Improving and enhancing NWP based wind power forecasts under Norwegian conditions

Pål Preede Revheim

Improving and enhancing NWP based wind power forecasts under Norwegian conditions

Doctoral dissertation for the degree *Philosophiae Doctor (PhD)* at the Faculty of Engineering and Science with specialization in Renewable Energy

University of Agder

Faculty of Engineering and Science

2015

Doctoral Dissertations at the University of Agder nr. 110

ISSN: 1504-9272
ISBN: 978-82-7117-796-6

Printed by the Printing Office, University of Agder
Kristiansand

Acknowledgements

When I started the work with this thesis three years ago I was completely new to the fields of energy meteorology and wind power forecasting. Today I know a little more, but most of all I know how much more I would like to know. The three years have passed very fast. I hope that I in the years to come will have the chance to dig deeper into some of the topics of this thesis, and perhaps even more to learn about a wider range of issues related to grid integration and the economy of renewable energy. There are many people who have been important in both for my work and in my life these three years, and I will use this opportunity to thank some of them.

First of all I would like to thank my main supervisor Professor Hans-Georg Beyer. He has let me benefit from his years of experience from wind power forecasting, while at the same time giving me enough time and space to try and fail and make my own experiences. The "world of science" has at times been a strange place, and I am very grateful to have had an experienced and steady guide who has made sure that I did not come completely off course.

I also want to thank Professor Mohan Lal Kolhe, who has been my co-supervisor. Even though he arrived at UiA later than expected, and as a result of that never got deeply involved in my work, it has been comforting to know that he would have been there for me if needed.

Further, I want to thank the Norwegian Meteorological Institute, and John Bjørnar Bremnes in particular, for providing me with forecast- and measurement data. This thesis would not have been possible without their helpfulness.

During my work I got the chance to spend some time at Universidade Federal de Santa Catarina in Florianopolis, Brazil. Thanks to Manfred-Georg Kratzenberg, Professor Ricardo Rüther and Professor Hans Helmut Zürn for hosting me. It was truly a memorable experience.

Fortunately, being a PhD student is not only about the work. I want to thank the PhD students at the Faculty of Engineering and Science for providing a nice social environment. I have met many great people, and will in particular like to thank Georgi, Charly, Nazmin, Gunstein, Abozar, Muhammad and Deepak who I have had the

pleasure of sharing office with the whole or parts of the period. Thanks for interesting discussions about science and life and the everyday problems of a PhD student.

Last I would like to thank my parents, my girlfriend Christine, my son Elias. Thanks to my parents for keeping Christine and Elias with company while I have been working long days and long weeks. Thanks to Christine for supporting and helping me, and making it possible for me to dedicate all my time to the thesis the last intense months. The importance of her contribution can hardly be overestimated. Thanks to Elias for staying happy and healthy his first half year. I'm looking forward to getting to know you better over the coming years.

Pål Preede Revheim
Grimstad, Norway
January 2015

Abstract

This thesis studies methods for improving and enhancing NWP based wind power forecasts for cases from the Western coast of Norway. This is an area with excellent wind conditions, but where the installed wind power capacity at present is limited. The area is characterized by a rugged coastline and complex terrain, which have earlier been shown to lead to high wind power forecast errors. The overall aim of the thesis is to study how this kind of conditions influence wind power forecast errors and to investigate how wind power forecast models can be made more resistant to the challenges these conditions pose.

The data basis for the thesis consists of wind observations and NWP wind forecast for 43 sites, all covering the period from January 1st 2009 to December 17th 2011. The observations and forecasts are transformed into synthetic wind power forecasts and observations by the use of a logarithmic height profile and a generic power curve.

Different methods of reducing the wind power forecast errors of the single sites and groups of sites are tested. When applied to "unseen" forecasts (i.e. independent test data) the simpler models tend to out-compete more complicated models of the same kind. This is caused by a combination of noisy data and sparse data for certain wind speeds and wind directions leading to models being easily over-fitted. Still, using a regression model based on information on spatial and temporal dependencies of the forecast errors of groups of sites, a reduction in the group forecast error of 49 % is obtained.

The group point forecasts are expanded into probabilistic forecasts using the post-processing method Bayesian Model Averaging (BMA). It is shown how BMA can be used to produce probabilistic forecasts for the lumped power output of groups of sites from the point forecasts for the single-site group members and historical observations and forecasts from a training period. Some ideas for further development of the method for wind power forecasting are presented.

Last, the issue of wind power ramps – large sudden changes in the wind power production – is addressed. Using a simple method to forecast ramps from wind power forecasts it is found that the methods earlier used to reduce the wind power forecast

errors for groups of sites also lead to an increased predictability of wind power ramps. Ramp forecasts are also made using the classification method Random Forests. The method is found to have some very desirable properties, but the current implementation of the method also has some serious problems that need to be solved before the method is a good option for wind power ramp forecasting.

Contents

Chapter 1 – Introduction and background

This chapter explains the background and relevance of the thesis, introduces the thesis objectives, and gives an outline of its structure. Section 1.1 gives a brief review of the position and development of wind power in the Europe and Norway. It is shown how the installed wind power capacity has grown rapidly in recent years, and how political strategies and targets aim at continuing this growth. Some of the challenges the large volumes of wind power causes are pointed out, and it is explained how accurate and reliable wind power forecasts can help solve these challenges. Section 1.2 presents the basic concepts of wind power forecasting, including the explanation of the scheme used to transfer wind speed forecasts into wind power forecasts. Section 1.3 introduces the main approaches to wind power forecasting found in the literature and show how the approach varies with the purpose of the forecast. An overview of the state-of-the-art for different types of wind power forecasts is given. This will be subject to further discussions and constitute a basis for the thesis objectives. Section 1.4 introduces the concept of forecast evaluation and presents some of the most commonly used forecast evaluation methods and criteria. These will be frequently used and referred to throughout the thesis. Section 1.5 presents the objectives of the thesis. Some shortcomings of the state-of-the-art in wind power forecasting and opportunities for further developments are pointed out, and these serve as motivation for the objectives. It is explained how the objectives can contribute to more accurate and reliable wind power forecasts, and how these will be approached. Section 1.6 gives an outline of the structure of the thesis.

As a part of the work with this Thesis there have also been written several papers and posters, and held a number of oral presentations. Some of these underlie sections of this thesis, while others are complementary to the thesis e.g. showing alternative applications of methods. A list of these papers, posters and oral presentations are found in Annex 2.

1.1 Background

Recent years there has been an increased focus on wind as a source of energy. This has resulted in a significant increase in installed wind power production capacity and market share. With today's strong political focus on mitigating climate change and becoming less dependent on fossil fuels this tendency is unlikely to change. According to the European Wind Energy Association (EWEA) the total installed wind power in Europe rose from ~66 GW in 2008 (EWEA 2010) to ~121 GW in 2013 (EWEA 2014a). According to future projections this will further increase to somewhere in between 166 GW and 217 GW by 2020, corresponding to a market share of between 12.7 % and 16.9 % (EWEA 2014b). The development in Norway have had a slightly slower pace, with an increase in the installed wind power from 429 MW in 2008 (EWEA 2010) to 811 MW in 2014 (NVE 2014).

However, the increasing shares of wind power also raise new challenges. Unlike conventional power plants, the production of wind power is to a large extent dependent on factors beyond human control, most important the magnitude of the wind. As the wind is highly variable, so is also the wind power production. A secure electricity supply requires that the electricity production mirrors the consumption as exactly as possible at all points of the grid and at all times. The power consumption, and its variations over the day, is from experience well known. In conventional power systems it is therefore possible to adjust the production to be in balance with the consumption (M. Lange 2003). Large shares of wind energy – as large shares of solar energy in the grid - make this balancing more difficult, as it is no longer just the consumption that is subject to variability.

Another major difference between wind and most conventional sources of energy is that wind cannot be stored. The consequence of this is that if the power system is balanced through curtailing the wind power, energy equivalent to the curtailment is lost.

Numerous different strategies for dealing with the challenges of wind power have been considered, including:

- Strengthening national and European grids (Van Hulle *et al.* 2009; Eriksen *et al.* 2005)

- New solutions for demand side management (Ipakchi & Albuyeh 2009; Strbac 2008)
- Energy storage (Estanqueiro *et al.* 2012; Denholm *et al.* 2010; Carrasco *et al.* 2006; Barton & Infield 2004)
- The introduction of an ancillary service market (de Boer *et al*. 2013; de Boer *et al.* 2012)
- Wind power forecasting (Alessandrini *et al.* 2012; Eriksen *et al.* 2005; Watson *et al.* 1994)

Wind power forecasting is a collective term for methods to predict future wind power output. The aim of a wind power forecast is to provide an end-user with an estimate of the available wind power at a given time in the future. Depending on the forecast horizon, wind power forecasts are used to optimize the operation of wind turbines (milliseconds to seconds), allocate production reserves (seconds to hours), optimize the scheduling of conventional power plants (hours), optimize the value of the produced electricity (0-48 hours) and for planning of maintenance etc. (days) (Giebel *et al.* 2011; Monteiro *et al.* 2009). The focus in this thesis will be on wind power forecasting for forecast horizons from 1 hour to 24 hours.

1.2 The basics of wind power forecasting

Wind turbines convert the kinetic energy of the wind into electric energy. The energy content of the wind available to a wind turbine is dependent on the rotor diameter of the wind turbine, the air density and the wind speed through the relation:

$$P = \frac{1}{2}\rho\pi R^2 v^3, \tag{1.1}$$

where ρ is the air density, R is the rotor diameter and v is the wind speed (see e.g. Gipe 2004, p. 30). The energy that actually can be converted is however limited to a fraction of the energy contents. The upper limit for this fraction is the Betz's limit $C_p = 16/27$ (≈0.593) (see e.g. Gipe 2004, p. 56), but for real-world turbines the fraction of extracted energy is lower. E.g. does the turbine manufacturer Enercon report turbine-efficiencies reaching 0.5 (Enercon 2010) at optimal conditions and the turbines of other producers are expected to be in the same region. The efficiency typically varies

with wind speed causing the turbines power output to deviate from the cubic characteristic (Equation 1.1).

How the power output varies with the wind speed is described by a characteristic curve, normally referred to as the power curve. This curve summarizes the conversion efficiency of the turbine, characteristics of the turbines generator and power electronics, the built-in control systems etc. The exact shape of the power curve is dependent on the wind turbine model, but the main characteristics are common to all power curves. Figure 1.1 shows a stylized example of a power curve.

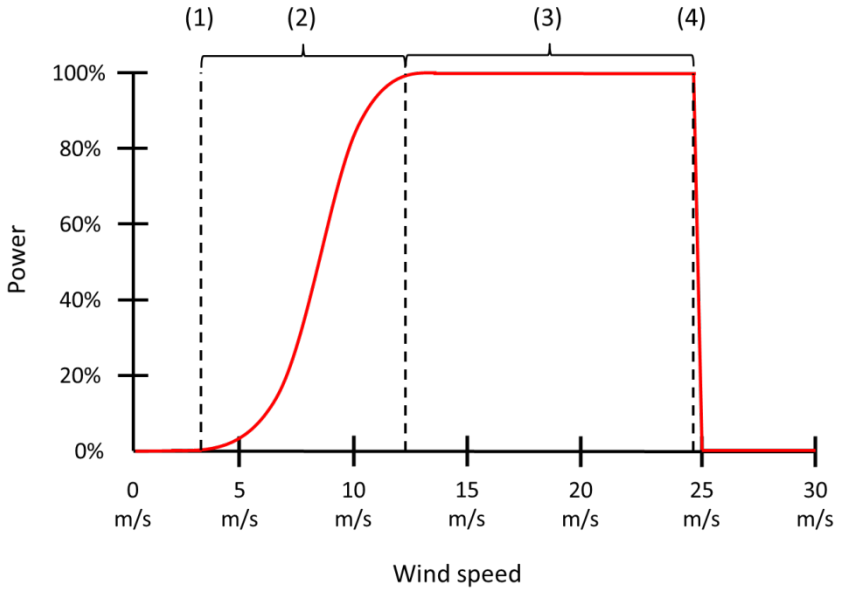

Figure 1.1 - Stylized example of a power curve describing the relation between wind speed and wind power for a wind turbine. (1) Shows the lower cut-in speed, (2) the steep part from cut-in to rated capacity, (3) the flat upper part and (4) the upper cut-out speed.

The power curve has four important features:

1. A lower cut-in speed. This is the lowest wind speed the turbine needs to transform the wind into energy. The value of the lower cut-in speed will to some extent depend on the wind conditions the turbine is intended for, but a typical value is around 3.5 m/s.
2. A steep, s-shaped part from zero production at the cut-in limit up to the turbine reaches its nominal capacity at a rated wind speed. The mean wind speed at the site of the turbine will be located somewhere in this part.

3. A flat upper part at rated capacity from the rated wind speed up to an upper cut-out limit.
4. An upper cut-out limit where the wind turbine is stopped for safety reasons. The upper cut-out limit is a part of the turbines control system and intended to protect the structure from large loads which may cause a shortened lifespan or failures.

The effect of the power curve is that the conversion from wind to power is highly non-linear. The change in power production a change in wind speed leads to will be dependent on the position in the curve; from small in the flat lower and upper parts, via large in the steep middle part to 100 % if the change in wind speed crosses the upper cut-out limit. Hence, the power curve inflates or deflates the prediction error of the wind speed according to its local derivative (M. Lange, 2003).

It should however be noted that the power curves supplied by the manufacturers are obtained from measurements – corrected for deviations from the standard air pressure - under standardized conditions referring to turbulence and wind speed variations with height above ground (see IEC standard 61400-12). In real applications the turbines will be subject to differing wind conditions, e.g. different turbulence level, and wind profiles affected by obstacles and variations in air pressure, disturbing the sharp, deterministic relationship between wind and power as shown in Figure 1.1. Wind power forecasting thus can be regarded as two processes; one in which the wind resources available to the wind turbines is determined, downscaling, and one in which the wind resources are transformed into power (Siebert 2008).

1.3 Wind power forecasting methods

Wind power forecast method can be formulated in the very general way:

$$\hat{p}_{t+k|t} = f(\theta_t), \tag{1.2}$$

where $\hat{p}$ is the wind power forecast made at time t for forecast horizon $t+k$ and θ_t is the set of explanatory variables (Pinson 2006). This leaves two challenges to the forecaster:

1. To determine the function f that gives the best description of the relation between $\hat{p}_{t+k|t}$ and θ_t.
2. To determine which variables that should be included in θ_t.

A large number of models for wind power forecasting have been proposed in the literature, of which many are also in operational use. It is not within the scope of this chapter to provide a complete overview of these. For a review of the history of wind power forecasting it is referred to Costa *et al.* 2008, for two extensive reviews of the state-of-the-art in wind power forecasting it is referred to Giebel *et al.* (2011) and Monteiro *et al.* (2009) and for a list of operational wind power forecast models to Foley *et al.* (2012) and Giebel *et al.* (2011). This chapter aims at giving an overview of the different types of forecast models, placing the theme of this thesis into a larger context and providing a basis for the objectives of the thesis that will be presented in Chapter 1.4. This entails that the presentation will focus on short term forecasting models, and that models built for very short forecast horizons (milliseconds to seconds) and for long forecast horizons (days) are omitted.

Methods for wind power forecasting can coarsely be classified by whether they include the output of some Numerical Weather Prediction (NWP) model or not (Siebert 2008). NWP models are, in short, sets of partial differential equations that from knowledge about the present state of the atmosphere allows computation of how meteorological variables, like wind speed and wind direction, will evolve with time (Pinson 2006). For an introduction to NWP models it is referred to Kalnay *et al.* (1998) and for information on operational NWP models – as applied to wind power - to Giebel *et al.* (2011) and Monteiro *et al.* (2009).

Methods not incorporating NWP inputs are mainly used for very short forecast horizons. This has two reasons; most important it has been shown that statistical approaches outperform NWP-based approaches for forecasts with lead-times up to a few hours (Giebel *et al.* 2011). Secondly NWP models with high temporal resolution are computationally expensive. Running NWP models with very frequent updates, as this application would require, thus would be either very expensive or not possible (Pinson 2012).

There is one noticeable exception from this, the persistence model. This is a very simple and naïve predictor which simply states that the future wind power production will be the same as the last known value, i.e.

$$\hat{p}_{t+k|t} = p_t, \tag{1.3}$$

where t is the time when the forecast is issued and k is the forecast horizon. The persistence model is very commonly used as a reference model, and may be hard to beat for forecast horizons up to a few hours (Monteiro *et al.* 2009). A few other persistence-like models have been proposed that also incorporate the long-term mean (Moerhlen 2004; Nielsen 1998), but none of these are anywhere near equally widespread as the basic persistence model. A list of other non-NWP wind power forecast models is found in Giebel *et al.* (2011).

Models including NWP outputs can be further divided into physical models and statistical models. The division is not very strict, as many of the physical models also utilize empirical data to model the power curve or correct the final output (model output statistics, MOS) and many of the statistical models use different physical principles for spatial refinement of the NWP output. These models are sometimes referred to as hybrid models. The division does however give a good description of the main principles behind the models. The physical models, as the name indicates, refine the NWP predicted wind speed into a power prediction through physical modelling of the site conditions and the conversion process. Statistical models, on the other hand, learn and model the relation between NWP predicted wind speed and measured power output from empirical data.

1.3.1 Physical models

In physical models the function f in Eq. 1.2 is based on the physical relation between $\hat{p}_{t+k|t}$ and θ_t and the explanatory variables in θ_t is limited to the NWP output and information about the power curve.

Physical models are normally split into two (or more) separate models organized as a hierarchy, where first the wind field at hub height around the wind farm is determined and thereafter the conversion from wind speed to wind power is computed. In the first step there are two distinct options; either is to combine modelling of the wind profile with the geostrophic drag law or to use computational fluid dynamics (CFD) (Pinson 2006). In the second step the wind speed is converted into power by the use of a theoretical or preferably an empirical power curve. It is also common practice to

account for systematic errors in the model (either in the NWP output or in the physical modelling) by Model Output Statistics (MOS) post-processing. This implies introducing a statistical element into the model thus, strictly speaking, making it a hybrid model.

Amongst the oldest and most well-established physical wind power forecast models are Prediktor developed at the Risø National Laboratory and Previento developed at the University of Oldenburg. Both models work in a similar manner utilizing the geostrophic drag-law derived wind together with a wind farm model accounting for terrain and shadowing effects. Previento also offers possibilities to predict both for single wind farms and for larger areas and to assess the risk of relying on the wind power forecast (Focken *et al.* 2002a). More information on Prediktor is found in Landberg (1998) and Landberg (1999) and on Previento in Beyer *et al.* (1999) and Focken *et al.* (2001).

CFD-based models are mainly used in connection with complex terrain, and are due to the computational costs more commonly used for resource assessment than forecasting. More information on the use of CFD models for wind power forecasting is found in Magnusson & Wern (2001), Magnusson (2002) and Rodrigues *et al.* (2007).

1.3.2 Statistical models

In statistical models the function *f* and exploratory variables θ_t in Eq. 1.2 is selected so that some chosen loss function, a function of the forecast errors, is minimized when the model is applied to a training dataset with known outcomes.

Statistical models describe the relationship between explanatory variables, like NWP wind speed forecasts, and measurements of the observed wind power production. In contrast to the physical models the statistical models make a direct link between the NWP outputs (or an enhancement of these) and the observed wind power, thus there is no need for a separate step in which the wind speed is converted into power. Statistical models usually consist of an autoregressive part and a meteorological part. The autoregressive part accounts for the persistence of the large-scale processes influencing the wind speed, while the meteorological part accounts for the NWP forecasts (Pinson 2006). Based on the found relationships various statistical models, e.g. regression or classification models, are used to construct wind power forecasts.

The fitting of the models consists of an optimization problem, where the goal is to minimize the forecast error when applied to a training dataset.

Statistical models can be divided into structural and black-box types of models. Structural models try to make use of the analyst's expertise on the phenomena of interest, and the focus will therefore be on known structures and relations. Black-box models on the other hand require little knowledge of the subject and are constructed from data in a mechanical way. Also here in-between models are common. These aim at exploiting the strengths of both the structural and the black-box models and are normally referred to as grey-box models. (Pinson 2006)

Wind Power Prediction Tool (WPPT), developed at the Technical University of Denmark, is an example of a well-established statistical forecasting model, and has been in operational use since 1994. WPPT is a conditional semi-parametric model, which integrates information from NWPs and uses semi-parametric estimates of wind direction dependent power curves to transform forecasts of wind speed and wind direction into wind power (Nielsen *et al.* 2001). Another example of an operational statistical wind power forecast tool is Sipreolico, developed at the University Carlos III of Madrid. Sipreolico consists of several prediction models ranging from simple reference models to conditional non-parametric models. The models are used depending on the availability of data, and the results combined in a weighted average depending on the models recent performance (Sánchez 2006; Usaola *et al.* 2002).

A number of statistical models incorporate offsite data as predictor variables. The idea behind this is basically to try to benefit from information about the upstream wind field to correct downstream forecasts. Examples of forecast models incorporating offsite data are found in Larson & Westrick (2006), Tastu *et al.* (2011), M. Lange *et al.* (2008) and Wessel *et al.* (2009).

A wide variety of grey- and black-box methods are also presented in the literature and some are included in operational models. Examples include Support Vector Machines (Frías *et al.* 2009; Moon *et al.* 2004), Neural Networks (Kariniotakis *et al.* 1996a; Li *et al.* 2001) and fuzzy-logics (Kariniotakis *et al.* 1996b; Wu & Dou 1995). These models are trained over large collections of data using specific algorithms, and map the relation between the input and the wind power output. The models are attractive in the way that they are flexible and easy to adapt to different learning structures, but as with all black-box/ grey-box models the results should be inspected carefully as they are

able to find all kinds of relationships between the input and the output, also those with a lack of causal connection.

1.3.3 Probabilistic models

Point forecasts have been, and still are, the main focus in wind power forecasting. These have the advantage that they are very easy to interpret, but they will always be subject to a forecast error. Traditional point forecasts will provide no information on this error. The opposite to point forecasts is probabilistic forecasts. In probabilistic forecasts the forecasted unit $\hat{p}_{t+k|t}$ (Eq. 1.2) is issued as a prediction interval, a set of quantiles, a theoretical probability distribution or the like instead of as a single value. The benefits of probabilistic forecasts over point forecasts has been shown for numerous applications, including wind power trading (Zugno *et al.* 2012; Botterud *et al.* 2012a; Pinson *et al.* 2007), economic dispatch and unit commitment (Botterud *et al.* 2012b; Zhang *et al.* 2013; Ahlstrom *et al.* 2013) and optimization of storage (Wu *et al.* 2014; Castronuovo *et al.* 2014; Duque *et al.* 2011). A thorough explanation of the basic principles in probabilistic wind power forecasting is found in Monteiro *et al.* (2009).

The division between physical models and statistical models are also relevant for probabilistic models, but also here the division is not strict and the in-betweens are common.

The physical models are focused on determining the uncertainty in the wind forecasts and how this transforms to the wind power forecasts. The most common physical approach is so-called NWP ensembles. NWP ensembles use a number of different model runs to make a set of possible outcomes. The different runs are made by running the same NWP model with variations in the inputs or settings (Lang & McKeogh 2009), by running different NWP models with the same inputs or by combinations of the two approaches (Giebel *et al.* 2005). The set of possible outcomes is thereafter used as input to a point forecast model to create a set of possible wind power outcomes. The sets of wind power outcomes can be used as a basis for non-parametric methods for finding e.g. quantiles or as a basis for parametric statistical analysis.

Statistical models construct probabilistic wind power forecast directly without modelling the uncertainty of the wind. A simple way of creating a probabilistic

forecast is to look at historical forecast errors and assume that future forecast errors will behave accordingly. More advanced variants of this method include determining the uncertainties for the different meteorological situations and determining the uncertainties for different levels of expected power generation. To avoid non-continuous descriptions of the uncertainty smoothing-techniques like fuzzy-logics (Pinson & Kariniotakis 2010) and quantile regression (Bremnes 2004; Bremnes 2006) have been used.

In later years, there have been developed models where also the temporal interdependence structure of the forecasts for different look-ahead times is considered. This means that instead of e.g. issuing forecasts for single hours scenarios for multiple hours are issued. These forecasts are often referred to as *trajectory forecasts* and have been shown to be beneficial for amongst other time-dependent decision problems. More information on trajectory forecasts is found in Pinson *et al.* (2009a) and Tastu *et al.* (2014).

1.3.4 Regional forecasts

Regional forecasts are forecasts of the lumped wind power generation of multiple wind farms in a region or a country. For many larger users of forecasts, like Transmission System Operators (TSOs) and energy traders, this is the quantity of most interest.

The simplest method to forecast for larger regions, to forecast the output of each wind farm and add the predictions, has proven little advantageous (Monteiro *et al.* 2009). The wind farms of an area will not be subject to the same wind speed at all times, and the errors in the wind speed forecasts from NWP models will be both spatially and temporally distributed. As a result the normalized forecast error of a region will be smaller than the normalized forecast error for a single wind farm (Focken *et al.* 2002b). How much smaller the forecast error will be depends on the spatial correlation between the wind farms.

Lange & Focken (2005) and Focken *et al.* (2002b) studied the impact spatial extension of regions and the number of wind farms in a region has on the regional forecast error. Their analysis showed that the magnitude of the error reduction obtained by spatial smoothing only weakly depended on the number of wind farms and were mainly determined by the size of the regions.

Several publications study the effect of the number and location of reference wind farms on the expected power output of a whole region. Siebert and Kariniotakis (2006) and Siebert (2008) evaluate the impact of input selection on the accuracy of regional forecasting, and propose three methods to determine the best reference wind farm combination based on amongst other clustering, information theory and regressive power curves (RPC). Other approaches for selecting or constructing reference wind farms are found in Focken *et al.* (2001), Marti *et al.* (2003), Ernst *et al.* (2001) and Nielsen *et al.* (2002).

1.3.5 Ramp forecasts

Wind power ramp forecast are forecasts tailored for predicting sudden, large changes in the wind power production (increases or decreases). These are events that typically will be rare, but with a potentially very large impact (Kamath 2010; Greaves *et al.* 2009). Due to their rareness wind power ramps are often hard to predict and the available training data is limited.

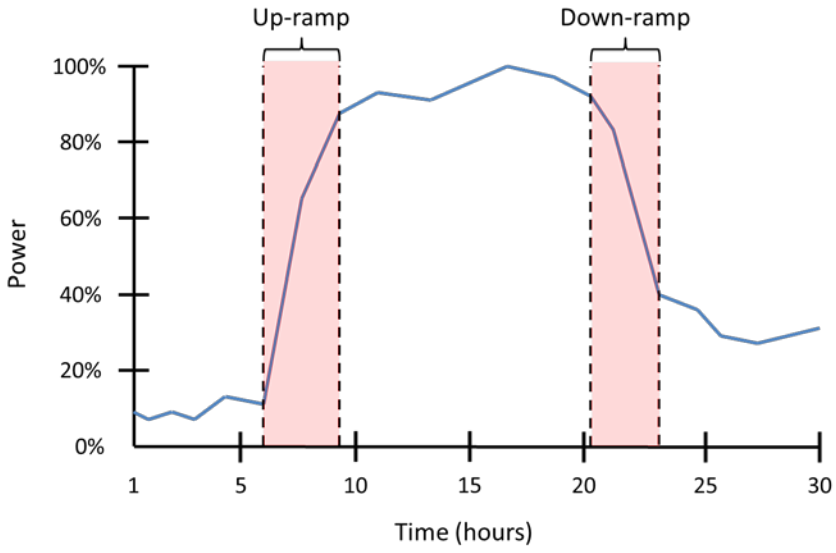

Figure 1.2 – Example of wind power up-ramp and down-ramp.

The amount of literature on ramp forecasting is still limited, but a few noteworthy approaches are found in the literature. Parkes *et al.* (2009) and Suzuki *et al.* (2012) use various data mining and pattern recognition techniques in order to use offsite data as a complement to NWP forecasts. The ramp forecasts are issued as discrete, deterministic

variables that indicate whether a ramp will take place or not within the next *k* hours. In Greaves *et al.* (2009) this method is also extended to including probability distributions for the timing of the ramps.

A somewhat different approach is used by Zheng and Kusiak (2009), where instead of forecasting ramps directly forecasts of ramp rates, the expected average change within some look-ahead time, are issued. The ramp rates are forecasted with the use of a multivariate time-series model based on historical measurements of six different variables and five different data-mining algorithms for model reduction.

Bossavy *et al.* (2010) propose two different methods for probabilistic ramp forecasts. Their first method defines two variables, the intensity of the nearest ramp and the time until the nearest ramp, and uses these as additional explanatory variables in probabilistic wind power forecasts. The method is shown to give more reliable forecasts for the higher quantiles without sacrificing sharpness. In their second model a similar approach is applied to NWP ensemble data to make forecasts of ramp timing and intensity for various look-ahead times up to $k = 70$ hours.

Zack *et al.* (2010) describes a system called ERCOT Large Ramp Alert System (ELRAS) where multiple sources of data are used to make probabilistic ramp rate forecast and hybrid deterministic-probabilistic ramp forecast. The system is based on a methodology that identifies different types of ramps and weather conditions and selects the forecast algorithm that over a training period has shown to have the better performance for the given conditions. Other examples where ramp probabilities are linked to specific weather systems are found in Cutler *et al.* (2007), Reikard (2010) and Couto *et al.* (2013)

A more detailed review of the ramp forecast models listed above and a few other wind power ramp forecast models are found in Ferreira *et al.* (2010) and Ouyang *et al.* (2013).

1.4 Forecast evaluation

Forecast evaluation is the process of verifying that a forecast does what it is intended to do. Evaluation of forecasts is essential for monitoring their accuracy, understanding their errors and improving the performance of the forecast model (Ebert *et al.* 2013).

Ideally wind power forecasts would be evaluated on the added value they contribute to the end user, such as added value (or reduced loss) for the power producers or reduction in energy imbalances for the TSOs, but in many cases this is practically very difficult or impossible. Other evaluation criteria that are easy to implement and highlight important aspects of the performance of forecasts are therefore commonly used.

Forecast evaluation is a large research area of its own. A wealth of different methods and parameters for forecast evaluation are found in the literature of which any are relevant to and have been applied to wind power forecasts. This section does not intend to give a complete overview of these. The aim here is limited to presenting some of the most commonly used methods and statistics for evaluation of wind power forecasts. For a thorough introductory text on forecast evaluation it is referred to Jolliffe & Stephenson (2003), for a thorough discussion on the use of forecast evaluation statistics to Mason (2008) and for the state-of-the-art in forecast evaluation to Casati *et al.* (2008) and Ebert *et al.* (2013) and references given therein.

1.4.1 Point forecasts

The error of a wind power forecast is defined as the deviance between the observed wind power and the forecasted wind power:

$$\varepsilon_{t+k|t} = p_{t+k} - \hat{p}_{t+k|t}, \tag{1.4}$$

where $\varepsilon_{t+k|t}$ is the forecast error at time $t+k$, p_{t+k} is the observed wind power at time $t+k$ and $\hat{p}_{t+k|t}$ is the wind power forecast issued at time t for time $t+k$. For easier comparison of the forecast errors of wind farms with different installed production capacity it is common practice to normalize the forecast errors to the rated installed capacity. This transforms the forecast errors to be in the interval [0, 1], which in many cases is a desirable property.

Some work has been done on standardizing the performance evaluation of wind power forecasts. Madsen *et al.* (2005) propose a protocol for evaluation of wind power point forecasts where bias, mean average error (MAE) and root mean squared error (RMSE) are recommended as a minimum. Bias is the average of the forecast errors, defined as:

$$BIAS_k = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_{t+k|t}, \quad (1.5)$$

where *N* is the number of observations in the evaluation period. A bias of 0 indicates that a forecast is perfectly precise, but it does not say anything about the accuracy of the forecast. Bias is also sometimes referred to as *systematic error*. MAE is the mean of the absolute values of the forecast errors, defined as:

$$MAE_k = \frac{1}{N}\sum_{t=1}^{N} |\varepsilon_{t+k|t}|, \quad (1.6)$$

where *N* is the number of observations in the evaluation period. MAE reflects both the systematic errors and the random errors. RMSE is the root of the mean of the squared forecast errors, defined as:

$$RMSE_k = \sqrt{\frac{1}{N}\sum_{t=1}^{N} (\varepsilon_{t+k|t})^2}, \quad (1.7)$$

once again with *N* being the number of observations in the evaluation period. RMSE has the advantages that it has the same unit as what is predicted and that it puts more weight on larger and more severe forecast errors. All three error measurements can be calculated similarly using normalized forecast errors. M. Lange (2003) shows how RMSE can be split into parts that shed light on different sources of forecast errors. In this thesis RMSE will be the main criteria for evaluation of point forecasts.

1.4.2 Probabilistic forecasts

Evaluation of probabilistic forecasts is not as intuitive and straight forward as for point forecast. One reason for this is that a probabilistic forecast, unlike point forecasts, cannot be evaluated by regarding single forecast-observation pairs. For example will there for a probabilistic forecast stating a 90 % probability that the next hours wind power production will be between 60 % and 70 % of rated output capacity be a 10 % probability that the observed wind power falls outside the interval. For this reason probabilistic forecasts needs to be evaluated following a distribution-centred approach where the forecast is evaluated on how consistent a predictive density corresponds to the distribution of observations (Tastu 2013).

Two central terms in evaluation of probabilistic forecasts are *reliability* and *sharpness*. Reliability means that the forecasted probabilities should be in accordance with the observed probabilities (Atger 1999). A simple way of controlling this is to check the correspondence between various theoretical confidence levels and how many of the forecasted values fall outside the corresponding prediction limits. For example, when evaluating a forecast with a confidence level of 90 % it should be expected that approximately 90 % of the observations in an evaluation set is covered by the forecast, given that the size of the evaluation set is sufficiently large. Sharpness, on the other hand, means that the uncertainty of the forecast should be as small as possible (Jolliffe & Stephenson 2003). For quantile forecasts this is measured as the average distance between the upper and lower quantile for a given confidence level. Notice that there to some extent is an inverse relationship between reliability and sharpness so that improving the reliability of a forecast in general will lead to a degradation of the sharpness and vice-versa (Juban *et al.* 2008). Reliability and sharpness are used to evaluate probabilistic forecasts in amongst other Bremnes (2004), Bremnes (2006), Nielsen *et al.* (2004) and Juban *et al.* (2008). A thorough presentation of reliability and sharpness including mathematical definitions is found in Monteiro *et al.* 2009.

Significant effort has also been put into scores that summarize quality aspects like reliability and sharpness into a single numerical value. This kind of score is needed in order to e.g. compare competing forecast models or make weighted combinations of multiple forecasts. Some of the scores have also been extended into versions usable for evaluation of trajectory forecasts. More information on evaluation scores for probabilistic and trajectory forecasts are found in Tastu (2013), Pinson *et al.* (2009b) and Pinson (2006).

1.4.3 Ramp forecasts

Probabilistic ramp forecasts can be evaluated using many of the same techniques as ordinary probabilistic forecasts. Deterministic forecasts, however, require dedicated techniques for evaluation.

The outcome of a deterministic forecast is often presented in the form of a *k*k contingency table* (sometimes also referred to as a classification/ misclassification matrix). This is a table (see Table 1.1) where the columns represent the actual observations and the rows the forecasts. The observations that are forecasted into the

correct class will then be on the diagonal of the table, while the forecast errors are off the diagonal. For more information on the properties of contingency tables it is referred to Agresti (2007). The contingency tables contain the full information about the forecast results presented in a structured way, but they have the disadvantage that they consist of *k***k* numbers. For applications lice automatic evaluation or comparison of different ramp forecast methods, this will be troublesome.

Table 1.1 – Schematic 3*3 contingency table. The numbers of observations in each category are represented by $n(F_i,O_j)$. N is the total number of observations. Correct predictions are on the diagonal ($i=j$, marked with green), incorrect predictions off the diagonal ($i \neq j$).

		Observed category			
		No ramp	Up ramp	Down ramp	Total
Forecasted category	No ramp	$n(F_1,O_1)$	$n(F_1,O_2)$	$n(F_1,O_3)$	$N(F_1)$
	Up ramp	$n(F_2,O_1)$	$n(F_2,O_2)$	$n(F_2,O_3)$	$N(F_2)$
	Down ramp	$n(F_3,O_1)$	$n(F_3,O_2)$	$n(F_3,O_3)$	$N(F_3)$
	Total	$N(O_1)$	$N(O_2)$	$N(O_3)$	N

A large number of evaluation metrics for contingency tables, where the *k***k* numbers are reduced to one single value, have been proposed. These can roughly be divided into simple evaluation metrics that only aims at highlighting one property of the forecast, and skill scores that aim at summarizing the totality of the contingency table into one score.

The most commonly used simple evaluation metrics are *sensitivity*, *specificity* and *frequency bias*. Sensitivity measures the probability of predicting a ramp given that a ramp is occurring with a range from 0 to 1, where 1 is perfect sensitivity. A sensitivity of 1 indicates that the forecast is perfect in predicting when ramps are not occurring, but it gives no indication of the probability of a ramp actually occurring given that a ramp is forecasted. Specificity measures the probability of not predicting a ramp given that no ramp is occurring. Specificity also has a range from 0 to 1, with 1 as perfect specificity indicating that the forecast is perfect in avoiding forecasting ramps when ramps are not occurring. Frequency bias measures how the frequency of predicted ramps corresponds to the frequency of observed ramps. Frequency bias has a range from 0 to infinity, with 1 indicating a perfect match. Bias-values lower than 1 indicate a tendency of under-predictions, while bias values higher than 1 indicate a tendency of over-predictions. In this thesis sensitivity, specificity and frequency bias will be amongst the metrics used to evaluate ramp forecasts.

The simple metrics have advantages in that they are easy to calculate, easy to understand and that they shed light on some aspect of the forecast that might be important for some application of the forecast. However, they all have very clear disadvantages, e.g. are the sensitivity and specificity easily manipulated by adjusting the shares of ramp or no-ramp forecasts. All also focus on a limited part of the contingency table and neither reflects all cells of the table.

Skill scores aim at summarizing the totality of the contingency tables into one single value. A large number of such evaluation metrics have been proposed, but the vast majority of the metrics are developed for and has validity limited to *2*2* contingency tables (binary outcomes). In the case of wind power ramps, where it is of interest to differentiate between up-ramps (rapid increases), down-ramps (rapid decreases) and no ramps the dimensions of the contingency table will be *3*3*, and hence will these metrics be of little use. Murphy (1996) claims that it is now *"generally understood"* that there is no universally acceptable way of summarizing *k*k* tables, but that there are some alternatives available under different simplifying assumptions and/or restrictive conditions.

One of these is the Hanssen and Kuipers (HK) skill score (Pierce, 1884; Hansen and Kuipers, 1965), which for *k*k* tables are defined as:

$$HK = \frac{\frac{1}{N}\sum_{i=1}^{K} n(F_i, O_i) - \frac{1}{N^2}\sum_{i=1}^{K} N(F_i)N(O_i)}{1 - \frac{1}{N^2}\sum_{i=1}^{K}(N(O_i))^2}, \tag{1.8}$$

where N is the total number of observations, K the number of categories and $n(F_i,O_i)$, $N(F_i)$ and $N(O_i)$ are defined in Table 1. *HK* measures the fraction of correct forecasts compared to how it would be if the forecasts were based on random chance. *HK* has a range from -1 to 1, with 1 as a perfect score and 0 indicating no skill.

Thorough explanations of evaluation metrics for deterministic forecasts including presentations of a wide variety of skill scores are found in Stephenson (2000) and Jolliffe and Stephenson (2003). An overview of the state-of-the-art in evaluation of deterministic forecasts of rare events is found in Casati *et al.* (2008) and references given therein. Examples on the use of metrics for evaluation of ramp forecasts are found in Ferreira *et al.* (2010).

1.5 Thesis objectives

The focus of this thesis will be on improving and enhancing NWP based wind power forecasts under Norwegian conditions. *Improving* is here understood as correcting systematic errors in single site forecasts, while *Enhancing* is understood as increasing the accuracy and usability of forecasts through e.g. group forecasts, probabilistic forecasts and ramp forecasts. All data and cases used in the thesis are from the Western coast of Norway.

Although there are local variations, the area of study is characterized by a complex coastline with many fjords and islands of various sizes as well as many near-coastal mountains. The area has very good wind conditions, with many sites with 80 meter mean average wind speeds of more than 8 m/s (see e.g. NVE 2009), but the amount of installed wind power is still very limited. At present, as few as 20 wind farms, of which eight have an installed capacity of less than 5 MW, are spread out on nearly 3000 kilometres of coast line. Norway also has very large amounts of installed hydropower (30.96 GW compared to 811 MW of wind power) with a storage capacity of 84.3 TWh (all numbers for 01.01.2014 from www.nve.no); meaning that the need for wind power forecasts for production planning and grid balancing is limited. As a result of this, limited work has been done on validating the performance of existing wind power forecast methods or developing site-specific wind power forecasts for sites in Norway. There are though some noticeable exceptions. For example did Bremnes (2006) and Bremnes (2004) show how probabilistic forecasts made by quantile regression can increase the economic value of the wind power for a wind farm in Norway, and Berge et al. (2003) looked at methods for short-term predictions of wind energy production along the Norwegian coast. These forecasts were however limited to single wind farms.

As was shown in the previous sections a large number of wind power forecasting systems, based on a wide range of methods, are already available. Some of these (e.g. Previento (see e.g. Beyer *et al.* 1999 and Focken *et al.* 2001) and WPPT (see e.g. Nielsen *et al.* 2007a and Madsen *et al.* 2005) have proven successful over longer periods of operational use. Common to the majority are that they have been developed, tested and applied to cases in the western part of Central Europe or on the Iberian Peninsula. This is where the largest volumes of wind power capacity in Europe are installed, and thus also the areas where the need for- and benefits from wind power

forecasts are the highest. With some exceptions on the Iberian Peninsula, the majority of the wind farms in these areas are located in relatively flat and simple terrain.

However, if looking at some the areas with the largest potential for further expansions of the onshore wind power capacity, like the western coast of Norway, these are characterized by a complex terrain. Kariniotakis *et al.* (2004) and Martí *et al.* (2006) have shown that performance of the state-of-the-art wind power forecast models varies with terrain complexity, sites located in complex terrain obtaining significantly higher forecast errors than sites in less complex terrain. This calls for tailored models for complex sites.

One likely explanation for the differences in forecast model performance between complex and less-complex sites is the representativeness and the quality of the input data from NWP forecasts or other sources. Möhrlen (2004) found that using NWP models for wind power forecasts demands an accurate surface parameterization. Achieving this will be more difficult in complex terrain as the changes within the NWP gridpoints are likely to be larger.

The well-established forecast models are extensive and complicated models involving the estimation of many parameters. Given input data of high quality this means that they can be fine-tuned and reach a very high level of accuracy. However, given input data of lower quality, they might struggle obtaining stable parameter estimates. One solution to this would be to increase the quality of the input data in complex terrain through increasing the spatial resolution of the NWP models, but the computational costs of this solution would be high. An alternative solution would be to develop simpler, and through that more robust, wind power forecast models for these kinds of sites. This might involve simplifying existing forecast models gaining increasing stability at the expense of optimum accuracy, or developing new, simpler models. A third solution would be to increase the emphasis put on empirical data, thus taking maximum advantage of already gained experience from the respective sites.

