Modelling of Uncertainties in Wind Power Forecasts

Lauri Sääskilahti

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 18.4.2023

Supervisor

Prof. Ahti Salo

Advisor

M.Sc. (Eng.) Jyrki Keskinen

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the master's thesis

Author Lauri Sääskilahti

Title Modelling of Uncertainties in Wind Power Forecasts

Degree programme Master's Programme in Mathematics and Operations Research

Major Systems and Operations Research **Code of major** SCI3055

Supervisor Prof. Ahti Salo

Advisor M.Sc. (Eng.) Jyrki Keskinen

Date 18.4.2023 **Number of pages** 54 **Language** English

Abstract

The share of intermittent wind power generation is increasing in the power system. As weather-dependent, stochastic, non-dispatchable resource accurate forecasts of wind power production are necessary for efficient operation of both power system and electricity markets. There are several providers of such forecasts, but few give estimates on the reliability of their forecasts. Wind power forecast errors cause financial losses to wind producers and market participants, and there is a keen interest towards such reliability estimates of wind forecasts. In this thesis three different models are introduced to study the power production of single wind farm. The main modelling approach is to leverage numerical ensemble weather predictions to build probabilistic machine learning models. Forecast intervals are created with three different approaches with the goal of creating actionable insights about the power production uncertainty to support decision-making.

Keywords Wind power forecasting, energy modelling, uncertainty quantification, energy systems, renewable energy, numerical weather predictions, machine learning

Aalto University
School of Science

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Diplomityön tiivistelmä

Tekijä Lauri Sääskilahti

Työn nimi Tuulivoiman tuotantoennusteiden epävarmuuden mallintaminen

Koulutusohjelma Master's Programme in Mathematics and Operations Research

Pääaine Systems and Operations Research **Pääaineen koodi** SCI3055

Työn valvoja Prof. Ahti Salo

Työn ohjaaja DI Jyrki Keskinen

Päivämäärä 18.4.2023 **Sivumäärä** 54 **Kieli** Englanti

Tiivistelmä

Vaihtelevan tuulivoimatuotannon osuus energiantuotannossa kasvaa alati. Sääriippuvaisena, stokastisena ja heikosti hallittavana tuotantona tarkat tuulivoimatuotantoennusteet ovat välttämättömiä sekä energiajärjestelmän että -markkinoiden sujuvalle toiminnalle. Ennusteiden tarjoajia on useita, mutta harva tarjoaa arvioita ennusteidensa luotettavuudesta. Ennustevirheet aiheuttavat taloudellisia menetyksiä sekä tuottajille että muille markkinaosapuolille, minkä takia alalla on paljon kiinnostusta luotettavuusestimaatteja kohtaan. Tässä diplomityössä esitellään kolme mallia yksittäisen tuulipuiston tuotannon mallintamiseen. Mallintamisessa hyödynnetään sään parviennusteita pääasiallisena tapana mallintaa tuotannon vaihtelevuutta. Ennusteintervallit luodaan kolmella eri metodilla tarkoituksena tuottaa tuulivoimatoimijoille tietoa tuulivoimatuotannon epävarmuudesta päätöksenteon tueksi.

Avainsanat Tuulivoiman ennustaminen, energiamallintaminen, epävarmuuden kvantifiointi, energiajärjestelmät, uusiutuva energia, sään parviennusteet, koneoppiminen

Preface

I want to thank my instructor Jyrki Keskinen for his good guidance and Jukka Pajunen for his support and valuable counsel during the thesis. I also want to thank Jaakko Kleemola from Suomen Hyötytuuli Oy for kindly providing the wind farm data used in this thesis.

Helsinki, 18.4.2023

Lauri Sääskilahti

Contents

Abbreviations

Abbreviations

ANN	Artificial Neural Network
ARIMA	Auto-Regressive Integrated Moving-Average Model
CNN	Convolutional Neural Network
ECMWF	European Centre for Medium-Range Weather Forecasts
FMI	Finnish Meteorological Insitute
GAN	General Adversarial Neural Network
k-NN	k-Nearest Neighbour
LIDAR	Light Detection and Ranging Measurements
LSTM	Long-Term Short-Term Memory Neural Network
LUBE	Lower Upper Bound Estimation
MEPS	The MetCoOp Ensemble Prediction System
MET Norway	Norwegian Meteorological Institute
MLP	Multi-Layer Perceptron
MW	Megawatt, $10^6 \frac{J}{s}$
MWh	Megawatt hour, $3600 \cdot 10^6 J$
NN	Nearest Neighbour
NWP	Numerical Weather Prediction
PCA	Principal Component Analysis
RES	Renewable Energy Source
SCADA	Supervisory Control And Data Acquisition System
PC	Power Curve
VAR	Vector Auto-Regressive Model
WT	Wind Turbine
WTPC	Wind Turbine Power Curve

1 Introduction

1.1 Background

The share of wind power in electricity production is rapidly rising. In 2020 approximately 12 % of Finnish domestic electricity production, equivalent to 10 % of domestic consumption, was produced by wind. By the end of June 2022, the cumulative installed wind power capacity had reached 4000 MW, with 784 MW installed in the first half of 2022 alone, indicating rapid rise in wind power production. [1] Rising wind power share creates risks for energy supply security as wind generation is non-dispatchable and varies greatly depending on the weather conditions. Varying wind production levels also change the electricity markets, as it increases the importance of intraday markets [2] and is one of the primary motifs for the transition to 15-minutes trading intervals, which is planned to occur in the Nordics in 2023 [3].

In variable production market environment, accurate wind power forecasts are needed for the effective operation of transmission system and power markets. There are several providers of power production forecasts and many market participants have developed their own forecast tools. However, many models yield single point estimates, and lack metrics to measure their own reliability. This is problematic, as in underproduction scenarios the producer has to buy the deficit from the electricity markets to fulfill the contracts. Thus, the lack of reliability information complicates the decision-maker's work in managing market risks as they have insufficient information on how to best optimize wind farm and market operations to minimize costs in underproduction scenarios. On the other hand, overproduction scenarios are not desired, as they may lead suboptimal pricing and market transactions, although the risks involved are significantly smaller than in underproduction scenarios.

Currently many wind power producers and other power market participants aggregate forecasts from multiple sources to get more robust estimates. Even though the error in estimate decreases, this does not solve the issue of missing reliability information. The energy crisis has increased the price of electricity and thus also the price of forecast errors have also increased. In times of high prices, information about own production risks are especially valuable and can yield significant financial savings for market participants. Moreover, the transmission system operator may gain from the ability to identify underproduction scenarios and their associated risks.

This thesis studies only forecast uncertainty. The reasoning is that even though this is highly related to market risks, the market risks themselves are far more complicated and depend on many things, real and perceived, and often require knowledge of enterprises' confidential information. Thus, this thesis focuses only on production variability and aims to provide insights which can be utilized as input for market participant's decision making process.

For this thesis, representatives of four different organizations were interviewed: Helen and another large energy producer which chose to remain anonymous, the Finnish transmission system operator Fingrid, and the wind-energy producer Suomen Hyötytuuli. Interviewed organizations have different interests in energy markets, but all interviewees shared the interest in wind power uncertainty modelling. In the interviews, the lack of information about the uncertainty of the power forecasts was acknowledged by all interviewees.

In the Nordic electricity markets, trading takes place between 1 and 36 hours before delivery. In the interviews, this forecasting horizon was of most interest, and is for this reason chosen for this thesis, with emphasis on forecasting one to three hours into the future.

Suomen Hyötytuuli estimates that large majority of the incurred balancing costs is generated by few hours with high prices. The same is indicated by a large study [4]. When information on the wind power forecast accuracy can be coupled with knowledge on potentially high prices, wind producers could more efficiently control their market risks.

1.2 Research Objectives

Based on the interviews, there is clearly a demand for a tool, that could provide uncertainty information about the wind power forecasts. This kind of information could support the decision-making process, providing value to energy system participants, which has been noted in the literature as well [5].

Such a tool is difficult to build, as this would require data about the errors of such forecasts, which are often proprietary information. With the absence of usable forecasts, wind power is first modelled with three different models in order to obtain forecasts. These forecasts and their errors are then studied. Wind power uncertainty is finally modelled by creating forecast intervals which capture the variability in production.

Three models were created to model the production variability. The key factor in uncertainty modelling is the weather uncertainty, which is captured by ensembles of numerical weather predictions. The effect of weather and weather uncertainty is quantified and one of the goals is to provide a basic evaluation on the limits of uncertainty modelling given the accuracy of weather predictions.

The uncertainty of wind power is assessed probabilistically through an analysis of forecast intervals. The scope of this thesis is limited to the production of a single operational wind farm. Also the power production is studied only on operational wind turbines, as research on the availability of wind turbines is not feasible because it would significantly expand the scope of this thesis.

For modelling the production uncertainty, the production mean has to be first modelled. The mean of production forecasts is not itself the primary objective secondary importance in this thesis, but it is mandatory step in model building and considerations about it are discussed.

In this thesis earlier research is covered in Chapter 2. Chapter 3 introduces the data used in this thesis and metrics for evaluating the modelling results. In Chapter 4 the models are explained in detail, whereas in Chapter 5 the results are analysed. Discussion about the chosen methods and results are presented in Chapter 6. Lastly, Chapter 7 consists of the summary of this thesis.

2 Modelling and Forecasting Wind Power Generation

2.1 Wind Power Technology

Wind power technology harvests the kinetic energy of flowing air masses by converting it into electricity. There are several types of wind power plants, but the dominant one is the three-rotor design [1]. This thesis focuses solely on three-rotor power plants.

A wind turbine consists of four major parts: rotor blades, the nacelle, which contains the power generator, supporting mast and the foundations. Modern new wind turbines are typically 140 - 175 meters high with a rotor diameter of 130 - 160 meters and rated power of 4 - 5 MW. Offshore wind turbines are typically larger than land-based, and can have rated power of as much as 10 MW. [1] The trend in wind turbines has been to build larger turbines, as these can harvest the wind energy from higher altitudes, where winds are steadier and stronger, generating larger and more stable power output.

The kinetic energy from wind is given by the equation

$$P_{kin} = \frac{1}{2}\dot{m}u^2, \tag{1}$$

where u is the wind speed and $\dot{m}$ is the mass flow through the turbine and is given by

$$\dot{m} = \rho A u. \tag{2}$$

Here ρ is the density of air and A is the sweep area of the rotor blades. Kinetic energy absorbed from wind into the kinetic energy of the turbine is given by multiplying the kinetic energy of wind by power coefficient c_p. The power coefficient is not a constant, but instead a function of wind speed as the turbine control system automatically adjusts the angle of rotor blades to produce the desired power output. The theoretical maximum for the power coefficient is $\frac{16}{27} \approx 59,3\%$ which is known as Betz's law [6].

The relationship between wind speed and generated wind power is modelled as wind power curve (WPC) (see Figure 1). The curve is divided into four zones: in zone I the wind speed is below the cut-in speed and too low to generate any power, in zone II the power output rises cubically along with the increasing wind speed, in zone III the turbine has reached its rated wind speed and produces power steadily at maximum output by adjusting the rotor pitch angles until wind speed reaches cut-out speed at zone IV when the turbine is shut down for safety reasons.

The wind turbine is controlled automatically by the supervisory control and data acquisition (SCADA) system. SCADA is responsible for example for adjusting the blade angles, shutting down the turbine and collecting production data.

There are also other factors affecting the wind turbine production by causing losses when compared to the ideal case. These factors are more thoroughly studied in Section 4.2.

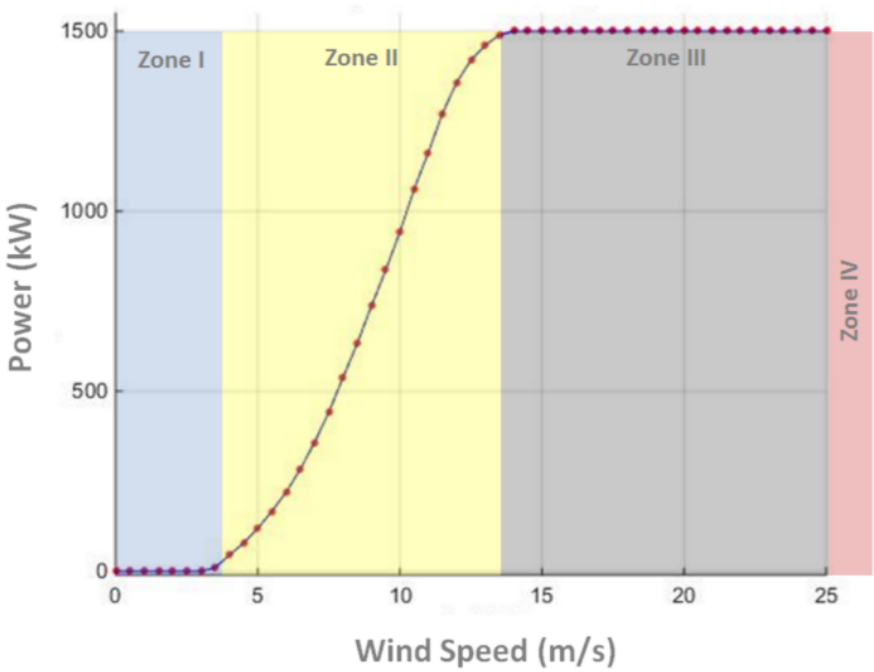

Figure 1: Wind turbine power curve. The four operating zones of wind turbines are highlighted. Figure taken from [7].

2.2 Weather Forecasts

Wind power is entirely weather dependent and thus weather forecasting plays a major role in wind power forecasting. Indeed, the main source of error for short- and medium-term wind power forecasting models are the inaccuracies in weather forecasting [8]. Thus, accurate and versatile weather forecasts are essential for successful wind power forecasting.

The atmosphere is fluid, and its state is typically modelled with partial differential equations. By solving these equations with respect to time, the atmospheric state can be calculated for future timesteps. However, these equations are not analytically solvable and need to be discretized. The outputs of these numerical equation models are called numerical weather predictions (NWPs). [8]

Numerical weather models typically span large areas, like the entire globe or the Nordic region, and consist of equally spaced grid points. The equations are solved for each grid point, the resolution varying from one kilometer to tens of kilometers depending on the weather model. NWPs are computationally extremely heavy and for this reason they are usually run only few times a day. For example Norwegian Meteorological Institute offers new forecasts once every six hours.

The atmosphere is a chaotic system, and models are sensitive to initial conditions. For this reason measured weather data acquired from weather stations, weather balloons, sailing ships and satellites are incorporated into the model. This process is called data assimilation and it helps to increase the model accuracy. Statistical post-processing can be also be applied to the results to remove biases and increase performance. [8]

As the initial state of the atmosphere cannot be ascertained and because all weather models have some shortcomings, meteorologists have addressed the problem of weather forecast error by producing *ensemble* forecasts. An ensemble forecast is a collection of forecasts produced by models that have been initialized with altering initial states and/or model parameterizations. A single forecast is called an ensemble member. The strength of the ensemble method is that it captures the uncertainty in weather forecasts by producing a probability distribution. Weather ensemble forecasts are extremely useful for wind power forecasting. Leveraging their information for wind power models increases the performance as well as produces naturally probabilistic forecasts [8]. An example of probability distributions generated by ensemble forecasts is illustrated in Figure 2.

