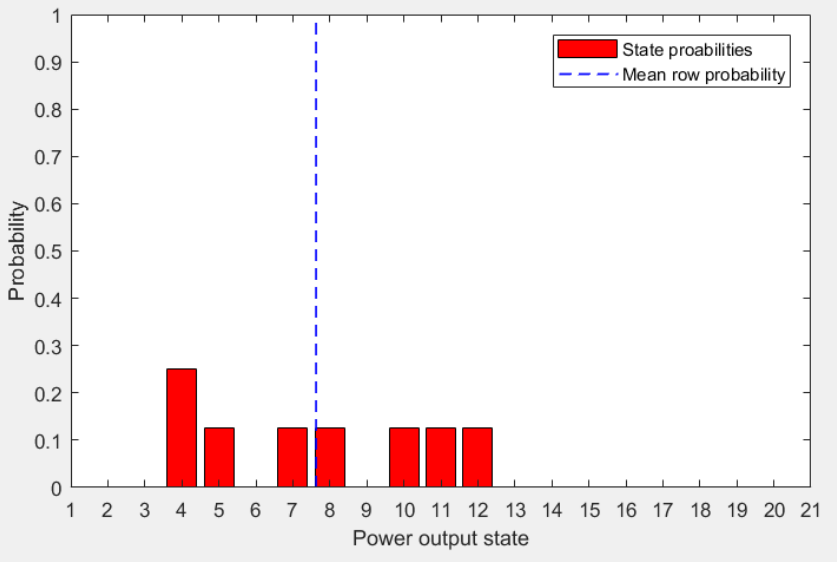

Figure 3.10: Row 66 of the transition matrix of MCM1 for Kjøllefjord, trained with data from January to November, corresponding to the combination of input states $P_{t_0} = 6$ and $\bar{U}_{t_2} = 6$.

3.3.6 Meteorological data-customized power curve function

In practice, the instantaneous power output will not necessarily follow the idealized relationships of the power curves described in section 3.3.3. These power curves are averages of empirical data, and meso- and micro-scale interference can influence the operation of the turbines (Rosen & Sheinman, 1996). Applying raw meteorological forecast data to these power curves, may lead to an overestimation in predicted power production, as the power curves produced by the WindPRO software are not developed for this purpose. The strongest variability is where the power curves of turbines is very sensitive to changes, such

as in the vicinity of cut-in wind speed and where they approach their rated power (Früh, 2012). The 1 m/s intervals of input predicted wind speed, combined with prediction error and uncertainty in the meteorological data, can result in substantial errors in the predicted power production.

A potential improvement in prediction performance, could be to use meteorological prediction data combined with actual power production data, to create an independent power curve. As for the Markov chain models, the leave-one-out method, described in Section 3.3.4, will be used to provide both the power curve functions and new independent verification data. Each verification period will consist of one month of data, with the remaining 11 months being the basis of the power curves. Error measures will only be based on the verification period, so that the predicted power output for the whole year is obtained using a power curve function based on independent data. The input parameters will be the predicted wind speed and direction in two hours, $\widehat{U}_{t_2}$ and $\widehat{D}_{t_2}$, along with the actual output power in two hours, P_{t_2}. The output parameter is the predicted power production in two hours, $\hat{P}_{t_2}$.

The power production data and associated predicted wind speed for each training period, is sectioned into 12 wind direction sectors at 30° intervals. Meaning, for each wind park there will be 12 power curve functions for each verification period, where the predicted wind direction determines which function to be applied, and the predicted wind speed determines the predicted output power. A sixth degree polynomial is fitted to a scatter plot of all observations in the given training period for each wind direction sector. This is obtained by polynomial regression using the method of least squares, described in section 2.3.2. The coefficients of the polynomial are established according to Equation (2.18). The degree of the polynomial regression, is chosen on the basis of optimizing the model output NRMSE.

Polynomial regression is limited by its global nature, where the fitted value of output power at a given wind speed depends strongly on all data values, even those wind speeds that are far from the given wind speed. Thus, it is not easy to attain a functional form in a specific wind speed region, without impairing the integrity of the fitted curve in other regions (Shokrzadeh, Jozani, & Bibeau, 2014). The sensitivity of polynomial regression to anomalies or outliers within the data can be reduced by smoothing the data before fitting the polynomial. Locally Weighted Scatterplot Smoothing, or LOWESS, is applied to the scatter plot of the power data and wind speed data. This is a robust nonparametric scatter plot smoother, which implements a form of local linear regression. Each data point is produced by a weighted regression fit to neighbouring data points, with weights declining as the value gets further from the value of the focal point (Fox & Weisberg, 2018). In assigning low weights to observations which generate large residuals, the method provide resistance against outliers. The degree of smoothness depends on the *span*, describing the fraction of neighbouring data included in each local regression. The *span* is chosen to optimize the model output NRMSE. The LOWESS smoother is applied to four wind speed intervals of the data separately, 0 m/s to cut-in speed, cut-in speed to rated speed, rated speed to cut-out speed, and wind speeds above cut-out speed. The resulting power curve used to predict the output power through December for Kjøllefjord wind park is shown in Figure 3.11. This example is for wind directions between 330 and 360 degrees where, according to the leave-one-out method, consists of data obtained for January to November.

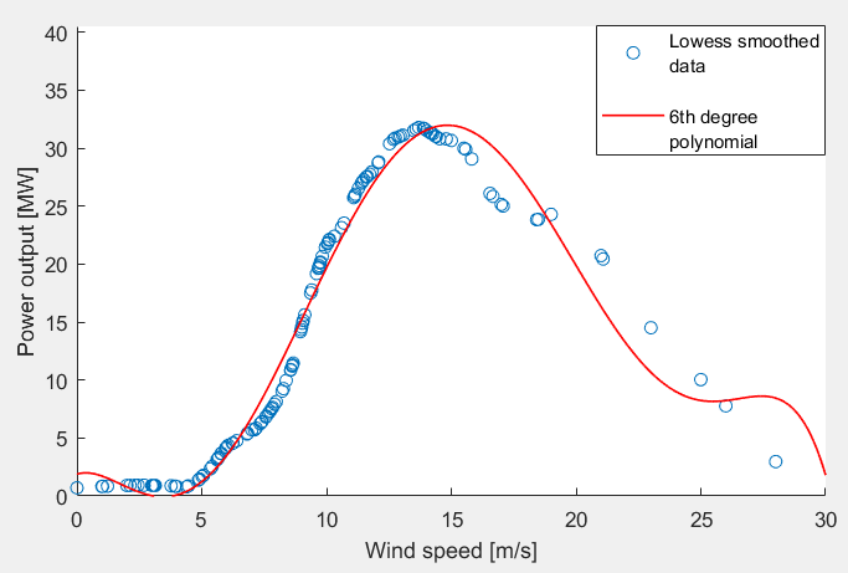

Figure 3.11: Power curve function for wind directions between 330 and 360 degrees for Kjøllefjord wind park. The data included is the predicted wind speed and output power for January to November 2017.

Implementing the meteorological data-customized power curve function to short-term wind power prediction for a given verification period, the predicted wind direction determines to which function the predicted wind speed is entered. Thus, producing a predicted power output for the given moment in time.

3.3.7 Combination of models

In some cases it might be beneficial to combine models in order to optimize the performance. The diversification of combining models may reduce the weaknesses of the individual models, by apprehending multiple aspects of the data from different modelling approaches. Especially regarding state-dependent models, when the performance of the model depends on the verification conditions being the same as the training conditions.

When combining models in this study, a linear combination with scaling constant a, and coefficients $k_1 = (1 - a)$ and $k_2 = a$, will be applied. According to Equation (2.19), we have

$$\hat{P}_{M1,M2} = (1 - a)\hat{P}_{M1} + a\hat{P}_{M1} \tag{3.7}$$

where M1 and M2 are two separate prediction models. The performance is optimized by calculating the NRMSE for all values of a in the evenly spaced interval [0,1], with 0.01 spacing between each value. The optimal value of a is chosen for when the lowest possible NRMSE is achieved.

4 Results

4.1 Wind roses

The MEPS weather forecast data for the time period between 1. January 2017 and 31. December 2017, containing hourly point prediction data of wind speed and direction, is used to construct and illustrate the wind distribution at all five wind parks. For lead time 0 of the meteorological forecast data, the wind distributions are illustrated by wind roses given in Figure 4.1 through Figure 4.5, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively. The wind speed and direction constituting the wind roses are the measurements done at the time of prediction, 4 times each day through the whole time period. The wind roses are divided into 16 directional segments, with each representing a 22.5° interval, centred at the designated direction. The segments represents the four cardinal directions, with the four respective intercardinal directions, and eight secondary intercardinal directions. Figure 4.6 through Figure 4.10 illustrates the seasonal variations of the sites. The data is split into a winter half-year, consisting of data for January to the end of March and October to the end of December, and a summer half-year, consisting of the remaining months. Considering the data analysis of the sites include only one year of data, one cannot conclusively state that the wind regimes presented in this study is a regular occurrence at all five wind parks.

For Havøygavlen wind park, the wind rose displays a rather even wind distribution, shown in Figure 4.1. However, the most common wind direction is observed from the south-southwest (SSW), seemingly corresponding to the less resistant strait between the mainland and Rolvsøya in the west. The occurrence of high wind speeds are largely originating from the western semicircle, and fairly uniformly distributed in this region. Few winds were observed from the directional region between north (N) and northeast (NE), and from the south-southeast (SSE).

Where the former is likely the result of Hjelmsøya blocking or redirecting the wind flow originating from this direction, and the latter is presumably a similar consequence of the mountain range located on the mainland in the aforementioned direction. The islands, straits and mountain range described are shown in Figure 3.1, with the terrain elevation indicated by colour. Furthermore, Figure 3.2 indicates a near perpendicular relation between the two parallel lines in which the wind turbines are placed, and the dominant wind direction (SSW). In that case minimizing the wake effects, and justifying the positioning of the wind turbines. The seasonal variation at Havøygavlen wind park is displayed in Figure 4.6. This illustrates a more even distribution of lower wind speeds during the summer half-year, while the winter half-year exhibit higher wind speeds in general, dominated by the south-southwestern wind direction.

Figure 4.2 illustrates the wind regime for the surrounding area of Kjøllefjord wind park. The wind rose displays a rather even wind distribution, with the exception of one dominant wind direction, south-southwest (SSW), comprising 17.1% of all measured wind directions. Given the size of the adjacent directional segments, southwest (SW) and south (S), one might assume they are subject to the same topographic effect. From the description of the surrounding topography in section 3.1.2, and by inspecting Figure 3.1, the dominant wind direction is seemingly a result of the large north-south going fjord, Laksefjorden, running along the west side of the peninsula. In general, higher wind speeds appear to be rarely occurring for the site altogether. Figure 4.7 illustrates the seasonal variation at Kjøllefjord wind park. The wind roses demonstrate a more even distribution of lower wind speeds during the summer half-year, while the winter half-year exhibit higher wind speeds in general, dominated by the south-southwestern wind direction.