The main aim of this thesis is to assess and develop methods for improving and enhancing NWP wind power forecasts that is suitable for Norwegian conditions, as described in the first paragraph of this section. This involves establishing correction schemes improving the quality of the input data to statistical forecast models, proposing a method for gathering single sites into groups, investigating potential predictors for group forecasts, making adaptions to existing spatio-temporal forecast

models to increase the robustness and to propose a simple and robust model for probabilistic forecasts of group wind power production. Finally, situations resulting in very large wind power forecast errors will be addressed, also here with the aim to simplify existing forecast models or to propose a new and simpler model.

1.6 Thesis outline

The thesis is structured as follows:

- **Chapter 1** (the present chapter) gives an introduction to the field of wind power forecasting. The state-of-the-art within wind power forecasting and forecast evaluation is presented and used as a basis for the presentation of the thesis objectives.
- **Chapter 2** presents the data used in the thesis. The data consists of two datasets, one set of wind observations and one set of NWP wind forecasts covering the time-period 01.01.2009 to 17.12.2011. It is explained how the wind measurements and forecasts are transformed into synthetic wind power measurements and forecasts, and sources of uncertainty inherent in the data and the transformations are discussed.
- **Chapter 3** explores the single-site wind power forecast errors and the driving forces behind these. Systematic contributors to forecast errors are examined, and correction schemes to remove their influence are proposed and tested. The chapter constitutes a preliminary investigation of the data, with a primary aim to enhance the quality of the input data for the models used in later chapters.
- **Chapter 4** expands the approach from Chapter 3 into looking at the wind power forecast error of the lumped output of groups of sites. It is investigated how considering groups of sites instead of single sites reduce the relative forecast error, and which factors that contribute to the reduction. Further are spatio-temporal dependencies within and between groups of sites explored and some models capable of modelling the dependencies tested.
- **Chapter 5** builds on the forecasts for groups of sites from Chapter 4 and investigates the possibilities for expanding these into probabilistic forecasts. The method Bayesian Model Averaging is used to make probabilistic group forecasts based on the point forecasts for the group members.

- **Chapter 6** discusses wind power ramps and challenges related to these. The effect applying different processing schemes to wind power forecasts has on ramp predictability is studied, and the possibilities of using the classification method Random Forests to make categorical ramp forecasts is investigated.
- **Chapter 7** draws overall conclusions from the present work and gives perspectives for further research.

Chapter 2 – Data

The data used in the thesis consist of two sets, one of wind observations and one set of NWP wind forecasts, covering almost 36 months from January 1st 2009 to December 17th 2011.

A direct comparison of the two datasets may give a measure of the quality of the wind forecasts. However, as explained in Section 1.2, this cannot be directly translated into a measure of the quality of the forecast for use in wind power forecasting. In order to reflect the effect of the power curve (see Figure 1.1) the wind speed observations needs to be transformed into synthetic wind power observations. All forecast models in this thesis are based on a set of "raw" wind power forecasts made by applying the same transformations to the wind speed forecasts.

Section 2.1 presents details about how the wind speed observations are measured and discusses uncertainties inherent in the measurements. The process of choosing which sites to include in the examples in the thesis is explained and the reasons for choosing those sites are reasoned for. Section 2.2 presents details about how the wind forecasts are generated and discusses uncertainties in the NWP forecast model. Section 2.3 explains in detail how the wind speed observation and wind speed forecasts are transformed into synthetic wind power production and "raw" wind power forecasts.

2.1 Wind measurements

All wind speed and wind direction measurements used in this thesis stem from synoptic weather stations operated by the Norwegian Meteorological Institute (MET), openly available from their online service eklima.met.no. Initially wind speed and direction measurements from 225 sites spread all over Norway were considered. These were subject to a reduction process were sites not fulfilling the following criteria were discarded from the further analysis:

- Measurements of both wind speed and wind direction should cover all three years. Minor gaps in the time series are however accepted.
- Measurements should be made at 10 meter a.g.l. (or higher).

- Measurements should be made automatically (not visually) and reported for every hour.

In addition to this, a qualitative assessment was performed selecting the sites located in areas were wind power is a viable alternative. This involved discarding sites in urban areas, in the bottom of valleys, in very complex terrain, in low-wind areas (NVE 2009) etc. As the wind resources are best in the western part of Norway and all wind farms up to now have been built in this area the few remaining sites in eastern Norway were also discarded.

2.1.1 Accuracy of wind measurements

Only 43 of the 225 sites meet the described requirements (see map in Fig. 2.1). Common to these is that the wind speeds and wind directions are measured automatically at 10 meter a.g.l. with the use of either a combination of a wind vane and a cup anemometer or a sonic anemometer. The measurements are made every second and the average of the previous 10 minutes is reported at an hourly basis. The instrumentation and siting of the measurement masts are in accordance with the recommendations from the World Meteorological Organization (WMO) (WMO 2010). This implies amongst other:

- a wind speed range covering at least 0.5 m/s to 65 m/s
- a wind speed resolution of 0.5 m/s or better
- a minimum accuracy of ± 0.5 m/s for wind speeds <5 m/s and 10 % of the measured value for wind speeds over 5 m/s
- a wind direction resolution of ± 5º or better
- a minimum wind direction accuracy of ± 5º

The uncertainties in the wind speed measurements are related to the measurement methods and the measurement sites. As noted in Section 2.1 the measurements are performed in accordance with recommendations from WMO, which defines guidelines for both the siting of the measurement masts and for their instrumentations, including limits for minimum accuracy. The main challenge with the measurements is that they are performed at 10 meter a.g.l. This means that the influence of the local orography, land cover and nearby buildings will be different from at higher altitudes. Apart from the use of roughness lengths in the logarithmic transformation, no kind of refinement

to local conditions has been applied to the measurements. The selection of sites has however been done to minimize this potential problem through choosing sites in open spaces. Nevertheless, this clearly is a source of uncertainty and a potential source of error.

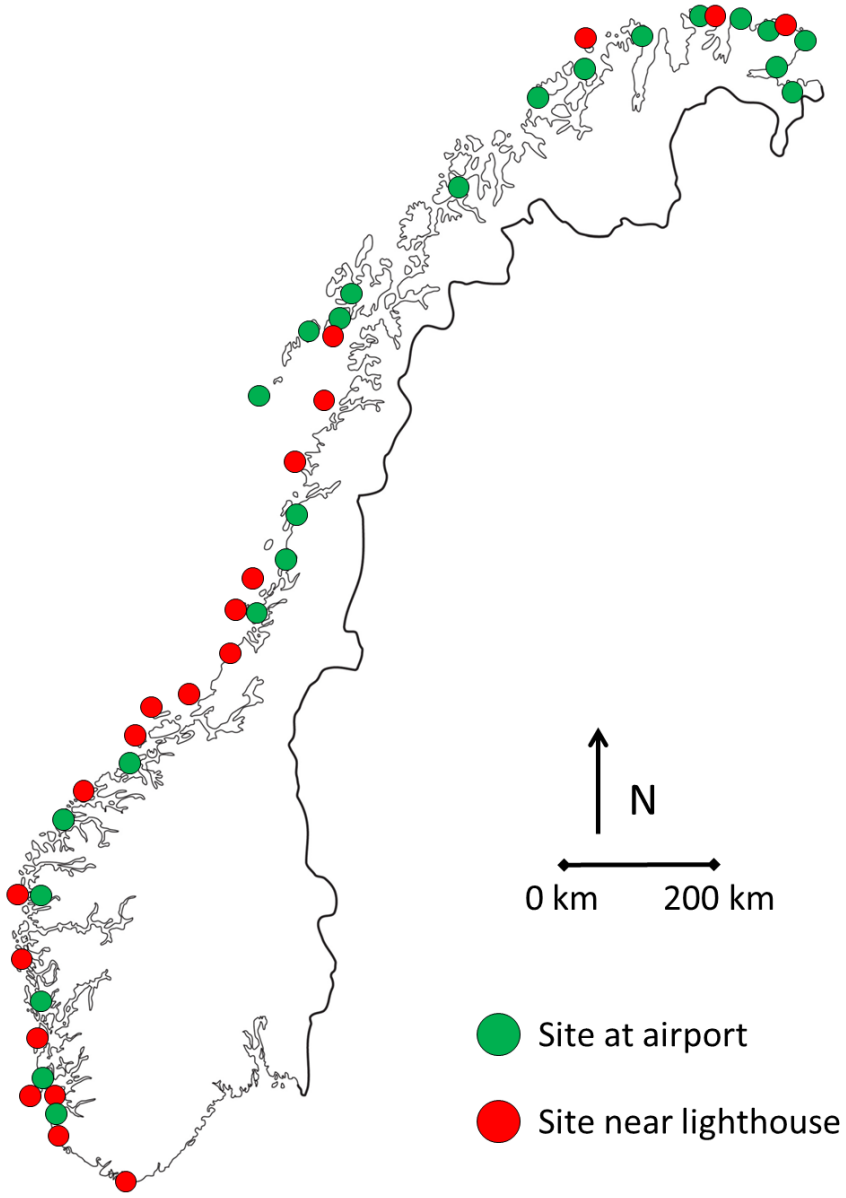

Figure 2.1 – Map of the 43 measurement sites used in the thesis. Green dots indicate sites located at airports and red dots indicate sites located near lighthouses. All measurements are made at 10 meter above ground level. Map from http://atlas.cappelen.no, © J.W: Cappelens Forlag AS, 2005

2.2 Forecasts

All the forecast data used in the thesis are provided by the NWP-model High Resolution Limited Area Model (HIRLAM) (Udén *et al.* 2002), as operated by MET (Haugen *et al.* 2008). The model has a spatial resolution of 4*4 km and comprises 60 layers in the vertical dimension. Boundary values are taken from a HIRLAM model with a spatial resolution of 8*8 km. The model is initialized with boundary values from the Integrated Forecasting System (IFS) global model with a spatial resolution of 0.25º run by the European Centre for Medium-Range Weather Forecasts (ECMWF). Forecast values for locations within the 4*4 km grid are calculated from the corresponding values at the four nearest grid-points through bilinear interpolation.

The forecasts are generated at 00 UTC in hourly steps for a look-ahead time of up to 48 hours. In our analyses and examples forecasted wind speed and forecasted wind direction, both given for 10 meter above ground level (a.g.l.), is used as the input forecast variables.

2.2.1 Accuracy of forecasts

The accuracy of the forecasts from the HIRLAM 4*4 km model has been verified four times a year against observations and has been compared to the accuracy of HIRLAM models with lower resolution (8*8 km and 12*12 km), IFS and the Unified Model (UM) with a resolution of 4*4 km. Results are found in Bremnes & Homleid (2009a-d; 2010a-c; 2011a-e). Even though there are both inter-annual and intra-annual variations the general conclusions from the verifications of the HIRLAM 4*4 km 10 meter a.g.l. wind speed forecasts are that strong winds are underestimated, weak winds are overestimated and there is an overestimation of the wind speed if averaging over all of Norway. The accuracy of HIRLAM 4*4 km is found to be comparable or better than the other models.

Uncertainty in the forecasts is caused by both errors in the specification of the initial state of the NWP model as well as errors in the NWP model formulation itself (Ehrendorfer 1997). The initial state constitutes the link between the “model world” and the “real world” providing an image of the state of the atmosphere at the time of initialization. A complete overview of the state of the atmosphere is obviously beyond reach, so the image is represented by a number of measurements of meteorological

variables taken from a large number of sites and sources. These sources will be subject to uncertainty regarding their accuracy and their representativeness. The formulation of the NWP models represents both a simplification and a discretization of the physical processes in the atmosphere. Although the models are very extensive and well tested, due to the complexity of the system there will always be a discrepancy between the model and the "real world".

A third source of errors in the NWP forecasts is the spatial resolution of the model. According to Möhrlen (2004) using NWP models for wind power forecasting demands high accuracy of the surface parameters. This implies that the demand for small grid spacing also will be high, hence that the accuracy of the forecasts will benefit from a high spatial resolution. In semi-complex terrain, as most of the sites in this thesis are located in, a distance of 4 km might involve significant changes in the terrain, and there is no guarantee that a bilinear interpolation is the best representation of a point within the grid. This is in accordance with the verification results of Bremnes & Homleid (2009a-d; 2010a-c; 2011a-e) when comparing HIRLAM models with a spatial resolution of respectively 12*12 km, 8*8 km and 4*4 km. It is however contrary to what was found in Bremnes & Giebel (2014), where it was concluded that the wind power forecasting skill did not seem to improve when using high resolution NWPs. It should though be noted that the study was based on three sites in very simple, yet coastal-near, terrain. Midtbø et al. (2011) found that a high resolution (100*100 m) computational fluid dynamics (CFD) model named SIMRA led to improvements over coarser resolution NWP models for some of the sites included in our examples. This approach would to a large extent solve the interpolation issues, but the computational costs would be high.

2.3 Transformation from wind speed to wind power

Both the measured and forecasts are given as wind speed at 10 meter a.g.l. In order to reflect the measured and forecasted wind power production the wind speed first needs to be transformed to the assumed hub height and then into to wind power.

With respect to the available data base for the transformation, only a simple approach using the basic logarithmic wind profile (see below) according to a unique roughness

length is used here. This neglects terrain effects on the wind profile and deviations of the thermal stratification of the atmosphere from neutral.

Assuming a neutral atmosphere the height transformation can be performed with the use of a logarithmic profile (Monin & Yaglom 1971). This is defined as:

$$u(h_2) = u(h_1)\frac{\ln\left(h_2/z_0\right)}{\ln\left(h_1/z_0\right)}, \tag{2.1}$$

where *u* is the wind speed, h_1 is the measurement height, h_2 is the target height and z_0 is a constant related to the surface roughness called *roughness length*. The validity of the logarithmic transformation is based on the assumption that the atmosphere is neutral. This is most likely not the case for the majority of the observations, hence is there uncertainty related to the vertical extrapolation. A correction of the vertical extrapolation according to the stability of the atmosphere would be desirable, but this information is not available.

The roughness length depends on the surface cover, land use, obstacles etc. A subjective way of estimating the roughness length is to perform a visual survey of the terrain surrounding the measurement site and compare this to a table of terrain classifications (WMO 2010). Here the visual inspection is performed on the basis of aerial photos from Google Earth and kart.finn.no and compared to a table of terrain classifications in Davenport *et al.* (2000). A list of sites and the corresponding assumed roughness lengths is found in Annex 1.

To reflect the non-linear effect of the power curve the height-transformed wind speed measurements and forecasts are transformed into synthetic wind power by the use of a generic power curve. Here the theoretical power curve of a Siemens SVT-2.3 is used as the power curve (Siemens 2009). The Siemens SVT-2.3 is a very widely used 2.3 MW modern 3-blade upwind wind turbine with a power curve that has all the common characteristics of power curves (see Fig. 1.1). Transforming wind to power by the use of a theoretical power curve means that the uncertainty in the transformation process (see Section 1.2) is not directly reflected. Parts of this will, however, be included through the effect of wind direction on the wind speed forecast error.

Chapter 3 – Single site forecast uncertainty

In this chapter factors explaining the single-site wind power forecast errors are explored. The forecast errors are calculated following Equation 1.4. The main focus of this chapter is on the systematic part of the forecast errors, the forecast bias, and how this can be reduced. Section 3.1 constitutes an exploratory investigation of which factors that contribute to a large forecast bias. Sections 3.2 discuss how the findings from Section 3.1 can be used in bias correction schemes to reduce the forecast error and provide later models with input data of higher quality. Different bias correction schemes are tested and evaluated. Section 3.3 draws some general conclusions on the factors influencing the single site forecast errors and raise some ideas for other options for correction schemes.

3.1 Exploratory analysis of forecast errors

This section aims at uncovering factors that contribute to forecast bias and through that to the forecast uncertainty. Five different factors are considered; terrain complexity, surface roughness length, look-ahead time, wind direction and wind speed. Notice that as the NWP forecasts used in the thesis are issued once a day (at 00:00 CET) the look-ahead time corresponds to the time of day for which the forecast is issued. For the latter three the site-to-site differences are very large and not possible to visualize in one figure. To provide details and figures of all 43 sites will be excessive, so in order to explain and visualize the main findings three sites will be used as examples. These are:

- Hammerfest – Site located at an airport in the very far north of Norway.
- Rørvik – Site located at an airport at a large island (~100 km^2) off the western coast in the middle of Norway.
- Utsira – Site located by a lighthouse on a small island (~6 km^2) off the western coast of southern Norway.

3.1.1 Terrain complexity and surface roughness length

Two factors likely to influence the accuracy of the forecasts are the terrain complexity and the surface roughness length. As described in Section 2.2 terrain complexity influence forecast uncertainty through an increasing demand for accurate surface parameterization and through the bilinear interpolation from the model grid points. The influence of terrain complexity on wind power forecast accuracy has been studied by amongst other Marti *et al.* (2001) and Kariniotakis *et al.* (2004). Both found that the forecast accuracy was reduced with increasing terrain complexity. M. Lange (2005) found that the bias of wind speed predictions depend on terrain complexity, with more complex sites having a negative bias (e.g. an under-prediction of the wind power).

One way of reducing the uncertainties caused by complex terrain is by spatial refinement through downscaling of the NWP forecasts using a high-resolution CFD model (see e.g. Magnusson & Wern 2001, Mana *et al*. 2013 or Castellani *et al.* 2014). This method does however demand an extensive modelling of the terrain and tend to be computationally expensive. In cases where the uncertainty of the NWP forecasts used to initialize the CFD model is high, the CFD model might also end up inflating the uncertainty due to the additional computations. A less computationally demanding alternative to a CFD model is the wind transformation model WAsP (see e.g. Mortensen *et al.* 1993). Although WAsP has its main use as a wind resource estimation tool, it can also be used to analyse how the terrain, surface roughness and obstacles affect the wind speed and wind direction at a site. This, in turn, can be used to build correction schemes for the NWP forecasts. WAsP-based correction schemes will however be static, and will therefore not differ significantly in use from empirically-based correction schemes. Like for the CFD models, WAsP also requires extensive modelling of the terrain. For these reasons no spatial refinement through downscaling of the NWP forecasts used in this thesis has been performed.

The terrain complexity is represented by a ruggedness index value (RIX) (Bowen & Mortensen 1996). This is a measure of the percentage of the terrain within a certain radius that has an inclination exceeding 30 %. The RIX values used here are taken from a wind atlas for Norway made by Kjeller Vindteknikk for the Norwegian Water Resource and Energy Directorate (NVE 2009a). The values are given for 1*1 km and following recommendations from Berge *et al.* (2006) the radius for the calculations were set to 2 km. Details about the calculations of the RIX-values are found in NVE (2009b). The surface roughness length influences the forecast uncertainty through the

vertical extrapolation of the wind speed (Equation 2.1). Details about how the surface roughness lengths are found are given in Section 2.3.

Figures 3.1 and 3.2 show the wind power forecast bias as a function of look-ahead time split on respectively RIX-value (as calculated in NVE 2009a) and assumed surface roughness length (see Section 2.3 and Annex 1). The mean bias for each look-ahead time up to 24 hours is calculated and each of the 43 sites is plotted individually. For the RIX-values the sites are grouped into sites with RIX values lower and higher than 5 % and for the roughness length values in groups with roughness lengths higher and lower than 0.2. As is seen in Figure 3.1 there is a much larger variation in the bias amongst the sites with the higher RIX-values. However, there is no clear tendency of high-RIX sites having a negative bias, as was found in M. Lange (2005). Figure 3.2 shows that the sites with roughness lengths < 0.2 tends to have a slightly positive bias, while sites with roughness lengths > 0.2 tends to have a slightly positive bias.

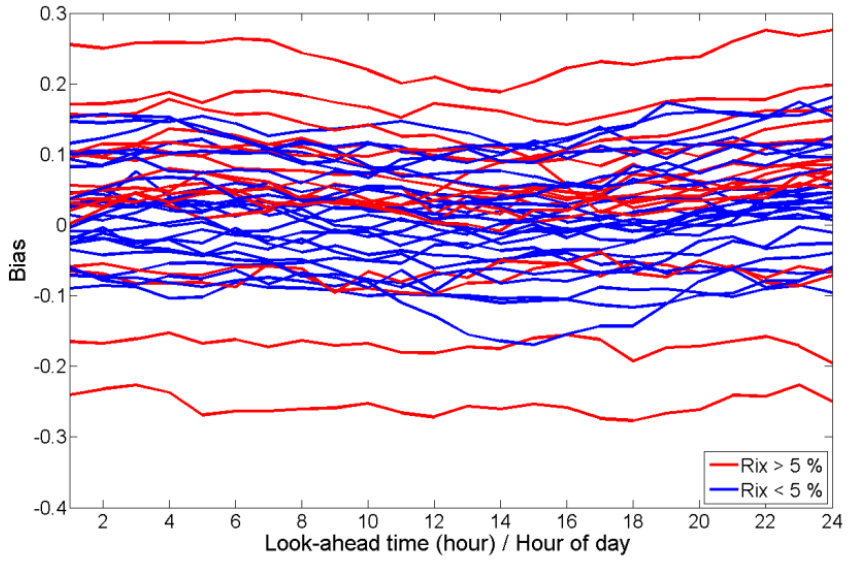

Figure 3.1 – Wind power forecast bias as a function of look-ahead time split on RIX-value. One line indicating each site for look-ahead times up to 24 hours. Red lines indicate sites with RIX-values > 5 % and blue lines sites with RIX-values < 5 %.

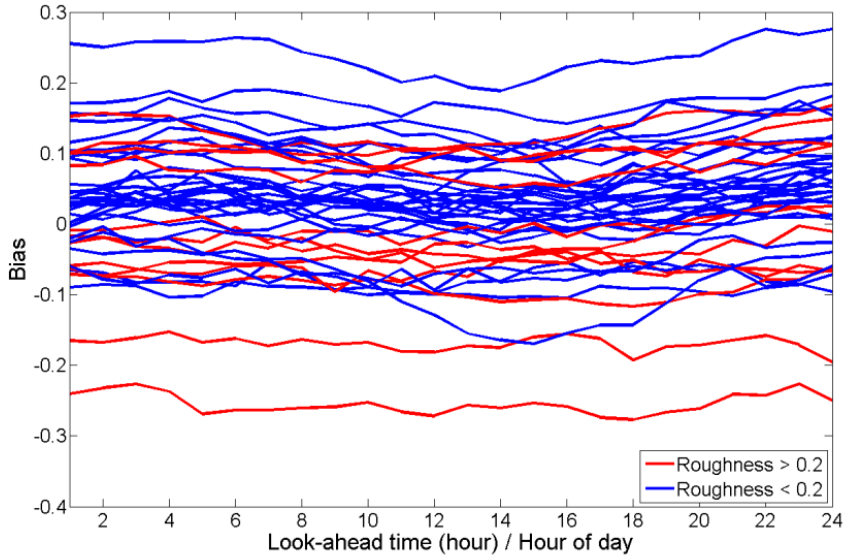

Figure 3.2 – Wind power forecast bias as a function of look-ahead time split on assumed surface roughness length. One line indicating each site for look-ahead times up to 24 hours. Red lines indicate sites with roughness lengths > 0.2 and blue lines sites with roughness lengths > 0.2.

3.1.2 Look-ahead time

Forecasts made for longer look-ahead times means that the NWP model must do calculations for larger numbers of time-steps and that the uncertainty of the forecasts increases. It is however not equally obvious that the increase in uncertainty will be systematic, hence that the forecast bias will increase. Notice also that as the NWP forecast used in this thesis is only run once a day it is not possible to make a clear distinction between bias caused by increasing look-ahead time and bias caused by the time for which the forecast is issued (e.g. night-day differences in surface heating).

For each look-ahead time the mean bias is calculated. The mean bias for all sites combined is shown by the blue line in Figure 3.12. This shows a V-shaped pattern indicating diurnal variations in the forecast bias and a slight increase in the bias with increasing look-ahead time. The site-to-site variations are however very large and as a consequence of this the sites should be assessed individually.

Figures 3.3 (Hammerfest) and 3.4 (Rørvik) shows a V-shaped pattern indicating diurnal variations in the forecast bias. The error bars (vertical red lines) indicate 95 % confidence regions for the forecast bias of each look-ahead time. Similar diurnal variations have been described by M. Lange (2003) for sites in Germany, where he

claims that this can be explained by imperfect modelling of the stability effects of the lower boundary layer by the NWP model. Hammerfest (Figure 3.3) has a positive bias fluctuating around 0.1 (10 %), while Rørvik has a positive bias in between 0.19 and 0.28 (19 % - 28 %) which is the highest of all the 43 sites. In Figure 3.5, describing the forecast bias of a site at the relatively small island Utsira, no obvious diurnal variation is found. This is also in accordance with the findings of M. Lange (2003), and can be explained by the sea dampening day-night variations in surface temperature. Utsira (Figure 3.5) are also distinguished by having a very low bias (< 0.02) centred on 0. From this it is very likely that the differences in bias over the day for Hammerfest and Rørvik is an effect of diurnal variations in the stability of the lower boundary layer, i.e. an effect of the time of day, and not an effect of the look-ahead time.

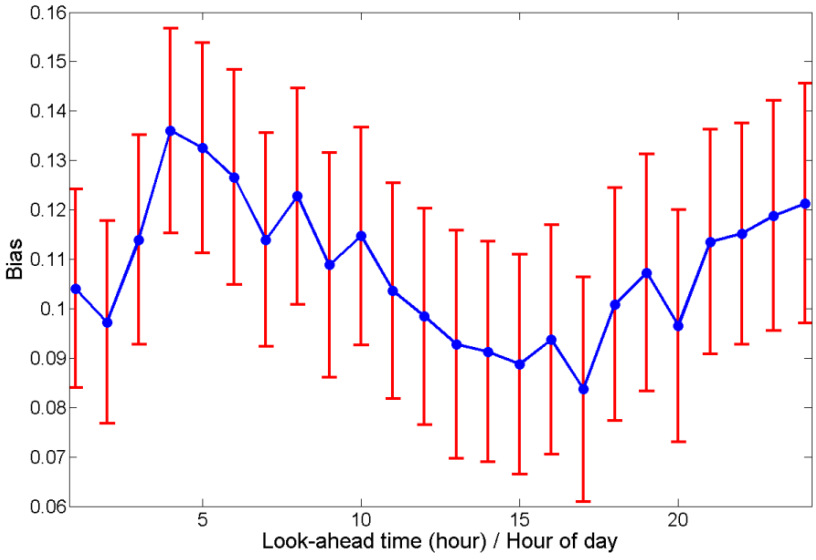

Figure 3.3 – Bias as a function of look-ahead time at Hammerfest. The red lines indicate 95 % confidence regions for the forecast bias for each look-ahead time. The plot shows clear signs of diurnal variations in the forecast bias with lower bias during daytime and higher bias at night. Early morning (04:00-05:00) and afternoon (14:00-15:00 and 17:00) are significantly different.

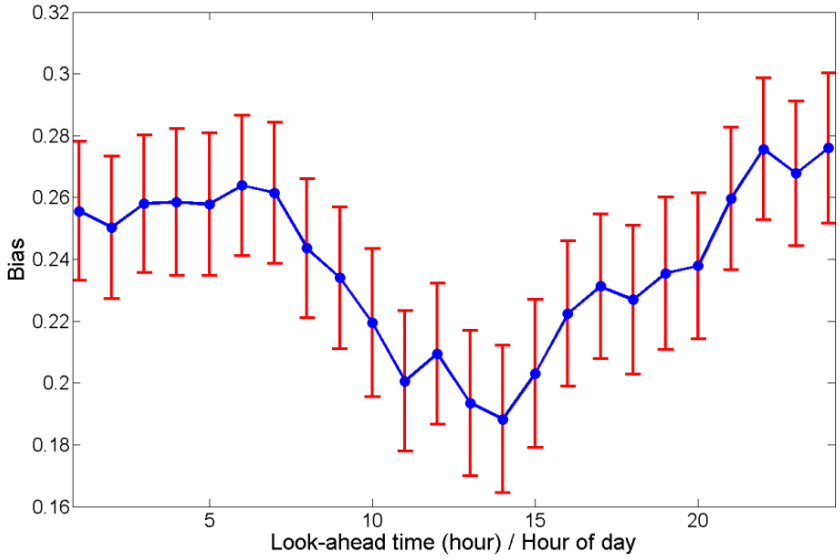

Figure 3.4 – Bias as a function of look-ahead time at Rørvik. The red lines indicate 95 % confidence regions for the forecast bias for each look-ahead time. The plot shows clear signs of diurnal variations in the forecast bias. Early morning (03:00-07:00) and late evening (21:00-24:00) have a significantly higher bias than midday (11:00-15:00).

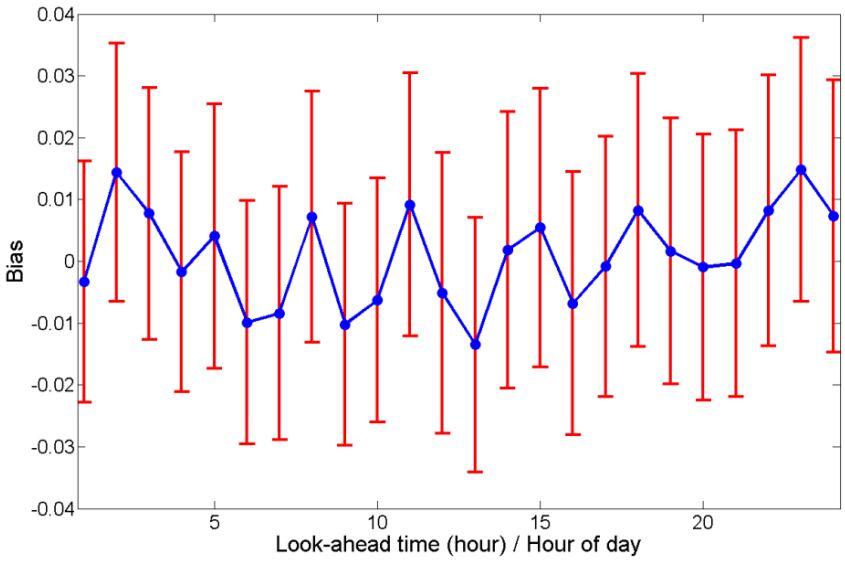

Figure 3.5 – Bias as a function of look-ahead time at Utsira. The red lines indicate 95 % confidence regions for the forecast bias for each look-ahead time. The plot shows a very low bias centred on 0. No significant differences in bias are found.

3.1.3 Wind direction

The surroundings of the sites cannot be assumed to be equal in all directions. Hence might the wind at a site be subject to differences in e.g. terrain complexity and surface roughness length upwind of the site depending on the wind direction. Considering the resolution of the NWP forecast model not all of this will be covered by the wind speed forecasts. This might lead to wind direction dependent differences in the wind power forecast accuracy which, if assuming that the terrain and surface cover is relatively static, it is likely to believe will be systematic. This will be expressed as a direction-specific forecast bias. Saleck & von Bremen (2007) report similar direction-dependent differences for the predicted wind speed for a wind farm in northern Germany, and this is also one of the effects that most physical wind power forecast systems, like *Prediktor* (Landberg 1998; Landberg 1999) and *Previento* (Beyer *et al.* 1999; Focken *et al.* 2002a), account for.

The directional bias is calculated as the mean bias for all observations where the forecast wind direction falls into one of eight equally sized wind sectors. For each wind sector a 95 % confidence region for the bias is also calculated (error bars, the vertical red lines in the figures). When these do not overlap the difference in forecast bias of two wind sectors are significantly different. Like for the time-dependent bias the site-to-site variations are very large so the sites are assessed individually. The three same sites as was used to illustrate the time-dependent bias are also used to illustrate the direction-dependent bias.

Figure 3.6 shows the direction-dependent bias at Hammerfest. As is seen the bias heavily depends on the wind direction, with a bias ranging from slightly under 0 (2-3 % under forecasting) for wind forecast wind directions 45º to 135º to between 0.2 and 0.25 (25-25 % over forecasting) for forecast wind directions 135º-225º. Figure 3.7 shows the direction-dependent bias at Rørvik. As expected from previous knowledge this in general is high and in a positive direction. The wind power forecast bias as Rørvik follows the same pattern as Hammerfest; a W-shape with peaks in the bias following the north-south axis and the troughs following the east-west axis. The wind power forecast bias is highest for wind directions between 135º and 180º, slightly exceeding 0.35 (35 % over forecasting), and lowest for wind directions between 45º and 90º and between 225º and 270º, both between 0.1 and 0.15 (10 %- 15 % over forecasting). Figure 3.8 shows the direction-dependent bias at Utsira. Also here the W-

shape is seen, but the bias is small and the difference between the peaks (0.04, 4 % over forecasting) and the troughs (-0.03, 3 % under forecasting) is limited.

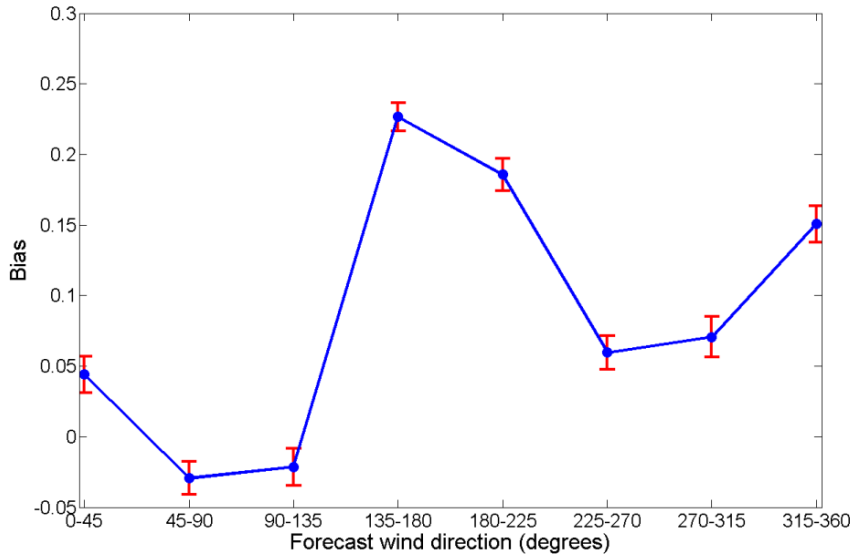

Figure 3.6 – Direction-dependent bias at Hammerfest. The red lines indicate 95 % confidence regions for the bias of each wind sector. The bias of the forecasted wind power range from -0.03 to 0.24 depending on the forecast wind direction.

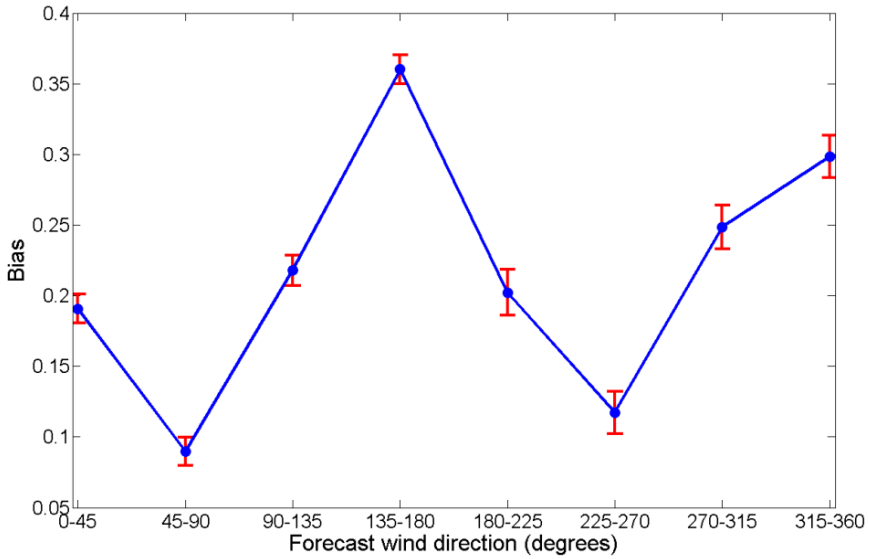

Figure 3.7 – Direction-dependent bias at Rørvik. The red lines indicate 95 % confidence regions for the bias of each wind sector. The bias of the forecasted wind power range from 0.11 to 0.36 depending on the forecast wind direction.

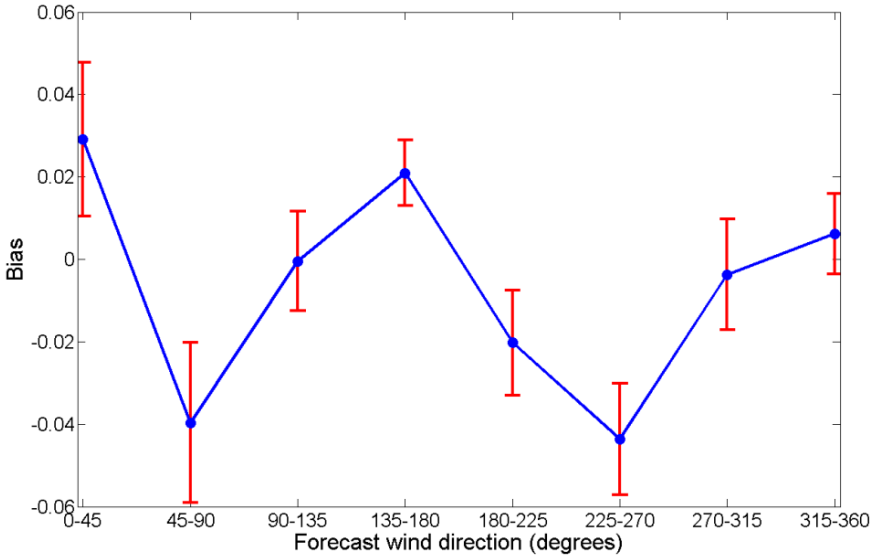

Figure 3.8 – Direction-dependent bias at Utsira. The red lines indicate 95 % confidence regions for the bias of each wind sector. The bias of the forecasted wind power range from -0.03 to 0.04 depending on the forecast wind direction. Notice that even though the error bars might seem much wider than in Figures 3.6 and 3.7 this is an effect of the scale of the y-axis.

3.1.4 Wind speed

Bremnes & Homleid (2009a-d; 2010a-c; 2011a-e) have documented that the HIRLAM NWP model has a tendency to over-forecast low wind speeds and under-forecast high wind speeds. Even though this cannot be directly translated into over- and under-forecasting of the wind power production, it provides an indication that there might be a systematic wind speed dependent bias also in the wind power forecasts.

To assess the wind speed dependency of the wind power forecast bias, the data at each site is split into four categories (0-5 m/s, 5-10 m/s, 10-15 m/s and 15+ m/s) depending on the forecast wind speed. For each site and category the average bias and 95 % confidence regions (error bars, the vertical red lines in the figures) for the average bias is calculated. Like for the time-dependent and direction-dependent bias the site-to-site variations are large, and the same three sites are used for examples.

Figure 3.9 shows how the wind power forecast bias depends on the forecast wind speed at Hammerfest. The bias is negative (-0.12 and -0.25) for low wind speeds (<5 m/s) and for very high wind speeds (15+ m/s) and positive (0.14 and 0.08) for intermediate wind speeds (5-15 m/s). Figure 3.10 shows how the wind power forecast bias depends on the forecast wind speed at Rørvik. The bias is positive for wind speeds

up to 15 m/s, with a clear peak (0.45) for wind speeds between 5 m/s and 10 m/s. For very high wind speeds (15+ m/s) there is a very large negative bias of approx. -0.8. It should however be noted that this is calculated from a very limited number of observations, hence is the uncertainty of this result large. Figure 3.11 shows how the wind power forecast bias depends on the forecast wind speed at Utsira. Also here the bias is negative (-0.09 and -0.02) for low wind speeds (<5 m/s) and for very high wind speeds (15+ m/s), and positive (0.05) for intermediate wind speeds (5-15 m/s). As this is a high mean wind site, wind speed forecasts exceeding 15 m/s is not uncommon. This contributes to the high wind speed bias being less extreme than for the two other sites. For all three sites the confidence regions are very narrow for wind speeds up to 15 m/s. This is an indication that the forecast bias is systematic, and thus that wind speed is a relevant factor to include in a correction scheme.

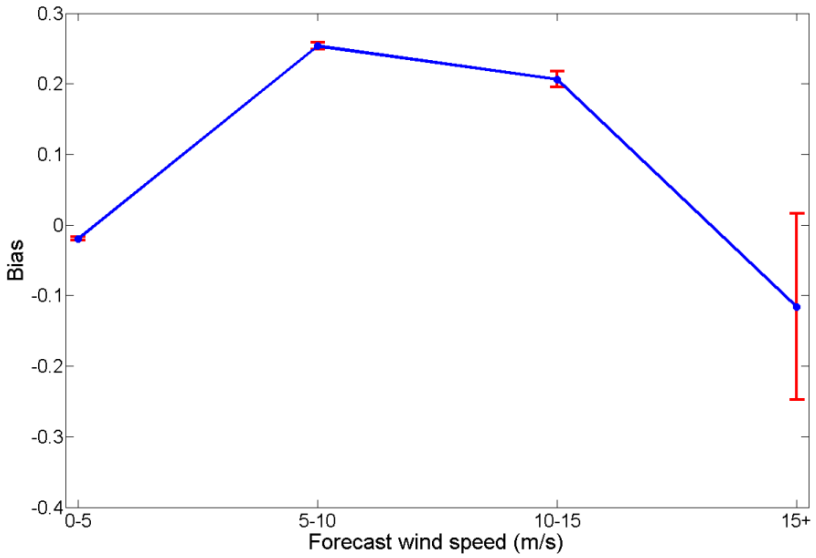

Figure 3.9 – Wind speed dependent bias at Hammerfest. The red lines indicate 95 % confidence regions for the bias of each wind speed category. The bias of the forecasted wind power range from -0.25 to 0.14 depending on the forecast wind speed. The confidence regions for the bias is narrow for wind speeds up to 15 m/s, and appear even more so as a result of the scale on the y-axis. The number of forecasts of wind speeds 15+ m/s is very limited resulting in large uncertainties and a wide confidence region for the forecast bias.

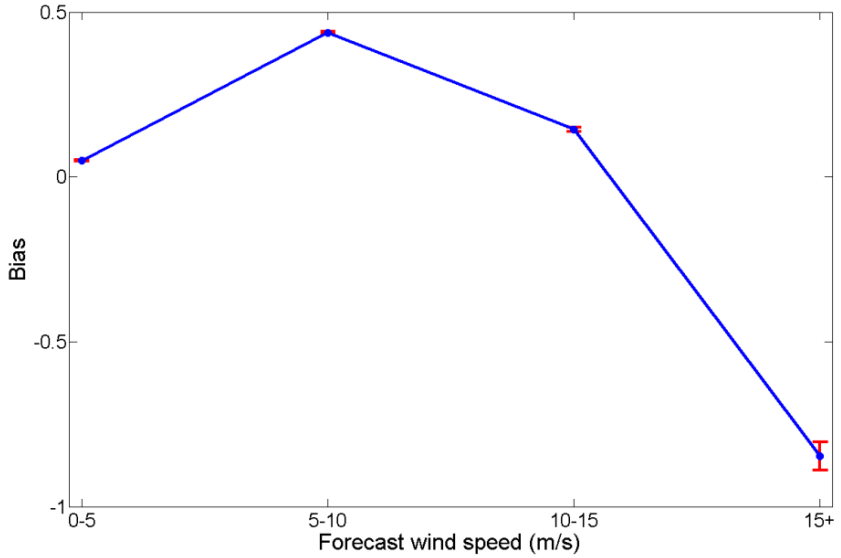

Figure 3.10 – Wind speed dependent bias at Rørvik. The red lines indicate 95 % confidence regions for the bias of each wind speed category. The bias of the forecasted wind power range from -0.8 to 0.45 depending on the forecast wind speed. The confidence regions for the bias is narrow for wind speeds up to 15 m/s, and appear even more so as a result of the scale on the y-axis. The number of forecasts of wind speeds 15+ m/s is very limited resulting in large uncertainties and a wide confidence region for the forecast bias.