Some NWP models include the European Centre of Medium Range Forecasts (ECMWF), Nordic and now outdated High Resolution Limited Area Model (HIRLAM) and HARMONIE-AROME which is currently used by the Nordic meteorological institutes. [9]

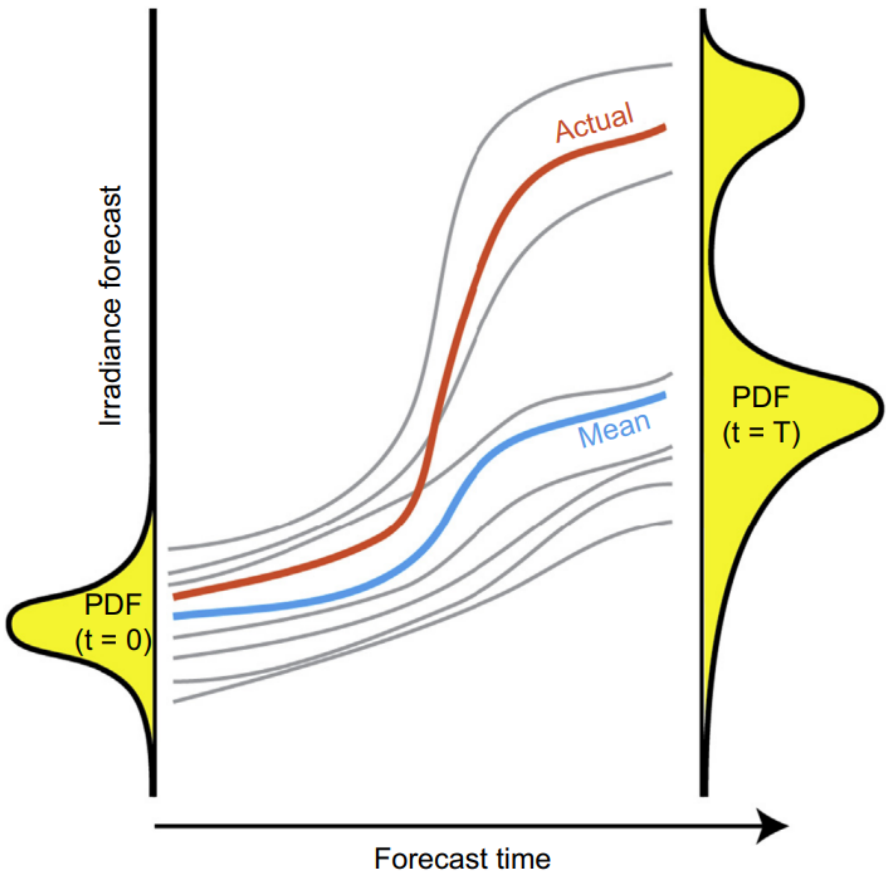

Figure 2: An illustration of probability distribution of solar irradiance generated by weather ensemble forecasts. Figure taken from [8].

2.3 Wind Power Forecast Methods

In [10], wind power forecasting horizons were divided into five categories: very short-term (seconds - 30 minutes), short-term (30 minutes - 6 hours), medium-term (6 - 24 hours), long-term (24 - 72 hours) and very long-term (72+ hours). Other formulations are also possible. The horizon of interest affects the choice of methods. Forecasting accuracy decreases with the increase of forecast horizon. [10]

There are numerous wind power forecasting methods spanning a huge number of different models. All sorts of approaches ranging from simple, deterministic equations to linear models (e.g. Autoregressive Integrated Moving Average and Vector Autoregressive models, Kalman filter) to complex artificial neural networks (ANN, e.g. Multilayer Perceptron, Convolutional Neural Network, Long-Short Term Memory, Generative Adversarial Network) with various combinations of several models into larger hybrid models and with different pre- and postprocessing steps (e.g. Principal Component Analysis, fuzzy logic, wavelet transform) have been tried. [10] gives a detailed analysis of many possible model structures found in academic literature.

In [8], typical wind power model flow is introduced as shown in Figure 3. First, SCADA data set on power production is acquired, then external weather variables from NWPs are gathered as well as technical knowledge from the wind park and its terrain. Then the data is possibly preprocessed to fit the task at hand and wind turbine power curve is constructed, whereafter that the model is ready to be used for forecasting. Lastly up-scaling from single turbine to wind park level and to a regional level encompassing multiple parks is done if needed. Kariniotakis [8] notes that some models leave out certain steps: SCADA data sets are not always available to all interested parties, when deterministic equation models combined with NWPs can bypass this, whereas ANNs often skip the explicit power curve modelling and up-scaling from turbines to park level, opting to directly estimate wind park production. It should be noted that few models in the literature incorporate information about the surrounding terrain of the wind park, even though the effect has been noted. This could be due to lack of knowledge of fluid dynamics or due to the increased workload this entails.

Several ways to describe and categorize different methods have been proposed. For example, models can be classified into parametric and non-parametric models [11], [12]. Parametric models rely on assumptions on data distribution, whereas non-parametric models do not, as they are typically data-driven and can handle arbitrary distributions. Parametric models often also implicitly incorporate more domain expertise in the model through the parameters given by experts.

The division between numerical weather prediction models (NWP-models), which use external weather prediction data as inputs, and pure statistical models, which only need historical production data, can be made too. However, purely statistical models are mostly only used only in (very) short-term forecasting, as after 3-6 hours the NWP-models outperform purely statistical ones. [8] Models can also be divided along whether they are deterministic or probabilistic. Naturally another distinction between white-box models and black-box models can be made too.

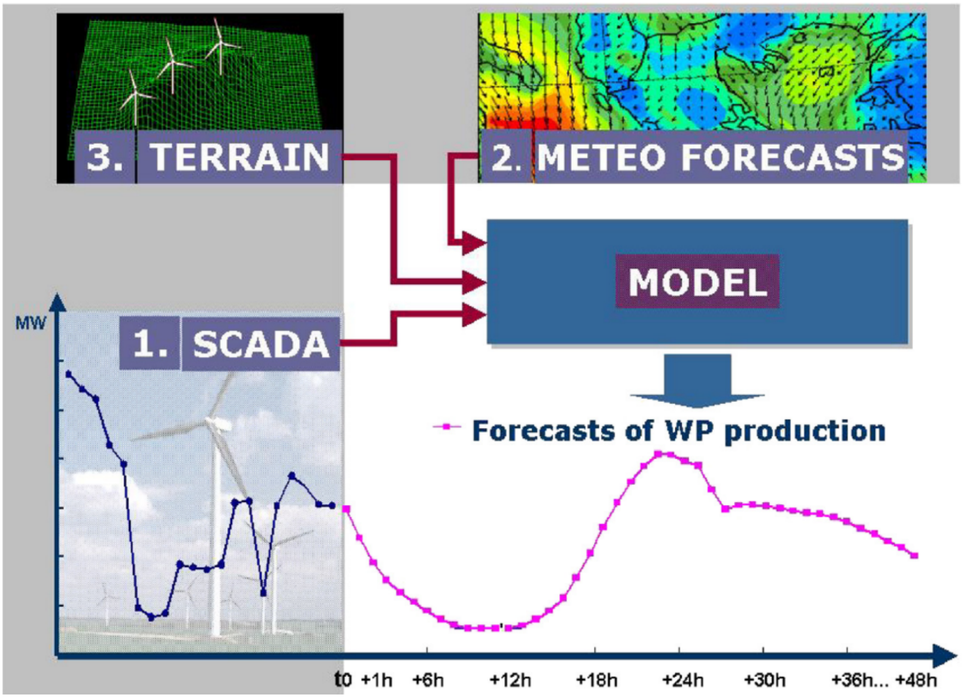

Figure 3: Typical workflow and model ingredients for wind power production forecast model. Not all models incorporate all of the given elements however. Figure taken from [8].

The aforementioned divisions are valid but not comprehensive. Instead, they should be thought as model features which together characterize different modelling approaches. This point-of-view, however, has not been explicitly expressed in the earlier literature reviews on wind power forecasting methods. One possible reason for this is the apparent similarity between different categorizations: typically incorporating large amount of domain expertise into the model results in parametric models, governed by deterministic relationships (unless stochastic part is explicitly declared) dependent on external NWPs and are easily interpretable due to their white-box model nature. Data-intensive, non-parametric and flexible, black-box machine learning models are at the other end of the spectrum. However, in reality the distinctions are not strict, as the most useful models are hybrids, which incorporate historical data, technical domain expertise, external weather forecasts and advanced statistical methods.

Based on this observation another, and perhaps a more intuitive way of describing different methods would be to present different methods based on their *data-intensity* and incorporated *domain-expertise*. According to [8], the more successful models tend to be gray-box hybrid models which synthesize both approaches. However, the latest research seems to focus more on data-intensive machine learning models, limiting the use of domain expertise into the variable space only.

Whether model is deterministic or probabilistic is depends less on the approach, and probabilistic models can often be generated at will. Deterministic models use a one-to-one mapping from weather variables and past data to the mean power output. Probabilistic models provide also the distribution of the power output. Probabilistic forecasts can be generated by presuming an error distribution - which can be learned from the data - or by running multiple future weather scenarios, obtained either from ensemble NWPs or generated through machine learning methods like Generative Adversarial Network, through a deterministic model to generate a probabilistic power output distribution.

In the literature both *wind turbine power curve modelling* and *wind power forecasting* are discussed [10] [11]. Though not the same thing, there is a significant overlap. Wind power forecasting typically tries to predict the future power output of wind farm using various methods. Wind turbine power curve modelling seeks to map the underlying causal relationships governing single turbine power production and describe power output as a function of technical information of the turbine, site and weather variables. When combined with weather forecasts and scaled to the level of wind park, wind turbine power curve models effectively become wind power forecast models.

In wind power forecasting, there are two types of forecasting errors: level error and phase error. Level error occurs when the temporal location of power output change is forecast correctly, but the exact power output is over- or underestimated. Phase error occurs when the temporal location of production change is incorrectly estimated. Phase errors can occur, for example, when weather forecasts have mispredicted the timing of weather fronts. Phase errors are important as they cause errors on several time points, and because weather fronts typically span large geographical areas, affecting several wind farms and thus possibly disrupting electricity markets. The errors are illustrated in Figure 4.

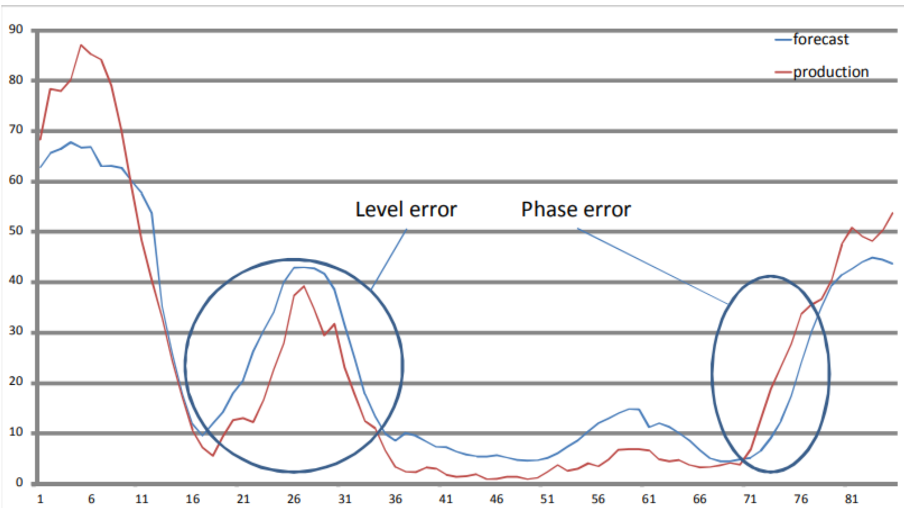

Figure 4: In wind power forecasting level error occurs when the level of production is wrongly predicted, and phase error when the temporal occurrence of output change is wrongly predicted. Figure taken from [13].

Wind power forecasting mostly focuses on mean forecasting, but supplemental uncertainty analysis is also important for practical operation of the energy system. Yan et al. [14] offer a recent, comprehensive qualitative review about wind power uncertainty, recognizing that albeit the academic literature has recognized the importance of uncertainty modelling, there are gaps in the communication and utilization uncertainty knowledge in end-user applications. This is very much in line with interviews conducted for this thesis.

Uncertainty can be presented and communicated probabilistically with probability density functions, quantiles and intervals or moments of distributions like mean and variance. It is also possible to use scenario forecasting or a risk index (e.g. a simple numerical scale from 1 to 5), which the end-user can interpret for the task at hand [15] [16]. The probabilistic approach seems to be the most typical in literature. It is used in this thesis in the form of quantile intervals.

For wind power interval forecasting two distinct approaches can be distinguished: estimating the mean and creating a forecast interval around it, and creating forecast intervals directly.

In the first approach it is rather typically assumed, that the prediction interval follows the Gaussian distribution. Due to the nature of wind, this assumption is not good for wind power forecasting [8]. This motivates the use of non-parametric methods for forecast interval creation [5]. Besides presuming a certain well-defined error distribution, intervals can be generated from previous forecast errors (see for example [5]).

Interval optimization is also typical when creating intervals from previous forecast errors, with the idea of instead of taking predefined quantiles, to find the shortest interval which contains the predefined probability mass. This method, however, has the drawback that if the tails of the distribution are long, the scenarios which are not covered by the intervals have even larger errors, a fact which has mostly been dismissed in the literature. As we have noted before, large forecast errors make up for larger percentage of the cost profile than their share of occurrences is.

For direct interval forecasting, the lower upper bound estimation (LUBE), first introduced in [17], is a popular way to create forecast intervals and has been applied to wind power forecasting for example in [18], [19] and [20]. The idea behind LUBE is that instead of training a Neural Network model to estimate the mean and variance, to directly have the Neural Network to output upper and lower bounds for the wind power output.

3 Data and Methods

3.1 Data Set

The data used for this thesis consists of weather forecasts provided by the Norwegian Meteorological Institute's open MetCoOP Ensemble Prediction System (MEPS) data set [21, 22] and wind farm SCADA data, partly supplemented by Finnish Meteorological Institute's weather data, from a single wind farm in western Finland provided by Suomen Hyötytuuli Oy.