Inspecting the wind regime at Nygårdsfjellet wind park, shown in Figure 4.3, we can observe a strong correlation with the surrounding topography. The wind rose displays two highly dominant wind directions, east (E) and west (W), comprising 44.5% and 17.6% of all measured wind directions respectively.

The wind regime of the site is evidently a result of the west-east going valley, with large mountains in the north and south, where the wind park is located. This is illustrated in Figure 3.1 and Figure 3.4. The most common wind direction, east (E), is also where the highest occurrence of high wind speeds are originating from. The seasonal variation at Nygårdsfjellet wind park is displayed in Figure 4.8. The winter half-year exhibit a larger occurrence of high wind speeds, than for the remaining six warmer months. This implies that there may be a correlation between high winds and cold temperatures at the site, further supporting the findings of Bilal (2016).

For Fakken wind park, the wind rose displays a wind distribution seemingly affected by the surrounding complex topography and nearby islands. The dominant wind directions demonstrated in Figure 4.4, are winds originating from the south-southwest (SSW) and the southeast (SE). The highest occurrence of high wind speeds is also observed from the southeastern wind direction. From the description of the surrounding topography in section 3.1.4, and by inspecting Figure 3.1, the dominant wind directions may be a result of the two large north-south going fjords in the south, Ullsfjorden and Lyngen, separated by the Lyngen Alps. Few winds were observed from the east (E) and west (W), including adjacent segments of the two cardinal directions. The former is likely the result of the high mountains and mountain ranges located west on Vannøya, shown in Figure 3.5. The latter is presumably the result of the mountainous island Arnøya, located to the east, providing shelter from winds blowing from this direction. For the remaining wind directions, the wind rose displays a rather even distribution, largely consisting of lower wind speeds. The seasonal variation at Fakken wind park is displayed in Figure 4.9. This illustrates a high occurrence of lower wind speeds originating from the open sea in the north (N) and north-northeast (NNE), during the summer half-year. While the winter half-year exhibit a wind pattern fairly dominated by the south-southwestern (SSW) wind direction.

Figure 4.5 illustrates the wind regime for the surrounding area of Raggovidda wind park. The wind rose displays a rather even wind distribution, and in general a very large occurrence of higher wind speeds compared to the other four wind parks analysed in this study. The most common wind directions are observed between and including the directional segments of northwest (NW) and south-southwest (SSW). The highest occurrence of high wind speeds is also located in this directional region, with winds from the southwest (SW) being the most frequent direction of high wind speeds. In fact, the mean wind speed of all measured winds originating from the southwest in the yearly data, is 10.2 m/s. The wind regime illustrated by Figure 4.5, with the easterly dominant wind direction region, is seemingly a result of the large north-south going fjord in the west, Tanafjorden, shown in Figure 3.6. The lower gradient terrain westward of the wind park, leading all the way down to the fjord, may provide less resistant pathways for the wind. Figure 4.10 illustrates the seasonal variation at Raggovidda wind park. The wind roses demonstrate a clear contrast in wind pattern between the two periods. There is a substantially higher occurrence of higher wind speeds during the winter half-year, mainly originating from the western semicircle. While the summer half-year exhibit a more even distribution of lower wind speeds.

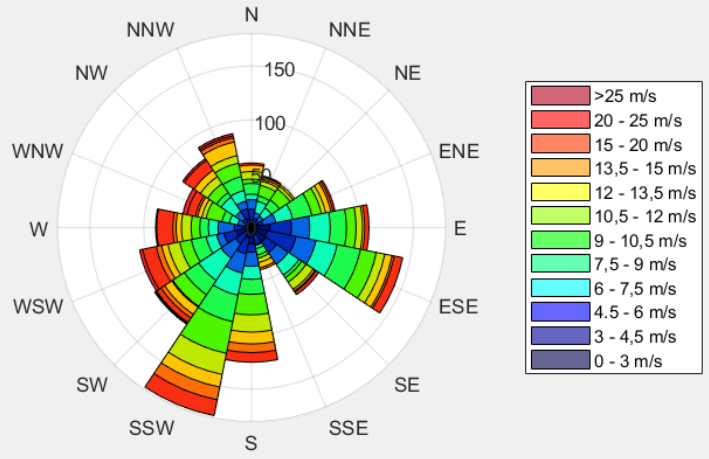

Figure 4.1: Wind rose for Havøygavlen wind park in the time period 1. January 2017 - 31. December 2017.

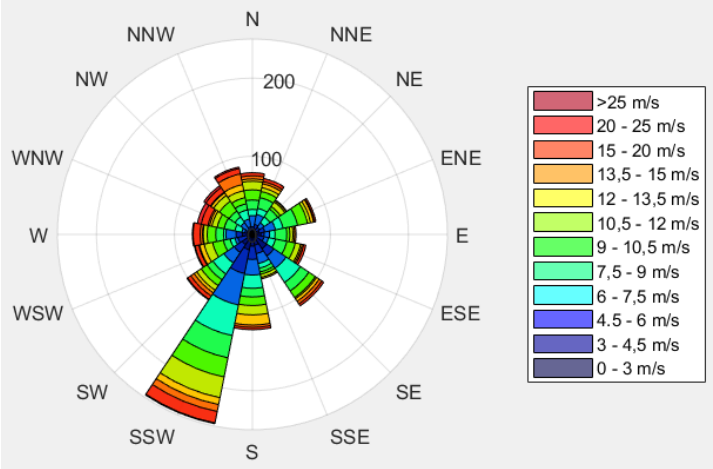

Figure 4.2: Wind rose for Kjøllefjord wind park in the time period 1. January 2017 - 31. December 2017.

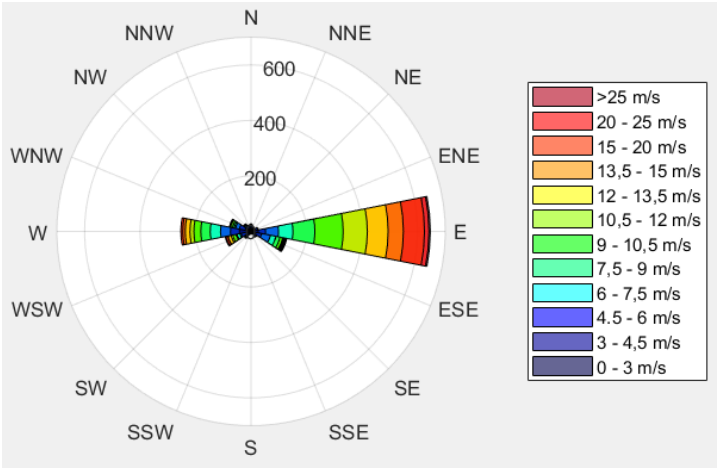

Figure 4.3: Wind rose for Nygårdsfjellet wind park in the time period 1. January 2017 - 31. December 2017.

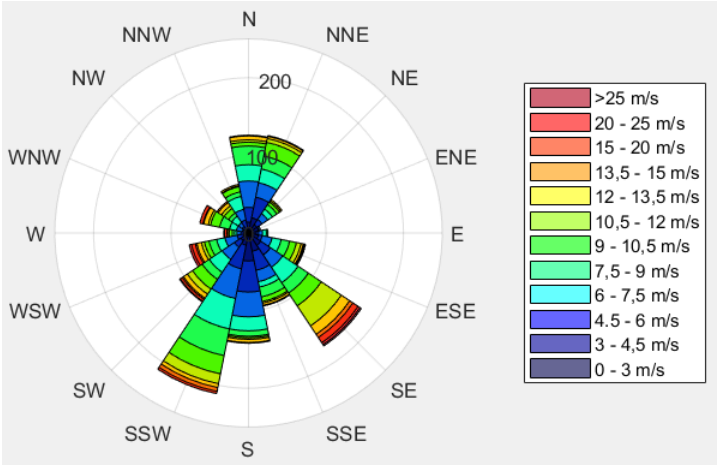

Figure 4.4: Wind rose for Fakken wind park in the time period 1. January 2017 - 31. December 2017.

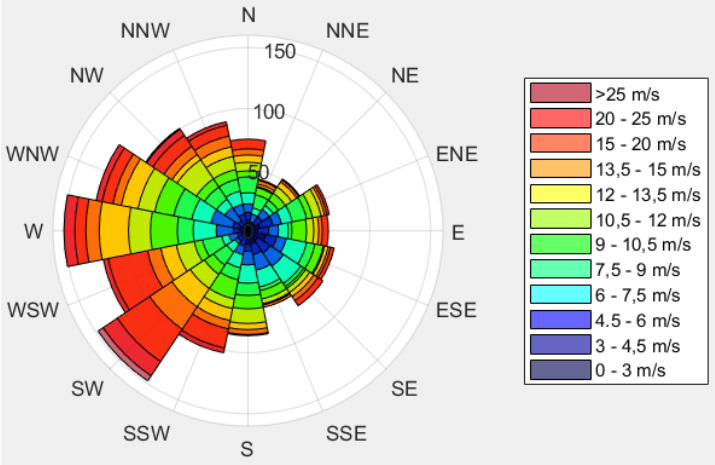

Figure 4.5: Wind rose for Raggovidda wind park in the time period 1. January 2017 - 31. December 2017.

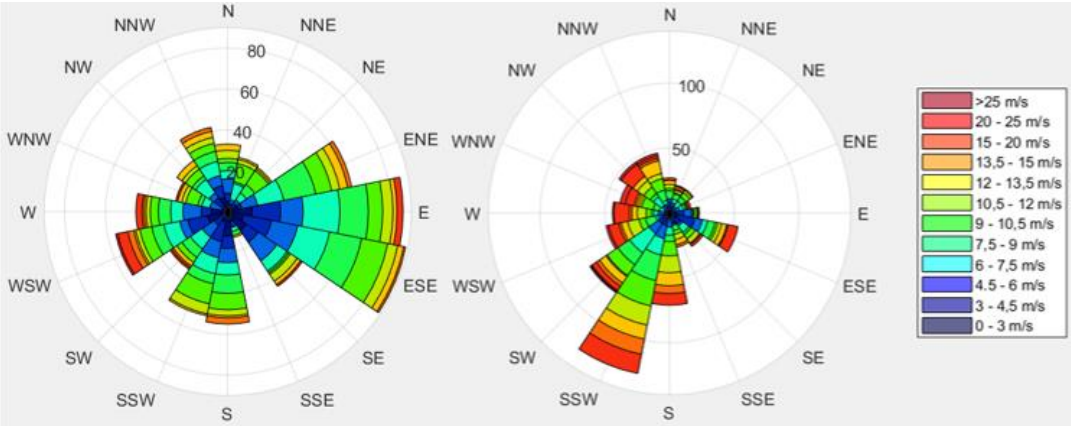

Figure 4.6: Wind roses for Havøygavlen wind park, with the data from 2017 split into summer half-year (left) and winter half-year (right).