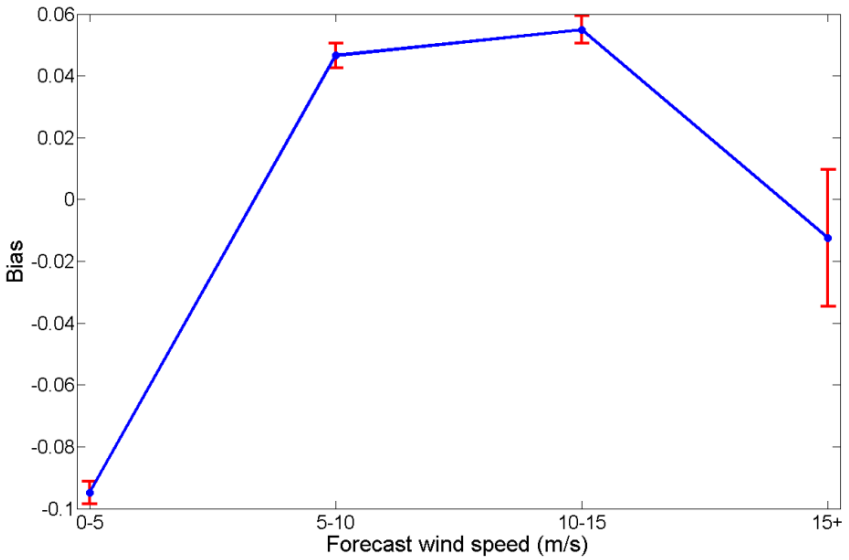

Figure 3.11 – Wind speed dependent bias at Utsira. The red lines indicate 95 % confidence regions for the bias of each wind speed category. The bias of the forecasted wind power range from -0.09 to 0.05 depending on the forecast wind speed. Notice the difference in the scale of the y-axis compared to Figures 3.9 and 3.10. The confidence regions for the bias are narrow for all four wind speed categories, but as in Figure 3.9 and 3.10 the uncertainty of the bias for forecasted wind speeds 15+ m/s is clearly higher than for the lower wind speed categories.

3.2 Bias correction

In Section 3.1 it was shown that the bias of wind power forecasts partly can be explained by a number of external factors. It was further shown that there were large site-to-site variations in both the importance of the different factors and in how they influenced the forecast bias. In this section the findings from 3.1 are turned into bias correction schemes that will be applied to the raw wind power forecasts in order to reduce their bias and through that their RMSE.

The bias corrections are made by calculating the systematic error in an empirical dataset (e.g. the average error for a certain look-ahead time (hour of day), wind direction etc.) and using these to correct future observations. If for example a specific look-ahead time shows a consistent pattern of over-forecasting, the bias correction scheme will deduct the average forecast error for that look-ahead time from future forecasts for the same look-ahead time.

As long as applied to the same data that the bias is calculated from ("seen data"), including more factors in a bias correction scheme will also lead to a lower bias. However, this will not necessarily be the case when later applying the bias correction scheme to new data ("unseen data"). For some of the factors some categories (e.g. wind directions and wind speeds) have very few observations. For these cases the observed bias might not be a systematic error, but can as likely be the result of a few extreme incidents. This reduces the generalizability of the bias estimates and might lead to problems with over-fitting. To control for this the dataset is split in two parts, where 80 % of the data is used to calculate the factor-dependent biases the correction schemes are based on and the remaining 20 % is used as an independent test set to check the results.

From Figures 3.1 and 3.2 it is clear that it is not possible to base a bias correction scheme on the sites RIX values or surface roughness. Even though this would have reduced the bias all sites combined, it would inevitably lead to some sites getting a higher bias than what they had initially. For other applications than the construction of a nation-wide forecast this would be a clear drawback. For the other factors the bias corrections can be calculated and applied per site.

Five different bias correction schemes are tested in order to reduce the systematic wind power forecast errors of the wind power forecasts presented in Chapter 2. These are:

- A simple version, where the average bias for all observations is used to correct the forecasts. The bias is calculated as:

$$BIAS = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_t, \quad (3.1)$$

where *N* is the total number of forecast-observation pairs and ε_t is forecast error of forecast *t*.

- An hourly version, where the average bias for each look-ahead time (hour of day) is used to correct the forecast. The hourly bias is calculated as:

$$BIAS_h = \frac{1}{n_{t\epsilon h}}\sum_{t\epsilon h} \varepsilon_t, \quad (3.2)$$

where *h* is the look-ahead time (*h = 1,…,24*), $n_{t\epsilon h}$ is the number of forecasts for look-ahead time *h* and ε_t is the forecast error of forecast *t*.

- A wind direction version, where the average bias for specified wind sectors (see Section 3.1.3) is used to correct the forecasts. The directional bias is calculated as:

$$BIAS_d = \frac{1}{n_{t\epsilon d}}\sum_{t\epsilon d} \varepsilon_t, \quad (3.3)$$

where *d* is an indicator variable for the forecast wind direction (binned into eight equally sized sectors, see Section 3.1.3), $n_{t\epsilon d}$ is the number of forecasts in wind direction sector *d* and ε_t is the forecast error of forecast *t*.

- A wind speed version, where the average bias for specified wind speed categories (see Section 3.1.4) is used to correct the forecasts. The wind speed bias is calculated as:

$$BIAS_s = \frac{1}{n_{t\epsilon s}}\sum_{t\epsilon s} \varepsilon_t, \quad (3.4)$$

where *s* is an indicator variable for the forecast wind speed (binned into four categories, see Section 3.1.4), $n_{t\epsilon s}$ is the number of forecasts in wind speed category *s* and ε_t is the forecast error of forecast *t*.

- A wind direction and wind speed version, where the two preceding correction schemes are combined and the forecasts corrected for both wind speed and wind direction bias. The wind direction and speed bias is calculated as:

$$BIAS_{d,s} = \frac{1}{n_{t\epsilon d,s}} \sum_{t\epsilon d,s} \varepsilon_t, \tag{3.5}$$

where d and s are indicator variables for the forecast wind direction (binned into eight equally sized sectors, see Section 3.1.3) and forecast wind speed (binned into four categories, see Section 3.1.4), $n_{t\epsilon d,s}$ is the number of forecasts that belongs to both wind direction sector d and wind speed category s and ε_t is the forecast error of forecast t.

Figure 3.12 shows the resultant forecast bias after applying the five bias correction schemes to the same data as they are built from. The bias of the raw, uncorrected data is included as a comparison. The bias of all the 43 sites is aggregated into one series of bias values for look-ahead times up to 24 hours. Naturally, the hourly bias correction comes out with zero bias for all look-ahead times. Of the remaining four correction schemes the simple and wind direction based leads to a slightly lower bias than the wind speed based and the direction + speed based for most look-ahead times. All four correction schemes in average lead to significant reductions in bias over the raw forecasts, but for look-ahead times between 12 and 16 hours the raw forecasts have a comparable or lower bias.

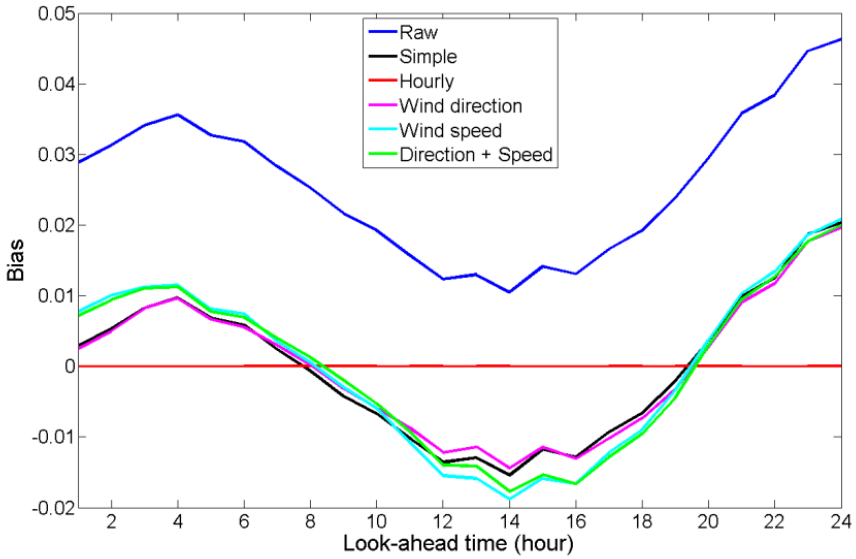

Figure 3.12 – Bias for the raw data (blue line) and the raw data corrected by the use of five bias correction schemes. The bias of the 43 sites is aggregated into one single line for each scheme. Naturally, the hourly bias correction (red line) has the lowest bias (0 for all look-ahead times), followed by the simple (black line) and wind direction (magenta line) bias correction. All bias correction schemes in average gives a significant reduction in the bias, but for some look-ahead times (12-16 hours) the raw forecast have a comparable or lower bias.

Figure 3.13 shows the resulting forecast RMSE after the bias corrections in Figure 3.12. All five bias correction schemes lead to a reduction in RMSE compared to the raw forecasts, but in contrast to the changes in resulting bias there are clear differences between the correction schemes. The simple and hourly correction schemes leads to the least improvements in RMSE (~0.01 compared to no bias correction). Both considering the obtained bias reductions of the hourly correction schemes and the findings in M. Lange (2003), this is surprising. The directional bias reduction scheme gives a reduction in RMSE of ~0.025, while the wind speed and direction + speed correction schemes gives reductions in RMSE of respectively ~0.035 and ~0.043. As expected the RMSE of all correction schemes and the raw data are increasing with increasing look-ahead time, but the internal order and spacing between the correction schemes remains more or less unchanged for look-ahead times up to 24 hours.

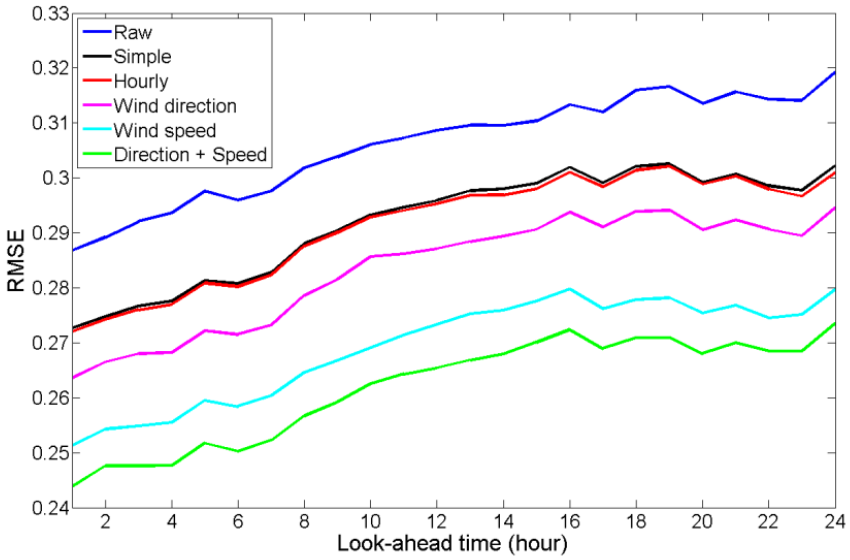

Figure 3.13 –RMSE for the raw data (blue line) and after applying the five bias correction schemes to the raw data. The bias of the 43 sites is aggregated into one single line for each scheme. The direction + speed correction (green line) gives the lowest RMSE, followed by the wind speed correction (light blue line), the wind direction correction (magenta line) and the hourly and simple correction (red and black lines).

Figure 3.14 shows the resulting bias after applying the bias correction schemes from Figure 3.12 to unseen data. The differences compared to Figure 3.12 are striking. The wind speed and direction + speed correction schemes actually lead to an increase in the bias compared to the uncorrected raw data. This is a clear indication that the correction schemes are more influenced by single observations than general and generalizable systematic patterns in the forecast errors. The directional bias correction scheme gives

a bias between -0.01 and 0.03 and has the lowest bias of the correction schemes for all look-ahead times up to 24 hours. The simple and hourly correction schemes give a bias that is comparable to that of the uncorrected raw data, but the bias is shifted from positive (over-forecasting) to negative (under-forecasting). For all correction schemes except for the hourly traces of a diurnal V-shape pattern is found.

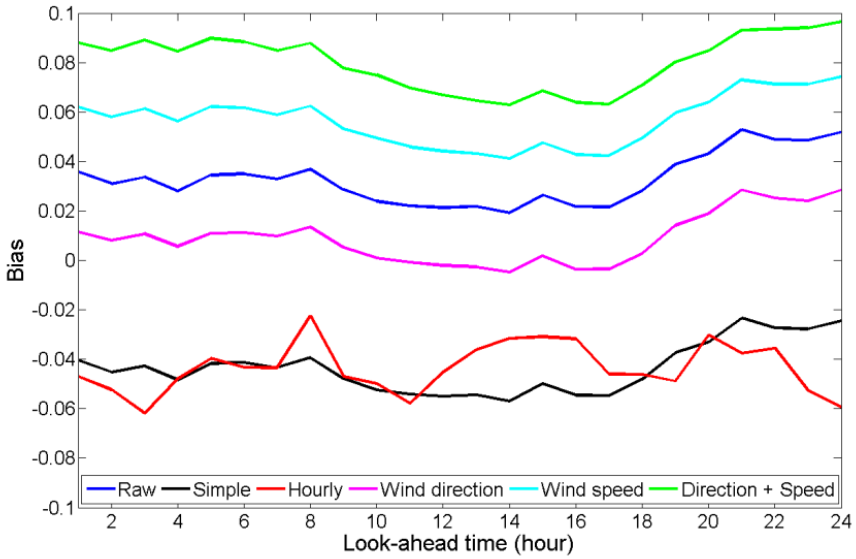

Figure 3.14 – Bias for the raw data (blue line) and the raw data corrected by the use of five bias correction schemes applied to new and unseen data. The bias of the 43 sites is aggregated into one single line for each scheme. In contrast to when the bias correction scheme is applied to the same data as the biases were calculated from the direction + speed scheme (green line) results in the highest bias. Also the wind speed correction scheme (light blue line) results in a higher bias than that of the unprocessed raw data. The simple and hourly correction schemes (black and red lines) have a bias at the same level as the unprocessed raw data, but shifted from positive to negative. The wind direction based bias correction scheme (magenta line) results in the lowest bias for all look-ahead times.

Figure 3.15 shows the resulting forecast RMSE after the bias corrections in Figure 3.14, that is, the bias correction schemes applied to unseen data. Also here the differences compared to the version where the bias corrections are applied to the same data as they are calculated from are striking. Like in Figure 3.14 the direction + speed corrections comes out with the poorest results, here in terms of the highest RMSE. This supports the assumption that the amount of available data is not sufficiently large to provide stable estimates of the bias for all the corrections (8 wind direction sectors * 4 wind speed categories), and as a result of this ends up introducing noise instead of correcting systematic errors. The simple and hourly bias correction schemes result in a RMSE at the same level or slightly higher than that of the uncorrected raw data. Notice that the RMSE of the uncorrected unseen data is slightly lower than that of the data the

correction schemes are based on. The wind speed based correction scheme leads to an improvement of the RMSE that is increasing with increasing look-ahead time. The largest reduction in RMSE is achieved by the wind direction based correction scheme. Data that have been bias corrected using this scheme has the lowest RMSE for all look-ahead times, with a consistent reduction in RMSE of ~0.02 compared to the unprocessed raw data for all look-ahead times. All five correction schemes and the uncorrected raw data have an increase in RMSE with increasing look-ahead time, even though this for the direction + speed scheme is not particularly evident.

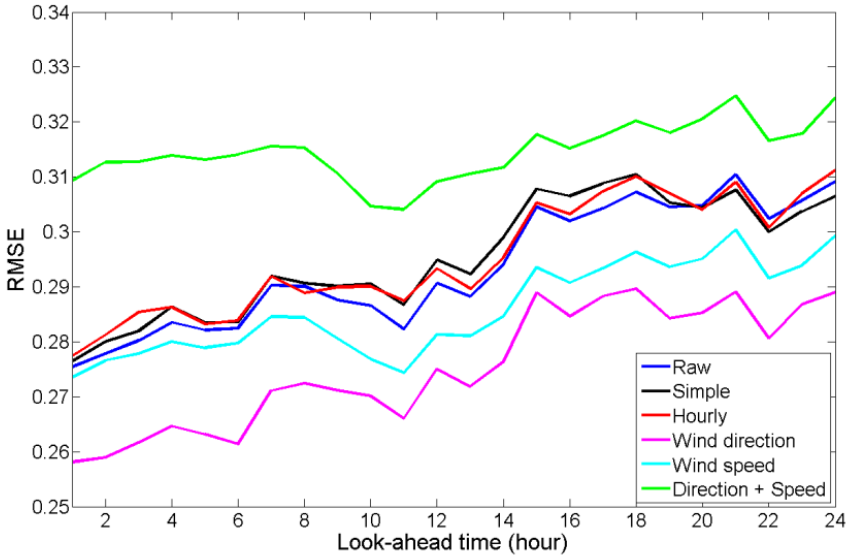

Figure 3.15 – RMSE for the new, unseen raw data (blue line) and after applying the five bias correction schemes to the unseen data. The bias of the 43 sites is aggregated into one single line for each scheme. The wind direction correction (magenta line) gives the lowest RMSE, followed by the wind speed correction (light blue line). The simple and hourly correction schemes (black and red lines) result in a RMSE at the same level as the uncorrected data. The direction + speed correction (green line) results in the highest RMSE for all look-ahead times. This is a clear indication that the amount of data is not sufficiently large to get stable estimates for all the corrections and that the correction scheme ends up adding noise rather than removing systematic errors.

3.3 Conclusions

Multiple external factors contribute to the wind power forecast uncertainty. When these contributions are systematic they can be accounted for through bias correction schemes. In this chapter it was found that the look-ahead time, the wind direction and the wind speed gave systematic contributions to the forecast errors. It was further found that the site-to-site variations both in the importance of the different factors and the magnitude of their contributions were large, hence that the bias needed to be

corrected for each site individually. This was illustrated by three sites; Hammerfest, Rørvik and Utsira.

Even though a bias correction scheme accounting for both wind direction and wind speed were found to give the largest reduction in RMSE when applied to the same data as the bias was calculated from, the performance of this scheme when applied to new, unseen data was poor. The general cause of this is lack of data from high wind speeds and non-prevailing wind directions, which means that the estimates that are supposed to be of systematic patterns are based on too few observations to be generalized. As a consequence of this they end up adding noise. The bias correction schemes should therefore be kept as simple, and hence robust, as possible.

Both in the bias of the raw, uncorrected data and in the bias of the corrected data a V-shaped diurnal pattern is found. This has earlier been described by M. Lange (2003) and explained by imperfect modelling of the stability effects of the lower boundary layer by the NWP model. None of the bias correction schemes (except for hourly bias corrections) are able to fully remove this effect, and when trying to combine other correction schemes with the hourly correction problems with lack of data and instable bias estimates quickly occur.

The bias correction scheme based on the forecast wind directions (split into eight equally sized sectors) were found to give in average the largest reduction in bias and also resulting in the in average lowest RMSE (see Figures 3.14 and 3.15). For look-ahead times up to 24 hours this correction scheme gave a reduction in the average RMSE of ~2 percentage points (6,5 % reduction). The change in single site RMSE (average of all look-ahead times) varied between an increase of ~0.2 percentage points and a decrease of ~11 percentage points. A comparison between the single site RMSE of the raw, uncorrected data and the RMSE of the data after the directional bias correction is shown in Figure 3.16. The output of the directional correction scheme will be used as input data in all examples later in the thesis.

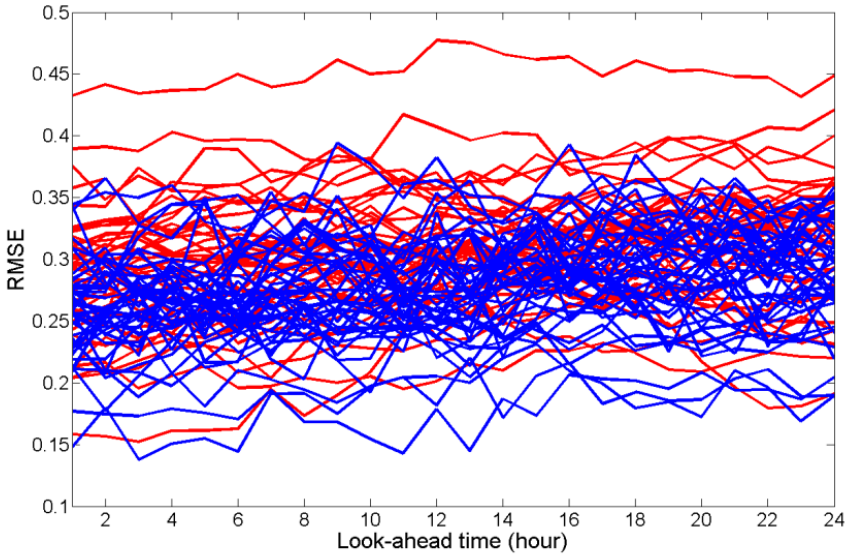

Figure 3.16 - Wind power forecast RMSE as a function of look-ahead time. The red lines indicate the RMSE of the raw, uncorrected data, while the blue lines indicate the RMSE of the data corrected with the directional bias corrections scheme. Notice that the bias correction schemes do not lead to improvements in terms of reduced RMSE for all sites and all look-ahead times

When working with measurements from as low heights as 10 meter a.g.l. it is to be expected that the measurements are influenced by local orography and obstacles. The same effects are not likely to be captured by a NWP model with resolution of 4 km * 4 km. Hanna & Yang (2001) argue that some of the errors in NWP models are caused by unrepresented surface properties. Even though the selection of sites has been performed to minimize this problem (see Section 2.1) it would be naïve not to think that it is still present. The directional bias correction will to some extent account for the local effects, but it will not be able to adjust for differences in these effects depending on e.g. wind speed. A common approach to handle these issues is by downscaling; nesting a higher resolution model in the NWP model. This is widely used as the first step in physical and hybrid forecast models (see Section 1.3.1). It is likely that this kind of an approach would reduce the systematic errors caused by orography and potentially also obstacles (see examples in e.g. Möhrlen 2004, Davis *et al.* 2010 and Hashimoto *et al.* 2007). Möhrlen (2004), however, also finds that an increase in horizontal resolution comes at the expense of higher risk of phase errors, and as a consequence of this the overall improvements in e.g. RMSE are limited.

Chapter 4 – Group forecasts

In the previous chapter the focus was on analysing and reducing the systematic parts of the single site wind power forecast error. Here, this is expanded into looking at the overall forecasts error of groups of sites. For many end users of forecasts, like transmission system operators (TSOs) (see e.g. B. Lange *et al.* 2006, Holttinen 2004 and Holttinen *et al.* 2013) and electricity traders (see e.g. Parkes *et al.* 2006), the main forecast quantity of interest is the combined electricity production affecting a section of the distribution grid or a bidding area. Forecasts of the aggregated wind power output of groups of sites will then be more relevant than single site forecasts.

Section 4.1 investigates the reductions in relative forecast error that can be obtained through forecasting the lumped output of groups of sites and which factors that contribute to large reductions of the forecast error. Section 4.2 investigates spatio-temporal dependencies of the forecast errors within and between groups of sites. The section is partly based on Revheim & Beyer (2012). Section 4.3 discusses how the findings on spatial and temporal dependencies can be used to obtain further reductions of group forecast errors. Sections 4.3.1 - 4.3.3 are partly based on Revheim & Beyer (2012) and Section 4.3.4 on Revheim & Beyer (2013a). Section 4.4 draws some general conclusions on the benefits of group forecasts and how different models can maximize those benefits. All examples and figures in the chapter are for forecast with 1 hour look-ahead time.

4.1 Spatial smoothing effects

The term *spatial smoothing effects* is used to describe the reduction in relative forecast error that is obtained by forecasting the lumped output of multiple, geographically dispersed sites compared to forecasting the output of a single site. By integrating over a larger region, the forecast errors and fluctuations of the single site measurements to some extent cancel out. The existence of spatial smoothing effects is well recognized and have been studied and demonstrated for Northern Germany (Focken *et al.* 2002b; M. Lange 2003), UK (Parkes *et al.* 2006), Spain (Parkes *et al.* 2006), Eastern Canada

(Han & Chang 2010), Northern Japan (Nanahara *et al.* 2004) and Denmark (Girard & Allard 2013).

From introductory statistics it is known that the sum of the variance of two or more sites depends on the variances as well as the covariance of the sites (see e.g. Johnson & Wichern 2002). In terms of spatial smoothing effect this means that the magnitude of the reduction in forecast error for a group of sites depends on the cross-correlation of the forecast errors of the group members. For a more thorough explanation including formal definitions it is referred to Focken *et al.* (2002b). In Focken *et al.* (2002b) it is shown that two main drivers of the spatial smoothing effect are the number of sites within the region and the spatial size of the region. It is intuitive that the cross-correlation will decrease with increased distance, but it is likely that local conditions, like the terrain complexity and local meteorological conditions, also will have an influence on this relation. In the following the spatial smoothing effects, including the influence of the number of sites within a group and the spatial size of the group, will be validated for the case western Norway.

4.1.1 Effect of the spatial group size

To assess the influence of spatial group size, first the cross-correlation of the forecast errors for all possible pairs of the 43 sites is calculated. The distances between the sites are calculated from the coordinates of each measurement site. In figure 4.1 the cross correlation of each pair of sites is plotted against the distance between the sites. Not surprisingly the pairwise cross-correlations are site specific; hence the variation on the y-axis. Nevertheless, there is a clear pattern that the cross-correlations decrease rapidly for the first 100 km, thereafter the reduction is declining towards 400 km, and from 400 km an upwards the cross-correlations reaches a stable level slightly above 0. The general pattern is the same as was found by Focken *et al.* (2002b), but for this case the cross-correlations decrease more rapidly with distance. This can most likely be explained by the differences in the complexity of the terrain.

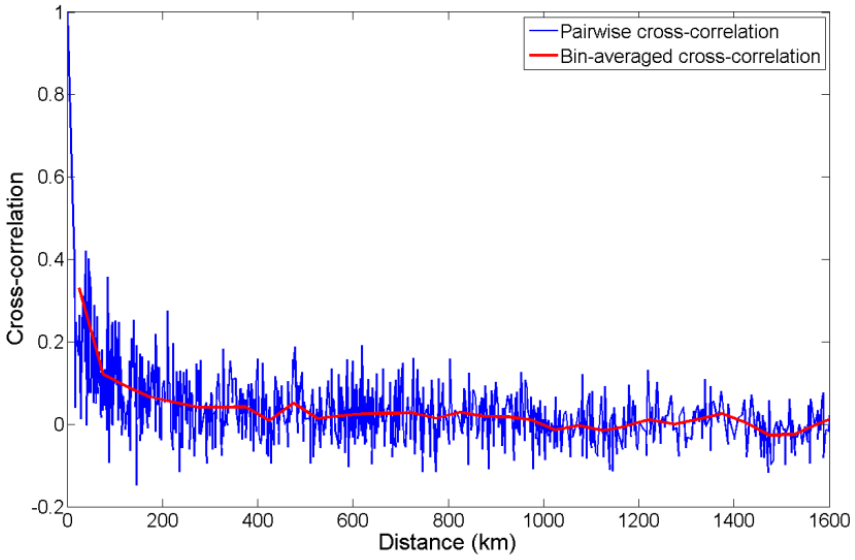

Figure 4.1 – Cross-correlation of forecast error for all possible pairings of the 43 sites plotted against distance between the sites (blue line) and the same cross-correlations averaged over 50 km bins (red line). The vertical deviations are caused by site-to-site variations in the pairwise cross-correlations. The graph shows that the cross-correlation is decreasing rapidly for the first 100 km, thereafter there is a decline in the decrease towards 400 km, and from 400 km and upwards the cross-correlation reaches a stable level slightly above 0.

In order to investigate how the reductions in cross-correlation with increasing distance translate to reduction in group forecast error, 10 000 groups of n sites (n = 2,…,8) are randomly selected from the 43 sites. For each of the groups the ratio between the standard deviation of the group and the mean standard deviation of the single sites are calculated ("smoothing ratio", lower value indicates a higher error reduction). The maximum distance between the sites in the groups (group size) is then averaged over 100 km bins. Notice that the group size measure, maximum distance, does not reflect how the sites are distributed within the groups. A very uneven internal distribution of sites, e.g. that all except one is clustered together, would most likely lead to less error smoothing than an even internal distribution. Considering the large number of randomly selected groups (10 000) it is however not likely that any of the bins are suffering from this potential problem. Figure 4.2 shows the spatial smoothing ratio for groups consisting of 2 to 8 sites plotted against the binned group sizes. For groups of all numbers of sites the same general pattern is found; a rapid decrease in the smoothing ratio for the first few hundred kilometers and thereafter a slower decline towards a stable level. For how many kilometers the rapid decrease lasts depends on the number of sites in the groups, from ~400 km for groups of 2 sites increasing to ~1000 km groups of 8 sites. Except for the groups consisting of 2 sites only, none of

the groups shows signs of the smoothing ratio reaching a stable level. It is also seen that there is a clear decrease in the smoothing ratio with an increasing number of sites in the groups. There are indications that the reduction caused by adding extra sites is decreasing with the number of sites in the groups. This will be explored further in Section 4.1.2.

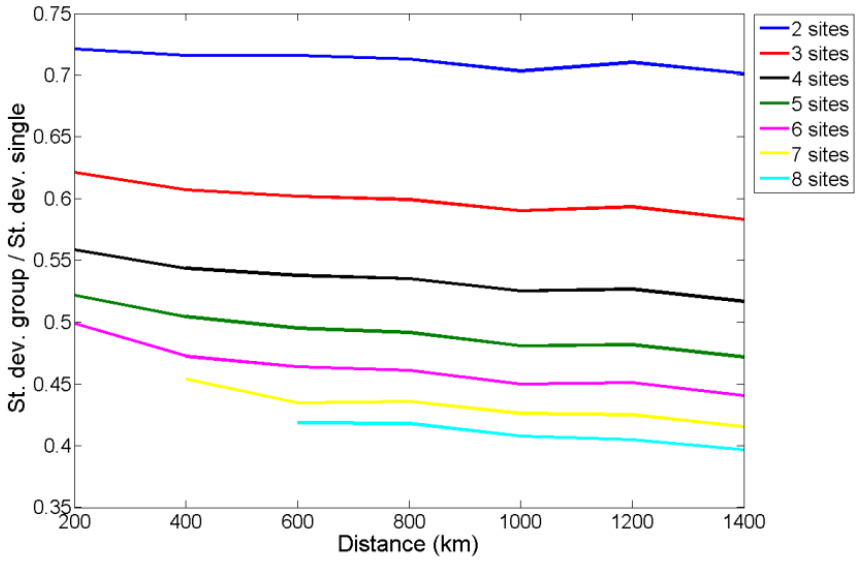

Figure 4.2 – Ratio between the standard deviation of groups of *n* (*n*= 2,…,8) members and the mean standard deviation of the single site members of the groups. For each number of group members 10 000 groups are built by randomly selecting from the 43 sites. The maximum distance between the group members are averaged over 100 km bins. Groups of all numbers of members show a rapid decrease for the first few hundred kilometers. The gain in smoothing ration from adding additional sites to the groups is decreasing with increasing number of group members.

4.1.2 Effect of the number of group members

Figure 4.2 indicated that the additional spatial error smoothing effect gained by adding more sites to a group of fixed spatial extent was decreasing with the number of sites already in the group. This makes logical sense as adding more sites means that the intra-site distances will be reduced, and thus that the average cross-correlation between the group members will increase (following the results from Figure 4.1). Focken *et al.* (2002b) refers to this as a *saturation level* and shows that the saturation level is decreasing with increasing spatial extent of the group.

Figures 4.3 shows the spatial smoothing effect for groups with a spatial extent of respectively 600–800 km and 1000–1200 km as a function of the number of group members. The groups are randomly selected, as described in Section 4.1.1, but here the

spatial extent has been kept constant and the numbers of group members have been varied. Missing values indicate that no combinations of *n* sites are possible within the defined limits for the groups' spatial extent.

The smaller groups (600-800 km, shown in red in the figure) show clear signs of a decline in the smoothing effect when adding more sites to the groups, but it is not possible to conclude that a saturation level is reached. From the 43 sites available it is not possible to create groups of more than 11 members with a spatial extent of 600-800 km, hence can it not be examined when a saturation level is reached. The larger groups (1000-1200 km, shown in blue in the figure) also shows a tendency of the smoothing ratio flattening out for the larger numbers of group members, but also here there is a decline throughout the graph. It therefore cannot be concluded with certainty when a saturation level will be reached. With the 1200 km upper limit it is not possible to construct groups of more than 16 members from the available data.

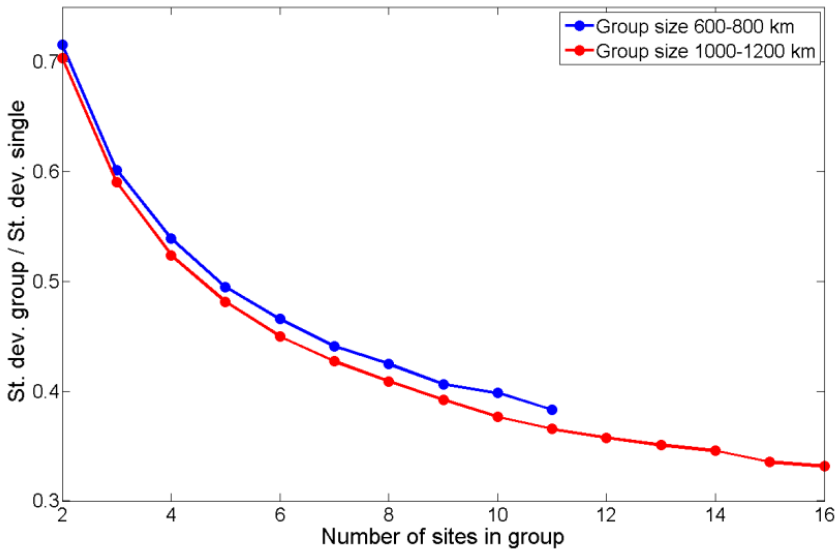

Figure 4.3 – Ratio between the standard deviation of groups and the mean standard deviation of the single site members of the groups. The spatial extent of the groups is fixed (600-800 km (red line) and 1000-1200 km (blue line)) and the number of sites within the groups is varied on the x-axis. Both group sizes show a clear decline in the ratio when adding more sites to the group, and both show also signs of flattening out for the larger numbers of group members. As it is not possible to create groups of more than respectively 11 and 16 sites within the group extent limits it is however not possible to conclude that saturation levels are reached.

4.2 Spatio-temporal dependencies within and between groups

In addition to the spatial dependencies discussed above, the forecast errors are also subject to *temporal dependencies*. In this section the temporal dependencies, and how they interact with the spatial dependencies, are investigated. The temporal dependencies can be divided into two types; dependencies within groups of sites (auto-correlations) and dependencies between groups of sites (cross-correlations). The dependencies within the groups give a measure of how much information knowledge about previous forecast errors for one group, for example the forecast errors of the last hour, contains about the forecast errors at a later time for the same group. The dependencies between groups give a measure of how much information knowledge about previous forecast errors at one group contains about the forecast error at another group at a later time.

Focken *et al.* (2002b) show that the auto-correlation of groups of sites is higher than the auto-correlation of single sites for a case from Germany. Tastu *et al.* (2011) show that the auto-correlation of a group of sites in Denmark is quickly declining, but still significant for delay times up to 3 hours. Tastu *et al.* (2011) also show that there is a clear cross-correlation between neighbouring groups of sites and that the magnitude of the cross-correlation and the time-lag of the maximum cross correlation are heavily dependent on the wind direction. Hence is there a propagation of forecast errors which is depending on the wind direction. In the following the findings from Germany and Denmark regarding the within-group auto-correlations and between-groups cross-correlations will be validated for the case western Norway. Special attention is paid to how the local conditions influence the duration of the auto-correlations and magnitude and propagation of the cross-correlations.

4.2.1 Formation of groups of sites

In order to assess the cross-correlations between groups, and go into details on how the wind direction affects this, the 43 sites needs to be organized in a set of permanent, mutually exclusive groups. The most obvious and natural way of grouping the sites would be to allocate the sites into the regions used by the Norwegian TSO Statnett, which also corresponds with bidding areas used by the Nordic power market Nord Pool (see map in Statnett 2013). From an operational perspective this would be

preferable, as these are the regions where the use of forecasts for e.g. grid balancing and electricity trading will take place. However, this would lead to only four regions covering the entire western coast of Norway, of which the northernmost would have a distance from south to north of more than 1000 km and cover 25 out of the 43 sites. As the focus of this thesis is on forecast methods rather than practical implementation of forecasts it would be beneficial to have more groups, more evenly sized groups and groups with a lesser spatial extent. The groups used in this thesis do therefore not have a direct connection to the assumed practical use of the forecasts, but they will be more suitable for testing and illustrating forecast methods.

The 43 sites were gathered into groups through Ward linkage clustering (see e.g. Johnson & Wichern 2002) based on locations (longitude and latitude coordinates) and the matrix of cross-correlation of the pairwise single-site forecast errors. The data basis for the clustering procedure was chosen to produce as homogenous clusters as possible with regards to location and meteorological conditions. Numerous clustering methods were considered, and Ward linkage chosen based on its ability to produce groups with fairly similar numbers of group members. For a thorough presentation of different clustering methods, including details about their advantages and disadvantages, it is referred to Everitt (1993). A constraint was put on the group formation process to ensure that the groups are clearly separated. This was implemented as weights added to the cross-correlation matrix in order to adjust the relative importance of e.g. location and cross-correlation. Various numbers of groups were tried, and the final number of seven groups chosen as this showed to give a reasonable compromise of clearly separated groups, relatively even numbers of sites within the groups and the spatial spread of the groups. The seven groups and their location are shown in Figure 4.5. The groups and numbering in Figure 4.5 will be used in all examples throughout the remaining of this thesis.

Table 4.1 – Key statistics for the seven groups in Figure 4.5. All groups benefit from a substantial spatial smoothing effect, reducing the standard deviation of the forecast errors by 40 % to 54 %.

	Spatial extent of group	Mean st. dev. single site	St. dev. group	St. dev. group / mean st. dev. single
Group 1	276.5 km	0.23	0.11	0.48
Group 2	250.0 km	0.28	0.15	0.53
Group 3	141.9 km	0.26	0.15	0.60
Group 4	355.2 km	0.28	0.13	0.46
Group 5	138.9 km	0.28	0.17	0.59
Group 6	204.4 km	0.29	0.14	0.48
Group 7	245.9 km	0.28	0.16	0.56

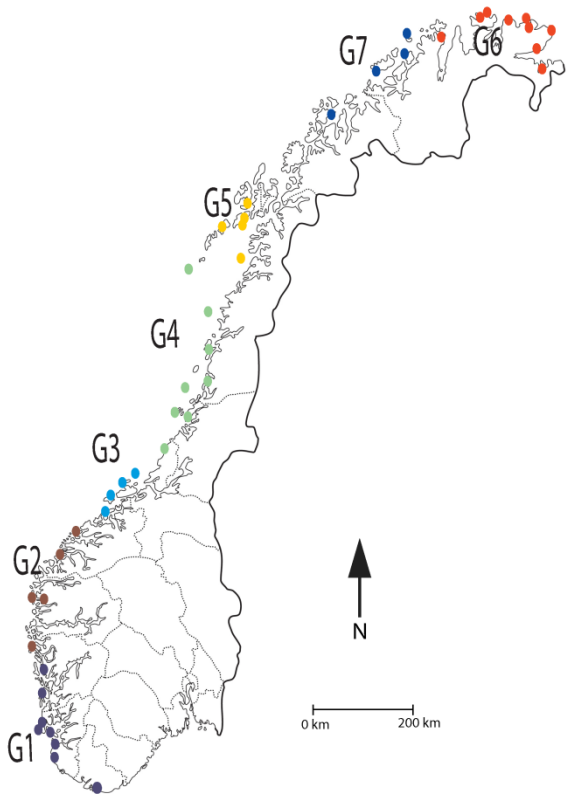

Figure 4.5 – Map of the 43 sites split into 7 groups. The groups are created through a Ward linkage clustering process based on location of the sites and the cross-correlation of the wind power forecast errors. The group numbers in the map (G1-G7) will be referred to in examples throughout the chapter. Map from http://atlas.cappelen.no, © J.W: Cappelens Forlag AS, 2005

4.2.2 Temporal dependencies within groups

The temporal dependencies within the groups, normally referred to as the *auto-correlations*, measure the degree of similarity between a time-series of measurements and a time-lagged version of the same time-series over successive time intervals. Here it is looked at time series of group wind power forecast errors, hence will the auto-correlation measure how much information knowledge about the forecast error at one time contains about the forecast error at a later time. Auto-correlation is defined on the interval [-1, 1], where -1 indicates a perfect negative dependence, 1 indicates a perfect positive dependence and 0 indicates that no linear dependence is present. If high positive (or negative) auto-correlations are present, knowledge about previous forecast errors can be used to correct later forecasts. Focken *et al.* (2002b) and Tastu *et al.* (2011) have studied the temporal dependencies of the forecast errors within groups of wind farms for respectively Germany and Denmark. Focken *et al.* (2002b) found noticeable auto-correlations for time-lags of 24 hours. Tastu *et al.* (2011) finds significant auto-correlations for time-lags up to 3 hours.

Figure 4.6 shows the auto-correlation for the seven groups in Figure 4.5 for time-lags up to 24 hours. It is seen that there exists a clear auto-correlation of the forecast errors for the first hours of time-lag for all groups. There are some differences between the groups, Groups 4 and 7 having higher auto-correlations for time-lags of more than 12 hours and Group 3 having a lower auto-correlation for time-lags lower than 20 hours. For the ideal forecast model the forecast errors would be completely random, i.e. would the auto-correlation be 0 for all time-lags. When this is not the case here, this is a sign that information regarding the temporal dependencies within the groups (the auto-correlations) can be used to reduce the forecast error of the group forecasts.

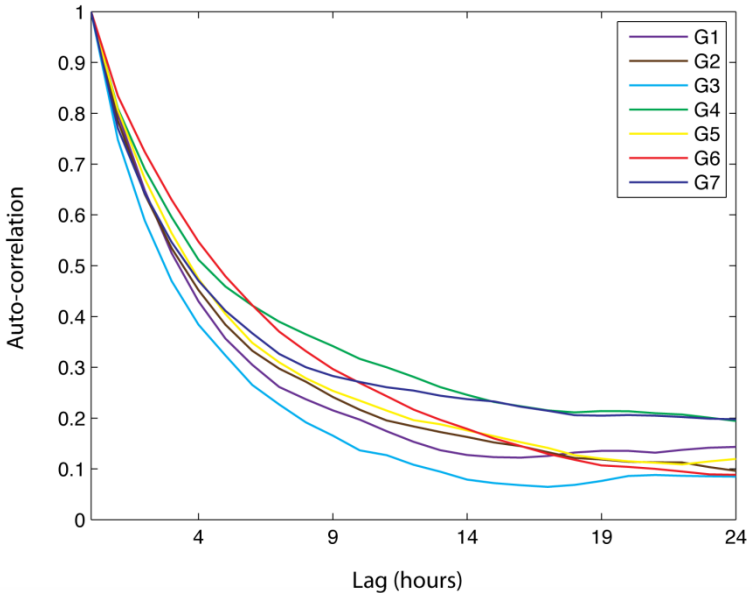

Figure 4.6 – Auto-correlations of the time-series of wind power forecast errors for Groups 1-7 for time-lags up to 24 hours. The group numbers are as shown in Figure 4.5. All groups show a rapid decrease in the auto-correlation for the first 5 hours, flattening out towards 24 hours. Some differences between the groups are noticeable, Groups 4 and 7 having the highest auto-correlations and Group 3 having the lowest auto-correlations.