MEPS is a NWP forecast model operated by the Norwegian Meteorological Institute and its products are freely available for public. It spans the geographical region of Scandinavia and Finland with a grid resolution of 2,5 km. MEPS contains 30 ensemble members (starting from February 2020, before that 10 ensemble members) and has forecast horizon of 66 hours, with mostly 1-hour resolution. For this thesis, forecast horizon of 36 hours was used. The data sets contain information about important weather features like pressure levels, temperature, wind speed and direction, wind gust, humidity and precipitation. Only subset of all variables found in MEPS was used. Selected variables are described in Table X 1. The grid point closest to the center of the wind farm was used for the weather data. Alternatively three or four closest points surrounding the wind farm could have been chosen and the weather conditions interpolated from these, but the this was omitted as it was deemed that the possibly increased accuracy would not justify the increased workload.

From the zonal and meridional winds, the total wind speed and direction was calculated. The ensemble forecasts were compared with the on-site weather measurements from the SCADA data. Biases in the weather forecasts were then corrected for wind speed and temperature. The wind speed was extrapolated to the turbine nacelle height using the the log wind profile extrapolation method [23]. To decide

MEPS Dataset Variables	Quantity
Zonal wind at 10 m altitude	m/s
Meridional wind at 10 m altitude	m/s
Wind gust speed	m/s
Air temperature at 2 m altitude	$^{\circ}C$
Atmosphere boundary layer thickness	m
Surface air pressure	hPa
Roughness length for momentum	m
Cloud area fraction	%
Low type cloud are fraction	%
Fog area fraction	%
Relative humidity at 2 m altitude	%
Precipitation amount	mm
Snowfall amount	mm

Table 1: Table of weather variables used from the MEPS open data.

the optimal extrapolation parameters, the squared difference to the measured wind speed was minimized, yielding the extrapolation parameter of 0.75. It should be noted that this is not perfect method, as the wind behaviour at 10 meter altitude and nacelle height can at times differ significantly, but with the absence of better data or approach this was the method chosen. The temperature and air pressure was then bias-corrected as well. The wind direction also has a near constant error of around 30 degrees, and was bias-corrected to better match the measured wind speed.

New variables were created for the MEPS data by calculating the air density of dry air by using the ideal gas law

$$\rho = \frac{pM}{RT}, \tag{3}$$

where p is absolute pressure in Pa, M is the molar mass of dry air 0.02896524 $kg \cdot mol^{-1}$, R the ideal gas constant 8.3145 $J \cdot K^{-1} \cdot mol^{-1}$ and T the temperature in Kelvin, see [7]. The effect of relative humidity was deemed negligibly small and omitted in order to simplify the calculations. From the air density and the wind speed the kinetic energy of wind was calculated with equations 1 and 2. Last, as neither of the models proposed in Chapter 4 utilize the temporal structure of the time series, instead treating each observation as independent, the MEPS data which was corrupted or not attainable due to server-side errors were removed for simplicity.

The wind farm data used comes from Nikkarinkaarto wind farm in Raahe, western Finland. The wind farm was established in 2016 and has 10 Vestas V126 turbines, with turbine rated power of 3,45 MW, nacelle height of 137 meters and rotor diameter of 126 meters. The wind farm is located on flat terrain, mostly surrounded by fields, with no major wind obstacles.

The data spans the years 2018-2020 and has a temporal resolution of 5 minutes. The wind park data has the variables timestamp, generated power, measured wind speed, measured wind direction, measured ambient temperature and turbine availability for each turbine. The data was of good quality and had very few missing values. The data does not include cut-off scenarios, as wind speeds above 22.5 m/s were not recorded for this time period, and thus the models created not are effective at studying cut-off scenarios.

The few incomplete entries were simply removed from the data set, as preserving the time series structure was not necessary. The MEPS data which was corrupted or not attainable due to server-side errors were also simply removed for simplicity. Some other preprocessing procedures were applied to the SCADA data, namely:

- Correction of systematic error in temperature measurement on two of the turbines
- Times were transformed to sine/cosine representation
- Aggregation to hourly resolution
- The data was split to train, validation and test sets with 70-15-15 % ratio.

Specifically, the time variables were transformed into cyclical variables using sine and cosine transformations for both time of the day and day of the year with the equations 4 and 5. The logic is that cyclic variables better represent the natural passing of time, as there is no cut-off point at 00:00. If cyclic encoding is not applied, the hours 23 and 01 would seem very dissimilar, which inaccurately describes the true temporal difference of only 2 hours. This typical encoding method has been applied for example, in [24],

$$h_{sin} = sin(\frac{2\pi h}{24}) \tag{4}$$

$$h_{cos} = cos(\frac{2\pi h}{24}). \tag{5}$$

Similar encoding was done for the day of the year. These encodings were also applied to MEPS data set.

Because the wind farm records a limited number of weather measurements, weather forecasts with one hour forecast horizon originally made by FMI were used to complement for this deficit. The forecast data was assumed to be accurate enough and used as a replacement for the missing variables for data matching purposes and calculating less important equation terms for the models. Treating estimated data as measurements incorporates some errors, but this was deemed both acceptable and necessary in this context.

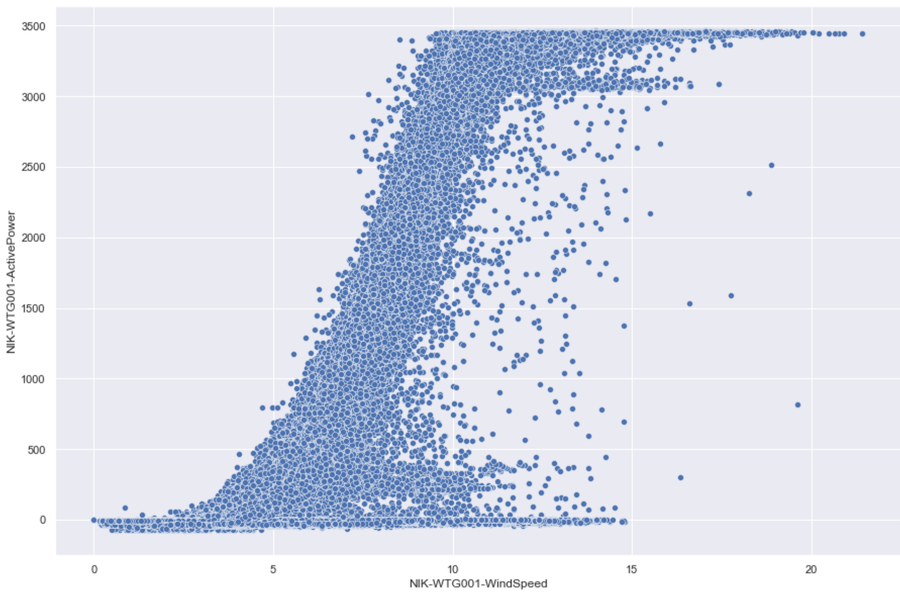

Figure 5: Scatter plot of wind power production as a function of measured wind speed. Both corrupted data points and large deviations from the mainline production can be observed.

3.2 Model Performance Metrics

Statistical performance metrics are necessary to estimate the accuracy and goodness of the chosen models. In choosing error measures, it is important to measure the model in metrics that best relate to the actual variable of interest; here it is balancing cost of forecast errors for the wind producer. It is natural to assume that in the market context larger deviations have larger negative impact than what would be the equivalent sum of many small errors. Suomen Hyötytuuli has also stated that most of their balancing costs come from relatively few instances, indicating that larger deviations dominate the cost profile. Thus, in this thesis, it is beneficial to give more weight for evaluation metrics that penalize observations with larger deviation from the mean more, such as squared error metrics, when evaluating the performance of forecast mean.

In this thesis the forecast mean is evaluated with the Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y} - y)^2}, \tag{6}$$

where $\hat{y}$ is the predicted power and y is the real, measured power output.

Also Mean Absolute Error (MAE),

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|, \tag{7}$$

has its merits as a metric. This was not used as an objective function in fitting the models unlike the squared error metric, but MAE is possibly more intuitive to interpret for human reader due to its linear nature.

The coefficient of determination, typically called R^2, describes the proportion of variation that the regressors explain. It is the square of Pearson correlation coefficient and alternatively defined as one minus the sum of squared residuals divided by total sum of squares (i.e., total variation),

$$R^2 = 1 - \frac{\sum_i^N(\hat{y}_i - y_i)^2}{\sum_i^N(y_i - \bar{y}}, \tag{8}$$

where N is the number of data points and $\bar{y}$ is the mean of the dependent variable.

The errors in forecasts are not sign independent: overestimating the production resulting in negative residual has stronger consequences than underestimating, as the surplus can often be sold at profit or curtailed at zero cost. For simplicity and the difficulty of implementing such asymmetric cost function, this piece of information is not incorporated into performance metrics, and instead only visually and verbally analysed in the Results section.

The constructed power production distributions are evaluated by standard deviation, prediction interval width D and the prediction interval coverage percentage (PICP), which measures the share of test points falling inside the 95 % prediction interval. Ideally, we would want to minimize both the standard deviation, the interval width D and maximize PICP, but these objectives are contradictory. For this reason all metrics are studied together to find a balanced outcome.

For a test set of power production realizations X, with size N, the coverage is given by

$$\mathrm{PICP}(X,N) = \frac{\sum_{i=1}^{N} I(x_i)}{N}, \tag{9}$$

which yields a value between 0 and 1 and where $I(x)$ is an indicator function:

$$I(x) = \begin{cases} 1, x \in Interval \\ 0, x \notin Interval. \end{cases}$$

Another metric used is what we call a D-score. It is simply the width of the interval and is measured in kWh/h. A smaller $D-$score is better. This is objective is contradictory to PICP, as the widening of the interval typically increases the PICP score, and these two need to be examined together.

4 Models

We consider three different models for the task. The first model is a simple statistical nearest-neighbour model. Its idea can be described as a similar day analysis: similar observations from the past, based on different weather variables, are collected. A single-point estimate for the mean of the power production is calculated as a weighted average.

The second model is Random Forest model. Random forest is an ensemble model which consists of a set of decision trees. Decision trees divide the variable space into several sections, which are estimated with constant functions. The predictions are obtained by taking the mean of estimates produced by all trees.

The third model utilizes the technical domain knowledge of wind and wind turbine technology by describing the power production with deterministic mathematical equations. We call this model the equation model. The model takes as input parameters turbine specific technical information and weather data. The aim is to construct an ideal power curve for the production and then have additive/multiplicative components which adjust the ideal production to real world situations with power losses.

The modelling process is shown in Figure 6.

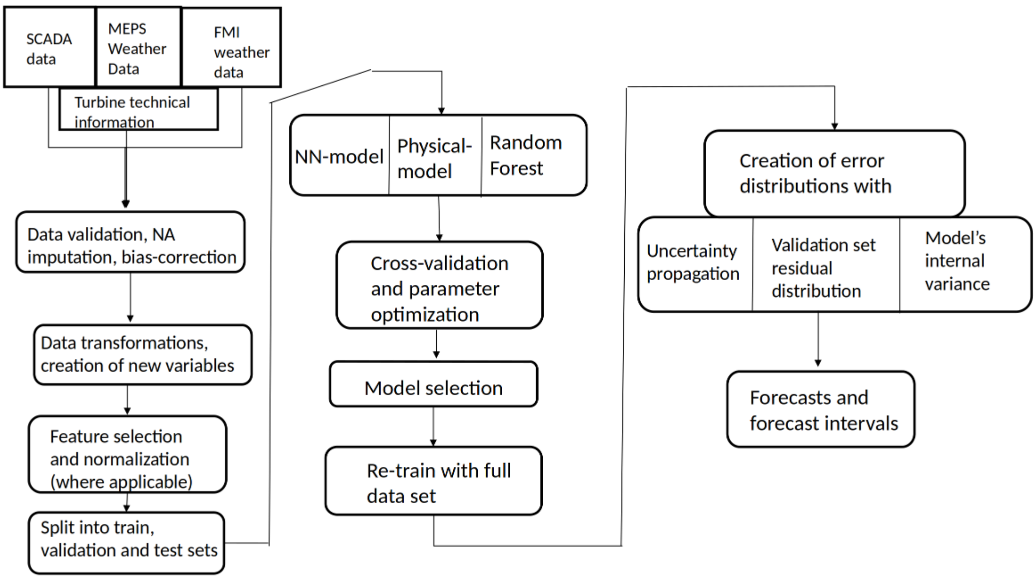

Figure 6: Visual presentation of the modelling workflow.

4.1 Nearest Neighbour Model

Nearest neighbour methods are generic statistical algorithms, originally introduced by Fix and Hodges in 1951 [25] in the form of k-nearest neighbours (k-NN) method. The idea behind the k-NN is simple: for data points with missing variable values, the training set is searched to identify k nearest observations in m-dimensional space based on the available variables and the missing data points are inferred by aggregating over the values of the k identified data points. Nearest neighbour methods can be thought as the mathematical equivalent of similar case analysis, where (almost) similar observations from the past are studied and the new data is assumed to behave accordingly.

In the k-NN algorithm, the parameter k is a predefined hyperparameter. The optimal value of k can be learned through model cross-validation. The closeness of an observation is defined by a distance metric. A typical choice of metric is the Euclidean distance which for points p and q is given by

$$(p, q) = \sqrt{\sum_{i=1}^{M} (p_i - q_i)^2}, \tag{10}$$

where M is the number of dimensions. However, any other norm can also be used. The typical aggregating methods over the neighbours are the simple mean and weighted mean. The methods are applicable for both regression and classification.

In this thesis, we match predicted weather conditions with past observations and calculate the predicted power output as a weighted mean of the past observations. The benefits of the approach are straightforward and easy implementation, computational efficiency and that the approach models the spread of the power production directly, without the need to explicitly model wind turbine operation. Another characteristic is that little prior knowledge of wind power is required, as the results are based entirely on past observations: the advantage is that all predictions have a degree of reliability as they are based on actual realizations. On the other hand, the disadvantage is that the data is required to be comprehensive, as novel operating circumstances are poorly processed without similar data.

In the model implementation, the MEPS weather data was treated as correct weather predictions, and was matched with historical data from the site. For this only the weather variables available for the wind farm were used, and some of the MEPS variables were thus left out. This ignores the weather forecast errors for the moment, but has the advantage of allowing the separation of weather forecast error and model error.

The model was implemented using the scikit-learn library [26] of the Python programming language. The implementation was programmed to use the weighted Euclidean distance in the nearest neighbour search.

Power production for the wind farm was predicted for all ensemble members individually and then the results were aggregated for each data point. Forecast intervals were constructed by taking the minimum and maximum ensemble predictions for each data point, and the interval was compared to the test data to calculate the PICP and D scores.

The nearest neighbour method does not utilize the time series structure of the data, but instead treats each observation as independent of all others. One possible way to improve its performance could be to match weather sequences instead of individual data points. This, however, is left as a note for future research.

4.2 Random Forest Regressor

Random forest is an ensemble method which uses a collection of randomized decision trees to create predictions. First introduced in 1995 by [27], Random Forest is a non-linear black-box method that can be used for both classification and regression.

A Random Forest is created by building a collection of trees, each of which is constructed by randomly selecting a subset of variables and using the bootstrap aggregation method (bagging) to create a training set for the tree. The point in bagging is that randomly sampling a unique training set for each tree de-correlates the trees increasing its forecast performance. The final output of the forest is the aggregate of each individual tree - hence the name Random Forest.