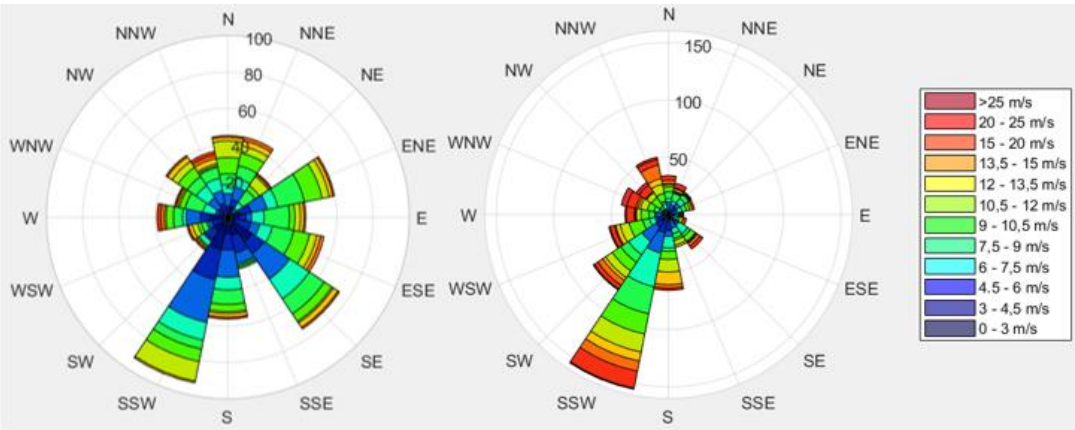

Figure 4.7: Wind roses for Kjøllefjord wind park, with the data from 2017 split into summer half-year (left) and winter half-year (right).

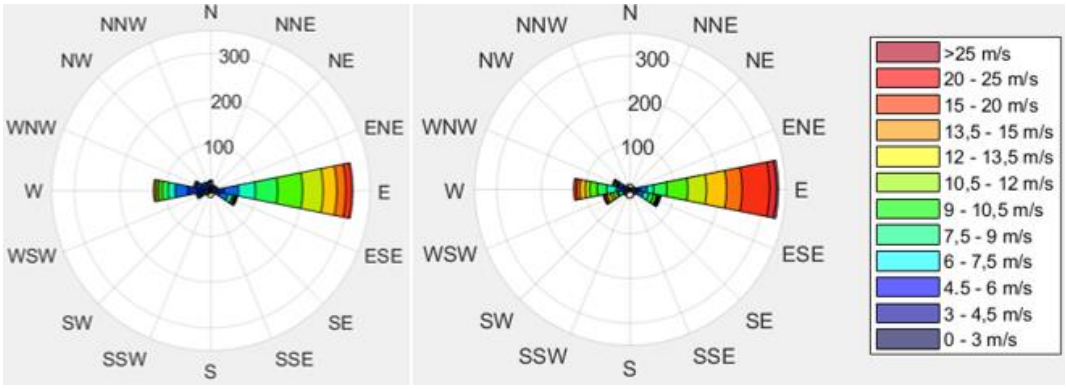

Figure 4.8: Wind roses for Nygårdsfjellet wind park, with the data from 2017 split into summer half-year (left) and winter half-year (right).

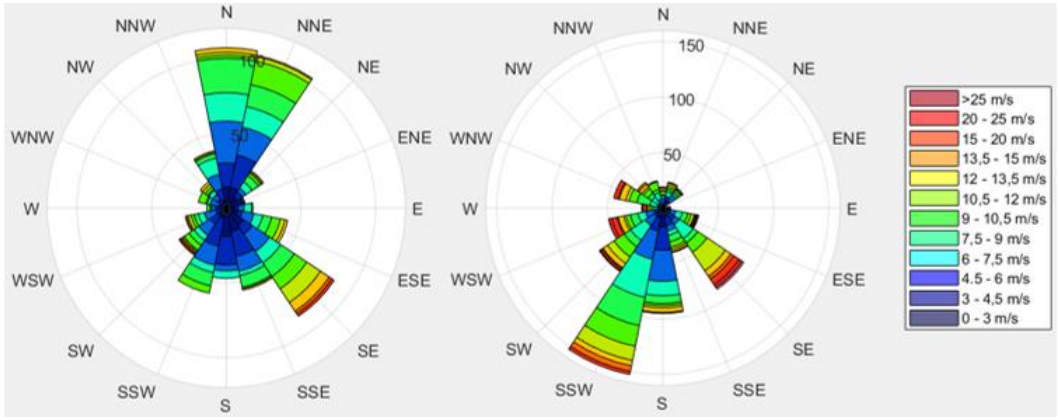

Figure 4.9: Wind roses for Fakken wind park, with the data from 2017 split into summer half-year (left) and winter half-year (right).

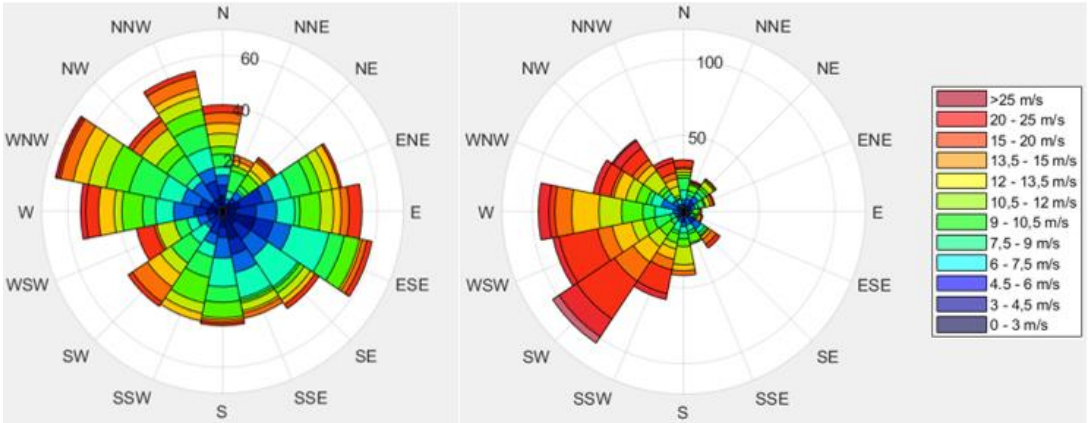

Figure 4.10: Wind roses for Raggovidda wind park, with the data from 2017 split into summer half-year (left) and winter half-year (right).

4.2 Power curve models

The site-specific power curve function provided by Kjeller Vindteknikk AS, henceforth referred to as the PCF, along with the meteorological data-customized power curve function, henceforth referred to as the MET-PCF, is illustrated in Figure 4.13 through Figure 4.17 for all five wind parks. The colour axis scaling is the same for both the MET-PCF plot and the site-specific power curve plot, ranging from zero to the rated power of the given wind park. The MET-PCF is developed according to the description in section 3.3.6, with the exception of all 12 months of data being the basis of the power curves. The prediction performance of the PCF and the MET-PCF are given in terms of the NRMSE and the improvement parameter I, in Table 4.1 through Table 4.5, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively. The output parameter is chosen to be the predicted output power in two hours ($\hat{P}_{t_2}$) for both models, and their performance is measured by the extent in which they can improve on the persistent model.

The immediately observed disparity between the two models, is that the PCF produces a substantially larger power output in the region of rated wind speeds, than the MET-PCF. The illustration of one of the power curve functions of the MET-PCF at Kjøllefjord, by Figure 3.11, displayed some obvious errors. The shape of the power curve function indicates power production at wind speeds below cut-in wind speed, and beyond cut-out wind speed. This tendency is also seen in the illustrations given by Figure 4.13 through Figure 4.17, for all five wind parks. In the boxes below cut-in wind speed, where the PCF indicates no power production represented by the dark blue colour, the MET-PCF produces scattered lighter blue colours for different wind speeds and directions. And where the PCF reaches cut-out wind speeds and displays rather abrupt transitions to zero power output, the MET-PCF displays a slower transition, occasionally indicating power production reaching wind speeds as far as 30 m/s.

These errors may imply a weakness in the polynomial regression fitting of the data. Also, the meteorological forecast data might not accurately reflect the actual observed wind at the site. A possible solution to the former issue could be to only define the power curve functions within the respective cut-in and cut-out wind speeds, possibly with an additional margin of forecast error. More data could also improve the power functions by obtaining more appropriate and favourable curve patterns.

The prediction performance of both the PCF model and the MET-PCF model are comparatively bad for all five wind parks, with the persistent model outperforming both models. However, the MET-PCF model performs substantially better than the PCF model. Observing the comparisons of the two models in Figure 4.13 through Figure 4.17, and considering the training data used to develop the MET-PCF models is closely associated to the data used as input in the PCF models, it is evident that the large NRMSE of the PCF model is a case of overestimation. For instance, the bias of the PCF model at Havøygavlen is 5.56 MW, constituting 13.73% of installed capacity of the wind park. However, the WindPRO software used to produce the direction-specific power curves applied in the PCF model, assumes a power production where all turbines at the wind park are working properly at all times, and is evidently not created for the purpose employed in this study. Anything affecting the healthy operation of the wind park is not accounted for, leaving the insertion of raw data impractical, and does not reflect a conventional power production for a wind park. A way of removing the apparent bias and increase the prediction performance, may be to introduce some sort of scaling to the predicted output power. Either a direct approach, by scaling resulting power output, or an indirect approach, by scaling the input verification data to optimize the model output NRMSE.

By temporal error analysis of the MET-PCF model, it is evident that the pure environmental approach of the model, using only meteorological forecast data as input, leads to a vulnerability to downtime periods and performance issues affecting the healthy operation of the wind park.