4.2.3 Spatio-temporal dependencies between groups

In addition to the temporal dependencies within the groups it is also likely to believe that there are dependencies between the groups, normally referred to as *cross-correlations*. These can be calculated in the same way as the auto-correlations, only here the time-lagged time-series comes from another group. As for the auto-correlations, the cross-correlations measure how much information knowledge about the forecast errors of one group at one time contains about the forecast errors of another group at a later time. The cross-correlations are defined on the interval [-1, 1], where -1 indicates a perfect negative dependence, 1 indicates a perfect positive dependence and 0 indicates that no linear dependence is present. If a high positive (or negative) cross-correlation is present, knowledge about the previous forecast errors of one group can be used to correct later forecast for another group (the one for which the cross-correlation is high). Temporal dependencies between groups of wind farms have earlier been studied by Tastu *et al.* (2011) for a case from Denmark, in which it was found that the cross-correlation of a focus-group and four other groups were clearly positive for the first few hours. It was also found that for two of the groups the highest

cross-correlations were found for 1-2 hours of time-lag, indicating a spatio-temporal propagation of the forecast errors.

Here, the cross-correlation between all pairs of groups for time-lags up to 24 hours is calculated. With one exception (Group 6, see Figure 4.5) all the groups show a clear cross-correlation with their neighbouring groups, but of varying magnitude. The exception would be interesting for further studies, but won't be pursued further here. The strongest cross-correlations are found for Group 4. This group will be used as an example for the remainder of the chapter. From Figure 4.7 it is clear that Group 4 has the highest cross-correlation with Group 3, followed by Groups 5 and 7. This is an intuitive result as these are neighbouring groups. Group 4 is also positively correlated to the rest of the groups, but the cross-correlations are of less magnitude. The cross-correlation is highest at time-lag 0, but especially for Group 3 the cross-correlation remains high for time-lags up to several hours.

As with the auto-correlations, the ideal forecast model wouldn't show any clear patterns in the cross-correlation of the forecast errors. The present cross-correlations thus are a sign that there is a potential for improvements of the group forecasts through forecast models including information about the temporal dependencies between the groups.

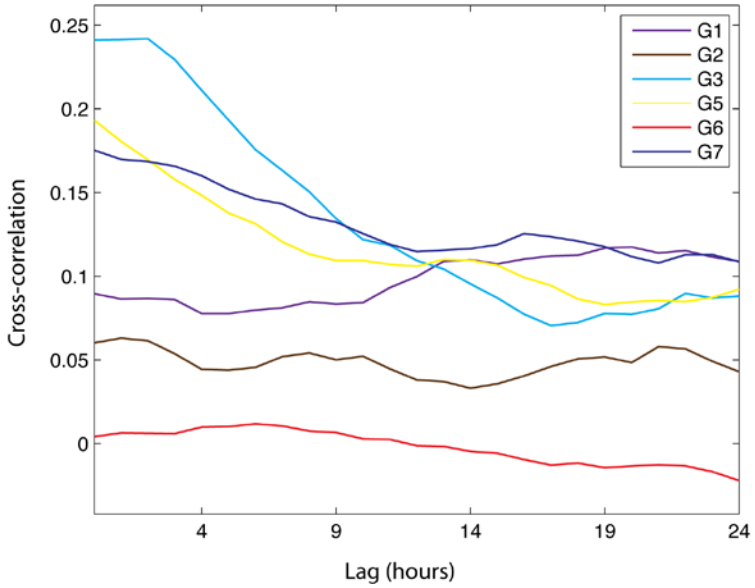

Figure 4.7 – Cross-correlations between the time-series of forecast errors for the power output of Group 4 and Groups 1-7 for time-lags up to 24 hours. The group numbers are as shown in Figure 4.5. As expected the cross-correlation is highest between Group 4 and the two neighbours Groups 3 and 5.

4.2.4 Influence of wind direction on spatio-temporal dependencies

As a further refinement of the analysis of the dependencies between groups, the influence of wind direction on the cross-correlation of the forecast errors is studied.

For this purpose the lumped group wind power forecasts are divided into groups dependent on the forecast wind direction in the respective groups. Four wind sectors are defined, 0º to 90º, 90º to 180º, 180º to 270º and 270º to 360º. The sectors are here defined on the basis that they are easy to relate to. The possibilities of optimizing the definition of the wind sectors on basis of the forecasts will be discussed in Section 4.4.3. For each wind sector the cross-correlations of the forecast error between group of study and the neighbouring groups is calculated. As noted in Section 4.2.3 Group 4 from Figure 4.5 will be used in the examples.

Figure 4.8 shows that the highest cross-correlations between Groups 4 and 3 are found for wind sector $[90°, 180°\rangle$ with time-lags of 1 - 3 hours. This is a sign that there propagation of the forecast errors might be present (forecast errors propagating from south to north). The results correspond well with the results of the general cross-correlation shown in Figure 4.7. Wind sector $[180°, 270°\rangle$ shows the same pattern, but with some additional hours delay and a maximum of less magnitude. The last two sectors also show a positive cross-correlation, but either quickly decreasing or off less magnitude.

Figure 4.9 shows that Groups 4 and 5 have the highest cross-correlation for wind sector $[180°, 270°\rangle$ at time-lag 0. As with Group 3 all the sectors have a positive cross-correlation, but no clear sign of increasing cross-correlation for the first hours is found.

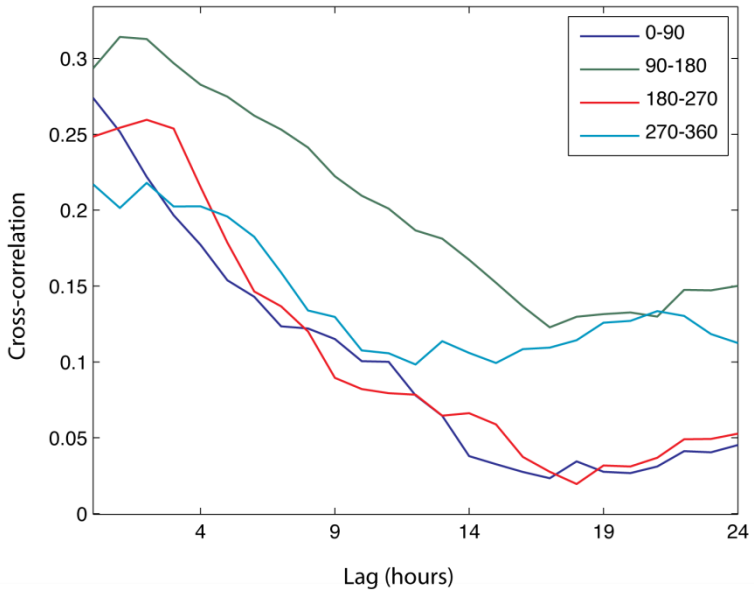

Figure 4.8 - Cross-correlations between the time-series of forecast errors for the power output of Group 4 and Group 3 split on forecast wind direction for Group 4. The group numbers are as shown in Figure 4.5. It is seen that there are clear differences in cross-correlation depending on the wind direction. Wind directions between 90° and 180° gives the highest cross-correlations, and shows also signs of a 2-3 hour delay in the peak cross-correlation. In this situation Group 4 will be downwind of Group 3.

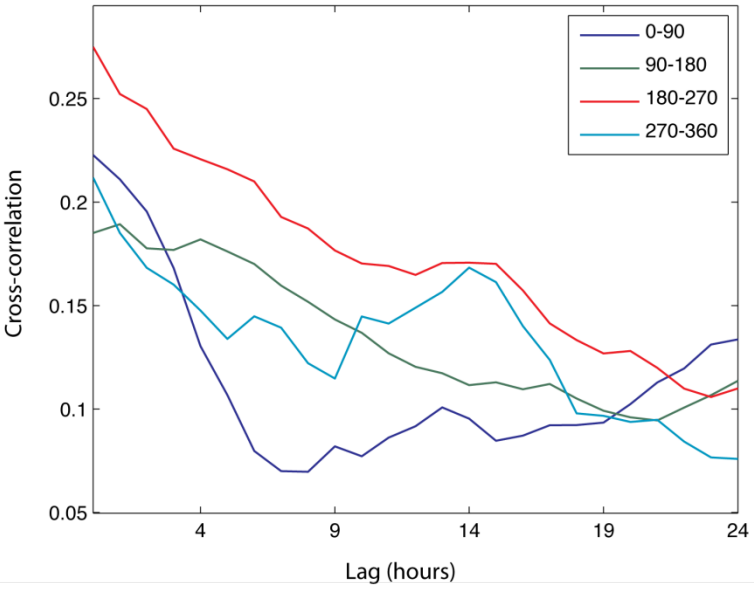

Figure 4.9 - Cross-correlations between the time-series of forecast errors for the power output of Group 4 and Group 5 split on forecast wind direction for Group 4. The group numbers are as shown in Figure 4.5. It is seen that there are clear differences in cross-correlation depending on the wind direction. Wind directions between 180° and 270° gives the highest cross-correlations. In this situation Group 4 will be downwind of Group 5.

4.3 Spatio-temporal regression models

In the previous sections it was shown that there are profound spatio-temporal dependencies within (see Figure 4.6) and between (see Figure 4.7) the groups of sites that are neither accounted for in the NWP forecasts or in the correction schemes presented in Chapter 3. It was also found indications that the forecast errors are transferred from one group to another, and that the presence of this transfer depends on the wind direction (see Figures 4.8 and 4.9). The focus of this section is on how these spatio-temporal dependencies can be modelled and used to reduce the group forecast errors.

Different types of forecast models benefiting from spatio-temporal information has been demonstrated for Germany (Papke *et al.* 1993; B. Lange *et al.* 2008; Wessel *et al.* 2009), sites in Greece Damousis *et al.* (2004), west-coast USA (Larson & Westrick 2006; Gneiting *et al.* 2006; Hering & Genton 2010) and Denmark (Tastu *et al.* 2010; Tastu *et al.* 2011). There are major differences between the models both in complexity and forecast horizon, but the basic idea behind all is that there is a relation between the forecast error at one place and time and the forecast error at another place at a later time, and that the driving forces of this relationship can be modelled and used to improve the forecasts.

Tastu *et al.* (2011) propose different regression models able of capturing and utilizing spatio-temporal dependencies in regional forecasts. In Tastu *et al.* (2010) it is however pointed out that the validity of the results are limited to the same terrain as in Denmark. As described in Section 1.5 this is far from the situation in Western Norway. Here, two of the models in Tastu *et al.* (2011) are tested on cases from Western Norway. The third and most complex model is left out due to the fact that it involves the modelling of the interaction between wind direction and wind speed. This is an intuitive relationship which most likely will give positive contributions, but with the data available in this thesis it has not been possible to obtain stable estimates of this effect (see Section 3.2).

4.3.1 Linear models

The simplest model that will be used is a linear regression model including an auto-regressive term and one or more explanatory variables. This model aims at using the

information from Figures 4.6 and 4.7 to explain and reduce the forecast errors. The general form of this model is given by:

$$y_t = \beta_0 + \sum_{l=1}^{p} \beta_l \, y_{t-l} + \sum_{i=1}^{n} \sum_{j=1}^{k_i} \beta_{i,j} \, x_{i,t-j} + \varepsilon_t, \tag{4.1}$$

where y_t is explained by its p previous auto-regressive values and by n input variables, each up to time lag k_i, $i = 1,\ldots,n$. ε_t is a random variable with zero mean representing the noise not explained by the model (Tastu *et al.* 2011). The estimation of the model is done with least-squares estimation. For a more thorough presentation of the model, the theory behind and the parameter estimation it is referred to Box *et al.* (2008) or Chatfield (2003).

4.3.2 Regime-switch models

An extension of the basic linear models is to allow the regression coefficients to vary depending on the forecast wind direction. Figures 4.8 and 4.9 show that the magnitude of the cross-correlation not only depends on the time, but also on the wind direction, and this is information it would be desirable to be able to capture in the models. The wind sectors defined in the analysis of the influence of wind direction on spatio-temporal dependencies (see Section 4.2.4) is used as regimes, and different regression models are built for each regime. This gives the model:

$$y_t = \beta_0^{s_t} + \sum_{l \epsilon L_y^{s_t}} \beta_l^{s_t} \, y_{t-l} + \sum_{i=1}^{n} \sum_{l \epsilon L_y^{s_t}} \beta_{i,j}^{s_t} \, x_{i,t-j} + \varepsilon_t, \tag{4.2}$$

where

$$s_t = \begin{cases} 1, if \ \theta_t \epsilon R_1 \\ 2, if \ \theta_t \epsilon R_2 \\ 3, if \ \theta_t \epsilon R_3 \\ 4, if \ \theta_t \epsilon R_4 \end{cases}.$$

θ_t serves as the external signal determining the regime-switches, with t as the time index (Tastu *et al.* 2011). y_t is the response variable, here the one-hour ahead forecast error. $x_{i,t-j}$ are the forecast errors for Group i at lag j and ε_t is a random variable with zero mean representing the noise not explained by the model. $L_y^{(s_t)}$ and $L_{x_i}^{(s_t)}$ are sets of

integers defining the input-lags of the model. The superscript s_t indicates that these sets of integers might be different for the different wind sectors. The $\beta_{j,i}^{(s_t)}$ are the linear coefficients to be estimated for each regime. R_1 to R_4 represents the same wind sectors as defined in Section 4.2.4. The estimation of the model is done by least-squares estimation for each regime. For a more thorough presentation of the model, the theory behind and the parameter estimation it is referred to Tong (1990), Pinson *et al.* (2008) and Tastu *et al.* (2011).

4.3.3 Results from spatio-temporal models

The models presented in Sections 4.3.1 and 4.3.2 are applied to the forecasts from Groups 2, 3, 4, 5 and 7 (see Figure 4.5). Groups 1 and 6 are omitted because they only have neighbour-groups on one side, hence will it not be possible to model the propagation of forecast errors from respectively south and north as one of these directions are without available measurements. For all models the number of external inputs is limited to time-lags up to six hours from the two nearest neighbouring groups and six hours for the auto-correlation term. As in the previous sections, Group 4 will be used for examples.

To avoid over-fitting, the models are evaluated by comparing Akaike weights. This is a parameter for comparison of models penalizing over-parameterized models based on the Akaike Information Criterion (AIC), with a simple and easily understandable interpretation. For a thorough presentation of the properties of AIC-values see Akaike (1974). Using the relation $\mathcal{L}(model|data) \propto exp(-0.5 * \Delta_i)$, where $\Delta_i = AIC_i - min(AIC)$, the Akaike weights are calculated as:

$$w_i = \frac{exp(-0.5*\Delta_i)}{\sum_{r=1}^{R} exp(-0.5*\Delta_r)}, \tag{4.3}$$

where R is the total number of candidate models (Wagenmakers & Farrell 2004). The $w_i$values can be interpreted as the probability that model i is the best model, given the data and the set of candidate models. As a general rule of thumb the confidence set of candidate models should include models with Akaike weights within 10 % of the highest scoring model.

Results from the estimation of the linear regression model (see Equation 4.1) applied to the forecasts for Group 4 is shown in Table 4.2. The table shows the structure of the models together with the resultant RMSE and Akaike weight for the three best performing models (L1, L2 and L3) and the simplest model with a RMSE comparable to that of the best model (L4). *Lags Y* indicate the time-lags used in the auto-correlation term, *Lags G3* indicates the time-lags used for the cross-correlation with neighbour Group 3 and *Lags G5* indicates the time-lags for the cross-correlation with neighbour Group 5. From Table 4.2 it is clear that the differences between the models are very small. Even though the Akaike weights indicate that model L1 (the best performing model) is about 36 times as likely as model L4 (the simplest model) to be the best model, this hardly gives any visible difference in terms of RMSE.

Table 4.2 – Results from the linear regression model. The table shows the structure of the models together with the resultant RMSE and Akaike weight for the three best performing models (L1, L2 and L3) and the simplest model with a RMSE comparable to that of the best model. "Lags Y" indicate the time-lags used in the auto-correlation term, "Lags G3" indicates the time-lags used for the cross-correlation with neighbour Group 3 and "Lags G5" indicates time-lags for the cross-correlation with neighbour Group 5.

Model	Lags Y (hr)	Lags G3 (hr)	Lags G5 (hr)	RMSE	w_i
L1	1, 2, 3	1, 2, 4	1	0.0859	0.0756
L2	1, 2, 3	1, 2	1	0.0859	0.0642
L3	1, 2, 3	1, 2, 4	1, 4	0.0859	0.0374
L4	1, 2, 3	1	1	0.0859	0.0021

Results from the estimation of the regime-switch model (see Equation 4.2) applied to the forecasts for Group 4 is shown in Table 4.3. The table contains the same elements and is structured the same way as Table 4.2. Individual linear regression models are built for each wind sector (see Section 4.2.4), and resultant RMSE and Akaike weights are calculated for each of the sector-dependent models. For each of the wind sectors the results for the best model in terms of Akaike weights (D1) and the simplest model with a comparable RMSE (D2) are shown. Table 4.3 shows that there is a clear difference between the different wind-sectors, both in terms of the regressors included in the models and the sectors RMSE. There also is a slight difference between the "best" and the "simplest" model, but this is still very small and most likely negligible for practical purposes. Both models are an improvement compared to the simple linear models (see Table 4.2), but also these differences are small.

Table 4.3 – Results from the regime-switch models. The table shows the structure of the models together with the resultant RMSE and Akaike weight for the three best performing models in terms of Akaike weights (D1) and the simplest model with a RMSE comparable to that of the best model (D2). "Lags Y" indicate the time-lags used in the auto-correlation term, "Lags G3" indicates the time-lags used for the cross-correlation with neighbour Group 3 and "Lags G5" indicates the time-lags for the cross-correlation with neighbour Group 5.

Model	Direction	Lags Y (hr)	Lags G3 (hr)	Lags G5 (hr)	RMSE	w_i
D1	0-90	1, 3, 4	2, 3, 4	1, 3	0,0711	0,0358
	90-180	1, 2, 3	1, 2, 4	2	0,0848	0,0339
	180-270	1, 2	1, 2	1	0,0914	0,0278
	270-360	1, 2, 4	2	4	0,0913	0,0151
	Overall				0,0853	
D2	0-90	1, 3, 4	2	1, 3	0,0712	0,0122
	90-180	1, 2, 3	1	2	0,0849	0,0133
	180-270	1, 2	1	1	0,0914	0,0053
	270-360	1, 2	1	4	0,0914	0,0098
	Overall				0,0854	

To better assess the performance of the models, and as an additional control for over-fitting, the models are tested on "unseen" data through a cross-validation routine (see e.g. Martinez & Martinez 2002). The data are split in three parts of equal size, from which two parts are used to build the regression models, and the last part are used to evaluate the performance of the models. Results from the cross-validation verification are shown in Table 4.4. From the table it is clear that in terms of forecast RMSE all the regression models give a large improvement compared to the directional-corrected NWP forecast. The reduction in forecast error is between 48.4 % and 49 %. The results also show that the simplest linear model, L4 (see Table 4.2), outperform the more complex models after the cross-validation. This is an indication that the effects captured by the more complex models are too weak to be generalized for new data. It should however be noted that there is a clear connection between the number of observations within each wind sector and the results from the cross-validation, sectors with fewer observations providing poorer results. This could indicate that there is a potential for improvements through a smarter choice of wind sectors. This will be further explored in Section 4.3.4.

Table 4.4 – Cross-validation verification results for the linear regression and regime-switch regression models compared to the original forecast for Group 4. L1, L4, D1 and D2 refer to model numbers used in Tables 4.2 and 4.3. The structure of the models with regards to the auto-correlation and cross-correlation terms (neighbour sites and time-lags included) is the same as in Tables 4.2 and 4.3.

	Forecast	**Linear regression**		**Regime-switch regression**	
Model		L1	L4	D1	D2
RMSE	0,1685	0,0861	0,0860	0.0887	0.0886

4.3.4 Importance of choice of wind sectors

In Section 4.2.4 it was shown that wind direction plays an important role in explaining the cross-correlation of groups of sites (see Figures 4.8 and 4.9). Thus, when building a post-processing model taking into account error propagation this ideally should be direction specific and may be sorted to wind regimes. However, when building such models and validating them against unseen data through cross-validation it was found that simple linear models, i.e. models that are not direction specific, lead to a lower RMSE (see Table 4.4). The main reason for this is assumed to be lack of data outside the prevailing wind directions, making it difficult to obtain stable and reliable parameter estimates for these wind directions. One solution to this problem would be to let the choice of wind sectors be data-driven. In this section the importance of choosing a good definition of wind sectors in direction specific regression models are investigated. It is further looked at the possibilities for data-driven selection of wind sectors. The section is partly based on Revheim & Beyer (2013a), and for a more extensive explanation of the methods used it is referred to this.

To assess the importance of the choice of wind sectors, n (n = 500) sets of random wind sectors (up to 4 sectors), in sum covering 360° with no overlaps, are generated. For each of the n sets of wind sectors regime-switch regression models (Equation 4.2) with various time-lag inputs for the auto-correlation and cross-correlation terms are built and the RMSE of each sector calculated. The lowest RMSE for each individual sector is combined into a measure of the RMSE for each set of wind sectors. This procedure results in n RMSE values, which are considered to be the range in model performance caused by the choice of wind sectors. To safeguard against over-fitting of the models a cross-validation routine is also performed (see e.g. Martinez & Martinez 2002), testing each model on “unseen” data. The n RMSE values from the cross-

validations are similarly considered to be the range of model performance caused by the choice of wind sectors.

The procedures described above involve fitting a very large numbers of models, of which the vast majority is discarded. Even for relatively few hours of time-lag and small *n* the calculations are time-consuming. As an alternative to this, the possibilities of creating data-driven wind sectors are investigated. The idea is that information contained in the data about wind directions, forecast errors etc. can be used to create groups of observations that show a similar behavior in terms of forecast error propagation. The wind directions of the members of each group will span out a wind sector that is used to determine the regime-switches in the regression models.

One way of doing this is by cluster analysis. Two different clustering techniques, complete linkage and Ward linkage, are tested (see e.g. Johnson & Wichern 2002). Complete linkage clustering has the property that it ensures that all members of a cluster are within some maximum distance (similarity) of each other. This is a desirable property for this application, as there is a need for well-defined and distinct wind sectors. Ward, on the other hand, minimizes the *loss of information* from joining two clusters, with *loss of information* defined as the increase in geometrical inner square distance between the members of the resulting clusters. Similarly to the complete linkage method this has a tendency to create well-defined and distinct clusters. A matrix consisting of the forecast errors and wind directions of the region of interest and the forecast error of the neighboring regions for time lags up to 3 hours is used as basis for the clustering processes.

Figure 4.10 shows the RMSE of the best regime-switch regression models for 500 wind sectors when the models are applied to the same data as they are built from (blue line). The RMSE of the sets of wind sectors selected through the clustering procedures are marked with red arrows. The RMSE of the set of wind sectors used in the regime-switch models in Section 4.3.3 (model D2) are included as a reference and also marked with a red arrow. Most of all, the figure shows that the importance of the choice of wind sectors in this example is very limited. The difference between the set of wind sectors that gives the lowest and the highest RMSE is less than 1 %. It should though be noted that even the set of wind sectors that gives the highest RMSE still involves a very slight improvement compared to the simple linear models in Table 4.2. The red arrows indicate that the wind sectors selected by the cluster analysis lead to a RMSE that is comparable or slightly lower than that of the reference wind sectors. The set of

wind sectors selected through the complete linkage clustering is amongst the 1 % of the sets of wind sectors that gives the lowest RMSE. The set selected through the Ward linkage clustering gives the same RMSE as the reference sectors.

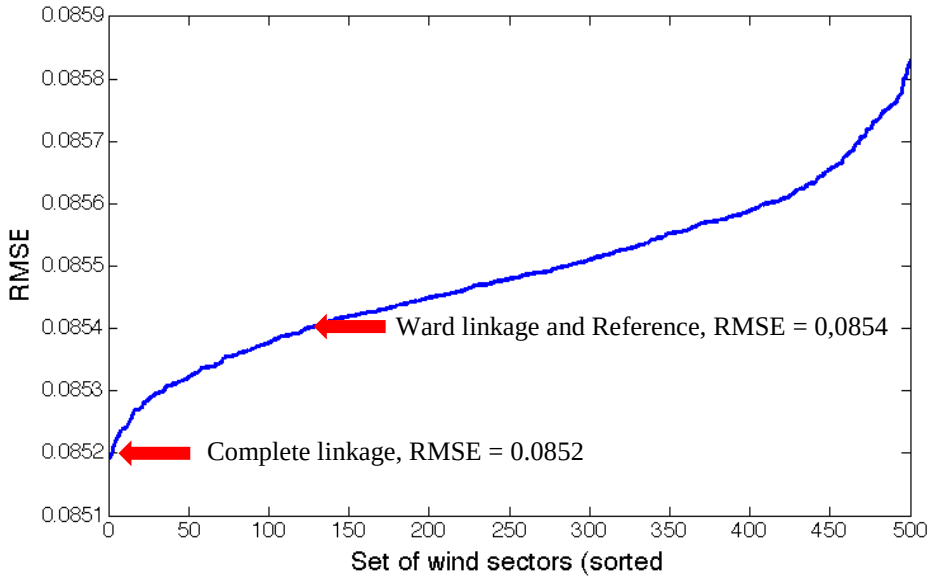

Figure 4.10 – RMSE of the best performing regime-switch regression models for 500 randomly chosen sets of wind sectors. The sets of sectors are sorted from lowest to highest RMSE. All sets contain up to four wind sectors, in sum covering 360 degrees. The RMSE range from 8.52 % to 8.58 % depending on the set of wind sectors. Results for wind sectors defined by the data-driven methods Ward linkage and complete linkage clustering and for a reference scenario (0°-90°, 90°-180° etc.) are shown with red arrows.

Figure 4.11 shows the RMSE of the best regime-switch regression models for the same 500 sets of wind sectors as in Figure 4.10, but here with the models evaluated on "unseen" data through cross-validation. The sets of wind sectors are ordered from lowest to highest RMSE. The ordering of the sets are therefore not exactly the same as in Figure 4.10, but there has not been made very large changes. The features of the figure are the same as in Figure 4.10. As expected there is a slight general increase in the RMSE for all sets of wind sectors. The difference between the set of wind sectors with the lowest and the highest RMSE have increased to more than 21 %. In the right side of the figure (from wind sector set no. 460 and up) there is a very steep increase in RMSE. These sets of wind sectors have in common that the from-shore wind direction (wind from east) is very wide and cover parts of the north-south winds. On the contrary, for the sets of wind sectors that gives the lowest RMSE one sector is collapsed into covering 0° and the three others cover the directions (winds from) north, south-west and south-east. The red arrows indicate that also here the clustering gives

wind sectors with a RMSE that is comparable or slightly better than the reference wind sectors. The Ward linkage clustering now produces the set of wind sectors that gives the lowest RMSE, while the complete linkage clustering gives a slightly higher RMSE than the reference sectors. The differences are however very limited.

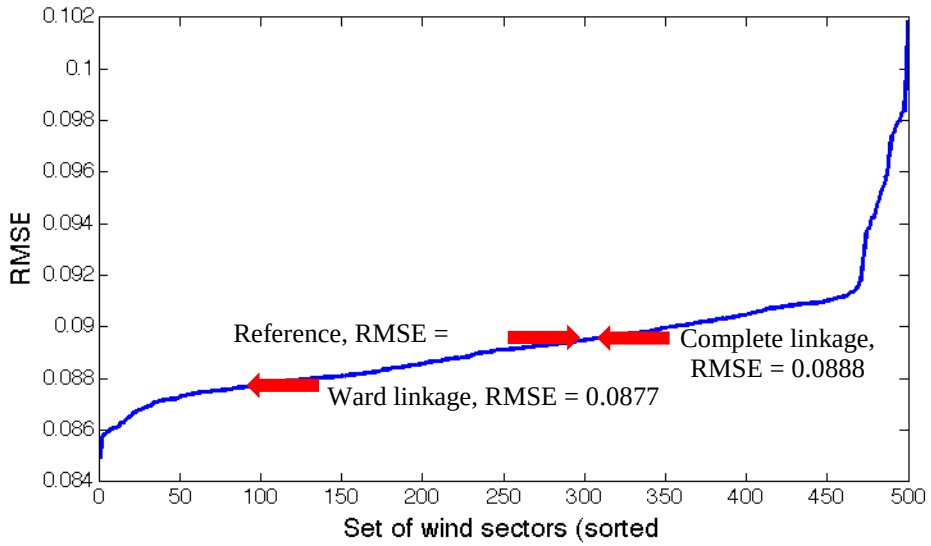

Figure 4.11 - RMSE of the same 500sets of wind sectors as in Figure 4.10, but here with models evaluated on "unseen" data through cross-validation. The sets of sectors are sorted from the lowest to the highest RMSE, i.e. are there some differences in the ordering compared to Figure 4.10. The RMSE range from 8.57 % to 10.19 % depending on the set of wind sectors. Results for wind sectors defined by the data-driven methods Ward linkage and complete linkage clustering and for a reference scenario (0°-90°, 90°-180° etc.) are shown with red arrows.

4.4 Conclusions

In this chapter the spatial and temporal dependencies within and between groups of sites have been studied, and it have been investigated how knowledge about these dependencies can be used to reduce the error of wind power forecast.

First it was demonstrated that there exists a clear spatial error-smoothing effect when considering groups of sites instead of single sites. When integrating over larger geographic areas and multiple sites the spatial smoothing effect leads to significant decreases in the relative wind power forecast error. In Figure 4.1 it was shown that the cross-correlation of the forecast errors for pairs of sites decrease quickly with the distance between the sites. This is well known from e.g. Focken *et al.* (2002b), Siebert (2008) and Girard & Allard (2013), but for the present case the cross-correlation seems to decline more rapidly with increasing distance. This is most likely caused by a more

complex terrain in Norway than in northern Germany and Denmark, but further studies will be needed to verify this. The spatial error-smoothing effect was found to depend on the spatial extent of the groups and the number of sites within the groups, with larger groups in terms of spatial extent or group members having a stronger error smoothing effect.

Thereafter, the presence of spatio-temporal dependencies in the forecast errors within and between groups of sites was identified. In Figure 4.6 it was shown that the auto-correlations are very clear for the first hours of time-lag. Also, the cross-correlations with neighbouring groups were found to be noticeable. When split on wind directions there were found signs of error propagation from upwind to downwind groups. The signs of forecast error propagation are much weaker for these cases than what was found in Denmark by Tastu *et al.* (2011). This has numerous reasons; first of all the groups in this case consist of few sites spread over large areas. This makes it very difficult to define one general and describing measure of the group forecast error, and thus also of how the forecast error propagates. A second reason is that the terrain both in and between the groups is more complex here than in the Danish case, increasing the spatial smoothing effects also for the groups. A third reason is that the groups are aligned in one line from southwest to northeast (see map in Figure 4.5). This makes it impossible to track forecast errors from northwest and southeast, as these directions are without upwind measurements.

Two different regression models were tested attempting to model the forecast errors. When both built from and applied to the full set of forecasts the more complex models came out as superior to the simpler models for both the simple linear and the regime-switching models (see Tables 4.2 and 4.3). The differences between the models were however small, and the number of candidate models within an approximate confidence set large. When tested against “unseen” data through cross-validation the simplest linear model (L4) came out with the lowest RMSE, followed by the more complex linear model (L1), the simplest regime-switching model (D2) and the more complex regime-switching model (D1) (see Table 4.4). This means that the directional modelling of the propagation of forecast errors does not give positive contributions to the forecasts. This is partly caused by the weak signs of error propagation, as mentioned and reasoned for above. Another important reason is lack of data for the wind directions outside the prevailing wind direction, leading to unstable parameter estimates for these wind directions. However, compared to the forecast error of the original forecast (directional bias-corrected NWP, see Section 3.2) all four models

gave a considerable improvement, with a reduction of the forecast RMSE from 16.85 % to between 8.60 % and 8.87 % (see Table 4.4).

The influence of the choice of wind sectors in the regime-switch regression models were assessed by calculating the resultant forecast RMSE for 500 different sets of wind sectors. When assessing the model performance on “unseen” data through cross-validation the differences between a good and a poor choice of wind sectors was found to be more than 21 % (see Figure 4.11). It was further shown that that it is possible to create data-driven wind sectors through clustering procedures. The improvements compared to a reference set of commonly used wind sectors (0°- 90°, 90°- 180° etc.) were however very limited, and for most applications it is likely to be as good a solution to choose wind sectors based on knowledge about prevailing wind directions and local orography.

Chapter 5 – Probabilistic forecasts

In Chapter 4 it was shown that forecasts for groups of geographically dispersed sites benefit from spatial smoothing effects, leading to a lower relative wind power forecast error than that of the single site group members. It is however likely to believe that the forecasts for some sites more accurately reflects the total output of the group. It will then be of interest to give these sites a greater influence over the group forecast. The use of reference sites is not a new concept. Siebert & Kariniotakis (2006) showed for a case from Denmark that the optimum number of reference wind farms for forecasting the combined output of 23 wind farms was as low as 3-5 wind farms. The use of reference sites for up-scaling from single-site forecasts to regional forecasts is also implemented in the operational models Wind Power Prediction Tool (WPPT) (see e.g. Nielsen *et al.* 2007a and Madsen *et al.* 2005) and RegioPred (see Marti *et al.* 2003).

The focus of Chapter 4 was on single-value point forecasts for groups of geographically dispersed sites. For numerous applications it has however been shown that probabilistic forecasts, forecasts of e.g. quantiles or probability density functions, have clear advantages over point forecasts (see references given in Section 1.3.3). In this chapter the possibilities of using an ensemble of single site point forecasts to create probabilistic forecasts for groups of sites is explored. One method for this, that might be able to both give uneven weights to the single sites depending on how well they empirically reflect the lumped group output and to give the a forecast output in the form of probability density functions (PDF), is the ensemble post-processing method Bayesian Model Averaging (BMA). BMA have earlier been successfully applied to make probabilistic forecasts for a wide variety of meteorological variables (see e.g. Sloughter *et al.* 2007, Bao *et al.* 2010 and Chmielecki & Raftery 2010). BMA have lately also received some attention for wind power purposes (see Sloughter *et al.* 2010, Li & Shi 2010 and Courtney *et al.* 2013), but up to now the applications have been limited to wind speed forecasts only.

Section 5.1 gives a brief introduction to the use of ensembles of forecasts in wind power forecasting and explains how the ensembles used in the examples in this chapter differs from the ensembles most commonly found in the wind power forecast literature. Section 5.2 presents the background and theory of the ensemble post-processing method Bayesian Model Averaging (BMA). Changes needed in earlier

applications in order for BMA to be applied for wind power forecasting is discussed. A step-by-step presentation of the building of a BMA model for wind power applications is thereafter given. In Section 5.3 the BMA post-processing model is applied to the 7 groups of sites defined in Section 4.2.1 (see map in Figure 4.1). Section 5.4 draws some conclusions on the applicability of BMA for wind power applications, discusses some weaknesses in the step-by-step procedure presented in Section 5.2 and presents some ideas for further possibilities within the BMA methodology.

5.1 Ensemble forecasts

Ensemble is a commonly used term in wind power forecasting. It normally refers to a group of different forecast (ensemble member forecasts) for the same future quantity. The main idea behind ensemble forecasts is that the ensemble members captures and reflects parts of the uncertainty inherent in e.g. the physical modelling of processes in the atmosphere in NWP forecast models (see Section 2.2), the parameterization of the state of the atmosphere (see Section 2.2) or the general predictability of the future given the situation at present. In wind power forecasting, ensembles of forecasts has two main uses; either the ensemble is averaged over to generate a more accurate and stable point forecast (see e.g. M. Lange *et al.* 2006, Nielsen *et al.* 2007b and Sánchez 2008a), or the distribution of the ensemble members is used to generate a probabilistic forecast (see e.g. Sack *et al.* 2012, Pinson & Madsen 2009 and Giebel *et al.* 2005).

A number of different ways of creating an ensemble is found in the literature. Their aim is to create multiple forecasts for the same quantity in a way that ensures that the single forecasts are as independent as possible (that the correlation between the ensemble members is as low as possible). An extensive and computationally costly way of creating a forecast ensemble is by making forecasts using different NWP models (see e.g. Vidal et al. 2010, Giebel et al. 1999 and Cali et al. 2008). This way the only common property of the ensemble members is the input measurements and potential similarities in the physical modelling of the atmospheric processes. A more common way of creating ensembles is to use one NWP model, but with different parameterizations (see e.g. Möhrlen 2004, Möhrlen & Jørgensen 2006 and Pinson & Madsen 2009). Other variants of ensembles are to use the output of the same NWP model with different initialization times (Landberg *et al.* 2002), to use the output of

different statistical post-processing models (see e.g. Dobschinski et al. 2008) or to build an ensemble from the observed output in similar situations from historic data (see e.g. Alessandrini *et al.* 2015). More examples of ensemble forecasts and uses of ensemble forecasts are found in Giebel *et al.* (2011) and Monteiro *et al.* (2009).

In the examples in the following sections the single sites forecasts within a group of sites are used as the ensemble members. The future quantity of interest is the lumped wind power output of the groups of sites. The term *ensemble* is thus used in the same way as in Focken *et al*. (2002b). This means that the forecast of each single site is treated as an ensemble member forecast of the normalized lumped output of the group of sites, and thus that the size of the ensemble for each group is the same as the number of members of the group. The ensemble members will then be the result of the same NWP model with the same parameterization and initialization time, but they will be geographically dispersed. Limitations in the spatial resolution of the NWP models, including the parameterization of surface orography, will contribute to the independence of the ensemble members (see Section 2.2). A somewhat similar approach was used by Moon *et al.* (2004), where the four surrounding grid-points were used as ensemble members to forecast the wind power at an off-grid site. The dependencies between the forecast errors of geographically dispersed sites were investigated in Section 4.1.1.

5.2 Bayesian model averaging

BMA is a statistical post-processing method for producing probabilistic forecasts from ensembles by weighting and combining competing forecasts. BMA was originally developed as a way to combine multiple statistical models (see e.g. Kass & Raftery 1995, Raftery *et al.* 1997 and Hoeting *et al.* 1999). Raftery et al. (2005) showed how BMA can be used for statistical post-processing of forecast ensembles, producing PDFs of Gaussian-distributed future weather quantities. Sloughter *et al.* (2010) further developed the method for non-Gaussian data fitting wind speed with a gamma distribution. Courtney *et al.* (2013) successfully applied a similar approach to a case from Ireland. Other uses include forecasts of precipitation (Sloughter *et al.* 2007), wind direction (Bao *et al.* 2010) and visibility (Chmielecki & Raftery 2010).

The BMA predictive PDF of a future quantity, here the future wind power production of a group of sites, is a weighted average of the group members PDFs. The weights can be interpreted as posterior probabilities and will here reflect how well the forecast for each single site reflects the normalized lumped group output over a training period. Each single site forecast f_k is associated with a component PDF, $g_k(y|f_k)$, that links the single-site forecast and the group output. The BMA predictive PDF of the future group wind power *y* is a combination of the component PDFs

$$p(y|f_1, \dots, f_K) = \sum_{k=1}^{K} w_k g_k(y|f_k), \tag{5.1}$$

where the BMA weights w_k are based on forecast *k*'s relative performance in the training period (Sloughter *et al.* 2010). *K* is the number of ensemble members (number of single sites within the group) and f_k the single-site forecasts. The BMA weights are probabilities, hence are they non-negative and sum to 1.

For building a BMA model for wind power forecasting there is the need of deciding on:

- A component PDF, $g_k(y|f_k)$, linking the single site forecasts and the observed lumped group production.
- A way of parameterizing the component PDF from the single site forecasts.
- A way of estimating the BMA weights from data from a training period.

This will be discussed in the following sections. As in the previous chapter, the case of Group 4 (see Figure 4.5) will be used in all examples. A detailed map of Group 4, including names of the single site group members, is shown in Figure 5.1. The names of the sites will be referred to later in the chapter.

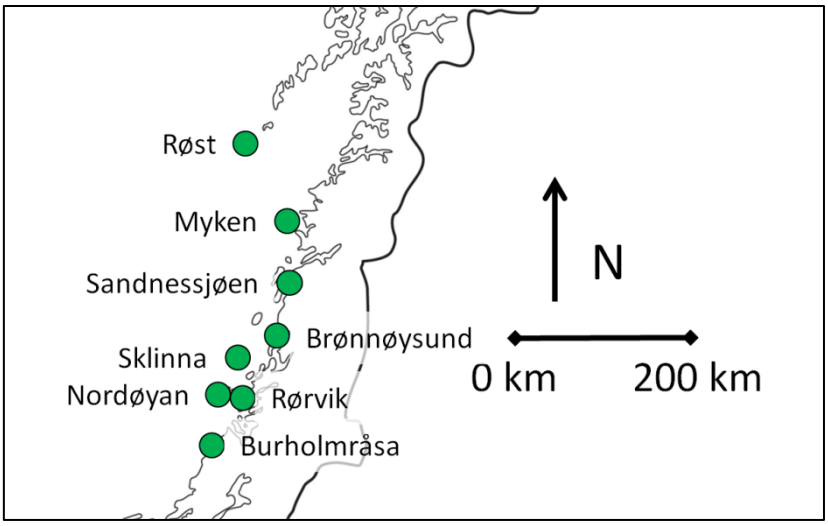

Figure 5.1 – Detailed map of Group 4 from Figure 4.5. The group consist of 8 sites along the Western coast of the middle part of Norway. Map from http://atlas.cappelen.no, © J.W: Cappelens Forlag AS, 2005

5.2.1 Choice of component PDF

In earlier applications of BMA for wind power applications the gamma distribution (see e.g. Miller & Miller 2004, p. 202) has been used to model the distribution of the wind speeds (Sloughter *et al.* 2010; Courtney *et al.* 2013). In Revheim & Beyer (2013b) it was shown that the gamma distribution is a good choice to model the probability density, if the objective is to make a probabilistic forecast of the mean wind speed of a group of sites from the single site wind speed forecasts. When applying the method and Gamma PDF with forecasted wind speeds as input and lumped group wind power as output, it was however found that it severely under-predicted the wind power.

Concerning the model used to characterize the PDF of wind power forecast errors, a literature study a wide variety of models is suggested. Doherty & O'Malley (2005) and Pappala *et al.* (2009) use a Normal (Gaussian) distribution to describe the distribution of the wind power forecast errors, Justus *et al.* (1976), Dietrich *et al.* (2009) and Bradbury (2013) use a Weibull distribution and Bofinger *et al.* (2002), Sánchez (2008b) and Blaudszuweit *et al.* (2008) use a Beta distribution. Lately some more exotic distributions have also appeared, including the Cauchy distribution (Hodge & Milligan 2011), generalized logit-Normal distribution (Pinson 2012), Hyperbolic distribution (Hodge *et al.* 2012), censored Normal distribution (Tastu *et al.* 2012) and a "Versatile" distribution (Zhang *et al.* 2013). There is up to now no general agreement

on one universal PDF for describing the distribution of wind power forecast errors, although many authors agree that the normal distribution in most cases is a poor choice (Hodge *et al.* 2012; Pinson 2012; Blaudszuweit et al. 2008). Hodge *et al.* (2012) notes that there are large site-to-site variations in the distribution of forecast errors, and that this calls for site-specific assessments where the PDF is chosen on the basis of the available data.