One of advantages of the Random Forest is that because it is based on decision trees, it can produce variable importance scores. In scikit-learn these are calculated as Gini importance [26]. Another advantage is that as the forest aggregates over a host of forecasts, the distribution of these forecasts can also be studied besides the mean estimate. Random forests are commonly used models, and the standard scikit-learn implementation was used for building the model in this thesis.

As for the k-NN model, the model was at first trained with measurement data and for test predictions MEPS weather predictions was given as input. Again, this had the advantage of separating the weather forecast error and model error.

Because the Random Forest model performed best, as reported in Chapter 5, this model type was subjected to further study, namely to be trained directly with MEPS data. In this approach the weather and model uncertainties are no longer separable, but as the model is trained with uncertain real world data, allowing the model to learn about the uncertainty of the MEPS weather forecast data. It can also handle weather variables not measured/not measurable at the site, thus utilizing larger volumes of data. The assumption is, that the model could learn possibly some systematic performance issues in the weather forecasts, could learn the relationship between the weather forecast location and the wind farm, which differ both in geographical coordinates and altitude, which had thus far been ignored. The model can thus be assumed to implicitly post-process the weather forecasts.

This model actually consists of a collection of Random Forest models, one for each time horizon. The reasoning is, that the weather forecast uncertainty increases with the increase of forecast horizon, and thus the most reliable results should be obtained when training and testing horizons are the same. As there is multiple ensemble forecasts for each time point, the model was evaluated both with all ensembles and the means of ensembles as training data. The logic behind using means is that is lowers the computations significantly and might produce more robust training data.

4.3 Equation Model

Wind power has been extensively described by physical equations in the literature. The power of wind can be calculated with Equations (1) and (2). The kinetic energy is then converted into kinetic energy of the turbine and further into electricity. These transformations cause losses and can be modelled by multiplying the kinetic energy of wind by conversion coefficients c_p for kinetic energy capture and c_e for electricity transformation. The full electricity output of wind turbine is given by the equation

$$P_e = \frac{1}{2}\rho A u^3 c_p(u) c_e, \tag{11}$$

where ρ is the air density, A the rotor sweep area, u the wind speed, c_p the power coefficient for the conversion of the mechanical energy of wind to the mechanical energy of the rotor and c_e the electrical conversion coefficient.

This is the theoretical equation for wind power generation and can be thought of as the ideal production power. In reality there are typically power losses and the experienced production is smaller. Factors causing losses are turbulent winds, which cause suboptimal wind energy capture, ice forming on the blades (icing), which changes the aerodynamic properties of the blades and can lead even to complete turbine shutdown, wake effect caused by some turbines of the wind park shading others, errors in wind direction measurements and turbine positioning and technical malfunctions to name some.

Potential approaches to model the (ideal) wind turbine power curve are extensively described in [11] and [12]. Possible methods for fitting empirical power curves include but are not limited to polynomials, exponential functions, logistic functions and kernel regression.

The aerodynamic efficiency of the rotor blades η_b, the mechanical efficiency of the motor η_m and the electrical efficiency of the generator η_e effect the total efficiency of the turbine. These can be represented together as the coefficient of power c_p [28]. We assume that both the mechanical and electrical efficiency depend only on the technical design of the turbine and are thus constant, whereas the aerodynamic efficiency is changed when the rotor blades' pitch angles are changed according to the wind speed.

To model wind turbulence is difficult, because turbulence can be very local and chaotic, not well described by weather models and typically not provided in weather services. Modelling icing is also difficult, because ice changes the complex aerodynamic properties of the blades that are not well captured by mathematical equations even given the ice level is well measured - which it usually is not. Weather services often also provide icing forecasts, although for our historical data from 2018-2020 this is limited. For these reasons the model was simplified by leaving icing outside of its scope.

Modern wind turbines are built to automatically adjust the rotor positioning to directly face the wind. Due to errors in wind direction measurements ideal positioning is not always achieved which causes power loss and can be modelled by taking the cosine of the error angle $u_{eff} = u_0 cos(\alpha)$, [29]. In principle, this could be estimated with accurate light detection and ranging measurements (LIDAR), but this data is not included in our data set, and must thus be left outside the model.

Typically, wind turbines are optimally spaced to prevent the wake effect. The wake effect occurs when wind comes from a direction that causes turbines to shade others. Some of the energy of the wind is absorbed by the shading turbines and the air flow downstream becomes turbulent [30], leading to efficiency losses in the shaded turbines. This could be modelled, but is deliberately left outside the scope of this thesis to simplify the modelling task. It is noted as a topic for potential future research.

Summarizing the above notes, the full equation would get the following form:

$$P_e = \frac{1}{2}\rho A u^3 c_p(u) L_{icing}(T, hum, cloud) L_{wake}(\alpha) L_{turb}(\alpha) L_{merr} L_{mal}, \tag{12}$$

where α is the wind direction, L_{icing} is losses due to icing which are affected by temperature, humidity and cloudiness, L_{wake} is losses due to wake effect, L_{turb} is losses due to turbulence, L_{merr} is losses due to measurement error and L_{mal} is losses due to technical malfunctions.

With our limitations the ideal power curve equation is simplified to

$$P_e = \frac{1}{2}\rho A u^3 c_p(u). \tag{13}$$

Now the only unknown is the power coefficient function $c_p(u)$. This can be calculated by moving the other terms to the left side. The results are depicted in Figure 7. The red horizontal line presents the theoretical maximum given by Betz's law. All points above this line were dismissed as measurement errors. Data from all 10 wind turbines was used, as they share identical technical properties.

The function for the ideal c_p was estimated as a function of kinetic energy. Wind speed was also considered as the independent variable, but this lead to larger spread as it does not include the effect of air density on the estimation, so kinetic energy was chosen instead. The data was binned and the 85th percentile was taken from each bin. 85th percentile was chosen after some manual inspection as it gives good performance unhindered by large losses and it is devoid of overperforming outliers. For the selected data points a kernel regression model was finally fitted. The resulting power coefficient function is presented in Equation (8).

This presents the equation model which takes as input kinetic energy of wind and air density and outputs the production of single turbine. Production for the entire wind farm is attained by multiplying the result for one wind turbine by ten.

The result describes the power coefficient behaviour for ideal production and is not applicable in the presence of losses. One possibility is to multiply the ideal curve with a black-box loss term $L(T, \alpha, humidity, ...)$ which can be estimated with a neural network, which would lead to a gray-box model. This was also one of the original intentions in this thesis, but due to largely increased workload was left as a prospect for future research.

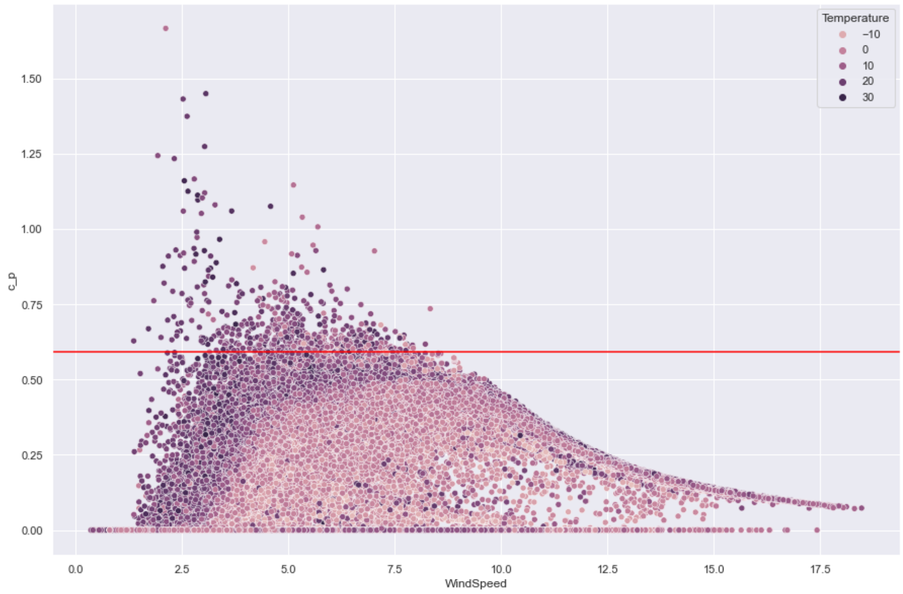

Figure 7: A scatter plot of power coefficient as the function of wind speed. Colouring shows the temperature in Celsius degrees. Red horizontal line is the theoretical maximum coefficient as given by Betz's law.

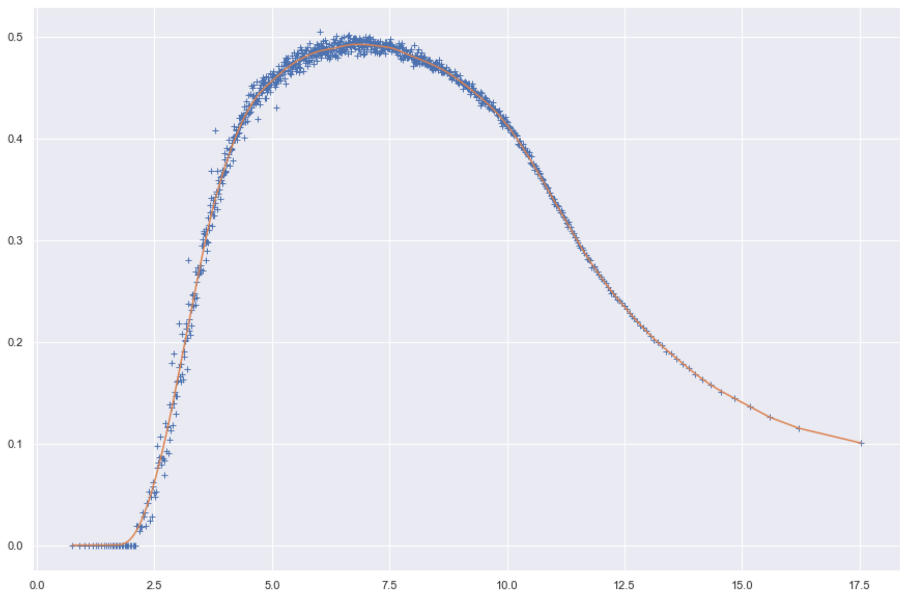

Figure 8: Kernel regression estimation of the ideal power curve for the turbines.

5 Empirical Results

In Chapter 5.1 each of the three model types (k-NN, Random Forest and Equation Model) trained with measurement data were examined first with wind power park availability corrections present. The models are evaluated by their ability to forecast the mean of the power production. Of these models, the Random Forest performed best and was thus subjected to a more thorough examination in Chapter 5.2 by training it directly with weather predictions. In Chapter 5.3 three methods for creating forecast intervals are introduced and the results are studied.

To ensure the comparability to the measured weather data, we use consistently the shortest forecast horizon of one hour for comparing the results. Exception for this is uncertainty propagation in Chapter 5.3.3 where forecast horizon of three hours was used because one hour horizon is consistently missing some weather variables.

5.1 Evaluating the Forecast Performance with Measured Weather Data

5.1.1 Nearest Neighbour Model

The best model was selected from a set of hyperparameters through a cross-validation procedure. A feasible choice of hyperparameters was selected using the PICP-score, $D-$score and RMSE as criteria based on the modeller's view. The selected model was a k-NN model with 20 neighbours that used the weighted Euclidean distance with the weights being the absolute values of Pearson correlation coefficients between independent variables and the power output of one of the turbines. The values for the weights are shown in Table 2. The results were calculated for forecast horizon of one hour. Other horizons could have been chosen as well, but it was deemed that this horizon best corresponds to predictions made with measurement data.

The scores for best k-NN model for selected time horizons are shown in Table 3. In general, the forecast results tend to become worse as forecast horizon increases and there is more uncertainty. This effect can be seen in the increase of RMSE and the $D-$score. Because of the larger forecast interval the PICP score increases, though this is not related to better performance.

The effect of forecast horizon is not all that drastic, as shown in Figure 9. In other words, the shorter forecast horizon does not provide that much better forecasts. The reasons for this are most likely that the MEPS data does not have the newest weather measurements as well as the model lacking auto-correlation structure i.e. it does not utilize knowledge of current production. The model is not optimally constructed for short-term forecasting.

Though this thesis was restricted to studying the variability of operational turbines, a quick check about the effect of availability correction was performed. Without the availability correction the RMSE was 8684 kWh/h for one hour forecast horizon, whereas with correction RMSE was 8648 kWh/h. The difference is small, almost negligible. From this we can conclude, that unexpected dysfunctions of the turbines account for very small part of the total forecast errors over a long time period.

Table 2: The absolute values of correlation coefficients of control variables with power output of one of the wind turbines. Values are used as weights for the distance metric in nearest neighbour matching. FMI refers to the variable being the one hour prediction produced by Finnish Meteorological Institute instead of being measured.

Variable	Correlation
Mean Ambient Temperature (measured)	0.070772
Wind Speed (measured)	0.887680
Wind Direction (measured)	0.051706
Time of day, sine	0.041399
Time of day, cosine	0.118889
Time of year, sine	0.006463
Time of year, cosine	0.181000
Air Density (FMI)	0.048490
Air Pressure (FMI)	0.152657
Air Temperature (FMI)	0.109525
Humidity (FMI)	0.092030

Table 3: Test results for the k-NN model with selected time horizons.

	RMSE	MAE		R^2 D-score	PICP
1h	8648	5986	33.7 %	7072	36.8 %
6h	9357	6294	31.1 %	8367	42.8 %
24h	9602	6404	27.7 %	9197	46.1 %

In a sense, this is an encouraging discovery, as in actual operation setting these disturbances are difficult to account for.

For quantifying the effect of weather forecast uncertainty on the power forecasts, the model was tested with both the weather forecast data with forecast horizon of one hour and the actual weather measurements at site. The unexplained variance in the forecasts based on the actual measurements consists of measurement errors and model error, whereas power predictions based on weather forecasts also include the error created by weather forecasting uncertainty. For actual measurements MAE was 1239 kWh/h and for weather-based predictions 5986 kWh/h. This means that 82.8 % of the absolute deviations are due to the external weather forecast uncertainty, whereas the combined share of measurement and model errors was only 17.2 %. The error sources are illustrated in Figure 10.