When these circumstances coincides with the input meteorological forecast data indicating normal power production, an overestimation is produced. To counteract and reduce prediction errors regarding such behaviour, a linear combination of the persistent model and the MET-PCF model is formed, henceforth referred to as the MET-PCF + PM. According to Equation (2.19), MET-PCF + PM is the linear combination the predicted power output using the MET-PCF model and the persistent model with $m = 2$. A scaling constant a is used to optimize the performance by calculating the NRMSE of the linear combination, with coefficients $k_1 = (1 - a)$ and $k_2 = a$, so that

$$\hat{P}_{MET-PCF+PM} = (1 - a)\hat{P}_{MET-PCF} + a\hat{P}_{PM} \tag{4.1}$$

The NRMSE is calculated for all values of a in the evenly spaced interval [0,1], with 0.01 spacing between each value. The optimal value of a is then chosen for when the lowest possible NRMSE is achieved, for each of the five wind parks. Figure 4.11 shows the plot of the NRMSE of the model MET-PCF + PM at Kjøllefjord wind park, for all values of the scaling constant a. The NRMSE of the MET-PCF model and the persistent model, are consequently found at a = 0 and a = 1 respectively.

Figure 4.12 shows how the overestimation of the MET-PCF model is reduced by including the persistent model, in a period where the conditions indicated by the meteorological forecast data indicates a larger power production than what is actually produced. The performance of the MET-PCF + PM model is given in terms of the NRMSE and the improvement parameter I, in Table 4.1 through Table 4.5, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively. The output parameter is chosen to be the predicted output power in two hours ($\hat{P}_{t_2}$), and the performance of the model is measured by the extent in which it can improve on the persistent model.

The MET-PCF + PM model performs remarkably well for all five wind parks. The improvement parameters for each wind park are 16.08%, 11.33%, 17.38%, 16.91% and 19.30%, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively. This signifies a substantial gain of using the MET-PCF + PM, compared to the persistent model used as reference. The extensive reduction in NRMSE by the inclusion of the persistent model to the MET-PCF model, through linear combination, suggests that the weakness to the MET-PCF model previously reviewed, has been diminished considerably.

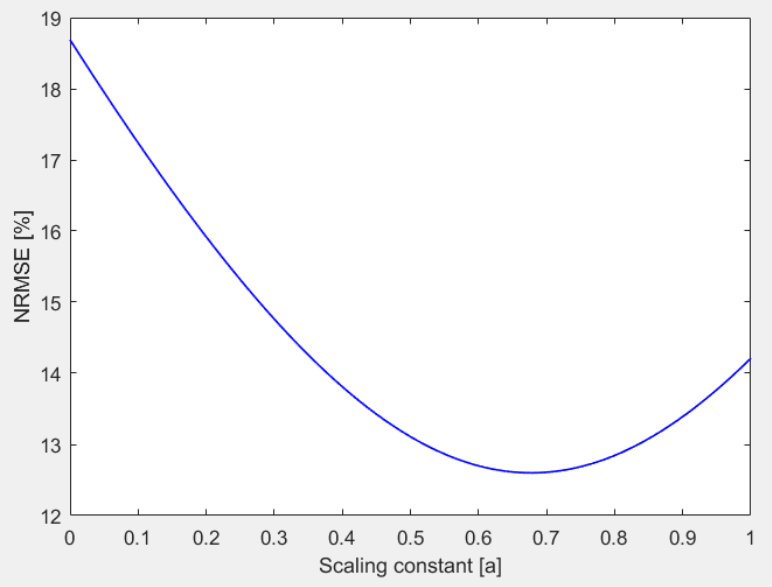

Figure 4.11: Plot of the NRMSE of the model MET-PCF + PM at Kjøllefjord wind park, for all values of the scaling constant 'a' in the linear combination (1-a)MET-PCF + aPM.

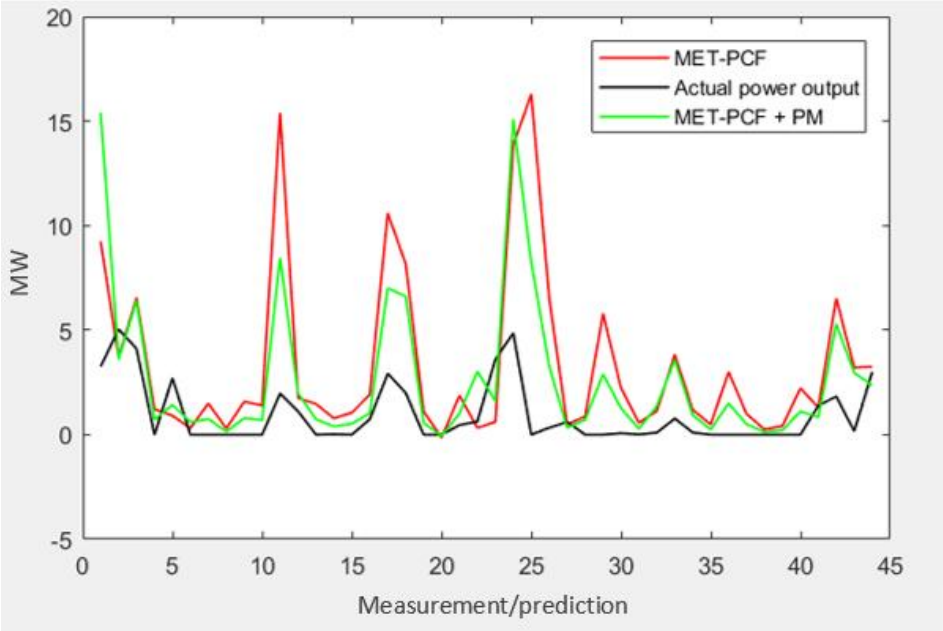

Figure 4.12: Plot comparing the actual power output at Kjøllefjord wind park 7. November - 18. November, to the predicted power output using MET-PCF and MET-PCF + PM.

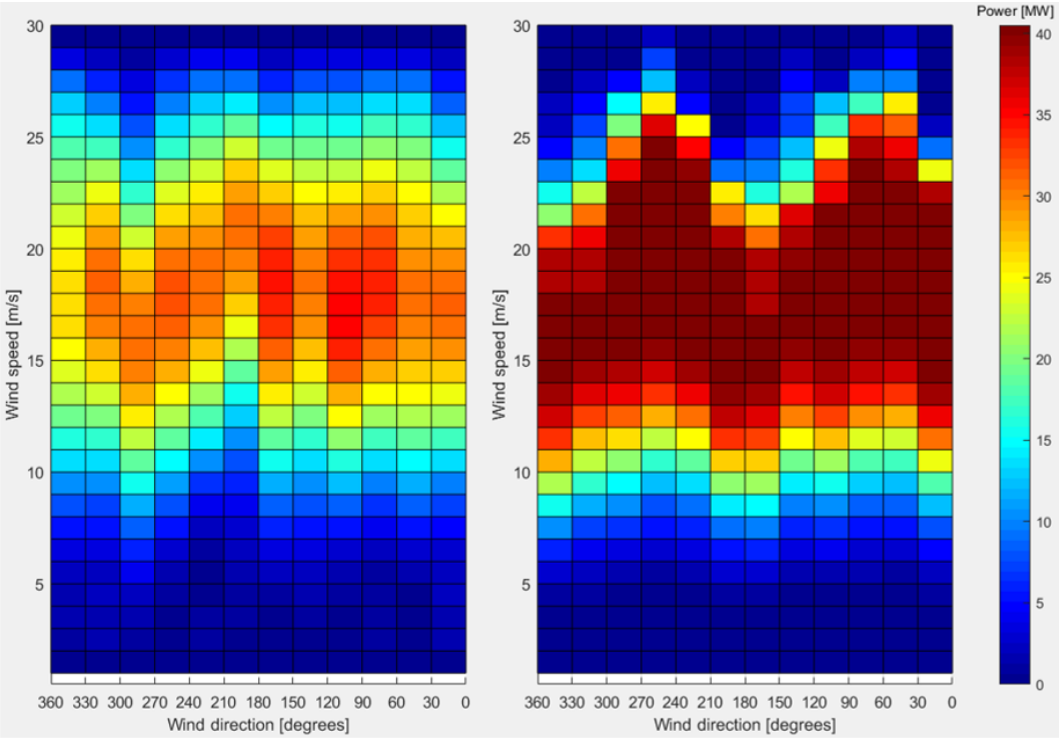

Figure 4.13: The predicted power output at Havøygavlen wind park for a given wind speed and direction, for the MET-PCF (left) and the site-specific power curve provided by Kjeller Vindteknikk AS (right).

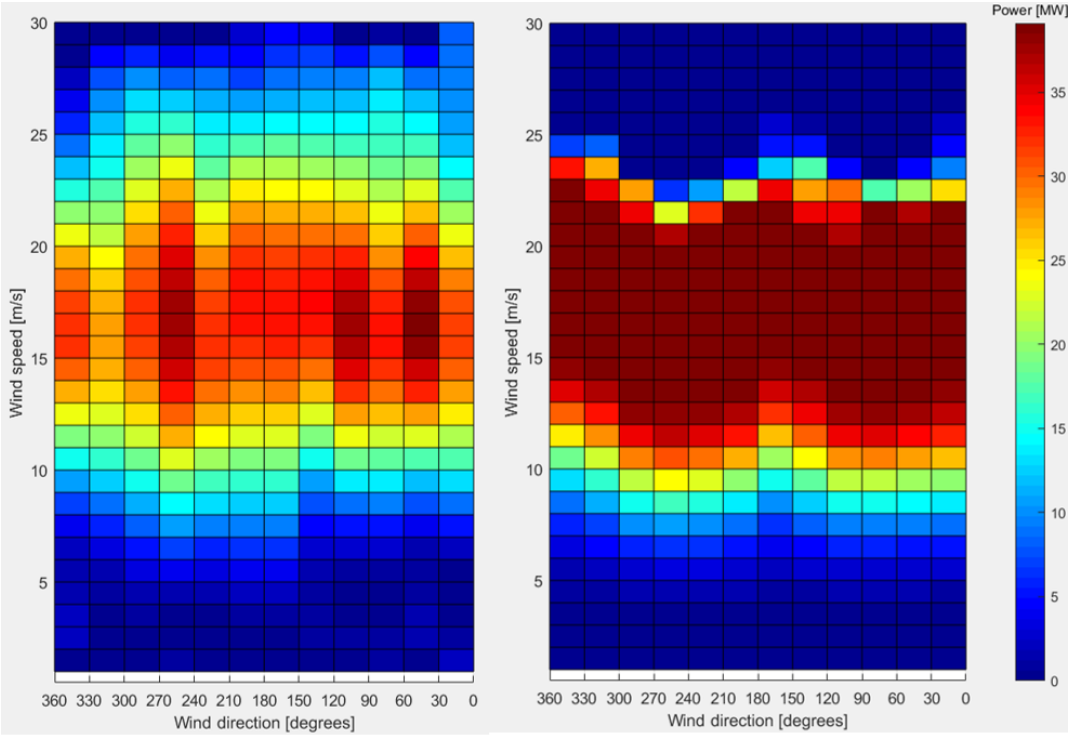

Figure 4.14: The predicted power output at Kjøllefjord wind park for a given wind speed and direction, for the MET-PCF (left) and the site-specific power curve provided by Kjeller Vindteknikk AS (right).