The fit of potential component PDFs are tested one site at a time. For each site the observed group wind power is divided into 10 bins depending on the corresponding forecasted wind power (0 % - 10 %; 10 % - 20 % etc.). The data in each bin will then show the span in observed wind power that can be expected e.g. for forecasts between 0 % and 10 % of the power output. For each bin the fit of the PDF is assessed by a quantile-quantile plot (see e.g. Martinez & Martinez 2002) of the observed wind power (conditional of forecast bin) against the theoretical quantiles of the PDF. As the PDFs will also be used to estimate the BMA weights it is seen as beneficial to keep the PDF as simple as possible, and the focus of the investigation is therefore on simple and well-known distributions. Five different PDFs are tested; Normal, Exponential, Weibull, Beta and Gamma (see e.g. Miller & Miller (2004) for a thorough presentation of the distributions). Not surprisingly the Normal, Exponential and Gamma distributions show a general poor fit. With some exceptions both the Weibull and Beta distributions show a good fit to the data, with Beta being the better. The exceptions are most likely caused by a low number of forecasts in the upper bins (especially 70 % - 80 % and 80 % - 90 %) for some of the sites.

Examples of quantile-quantile plots of the Beta distribution fit for three different wind power forecast bins are shown in Figures 5.2 to 5.4. The figures show the fit of the Beta distribution to data from the site Røst in Group 4 (see map in Figure 5.1), but are also representative for the vast majority of the other sites. With a perfect fit, i.e. full correspondence between the distribution of the data and the theoretical distribution, the quantile-quantile plots shall display a straight line with inclination x=y (dashed red line). All three figures show a generally good fit to the data, Figures 5.3 (forecast bin 40 % - 50 %) and 5.4 (forecast bin 90 % - 100 %) being nearly perfect. Figure 5.2 (forecast bin 0 % - 10 %) shows clear, but still limited, deviations. The deviation is caused by a heavy lower tail, again caused by a high frequency of observations with zero production.

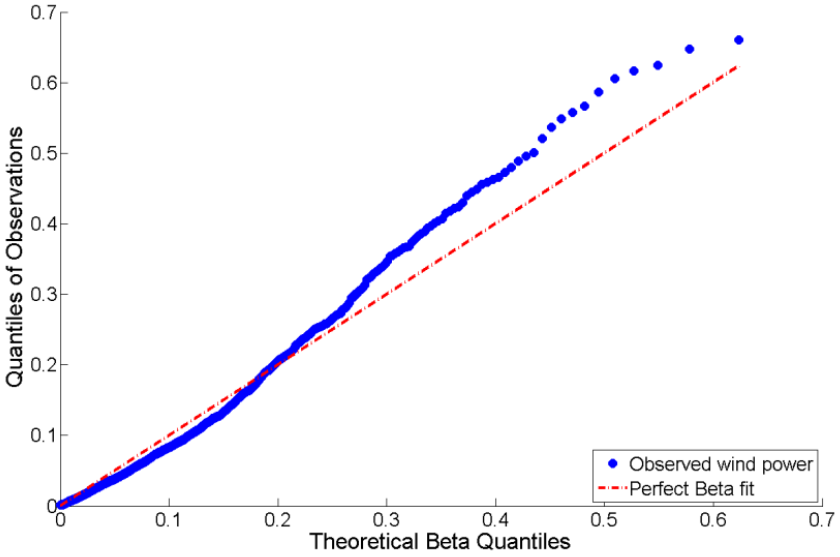

Figure 5.2 – Beta- distribution quantile-quantile plot for the observed wind power conditional on the forecasted wind power being between 0 % and 10 %. The data stem from the site Røst (see Figure 5.1). With a perfect fit the blue dots should be aligned on the dashed red line. The figure shows that the distribution of the data deviates from the theoretical beta distribution, but the deviation is not very severe. The deviation is caused by a high frequency of observations with zero production.

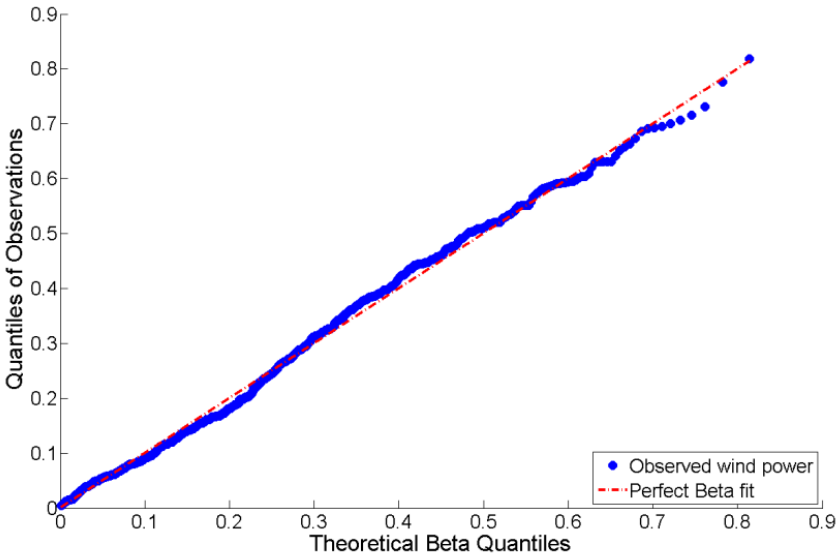

Figure 5.3 – Beta- distribution quantile-quantile plot for the observed wind power conditional on the forecasted wind power being between 40 % and 50 %. The data stem from the site Røst (see Figure 5.1). With a perfect fit the blue dots should be aligned on the dashed red line. The figure shows that the correspondence between the distribution of the data and the theoretical Beta distribution is very good.

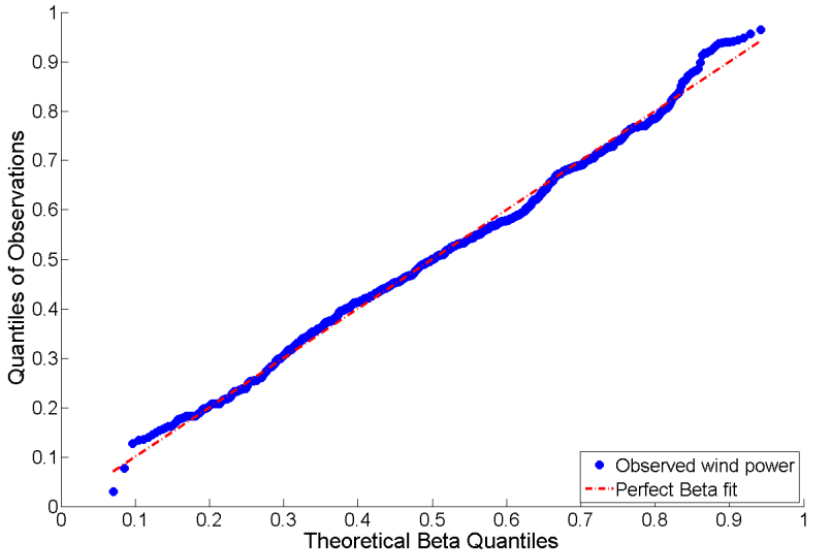

Figure 5.4 – Beta- distribution quantile-quantile plot for the observed wind power conditional on the forecasted wind power being between 90 % and 100 %. The data stem from the site Røst (see Figure 5.1). With a perfect fit the blue dots should be aligned on the dashed red line. The figure shows that the correspondence between the distribution of the data and the theoretical Beta distribution is very good.

Based on the assessments of the quantile-quantile plots the Beta distribution is chosen as the component PDF $g_k(y|f_k)$ in the following. For forecasting normalized wind power the beta-distribution has the advantage that it is limited to the interval 0 to 1. A second advantage is that the distribution parameters can easily be estimated from the data making potentially slow and heavy numerical techniques for parameter estimation unnecessary.

The component PDF, $g_k(y|f_k)$, in Equation 5.1 is then defined as:

$$g_k(y;\alpha_k,\beta_k) = \frac{\Gamma(\alpha_k+\beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} y^{\alpha_k-1}(1-y)^{\beta_k-1}, \tag{5.2}$$

where $0 \leq y \leq 1$, $\alpha, \beta > 0$ and k is an index for the single site. For more information on the properties of the Beta distribution it is referred to e.g. Miller & Miller (2004).

5.2.2 Estimation of the parameters of the component PDF's

With the model for the component PDF in place, the next issue is how to estimate the parameters of the Beta distribution from the wind power forecasts. In Sloughter *et al.* (2010) the parameters of the Gamma distribution are found through an exploratory analysis of the relation between the observed wind speed and the forecasted wind speed. This procedure is here used as the starting point for estimating the required parameters of the target Beta distributions. Thus, two unknown parameters, α_k and β_k (see Equation 5.2), are to be estimated for each single site ensemble member *k*.

As indicated in the previous section the parameters of the Beta distribution can be estimated from the mean and variance of sample data using the following relations:

$$\alpha_k = \left(\frac{1-\mu_k}{\sigma_k^2} - \frac{1}{\mu_k}\right) * \mu_k^2 \tag{5.3}$$

and

$$\beta_k = \alpha_k * \left(\frac{1}{\mu_k} - 1\right), \tag{5.4}$$

where μ is the mean of the data, σ^2 is the variance of the data and *k* is an index for the single site (see e.g. Miller & Miller 2004). In order to estimate the Beta parameters from the data it is therefore sufficient to find a relation between the mean and the variance of the single site forecasts and the group observations.

Figure 5.5 shows the relation between the single site forecasted wind power (represented by the midpoint of the 10 forecast bins from Section 5.2.1) and the means of Beta distributions fitted to the observed group wind power, conditional on the forecast bin. The figure shows all sites in all groups from Figure 4.1. Even though there is a lot of vertical scatter a first order linear trend is very visible (indicated with the dashed red line). The scatter is caused by the many different sites plotted in the same figure and indicates that each site needs to be assessed individually. In Figures 5.6 and 5.7 two of the sites from Figure 5.5, Røst (Figure 5.6) and Brønnøysund (Figure 5.7) are plotted individually. In both figures a clear first order linear relation between the forecast and observed wind power is seen. Similar results as in Figures 5.6 and 5.7 are also found for most other sites. From this the mean parameter in Equations 5.3 and 5.4 can be estimates by

$$\mu_k = b_{0k} + b_{1k} f_k, \tag{5.5}$$

where b_{0k} and b_{1k} are regression parameters estimated from the forecast-observation pairs from the training period and f_k is the forecast wind power production of site *k*. The training period is a sliding window of fixed size, and the parameters are re-estimated for each training period. To enable the estimates to be time-adaptive it is desirable to keep the length of the training period as short as possible. Still, it needs to be sufficiently long to enable stable parameter estimates. Here 30 days is found to be a good compromise for the length of the training period.

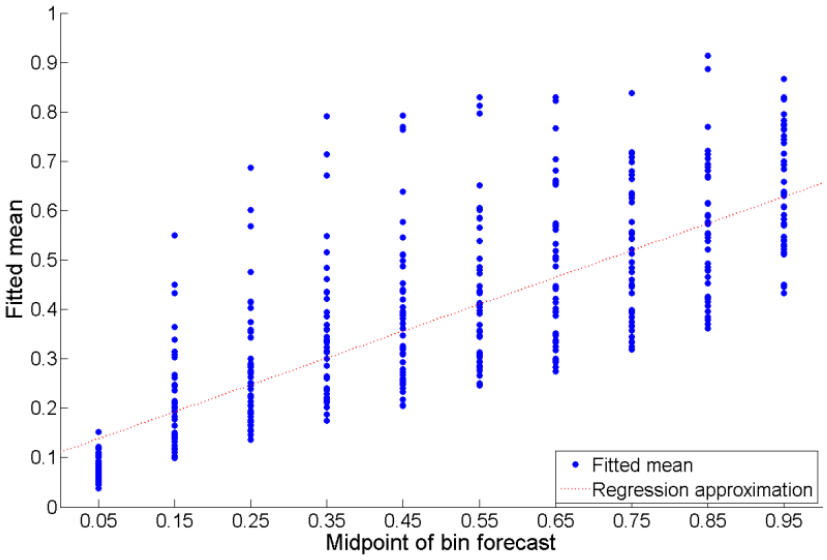

Figure 5.5 – Relation between binned forecast wind power and the mean of Beta distributions fitted to the observed wind power, conditional on forecast bin. All sites from all groups in Figure 4.1 are plotted. A clear first order linear trend is seen (indicated with the dashed red line). The vertical scatter is caused by differences between the sites, indicating that the sites need to be assessed individually.

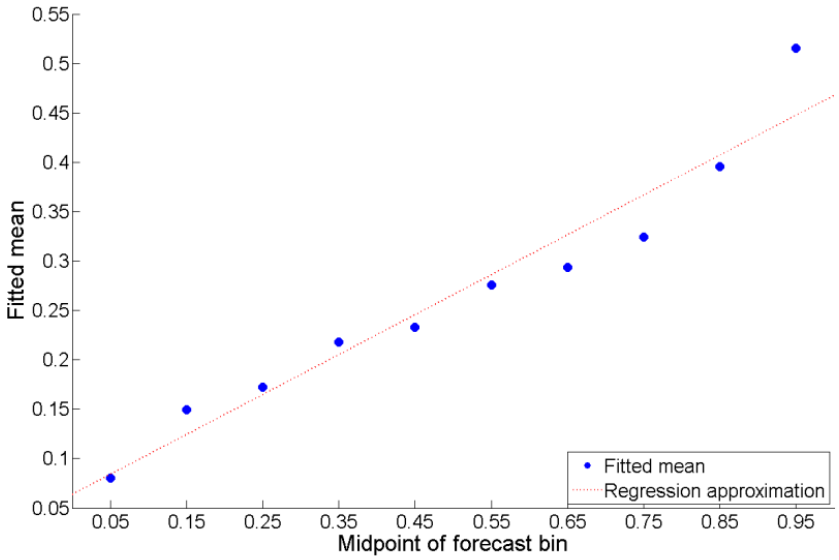

Figure 5.6 – Relation between binned forecast wind power and the mean of Beta distributions fitted to the observed wind power, conditional on forecast bin. The figure shows the site Røst (see Figure 5.1). The plot shows a clear first order linear relation (indicated with the dashed red line).

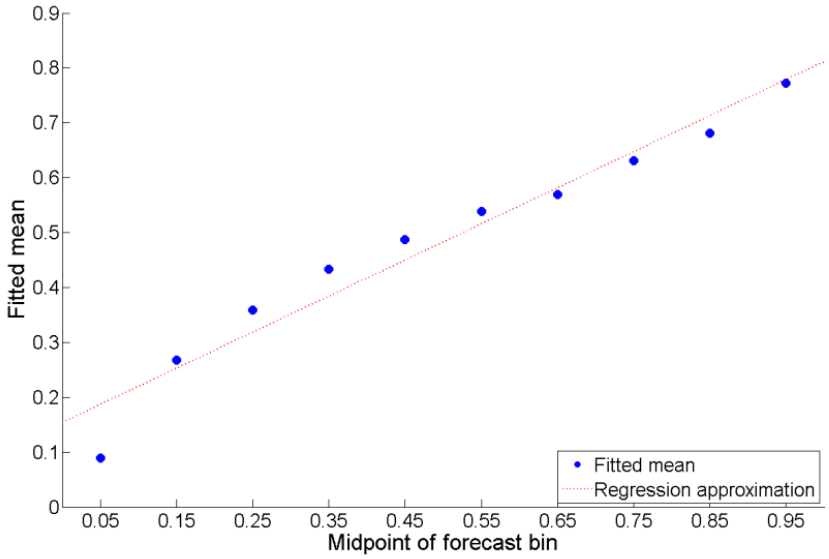

Figure 5.7 – Relation between binned forecast wind power and the mean of Beta distributions fitted to the observed wind power, conditional on forecast bin. The figure shows the site Brønnøysund (see Figure 5.1). The plot shows a clear first order linear relation (indicated with the dashed red line).

Figure 5.8 shows the relation between the single site forecasted wind power (represented by the midpoint of the 10 forecast bins from Section 5.2.1) and the standard deviation of Beta distributions fitted to the observed group wind power,

conditional on the forecast bin. The figure shows all sites from all groups in Figure 4.1. Even though there also here is a lot of vertical scatter, clear signs of a second order linear relation is seen (indicated with the dashed red line).

Considering the shape of the power curve (see Figure 1.1), a second order relation with a peak around 0.5 would seem sensible as the steep middle part of the power curve is likely to inflate the standard deviation. In Figures 5.9 and 5.10 two of the sites from Figure 5.8, Røst (Figure 5.9) and Brønnøysund (Figure 5.10) are plotted individually. Figure 5.9 shows the same second order linear relation as in Figure 5.8, but with the peak at slightly higher forecasted values. Figure 5.10, on the other hand, shows no clear signs of any relations. The two figures give a good description of the relation between the forecasted wind power and the standard deviation of the fitted Beta distributions for the other 41 sites; for some sites there is a clear second order linear relation, while for other sites there is no obvious relation. Unlike for the means it is therefore not possible to derive one expression for how the standard deviation parameter in Equations 5.3 and 5.4 can be estimated from the forecasts. Two different strategies are therefore applied:

1. To derive an expression for the standard deviation with the binned forecast-observation pairs from all sites pooled. This involves fitting a second order expression to the fitted standard deviations shown in Figure 5.8. The expression for this is given by

$$\sigma_k = c_0 + c_1 f_k + c_2 f_k^2, \tag{5.6}$$

   where c_0, c_1 and c_2 are regression parameters estimated from all forecasts and observations from all sites and f is the forecast wind power production of site k.

2. To make a look-up table of the standard deviations for each forecast bin for each site.

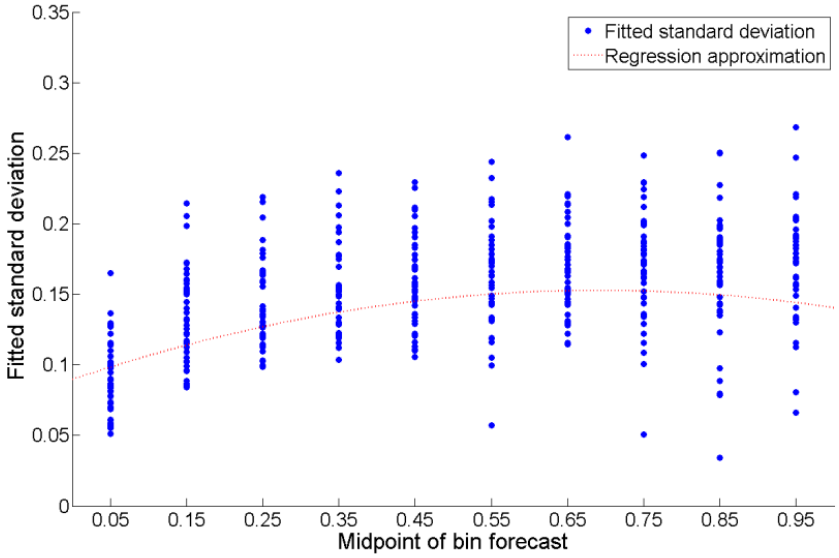

Figure 5.8– Relation between binned forecast wind power and the standard deviation of Beta distributions fitted to the observed wind power, conditional on forecast bin. All sites from all groups in Figure 4.1 are plotted. A clear second order linear trend is seen (indicated with the dashed red line). Considering the shape of the power curve (see Figure 1.1) this seems sensible. The vertical scatter is caused by differences between the sites.

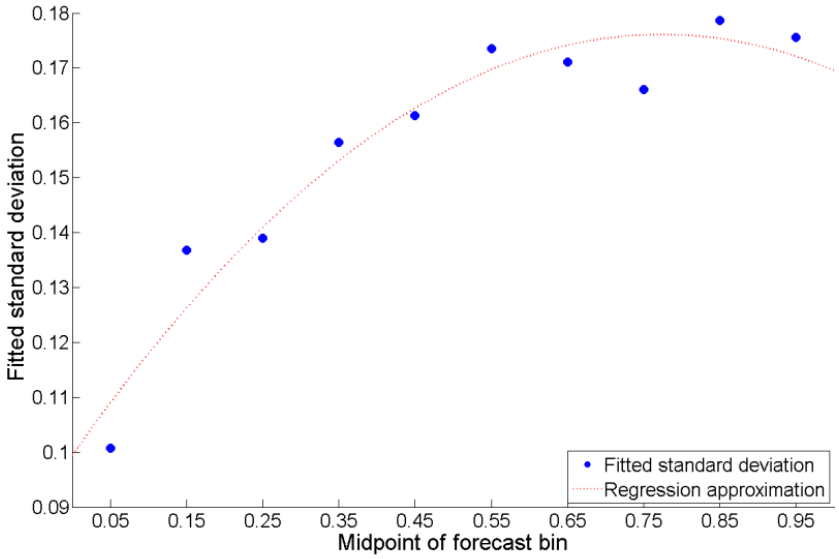

Figure 5.9 – Relation between binned forecast wind power and the standard deviation of Beta distributions fitted to the observed wind power, conditional on forecast bin. The figure shows the site Røst (see Figure 5.1). A clear second order linear trend is seen (indicated with the dashed red line).

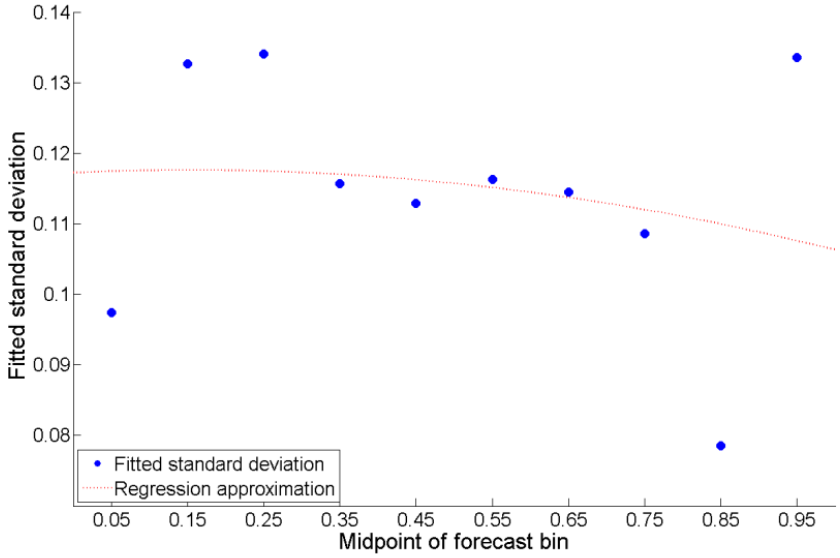

Figure 5.10 – Relation between binned forecast wind power and the standard deviation of Beta distributions fitted to the observed wind power, conditional on forecast bin. The figure shows the site Brønnøysund (see Figure 5.1). No clear relationships between the forecast and the observed standard deviation are seen. The dashed red line indicates the fit of a second order linear relation (as was found in Figures 5.8 and 5.9), but as is clearly seen this does not give a good approximation of the data.

5.2.3 Estimation of BMA weights

With the component PDF and the expressions for how to estimate the PDF parameters from the training data in place, the last part remaining of the BMA model is the BMA weights w_k (see Equation 5.1). The BMA weights reflect forecast *k*'s relative performance in the training period, and as the name indicates acts as weights deciding the importance of the respective sites component PDF on the BMA predictive PDF. Similar to the regression parameters for the mean of the Beta distribution (Equation 5.5), the BMA weights are estimated from the forecast-observation pairs from the training period. The same training periods as for the regression parameters are used (30 day sliding window).

Assuming independence in the forecast errors in space and time the BMA weights can be estimated by maximum likelihood estimation (see e.g. Johnson & Wichern 2002) from the training data through the log-likelihood function

$$\ell(w_1, \dots, w_k; \theta) = \sum_t \log p\left(y_t \middle| f_{1,t}, \dots, f_{K,t}\right) = \sum_t \log\left(\sum_{k=1}^{K} w_k g_k(y|f_k)\right), \quad (5.7)$$

where the sums extends over all times *t* and ensemble members *K* in the training data and θ represents the parameters of the component PDFs (Sloughter *et al*. 2010). $g_k(y|f_k)$ is the component Beta PDF as defined in Equation 5.2, with parameters α_k and β_k as defined in Equations 5.3 and 5.4 and explained in Section 5.2.2. For the Beta distribution normal maximum likelihood estimation is only possible when the observation *y* lies within the open interval 0 to 1. Here lumped group wind power productions of both 0 % (0) and 100 % (1) are present, thus is it not possible to calculate the likelihood function for all values of *y*. This is solved by computing modified likelihoods where the zeroes and ones are treated as if they were respectively left-censored at a value close to 0 and right-censored at a value close to 1, as implemented in the Matlab-function *betalike* (Mathworks 2014).

5.3 Application results

The BMA model is tested on the 7 groups of sites defined in Section 4.2.1 (see map in Figure 4.1) with single site forecasts corrected with the directional bias correction scheme described in Section 3.2. For each group the single site forecasts are used as ensemble members (f_k, with *k* being the number of sites in the group) and the lumped power output from the group is used as the target value *y*. As in the previous examples all measures are normalized to a scale from 0 to 1. For all groups the BMA model is built using the Beta function as component PDF and with parameters estimated as described in Sections 5.2.2 and 5.2.3. It soon became clear that to estimate the standard deviations of the fitted Beta distributions by second order linear regression (Equation 5.6) did not give a good description of the true standard deviation for all sites, and thus that the estimated parameters of the Beta distribution (Equations 5.3 and 5.4) became unreliable (and in some instances out of range). The standard deviations were therefore decided by building site-specific look-up tables (option 2 in Section 5.2.2) with the same 10 bins as was used in the exploratory analyses of Section 5.2.2. For all sites the length of the training period, from which the Beta parameters (except the standard deviations) and BMA weights are calculated, were set to 30 days.

Figure 5.11 shows the unweighted Beta component PDFs for the 8 sites in Group 4 (see maps in Figures 4.1 and 5.1) for the 1 hour look-ahead group forecast for February 14th 2009. The parameters of the beta distribution are fitted on the basis of training data from January 15th 2009 to February 13th 2009 as well as on the site-

specific look-up tables based on all available data (for the standard deviations in Equations 5.3 and 5.4). A good agreement of the PDFs for 7 out of the 8 sites is seen, Røst being the site that clearly deviates from the others. This is not an uncommon situation for the group, and is most likely caused by Røst's location (on a small island far off the coast, see Figure 5.1) that clearly deviates from the rest of the sites in the group.

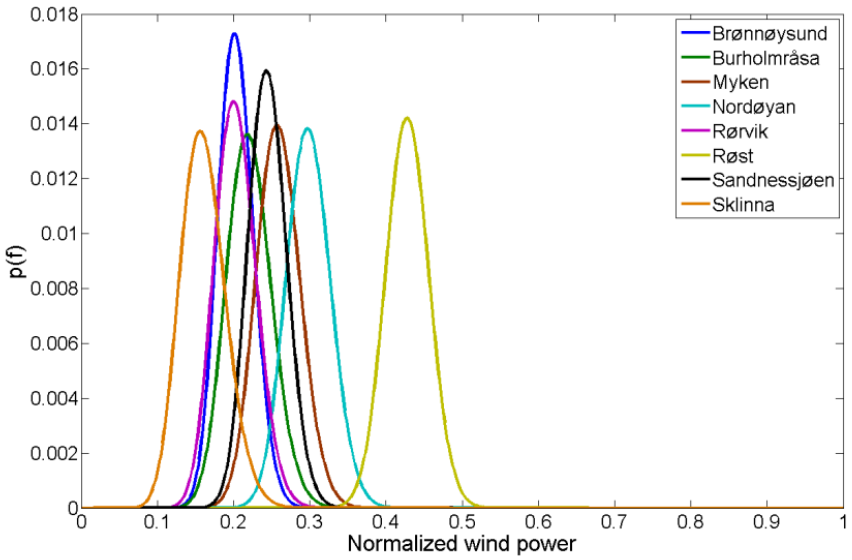

Figure 5.11 – Unweighted Beta component PDFs for the 8 sites in Group 4 (see maps in Figures 4.1 and 5.1) for the 1 hour look-ahead time group forecasts for February 14th 2009. The parameters of the Beta distribution are fitted on the basis of training data from the 30 previous days. The figures show that there is a good agreement of the PDFs for 7 out of the 8 sites, Røst (yellow line) being the one that deviates.

Figure 5.12 shows the component PDFs for the same 8 sites as in Figure 5.11 weighted with the BMA weights w_k reflecting the sites relative performance in the training period, together with the BMA predictive PDF (red line). The true observed lumped group wind power is marked with the scattered black line. The figure shows that five of the sites (Sklinna (orange), Rørvik (magenta), Brønnøysund (blue), Burholmråsa (green) and Nordøyan (cyan)) are found to give a good performance in the training period and are given more weight than the remaining 3 sites. Myken (brown) and Sandnessjøen (black) also have some influence on the resultant PDF, but are clearly weighted down. Røst, as was found to clearly deviate from the other 7 sites in Figure 5.11 is weighted down to practically 0, and can hardly be seen in the figure (just visible along $p(f) = 0$ for normalized wind power between 0.2 and 0.25). As explained in the section above this does not come as a surprise and can be explained with the

location which makes it unrepresentative of the group. The BMA predictive PDF for the group ("Group", marked in red) is the sum of the component PDFs (see Equation 5.1). This is the output of the forecast and is to be interpreted as the probability distribution of the lumped group wind power given the single site forecasts. The distribution has its peak just right of the true observed power (scattered black line).

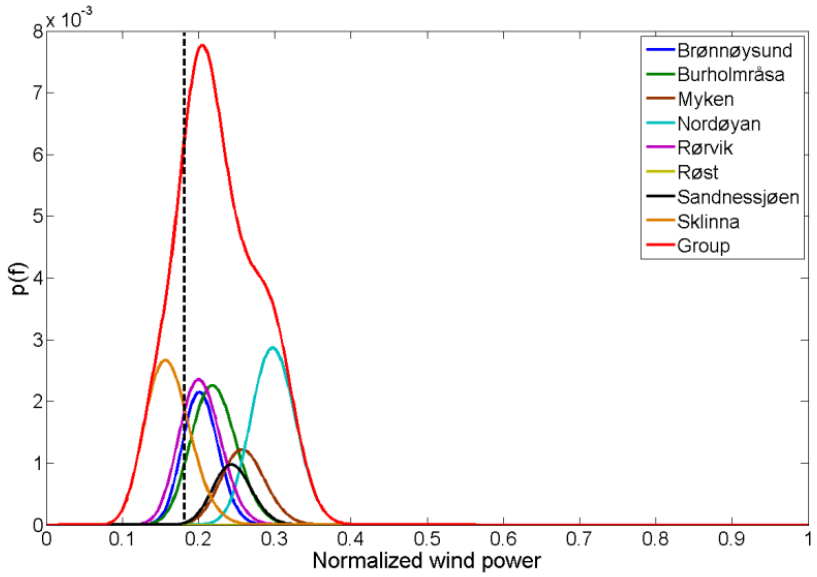

Figure 5.12 – Weighted Beta component PDFs for the 8 sites in Group 4 (see maps in Figures 4.1 and 5.1) fitted to 1 hour look-ahead time forecasts for February 14th 2009. The BMA predictive PDF of the lumped wind power output of the group is shown by the red line, and the true observed group wind power with the scattered vertical black line. The BMA weights reflect the single site forecasts relative performance over the 30 previous days. The figure shows that five of the sites (Sklinna (orange), Rørvik (magenta), Brønnøysund (blue), Burholmråsa (green) and Nordøyan (cyan)) are found to give a good performance in the training period and are given more weight than the remaining 3 sites. Myken (brown) and Sandnessjøen (black) also have some influence on the resultant PDF, but are clearly weighted down. Røst, as was found to clearly deviate from the other 7 sites in Figure 5.11 is weighted down to practically 0, and can hardly be seen in the figure (just visible along p(f) = 0 for normalized wind power between 0.2 and 0.25). The BMA predictive PDF for the group ("Group", marked in red) is the sum of the component PDFs (see Equation 5.1). This is the output of the forecast and is to be interpreted as the probability distribution of the lumped group wind power given the single site forecasts. The distribution has its peak just right of the true observed power (scattered black line).

Figures 5.13 and 5.14 show examples of the BMA predictive PDF used to make probabilistic forecasts in the form of prediction intervals for the future lumped group power output of Group 4 (see Figure 4.1). Figure 5.13 shows the forecasts for one week starting February 16th 2009 and Figure 5.14 the forecasts for one week starting November 18th 2009. In both figures forecast look-ahead times from 1 to 24 hours (forecasts for all hours of the day) are included to make the figures easier to read. Forecasts are made for four different sets of confidence levels, 50 % (green), 75 %

(cyan), 90 % (blue) and 95 % (black). The prediction limits for each time-step correspond to the respective quantiles of the BMA predictive PDF (as shown in red in Figure 5.12). The red dots represent the true observed group wind power. One problem that occurred when making the prediction limits, is that the value of the Beta distribution never can reach exactly 0 (0 % group power output) or 1 (100 % group power output). Zero-production is common in the observed data, thus is it necessary to do adjustments in order to enable the forecasts to also cover these observations. This is done in a quick and simple, though not particularly elegant, way by setting a limit for when a beta value is respectively small or large enough to be regarded as a 0 or 1. A lower limit of 0.025 was found to give good results. An upper limit was found to be of less importance as group power outputs of 1 (100 % output from all sites in the group) are extremely rare.

In Figure 5.13 and Figure 5.14 the prediction intervals show a good coverage of the observations. The most visible problems are seen when there are large sudden changes in the observed power, as is the case on February 21st (Figure 5.13) and on November 18th and November 22nd (Figure 5.14). In these instances the confidence intervals tends to miss the timing of the increase or decrease (phase errors) or to miss the amplitude of the change (amplitude errors). The figures also show that the width of the confidence intervals varies with time. This is caused by variations in the shape and spread of the component PDFs and is a desirable property.

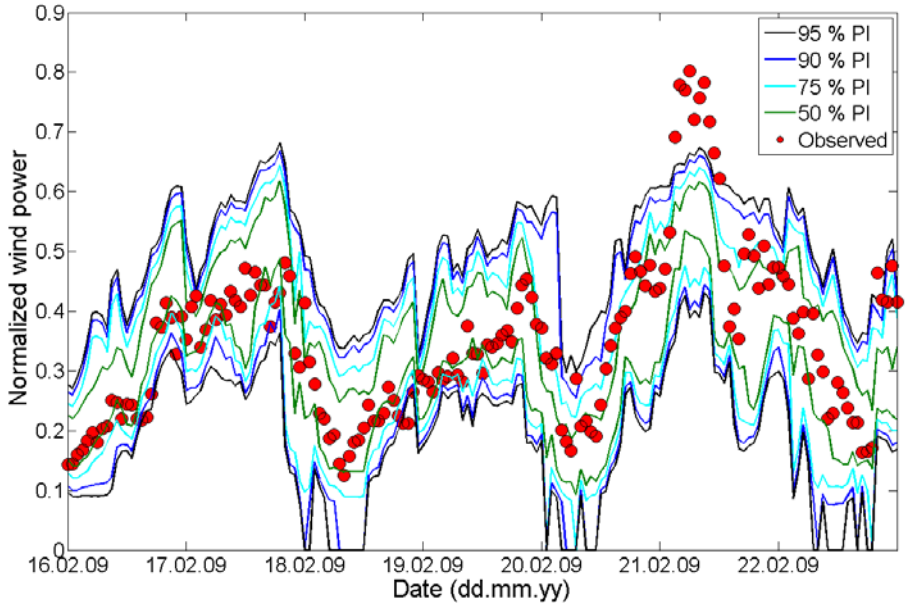

Figure 5.13 – BMA predictive PDF used to make probabilistic prediction interval (PI) forecasts for the future lumped group power output of Group 4 (see Figure 4.1). The figure shows the forecasts for one week starting February 16th 2009. Forecast look-ahead times from 1 to 24 hours (forecasts for all hours of the day) are included to make the figure more intuitive. Forecasts are made for four sets of confidence levels, 50 % (green), 75 % (cyan), 90 % (blue) and 95 % (black). The prediction limits for each time-step correspond to the respective quantiles of the BMA predictive PDF (as shown in red in Figure 5.12). The red dots represent the true observed group wind power.

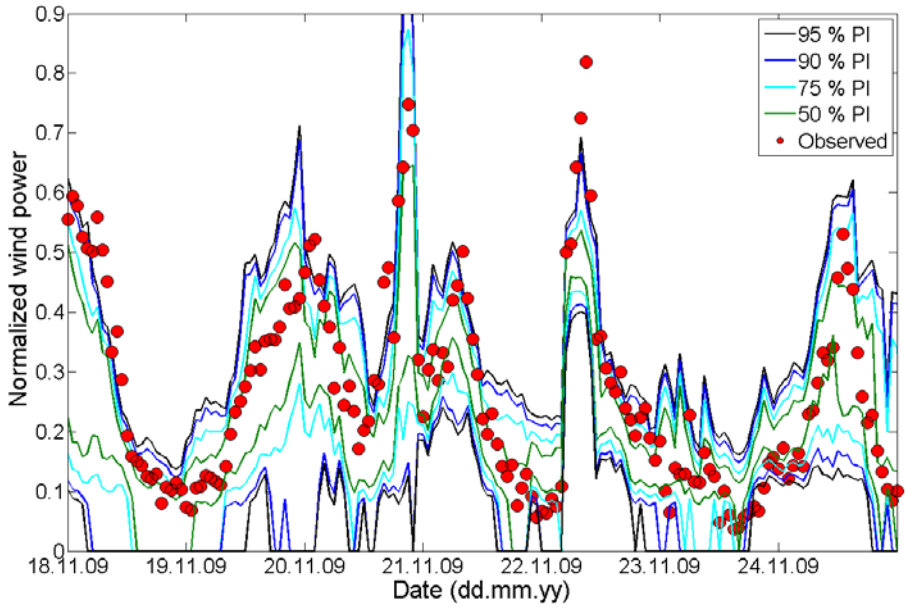

Figure 5.14 - BMA predictive PDF used to make probabilistic prediction interval (PI) forecasts for the future lumped group power output of Group 4 (see Figure 4.1). The figure shows the forecasts for one week starting November 18th 2009. Forecast look-ahead times from 1 to 24 hours (forecasts for all hours of the day) are included to make the figure more intuitive. Forecasts are made for four sets of confidence levels, 50 % (green), 75 % (cyan), 90 % (blue) and 95 % (black). The prediction limits for each time-step correspond to the respective quantiles of the BMA predictive PDF (as shown in red in Figure 5.12). The red dots represent the true observed group wind power.

Figures 5.11-5.14 only shows a few examples of the performance of the BMA model applied to Group 4 (see Fig. 4.1). Even though the excerpts are representative for the majority of the other groups and times, the model performance should also be assessed from a wider perspective. To do so, BMA predictive PDFs were made for the lumped group power of all 7 groups for all hours in the available data. Similar to in Figures 5.13 and 5.14, the forecasts are issued in the form of prediction intervals with four sets of confidence levels (50 %, 75 %, 90 % and 95 %). For each set of prediction limits the number of observations that respectively falls below the lower limit, is covered by the interval, and that is above the upper limit is counted. Ideally the prediction intervals shall cover the same share of the observations as the confidence level indicate (e.g. a 50 % prediction interval shall cover 50 % of the observations).

Table 5.1 shows the shares of observations that are covered by the BMA prediction intervals for 1 hour look-ahead time (forecast for 01:00 CET, one observation per group per day). The results from all the 7 groups are included in the table. The table shows that the three higher confidence levels give a fairly accurate coverage, with deviations of only 2.1 %, 0.5 % and 2.9 %. The prediction intervals with confidence level 50 % has a too high coverage (11.2 % over-coverage), indicating that the prediction interval is too wide. In terms of the shape of the BMA predictive PDF (Figure 5.12) this means that the BMA predictive PDF in average is too little peaked (too low kurtosis). It is also noticed that the share of observations that falls outside the limits is very unevenly distributed, the percentages above the upper limit being significantly higher than the percentages below the lower limit. One reason for this is the scarcity of data in the higher forecast bins (forecasts bins 70 % - 80 % and 80 % - 90 %) causing the estimates of the standard deviation of the Beta distribution to be less accurate.

Table 5.1 – Percentage of observations covered by prediction intervals (PI) for forecast for 1 hour look-ahead time (1 observation per group per day). For the ideal forecast the share of observations covered by the prediction intervals shall match the confidence level of the interval (e.g. 50 % coverage for a 50 % PI). The three higher confidence levels has a fairly accurate coverage, the lowest confidence level has a too high coverage indicating that these prediction intervals in general are too wide. Notice also that the observations not covered by the prediction intervals are very unevenly distributed.

	Under lower limit	Covered by PI	Over upper limit
95 % PI	1.2 %	92.1 %	6.7 %
90 % PI	2 %	90.5 %	7.5 %
75 % PI	6 %	77.9 %	16.2 %
50 % PI	14.2 %	61.2 %	24.5 %

Table 5.2 corresponds to Table 5.1, but here all look-ahead times are included (all hours of the day, 24 observations per group per day). The results are very similar to those of the 1 hour look-ahead time forecasts, and show the same strengths and weaknesses.

Table 5.2 – Percentage of observations covered by prediction intervals (PI) for forecasts for look-ahead times from 1 to 24 hours (24 observations per group per day). For the ideal forecast the share of observations covered by the prediction intervals shall match the confidence level of the interval (e.g. 50 % coverage for a 50 % PI). The three higher confidence levels has a fairly accurate coverage, the lowest confidence level has a too high coverage indicating that these prediction intervals in general are too wide. Notice also that the observations not covered by the prediction intervals are very unevenly distributed.

	Under lower limit	Covered by CI	Over upper limit
95 % CI	2.4 %	90.5 %	7.1 %
90 % CI	3.3 %	87.7 %	9.0 %
75 % CI	6.4 %	78.8 %	14.8 %
50 % CI	14.4 %	62.3 %	23.2 %

5.4 Conclusions

In this chapter the possibilities for using the statistical post-processing method Bayesian Model Averaging (BMA) to make probabilistic forecasts of the lumped wind power output of groups of sites from the single site group members point forecasts have been explored. Following the approach of Sloughter *et al.* (2010) it was found that the lumped group wind power could be described by the Beta distribution (see Equation 5.2). The Beta distribution has the advantage that the parameters can be estimated from the mean and the standard deviation of sample data (see Equations 5.3 and 5.4). It was shown that the mean of the Beta distribution can be estimated from the single site forecast data by first order linear regression (see Equation 5.5 and Figures 5.6 and 5.7). For some of the sites the standard deviation of the Beta distribution could be estimated from the forecasts by second order linear regression (see Equation 5.6 and Figure 5.9), but for other sites this gave a very poor fit (see Figure 5.10). The standard deviations of the Beta distributions were therefore calculated for 10 bins from all available data from each site and organized as look-up tables. This is not an ideal solution, as it involves sacrificing parts of the time-adaptive properties of the BMA method, and it was also found that some of the bins for some sites suffered from scarcity of data causing the standard deviation estimates to most likely be inaccurate. The BMA weights were estimated from the log-likelihood function (Equation 5.7) by

maximum likelihood estimation. Both the means of the Beta distribution and the BMA weights were estimated from forecast-observation pairs from a training period acting as a sliding window. A training period of 30 days was found to be a good compromise of time-adaptivity and stability in the parameter estimates.