The number of ensemble members provided by MEPS changed from 10 to 30 in February 2020. There is a noticeable difference between the number of ensembles used in the power forecast. The spread of wind speed is better captured 30 ensemble members. As shown in Table 4, the model predicted the mean with same level of confidence with both numbers of ensembles, but for 30 members the D- and

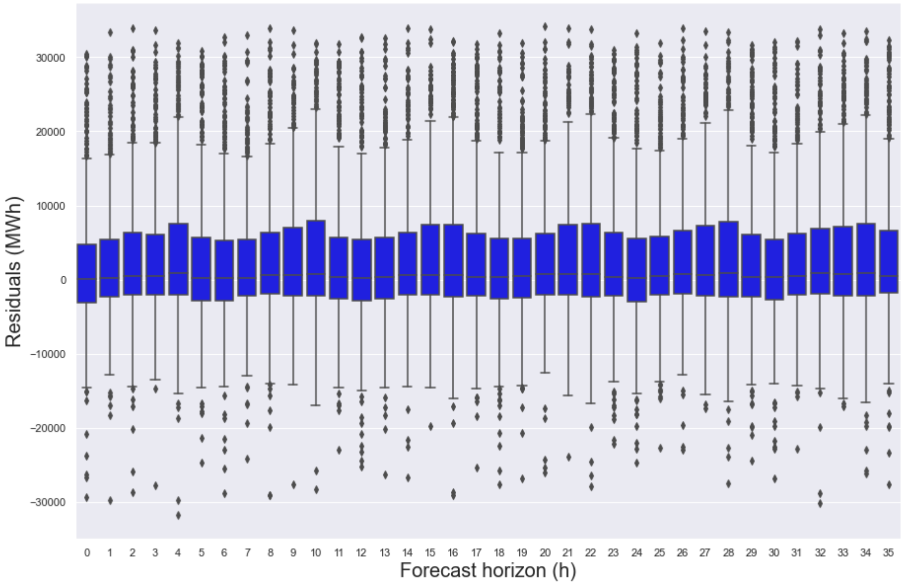

Figure 9: Boxplots of forecast residuals for each forecast horizon.

PICP-scores were higher. In other words, the model with describes the mean of the production accurately regardless of the number of ensemble members used, but 30 members better capture the spread of the weather scenarios.

Table 4: Test scores for different numbers of ensembles used.

Ensemble members:	RMSE	R^2	MAE	D-score	PICP
10	8675.9	34.0 %	6001.5	5831.7	30.5 %
30	8597.2	32.8 %	5958.0	10323.3	53.3 %

The k-NN model performs differently in different wind speed regions, see Figure 11. For small wind speeds the maximal forecast error resembles the power curve. At rated wind speed (10 m/s) the error starts to decrease - this is due to the turbine adjusting its power coefficient to produce steady stream of power, meaning that changes in wind speed do not significantly affect the power output as long as the effective wind speed is within the rated power range of [10, 22.5 m/s].Underpredictions occur almost completely on wind speeds less than 10 m/s. This is a useful result because underpredictions are more harmful than overpredictions for the wind power operator.

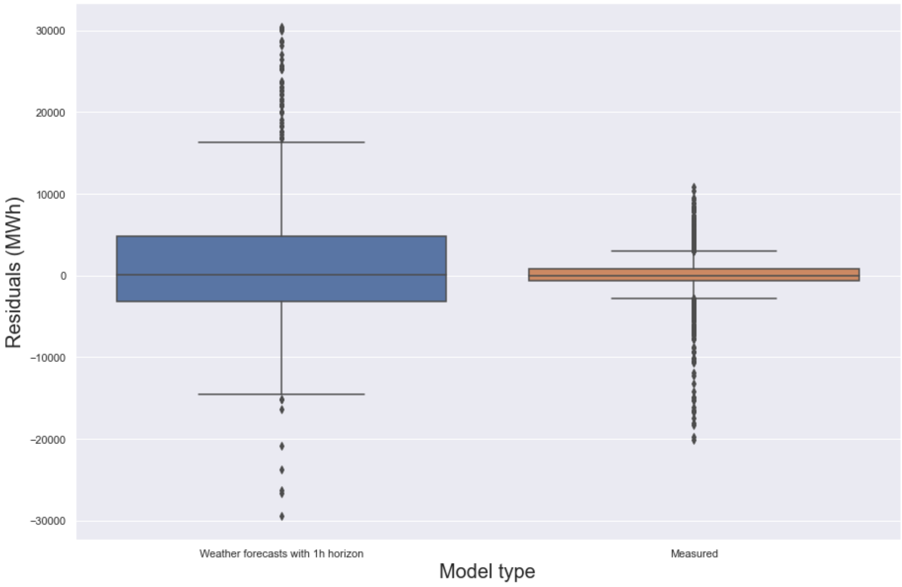

Figure 10: Boxplot of forecast residuals generated both by weather forecast inputs and actual measurement inputs. The variance in the measurement-based residuals are attributed to measurement errors and inherent model error, whereas weather-based model incorporates the weather forecast uncertainty, which is the most significant source of error.

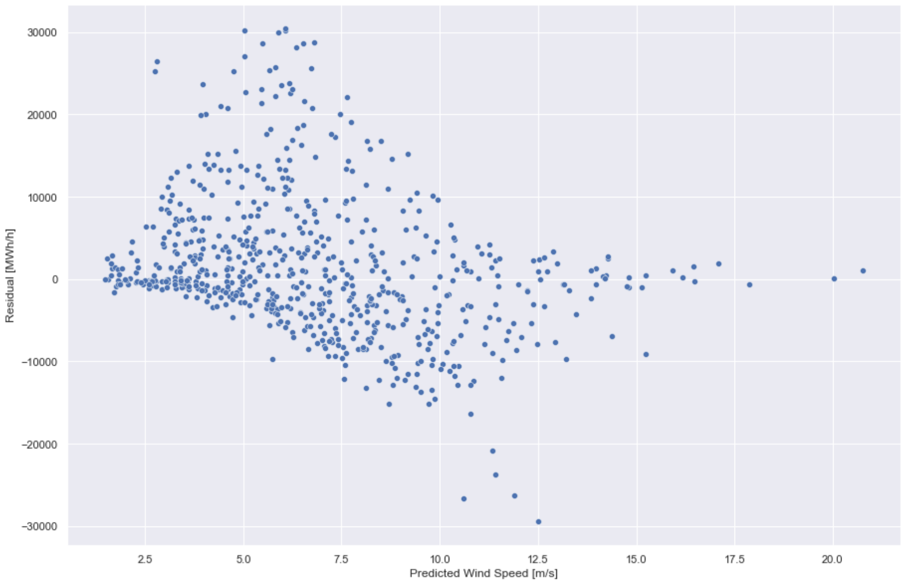

Figure 11: Model residuals as a function of wind speed. Negative residuals correspond to underprediction and positive residuals to overprediction. When the wind speed reaches the rated power interval [10, 22.5 m/s], the error starts to decrease significantly and no underpredictions occur. Underpredictions are encountered only for lower wind speeds.

5.1.2 Random Forest Regressor

The model hyperparameters were obtained by executing a random search on a large hyperparameter space to narrow it down, and then performing a more thorough grid search in the narrowed space. The metric used for evaluation was RMSE.

The attained variable importances are presented in Table 5. Wind speed is by far the dominant variable presenting 76 % importance.

Table 5: Variable importances for Random Forest Regressor.

Variable	Variable importance
Mean Ambient Temperature	0.039
Maximum Wind Speed	0.760
Mean Wind Direction	0.027
Time of the day, sine	0.006
Time of the day, cosine	0.010
Time of the year, sine	0.032
Time of the year, cosine	0.046
Air Density	0.017
Air Pressure	0.024
Air Temperature	0.020
Humidity	0.018

The model was tested with one hour forecast horizon. The test scores are presented in Table 6. Now the test scores are better than with k-NN model, but not much superior.

The histogram of test residuals is presented in Figure 13. The test residuals are plotted as a function of predicted wind speed in 12. The pattern is similar as with k-NN model, the residuals increasing as wind speed reaches zone II of the power curve, and starting to decrease as it reaches zone III.

Table 6: Test score for Random Forest Regressor with forecast horizon of one hour.

RMSE	MAE	R^2	D-score	PICP
8445 kWh/h	5819 kWh/h	36.8 %	6537 kWh/h	32.1 %

The uncertainty sources were quantified by comparing the model with results from measurement data. The MAE score was 804 kWh/h for measurement data. This part was explained by the model and measurement errors whereas when the weather forecast uncertainties were incorporated the total error was 5822 kWh/h. This means that 86.2 % of the result errors were explained by the weather forecast errors. This is illustrated in Figure 14. Again, the main source of uncertainty lies in the obtained weather forecasts, not in the model itself.

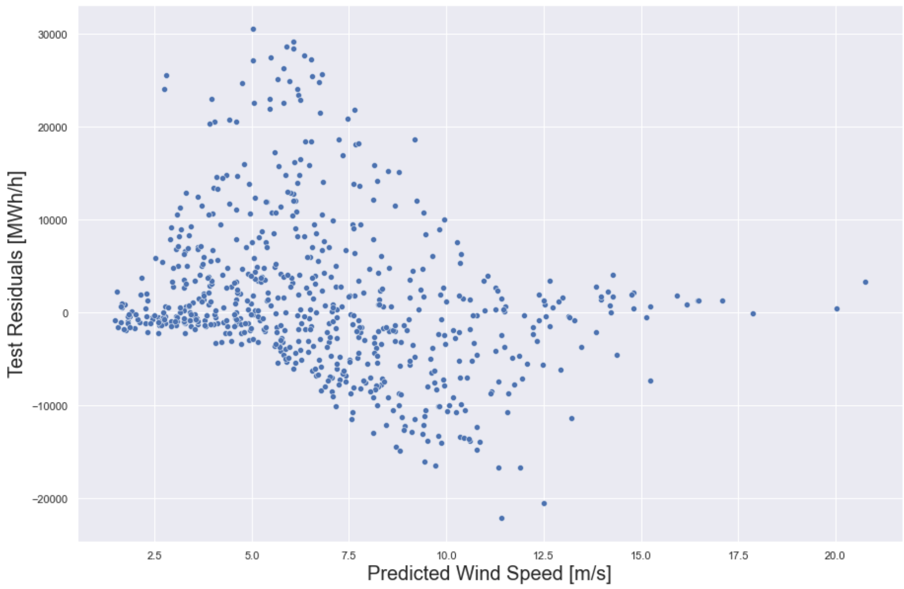

Figure 12: Scatter plot of residuals for Random Forest model as a function of wind speed for forecast horizon of one hour. Negative residuals correspond to underprediction and positive residuals to overprediction.

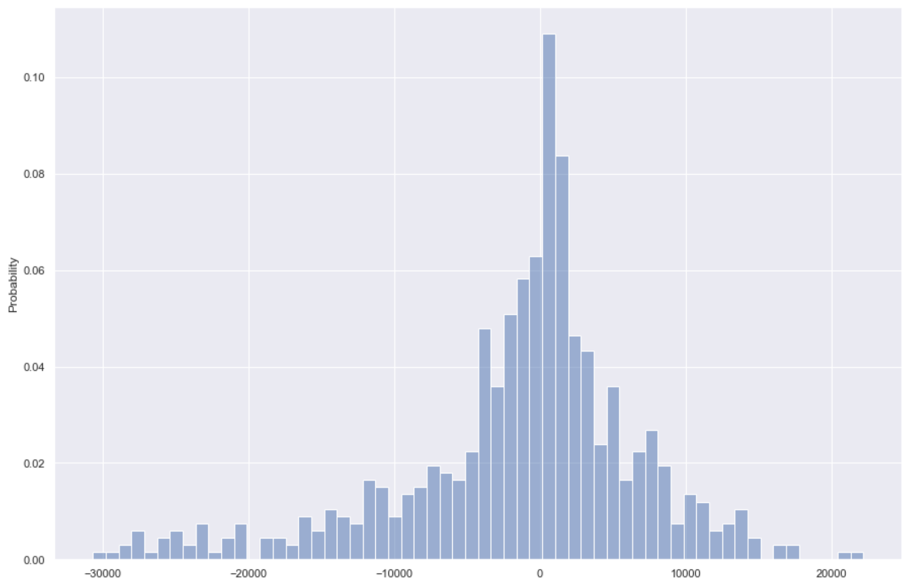

Figure 13: Histogram of forecast residuals with Random Forest Regressor and MEPS weather data with forecast horizon of one hour.

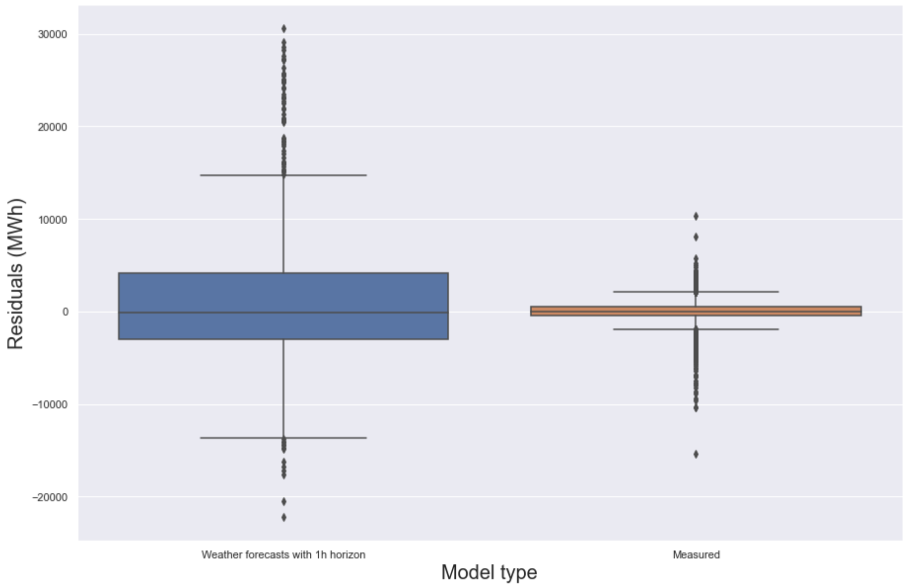

Figure 14: Boxplot of forecast residuals generated both by weather forecast inputs and actual measurement inputs. The variance in the measurement-based residuals are attributed to measurement errors and model error, whereas weather-based model incorporates the weather forecast uncertainty, which is the most significant source of error.

5.1.3 Equation Model

When the equation model for ideal production was used without taking into account the loss terms, the obtained residuals showed a high degree of spread and nearly consistent over-prediction as is evident in Figure 15.

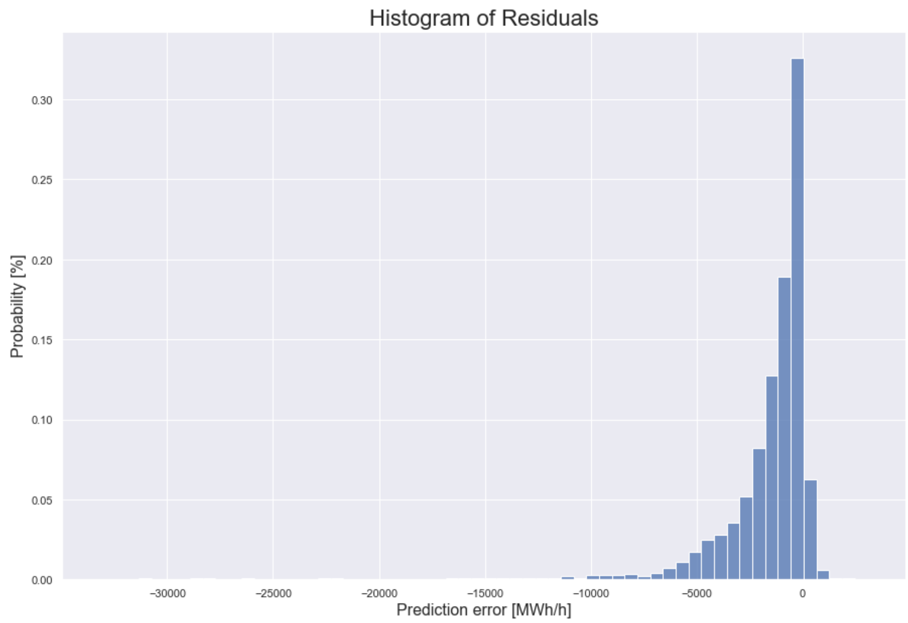

Figure 15: Histogram of the residuals generated by the equation model. Significant over-prediction is present.