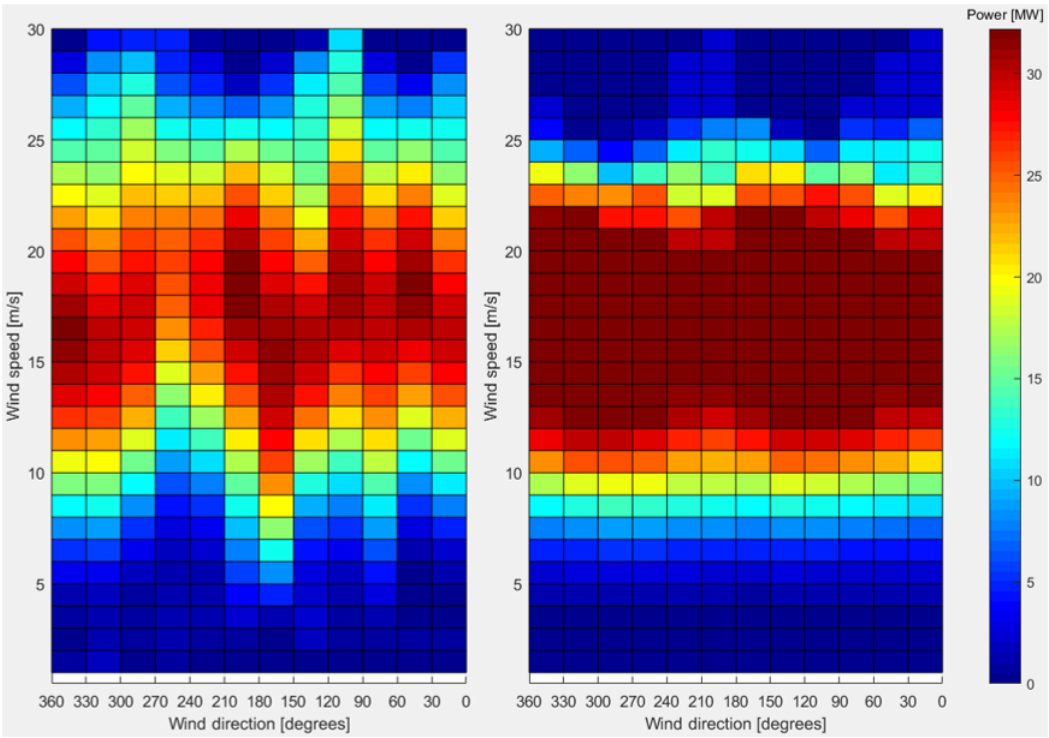

Figure 4.15: The predicted power output at Nygårdsfjellet wind park for a given wind speed and direction, for the MET-PCF (left) and the site-specific power curve provided by Kjeller Vindteknikk AS (right).

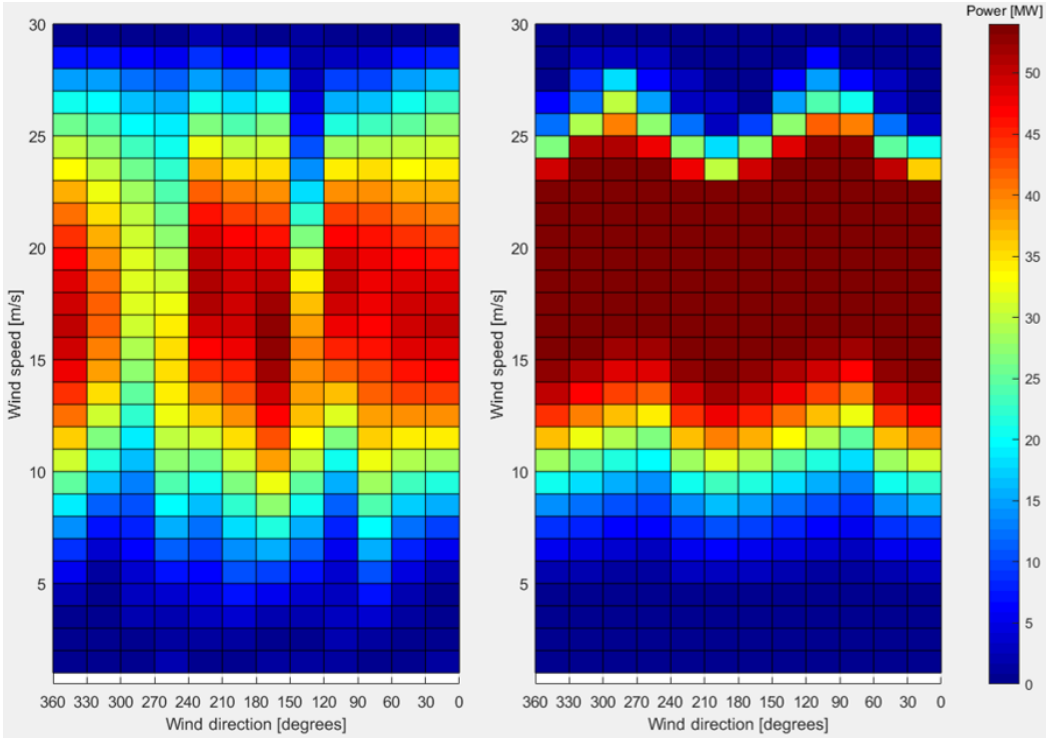

Figure 4.16: The predicted power output at Fakken wind park for a given wind speed and direction, for the MET-PCF (left) and the site-specific power curve provided by Kjeller Vindteknikk AS (right).

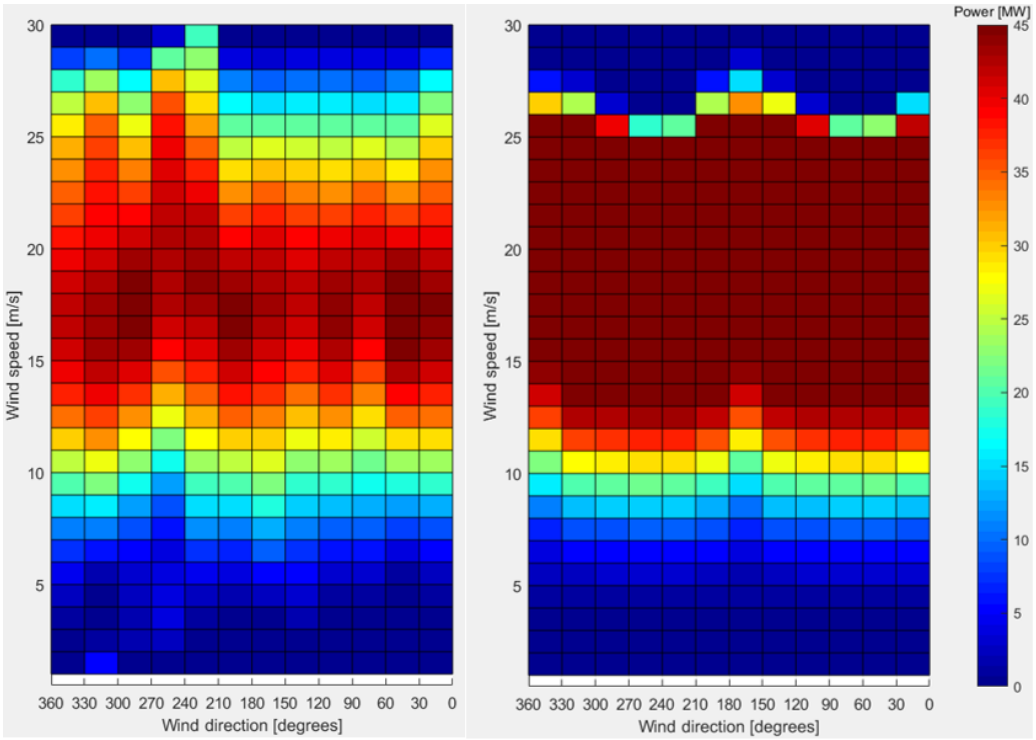

Figure 4.17: The predicted power output at Raggovidda wind park for a given wind speed and direction, for the MET-PCF (left) and the site-specific power curve provided by Kjeller Vindteknikk AS (right).

4.3 Markov chain model

A total of 5 Markov chain models with different combinations of input parameters where tested for all five wind parks, according to the method described in section 3.3.5, using the leave-one-out method described in section 3.3.4. The output state parameter is chosen to be the predicted power in two hours ($\hat{P}_{t_2}$) for all Markov chain models.

A persistent model (PM) with $m = 2$, has been run for the whole year, according to Equation (3.3). This implies that the model uses the average power output over the last hour as the forecasted power output of between the next hour and two hours into the future. The persistent model will serve as a reference model to all formulated methods, for all five wind parks. For every model, the normalized root mean square error (NRMSE) for the whole year is calculated. The improvement parameter *I* is calculated according to Equation (2.25), with the persistent model as reference and NRMSE as the evaluation criterion. The results are shown in Table 4.1 through Table 4.5, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively.

Table 4.1: Overview of the performance of the persistent model (PM), the power curve function model (PCF), the Markov chain models (MCM), the meteorological data-customized power curve function model (MET-PCF), and the combined MET-PCF and persistent model (MET-PCF + PM) at Havøygavlen, in terms of the NRMSE and the improvement parameter I.

Model	**Input parameters**	**Output parameters**	**NRMSE (%)**	**Improvement parameter I (%)**
Havøygavlen				
PM	P_{t_0}	$\hat{P}_{t_2}$	13.62	
PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	25.73	-88.91
MCM 1	$\hat{U}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.35	1.98
MCM 2	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	12.81	5.95
MCM 3	$\hat{U}_{t_2}$, $\hat{T}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	12.78	6.17
MCM 4	$\hat{U}_{t_2}$, $\hat{p}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.00	4.55
MCM 5	$\hat{U}_{t_2}$, $\hat{U}_{t_0}$, P_{t_0}	$\hat{P}_{t_2}$	12.74	6.46
MET-PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	14.65	-7.56
MET-PCF + PM	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	11.43	16.08

Table 4.2: Overview of the performance of the persistent model (PM), the power curve function model (PCF), the Markov chain models (MCM), the meteorological data-customized power curve function model (MET-PCF), and the combined MET-PCF and persistent model (MET-PCF + PM) at Kjøllefjord, in terms of the NRMSE and the improvement parameter I.

