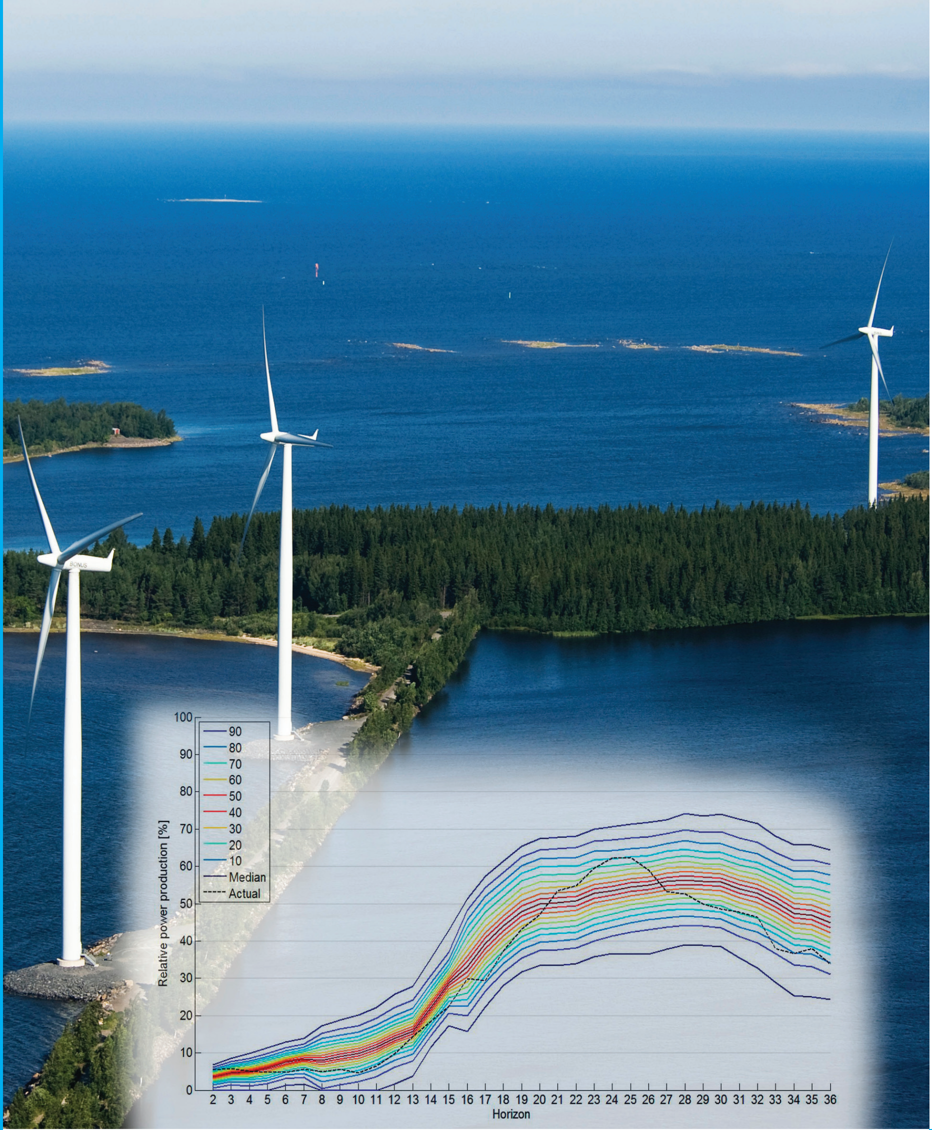
SCIENCE • TECHNOLOGY • RESEARCH HIGHLIGHTS • VISIONS

95



Wind power forecasting accuracy and uncertainty in Finland

Hannele Holttinen | Jari Miettinen | Samuli Sillanpää

VTT

VTT TECHNOLOGY 95

Wind power forecasting accuracy and uncertainty in Finland

Hannele Holttinen, Jari Miettinen & Samuli Sillanpää

ISBN 978-951-38-7985-3 (URL: http://www.vtt.fi/publications/index.jsp)