Table 7: Test score for Equation Model.

RMSE	MAE	R^2	D-score	PICP
9542 kWh/h	6964 kWh/h	19.3 %	859	11.1 %

The test scores are shown in Table 7. For ensemble predictions, the MAE for the mean of the ensembles for one hour forecast horizon was 6964 kWh/h whereas for measured weather data it was 3459 kWh/h. Again, 6964 kWh/h represents the combined model and weather error, whereas 3459 kWh/h represents the model error. This means that 50.3 % of the result errors were explained by the weather forecast errors. This share is significantly lower than for the other models considered which is explained by the rather high model error.

The 6964 kWh/h for MAE is very high and indicates severe shortcomings of the model. The 3459 kWh/h for perfect weather information is not too good either. In consequence, the ideal model oversimplifies the phenomenon because it leaves out loss terms that are necessary for acceptable model performance.

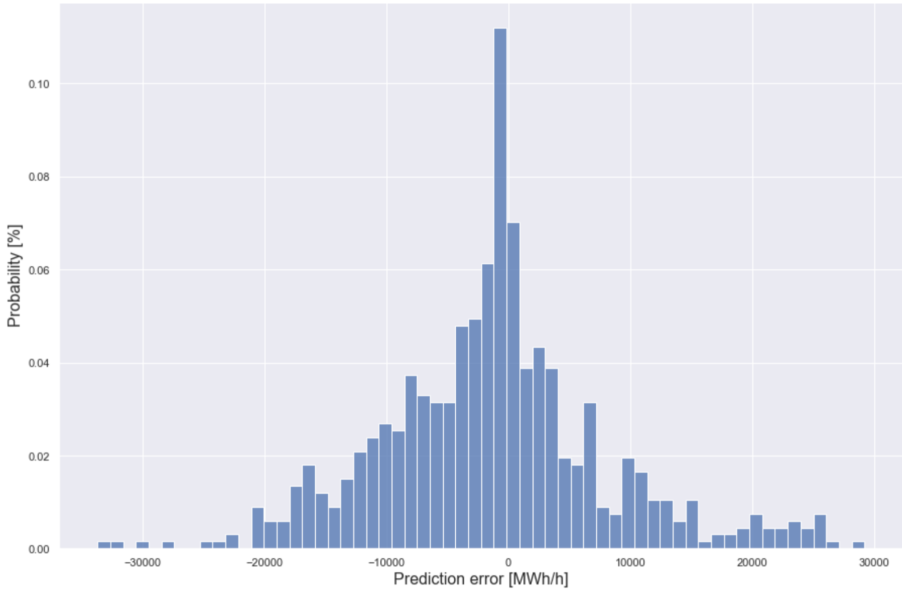

Figure 16: Forecast residuals for the test set of the ideal power curve model for the MEPS data. Residuals are calculated from the mean of the ensemble forecasts. The forecast errors are extremely large, as the maximum output of the wind farm is 34.5 MW.

5.2 Evaluating the Prediction Performance of Random Forest Model Trained with Weather Forecasts

The best model parameters were again attained by doing a random search of the parameter space complemented by a grid search the same way as for model trained with measurement data. The variable importances are presented in the Table 8. The metric used for evaluation was RMSE.

The model consists of 36 independent models, one for each forecast horizon. The evaluations were done for one hour forecast horizon. Unlike for the model trained with measurement data, wind speed is no more the vastly dominant variable, although it is the most important single variable. Training with both the means of ensembles and all ensembles were tested. Of these the use of full ensembles provided better results and was thus the focus of analysis.

Results on model performance with selected time horizons are in Table 10. For different number of ensembles used the results are in Table 11. For 30 members used the forecast interval is wider and PICP score higher, meaning that more ensembles better describe the weather distribution as mentioned earlier. Interestingly now there is also a significant improvement in the forecast performance of the mean - in contrast with k-NN model where there was no improvement. This is an indication that with more ensemble members more uncertainty can be captured, learned and utilized in the prediction step.

Figure 19 presents the test set residuals as a function of predicted wind speed. Unlike for other models, there is no longer any clear pattern. Instead the residuals

Table 8: Variable importance scores calculated for Random Forest trained with MEPS data.

Variable	Variable importance
Atmosphere boundary layer thickness	0.101
Surface air pressure	0.031
Cloud area fraction	0.015
Low type cloud area fraction	0.034
Fog area fraction	0.011
Relative humidity at 2 meter altitude	0.040
Roughness length for momentum	0.028
Precipitation	0.013
Snowfall	0.011
Time, sine of hour of day	0.021
Time, cosine of hour of day	0.055
Time, sine of day of the year	0.052
Time, cosine of day of the year	0.097
Wind direction	0.047
Air temperature nacelle height	0.046
Air pressure at nacelle height	0.031
Air density at nacelle height	0.032
Wind speed at nacelle height	0.205
Wind speed of gust at nacelle height	0.131

Table 9: Test score for Random Forest Model trained with MEPS with a forecast horizon of one hour.

RMSE	MAE	R^2	D-score	PICP
5308 kWh/h	4062 kWh/h	75.0 %	4000 kWh/h	25.3 %

look like white-noise, with heterogeneous variance. This is good as there is no longer any indication of a major model bias.

In Figure 18, the two-dimensional density estimation of wind power forecast residuals as a function of wind speed is shown. The red line presents the moving average of the residuals, illustrating the small change in average prediction error in different regions. This figure is in line with the power curve: in Zone II the residual distribution is wide, whereas in other zones it is much more manageable from practical viewpoint, as was echoed in earlier results.

Table 10: Test results for the Random Forest model trained with MEPS data with selected time horizons.

	RMSE (kWh/h)	MAE (kWh/h)	R^2 (%)	D-score (kWh/h)	PICP (%)
1h	5308	4062	75.0 %	4000	25.3 %
6h	5491	4313	76.3 %	4847	26.0 %
24h	5858	4551	73.1 %	5299	25.3 %

Table 11: Test scores for different numbers of ensembles used.

Ensemble members:	RMSE (kWh/h)	MAE (kWh/h)	D-score (kWh/h)	PICP (%)
10	5539	4288	3550	20.8 %
30	4673	3500	5140	36.3 %

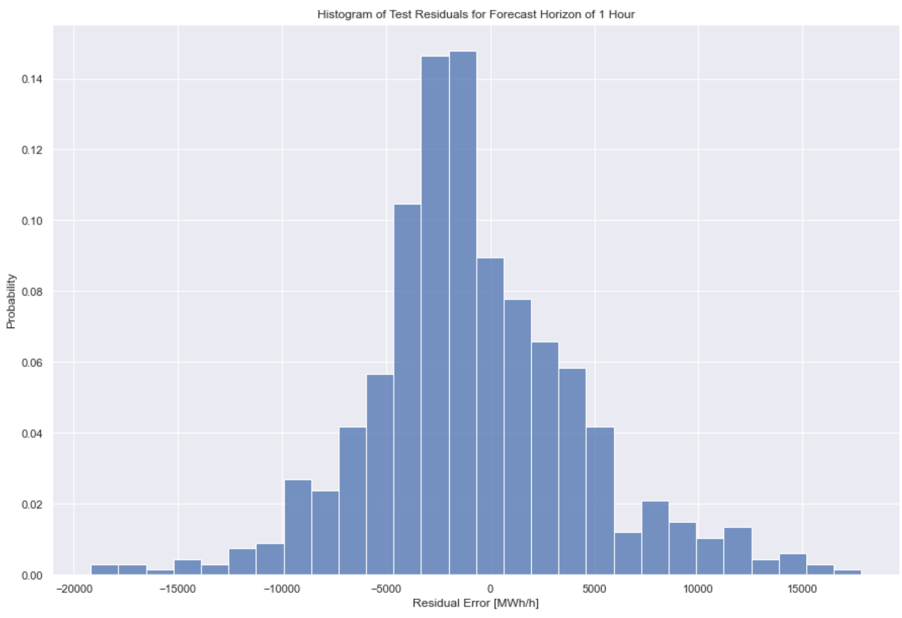

Figure 17: Histogram of residuals for forecast horizon of 1 hour.

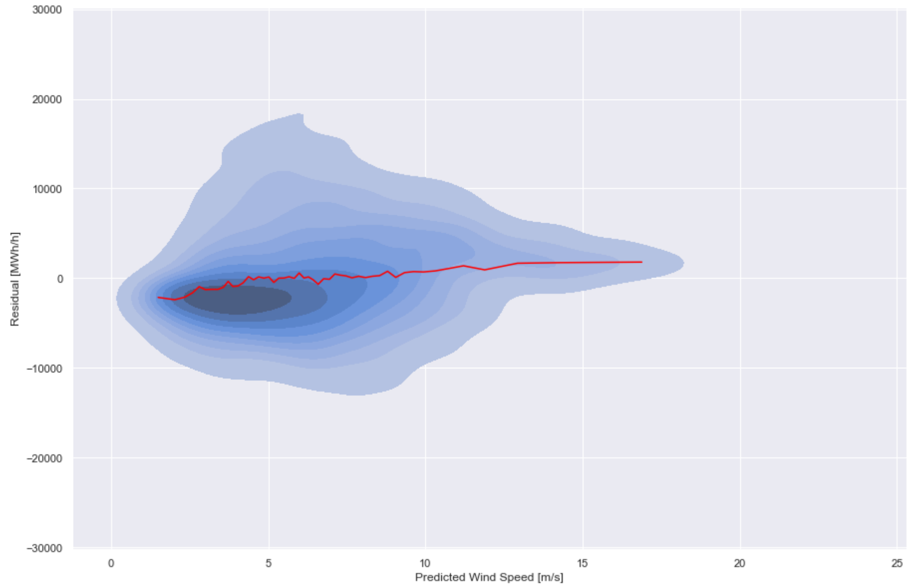

Figure 18: Plot of kernel density estimation for the distribution of forecast residuals as the function of the MEPS predicted wind speed. The red line is the mean of the residuals and highlights the regime shift of the residuals with the increase of the wind speed.

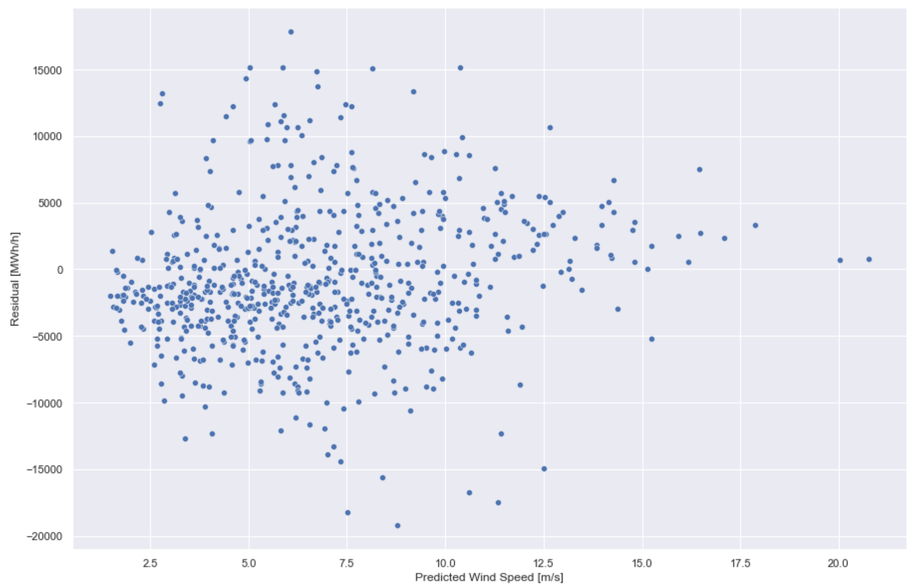

Figure 19: Scatterplot of forecast residuals for forecast horizon of 1 hour.

5.3 Creation and Analysis of the Forecast Distributions

The previous section identified the Random Forest model trained directly with weather forecasts as the best model for power production modelling. Now forecast intervals are created for this model. Three different methods are presented. For uncertainty propagation method Random Forest with measurement data was used instead for reasons presented in more detail in its corresponding subsection.

5.3.1 Forecast Intervals from Validation Set Residuals

The data set was split into training set, validation set and test set with 70-15-15 split as described in Chapter 3.1. The model was first trained with training data and predictions were made for validation set. Model was then retrained with both train and validation sets and forecasts were made for the test set. It was assumed that the error distribution of the test set would be similar to the residual distribution of the validation set. In essence, it was assumed that the errors of future forecasts would be similar to errors of past forecasts, and past residual distribution was used for creating forecast distributions.

The residual distribution of validation set was furthermore conditioned by finding a fixed number of most similar past weather situations with the k-NN method similar to described in Chapter 4.1. This way only similar weather scenarios of the past would be taken into account avoiding scenarios where, for example, weather forecast indicates wind speed of 10 m/s but error distribution includes completely windless weather conditions. For this, the weather variable space was normalized and k-NN was performed. Varying the number of neighbours did not affect the results, so a reasonable number of 30 neighbours was chosen. The search was performed with the weighted Euclidean distance metric, where the weights of the variable importances were taken from the model trained with only training data set and which were very close to the ones presented in Table 8. Finally, the ensemble members were aggregated and forecast distribution was created from all unique neighbours found.

Next forecast intervals were taken from the distribution with two different approaches. The first approach was to assume a Gaussian error distribution, estimate the standard deviation and take upper and lower limits with ± 2 standard deviations away from the mean forecast. In the second approach the 5th and 95th percentiles were taken directly.

The fact that the interval was constructed of residuals neglects the boundary conditions for minimum and maximum production. For this reason, the intervals where clipped to adhere to physical limitations of the wind farm i.e. production should stay between 0 and 34 500 kWh/h.