Model	**Input parameters**	**Output parameters**	**NRMSE (%)**	**Improvement parameter *I* (%)**
Kjøllefjord				
PM	P_{t_0}	$\hat{P}_{t_2}$	14.21	
PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	25.34	-78.33
MCM 1	$\hat{U}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.93	1.97
MCM 2	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.52	4.86
MCM 3	$\hat{U}_{t_2}$, $\hat{T}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.87	2.39
MCM 4	$\hat{U}_{t_2}$, $\hat{p}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.66	3.87
MCM 5	$\hat{U}_{t_2}$, $\hat{U}_{t_0}$, P_{t_0}	$\hat{P}_{t_2}$	13.81	2.81
MET-PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	18.69	-31.53
MET-PCF + PM	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	12.60	11.33

Table 4.3: Overview of the performance of the persistent model (PM), the power curve function model (PCF), the Markov chain models (MCM), the meteorological data-customized power curve function model (MET-PCF), and the combined MET-PCF and persistent model (MET-PCF + PM) at Nygårdsfjellet, in terms of the NRMSE and the improvement parameter I.

Model	**Input parameters**	**Output parameters**	**NRMSE (%)**	**Improvement parameter *I* (%)**
		Nygårdsfjellet		
PM	P_{t_0}	$\hat{P}_{t_2}$	15.36	
PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	26.37	-71.68
MCM 1	$\hat{U}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	14.90	2.99
MCM 2	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	14.55	5.27
MCM 3	$\hat{U}_{t_2}$, $\hat{T}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.93	9.31
MCM 4	$\hat{U}_{t_2}$, $\hat{p}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	14.18	7.68
MCM 5	$\hat{U}_{t_2}$, $\hat{U}_{t_0}$, P_{t_0}	$\hat{P}_{t_2}$	14.10	8.20
MET-PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	17.82	-16.02
MET-PCF + PM	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	12.69	17.38

Table 4.4: Overview of the performance of the persistent model (PM), the power curve function model (PCF), the Markov chain models (MCM), the meteorological data-customized power curve function model (MET-PCF), and the combined MET-PCF and persistent model (MET-PCF + PM) at Fakken, in terms of the NRMSE and the improvement parameter I.

Model	**Input parameters**	**Output parameters**	**NRMSE (%)**	**Improvement parameter *I* (%)**
Fakken				
PM	P_{t_0}	$\hat{P}_{t_2}$	16.03	
PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	20.55	-28.20
MCM 1	$\hat{U}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	15.34	4.30
MCM 2	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	14.51	9.48
MCM 3	$\hat{U}_{t_2}$, $\hat{T}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	15.38	4.05
MCM 4	$\hat{U}_{t_2}$, $\hat{p}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	15.03	6.24
MCM 5	$\hat{U}_{t_2}$, $\hat{U}_{t_0}$, P_{t_0}	$\hat{P}_{t_2}$	14.80	7.67
MET-PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	16.88	-5.30
MET-PCF + PM	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.32	16.91

Table 4.5: Overview of the performance of the persistent model (PM), the power curve function model (PCF), the Markov chain models (MCM), the meteorological data-customized power curve function model (MET-PCF), and the combined MET-PCF and persistent model (MET-PCF + PM) at Raggovidda, in terms of the NRMSE and the improvement parameter I.

Model	**Input parameters**	**Output parameters**	**NRMSE (%)**	**Improvement parameter *I* (%)**
Raggovidda				
PM 2	P_{t_0}	$\hat{P}_{t_2}$	15.49	
PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	21.32	-37.64
MCM 1	$\hat{U}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	14.12	8.84
MCM 2	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	14.06	9.23
MCM 3	$\hat{U}_{t_2}$, $\hat{T}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.82	10.78
MCM 4	$\hat{U}_{t_2}$, $\hat{p}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	13.63	12.01
MCM 5	$\hat{U}_{t_2}$, $\hat{U}_{t_0}$, P_{t_0}	$\hat{P}_{t_2}$	13.89	10.33
MET-PCF	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$	$\hat{P}_{t_2}$	16.54	-6.78
MET-PCF + PM	$\hat{U}_{t_2}$, $\hat{D}_{t_2}$, P_{t_0}	$\hat{P}_{t_2}$	12.50	19.30

Table 4.6: Average improvement parameter for each Markov chain model of all five wind parks, and the average amount of predictions where the modified persistent model is applied.

Model	*Average improvement parameter I (%)*	*Average amount of times the modified persistent model is applied (%)*
MCM 1	4.02	8.26
MCM 2	6.96	57.17
MCM 3	6.54	45.08
MCM 4	6.87	41.37
MCM 5	7.09	32.79

The Markov chain models perform rather well at all five sites evaluated, by means of attaining lower NRMSE than the persistent model. Comparison between the models shows that MCM 1 obtained the highest NRMSE in the evaluated time period. The arithmetic mean of all improvement parameters obtained for each Markov chain model at all sites, and the average amount of predictions where the modified persistent model is applied, is shown in Table 4.6. MCM 1 is the only Markov chain model with only two input parameters, whereas the remaining 4 models have three. This may indicate that the particularity of having three instead of two input parameters, has a favourable impact on the prediction performance. However, the resulting increase in transition matrix size, leads to substantially higher frequencies of the Markov chain models applying the modified persistent model. This is demonstrated in Table 4.6. The best overall performing Markov chain model, MCM 5, was also the model where the modified persistent model was applied fewest times overall, out of the models with three input parameters. In general, the differences in average improvement with respect to the persistent model, of all Markov chain models with three input parameters, is quite small.

Comparison between prediction performances of the different sites shows that Fakken has the highest NRMSE values, whereas the lowest NRMSE values are obtained at Havøygavlen wind park. This may indicate a more rapid change of wind pattern at Fakken, compared to Havøygavlen. The highest and lowest NRMSE attained by the persistent model also occurred at Fakken wind park and Havøygavlen wind park, respectively. This supports the aforementioned claim, but also suggests an evident correlation between the Markov chain models and the persistent model.

The best performing Markov chain model, in terms of the improvement parameter I at a single site, was the MCM 4 at Raggovidda wind park, with 12.01%. This model applied the persistent model 630 times out of 1440 predictions, corresponding to 43.75%, at Raggovidda. Removing all predictions in which the modified persistent model was used, and calculating the NRMSE of the remaining 810 predictions, yields 11.59 % using solely the Markov chain model itself.

Applying the persistent model on the same 810 hours of output power, produces a NRMSE value of 12.08 %. These results provide some clarity regarding how this specific Markov chain model performs as an independent stand-alone model, at Raggovidda wind park. However, the scale-dependency of the NRMSE, leaves comparisons across data sets that have different scales invalid. As the number of predictions made in the given time period that is being evaluated are almost halved, it would therefore not be applicable to compare this result with the other results in Table 4.5.

In terms of the improvement parameter *I*, the weakest performance for the Markov chain models is observed at Kjøllefjord wind park. The amount of predictions where the modified persistent model is applied at Kjøllefjord, exhibit a negligible deviation compared to the average amount of all wind parks. However, inspecting the actual power output of all wind parks through the whole prediction period, Kjøllefjord wind park holds the longest continuous periods with zero power output. During these periods it is very likely that the Markov chain models frequently apply the modified persistent model, as the given combinations of input parameters rarely occurs in the training period. Given that the predicted wind speed state indicates power production, the Markov chain model will produce a non-zero output, while the persistent model produces zero power output. Therefore, the gain of modifying the applied persistent model in the Markov chain models, is lowest at Kjøllefjord wind park. Although the modification of the persistent model applied in the Markov chain models produced lower NRMSE for all models and wind parks, this observation exposes an evident weakness to the simplistic modification matrix, illustrated by Figure 3.9.

The overall improvement with respect to the persistent model, is highest at Raggovidda wind park. The relatively high NRMSE of the persistent model demonstrates a high wind variability at the site, along with the persistent model's inability to model future substantial shifts to the meteorological parameters, which the Markov chain models are able to capture. Thus improving the NRMSE considerably when there are major changes in wind behaviour.

An interesting observation at Nygårdsfjellet, is the high performance of MCM 3, with the forecasted temperature as the stand-out input parameter. In terms of average improvement for all five wind parks, MCM 3 is the weakest performing Markov chain model with three input parameters, given by Table 4.6. However, the increase in mean wind speed in the winter months at Nygårdsfjellet, illustrated by Figure 4.8, may indicate that the inclusion of temperature enables the model to identify these seasonal differences, and therefore increase prediction performance. In contrast, the inclusion of wind direction as stand-out input parameter by MCM 2, has a significantly smaller effect due to the bi-directional wind flow at the site.

Comparing the performances of the different short-term wind power prediction models developed in this study, it is very interesting that the linear combination of the MET-PCF model and the persistent model significantly outperforms all Markov chain models at all five wind parks. Especially since all Markov chain models implement the output power over the last hour, P_{t_0}. Frequent testing and error analysis during the progress of developing the Markov chain models, has led to several modifications with the intention of minimizing different weaknesses to the model. However, the amount of times the modified persistent model is applied, given by Table 4.6, demonstrates the sparsity of the transition matrices and suggests further shortcomings to the models. The obvious solution to this concern, would be to increase the amount of training data. This would not eliminate the issue, but minimize sparsity and provide more profound probabilistic forecasts.

Another approach is to fill the vacant rows in the transition matrices with a row corresponding to an adjacent state for one of the input parameters. What input parameter to be altered in order to fill the empty or insufficient rows, and whether one should apply the state above or beneath, should be tested and chosen on the basis of optimizing the model output NRMSE. In doing so, all rows in the transition matrices would be sufficiently filled and able to provide predicted power outputs, without applying the modified persistent model.

For the Markov chain models, there will always be a loss in accuracy by converting the data into corresponding states, and when producing the predicted power production from the corresponding predicted power states. This accuracy loss is not present in the MET-PCF + PM model. Hence, another objective regarding the use of Markov chains in wind power prediction models, is to optimize the trade-off between transition matrix sparsity and accuracy loss by conversion, regarding parameter state selection and distribution. To optimize prediction performance, one would have to identify the ideal number of states and their distribution, for each parameter, at each specific location.