VTT Technology 95

ISSN-L 2242-1211
ISSN 2242-122X (Online)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT
PL 1000 (Tekniikantie 4 A, Espoo)
02044 VTT
Puh. 020 722 111, faksi 020 722 7001

VTT
PB 1000 (Teknikvägen 4 A, Esbo)
FI-02044 VTT
Tfn +358 20 722 111, telefax +358 20 722 7001

VTT Technical Research Centre of Finland
P.O. Box 1000 (Tekniikantie 4 A, Espoo)
FI-02044 VTT, Finland
Tel. +358 20 722 111, fax +358 20 722 7001

Wind power forecasting accuracy and uncertainty in Finland

Hannele Holttinen, Jari Miettinen & Samuli Sillanpää. Espoo 2013. VTT Technology 95. 60 p. + app. 8 p.

Abstract

Wind power cannot be dispatched so the production levels need to be forecasted for electricity market trading. Lower prediction errors mean lower regulation balancing costs, since relatively less energy needs to go through balance settlement. From the power system operator point of view, wind power forecast errors will impact the system net imbalances when the share of wind power increases, and more accurate forecasts mean less regulating capacity will be activated from the real time Regulating Power Market.

In this publication short term forecasting of wind power is studied mainly from a wind power producer point of view. The forecast errors and imbalance costs from the day-ahead Nordic electricity markets are calculated based on real data from distributed wind power plants. Improvements to forecasting accuracy are presented using several wind forecast providers, and measures for uncertainty of the forecast are presented.

Aggregation of sites lowers relative share of prediction errors considerably, up to 60%. The balancing costs were also reduced up to 60%, from 3 €/MWh for one site to 1–1.4 €/MWh to aggregate 24 sites. Pooling wind power production for balance settlement will be very beneficial, and larger producers who can have sites from larger geographical area will benefit in lower imbalance costs. The aggregation benefits were already significant for smaller areas, resulting in 30–40% decrease in forecast errors and 13–36% decrease in unit balancing costs, depending on the year. The resulting costs are strongly dependent on Regulating Market prices that determine the prices for the imbalances. Similar level of forecast errors resulted in 40% higher imbalance costs for 2012 compared with 2011.

Combining wind forecasts from different Numerical Weather Prediction providers was studied with different combination methods for 6 sites. Averaging different providers' forecasts will lower the forecast errors by 6% for day-ahead purposes. When combining forecasts for short horizons like the following hour, more advanced combining techniques than simple average, such as Kalmar filtering or recursive least squares provided better results.

Two different uncertainty quantification methods, based on empirical cumulative density function and kernel densities, were analysed for 3 sites. Aggregation of wind power production will not only decrease relative prediction errors, but also decreases the variation and uncertainty of prediction errors.

Keywords wind power, wind energy, forecasting, uncertainty, electricity market, imbalance costs

Preface

This publication is part of SGEM research programme work package WP5 managing variable renewables, with part financing from Nordic TFI programme project Icewind. The report is based on real data on wind power production and wind speed forecasts for more than 20 sites in Finland, for 3 years. In addition to analyses on the forecast errors and imbalance costs, there is a summary of work on improving the forecast accuracy by combining several meteorological forecasts, published as a thesis for Helsinki University. The publication also presents work on possibilities to provide probability intervals to simple point forecasts.

One of the authors, Samuli Sillanpää, has moved to Helsinki University before the publishing of this report.