For the standard deviation method, intervals were created for a forecast horizon of one hour, and for a 95 % interval, a mean interval width of 18309 kWh/h with a corresponding PICP-score of 88.2 % was attained. When the 5th and 95th percentiles were taken instead, the mean interval width of 14495 kWh/h and corresponding PICP-score of 98.8 % was achieved. For single horizons the sample sizes were small and this affects the results, but even from this we can conclude that the Gaussian

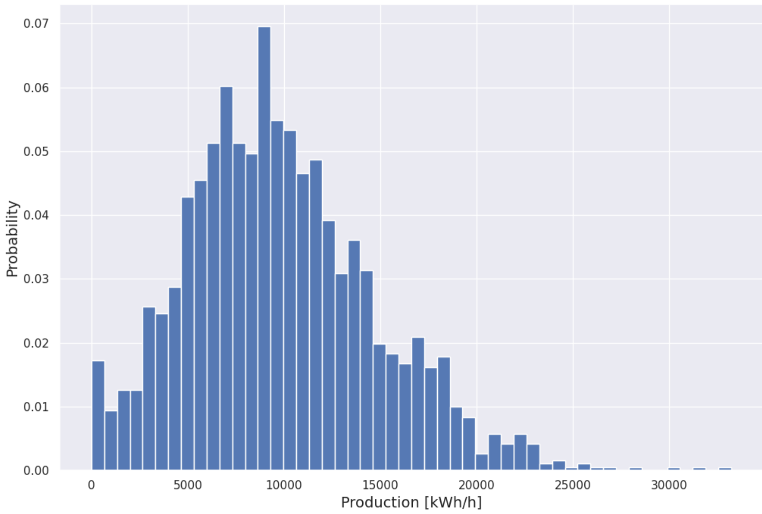

Figure 20: Histogram of forecast interval widths for the test set and forecast horizons of 0-3 hours. the designated coverage percentage is 90 %.

error assumption is not good, as it yields wider intervals with lower PICP-score than custom percentiles. The non-Gaussian nature is also seen in the Figure 17. For these reasons modelling the forecast distribution as Gaussian was not researched further and instead direct percentile approach was used.

Table 12: Aggregate statistics for the test set intervals created from validation set's residuals for concatenated time horizons of 0-3 hours.

Designated Coverage (%)	Mean Width (kWh/h)	Median Width (kWh/h)	PICP (%)
80 %	9906	9389	83.8
90 %	12173	11699	89.7

The result of 98.8 % PICP score was achieved for one hour time horizon, which had a small sample size, and cannot be considered a concluding result. This was addressed by concatenating horizons to larger units to give more robust estimates. For testing, horizons of 0-3 hours were concatenated both for the validation set's residuals and for the test set. The histogram of interval widths is shown in Figure 20.

The test statistics for designated coverage percentages of both 80 % and 90 % are reported in Table 12. The PICP scores were close to the designated coverage percentages. This is a good result, indicating that the assumption of similar past and future error distributions holds and the method captures the properties of error distribution as intended.

The mean width for 90 % intervals was 12 173 kWh/h or 35 % of the wind farm's rated power for the entire test data set. When conditioned and binned with respect to predicted wind speed, the mean interval widths are shown in Figure 21. The interval widths start to decrease when predicted wind speed reaches the turbine's rated wind speed of 10 m/s. For wind speeds of over 20 m/s the results are unreliable as there were very few observations available.

The PICP score as a function of predicted wind speed is shown in Figure 22. The PICP score decreases as wind speed increases. For zones I and II is this expected behaviour. The decrease in zone III seems counterintuitive, but is explained by the decreasing interval widths of Figure 21. The model performs worst in zone II: PICP-score is low, yet the interval width is high.

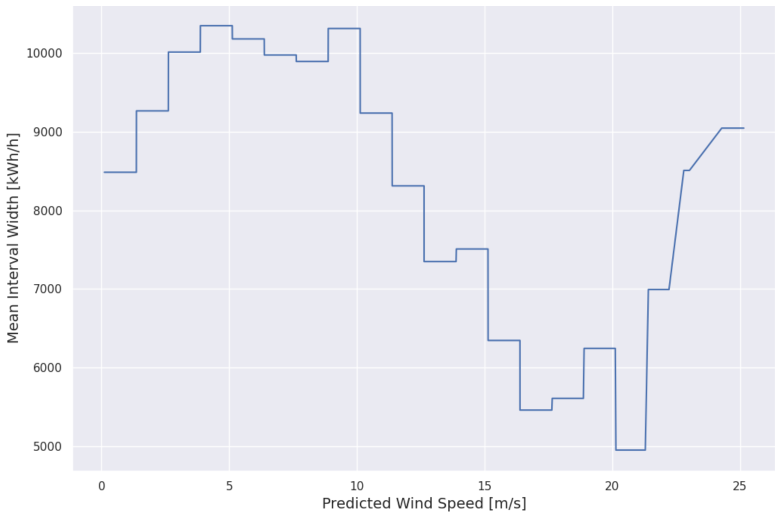

Figure 21: Mean interval width as a function of predicted wind speed. In zone II the interval width is higher than in other regions. For speeds above 20 m/s the estimates are unreliable as only very few observations were present.

The method does not perform uniformly well, and a closer examination of the data revealed that there are some intervals, where the width is zero. These are roughly 4.5 % of values after individual ensemble members have been aggregated for each time stamp. Also some intervals did not include the predicted value itself, indicating errors in the numerical calculation step. These results could not be corrected at the time of writing. In the real case the obtained PICP scores could be higher than reported here.

An illustrative time series plot is in Figure 23. It shows shows how uncertainty information could be communicated to the decision-maker. In the plot the realized power production is in red, and the blue fill describes the 90 % confidence region.

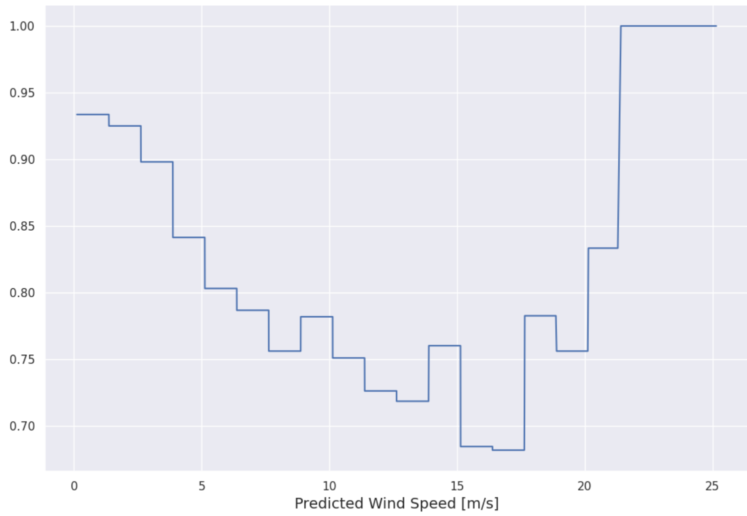

Figure 22: Percentage of realizations falling inside forecast interval as a function of wind speed. The score is mostly above 70 %. For speeds above 20 m/s the estimates are unreliable, as only very few observations were present.

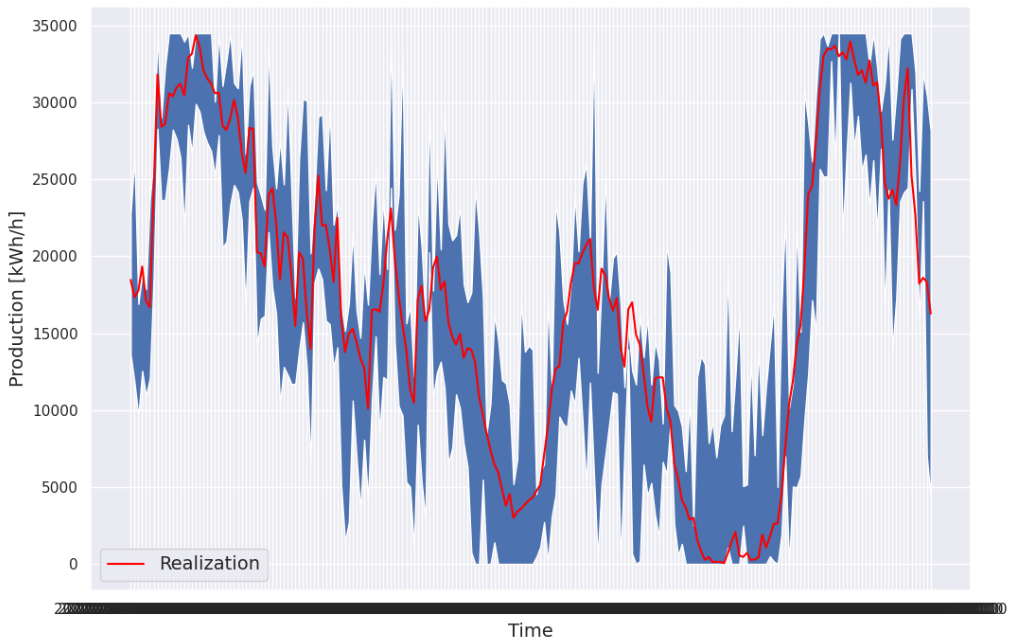

Figure 23: A time series plot for a period of two weeks. Red line is the realized wind power production and blue fill is the 90 % forecast interval.

5.3.2 Forecast Intervals from Individual Forecasts of Each Decision Tree

A second approach to create forecast intervals was to use the forecasts of each individual tree of Random Forest. Instead of simply aggregating the forecasts, the output of each tree was ordered and the 5th and 95th percentile were taken. The benefit of this approach was that it was easily implemented. On the other hand, this method requires the chosen model to be ensemble model i.e. a model which aggregates the estimate over many individual regressors.

Again forecast horizon of one hour was used to create statistics for the test set. The interval widths are presented in Figure 24. The corresponding PICP score for 90 % interval is 91.6 %. The mean interval width is 20960 kWh/h. According to PICP score the distribution describes the variability of production quite decently. The interval width of nearly 21 MWh/h is, however, quite large and as it is nearly two-thirds of the capacity, the average production being around one-third. The intervals are too wide to be of much real use in decision-making. Compared to intervals created from validation set residuals this method yields worse results.

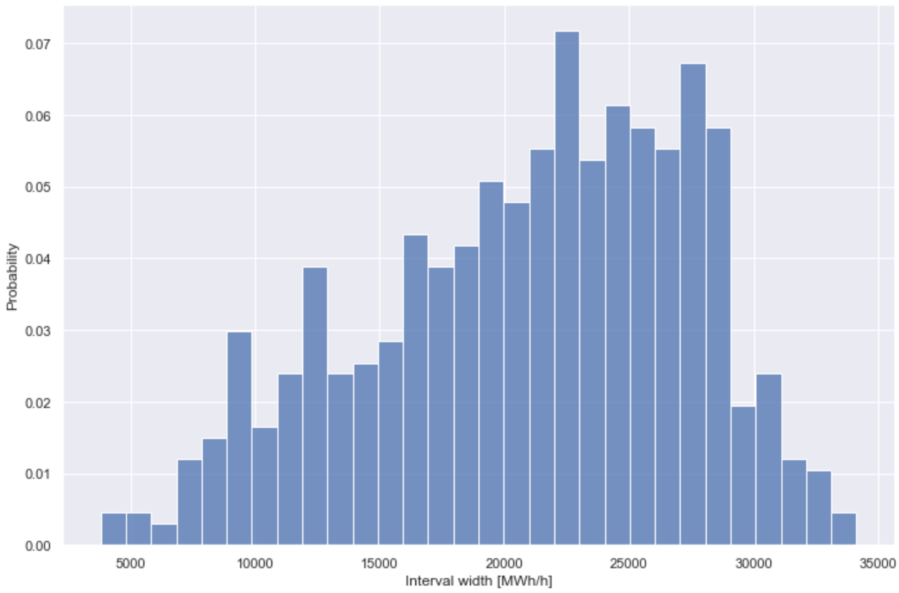

Figure 24: Histogram of forecast interval widths created from forecasts of individual trees.

5.3.3 Weather Forecast Uncertainty propagation

The third attempt to construct forecast distributions was the so-called uncertainty propagation as described in [31]. Basicly the idea is that if there is sufficient information about the distribution of the independent variable X and its relation to the dependent variable y, that is $f(X) = Y$, it is possible to estimate the distribution of Y by simulating multiple instances from the distribution of X and propagating these through the function $f(X)$. This is a simple way to utilize Monte Carlo simulation for forecast distribution generation and is easily implemented.

In our case, the set of independent variables in Random Forest model trained with MEPS data has 19 dimensions. The sampling from multidimensional distribution is a complicated statistical problem in itself. The Random Forest model trained with measurement data has only 10 variables and of these variables the wind speed was vastly dominant with 76 % regression importance score. For these reasons, the Random Forest model trained with measurement data was chosen. This made replacing the multidimensional distribution with just one-dimensional wind speed distribution a viable simplifying operation.

The distribution of wind speed was created for each forecast first by conditioning the training weather data set for the forecast horizon and taking forecasts where the predicted wind speed was within ± 0.1 m/s. For this training weather data the error to actual wind measurements at site were calculated. Other weather variables were assumed to be constant. Finally, this distribution was sampled with replacement a combined 100 times and these simulated error quantities were added to the weather forecast before being run through the power forecast model. Finally ensembles were aggregated to produce results for each time point.

Because forecast horizon of one hour is consistently missing some weather variables, forecast horizon of three hours was used instead. On some occasions no suitable past observations were found and the sampling distribution was empty. These cases were simply omitted from the data for test purposes.

A histogram of the resulting forecast interval widths is in Figure 25. The mean interval width for 90 % forecast intervals was 22641 kWh/h. The PICP score was 75.5 %. When the minimum and maximum were values were studied instead, the PICP score increased to 79.2 %. Considering the quantiles taken and the large intervals widths the PICP score is low and does not describe the forecast variance accurately. Thus we conclude that the uncertainty propagation method, though scientifically sound, does not suit the task of creating forecast intervals for wind power forecasting.

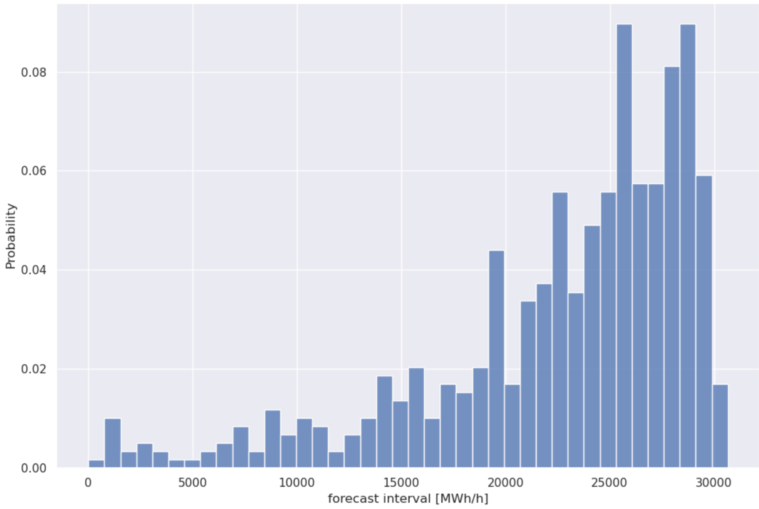

Figure 25: Histogram of forecast interval widths for uncertainty propagation with forecast horizon of three hours.