The implementation of the persistent model is quite different in the Markov chain models, and in the MET-PCF + PM model. The importance of the persistent model in Markov chain models, is demonstrated by the diagonal shape in Figure 3.8. Where the first 12 rows indicates the 12 wind speed states as the output power state over the last hour was 1. The next 12 rows indicates the 12 wind speed states as the output power state over the last hour was 2, and so on. Whereas for the MET-PCF + PM model, the linear combination of the persistent model with the MET-PCF model, produced a substantial decrease in NRMSE for all wind parks. However, the weighting of the persistent model is optimized in the MET-PCF + PM model, by calculating the NRMSE for all values of the scaling constant a, as described in section 4.2. The optimized scaling constant for each wind park, is restricted to the available dataset. Therefore, the Markov chain models are conceivably more robust when applied to new independent data.

4.3.1 Turbine-wise prediction

For the turbine-wise power prediction, the Markov chain model, MCM 1, is tested for all 15 Nordex N80 2.5 MW wind turbines at Havøygavlen wind park. This model includes the input parameters P_{t_0} and $\widehat{U}_{t_2}$, which is the power output state over the last hour and the predicted wind speed state in two hours, respectively. The leave-one-out method described in section 3.3.4, provides both training data and verification data, and the output state parameter is chosen to be the predicted power in two hours ($\hat{P}_{t_2}$). A persistent model (PM) with m=2, has been run for the whole year for each turbine and for all 15 turbines as a whole, according to Equation (3.3). For each turbine and for all 15 turbines as a whole, the normalized root mean square error (NRMSE) for the whole year is calculated. The improvement parameter *I* is calculated according to Equation (2.25), with the persistent model as reference and NRMSE as the evaluation criterion. The results are shown in Table 4.7.

The differences between the turbines in terms of the NRMSE, are in general quite small, but with a few exceptions. The lowest NMRSE values obtained by the Markov chain model, is observed for turbines HAVWTG005 and HAVWTG009, with 13.40% and 12.90%, respectively. The two turbines are located in the middle of each of the separate parallel lines in which the turbines are arranged, shown in Figure 3.2. This may indicate a slower change of wind pattern for the turbines in the middle. The middle turbines may experience wind speed deficit due to the wake effects imposed by the adjacent turbines. The same effect might possibly shelter the turbines in the middle from the largest fluctuations in power production, by the reduced range in power output, induced by the power production deficit. The biggest improvement with respect to the persistent model, for all 15 turbines, is observed for the southernmost turbine, HAVWTG005. Whereas for turbine HAVWTG009, the Markov chain model is outperformed by the persistent model.

Table 4.7: Overview of the performance of the Markov chain model 1 and the persistent model, for each of the 15 Nordex N80 wind turbines at Havøygavlen, and all 15 turbines as a whole, in terms of the NRMSE.

Wind turbine	**NRMSE MCM 1 (%)**	**NRMSE PM (%)**	**Improvement parameter *I***
HAVWTG001	15.02	15.54	3.35
HAVWTG002	15.34	15.63	1.86
HAVWTG003	14.70	15.02	2.13
HAVWTG005	13.40	14.03	4.49
HAVWTG006	15.13	15.36	1.50
HAVWTG007	15.20	15.51	2.00
HAVWTG008	13.54	14.07	3.77
HAVWTG009	12.90	12.85	-0.39
HAVWTG010	15.14	15.41	1.75
HAVWTG011	15.41	15.77	2.28
HAVWTG012	16.53	16.56	0.18
HAVWTG013	15.03	14.76	-1.83
HAVWTG014	14.65	15.11	3.04
HAVWTG015	14.94	14.97	0.20
HAVWTG016	17.75	17.81	0.34
All wind turbines (15 x Nordex N80)	**10.60**	**11.53**	**8.07**

By observing the wind rose for Havøygavlen wind park, in Figure 4.1, the most common wind direction, south-southwestern (SSW), is incident perpendicular to the two parallel lines of turbines. Hence, the northernmost turbine, HAVWTG009, may experience more turbulent wind originating from this direction, due to wake effects. And consequently, the performance of the Markov chain model is reduced.

The highest NMRSE values is observed for turbines HAVWTG012 and HAVWTG016, with 16.53% and 17.75%, respectively. Both these turbines are located on the eastern far end of Havøya, as displayed in Figure 3.2. This may indicate a more rapid change of wind pattern at the edge of the island. The steep cliffs down to the ocean, surrounding these two turbines, may produce a turbulent wind pattern, causing challenging prediction conditions.

Comparison between the NRMSE of the Markov chain model and the persistent model for all 15 wind turbines as a whole, shows an improvement of 8.07% in favor of the Markov chain model. Comparing the turbine-wise prediction results with the performance obtained for the whole wind park, given in Table 4.1. The NRMSE values of the Markov chain model and the persistent model has decreased from 13.35% to 10.60% and from 13.62% to 11.53%, respectively. Given that the models are trained and tested using different power data, the substantially lower NRMSE for the persistent model indicates a smaller variability in the on-site measured power output, used in the turbine-wise prediction. The removal of HAVWTG004 may also cause further dissimilarity between the two approaches. However, the MCM 1 only managed to attain an improvement parameter of 1.98% for the whole park, given in Table 4.1, opposed to improvement of 8.07% by the turbine-wise prediction. The substantial gain in improvement by the turbine-wise approach, might be explained by its ability to determine if one or several specific turbines are not producing normally, rather than identifying such events as a slight overall production deficit. Furthermore, random errors of each turbine may equalize each other, so that when combined and normalized by their total capacity, the NRMSE is considerably reduced.

5 Conclusion

A statistical analysis was performed for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda wind park, based on meteorological forecast data provided by the Norwegian Meteorological Institute, for the time period 1. January 2017 – 31. December 2017. The wind regime during this time period, along with seasonal variations, were presented in the form of wind roses. The results displayed larger occurrences of higher wind speeds during the winter half-year compared to the summer half-year, for all five wind parks. At Havøygavlen the analysis demonstrated a rather even wind distribution through the whole year, with the occurrence of high wind speeds largely originating from the western semicircle. The most common wind direction was observed from the south-southwest, seemingly corresponding to the strait between the mainland and Rolvsøya in the west. At Kjøllefjord the analysis demonstrated a dominant wind direction from the south-southwest, with an even distribution of the remaining wind directions. The dominant wind direction were found be an effect brought about by the large north-south going fjord, Laksefjorden, running along the west side of the peninsula. At Nygårdsfjellet the analysis established the existence of a bi-directional wind regime, with winds originating primarily from the east and the west, corresponding to the west-east going valley in which the wind park is situated. High wind speeds were found to be more frequently originating from the east. The most common wind directions at Fakken wind park, were found to be originating from the south-southwest and the southeast, with the latter being subject to the largest occurrence of high wind speeds through the year. The dominant wind directions were observed to be the likely result of the two large north-south going fjords in the south, Ullsfjorden and Lyngen. The analysis regarding seasonal variations at Fakken wind park, demonstrated a high occurrence of lower wind speeds originating from the open sea in the north and north-northeast during the summer half-year. While the winter half-year was fairly dominated by south-southwestern winds.

At Raggovidda the analysis established the existence of frequently occurring high winds throughout the year. The most common wind directions were found to be originating from between and including the directional segments of northwest and south-southwest, with the largest occurrence of high wind speeds observed from the southwest. The easterly dominant wind directions were observed to be the likely result of the large north-south going Tanafjorden in the west. The analysis regarding seasonal variations found a large occurrence of higher wind speeds originating from the western semicircle during the winter half-year, compared to the more even distribution of lower wind speeds during the summer half-year.

A meteorological data-customized power curve function was developed and tested at each of the five wind parks evaluated in this study, using the on-site power output and forecasted wind speed and direction. The performance was measured in terms of the NRMSE, and compared with a persistent model by an improvement parameter, and a direction-specific power curve function, for each wind park. The model was found to be outperformed by the persistent model at all five wind parks, with improvement parameter values of -7.56%, -31.53%, -16.02%, -5.30% and -6.78%, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively. A linear combination of the meteorological data-customized power curve function model and the persistent model, optimized by a scaling constant, was developed and similarly tested. The linear combination of these models were found to increase the performance substantially at all five wind parks, improving on the persistent model by 16.08%, 11.33%, 17.38%, 16.91% and 19.30%, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively, in terms of the improvement parameter.

Five Markov chain models were trained and tested using different sets of input parameters, such as wind speed, wind direction, temperature, surface air pressure and power output. Their performances were measured in terms of the NRMSE, and compared with that of a persistent model, by an improvement parameter. All Markov chain models were found to have lower NRMSE than the persistent model, for all five wind parks.

At Havøygavlen the best performing Markov chain model had a NRMSE of 12.74%, compared to the 13.62% of the persistent model. This model uses the forecasted wind speed in two hours and the current hour, together with the output power over the last hour as input. At Kjøllefjord and Fakken, the best performing Markov chain model had a NRMSE of respectively 13.52% and 15.51%, compared to the 14.21% and 16.03% of the persistent model, respectively. This model uses the forecasted wind speed and wind direction in two hours, together with the output power over the last hour as input. At Nygårdsfjellet the best performing Markov chain model had a NRMSE of 13.93%, compared to the 15.36% of the persistent model. This model uses the forecasted wind speed and temperature in two hours, together with the output power over the last hour as input. At Raggovidda the best performing Markov chain model had a NRMSE of 13.63%, compared to the 15.49% of the persistent model. This model uses the forecasted wind speed and surface pressure in two hours, together with the output power over the last hour as input. The improvement parameter with reference to the persistent model, of the best performing Markov chain model at each wind park, was found to be 6.17%, 4.86%, 9.31%, 9.48% and 12.01%, for Havøygavlen, Kjøllefjord, Nygårdsfjellet, Fakken and Raggovidda, respectively.

A turbine-wise prediction was performed for 15 turbines at Havøygavlen wind park, by the application of a Markov chain model using the forecasted wind speed in two hours, and the power output over the last hour as input. The performance was measured in terms of the NRMSE, and compared with that of a persistent model, for each turbine and for all 15 turbines as a whole. Out of the two parallel lines of turbines, the northernmost was found to exhibit, in general, lower improvement by the use of the Markov chain model, compared to the persistent model. This was identified as a result of the wake effects imposed by the dominant wind direction. Additionally, the Markov chain model was found to achieve the highest NRMSE for the two turbines located on the eastern far end of Havøya, indicating a more rapid change of wind pattern at the edge of the island. For all 15 wind turbines combined, the NRMSE of the Markov chain model and

the persistent model were found to be 10.60% and 11.53%, respectively. Thus yielding an improvement parameter value of 8.07%. The substantial gain in improvement by the turbine-wise approach, compared to the improvement parameter of 1.98% using the same Markov chain model for the whole park, was identified as a result of its ability to distinguish between turbine-wise downtime and a slight overall production deficit.