Four different prediction intervals were made for all 7 groups (see Figure 4.1) and how well they covered the observed lumped power output of the groups assessed. It was found that prediction intervals with confidence levels 95 %, 90 % and 75 % fairly accurately covered the same share of the data (for the ideal forecast the coverage rate should match the confidence level). For confidence level 50 % the prediction intervals covered a too high share of the observations, indicating that the BMA predictive PDFs was too little peaked (too low kurtosis). It was also found that the distribution of the observations that fell outside the prediction limits was very unevenly distributed for all confidence levels. One reason for this was scarcity of data in some of the forecast bins with subsequent unreliable estimates of the standard deviation of the Beta distribution.

Another problem that was encountered was that the Beta distribution cannot reach exactly 0 (0 % group power output) or 1 (100 % group power output). As especially hours of 0 % production is common in the data (100 % production is very rare and therefore of less importance) this means that the prediction limits were unable to cover a significant share of the observations. This was solved by setting a lower threshold limit (0.025) for which lower Beta values were considered as 0. This was a quick and easy solution to the problem, but it was not particularly elegant. A potential solution to this problem would be to follow an approach similar to that of Chmielecki & Raftery (2010), where a two-component PDF consisting of respectively a logistic distribution and a Beta distribution was used to forecast visibility. Another option could be to use a censored Normal distribution (see e.g. Pinson 2012, Tastu *et al.* 2012 and Messner *et al.* 2013). This option might also solve the problems encountered with an in average too little peaked BMA predictive PDF (too high coverage for confidence level 50 %) and the problems with the very uneven distribution of the observations outside the prediction limits. Caution must however be exercised when trying to fit more complex PDFs as these are likely to require more training data in order to obtain reliable parameter estimates.

Chapter 6 – Wind power ramps

Up to now, this thesis has focused on "the bigger picture", minimizing some function of the forecast errors (for example the RMSE) over longer evaluation periods. However, as was seen in Figures 5.13 and 5.14, large and rapid changes in the wind power production are particularly hard to forecast and often lead to large forecast errors. These situations are often referred to as *wind power ramps* (see Figure 1.2 and Section 1.3.4).

Wind power ramps should be regarded as extreme events. They are relatively rare, but when they occur they have a potentially large impact on the power system (Kamath 2010; Greaves *et al.* 2009). The potential impact makes accurate forecasts of ramps of great interest and of high potential benefit. On the other hand the rareness makes forecasting ramps difficult. Obviously, this means the amount of data for model building will be limited. It is also well known that forecast models based on known probability distributions and trained to minimize e.g. RMSE, like the BMA-model presented in Chapter 5, will tend to have a poorer fit near the upper and lower tails. This makes sense when trying to make the "in average" best forecast, but it is not an optimum strategy for making accurate forecasts of extreme events like wind power ramps. As a result of this, dedicated ramp forecasts have received increasing attention later years. An overview of ramp forecast models found in the literature is given in Section 1.3.4 and references given therein. In this chapter the wind power ramps will be regarded as categorical incidents that are either occurring or not occurring, and the ramp forecasts used to forecast the occurrence of the three events "no ramp", "up-ramp" (large increase of wind power) and "down-ramp" (large reduction of wind power). The focus of the chapter will be on forecasts of ramps in the lumped wind power output of groups of sites. As in the previous chapters Group 4 (see maps in Figures 4.1 and 5.1) will be used for examples.

Common to the ramp forecast methods listed in Section 1.3.4 is that they are specialized ramp forecast made separate of the "ordinary" wind power forecast. In situations with a need for e.g. rapid updates or with very large amounts of input data (e.g. with a very high spatial resolution), limiting the amount of duplication of effort would be beneficial. In this chapter the issue of wind power ramp forecasting is approached in two ways. First it is examined whether the predictability of wind power

ramp forecasts can be enhanced by applying correction schemes to NWP wind power forecasts. As have been shown in the previous chapters, wind power forecasts can be enhanced through e.g. bias correction schemes (Chapter 3) and regression models (Chapter 4). There is good reason to believe that also wind power ramp forecasts will benefit from being based on input forecasts corrected with similar schemes. Thereafter the possibility of using the technique Random Forests for ramp forecasting is investigated. A few examples on the use of Random Forests for wind power purposes are found in the literature, with the Random Forests used as an alternative to regression models for forecasting wind power ramp rates (Zheng & Kusiak 2009), the timing of wind power ramps (Bossavy *et al.* 2010) and the wind power production for the next 1-3 days (Fugon *et al.* 2008; Natenberg *et al.* 2013). Natenberg *et al.* (2013) found Random Forests to train quickly, to be resistant against over-fitting and to be good at handling non-linear relationships in the input data. Random Forests can however also be used as a classification technique providing categorical ramp forecasts.

Section 6.1 discuss when a change in the wind power is large enough and rapid enough to be considered a wind power ramp, and presents methods for identifying ramp events from time-series of historical wind power measurements. Section 6.2 investigates the effect applying different types of correction schemes to NWP wind power forecasts has on wind power ramp predictability. Examples of how the correction schemes affect the time-series of wind power forecasts are shown, and the effect this has on the ramp predictability is evaluated by different evaluation metrics (see Section 1.4.3). Section 6.3 investigates the possibilities of using Random Forest as a classification technique to predict wind power ramps and give early-warnings of when wind power ramps will occur. A step-by-step presentation of the Random Forests model building is given, and the results obtained through applying the method to wind power forecasts from Group 4 (see maps in Figures 4.1 and 5.1) are presented. As in Section 6.2, how corrections of the NWP wind power forecasts affects how well the Random Forest model predicts ramps is discussed. Section 6.4 draws some conclusions on how correction schemes for NWP wind power forecasts can increase wind power ramp predictability and presents some ideas for changes in the presented Random Forest model to increase its performance for ramp predictions.

6.1 Identification of wind power ramps

Transforming a wind power forecast into a wind power ramp forecast or a series of observed wind power measurements into a series of observed ramps and non-ramps requires a way of identifying when ramps are happening. In the literature there is no consensus about a standard formal definition of a wind power ramp (Zheng & Kusiak 2009). A main reason for this is that ramps are primarily described by the consequences of the change in power production, and that this will vary depending on location, installed wind power capacity, the flexibility of the grid, other energy sources connected to the grid etc. (Cutler *et al.* 2007). In other words, a rapid change in the wind power production that requires major efforts to deal with at one location and therefore is considered a ramp can at another location be within the limits of the flexibility of the system and therefore the change is not seen as a ramp. There are good practical reasons for this, so the lack of consensus about a definition cannot be considered a problem in itself, but it has the consequence that a wide variety of definitions are used. These range from very simple definitions that only takes the total change of wind power production over a time period into consideration (e.g. Kamath 2010) via definitions that focus on the rate of change (e.g. Zheng & Kusiak 2009) to more complicated definitions involving filtering of the wind power signal (e.g. Bossavy *et al.* 2010) or dedicated algorithms (e.g. Florita *et al.* 2013 and Zhang *et al.* 2014).

It is not within the scope of this chapter to discuss or assess different ramp definitions. For the exemplifications a slight alternation of the simple definition from Kamath (2010) is chosen:

$$P_{val} = P(t + \Delta_t) - P(t), \tag{6.1}$$

where $P(t)$ is the normalized lumped group power at time t and Δ_t is a pre-defined time increment, here set to 3 hours. P_{val} is compared to a pre-defined *ramp change limit* P_{lim}, that defines how large the change in wind power production needs to be from t to $t + \Delta_t$ in order to be considered a ramp. If $P_{val} > P_{lim}$ it is considered to be an up-ramp at time $t + \Delta_t$ (hence should an up-ramp be forecasted at time t), and if $P_{val} <\div P_{lim}$ it is considered to be a down-ramp at time $t + \Delta_t$ (hence should a down-ramp be forecasted at time t). Comparisons of different wind power ramp definitions are found in Ferreira *et al.* (2010) and Zhang *et al*. (2014).

For operational purposes the ramp change limit P_{lim} should be decided from knowledge about the properties of the respective energy system, but as the examples in this paper are based on constructed data this is not possible here. Instead three arbitrary values (0.2, 0.3 and 0.4) for P_{lim} are chosen to test if the results are consistent for different values of P_{lim}. The values decide how large the relative change in wind power production in Δ_t needs to be in order for the change to be considered a ramp, e.g. does a P_{lim} of 0.3 mean that an increase in the wind power production of more than 30 % within 3 hours is considered an up-ramp and a decrease of more than 30 % within 3 hours is considered a down-ramp. Changes of less than 30 % are considered no ramps. In all examples in the chapter the true observed ramps are identified by applying the ramp definition from Equation 6.1 with the parameterization presented above (20/30/40 % change within 3 hours) to each step of the time-series of observed lumped group wind power. The number of observed ramps and no ramps together with the share of observed ramps in the data are shown in Table 6.1. A visualization of how ramps are identified from the time-series of wind power forecasts and observations is given in Figures 6.1 and 6.2.

Table 6.1 – Number of true observed ramps and no ramps when applying the ramp definition in Equation 6.1 with ramp change limits P_{lim} 0.2, 0.3 and 0.4 to the time-series of observed wind power production from Group 4 (see Figure 4.1).

Ramp change limit	Up-ramps	Down-ramps	No ramps	Share of ramps
0.2	796	833	12322	0,117
0.3	221	229	13501	0,032
0.4	61	62	13828	0,009

6.2 Effect of correction schemes on ramp predictability

In this section it is investigated how using the models presented in Chapters 3 and 4 to correct the raw NWP forecasts influence the predictability of wind power ramps. In Chapters 3 and 4 it was shown that these lead to reductions in the forecast RMSE – an improvement of the "in average" forecast. Here it is checked if the correction schemes also improve the wind power forecasts in ramp situations, making wind power ramps easier to predict. The section starts with a look at the changes the correction schemes leads to in the NWP wind power forecasts, before moving on to how these changes affect the predictability of wind power ramps.

The ramp predictability of four different wind power forecasts, one uncorrected and three corrected, are assessed and compared:

- An uncorrected forecast where the raw NWP single sites forecasts are gathered into groups as described in Section 4.2.1.
- A forecast where the raw NWP forecasts are corrected by the directional bias correction scheme derived in Section 3.2 and gathered into groups as described in Section 4.2.1.
- A directional auto-correlation regression group forecast, as described in Section 4.3.2, but where only the auto-correlation terms are included (i.e. without the use of off-site information).
- A spati-temporal regression group forecast, presented as forecast D2 in Tables 4.3 and 4.4 (i.e. including off-site information).

All forecasts are based on the forecast data presented in Chapter 2, and all examples given are for Group 4 (see Figures 4.1 and 5.1).

The wind power ramp forecasts are made from the wind power forecasts by applying Equation 6.1 with the parameterization presented above (20/30/40 % change within 3 hours) to each step of the time-series. The observed historical measurements are used as $P(t)$ and the forecasts for 3 hours later as $P(t + \Delta_t)$. The forecasted ramps are compared to the true observed ramps, identified from the time-series of measurements as explained in Section 6.1.

Figure 6.1 shows the observed wind power production (magenta) together with the uncorrected- (green), directional bias corrected- (black), auto-correlation- (blue) and spatio-temporal regression (red) forecasts for three days from June 13th 2010 to June 16th 2010. Special attention is paid to larger amplitude errors (over- or under-predictions) and phase errors (time-shifts in forecasted changes) as this will have a large influence in the ramp identification. The uncorrected forecast differs significantly from the observed production with large amplitude errors (over-prediction) at both peaks. The corrected forecast also suffers from amplitude errors, but these are in general less frequent and of less magnitude. This pattern is representative for the entire forecast period used in this thesis. In front of the second peak there are clear signs of a phase error for all forecasts, but once again the error is more severe for the uncorrected forecast, with a more consistent phase error and ending in an amplitude error. Phase errors are less common than amplitude errors, but

also here the general pattern is that the uncorrected forecast is more prone to phase errors than the corrected forecast. Figure 6.2 shows the forecasts in Figure 6.1 transformed into a forecast of the change in wind power production over the next 3 hours (See Equation 6.1). The horizontal dashed lines indicate ramp change limits of ± 0.3. As is seen the problems with the uncorrected forecast in Figure 6.1 results in predictions of false ramps at both peaks. On the second peak the directional bias-corrected wind power forecast also forecasts a false ramp. Notice also that all forecasts except the directional bias corrected fail in forecasting the down-ramp late in the day June 15th.

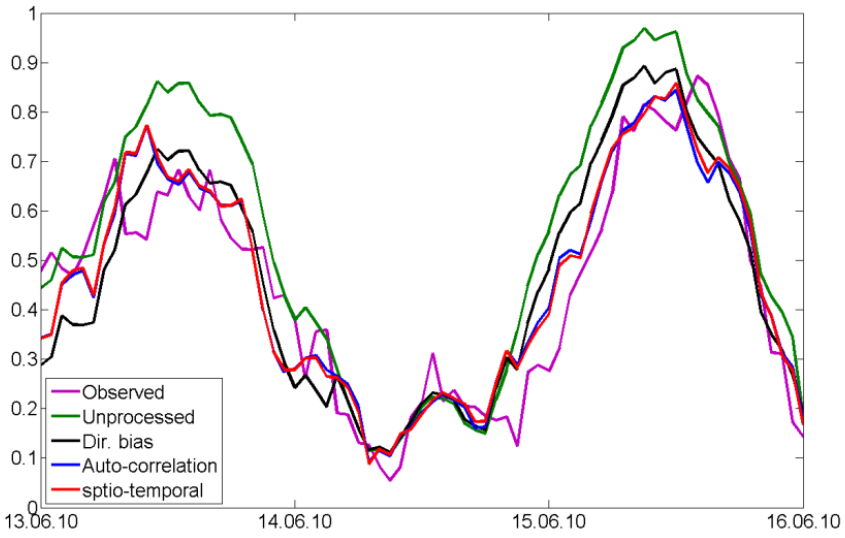

Figure 6.1 - Observed wind power production (magenta) together with the uncorrected- (green), directional bias corrected- (black), auto-correlation- (blue) and spatio-temporal regression (red) forecasts for three days from June 13th 2010 to June 16th 2010. At both peaks the uncorrected forecast has a very clear amplitude error. In front of the second peak all forecasts shows signs of phase error, forecasting the large increase in wind power production some hours too early.

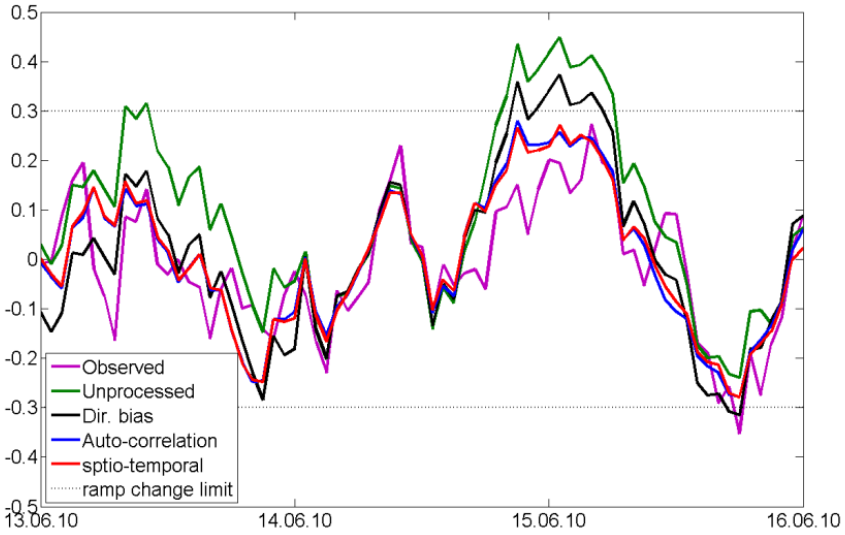

Figure 6.2 - Observed change in wind power production (magenta) together with the uncorrected- (green), directional bias corrected- (black), auto-correlation- (blue) and spatio-temporal regression (red) forecasts of the change for three days from June 13th 2010 to June 16th 2010. The change in wind production is calculated over 3-hour time intervals, following Equation 6.1. The horizontal dashed lines indicate a change of ±30 % within a 3 hour period, here defined as a ramp given a 30 % (0.3) ramp change limit. It is seen how the amplitude errors in the uncorrected forecasts results in an over-prediction of wind power up-ramps. Late in the day 15.06.10 there is a down-ramp that only the directional bias corrected forecast are able to capture.

To answer the core question of this section, how the correction schemes affects the predictability of wind power ramps, four different evaluation metrics, sensitivity, specificity, frequency bias and Hanssen and Kuipers skill score (HK) are considered. For a more thorough presentation of the evaluation metrics it is referred to Section 1.4.3 and Jolliffe & Stephenson (2003).

Figure 6.3 shows the sensitivity of the ramp forecasts. This corresponds to the probability of forecasting a ramp when a ramp actually occurs. The figure shows that the sensitivity is decreasing with increasing amount of correction of the forecasts, i.e. that more extensive correction schemes gives a lower probability of detecting ramps. This might seem contra-intuitive, but has its reason in that the metric do not account for over-forecasting of ramps (see Figure 6.5). Forecasting all observations as ramps will thus give an optimum sensitivity even though the forecast will be without value. Notice also that the number of true observed ramps for ramp change limit 0.4 is very low and as a result of this small changes in the number of detected ramps cause large impacts on the metric. For ramp change limit 0.4 the drop in sensitivity from the auto-

correlation regression model to the spatio-temporal regression model is caused by one single missed ramp.

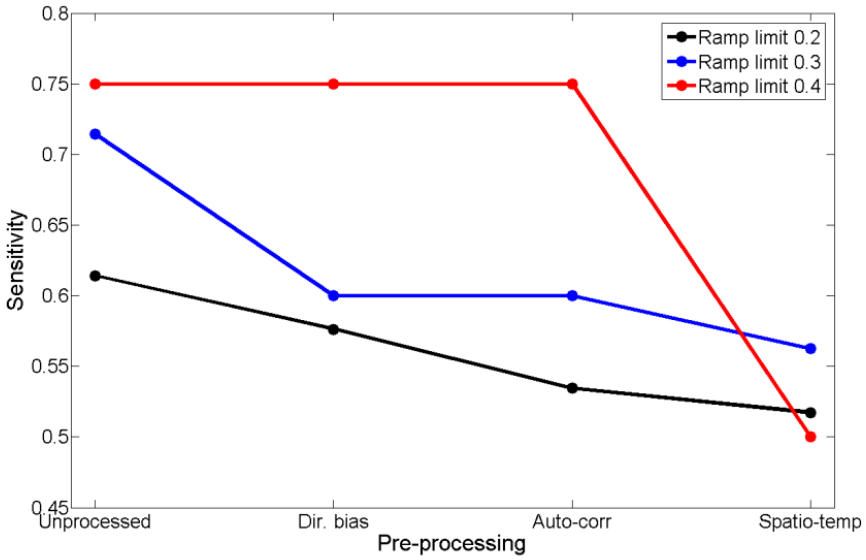

Figure 6.3 - Sensitivity, the probability of detecting a ramp that is happening, for ramp change limits 0.2 (black), 0.3 (blue) and 0.4 (red). Optimum sensitivity is 1, which means that all observed ramps are forecasted. Notice that for limit 0.4 the numbers of observed ramps are very low. As a result of this, small changes in the number of detected ramps cause large impacts on the metric. This is the cause of the large drop for the spatio-temporal model for limit 0.4, where the reduction in ramps detected is only 1.

Figure 6.4 shows the specificity of the ramp forecasts. This corresponds to the probability of avoiding forecasting a ramp when no ramp is occurring. As is seen the ranking of the correction schemes is the opposite of that of the sensitivities in Figure 6.3. This is expected, as a high sensitivity tends to lead to a low specificity and the other way around. It is also noticed that there is practically no gain from the spatio-temporal regression model compared to the auto-correlation regression model, and for ramp change limit 0.2 even a slight decrease. Seen in connection with Figure 6.3 this gives clear indices that the increase in ramp predictability for this case does not increase by including off-site information. Notice, however, that also this metric can be fooled into giving a perfect score, now by forecasting all observations as non-ramps. As a result of this it should not be used isolated. The problems with few true observed ramps do not get as big an impact on the results for the specificity as for the sensitivity as it is the non-ramps that have the most influence on the metric.

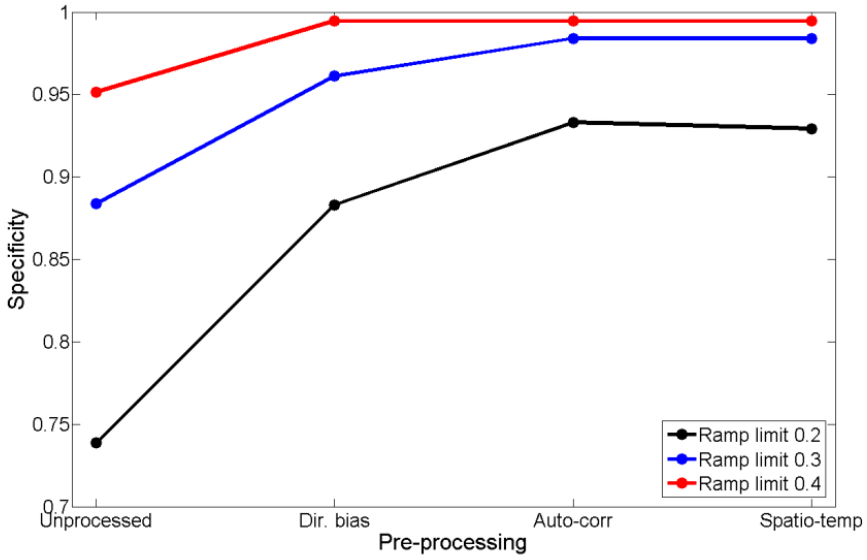

Figure 6.4 - Specificity, the probability of avoiding predicting false ramps, for ramp change limits 0.2 (black), 0.3 (blue) and 0.4 (red). Optimum specificity is 1, which means that no ramps are forecasted when a ramp is not occurring.

Figure 6.5 shows the frequency bias of the ramp forecasts. This is a measure of how the numbers of forecasted ramps correspond with the number of true observed ramps. Notice that the metric does not provide information on whether the ramps are forecasted at the correct time, but only on how well the number of forecasted and observed ramps corresponds when summed up over the entire evaluation period. As suspected from Figure 6.3 the ramp forecasts based on the uncorrected wind power forecast shows a very high frequency bias, forecasting between 300 % (for ramp limit 0.2) and nearly 800 % (for ramp limit 0.4) too many ramps. The tendency of over-forecasting is greatly reduced when basing the ramp forecasts on the bias-corrected wind power forecasts, and even further reduced when basing the ramp forecasts on the auto-correlation (with the exception of ramp limit 0.4) or spatio-temporal regression forecasts. All ramp forecasts have a positive frequency bias indicating an over-estimation of the probability for ramps, but for the ramp forecasts based on the spatio-temporal regression wind power forecasts the frequency bias is as low as 10 % to 20 %.

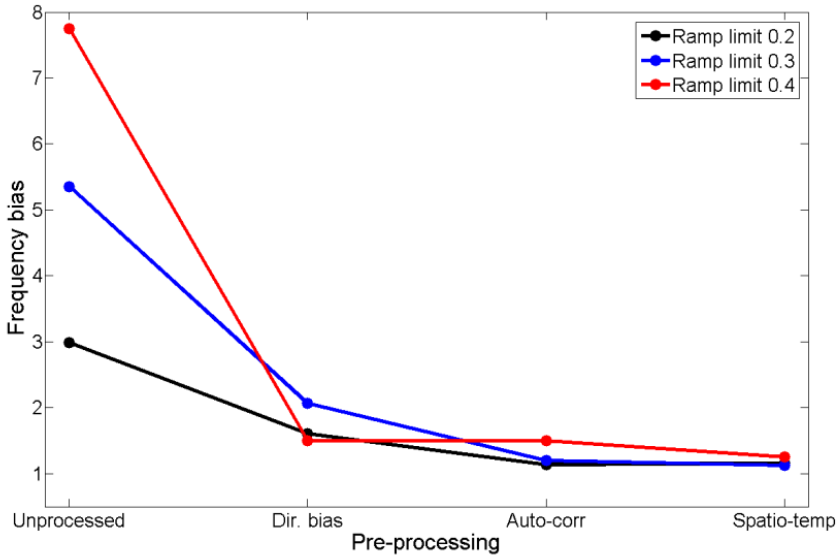

Figure 6.5 - Frequency bias, how the numbers of predicted ramps correspond to the number of observed ramps, for ramp change limits 0.2 (black), 0.3 (blue) and 0.4 (red). Optimum frequency bias is 1, which means that the same number of ramps is forecasted as is observed.

In contrast to the simple evaluation metrics that only measure one property at a time, HK aim at summarizing all aspects of the ramp forecasts into one value. As explained in Section 1.4.3, the HK scores measures a forecast strategy's relative improvement over a random forecast strategy. As all four datasets here are subject to the same forecast strategy (see Equation 6.1), the differences obtained here will be caused by how the different correction schemes influence the predictability of wind power ramps. Figure 6.6 shows the HK scores for the ramp forecasts. For all three ramp change limits the HK scores show a clear positive effect of the correction schemes, with the uncorrected NWP ramp forecast receiving the lowest HK scores. For ramp change limits 0.2 and 0.3 there is also a clear gain from choosing the more extensive regression correction schemes, with the auto-correlation regression achieving slightly higher HK scores than the spatio-temporal regression. For ramp limit 0.4 the directional bias correction and the auto-correlation regression models achieve the highest HK scores.

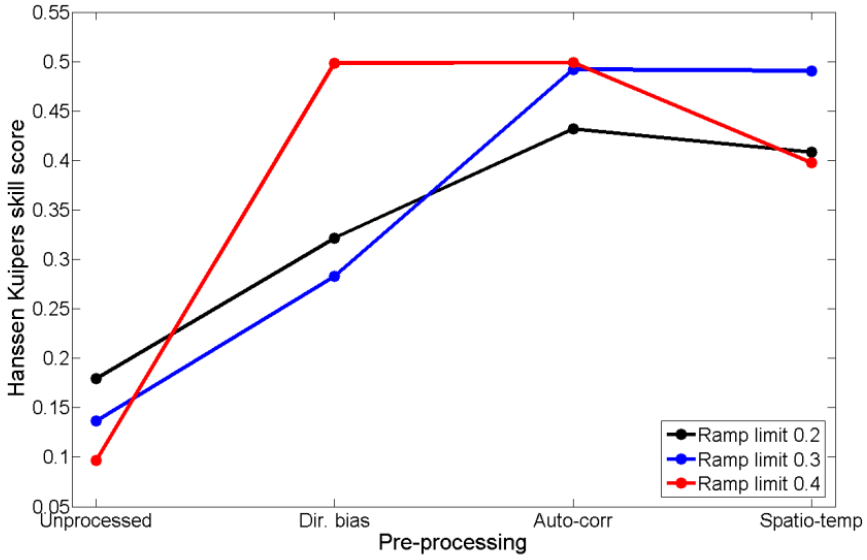

Figure 6.6 - Hanssen & Kuipers skill score, the relative improvement compared to a random prediction strategy, for ramp change limits 0.2 (black), 0.3 (blue) and 0.4 (red). The score has a range of -1 to +1, with 0 representing no skill.

6.3 Random Forests ramp forecasts

In this section the method Random Forests, as introduced by Breiman (2001), will be used as an alternative method to forecast wind power ramps. As described in Section 6.1 the ramps are seen as categorical events that are either occurring or not occurring (3 classes, “down-ramp”, “up-ramp” and “no ramp”). The Random Forests is therefore used as a classification technique, forecasting the class of sets of input data. As explained in the introduction of the chapter there is a handful of examples of previous use of Random Forests for wind power forecasting, including one example where it with questionable success was used to forecast wind power ramp rates (Zheng & Kusiak 2009). Natenberg *et al.* (2013), however, found that Random Forests has advantages over other machine learning methods in that it trains quickly, has a high resistance to outliers and over-fitting and that it handles nonlinearities in the input data well. All three are qualities that will be very beneficial in a wind power ramp forecast. In the mentioned examples Random Forests was used as a regression model, and not a classification model. Examples of successful use of Random Forests as a classification method are found for a wide range of other classification and pattern recognition problems (see e.g. Gislason *et al.* 2006, Jiang *et al.* 2007 and Dubath *et al.* 2011).

The basis of a Random Forest is a classification tree (see e.g. Hastie *et al.* (2009) for a thorough presentation of classification trees). This is a hierarchical structure that identifies and learns patterns in empirical data. Starting with a set of training data with known class outcomes, the first step of building a classification tree is to decide on which variable and which value that should be used to split the data into two groups. The split should be done so that it minimizes (or maximizes) the value of some evaluation metric when the forecasted class of each of the groups are compared to the known outcomes of the group members. A number of different evaluation metrics have been proposed, but the clearly most commonly used is the so-called Gini impurity (see e.g. Breiman *et al.* 1984). The Gini impurity measures the diversity of the known classes represented in each group (forecasted class), and the aim is thus to choose splits so that this is minimized. This process is repeated for each of the resultant groups until all observations are forecasted into the correct class or some other termination criteria are met. An illustration of the process is given in Figure 6.7. The set of variable-value pairs used to split the data at each node, usually referred to as splitting rules, identify patterns in the data and can later be used to forecast the classes of new sets of data.

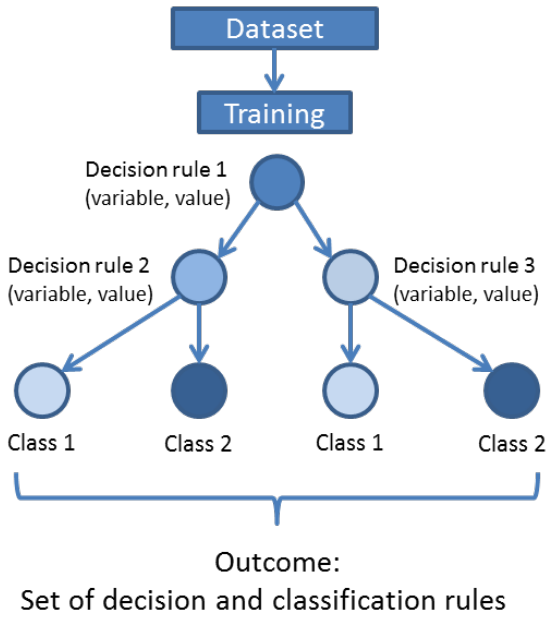

Figure 6.7 – Schematic illustration of a classification tree. The training dataset is split into smaller groups through deciding on which split that maximize (or minimize) the value of some evaluation metric when the forecasted class of the resultant groups are compared to the known class outcome of the group members. The process is repeated until all observations are forecasted into the correct class. The output of the tree is a hierarchical set of decision rules (variable-value pairs) that can be used to forecast the classes of new sets of data.

Classification trees have the ability to capture complex structures in the data. They are amongst other able of handling circular input data (wind directions) without the need for extensive modelling or the definition of wind sectors (see Sections 4.2.4 and 4.3.4). In contrast to many other machine learning methods, like neural networks, it is also a pure white-box technique, meaning that all parts of the model has an interpretation that can be understood and, if needed, corrected based on previous knowledge. Unfortunately are classification trees also very noisy and hence tend to have a high variance. Random Forests is a technique to handle this problem through constructing a large collection of de-correlated classification trees by building classification trees from random samples from the training data. Averaging over the total number of constructed trees, *B*, reduces the variance by σ^2/B for as long as the trees remain independent. Random forests also have the advantage that it has a "built-in" cross-validation procedure referred to as "out-of-bag" (OOB). This means that the model is controlled against unseen data at the same time as it is built and hence is it not necessary to cross-validate the model to control for over-fitting.

As in the previous sections all ramp forecasts are issued for the next 3 hours, answering the question: "will the change in group wind power exceed the ramp change limit (here 0.2/ 0.3/ 0.4) over the next 3 hours?"

6.3.1 Model building

The building of a random forest classification model goes through five steps (for a more thorough explanation see Breiman (2001)):

1. Draw a random set Z^* of *n* samples from the complete set of training samples N.
2. Randomly select a subset of *m* variables from the total *p* variables available.
3. Build a classification tree based on Z^* and the *m* selected variables (see Figure 6.7), but in contrast to an ordinary classification tree randomly select *m* new variables at each split.
4. Repeat 1) – 5) *B* times to build a forest of *B* trees.
5. Output the ensemble of trees $\{T_b\}^B$.

The predicted class for a new set of observations *x* is:

$$\hat{C}_{RF}^{B}(x) = majority\ vote\{\hat{C}_b(x)\}^B, \tag{6.2}$$

where $\hat{C}_b(x)$ is the class prediction random forest tree number b. A schematic visualization of the process is given in Figure 6.8.

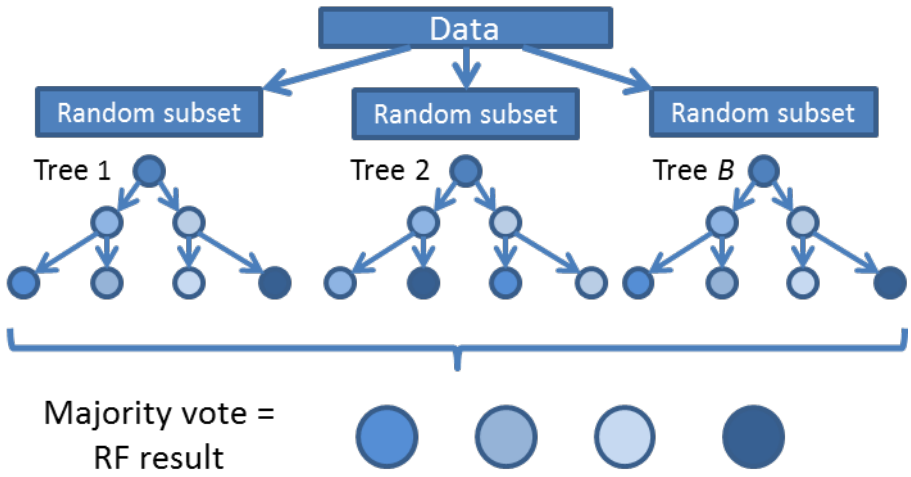

Figure 6.8 – Schematic figure of a Random Forest. For each of the *B* trees in the forest a random subset is drawn from the training data. The *B* trees is built the same way as shown in Figure 6.7 with the exception of the random sampling of variables in point 3). The forecast class of all observations are decided by a majority vote where each tree casts one vote for the class of the observation.

In the above description there are three parameters that need to be decided, the number of randomly selected samples (*n*), the number of randomly selected variables (*m*) and the number of threes in the forest (*B*). As the number of events (wind power ramps) is relatively small compared to the number of observations *n* is decided by splitting *N* in four equal parts of which three is allocated to n and the last is left as "new" observations. *m* and *B* are found in an exploratory manner by building random forests for different values of *m* with a large *B*. *m* is chosen so that the OOB forecast error of the "new" observations is minimized, while *B* is set to the lowest value for which the OOB forecast error of the "new" observations has reached a stable level. With all parameters in place the full model is fitted.

As in most multivariate problems it is to be expected that the predictor variables are of variable importance and that there most likely are variables that can be omitted without decreasing the models predictive performance. For random forests the contribution of each predictor variable can be calculated as the increase in forecast error that occurs when the values of the variable are randomly permuted across the

"new observations". This corresponds to making a new forecast without the systematic contribution of the variable. The increase is calculated for all *b*, and averaged over the total number of trees, *B*, to form a metric for the variable importance. Model reductions can then be performed by defining a lower limit for the predictor variables importance to remain in the model. The value of the lower limit for variable importance is defined in a similar way as the number of randomly selected variables *m*, by testing different values and selecting the one that gives the model with the lowest OOB forecast error when applied to the "new" observations. Thorough descriptions of Random Forests are found in Breiman (2001) and Hastie *et al.* (2009).

6.3.2 Data preparation

The input data to the Random Forest models will here consist of a combination of wind power forecasts and wind direction forecasts for the group of interest (for which the ramp forecast is issued) as well as wind power measurements and forecasts from the nearest neighbouring groups. Four input datasets are made, based on the same four variants of wind power forecasts that were described in Section 6.2 (uncorrected, directional bias corrected, auto-correlation regression and spatio-temporal regression). The neighbouring groups are included based on the assumption that there is a high probability that the ramp event one wants to forecast occurred at a downwind site at an earlier time, hence that the ramps are subject to spatial propagation from downwind sites to upwind sites. Logically, this makes sense, and it is also a common assumption in both wind power ramp forecast models (see e.g. Bossavy *et al.* 2010 and Collier *et al.* 2013) and wind power forecast models (see e.g. Section 4.3 and Tastu *et al.* 2011). For a presentation of the groups of sites see Section 4.2.1 and Figure 4.1.

For the group of interest the following predictor variables are included in the Random Forest ramp forecast model (see also Figure 6.9):

- The observed wind power at time *t* (the time the ramp forecast is issued)
- The observed wind direction at time *t*
- The change in wind power from time *t* (observed) to time *t+1* (forecast)
- The change in wind power from time *t* to time *t+2* (forecast)
- The change in wind power from time *t* to time *t+3* (forecast)

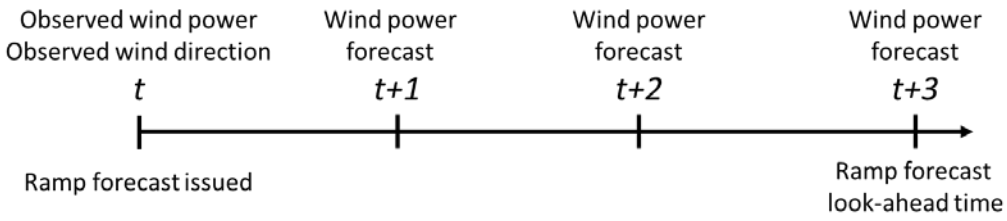

Figure 6.9 – Structure of the input data to the Random Forest ramp forecast model from the group for which the forecast is issued. The ramp forecast is issued at time *t* aiming to answer the question: "Will the change in wind power exceed the ramp change limit over the next 3 hours?"

For the neighbouring groups (two groups) the following predictor variables are included in the Random Forest ramp forecast model (see also Figure 6.10):

- The change in wind power from time *t* (observed) to time *t+3* (forecast)
- The change in wind power from time *t-1* (observed) to time *t+2* (forecast)
- The change in wind power from time *t-2* (observed) to time *t+1* (forecast)

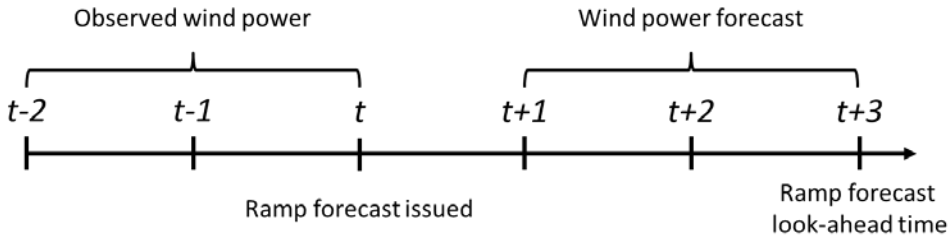

Figure 6.10 – Structure of the input data to the Random Forest ramp forecast model from the neighbouring groups (off-sites). The ramp forecast is issued at time *t*.

For each group of interest this gives four datasets, each consisting of 11 predictor variables and approximately 13800 samples.

6.3.3 Application results

From each of the datasets described above a Random Forest ramp forecast model is built following the step-by-step description in Section 6.3.1. As in the previous chapters and sections Group 4 (see Figures 4.1 and 5.1) is used for exemplifications. The ramp forecast is issued as a categorical forecast with three categories, where the

forecasted change in group wind power over the next three hours is placed into one of the three classes “down-ramp”, “up-ramp” or “no ramp”. The series of ramp forecasts are compared to corresponding series of true observed ramps (and no ramps) identified by Equation 6.1 with ramp change limits P_{lim} of 0.2 (20 % change in wind power within 3 hours) and 0.3 (30 % change in wind power within 3 hours). With a ramp change limit of 0.4 (40 % change in wind power within 3 hours, as was used in Section 6.2) the true observed ramps becomes too rare for the Random Forest models to be estimated properly. This problem will be discussed further in the conclusions of the chapter. The Random Forest ramp forecasts are assessed using the same four metrics as was used to check the influence of wind power forecast correction schemes on ramp predictability in Section 6.2; Sensitivity, specificity, frequency bias and Hanssen and Kuipers skill score (HK). A more thorough presentation of the metrics is given in Section 1.4.3 and Jolliffe & Stephenson (2003). The ramp forecasts from Section 6.2 are included as a reference as these show the corresponding ramp forecasts without the extra effort of the Random Forest modelling.

Figure 6.11 shows the sensitivity of the Random Forest ramp forecasts. The corresponding sensitivity of the ramp forecasts made by applying Equation 6.1 to the different wind power forecasts in Section 6.2 is included as a reference with dotted lines. The sensitivity measures the probability of forecasting a ramp when a ramp actually occurs. As is clearly seen the sensitivity of the Random Forest ramp forecasts is generally low, and much lower than the sensitivity of the reference forecasts. For ramp limit 0.3 the Random Forest ramp forecasts based on the auto-correlation and the spatio-temporal regression wind power forecast have a sensitivity close to zero, indicating that very few ramps are correctly forecasted. As noted in Section 6.2 cautions should though be exercised when interpreting the sensitivity metric as it is easily manipulated by over-forecasting of ramps. As shown in Figure 6.5 this explains parts of the high sensitivity of the reference ramp forecasts based on the uncorrected and directional bias corrected wind power forecasts, but it does not explain the high sensitivity of the reference ramp forecasts based on the auto-correlation and the spatio-temporal regression wind power forecasts.

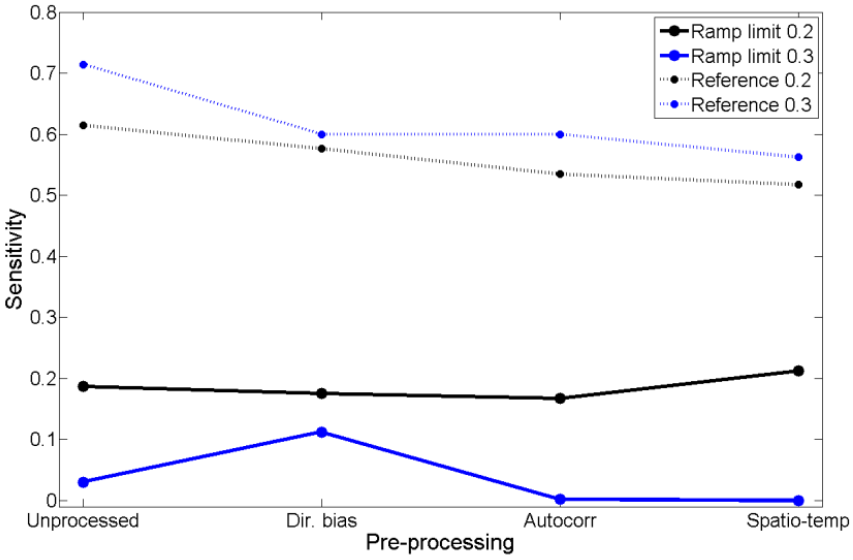

Figure 6.11 – Sensitivity, the probability of detecting a ramp that is happening, for ramp change limits 0.2 (black) and 0.3 (blue). The sensitivity of the ramp forecasts made by applying Equation 6.1 to the different wind power forecasts in Section 6.2 are included as a reference with dotted lines. Optimum sensitivity is 1, which means that all observed ramps are forecasted. The sensitivity of the Random Forest ramp forecasts is generally low, and much lower than the sensitivity of the reference forecasts. For ramp limit 0.3 the Random Forest ramp forecasts based on the auto-correlation and the spatio-temporal regression wind power forecast have a sensitivity close to zero, indicating that very few ramps are correctly forecasted.