6 Discussion

In this thesis, wind power production was first modelled and then methods of creating forecast intervals were examined. As for the latter, serious shortcomings in industry application were noted in the introduction, and this thesis has aimed to address this problem. Unfortunately, however, due to the proprietary nature of wind power forecast models used in industry, it was necessary to create a full wind power production model before being able to study the production uncertainty itself. This increased the workload significantly, limiting the efforts in uncertainty modelling. It is easy to conclude that the proprietary nature of many forecast models and the financial interests associated to them are serious hindrances in the way of collaboration and understanding wind power uncertainty in the energy sector at large.

The best model was Random Forest model trained directly with MEPS weather forecasts and forecast intervals constructed from the residuals of former forecasts. The model achieved RMSE of 5308 kWh/h and MAE of 4062 kWh/h for mean forecasting and one hour forecast horizon. The mean width for 90 % forecast intervals was 12173 kWh/h with 89.7 % PICP score.

The main source of uncertainty stems from inaccuracies in weather forecasts. For Random Forest model trained with measurement data the weather forecast errors were vastly dominant source of uncertainty, totalling 86 % of errors.

The problem of inaccurate weather forecasts is commonly tackled by using ensemble forecasts. Also machine learning models can be used to post-process weather forecasts as was done here. These methods, however, are far from conclusive and weather uncertainty remains a large, unresolved problem and the accuracy of weather forecasts being a limiting factor for wind power forecasting.

Figures 1 and 18 present the results visually. The wind turbine properties dampen the effect of weather fluctuations on regions other than zone II, which is the most problematic domain. As power output rises approximately cubically in zone II, small errors in wind speed forecasts can introduce significant errors in the model, as the residuals show in Figure 18. This result is in line with our initial expectations and previous literature [8].

Besides the forecast intervals, a functional mean forecasting model was attained as a side-product. The model turned out to be have merits of its own in the mean forecasting even separately from the uncertainty modelling.

For forecast intervals only the method of creating intervals from past residuals provided decent results. Whether the intervals and PICP scores are good enough to be of use for decision-making depends on the electricity market situation and the decision-makers market strategies. Thus a definite assessment of the results cannot unfortunately be given here, and the results should be analyzed from the viewpoint of practical market operations. The model has not been tested in real-time production, but the results are encouraging and the value of the results should be studied from the viewpoint of wind power operator and/or electricity market participant.

Next step for a practical decision-making tool would be to produce a more detailed analysis of the conditions of the performance. From a theoretical viewpoint, this might not be necessary, but for practical purposes it is important. One approach would be to run a larger test set with varying weather conditions through the model, record the performance after which they would be used as samples for future performance - in a similar way validation set's residuals were used as test set's error distribution. The underlying assumption is that the model does not perform uniformly well in all weather conditions, and likely there are some weather scenarios where the model's average performance or variability is significantly lower/higher. Already this thesis has presented similar analysis based on predicted wind speed, but so far ignoring other weather variables. A decision tree or a clustering method could be fitted to recognize these scenarios and serve as a lookup table for future predictions. After this, the proposed model would be able to predict the wind power mean, its confidence intervals and give self-reflection based on past forecast accuracy. This would yield actionable insights to wind power uncertainty and support decision-making processes.

Many ways to improve the model performance can be suggested, as indicated by the vast literature about wind power forecasting. Some possible improvements are:

- Time series structure was intentionally left out of the model to simplify the modelling. Introducing this increases the amount of information, presumably producing better performance
- Different weather data preprocessing procedures could be tested. In this thesis all variables were normalized to have the same scale, but no dimension reduction was done. For example Principal Component Analysis could be used to remove the variables with little practical value to improve data matching and model performance
- Wind speed extrapolation from 10 meter altitude to nacelle height was done based on deterministic formula. Wind speed differences in different altitudes are not trivial and the method can be assumed to introduce errors into the model. However, with the absence of better method this was used, though more refined estimates could improve the model performance
- MEPS data lacked post-processing. Most importantly, the use of weather measurements in the form of assimilated weather data was not utilized in this thesis, although it can be presumed to improve the forecast performance, especially on shorter forecast horizons
- The wind gust information was not extensively utilized. The acquired measurement data was a 5-minute mean, which smoothed out information about sudden gusts. This increased information could slightly improve the performance
- The importance of cyclic time variables was not extensively tested. Whether removing this information from weather scenario matching would improve the results is left for further research. They were also weighted using linear correlation coefficient, which might be suboptimal
- The effects of icing were deliberately left out of the model, and incorporating them could explain some of the large overpredictions in winter time

- Weather matching algorithm was k-NN, and this could possibly be improved. Other matching algorithms possibly coupled with non-linear transformations could be tried out.

This thesis has emphasized the importance and need for uncertainty modelling in wind power forecasts, which, although recognized by most energy sector participants, have not received the necessary attention in the energy industry. The significance of wind power uncertainty modelling will only increase in the future and more research both into the uncertainty modelling and its practical implementation are needed.

7 Summary

In this thesis wind power forecast uncertainty was studied. The keen interest towards more effective wind power uncertainty modelling shared by the interviewees and the reasons behind this were uncovered in the introduction. It was concluded, that with the rising share of intermittent wind power, there is a clear need for a tool for wind power uncertainty management which could provide actionable insights into the decision-making processes of energy producers and electricity market participants.

In Modelling and Forecasting Wind Power Generation the basics of wind power technology and the importance of weather forecasting as an essential component of wind power forecasting were introduced to the reader, as well as the scientific background behind weather forecasting. Furthermore the current academic research to wind power modelling was summarized and the process of wind power modelling was broken down. Different classes of models were categorized, with the notion of complex deep learning models being the current focus of research.

The wind farm SCADA data and its FMI supplements along with the MEPS weather forecasts were then introduced in detail. The necessary transformations and other preprocessing procedures to the data were explained. Also the metrics for evaluating the models were shown. In Models Chapter, the three different models - k-NN, Random Forest regressor and the Equation Model - were explained and their respective pros and cons discussed.

The three different models were evaluated on their ability to predict the mean of wind power production. These models were trained with measurement data, and the error sources were divided into external weather forecast and model errors, with the weather forecast errors being the vastly dominant source of uncertainty, totalling 86 % of errors for Random Forest model. The Random Forest model performed best of the three, and was further studied with it being trained directly on weather forecasts. This allowed the algorithms to directly learn the relationship between uncertain weather forecasts and actual power output, yielding significantly better results than earlier.

Forecast intervals were created for Random Forest model trained on weather forecasts. Uncertainty propagation method, creating a distribution from each individual regressor of the ensemble model and using validation set's residuals as test set's errors were studied. The last method provided the best results with mean interval width of 12173 kWh/h and PICP score of 89.7 %.

Lastly, the significance and applicability of the findings were analysed, concluding that the largest source of errors are weather forecast errors. This limits the possibility for improvement significantly. Modelling the variability of production is closely tied to the behaviour of the turbine's power curve: in zones I, III and IV usefully accurate results can be obtained, but zone II with its rising power output is difficult to model accurately. However, even incomplete information is useful for decision-makers in practice, and the practical applicability of the results should be studied. Limitations and possible future improvements were also discussed.

Quantification of wind power forecast uncertainty is an important subject which has often been at least partly neglected in practice. This thesis showed that there is need for more comprehensive modelling of wind power uncertainty, shared by all electricity market participants, as wind power uncertainty quantification would provide information for better decision-making. With the share of wind power in energy systems rising, uncertainty modelling will only gain in importance. There is room for both academic research and practical applications of uncertainty modelling in the future.

References

[1] Suomen Tuulivoimayhdistys. *Talvella tuulee eniten - Suomen Tuulivoimayhdistys.* [Online; accessed May 2022]. URL: https://tuulivoimayhdistys.fi/tietoa-tuulivoimasta-2.

[2] Petr Spodniak, Kimmo Ollikka, and Samuli Honkapuro. "The impact of wind power and electricity demand on the relevance of different short-term electricity markets: The Nordic case". In: *Applied Energy* 283 (2021), p. 116063.

[3] Fingrid. *Varttitase eli 15 minuutin taseselvitysjakso.* [Online; accessed July 2022]. Apr. 2018. URL: https://www.fingrid.fi/sahkomarkkinat/markkinoiden-yhtenaisyys/pohjoismainen-tasehallinta/varttitase/.

[4] Corinna Möhrlen et al. "How do Humans decide under Wind Power Forecast Uncertainty — an IEA Wind Task 36 Probabilistic Forecast Games and Experiments initiative". In: *Journal of Physics: Conference Series* 2151 (Jan. 2022), p. 012014. DOI: 10.1088/1742-6596/2151/1/012014.

[5] Pierre Pinson and George Kariniotakis. "Conditional Prediction Intervals of Wind Power Generation". In: *IEEE Transactions on Power Systems* 25.4 (2010), pp. 1845–1856.

[6] Trevor M Letcher. *Wind Power Fundamentals.* 1st edition. London, [England]: Academic Press, 2017. ISBN: 9780128094518.

[7] Miguel Á Rodrıguez-López, Emilio Cerdá, and Pablo del Rio. "Modeling wind-turbine power curves: Effects of environmental temperature on wind energy generation". In: *Energies* 13.18 (2020), p. 4941.

[8] Georges Kariniotakis. *Renewable Energy Forecasting: From Models to Applications.* Woodhead Publishing, 2017.

[9] Finnish Meteorological Institute. en. [Online; accessed June 2022]. URL: https://en.ilmatieteenlaitos.fi/numerical-weather-prediction.

[10] Inci Okumus and Ali Dinler. "Current status of wind energy forecasting and a hybrid method for hourly predictions". In: *Energy Conversion and Management* 123 (2016), pp. 362–371.

[11] M Lydia et al. "A comprehensive review on wind turbine power curve modeling techniques". In: *Renewable and Sustainable Energy Reviews* 30 (2014), pp. 452–460.

[12] Yan Jie et al. "Uncertainty estimation for wind energy conversion by probabilistic wind turbine power curve modelling". In: *Applied Energy* 239 (2019), pp. 1356–1370. ISSN: 0306-2619.

[13] Hannele Holttinen, Jari J. Miettinen, and Samuli Sillanpää. *Wind Power Forecasting Accuracy and Uncertainty in Finland.* VTT Technology 95. Finland: VTT Technical Research Centre of Finland, 2013.

[14] Yan Jie et al. “Uncovering wind power forecasting uncertainty sources and their propagation through the whole modelling chain”. In: *Renewable and Sustainable Energy Reviews* 165 (2022), p. 112519. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2022.112519. URL: https://www.sciencedirect.com/science/article/pii/S1364032122004221.

[15] Yan Jie et al. “Reviews on uncertainty analysis of wind power forecasting”. In: *Renewable and Sustainable Energy Reviews* 52 (2015), pp. 1322–1330. DOI: https://doi.org/10.1016/j.rser.2015.07.197. URL: https://www.sciencedirect.com/science/article/pii/S1364032115008552.

[16] Yao Zhang, Jianxue Wang, and Xifan Wang. “Review on probabilistic forecasting of wind power generation”. In: *Renewable and Sustainable Energy Reviews* 32 (2014), pp. 255–270.

[17] Abbas Khosravi et al. “Lower upper bound estimation method for construction of neural network-based prediction intervals”. In: *IEEE Transactions on Neural Networks* 22.3 (2010), pp. 337–346.

[18] Hao Quan, Dipti Srinivasan, and Abbas Khosravi. “Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals”. In: *IEEE Transactions on Neural Networks and Learning Systems* 25.2 (2014), pp. 303–315. DOI: 10.1109/TNNLS.2013.2276053.

[19] Abdollah Kavousi-Fard, Abbas Khosravi, and Saeid Nahavandi. “A New Fuzzy-Based Combined Prediction Interval for Wind Power Forecasting”. In: *IEEE Transactions on Power Systems* 31.1 (2016), pp. 18–26. DOI: 10.1109/TPWRS.2015.2393880.

[20] Yanxia Shen, Xu Wang, and Jie Chen. “Wind Power Forecasting Using Multi-Objective Evolutionary Algorithms for Wavelet Neural Network-Optimized Prediction Intervals”. In: *Applied Sciences* 8.2 (2018). ISSN: 2076-3417. DOI: 10.3390/app8020185. URL: https://www.mdpi.com/2076-3417/8/2/185.

[21] Malte Müller et al. “AROME-MetCoOp: A Nordic Convective-Scale Operational Weather Prediction Model”. In: *Weather and Forecasting* 32.2 (2017), pp. 609–627. DOI: 10.1175/WAF-D-16-0099.1. URL: https://journals.ametsoc.org/view/journals/wefo/32/2/waf-d-16-0099_1.xml.

[22] Inger-Lise Frogner et al. “Convection-permitting ensembles: Challenges related to their design and use”. In: *Quarterly Journal of the Royal Meteorological Society* 145.S1 (2019), pp. 90–106. DOI: https://doi.org/10.1002/qj.3525. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3525. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3525.

[23] *Vertical Extrapolation of Wind Speed*. en. [Online; accessed July 2022]. URL: https://windroseexcel.com/guides/vertical-extrapolation-of-wind-speed/.

[24] Alex Tolonen. "Forecasting short-term valuation of regional hydropower generation with machine learning". Master's Thesis. Aalto University. School of Electrical Engineering, 2022, pp. 68+4. URL: http://urn.fi/URN:NBN:fi:aalto-202203272587.

[25] Evelyn Fix and Joseph Lawson Hodges. "Discriminatory analysis. Nonparametric discrimination: Consistency properties". In: *International Statistical Review/Revue Internationale de Statistique* 57.3 (1951), pp. 238–247.

[26] Fabian Pedregosa et al. "Scikit-learn: Machine learning in Python". In: *Journal of Machine Learning Research* 12.Oct (2011), pp. 2825–2830.

[27] Tin Kam Ho. "Random decision forests". In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. 1995, pp. 278–282. DOI: 10.1109/ICDAR.1995.598994.

[28] Karoliina Hämäläinen et al. "Meteorological Solutions to Support Wind Energy Production in Finland." PhD thesis. 2021. ISBN: 978-952-336-144-7.

[29] Lauri Kuusisto. "Tuulivoimalan roottorin suuntausvirheen vaikutus liikevaihtoon sekä virheen havaitseminen Lidar-mittauksella". Master's Thesis. Lappeenranta-Lahti University of Technology, 2021. URL: https://urn.fi/URN:NBN:fi-fe202101283078.

[30] Danish Wind Industry Association. [Online; accessed June 2022]. URL: http://xn--drmstrre-64ad.dk/wp-content/wind/miller/windpower%5C%20web/en/tour/wres/wake.htm.

[31] Shuai Guo. *Using Monte Carlo to quantify the model prediction error.* [Online; accessed September 2022]. URL: https://towardsdatascience.com/how-to-quantify-the-prediction-error-made-by-my-model-db4705910173.