The discoveries made in this study emphasizes the potential of short-term wind power prediction models based on different approaches in time series analysis. Specific measures of improving the performances of the different models developed in the study, has been identified. The findings has also demonstrated the advantage of combining different prediction models to improve performance.

5.1 Further research

Several measures of improving the performances of the different prediction models developed in this study has been discussed. The performance of the rather simplistic linear combination of the meteorological data-customized power curve function model and the persistent model, could be improved by only defining the power curve functions within the respective cut-in and cut-out wind speeds, possibly with an additional margin of forecast error. But most importantly, by including more data for the development of power curve functions. The high dependency to the persistent model, suggests that the performance will decrease rapidly as the lead time increases. Hence, the weighting of the stand-alone meteorological data-customized power curve function model, will be increasingly more significant.

For the Markov chain models, the main objective would be to fill the vacant rows in the transition matrices. The inclusion of more training data would minimize the sparsity and provide more profound probabilistic forecasts. However, filling the vacant rows in the transition matrices with a row corresponding to an adjacent

state for one of the input parameters, would provide sufficiently filled transition matrices, enabling the model to provide stand-alone predicted power outputs. Optimization of the trade-off between transition matrix sparsity and accuracy loss by conversion, regarding parameter state selection and distribution, constitutes another potential project that can benefit the prediction models based on Markov chains. To optimize prediction performance, one would have to identify the ideal number of states and their distribution, for each parameter, for the specific location.

It would also be very interesting to develop a wind power prediction model configured for on-line operation, using an automated model algorithm which collects the desired meteorological forecast data, and logs the relevant power outputs that are to be used as input for future predictions. Thus producing a constant power production output with desired lead time, which can be used to provide improved hourly delivery commitments to the power grid. Additionally, in the case of Markov chains it would be beneficial to implement automatic optimization of parameter state selection and distribution. Preferably through a machine learning algorithm, tuning the model according to the characteristics of the specific site. The numerical weather prediction system MEPS, providing the meteorological forecast data applied in this study, is run four times daily at 00, 06, 12, 18UTC. Given that the operational power prediction model would have to constantly apply the last available forecast, the model would have to use forecasts of continually increasing lead time, until the next forecast is processed and available, to provide hourly delivery commitments to the power grid.

Bibliography

Andrews, J., & Jelley, N. (2013). *Energy science: principles, technologies, and impacts.* Oxford university press.

Anton, H., & Rorres, C. (2015). *Elementary Linear Algebra* (11. ed.). Wiley.

Barthelmie, R. J. (2007). Modelling and measurements of wakes in large wind farms. *Journal of Physics: Conference Series*(75).

Bilal, M. (2016). *Wind Energy at Nygårdsfjellet - Norway.* A dissertation for the degree of Philosophiae Doctor, University of Tromsø, Department of Physics and Technology, Tromsø.

Birkelund, Y., Alessandrini, S., Byrkjedal, Ø., & Monache, L. D. (2018). Wind power prediction in complex terrain using analog ensembles. *Journal of Physics: Conference Series, 1102*, 012008. doi:10.1088/1742-6596/1102/1/012008

Bremnes, J. B., & Giebel, G. (2017). *Do regional weather models contribute to better wind power forecasts?* The Norwegian Meteorological Institute.

Brokish, K., & Kirtley, J. (2009). Pitfalls of modeling wind power using Markov chains. *2009 IEEE/PES Power Systems Conference and Exposition*, (pp. 1-6).

Byrkjedal, Ø., & Åkervik, E. (2009). *Wind map of norway (in Norwegian) Report 9.* The Norwegian Water Resources and Energy Directorate.

Carpinone, A., Giorgio, M., Langella, R., & Testa, A. (2015). Markov chain modeling for very-short-term wind power forecasting. *Electric Power Systems Research 122*, 152–158.

Chai, T., & Draxler, R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *eoscientific Model Development*(7(3)), pp. 1247-1250.

Cheng, H., Garrick, D., & Fernando, R. (2017, May 1). Efficient strategies for leave-one-out cross validation for genomic best linear unbiased prediction. *Journal of Animal Science and Biotechnology*(Vol.8), pp. 1-5.

Foley, A., Leahy, P., & Mckeogh, E. (2010, June). Wind power forecasting & prediction methods. *9th Conference on Environment and Electrical Engineering, EEEIC 2010*, pp. 61-64.

Fowler, P. C. (2012). *Understanding the Strengths and Weaknesses of a New-generation Numerical Weather Prediction Model for Application to Short-term Wind Energy Prediction.* Lubbock: Texas Tech University.

Fox, J., & Weisberg, S. (2018). *An R Companion to Applied Regression.* SAGE Publications Inc.

Früh, W.-G. (2012, March). Evaluation of simple wind power forecasting methods applied to a long-term wind record from Scotland. *International Conference on Renewable Energies and Power Quality*.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. (2011). The state-of-the-art in short-term prediction of wind power: A literature overview. *ANEMOS. plus*.

Heinermann, J., & Kramer, O. (2016, April). Machine learning ensembles for wind power prediction. *Renewable Energy*(89), pp. 671-679.

Hocaoglu, F. O., Gerek, O. N., & Kurban, M. (2008, 5). The Effect of Markov Chain State Size for Synthetic Wind Speed Generation. *Proc. 10th Int. Conf. Probablistic Methods Applied to Power Systems*, (pp. 1-4).

Hodge, B., & Milligan, M. (2011, July). Wind power forecasting error distributions over multiple timescales. *IEEE Power and Energy Society General Meeting*, pp. 1-8.

Holttinen, H., Miettinen, J., & Sillanpää, S. (2013). *Wind power forecasting accuracy and uncertainty in Finland.* Espoo: VTT Technology 95.

Hyndman, R. J., & Koehler, A. B. (2006, October). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*(Vol.22), pp. 679-688.

Jacobsen, M. (2014). *Short-term wind power prediction based on Markov chain and numerical weather prediction models: A case study of Fakken wind farm.* Master's thesis, UiT The Arctic University of Norway.

Jin, J. (2017). *Wind Resource Assessment in Cold Climate Regions.* Master's thesis, UIT The Arctic University of Norway, Institute of Industrial Engineering, Tromsø.

Kondo, K., Tsuchiya, M., & Sanada, S. (2002). Evaluation of Effect of Micro-topography on Design Wind Velocity. *Journal of Wind Engineering & Industrial Aerodynamics*(90), pp. 1707–1718.

Køltzow, M. (2017, September 25). *MetCoOp Ensemble Prediction System (MEPS).* Retrieved from https://drive.google.com/file/d/0B-SaEtrDE91WWEJoNkJiUm5TNzg/view

Madsen, H. (2008). *Time Series Analysis (Vol. V. 72, Chapman & Hall/CRC texts in statistical science series).* Boca Raton: Chapman & Hall/CRC.

Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H. A., & Nielsen, T. S. (2005). A Protocol for Standardazing the performance. *Wind Engineering*, pp. 475-489.

Manwell, J., McGowan, J., & Rogers, A. (2009). *Wind energy explained : Theory, design and application* (2. ed.). Chichester, UK: Wiley.

Markov, A. A. (1907). Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain. *The Notes of the Imperial Academy of Sciences of St. Petersburg VIII Series, Physio-Mathematical College, XXII/9*.

Milligan, M., Schwartz, M., & Wan, Y.-h. (2003). *Statistical wind power forecasting models: Results for US wind farms.* Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States).

NASA. (n.d.). The Spherical Shape of the Earth. (M. Srinivasan, Ed.) Retrieved from https://sealevel.jpl.nasa.gov/overview/overviewclimate/overviewclimateshapeofearth/

NCAR. (n.d.). *Weather Research and Forecasting Model*. Retrieved from National Center for Atmospheric Research (NCAR): https://www.mmm.ucar.edu/weather-research-and-forecasting-model

Ragheb, M. (2016). *Orography and Wind Turbine Siting.* University of Illinois, Urbana, Illinois.

Ren, G., Liu, J., Wan, J., Guo, Y., & Yu, D. (2017, 10). Overview of wind power intermittency: Impacts, measurements, and mitigation solutions. *Applied Energy, 204*, 47-65. doi:10.1016/j.apenergy.2017.06.098

Rosen, A., & Sheinman, Y. (1996, January). The power fluctuations of a wind turbine. *Journal of Wind Engineering and Industrial Aerodynamics*, pp. 51-68.

Ross, S. M. (2014). *Introduction to probability models* (11th ed.). Academic press.

Shamshad, A., Bawadi, M. A., Hussin, W. M., Majid, T. A., & Sanusi, S. A. (2005). First and second order Markov chain models for synthetic generation of wind speed time series. *Energy, 30*, 693-708.

Shokrzadeh, S., Jozani, M. J., & Bibeau, E. (2014). Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods. *IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, VOL. 5*, 1262 - 1269.

Shpilrain, E. E. (2009). *Renewable Energy Sources Charged with Energy from the Sun and Originated from Earth-moon Interactions.* Eolss Publishers Company Limited.

The International Energy Agency. (2013). *Technology Roadmap - Wind Energy.* IEA.

The Norwegian Water Resources and Energy Directorate. (2019). *Analysis sheet for area 34 (in Norwegian).* NVE.

The Norwegian Water Resources and Energy Directorate. (2019, March 25). *Developed wind power stations (in Norwegian)*. Retrieved from https://www.nve.no/energiforsyning/vindkraft/utbygde-vindkraftverk/

The Norwegian Water Resources and Energy Directorate. (2019). *Proposal for National Framework for Wind Power (in Norwegian).* Oslo: NVE.

The Norwegian Water Resources and Energy Directorate. (2019). *Thematic map*. Retrieved from Map of all wind power parks in Norway: https://temakart.nve.no/link/?link=vindkraftverk

Weir, D. E. (2014). *Wind power - Production in 2013 (in Norwegian).* The Norwegian Water Resources and Energy Directorate.

Yoder, M., Hering, A. S., Navidi, W. C., & Larson, K. (2013). Short-term forecasting of categorical changes in wind power with Markov chain models. *WIND ENERGY, 17*, 1425–1439. doi:10.1002/we.1641

Yussuff, A. (2017). *Farm level power curves for wind integration studies considering turbine wake and complex terrain effects.* Master's thesis, The University of Edinburgh, Institute for Energy Systems, Edinburgh.

Østreng, E. (2014). *Analysis of Meteorological Data from the Kjøllefjord Wind Farm, and Characterization of Stability and Boundary Layer Height from WRF*. Master's thesis, Norwegian University of Life Sciences (NMBU), Ås.