Figure 6.12 shows the specificity of the Random Forest ramp forecasts. The corresponding specificity of the ramp forecasts made by applying Equation 6.1 to the different wind power forecasts in Section 6.2 is included as a reference with dotted lines. The specificity measures the forecasts ability to avoid forecasting a ramp when no ramp is occurring. As expected from Figure 6.11 the Random Forest ramp forecasts has a very high specificity. For both ramp change limits and all wind power forecasts the Random Forest ramp forecasts has a clearly higher specificity than the reference ramp forecasts. The difference is however decreasing with increasing ramp change limit and with more extensive corrections of the wind power forecasts. Like the sensitivity metric also the specificity metric is easily manipulated, now by under-forecasting ramps. This is a part of the reason for the high specificities of the Random Forest ramp forecasts, in particular for ramp change limit 0.3.

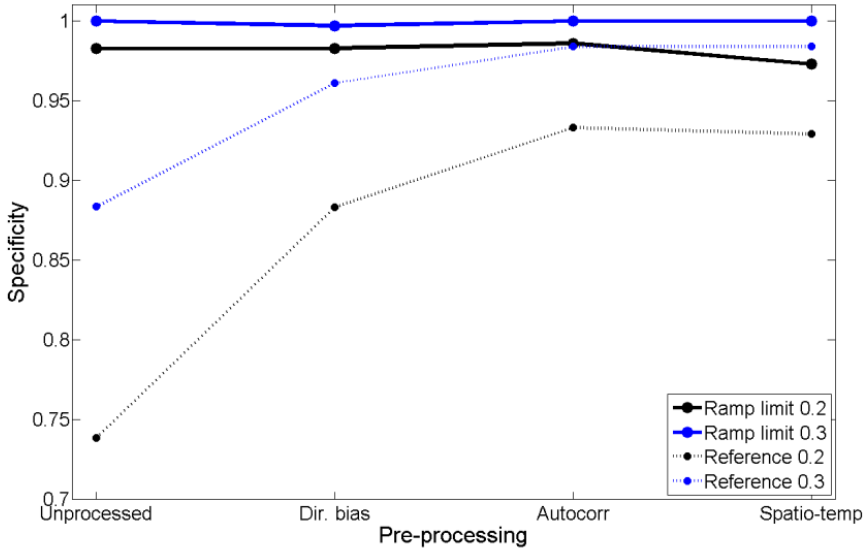

Figure 6.12 – Specificity, the probability of avoiding forecasting false ramps, for ramp change limits 0.2 (black) and 0.3 (blue). The specificity of the ramp forecasts made by applying Equation 6.1 to the different wind power forecasts in Section 6.2 are included as a reference with dotted lines. Optimum specificity is 1, which means that no false ramps are forcasted. The specificity of the Random Forest ramp forecasts is generally very high, but a major part of this can be explained by low numbers of forecasted ramps.

Figure 6.13 shows the frequency bias of the Random Forest ramp forecasts. The corresponding frequency bias of the ramp forecasts made by applying Equation 6.1 to the different wind power forecasts in Section 6.2 is included as a reference with dotted lines. The frequency bias measures how the numbers of forecasted ramps correspond with the number of true observed ramps, with an optimum value of 1. Notice that the metric does not provide information on whether the forecasted and observed ramps occur at the same time, it can thus not be used alone as a measure of the quality of a forecast. As expected from Figures 6.11 and 6.12 the Random Forest ramp forecasts has a very low frequency bias, meaning that far less ramps are forecasted than is observed. For ramp change limit 0.3 the Random Forest ramp forecasts based on the auto-correlation and spatio-temporal regression wind power forecasts has a frequency bias close to 0, indicating that very few ramps are forecasted. This corresponds well with and explains the results from Figures 6.11 and 6.12.

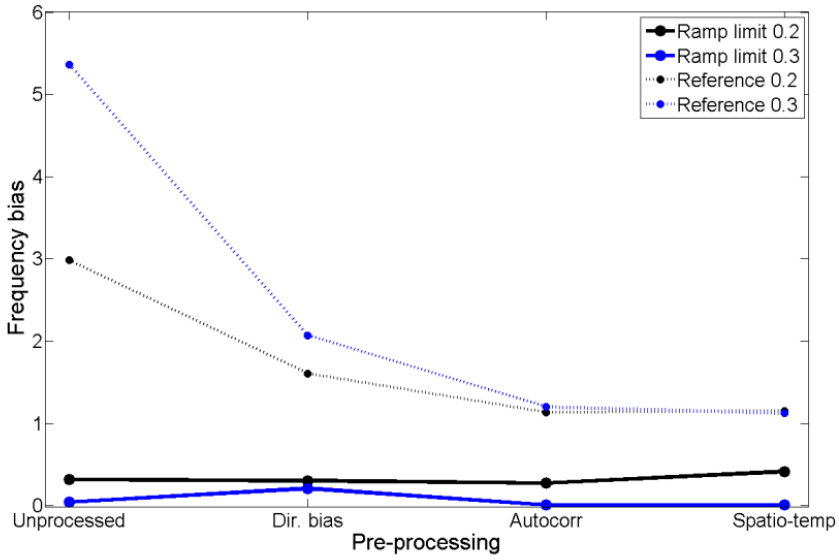

Figure 6.13 – Frequency bias, how the numbers of predicted ramps correspond to the number of observed ramps, for ramp change limits 0.2 (black) and 0.3 (blue). The frequency bias of the ramp forecasts made by applying Equation 6.1 to the different wind power forecasts in Section 6.2 are included as a reference with dotted lines. Optimum frequency bias is 1, which means that the same number of ramps is forecasted as is observed. For both ramp change limits the Random Forest ramp forecasts has a very low frequency bias, indicating that far less ramps are forecasted than are observed.

Figure 6.14 shows the HK score of the Random Forest ramp forecasts. The corresponding HK score of the ramp forecasts made by applying Equation 6.1 to the different wind power forecasts in Section 6.2 is included as a reference with dotted lines. In contrast to the simple evaluation metrics that only measure one property at a time, HK aim at summarizing all aspects of the ramp forecasts into one value, measuring a forecast strategy's relative improvement over a random forecast strategy. Somewhat surprisingly the figure shows that the Random Forest ramp forecasts obtain higher HK scores than the reference ramp forecast for ramp forecasts based on the uncorrected and directional bias-corrected wind power forecasts. The reason for this is the massive over-forecasting of ramps in the reference forecasts, as shown in Figure 6.5. For the ramp forecasts based on the auto-correlation and spatio-temporal regression wind power forecasts there is a large difference in HK score between the two ramp change limits. For ramp change limit 0.2 the Random Forest ramp forecast obtains slightly higher HK scores than the reference ramp forecast, while for ramp change limit 0.3 the reference forecast obtains much higher HK scores. For ramp change limit 0.3 the Random Forest ramp forecast based on the spatio-temporal regression forecast actually receive a negative HK score, meaning that the ramp

forecast performs worse than if randomly assigning the observations to the three classes.

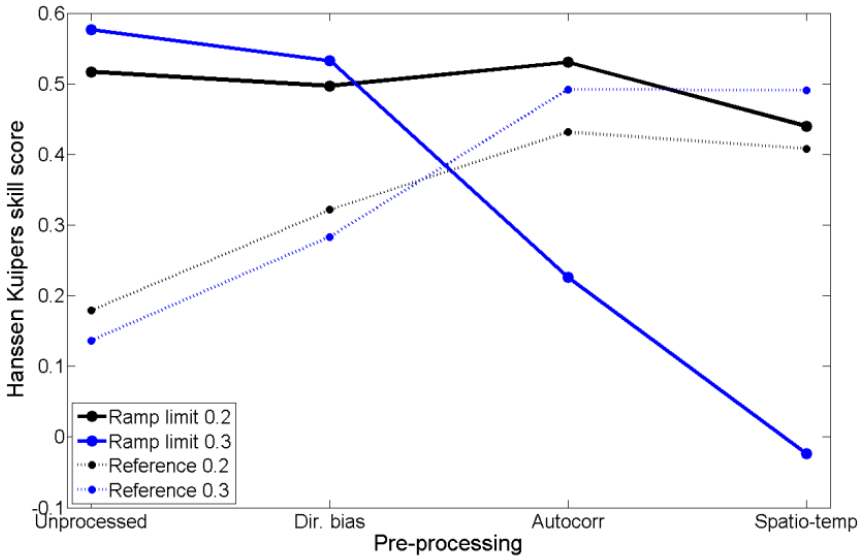

Figure 6.14 - Hanssen & Kuipers skill score, the relative improvement compared to a random prediction strategy, for ramp change limits 0.2 (black) and 0.3 (blue). The score has a range of -1 to +1, with 0 representing no skill, and can be interpreted as the relative improvement (or detriment) over a random forecast strategy. The Hanssen & Kuipers skill score of the ramp forecasts made by applying Equation 6.1 to the different wind power forecasts in Section 6.2 are included as a reference with dotted lines.

6.4 Conclusions

The focus of this chapter has been on forecasting large and rapid changes in the wind power production, normally referred to as *wind power ramps*. Wind power ramps are relatively rare and often hard to forecast, but they have a potentially very large impact on the power system and are therefore of high importance.

First it was examined whether applying different correction schemes to NWP wind power forecasts can increase the predictability of wind power ramps. A very simple definition of a wind power ramp (see Equation 6.1) was used both to make categorical ramp forecasts (“up-ramp”, “down-ramp” or “no ramp”) from time-series of wind power forecasts and to identify true observed ramps from a time-series of wind power measurements. The latter was used to evaluate the ramp forecasts based on the different correction schemes. Ramp forecasts based on three different corrected wind power forecasts (directional bias-correction (see Section 3.2), auto-correlation and spatio-temporal regression (see Section 4.3.2)) was compared to a ramp forecast based

on the raw NWP forecast using four different evaluation metrics. To test whether the results were consistent for different magnitudes of wind power ramps three different ramp change limits were used (20%/30%/40% change in group wind power within 3 hours).

For all three ramp change limits it was found that forecasts based on the corrected wind power forecasts gives improvements compared to the forecast based on uncorrected NWP forecasts. For ramp change limits 0.2 and 0.3 it was also found that the regression correction schemes (auto-correlation and spatio-temporal regression) comes out superior to the directional bias-correction. For ramp change limit 0.4 the image is not as clear and the directional bias-correction comes out as good as or slightly better than the spatio-temporal regression. The wind power ramp forecasts based on the uncorrected and directional bias-corrected wind power forecasts tends to give a massive over-forecasts of ramps, forecasting respectively 3 to 8 and 1.6 to 2.2 times as many ramps as is observed (see Figure 6.5). As a result of this, the percentage of correctly forecasted ramps is high (see Figure 6.3), but so is also the risk of forecasting false ramps (see Figure 6.4).

Thereafter the classification technique Random Forests was used to make ramp forecasts from the same four wind power forecasts (raw NWP, directional bias-corrected, auto-correlation and spatio-temporal regression). The results were compared to the ramp forecasts made by applying Equation 6.1 to the same wind power forecasts (used as a reference forecast). The Random Forest ramp forecasts was found to forecast far fewer ramps than observed (see Figure 6.13), a tendency that increased with increasing ramp change limit and with more extensive correction of the wind power forecasts. Based on the uncorrected and the directional bias-correction wind power forecasts, the Random Forest ramp forecast was found to give clear improvements over the reference ramp forecast. Based on the auto-correlation wind power forecast the improvements were limited and based on the spatio-temporal wind power forecast the Random Forests ramp forecast performed worse than the reference. The reason for this is most likely that the Random Forest model benefits from the same relations in the NWP wind power forecasts as is modelled by the spatio-temporal regression model. When this information is removed from the input data, the Random Forest models struggles with identifying patterns that give a high probability of wind power ramps. For both ramp change limits (0.2 and 0.3) the Random Forest ramp forecast was also found to give nearly 10 % higher improvement over a random ramp

forecast strategy than the reference forecast (HK score, see Figure 6.14). The main reason for this is the Random Forest models ability to avoid forecasting false ramps.

When trying to use the Random Forest model to forecast ramps exceeding ramp change limit 0.4 (40 % change in group wind power within 3 hours) the model forecasted all observations as no ramps. The main reason for this is the very low number of true observed ramps for this limit (less than 1 % of the observations, see Table 6.1), but it is also partly caused by how the Random Forest model is built. The OOB estimates (see Section 6.3) that are used to evaluate the model performance in the model building process are a measure of the misclassification rate and hence vulnerable to imbalanced data (i.e. data where the classes make up very different shares of the observations). In such situations classifying all observations into the majority class will guarantee a low misclassification rate (for ramp change limit 0.4 less than 1 %, see Table 6.1) and thus will the majority class be promoted over the minority classes. This is also an important part of the cause of the under-forecasting of ramps for ramp change limits 0.2 and 0.3. The same problem is also present in the HK evaluation metric, where likewise all correct and incorrect forecasts are given equal weight. The result of this is that the metrics put more emphasis on the majority class than the minority classes.

For an application like wind power ramp forecasting this is a major disadvantage, as it is of particular importance to forecast the minority classes (ramps) correctly. It is for example likely to believe that correctly forecasting the up-ramps and down-ramps would be of higher importance than correctly forecasting the no ramps. This calls for metrics that can give uneven weight to the different cells of the contingency table. One solution to this would be to assign monetary values to the different outcomes, to sum these over the test period and to use the sum as a criterion for evaluation. From an operational point of view this would be a realistic approach, but to set realistic values for the different outcomes would not be trivial.

Large variations in the ramp definitions, forecast horizons and presentation of wind power ramp forecasts makes comparison of the results in this chapter with results from most other ramp forecast strategies presented in the literature difficult. However, by allowing a small deviation in the ramp definition, results reported by Greaves *et al.* (2009) and Suzuki *et al.* (2012) are fairly comparable. In Greaves et al. (2009) a 3 hour forecast of ramps exceeding a 50 % threshold limit made for a portfolio of sites in UK is presented. The sensitivity of the ramp forecasts is found to be 0.5. In Suzuki *et al.*

(2012) the sensitivity of a 3 hour forecast for a single site in the Northern Great Plains, USA, is found to be 0.3. Both results are within the range of what was found when making ramp forecasts by applying Equation 6.1 to the corrected wind power forecasts in Section 6.2. It should though be noted that sites-specific conditions have a large influence on the results. In order to make direct comparisons of methods this would need to be adjusted for.

Chapter 7 – Conclusions and Perspectives

7.1 Conclusions

The present thesis has methods for improving and enhancing NWP based wind power forecasts for cases from the Western coast of Norway. The data basis for the thesis consisted of wind observations and NWP wind forecast for 43 sites, all covering the period from January 1st 2009 to December 17th 2011. The observations and forecasts were transformed into synthetic wind power forecasts and observations by the use of a logarithmic height profile and a generic power curve.

The Western coast of Norway is an area with excellent wind conditions, but where the installed wind power capacity at present is very limited. As a result of this, the efforts put into developing wind power forecasts for the area has up to now been sparse. The area is characterized by a rugged coastline and complex terrain. Kariniotakis *et al.* (2004) and Martí *et al.* (2006) have shown that performance of the state-of-the-art wind power forecast models varies with terrain complexity, sites located in complex terrain obtaining significantly higher forecast errors than sites in less complex terrain. Following Möhrlen (2004) one reason for this is that it is more difficult to achieve an accurate parameterization of the terrain surface in complex conditions and that this leads to a lower quality of the NWP forecast input. The overall aim of the thesis have been to study how this kind of conditions influence wind power forecast errors and to investigate how wind power forecast models can be made more resistant to the challenges these conditions pose.

First the wind power forecast errors of the single sites were investigated with the aim of establishing correction schemes capable of correcting systematic errors (bias) in the single site forecasts. It was found that the look-ahead time (time of day), the wind direction and the wind speed gave systematic contributions to the forecast errors. It was further found that the site-to-site variations were large, hence that the correction schemes needed to be decided for each site individually. A correction scheme accounting for both wind direction and wind speed was found to give the largest reduction in RMSE when applied to the same data as the forecast errors were calculated from, but when applied to new, unseen data the performance of this scheme was poor. This was caused by lack of data from high wind speeds and non-prevailing

wind directions, meaning that the estimates that are supposed to be of systematic errors are based on too few observations to be generalized. A correction scheme based on the forecast wind directions (split into eight equally sized sectors) were found to give in average the largest reduction in bias and also resulting in the in average lowest RMSE.

Thereafter the single sites were gathered into seven groups through a clustering procedure. It was demonstrated that the lumped wind power forecast errors of the groups were subject to a clear spatial error-smoothing effect compared to the single sites. This effect is well known from e.g. Focken *et al.* (2002b) Siebert (2008) and Girard & Allard (2013), but for the present case the smoothing effect seemed to increase more rapidly with the spatial extent of the group. A likely reason for this is the differences in terrain complexity between west-coast Norway and northern Germany and Denmark, but it might also a result of the low measurement height (10 meter a.g.l.).

Based on an assessment of the spatial and temporal dependencies of the forecast errors within and between the groups of sites, two different regression models (linear regression and regime-switch regression depending on forecast wind direction), both previously applied in Tastu *et al.* (2011), were tested in an attempt to model the spatio-temporal propagation of the forecast errors. When applied to the full dataset the more complex models came out as superior to the simpler models (see Tables 4.2 and 4.3), but when tested against "unseen" data through cross-validation the simplest linear model came out with the lowest RMSE (see Table 4.4). This means that the directional modelling of the propagation of forecast errors did not give positive contributions to the forecasts for the present case. This is partly caused by the weak signs of error propagation, which in its turn is caused by the large spatial extent of the groups and the complex terrain both within and between the groups. Another important reason is lack of data for the wind directions outside the prevailing wind direction, leading to unstable parameter estimates for these wind directions. However, compared to the forecast error of the directional-corrected NWP forecast, both models gave a considerable reduction of the RMSE (from 16.85 % to between 8.60 % and 8.87 %, see Table 4.4).

Attempting to solve the challenges with benefiting from the directional information, the influence of the choice of wind sectors in the regime-switch regression models was assessed and a method to derive data-driven wind sectors derived. When evaluating the model performance on "unseen" data through cross-validation the differences

between a good and a poor choice of wind sectors was found to be more than 21 % (see Figure 4.11). It was further shown that that it is possible to create data-driven wind sectors through clustering procedures, but the improvements from these compared to a reference set of commonly used wind sectors (0°- 90°, 90°- 180° etc.) were very limited. For most applications it is likely to be as good to choose wind sectors based on knowledge about prevailing wind directions and local orography.

In order to proceed from point forecasts to probabilistic forecasts, the possibilities for using the statistical post-processing method Bayesian Model Averaging (BMA) to make probabilistic forecasts of the output of groups of sites from the single site group members point forecasts was explored (see Chapter 5). It was found that the lumped group wind power could be described by the Beta distribution. The mean of the Beta distribution could be estimated from the single site forecast errors by linear regression (see Figures 5.6 and 5.7). No apparent way of estimating the standard deviation of the Beta distribution as a function of the single site forecast errors was found. This was therefore calculated from historic data and organized as site-specific look-up tables. This is not an ideal solution as it involves sacrificing parts of the time-adaptive properties of the BMA method, and it was also found that some sites suffered from scarcity of data from high-production situations, causing the standard deviation estimates to be inaccurate.

To evaluate the performance of the BMA probabilistic forecasts four different prediction intervals (50 %, 75 %, 90 % and 95 %) for the next-hour lumped group wind power was estimated and compared to the observed wind power output. It was found that the prediction intervals with confidence levels 95 %, 90 % and 75 % fairly accurately covered the corresponding share of the data (for the ideal forecast the coverage rate should match the confidence level). For confidence level 50 % the prediction intervals covered a too high share of the observations, indicating that the BMA predictive PDF is too little peaked (too low kurtosis). It was also found that the distribution of the observations that fell outside the confidence limits was very unevenly distributed for all confidence levels. One reason for this is the previously mentioned scarcity of data for some sites with subsequent unreliable estimates of the standard deviation of the Beta distribution. Another problem encountered was that the Beta distribution cannot reach exactly 0 (0 % group power output) or 1 (100 % group power output). As especially hours of 0 % production are common in the data (100 % production is very rare and therefore of less importance) this means that the confidence limits are unable to cover a significant share of the observations. This was

solved by setting a lower threshold limit (0.025) for which lower Beta values were considered as 0. This is a quick and easy solution to the problem, but not particularly elegant. Some ideas for other solutions are proposed in the outlook below.

Finally a special focus was put on wind power ramps – large and sudden changes in the wind power production – here defined as a 20/30/40 % change in lumped group wind power within 3 hours. First it was examined whether applying different correction schemes (directional bias-correction (see Section 3.2), auto-correlation and spatio-temporal regression (see Section 4.3.2)) to NWP wind power forecasts can increase the predictability of wind power ramps. A very simple definition of a wind power ramp (see Equation 6.1) was used both to make categorical ramp forecasts ("up-ramp", "down-ramp" or "no ramp") from time-series of wind power forecasts and to identify true observed ramps from a time-series of wind power observations. The latter was used to evaluate the ramp forecasts based on the different correction schemes. For moderate ramps (20 % and 30 % change) it was also found that the regression correction schemes (auto-correlation and spatio-temporal regression) gave higher ramp predictability than the directional bias-correction and the uncorrected reference forecast. For ramps exceeding 40 % change the image was less clear. The wind power ramp forecasts based on the uncorrected NWP and the directional bias corrected wind power forecasts gave massive over-forecasts of ramps, forecasting respectively 3 to 8 and 1.6 to 2.2 times as many ramps as was observed. This naturally led to a high percentage of correctly forecasted ramps, but it came at the expense of a very high number of false alarms.

Thereafter the classification technique Random Forests was used to forecast wind power ramps from the same four wind power forecasts. The Random Forest ramp forecasts was found to forecast far fewer ramps than was observed, a tendency that increased with increasing ramp change limit and with the more extensively corrected wind power forecasts (auto-correlation and spatio-temporal regression). Based on the uncorrected and the directional bias-corrected wind power forecasts the Random Forest ramp forecast was found to give clear improvements over the reference ramp forecast (the simple ramp definition from Equation 6.1). Based on the auto-correlation corrected wind power forecast the improvements were limited and based on the spatio-temporal wind power forecast the Random Forests ramp forecast performed worse than the reference. The reason for this is most likely that the Random Forest model benefits from the same relations in the NWP wind power forecasts as is modelled by the spatio-temporal regression model. When attempting to forecast ramps exceeding

40 % change, the Random Forests model forecasted all observations as no ramps. The reasons for this is the very low number of true observed ramps for this limit (less than 1 % of the observations) and the evaluation criteria used in the model building process (misclassification rate) that is vulnerable to imbalanced data (i.e. data where the classes make up very different shares of the observations). In such situations classifying all observations into the majority class will guarantee a low misclassification rate (for ramp change limit 0.4 less than 1 %) and thus will the majority class be promoted over the minority classes.

7.2 Perspectives

When the work on this thesis started it was the hope to obtain actual measurements of the wind power production from most, if not all, wind farms in Norway. Unfortunately this was not possible, and as a back-up solution a synthetic dataset was constructed from wind speed and wind direction measurements made at 10 meter a.g.l. at synoptic meteorological stations. For a thorough presentation of the data it is referred to Chapter 2. This had the advantage that the number of available sites was larger than it would have been if considering actual wind farms only, but it also meant introducing some major uncertainties that would not have been present in a real wind power dataset (e.g. height transformation and power curve transformation). That the measurements were taken from single sites also means that they are not subject to the intra-farm spatial smoothing that will be present in wind farms due to different locations of the turbines, upwind-downwind interactions etc. Late 2014 there were made an agreement between the Norwegian Water Resource and Energy Directorate (NVE) and the wind power companies in Norway on the release of production data from all Norwegian wind farms aggregated to 1 hour temporal resolution with a 6 months delay. This agreement came in place too late to be of benefit for this thesis, but as the availability of actual wind power observations grows, the conclusions of this thesis should be validated against real production data.

Another problem that was encountered on the data-side was the difficulty to determine whether systematic errors in the forecasts were caused by the look-ahead time or the time of day for which the forecast was issued (see Chapter 3). Although it is likely to believe that the majority of the temporal variations in the single site forecast errors for this case were caused by the time of day, access to more rapidly updated forecasts with

overlap would allow differentiating between these two sources of influence. This might also make it possible to derive improved correction schemes capable of removing more of the temporal variations in the forecast errors.

When expanding the single-site forecasts into forecasts for groups of sites, one obvious challenge was the low number of measurement sites (43 sites) spread out over a very large area (~3000 km). Even though the groups were kept compact, the spatial spreads of the groups were considerable (140 km – 355 km). This has the effect that the variations between the group members both in terms of wind power production and wind direction should be expected to be large, and as a result of this that it will be difficult to derive single measures of wind power production and wind direction that are representative for the whole group. In combination with complex terrain both within the groups and between the groups, this makes it difficult to model the propagation of forecast errors. Although there most likely are sources of data that have not been included in this thesis, it will not be possible to obtain the same density of measurements as is possible in e.g. Denmark. A strategy to still make use of information on forecast error propagation could be to use more dynamic groups of sites and let the composition of the groups vary with each time-step. This would make it possible to tailor groups of sites which at a given time are very similar with respect to wind power production and wind direction and to follow the propagation of forecast errors over shorter distances. The approach might however also lead to different group compositions containing partly the same sites and covering parts of the same area, i.e. will there be a need for a method to choose between competing forecasts and to combine forecasts for small groups into forecasts for larger areas.

The BMA approach for making probabilistic forecasts proved partly successful, but still has a large potential for further improvements. The two most obvious problems with the current implementation was the inability of the Beta distribution to model values of exactly 0 (no wind power production) or 1 (full wind power production) and the imbalance in the observations falling outside the prediction intervals (see Tables 5.1 and 5.2). A potential solution to these problems is to try the BMA-model with other distributions for the component PDFs. Candidates would include piecewise PDFs (see e.g. Chmielecki & Raftery 2010) and left-censored or interval-censored distributions (see e.g. Pinson 2012, Tastu *et al.* 2012 and Messner *et al.* 2013). Caution should however be exercised when trying to fit more complex PDFs as these are likely to require more training data to obtain reliable parameter estimates.

One of the most interesting aspects of BMA is the flexibility of the model. Any relevant source of data can be included without major adaptations, and it is also possible to make implementations where different component PDFs are assumed for different sources. There is neither any obstacle for implementing data-driven selection of the component PDFs, but as always additional complications makes the model more computationally expensive and might also increase the demand for training data. This is especially critical if the model parameters cannot be estimated analytically and numerical procedures are required. One obvious expansion of the BMA model in this thesis would be to use NWP ensemble forecast for each of the group members, combining the reference-site approach with an empirical optimization of the choice of NWP parameterization. Other options would be to build BMA models based on analog ensembles for the group members or on the output of regression models linking one or more single sites (and potential other variables) to the lumped group power output.

At last, making accurate forecasts of wind power ramps is an important, but very difficult task that for good reasons has received increasing attention later years. The Random Forests ramp forecast model applied in this thesis is an interesting model that has some desirable properties in its computational efficiency, interpretability and ability to handle non-linear relations in input data, but in the present implementation it was not particularly well suited for wind power ramp forecasting. The main problem with the model, the inability to handle very imbalanced data, has some potential solutions. Chen *et al.* (2004) discuss a number of up- and down-sampling techniques improving the forecast accuracy of the minority classes. Another way of handling this would be through an alternative evaluation criterion in the model building process, giving increased importance to accurate forecasting of the minority classes. For wind power ramp applications it would be natural to link such an evaluation criterion to the economic gain or loss from the forecast. Considering the variability of electricity prices, assigning realistic values to the different outcomes might however not be a trivial task.

References

Agresti, A (2007), *An introduction to categorical data analysis*, Wiley, Hoboken, New Jersey, USA

Ahlstrom, M *et al.* (2013), 'Knowledge Is Power: Efficiently Integrating Wind Energy and Wind Forecasts', *IEEE Power and Energy Magazine*, Vol. 11, Issue 6, pp. 45-52

Akaike, H (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatics Control*, Vol. 19, No. 6, pp. 716-723

Alessandrini, S *et al.* (2012), 'An Application and Verification of Ensemble Forecasting on Wind Power to Assess Operational Risk Indicators in Power Grids', in *Proceedings of the 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Lisbon, Portugal, pp. 91-94

Alessandrini, S *et al.* (2015), 'A novel application of an analog ensemble for short-term wind power forecasting', *Renewable Energy*, 76, pp. 768-781

Atger, F (1999), 'The Skill of Ensemble Prediction Systems', *Monthly Weather Review*, Vol. 127, pp. 1941-1953

Bao, L *et al.* (2010), 'Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction', *Monthly Weather Review*, Vol. 138, pp. 1811-1821

Barton, JP & Infield, DG (2004), 'Energy Storage and Its Use With Intermittent Renewable Energy', *IEEE Transactions on Energy Conversion*, Volume 19, Issue 2, June 2004, pp. 441-448

Berge, E *et al.* (2003), *Forecasting wind and wind energy production in Norwegian wind farms*, Kjeller Vindteknikk, KVT/EB/2003/005, Kjeller, Norway

Berge, E *et al.* (2006), 'An Evaluation of the WAsP Model at a Coastal Mountainous Site in Norway', *Wind Energy*, 9, 2006, pp. 131-140

Beyer, HG *et al.* (1999), 'Forecast of regional wind power output of wind turbines', in *Proceedings of the 1999 European Wind Energy Conference EWEC'99*, Nice, France, pp. 1070-1073

Blaudszuweit, H *et al.* (2008), 'Statistical Analysis of Wind Power Forecast Error', *IEEE Transactions on Power Systems*, Vol. 23, No. 3, pp. 983-991

Bofinger, S *et al.* (2002), 'Qualification of wind power forecasts', in *Proceedings of the 2002 Global Wind Power Conference GWPC'02*, Paris, France

Bossavy, A *et al.* (2010), 'Forecasting Uncertainty Related to Ramps of Wind Power Production', in *Proceedings of the European Wind Energy Conference and Exhibition EWEC'10*, Warsaw, Poland

Botterud, A *et al.* (2012a), 'Wind power trading under uncertainty in LMP markets', *IEEE Transactions on Power Systems*, Vol. 27, No. 2, pp. 894-903

Botterud, A *et al.* (2012b), 'Demand Dispatch and Probabilistic Wind Power Forecasting in Unit Commitment and Economic Dispatch: A Case Study of Illinois', *IEEE Transactions on Sustainable Energy*, Vol. 4, Issue 1, pp. 250-261

Bowen, AJ & Mortensen, NG (1996), 'Exploring the limits of WAsP the wind atlas analysis and application program', in *Proceedings of the European Wind Energy Conference EWEC'96*, Gothenburg, Sweden

Box, GEP *et al.* (2008), *Time Series Analysis: Forecasting and Control (4th ed.)*, Wiley, Hoboken, New Jersey, USA

Bradbury, LJS (2013), 'The probability density distribution for the power output from arrays of wind turbines and the intermittent nature of wind power', *Journal of Wind Engineering and Industrial Aerodynamics*, Volume 123, Part A, pp. 121-129

Breiman, L *et al.* (1984), *Classification and Regression Trees*, Wadsworth inc., Belmont, California, USA

Breiman, L (2001), 'Random Forests', *Machine Learning*, 45(2001), pp. 5-32

Bremnes, JB (2004), 'Probabilistic Wind Power Forecasting Using Local Quantile Regression', *Wind Energy*, Vol. 7, Issue 1, pp. 47-54

Bremnes, JB (2006), 'A Comparison of a Few Statistical Models for Making Quantile Wind Power Forecasts', *Wind Energy*, Vol. 9, Issue 1-2, pp. 3-11

Bremnes, JB & Giebel, G (2014), 'Does wind power forecasting skill depend on the spatial resolution of NWP models?', presentation at *IceWind Final Conference*, 3rd December 2014, Aarhus, Denmark. http://www.icewind.dtu.dk/~/media/Sites/Icewind/Final%20conf/4%20John%20ICEWIND_3.1_20141203.ashx [12.01.15]

Bremnes, JB & Homleid, M (2009a), *Verification of Operational Numerical Weather Prediction Models Desember 2008 to February 2009*, Met.no note no. 5-2009, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2009/filestore/report_200812-200902.pdf [05.11.14]

Bremnes, JB & Homleid, M (2009b), *Verification of Operational Numerical Weather Prediction Models March to May 2009*, Met.no note no. 17-2009, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-

_2013/Publikasjoner_2009/filestore/report_200903-200905.pdf [05.11.14]

Bremnes, JB & Homleid, M (2009c), *Verification of Operational Numerical Weather Prediction Models June to August 2009*, Met.no note no. 23-2009, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2009/filestore/note23_2009.pdf [05.11.14]

Bremnes, JB & Homleid, M (2009d), *Verification of Operational Numerical Weather Prediction Models September to November 2009*, Met.no note no. 28-2009, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2009/filestore/note28-2009.pdf [05.11.14]

Bremnes, JB & Homleid, M (2010a), *Verification of Operational Numerical Weather Prediction Models Desember 2009 to February 2010*, Met.no note no. 9-2010, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2010/filestore/report_200912-201002.pdf [05.11.14]

Bremnes, JB & Homleid, M (2010b), *Verification of Operational Numerical Weather Prediction Models March to May 2010*, Met.no note no. 11-2010, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2010/filestore/report_11_2010.pdf [05.11.14]

Bremnes, JB & Homleid, M (2010c), *Verification of Operational Numerical Weather Prediction Models June to August 2010*, Met.no note no. 15-2010, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2010/filestore/metno_note_15_2010.pdf [05.11.14]

Bremnes, JB & Homleid, M (2011a), *Verification of Operational Numerical Weather Prediction Models September to November 2010*, Met.no note no. 5-2011, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2011/filestore/note5_report_201009-201011.pdf [05.11.14]

Bremnes, JB & Homleid, M (2011b), *Verification of Operational Numerical Weather Prediction Models Desember 2010 to February 2011*, Met.no note no. 6-2011, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2011/filestore/note6_report_201012-201102.pdf [05.11.14]

Bremnes, JB & Homleid, M (2011c), *Verification of Operational Numerical Weather Prediction Models March to May 2011*, Met.no note no. 7-2011, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2011/filestore/note7_report_201103-201105.pdf [05.11.14]

Bremnes, JB & Homleid, M (2011d), *Verification of Operational Numerical Weather Prediction Models June to August 2011*, Met.no note no. 8-2011, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-

_2013/Publikasjoner_2011/filestore/note8_report_201106-201108.pdf [05.11.14]

Bremnes, JB & Homleid, M (2011e), *Verification of Operational Numerical Weather Prediction Models September to November 2011*, Met.no note no. 12-2011, Oslo, Norway. http://met.no/Forskning/Publikasjoner/Publikasjoner_1995_-_2013/Publikasjoner_2011/filestore/report_201109-201111.pdf [05.11.14]

Cali, Ü *et al.* (2008), 'Artificial neural network based wind power forecasting using a multi-model approach', in *Proceedings of the 7th International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms*, Madrid, Spain

Carrasco, JM *et al.* (2006), 'Power-Electronic Systems for the Grid Integration of Renewable Energy Sources: A Survey', *IEEE Transactions on Industrial Electronics*, Vol. 53, No. 4, August 2006, pp.1002-1016

Casati, B *et al.* (2008), 'Forecast verification: current status and future directions', *Meteorological Applications*, Vol. 15, pp. 3-18

Castellani, F *et al.* (2014), 'Wind Energy Forecasts in Complex Sites with a Hybrid Neural Network and CFD based Method', *Energy Procedia*, Volume 45, 2014, pp. 188-197

Castronuovo, ED *et al.* (2014), 'An integrated approach for optimal coordination of wind power and hydro pumping storage', *Wind Energy*, Vol. 17, Issue 6, pp. 829-852

Chatfield, C (2003), *The Analysis of Time Series: An Introduction (6th ed.)*, Chapman & Hall/CRC, New York, USA

Chen, C *et al.* (2004), *Using Random Forests to Learn Imbalanced Data*, Tech report 666, Department of Statistics, University of California, Berkley, USA, http://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf [07.01.15]

Chmielecki, RM & Raftery, A (2010), 'Probabilistic visibility forecasting using Bayesian model averaging', *Monthly Weather Review*, Vol. 139, pp. 1626-1636

Collier, C *et al.* (2013), 'Use of Offsite Data to Improve Short Term Ramp Forecasting', in *Proceedings of the 2013 European Wind Energy Association Annual Event*, Vienna, Austria

Costa, A *et al.* (2008), 'A review on the young history of the wind power short-term prediction', *Renewable and Sustainable Energy Reviews*, Vol. 12, Issue 6, pp. 1725-1744

Courtney, JF *et al.* (2013), 'High resolution forecasting for wind energy applications using Bayesian model averaging', *Tellus A*, 2013, 65, pp. 1-13

Couto, A *et al.* (2013), 'Impact of Weather Regimes on the Wind Power Ramp Forecasts', in *Proceedings of the 12th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*,

London, UK, pp. 123-128

Cutler, N *et al.* 2007, 'Detecting, Categorizing, and Forecasting Large Ramps in Wind Farm Power Output Using Meteorological Observations and WPPT', *Wind Energy*, Vol. 10, Issue 5, pp. 453-470

Damousis, IG *et al.* (2004), 'A Fuzzy Model for Wind Speed Prediction and Power Generation in Wind Parks Using Spatial Correlation', *IEEE Transactions on Energy Conversion*, Vol. 19, Issue 2, pp. 352-361

Davenport, AG *et al.* (2000), 'Estimating the roughness of cities and sheltered country', in *Proceedings of the 12th American Meteorological Society Conference on Applied Climetology*, Asheville, North Carolina, USA, pp. 96-99

Davis, C *et al.* (2010), 'Does Increased Horizontal Resolution Improve Hurricane Wind Forecasts?', *Weather and Forecasting*, 25(6), pp. 1826-1841

De Boer, W *et al.* (2012), 'A Comparison between Intraday Market and Capacity Market to Deal with Wind Power Forecasting Errors', in *Proceedings of the 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Lisbon, Portugal, pp. 421-426

De Boer, W *et al.* (2013), 'A New AS Market to Deal with Wind Power Forecasting Errors: Efficiency Assessment and Analysis of its Capability to Free up More Regulating Power', in *Proceedings of the 12th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, London, UK, pp. 475-480

Denholm, P *et al.* (2010), *The Role of energy storage with renewable electricity generation*, NREL Technical Report 1-2010, http://digitalscholarship.unlv.edu/renew_pubs/5 [22.10.14]

Dietrich, K *et al.* (2009), 'Stochastic unit commitment considering uncertain wind production in an isolated system', *4th Conference on Energy Economics and Technology*, Dresden, Germany

Dobschinski, J *et al.* (2008), 'Estimation of wind power prediction intervals using stochastic methods and artificial intelligence model ensembles', in *Proceedings of the German Wind Energy Conference DEWEK*, Bremen, Germany

Doherty, R & O'Malley, M (2005), 'A New Approach to Quantify Reserve Demand in Systems with Significant Installed Wind Capacity', *IEEE Transactions on Power Systems*, Vol. 20, pp. 587-595

Dubath, P *et al.* (2011), 'Random forest automated supervised classification of *Hipparcos* periodic variable stars', *Monthly Notices of the Royal Astronomical Society*, Vol. 414, No. 3, pp. 2602-2617

Duque, Á *et al.* (2011), 'Optimal operation of a pumped-storage hydro plant that compensates the imbalances of a wind power producer', *Electric Power Systems Research*, Vol. 81, No. 9, pp. 1767-1777

Ebert, E *et al.* (2013), 'Progress and challenges in forecast verification', *Meteorological Applications*, Vol. 20, Issue 2, pp 130-139

Ehrendorfer, M (1997), 'Predicting the uncertainty of numerical weather forecasts: a review', *Meteorologische Zeitschrift*, No. 6, pp. 147-183

Enercon (2010), *Enercon wind energy converters. Product overview*, Enercon 2010, http://www.enercon.de/p/downloads/EN_Productoverview_0710.pdf [29.10.14]

Eriksen, PB *et al.* (2005), 'System operation with high wind penetration', *IEEE Power & Energy Magazine*, 3(5), pp. 65–74

Ernst, B *et al.* (2001), 'Managing 3000 MW power in a transmission system operation centre', in *Proceedings of the European Wind Energy Conference EWEC'01*. Copenhagen, Denmark, pp. 890-893

Estanqueiro, A *et al.* (2012), 'Contribution of Energy Storage for Large-scale Integration of Variable Generation', in *Proceedings of the 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Lisbon, Portugal, pp.499-505

Everitt, BS (1993), *Cluster Analysis (3rd ed.)*, Edward Arnold, London, UK

EWEA (2010), *Wind in power – 2009 European statistics*, The European Wind Energy Association, February 2010. http://www.ewea.org/fileadmin/files/library/publications/statistics/Wind_In_Power_European_Statistics_2009.pdf [21.10.14]

EWEA (2014a), *Wind in Power – 2013 European statistics*, The European Wind Energy Association, January 2014. http://www.ewea.org/fileadmin/files/library/publications/statistics/EWEA_Annual_Statistics_2013.pdf [21.10.14]

EWEA (2014b), *Wind energy scenarios for 2020*, The European Wind Energy Association, July 2014, http://www.ewea.org/fileadmin/files/library/publications/scenarios/EWEA-Wind-energy-scenarios-2020.pdf [21.10.14]

Ferreira C *et al.* (2010), *A Survey on Wind Power Ramp Forecasting*, Report ANL/DIS-10-13, Argonne National Laboratory, Argonne, Illinois, USA

Florita, A *et al.* (2013), 'Identifying Wind and Solar Ramping Events', presented at *IEEE 5th Green Technologies Conference*, Denver, Colorado, USA, http://www.nrel.gov/docs/fy13osti/57447.pdf [13.01.15]

Focken, U *et al.* (2001), 'Previento – A wind power prediction system with an innovative upscaling algorithm', in *Proceedings of the European Wind Energy Conference EWEC'01*, Copenhagen, Denmark, pp. 826-829

Focken, U *et al.* (2002a) 'Previento – Regional wind power prediction with risk control', in *Proceedings of the 2002 Global Wind Power Conference*, Paris, France

Focken, U *et al.* (2002b), 'Short-term prediction of the aggregated power output of wind farms – A statistical analysis of the reduction of the prediction error by spatial smoothing effects', *Journal of Wind Energy and Industrial Aerodynamics*, Vol. 90, Issue 3, pp. 231-246

Foley, AM *et al.* (2012), 'Current methods and advances in forecasting of wind power generation', *Renewable Energy*, 37 (2012), pp. 1-8

Frías, L (2009), 'Support Vector Machines in the wind energy framework. A new model for wind energy forecasting', in *Proceedings of the 2009 European Wind Energy Conference EWEC'09*, Marseilles, France.

Fugon, L *et al.* (2008), 'Data mining for wind power forecasting', in *Proceedings of the 2008 European Wind Energy Conference EWEC'08*, Brussels, Belgium

Giebel, G *et al.* (1999), 'Relative Performance of different Numerical Weather Prediction Models for Short Term Prediction of Wind Energy', in *Proceedings of the European Wind Energy Conference EWEC'99*, Nice, France

Giebel, G *et al.* (2005), *Wind power prediction using ensembles*, Risø-R-1527(EN), Technical report, Risø National Laboratory, Roskilde, Denmark

Giebel, G *et al.* (2011), *The State if the Art in Short-Term Prediction of Wind Power, A Literature Overview, 2nd Edition*, ANEMOS.plus/SafeWind

Gipe, P (2004). *Wind Power*, James & James Ltd, London, UK

Girard, R & Allard, D (2013), 'Spatio-temporal propagation of wind power prediction errors', *Wind Energy*, Vol. 16, Issue 7, pp. 999-1012

Gislason, PO *et al.* (2006), 'Random Forests for land cover classification', *Pattern Recognition Letters*, Vol. 27, No. 4, pp. 294-300

Gneiting, T *et al.* (2006), 'Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method', *Journal of the American Statistical Association*, Vol. 101, No. 475, pp. 968-979

Greaves, B *et al.* (2009), 'Temporal Forecast Uncertainty for Ramp Events', in *Proceedings of the European Wind Energy Conference and Exhibition EWEC'09*, Marseille, France

Hanna, SR & Yang, R (2001), 'Evaluations of Mesoscale Moldels' Simulations of Near-Surface Winds, Temperature Gradients, and Mixing Depths', *Journal of applied Meteorology*,

Vol. 40, pp 1095-1104

Hanssen, AW & Kuipers, WJA (1965), 'On the relationship between the frequency of rain and various meteorological parameters', *Mededelingen van de Verhandlaugen*, 81, pp. 2-15

Han, Y & Chang, L (2010), 'A study of the reduction of the regional aggregated wind power forecast error by spatial smoothing effects in the Maritimes Canada', in *Proceedings of the 2010 IEEE Electric Power and Energy Conference (EPEC)*, Halifax, Canada

Hashimoto, A *et al.* (2007), 'Effects of Numerical Models on Local-Wind Forecasts over a Complex Terrain with Wind Farm Prediction Model', in *Proceedings of the European Wind Energy Conference EWEC'07*, Milano, Italy

Hastie, T *et al.* (2009), *The Elements of Statistical Learning. Data mining, Inference, and Prediction (2nd Ed.)*, Springer, New York, USA

Haugen, JE (2008), *Operasjonell Hirlam med 12km, 8 km og 4km gitter – beskrivelse av modelloppsett fra februar 2008 med verifikasjon fra parallellkjøringer*, Met.no note no. 1-2008, Oslo, Norway

Hering, AS & Genton, MG (2010), 'Powering Up With Space-Time Wind Forecasting', *Journal of the American Statistical Association*, Vol. 105, No. 489, pp. 92-104

Hodge, BM & Milligan, M (2011), 'Wind Power Forecasting Error Distributions over Multiple Timescales', in *IEEE Power & Energy Society General Meeting*, Detroit, Michigan, USA

Hodge, BM *et al.* (2012), 'Wind Power Forecasting Error Distributions: An International Comparison', in *Proceedings of the 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Lisbon, Portugal

Hoeting, JA *et al.* (1999), 'Bayesian Model Averaging: A Tutorial', *Statistical Science*, 14, pp. 382-401

Holttinen, H (2004), *The Impact of Large Scale Wind Power Production on the Nordic Electricity System*, Ph.D. Dissertation, Helsinki University of Technology, Helsinki, Finland

Holttinen, H *et al.* (2013), 'Methodologies to determine operating reserves due to increased wind power', *IEEE Power and Energy Society General Meeting (PES)*, Vancouver, Canada

Ipakchi, A & Albuyeh F (2009), 'Grid of the Future. Are We Ready to Transition to a Smart Grid?', *IEEE Power Energy Magazine*, Vol. 2, Issue 7, pp. 52-62

Jiang, P *et al.* (2007), 'MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features', *Nucleic Acids Research*, Vol. 35, No. 2, pp. 339-344

Jolliffe, IT & Stephenson, DB (ed.) (2003), *Forecast Verification. A Practioner's Guide in Atmospheric Science*, Wiley, Chichester, England

Johnson, RA & Wichern DW (2002), *Applied Multivariate Statistical Analysis, Fifth Edition*, Prentice Hall, Upper Saddle River, NJ, USA, pp. 76-77

Juban, J *et al.* (2008), 'Uncertainty Estimation of Wind Power Forecasts', in *Proceedings of the European Wind Energy Conference EWEC'08*, Brussels, Belgium

Justus, CG *et al.* (1976), 'Nation-wide assessment of potential output from wind power generators', *Journal of Applied Meteorology*, Vol. 15, Issue 7, pp. 673-678

Kalnay, E *et al.* (1998), 'Maturity of operational numerical weather prediction: medium range', *Bulletin of the American Meteorological Society*, 79(12), pp. 2753-2769

Kamath, C (2010), 'Understanding Wind Ramp Events through Analysis of Historical Data', in *Proceedings of the IEEE PES Transmission and Distribution Conference and Expo*, New Orleans, Louisiana, USA

Kariniotakis, G *et al.* (1996a), 'Wind power forecasting using advanced neural networks models', *IEEE Transactions On Energy Conversion*, Vol. 11, Issue 4, pp. 762-767

Kariniotakis, G *et al.* (1996b), 'A fuzzy logic and neural network based wind power model', in *Proceedings of the 1996 European Wind Energy Conference EWEC'96*, Gothenburg, Sweden, pp. 596-599

Kariniotakis, G *et al.* (2004), 'What performance can be expected by short-term wind power prediction models depending on site characteristics?', in *Proceedings of the European Wind Energy Conference EWEC'04*, London, UK

Kass, RE & Raftery AE (1995), 'Bayes Factors', *Journal of the American Statistical Association*, Vol. 90, pp. 773-795

Landberg, L (1998), 'A Mathematical Look at a Physical Power Prediction Model', *Wind Energy*, Vol. 1, Issue 1, pp. 23-28

Landberg, L (1999), 'Short-term prediction of the power production from wind farms', *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 80, Issue 1-2, pp. 207-220

Landberg, L, et al. (2002), 'Poor man's ensemble forecasting for error estimation', in *Proceedings of the AWEA 2002*, Portland, Oregon, USA

Lange, B *et al.* (2006), 'Wind power prediction in Germany – Recent advances and future challenges', in *Proceedings of the European Wind Energy Conference EWEC'06*, Athens, Greece

Lange, B *et al.* (2008), 'Improving Very Short-term Forecasts of Wind Power by Using an Online Wind Measurement Network', in *Proceedings of the European Wind Energy*

Conference EWEC'08, Brussels, Belgium

Lange, M (2003), *Analysis of the Uncertainty of Wind Power Predictions*, Ph.D. Dissertation, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Lange, M (2005), 'On the Uncertainty of Wind Power Predictions – Analysis of the Forecast Accuracy and Statistical Distribution of Errors', *Journal of Solar Energy Engineering*, Vol. 127, No. 2, pp. 177-184

Lange, M *et al.* (2006), 'Optimal Combination of Different Numerical Weather Models for Improved Wind Power Predictions', in *Proceedings of the 6th International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms*, Delft, Netherlands

Lange, M *et al.* (2008), 'Improving Very Short-Term Forecasts of Wind Power by Using an Online Wind Measurement Network', in *Proceedings of the European Wind Energy Conference EWEC'08*, Brussels, Belgium

Lang, SJ & McKeogh EJ (2009), 'Forecasting wind generation, uncertainty and reserve requirement on the Irish power system using an ensemble prediction system', *Wind Engineering*, Vol. 33, No. 5, pp. 433-448

Larson, KA & Westrick, K (2006), 'Short-Term Forecasting Using Off-Site Observations', *Wind Energy*, Vol. 9, Issue 1-2, pp. 55-62

Li, G & Shi J (2010), 'Application of Bayesian model averaging in modelling long-term wind speed distributions', *Renewable Energy*, Vol. 35, Issue 6, pp. 1192-1202

Li, S *et al.* (2001), 'Using neural networks to estimate wind power turbine generation', IEEE Transactions on Energy Conversion, Vol. 16, Issue 3, pp. 276-282

Madsen, H *et al.* (2005), 'A tool for predicting the wind power production of off-shore wind plants', *Wind engineering*, Vol. 29, Issue 6, pp. 475-489

Magnusson, M (2002), 'CFD as tool for wind forecasts', in *Proceedings of the First IEA Joint Action Symposium on Wind Forecasting Techniques*, Norrköping, Sweden, pp. 83-91

Magnusson, M & Wern, L (2001), 'Wind Energy Predictions using CFD and HIRLAM Forecast', in *Proceedings of the European Wind Energy Conference EWEC'01*, Copenhagen, Denmark, pp. 83-91

Mana, M *et al.* (2013), 'Short-term Forecasting of wind energy production using CFD simulations', poster at *EWEA Technology Workshop: Wind Power Forecasting*, Rotterdam, Netherlands

Marti, I *et al.* (2001), 'Prediction models in complex terrain', in *Proceedings of the European Wind Energy Conference EWEC'01*, Copenhagen, Denmark

Marti, I *et al.* (2003), 'LocalPred and RegioPred. Advanced tools for wind energy prediction in complex terrain', in *Proceedings of the European Wind Energy Conference EWEC'03*, Madrid, Spain

Martinez, WL & Martinez AR (2002), *Computational statistics handbook with matlab*, Chapman & Hall/CRC, Boca Raton, Florida, USA

Mason, SJ (2008), 'Understanding forecast verification statistics', *Meteorological Applications*, Vol. 15, pp. 31-40

MathWorks 2014, 'betalike', in *R2014b Documentation*, http://se.mathworks.com/help/stats/betalike.html [17.12.2014]

Messner, JW *et al.* (2013), 'Probabilistic wind power forecasts with an inverse power curve transformation and censored regression', *Wind Energy*, Vol. 17, Issue 11, pp. 1753-1766

Midtbø, KH *et al.* (2011), *Verification of wind and turbulence forecasts for the airports Mehamn, Honningsvåg, Hammerfest, Hasvik, Tromsø, Bardufoss, Evenes, Narvik, Svolvær, Leknes, Mo I Rana, Mosjøen, Sandnessjøen, Brønnøysund, Værnes, Ørsta-Volda, Sandane, Førde and Fagernes for the period 1.1.2011 to 31.08.2011*, Met.no note 11/2011, Oslo, Norway

Miller, I & Miller, M (2004), *John E. Freund's Mathematical Statistics with Applications (7th ed.)*, Pearson Prentice Hall, Upper Saddle River, New Jersey, USA

Monin AS & Yaglom AM (1971), *Statistical Fluid Dynamics: Mechanics of Turbulence, 1st Volume*, MIT Press, Cambridge, Massachussets, USA, p. 274

Monteiro, C *et al.* (2009), *Wind Power Forecasting: State-of-the-Art 2009*, Argonne National Laboratory ANL/DIS-10-1, Argonne, Illinois, USA

Moon, D *et al.* (2004), 'Support Vector Machines technology coupled with physics-based modelling for wind facility production forecasting', in *Proceedings of the 2004 Global Windpower Conference*, Chicago, Illinois, USA.

Mortensen, NG *et al.* (1993), *Wind atlas analysis and application program (WAsP), Vol. 2: User's guide*, Technical Report Risø-I-666(EN)(v.2), Roskilde, Denmark

Murphy, AH (1996), 'The Finley Affair: A Signal Event in the History of Forecast Verification', *Weather and Forecasting*, Vol. 11, No. 1, pp. 3-20

Möhrlen, C (2004), *Uncertainty in Wind Energy Forecasting*, PhD-thesis, University College Cork, National University of Ireland, Cork, Ireland

Möhrlen, C & Jørgensen JU (2006), 'Forecasting Wind Power in High Wind Penetration Markets using Multi-Scheme Ensemble Prediction Methods', in *Proceedings of the German Wind Energy Conference DEWEK*, Bremen, Germany

Nanahara, T *et al.* (2004), 'Smoothing Effects of Distributed Wind Turbines. Part 2. Coherence among Power Output of Distant Wind Turbines', *Wind Eergy*, Vol. 7, Issue 2, pp. 75-85

Natenberg, EJ *et al.* (2013), 'Application of a Random Forest Approach to Model Output Statistics for use in Day Ahead Wind Power Forecasts', Poster at the *93rd AMS Annual Meeting*, Austin, Texas, USA

Nielsen, HAa *et al.* (2004), 'Wind power ensemble forecasting', in *Proceedings of the 2004 Global Windpower Conference*, Chicago, Illinois, March 2004

Nielsen, HAa *et al.* (2007a), 'Improvement and automation of tools for short term wind power forecasting', in *Proceedings of the 2007 European Wind Energy Conference EWEC'07*, Milan, Italy

Nielsen, HAa *et al.* (2007b), 'Optimal combination of wind power forecasts', *Wind Energy*, Vol. 10, Issue 5, pp. 471-482

Nielsen, TS *et al.* (1998), 'A New Reference for Wind Power Forecasting', *Wind Energy*, Vol. 1, Issue 1, pp. 29-34

Nielsen, TS *et al.* (2001), 'Zephyr – The Prediction Models', in *Proceedings of the European Wind Energy Conference EWEC'01*, Copenhagen, Denmark

Nielsen, TS *et al.* (2002), 'Prediction of regional wind power', in *Proceedings of the 2002 Global Windpower Conference*, Paris, France

NVE (2009a), *Vindkart for Norge. Kartbok 2a: Terrengkompleksitet, RIX-verdier*, Available from: http://www.nve.no/Global/Energi/fornybar%20energi/Vindkraft/Kartbok2a_rev1.pdf?epslanguage=no [19.11.2014]

NVE (2009b), *Vindkart for Norge*, Rapport nr. 9/2009, NVE, Oslo, Norway. Available from: http://www.nve.no/Global/Publikasjoner/Publikasjoner%202009/Oppdragsrapport%20A%202009/oppdragsrapportA9-09.pdf?epslanguage=no [19.11.2014]

NVE (2014), *Vindkraft – Produksjon i 2013*, Rapport 20/2014, Norges vassdrags- og energidirektorat, Oslo, Norway, http://webby.nve.no/publikasjoner/rapport/2014/rapport2014_20.pdf [14.01.15]

Ouyang, T *et al.* (2013), 'A Survey of Wind Power Ramp Forecasting', *Energy and Power Engineering*, Vol. 5, No. 4B, pp. 368-372

Papke, U *et al.* (1993), 'Evaluation and Short-Time Forecasts of WEC-Power within the power grid of SCHLESWAG AG', in *Proceedings of the 1993 ECWEC*, Travemünde, Germany, pp. 770-773

Pappala, V *et al.* (2009), 'A Stochastic Model for the Optimal Operation of a Wind-Thermal Power System', *IEEE Transactions on Power Systems*, Vol. 24, pp. 940-950

Parkes, J *et al.* (2006), 'Wind Energy Trading Benefits Through Short Term Forecasting', in *Proceedings of the European Wind Energy Conference EWEC'06*, Athens, Greece

Parkes, J (2009), 'Temporal Forecast Uncertainty for Ramp Events', in *Proceedings of the European Wind Energy Conference EWEC'99*, Marseilles, France

Pierce, CS (1884), 'The Numerical Measure of the Success of Predictions', *Science*, Vol. 4, No. 93, pp. 453-454

Pinson, P & Kariniotakis, G (2003), 'On-line Assessment of Prediction Risk for Wind Power Production Forecasts', in *Proceedings of the European Wind Energy Conference EWEC'03*, Madrid, Spain

Pinson, P (2006), *Estimation of the Uncertainty in Wind Power Forecasting*, Ph.D. Dissertation, Ecole des Mines de Paris, Sophia Antipolis, France

Pinson, P *et al.* (2007), 'Trading Wind Generation from Short-Term Probabilistic Forecasts of Wind Power', *IEEE Transactions on Power Systems*, Vol. 22, No. 3, pp. 1148-1156

Pinson, P *et al.* (2008), 'Regime-switching modelling of the fluctuations of offshore wind generation', *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 96, Issue 12, pp. 2327-2347

Pinson, P *et al.* (2009a), 'From Probabilistic Forecasts to Statistical Scenarios of Short-term Wind Power Production', *Wind Energy*, Vol. 14, Issue 12, pp. 51-62

Pinson, P *et al.* (2009b), 'Skill forecasting from ensemble predictions of wind power', *Applied Energy*, Vol. 86, No 7-8, pp. 1326-1334

Pinson, P & Madsen, H (2009), 'Ensemble-based probabilistic forecasting at Horns Rev, Wind Energy', Vol. 12, Issue 2, pp. 137-155

Pinson, P & Kariniotakis, G (2010), 'Conditional prediction intervals of wind power generation', *IEEE Transactions on Power Systems*, Vol. 25, No. 4, pp. 1845-1856

Pinson, P (2012), 'Very short-term probabilistic forecasting of wind power with generalized logit-Normal distributions', *Journal of the Royal Statistical Society*, Series C 61(4), pp. 555-576

Raftery, AE *et al.* (1997), 'Bayesian Model Averaging for Linear Regression Models', *Journal of the American Statistical Association*, 92, pp. 179-191

Raftery, AE *et al.* (2005), 'Using Bayesian Model Averaging to Calibrate Forecast Ensembles', *Monthly Weather Review*, Vol. 133, Issue 5, pp. 1155-1174

Reikard, G (2008), 'Regime-Switching Models and Multiple Causal Factors in Forecasting

Wind Speed', *Wind Energy*, Vol. 13, Issue 5, pp. 407-418

Revheim, PP & Beyer HG (2012), 'Spatial-Temporal Analysis of Wind Power Forecast Errors for West-Coast Norway', in *Proceedings of the 11th International Workshop on Large Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Lisbon, Portugal, pp. 304-309

Revheim, PP & Beyer HG (2013a), 'Improving Wind Power Forecasts by a Post-Processing Scheme Using Spatial-Temporal Forecast Error Propagation - The Importance of the Appropriate Selection of Wind Direction Regime', in *Proceedings of the 2013 EWEA annual conference*, Vienna, Austria, http://proceedings.ewea.org/annual2013/allfiles2/1216_EWEA2013presentation.pdf [08.12.14]

Revheim, PP & Beyer HG (2013b), 'Using Bayesian Model Averaging for wind farm group forecasts', Poster at *EWEA Technology Workshop: Wind Power Forecasting*, Rotterdam, Netherlands

Rodrigues, A *et al.* (2007), 'EPREV – A Wind Power Forecasting Tool for Portugal', in *Proceedings of the European Wind Energy Conference EWEC'07*, Milano, Italy

Sack, J *et al.* (2012), 'From NWP Ensemble to Probabilistic Wind Energy Production Forecasts', in *Proceedings of the European Wind Energy Association Annual Event*, Copenhagen, Denmark

Saleck, N & Von Bremen, L (2007), 'Wind power forecast error smoothing within a wind farm', *Journal of Physics: Conference Series*, 75 (2007) 012051, pp. 1-8

Sánchez, I (2006), 'Short-term prediction of wind energy production', *International Journal on Forecasting*, Vol. 22, Issue 1, pp. 43-56

Sánchez, I (2008a), 'Adaptive combination of forecasts with application to wind energy', *International Journal of Forecasting*, Vol. 24, Issue 4, pp. 679-693

Sánchez, I (2008b), 'Adaptive density estimation of wind power predictions', in *Proceedings of the International Symposium on Forecasting ISF2008*, Nice, France

Siebert, N & Kariniotakis, G (2006), 'Reference wind farm selection for regional wind power prediction models', in *Proceedings of the European Wind Energy Conference EWEC'06*, Athens, Greece

Siebert, N (2008), *Development of Methods for Regional Wind Power Forecasting*, Ph.D. Dissertation, Ecole des Mines de Paris, Sophia Antipolis, France

Siemens (2009), *Outstanding efficiency. Siemens Wind Turbine SWT-2.3-93*, http://www.energy.siemens.com/mx/pool/hq/power-generation/wind-power/E50001-W310-A102-V6-4A00_WS_SWT-2.3-93_US.pdf [13.11.14]

Sloughter, JM *et al.* (2007), 'Probabilistic quantitative precipitation forecasting using Bayesian model averaging', *Monthly Weather Review*, Vol. 135, Issue 9, pp. 3209-3220

Sloughter, JM *et al.* (2010), 'Probabilistic wind speed forecasting using ensembles and Bayesian model averaging', Journal of the American Statistical Association, Vol. 105, Issue 489, pp. 25-35

Statnett (2013), *Nettutviklingsplan 2013. Nasjonal plan for neste generasjons kraftnett*, Statnett, Oslo, Norway, http://www.statnett.no/Global/Dokumenter/Prosjekter/Nettutviklingsplan%202013/Nettutviklingsplan%202013.pdf [02.12.2014]

Stephenson, DB (2000), 'Use of the "Odds Ratio" for Diagnosing Forecast Skill', *Weather and Forecasting*, Vol. 15, Issue 2, pp. 221-232

Strbac, G (2008), 'Demand side management: Benefits and challenges', *Energy Policy*, Vol. 36, Issue 12, pp. 4419–4426

Suzuki, A *et al.* (2012), 'Use of Offsite Data to Improve Short Term Ramp Forecasting', in *Proceedings of the 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Lisbon, Portugal, pp. 621-626

Tastu, J *et al.* (2010), 'Multivariate conditional parametric models for a spatio-temporal analysis of short-term wind power forecast errors', in *Proceedings of the European Wind Energy Conference EWEC'10*, Warsaw, Poland

Tastu, J *et al.* (2011), 'Spatio-Temporal Analysis and Modelling of Short-Term Wind Power Forecast Errors', *Wind Energy*, Vol. 14, Issue 1, pp. 43-60

Tastu, J *et al.* (2012), 'Spatio-temporal correction of probabilistic wind power forecasts', in *Proceedings of the 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Lisbon, Portugal

Tastu, J (2013), *Short-term wind power forecasting: probabilistic and space-time aspects*, PhD-thesis. DTU, Lyngby, Denmark

Tastu, J *et al.* (2014), 'Space-time trajectories of wind power generation: Parameterized precision matrices under a Gaussian copula approach'. *Lecture Notes in Statistics: Modelling and Stochastic Learning for Forecasting in High Dimension*. http://pierrepinson.com/docs/Tastuetal2013_windtrajectories_revised.pdf [02.11.14]

Tong, H (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford Statistical Science Series, Vol. 6, Oxford University Press, Oxford, UK

Udén, P et al. (2002), *HIRLAM-5 Scientific Documentation*, HIRLAM-5 Project, SMHI,

Norrköping, Sweden, http://hirlam.org/index.php/component/docman/cat_view/114-model-and-system-documentation/131-hirlam-documentation?Itemid=70 [05.11.14]

Usaola, J *et al.* (2002), 'Sipreolico: a wind power prediction tool for the Spanish peninsular power system operation', in *Proceedings of the European Wind Energy Conference EWEC'02*, Paris, France

Van Hulle *et al.* (2009), *Integrating Wind – Developing Europe's power market for the large-scale integration of wind power*, TradeWind technical report, February 2009

Vidal, J *et al.* (2010), 'Validation of a Wind Power Forecast System – The multi-model NWP ensemble strategy', in *Proceedings of the European Wind Energy Conference EWEC'10*, Warsaw, Poland

Wagenmakers, EJ & Farrell, S (2004), 'AIC model selection using Akaike weights', *Psychonomic bulletin & review*, Vol. 11, Issue 1, pp. 192-196

Watson, SJ *et al.* (1994), 'Application of wind speed forecasting to the integration of wind energy into a large scale power system', *IEEE Proceedings on Generation, Transmission and Distribution*, Vol. 141, No. 4, pp. 357-362

Wessel, A *et al.* (2009), 'Improving Short-Term Forecasts with Online Measurements', in *Proceedings of the European Wind Energy Conference EWEC'09*, Marseilles, France

WMO (2010), *Guide to Meteorological Instruments and Methods of Observation*. WMO-No. 8 (2008 edition, Updated in 2010). WMO, Geneva, Switzerland

Wu, G & Dou, Z (1995), 'Non-linear wind prediction using a fuzzy modular temporal neural network', in *Proceedings of Windpower'95 American Wind Energy Association Conference*

Wu, J *et al.* (2014), 'Statistical distribution for wind power forecast error and its application to determine optimal size of energy storage system', *International Journal of Electrical Power & Energy Systems*, Vol. 55, pp 100-107

Zack, JW *et al.* (2010), 'Development and Testing of an Innovative Short-Term Large Wind Ramp Forecasting System', in *Proceedings of the European Wind Energy Conference EWEC'10*, Warsaw, Poland

Zhang, J *et al.* (2014), 'Ramp Forecasting Performance from Improved Short-Term Wind Power Forecasting', presented at the *ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2014)*, Buffalo, New York, USA, http://www.nrel.gov/docs/fy14osti/61730.pdf [13.01.15]

Zhang, ZS *et al.* (2013), 'A Versatile Probability Distribution Model for Wind Power Forecast Errors and Its Application in Economic Dispatch', *IEEE Transactions on Power Systems*, Vol. 28, Issue 2, pp. 3114-3125

Zheng, H & Kusiak, A (2009), 'Prediction of Wind Farm Power Ramp Rates: A Data-Mining Approach', *ASME Journal of Solar Engineering*, Vol. 131, No. 3, pp. 031011-1 – 031011-8

Zugno, M *et al.* (2012), 'Trading wind energy on the basis of probabilistic forecasts both of wind generation and of market quantities', *Wind Energy*, Vol. 16, Issue 6, pp. 909-926

Annex 1 – Single site key statistics

Site	Assumed roughness	Approximate RIX	Number of observations (hours)	Mean wind speed 10 m	Mean wind speed 80 m	St. dev. 10 m	St. dev. 80 m
Berlevåg	0,1	5-10 %	33261	7,19	10,43	4,20	6,09
Brønnøysund	0,5	0-5 %	34414	4,29	7,26	2,85	4,82
Burholmråsa	0,25	0-5 %	29540	4,55	7,12	4,36	6,82
Båtsfjord	0,1	5-10 %	34198	6,90	10,02	4,55	6,61
Fedje	0,25	0-5 %	34671	6,84	10,70	3,82	5,97
Flesland	0,25	5-10 %	34683	3,65	5,71	2,52	3,94
Florø	0,25	5-10 %	34865	4,70	7,35	2,93	4,57
Fruholmen	0,03	0-5 %	32560	7,32	9,94	4,96	6,74
Halten	0,25	0-5 %	34515	7,65	11,96	4,16	6,50
Hammerfest	0,25	5-10 %	34873	4,25	6,64	2,78	4,34
Hasvik	0,5	5-10 %	34323	5,59	9,47	3,37	5,72
Haugesund	0,5	0-5 %	35010	5,44	9,21	3,00	5,09
Helligvær	0,1	0-5 %	34929	7,58	11,00	3,83	5,56
Honningsvåg	0,5	5-20 %	33108	5,71	9,67	3,73	6,31
Kirkenes	0,25	0-5 %	34531	4,34	6,78	2,64	4,13
Kristiansund	0,25	0-10 %	34680	3,58	5,60	2,19	3,42
Kvitsøy	0,25	0-5 %	34924	6,07	9,50	3,31	5,18
Leknes	0,25	0-5 %	34349	4,14	6,48	2,83	4,43
Lindesnes	0,5	5-10 %	34764	7,26	12,29	4,24	7,18
Makkaur	0,25	0-5 %	34133	6,02	9,42	4,26	6,67
Mehamn	0,25	5-10 %	34722	5,87	9,19	3,47	5,43
Myken	0,03	0-5 %	34850	7,99	10,85	4,17	5,66
Nordøyan	0,03	0-5 %	34932	8,74	11,87	4,40	5,98
Obrestad	0,1	0-5 %	34600	6,17	8,95	3,63	5,27
Ona	0,25	0-5 %	34013	6,86	10,72	3,97	6,22
Rørvik	0,5	5-10 %	34696	3,99	6,76	2,55	4,31
Røst	0,1	0-5 %	32487	6,33	9,19	3,47	5,04
Sandnessjøen	0,25	5-10 %	34951	3,93	6,15	2,46	3,84
Sklinna	0,25	0-5 %	33378	8,43	13,18	4,61	7,21
Skorva	0,25	5-10 %	31407	5,08	7,94	3,82	5,98
Slettnes	0,1	5-10 %	33731	7,01	10,17	3,87	5,61
Slaaterøy	0,25	0-5 %	31503	5,28	8,26	3,79	5,93

Sola	0,25	0-5 %	34882	4,11	6,43	2,88	4,50
Stokmarknes	0,1	5-10 %	34759	4,16	6,04	2,66	3,86
Sula	0,25	0-5 %	33852	5,45	8,53	3,48	5,44
Svolvær	0,25	5-20 %	34757	5,18	8,10	3,09	4,84
Tromsø	0,25	5-10 %	34301	4,31	6,73	3,15	4,93
Utsira	0,25	0-5 %	33030	7,19	11,24	4,52	7,07
Vadsø	0,25	0-5 %	34088	5,43	8,49	3,19	4,99
Vardø lufthavn	0,25	0-5 %	34897	7,51	11,75	4,06	6,35
Veiholmen	0,25	0-5 %	35020	6,30	9,85	3,59	5,61
Vigra	0,1	0-5 %	34353	4,78	6,94	2,52	3,66
Ytterøyane	0,1	0-5 %	34342	6,85	9,95	4,15	6,02

Annex 2 – List of papers, posters and oral presentations

Kratzenberg, MG *et al.* (2014), 'Improvement of NWP based short term Wind Power Forecasts by Post-processing using Artificial Neural Networks and Regression', in *Proceedings of the 13th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Berlin, Germany

Kratzenberg, MG *et al.* (2014), 'Random forest prediction in solar radiation forecasting schemes applied to handling of ramps and large forecast errors', in *Proceedings of the 4th International Workshop on Integration of Solar Power into Power Systems*, Berlin, Germany

Revheim, PP, Beyer, HG (2014), 'Improving wind power ramp predictability through pre-processing of NWP wind power forecasts', submitted to *Meteorological Applications*

Revheim, PP, Beyer, HG (2014), 'Offshore ramp forecasting using offsite data', poster at *EERA DeepWind 2014*, Trondheim, Norway

Revheim, PP, Beyer, HG (2014), 'Probabilistic Wind Farm Group Forecasts using Bayesian Model Averaging', presentation at *European Meteorological Society Annual Meetings 2014*, Prague, Czech Republic

Revheim, PP, Beyer, HG (2014), 'Random Forest Ramp Forecasts for Groups of Wind Farms', presentation at *European Meteorological Society Annual Meetings 2014*, Prague, Czech Republic

Revheim, PP, Beyer, HG (2014), 'Using Bayesian Model Averaging for Wind Power Forecasts', poster at *European Geosciences Union General Assembly 2014*, Vienna, Austria

Revheim, PP, Beyer, HG (2014), 'Using Random Forests for Wind Power Ramp Forecasting', in *Proceedings of the 2014 EWEA annual conference*, Barcelona, Spain

Beyer, HG, Revheim, PP (2013), 'An Option to Assign Confidence Intervals to Day Ahead Wind Power Forecasts - Example Based on Meteorological Conditions in Norway', in *Proceedings of the 12th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, London, UK

Revheim, PP, Beyer, HG (2013), 'Improving wind power forecasts by a post-processing scheme using spatio-temporal forecast error propagation - the importance of the appropriate selection of wind direction regimes', in *Proceedings of the 2013 EWEA annual conference*, Vienna, Austria

Revheim, PP, Beyer, HG (2013), 'Inter-annual variation of wind speed in southern Norway', presentation at the *11th European Conference on Applications of Meteorology (ECAM)*, Reading, UK

Revheim, PP, Beyer, HG (2013), 'Methods for Estimating Long-Term Average Wind Speed', in *Proceedings of the 12th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, London, UK

Beyer, HG, Revheim, PP (2013), 'Modelling probability distributions and resulting confidence levels of forecast errors in the day ahead power prediction for ensembles of wind farms in Norway', poster at the *2nd International Conference Energy & Meteorology (ICEM 2013)*, Toulouse, France

Revheim, PP, Beyer, HG (2013), 'Using Bayesian Model Averaging for wind farm group forecasts', poster at *EWEA Technology Workshop: Wind Power Forecasting*, Rotterdam, Netherlands

Beyer, HG, Revheim, PP (2012), 'Spatial-temporal analysis of wind power forecast errors for Norway', presentation at the *First International Workshop "Forecasting and Predictive Modelling for Renewable Energy"*, Bergen, Norway

Revheim, PP, Beyer, HG (2012), 'Assessment of Wind Power Related Wind Forecasts and Spatial Error-Smoothing Effects for a Case in South-Western Norway', presentation at *DEWEK 2012, 11th German Wind Energy Conference*, Bremen, Germany

Revheim, PP, Beyer, HG (2012), 'Spatial-Temporal Analysis of Wind Power Forecast Errors for South-Western Norway', poster at the *5th International Conference on Integration of Renewable and Distributed Energy Resources*, Berlin, Germany

Revheim, PP, Beyer, HG (2012), 'Spatial-Temporal Analysis of Wind Power Forecast Errors for West-Coast Norway', in *Proceedings of the 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, Lisbon, Portugal

Annex 3 – Errata

Page	Line	Old text	New text
Abstract	16	When applied to “unseen” forecasts the simpler models …	When applied to “unseen” forecasts (i.e. independent test data) the simpler models ...
2	2	Recent years there has been …	In recent years there has been …
2	7	… Europe raised from ~66 GW …	… Europe rose from ~66 GW …
3	15-16	(Giebel 2011; Monteiro *et al.* 2009)	(Giebel *et al.* 2011; Monteiro *et al.* 2009)
3	25	(see e.g. Gipe (2004, p. 30))	(see e.g. Gipe 2004, p. 30)
3	27	(see e.g. Gipe (2004, p. 56))	(see e.g. Gipe 2004, p. 56)
3	30	… turbines of other producers is expected …	… turbines of other producers are expected …
6	22 - 23	… for information on operational NWP systems to Giebel *et al.* …	… for information on operational NWP models – as applied to wind power – to Giebel *et al.* …
8	18-19	More examples of … in Giebel *et al.* (2011) and Monteiro *et al.* (2009) and references given therein.	[deleted text]
9	13	… which integrates information form NWPs …	… which integrates information from NWPs …
10	3-4	More examples of … in Giebel et al. (2011) and Monteiro et al. (2009) and references given therein.	[deleted text]
10	5	Point forecasts have been, and still is, …	Point forecasts have been, and still are, …
10	6-7	very easy to interpret, but in continuous situations, as is the case with wind power, they will …	very easy to interpret, but they will …
10	18	The division between physical models and statistical models are also …	The division between physical models and statistical models is also …
11	8	Later years, …	In later years, …
11	15-16	More examples of … in Giebel et al. (2011) and Monteiro et al. (2009) and references given therein.	[deleted text]
13	17-18	… make probabilistic ramp rate forecast and hybrid deterministic-probabilistic ramp forecast.	… make probabilistic ramp rate forecasts and hybrid deterministic-probabilistic ramp forecasts.
16	5	… forecasted values that falls outside the corresponding …	… forecasted values fall outside …
16	17	… is found in Monteiro *et al.* 2009.	… is found in Monteiro *et al* (2009).
17	5	For applications lice automatic …	For applications like automatic …
17	Caption Table 1.1	Correct predictions is on the diagonal …	Correct predictions are on the diagonal …
17	13	… simple evaluation metrics that only aims at …	… simple evaluation metrics that only aim at …
17	17	… *frequency* bias.	… *frequency bias*.
17	26-27	Bias-values lower than 1 indicates a	Bias-values lower than 1 indicate a

		tendency … while bias-values higher than 1 indicates a tendency …	tendency … while bias-values higher than 1 indicate a tendency …
19	9	… as well as many coastal-near …	… as well as many near-coastal …
19	28-29	(see e.g. Beyer et al. (1999) and Focken et al. (2001)) and WPPT (see e.g. Nielsen et al. (2007a) and Madsen et al. (2005))	(see e.g. Beyer et al. 1999 and Focken et al. 2001) and WPPT (see e.g. Nielsen et al. 2007a and Madsen et al. 2005)
19	32-33	… largest volumes of installed wind power capacity in Europe is installed…	… largest volumes of wind power capacity in Europe are installed …
20	32	… into groups, investigation potential predictors …	… into groups, investigating potential predictors …
21	26	… and which factors that contribute to the reduction.	… and which factors contribute to the reduction.
24	21	a minimum accuracy of 0.5 m/s for wind speeds …	a minimum accuracy of ± 0.5 m/s for wind speeds …
26	4	… as operated by the Norwegian Meteorological Institute (MET)…	… as operated by MET …
30	13	(see e.g. Magnusson & Wern (2001), Mana *et al*. (2013) or Castellani *et al.* (2014))	(see e.g. Magnusson & Wern 2001, Mana *et al*. 2013 or Castellani *et al.* 2014)
30	19	(see e.g. Mortensen et al. (1993))	(see e.g. Mortensen *et al.* 1993)
35	13	...accounts for.	… account for.
51	9	… with increasing distance translates to reduction …	… with increasing distance translate to reduction …
52	22	Figure 4.3 show the spatial …	Figure 4.3 shows the spatial ...
54	13	Focken *et al.* (2002b) shows that …	Focken *et al.* (2002b) show that …
54	14	Tastu *et al.* (2011) shows that …	Tastu *et al.* (2011) show that …
54	16	Tastu *et al.* (2011) also shows that …	Tastu *et al.* (2011) also show that …
60	4	… the lumped group wind power forecasts is divided …	… the lumped group wind power forecasts are divided …
66	10	(see e.g. Martinez & Martinez (2002))	(see e.g. Martinez & Martinez 2002)
68	12	(see e.g. Johnson & Wichern (2002))	(see e.g. Johnson & Wichern 2002)
73	12-13	(see e.g. Nielsen *et al.* (2007a) and Madsen *et al.* (2005))	(see e.g. Nielsen *et al.* 2007a and Madsen *et al.* 2005)
73	21	… output and to give the a forecast …	… output and to give the forecast …
73	25	(see e.g. Sloughter *et al.* (2007), Bao *et al.* (2010) and Chmielecki & Raftery (2010))	(see e.g. Sloughter *et al.* 2007, Bao *et al.* 2010 and Chmielecki & Raftery 2010)
73	26-27	(see Sloughter *et al.* (2010), Li & Shi (2010) and Courtney *et al.* (2013))	(see Sloughter *et al.* 2010, Li & Shi 2010 and Courtney *et al.* 2013),
74	1	… to be applied for wind power forecasting is discussed.	… to be applied for wind power forecasting are discussed.
74	4	(see map in Figure 4.1)	(see map in Figure 4.5)
74	12	… a group of different forecast …	… a group of different forecasts …
74	19-20	(see e.g. M. Lange *et al.* (2006), Nielsen *et al*. (2007b) and Sánchez (2008a))	(see e.g. M. Lange *et al.* 2006, Nielsen *et al*. 2007b and Sánchez 2008a)
74	21-22	(see e.g. Sack *et al.* (2012), Pinson	(see e.g. Sack *et al.* 2012, Pinson &

		& Madsen (2009) and Giebel *et al.* (2005))	Madsen 2009 and Giebel *et al.* 2005)
74	27	(see e.g. Vidal et al. (2010), Giebel et al. (1999) and Cali et al. (2008))	(see e.g. Vidal et al. 2010, Giebel et al. 1999 and Cali et al. 2008)
74	31-32	(see e.g. Möhrlen (2004), Möhrlen & Jørgensen (2006) and Pinson & Madsen (2009))	(see e.g. Möhrlen 2004, Möhrlen & Jørgensen 2006 and Pinson & Madsen 2009)
75	24-25	(see e.g. Kass & Raftery (1995), Raftery *et al.* (1997) and Hoeting *et al.* (1999))	(see e.g. Kass & Raftery 1995, Raftery *et al.* 1997 and Hoeting *et al.* 1999)
78	3	(Hodge *et al.* (2012); Pinson (2012); Blaudszuweit et al. (2008))	(Hodge *et al.* 2012; Pinson 2012; Blaudszuweit et al. 2008)
78	12	(see e.g. Martinez & Martinez (2002))	(see e.g. Martinez & Martinez 2002)
80	8	For forecasting, normalized wind power …	For forecasting normalized wind power …
81	14	(see e.g. Miller & Miller (2004))	(see e.g. Miller & Miller 2004)
84	2	Even though there also here are a lot of …	Even though there also here is a lot of …
86	19	(see e.g. Johnson & Wichern (2002))	(see e.g. Johnson & Wichern 2002)
87	6	… only possible when the observations *y* lies within …	… only possible when the observation *y* lies within …
89	22	Figures 5.13 and 5.14 shows examples …	Figures 5.13 and 5.15 show examples …
92	17	… indicating that these prediction interval …	… indicating that the prediction interval …
92	20-21	It is also noticed that the shares of observations that falls outside the limits are very unevenly distributed…	It is also noticed that the share of observations that falls outside the limits is very unevenly distributed …
93	21	… the mean of the beta distribution…	… the mean of the Beta distribution …
94	25-26	(see e.g. Pinson (2012), Tastu *et al.* (2012) and Messner *et al.* (2013))	(see e.g. Pinson 2012, Tastu *et al.* 2012 and Messner *et al.* 2013).
95	6	… often leads to large forecast errors.	… often lead to large forecast errors.
97	19-20	(e.g. Florita *et al.* (2013) and Zhang *et al.* (2014))	(e.g. Florita *et al.* 2013 and Zhang *et al.* 2014)
101	13	… Hanssen and Kuipers skill score (HK) is considered.	… Hanssen and Kuipers skill score (HK) are considered.
104	14	… uncorrected NWP ramp forecast receives the lowest …	… uncorrected NWP ramp forecast receiving the lowest …
105	22	(see e.g. Gislason *et al.* (2006), Jiang *et al.* (2007) and Dubath *et al.* (2011))	(see e.g. Gislason *et al.* 2006, Jiang *et al.* 2007 and Dubath *et al.* 2011)
106	10	(see e.g. Breiman *et al.* (1984))	(see e.g. Breiman *et al.* 1984)
108	21	… predictor variables are of various importance …	… predictor variables are of variable importance ...
108	24	… values of the variable is randomly permuted …	… values of the variable are randomly permuted …
109	22-23	(see e.g. Bossavy *et al.* (2010) and Collier *et al.* 2013)	(see e.g. Bossavy *et al.* 2010 and Collier *et al.* 2013)
109	24	(see e.g. Section 4.3 and Tastu *et al.* (2011))	(see e.g. Section 4.3 and Tastu *et al.* 2011)

110	Caption Figure 6.10	The ramp forecast is issued at time *t*. All	The ramp forecast is issued at time *t*.
113	11	The frequency bias measure of how the numbers …	The frequency bias measures how the numbers …
114	Caption Figure 6.13	… and 0.3 (blue).). The frequency …	… and 0.3 (blue). The frequency …
115	1	… worse than if assigning the observations to the three classes on random.	… worse than if randomly assigning the observations to the three classes.
119	19	The overall aim of the thesis …	The overall aims of the thesis …
121	33	As especially hours of 0 % production is common …	As especially hours of 0 % production are common …
124	31	(see e.g. Chmielecki & Raftery (2010))	(see e.g. Chmielecki & Raftery 2010)
124	32	(see e.g. Pinson (2012), Tastu *et al.* (2012) and Messner *et al.* (2013))	(see e.g. Pinson 2012, Tastu *et al.* 2012 and Messner *et al.* 2013)