Contents

List of symbols

α	probability quantile
e	prediction error
h	width
I	Indicator variable
K	Kernel function
N	Numeber of samples
p	power production
t	time
x	variable
y	variable
^	forecast

List of acronyms

AVG	Average
CDF	Cumulative Distribution Function
COM	Composite post prosessing
ECMWF	European Centre for Medium-range Weather Forecasts
FMI	Finnish Meteorological Institute
HIRLAM	High Resolution Limited Area Model
KDE	Kernel Density Estimation
NWP	Numerical Weather Prediction

OPT	Optimal combination
OUT	Outperformance
REG-NO	Regression
RLS	Recursive Least Square
SHARE	Fixed share
SMHI	Sveriges Meteorologiska och Hydrologiska Institute

1. Introduction

Wind power production and short term forecasting applications are still small in Finland. In 2012 wind power penetration was only 0.6% from the total electricity consumption [1]. However, in EU goals have been set to increase the share of renewable sources in EU to 20% before 2020. In order to reach the target, each member country has been given its own target. In Finland the amount of renewable energy sources from the energy consumption should be 38% before 2020 [2]. As part of this target, increasing the amount of wind power to 2500 MW (6 TWh, 6% of electricity consumption) is planned.

In Finland the guaranteed price subsidy is constructed in a way that wind power producers are responsible for their own balancing costs. Imbalances occurring due to forecast errors are settled in the balance settlement and will incur costs depending on the balancing price for the hour. Thus, forecast errors turn into costs and therefore market participant should strive for accuracy in the forecasts for hourly power production. This is challenging for wind power since the source is variable and non-dispatchable. Short term forecasting tools are essential in this process.

Quantification of wind power production uncertainty becomes more crucial as the wind power penetration increases. System operators (TSOs) would be interested to see how wind power production varies on different look-ahead hours, and what the probability of certain variation is. Wind power can be thought as a stochastic source of generation. Quantification of wind power production's uncertainty would help TSOs to understand predictability of wind power and it would also provide a tool to diminish the challenges that the variable nature of wind induces. Quantification of uncertainty would also help in assuring adequate balancing is available from the balancing market, called Regulating Power Market in the Nordic countries.

In this publication Section 2 explains the basics of wind power forecasting especially in day-ahead trading in the Nordic market. The data used in this publication is presented in Section 3. In Section 3.3 prediction errors in Finland 2010, 2011 and 2012 are analysed. The weight is given to the level of prediction errors in different geographical areas and the balancing costs induced for a producer. In Section 5 a method to combine Numerical Weather Prediction (NWP) models is presented in order to improve the accuracy of wind power predictions. In Section 6 methods to derive probability intervals to forecasts and effects of how aggregation of wind power production affects probability intervals are presented.

2. Short term forecasting of wind power

Wind power production forecasts are created by forecasting models. The purpose of a forecast model is to turn inputs, which are supplied to the forecast model, to *k*-steps ahead wind power forecast. Typically used inputs are lagged values of wind power production and wind forecast with other weather related variables by Numerical Weather Prediction (NWP) model. Also some temporal information such as what is the current time of day or season of a year are used. Model inputs are defined according to model needs. Usually, forecast models are tuned into a certain temporal horizon. For time span of minutes, it is highly unlikely that a model for making forecasts days ahead will perform well.

2.1 Different types of forecasting models

Forecast models can be distinguished by: forecasting interval and type of a forecast model. The type of a forecast model can be *physical*, *statistical* or a combination of these two. Temporal forecasting horizons can be separated mainly into *very short term*, *short term* and *medium term* forecasting.

Very short term forecasting is used when forecasting from minutes up to some hours ahead. This kind of information is especially needed for controlling wind turbines. The temporal horizon of short term forecasting is up to 72 hours ahead. In this temporal horizon forecast results are used by the TSOs and the energy traders. Wind power forecasting can be extended to cover following week, which would give additional help for scheduling maintenance plans. This forecast horizon can be called as a medium term. In this publication the weight is given to forecasts in day-ahead electricity trade and thus the temporal scale of this trade will be short term forecasting by definition [3] [4].

Forecast models of very short term forecasting differ from models of other temporal scales. It is quite convenient to use purely statistical models on very short term forecasting, since the models based on global or local NWP models cannot be used on such a short time scale. Also the variability wind power can be quite well explained by persistent and stochastic nature of wind power up to one hour ahead. However, in the forecast horizon above 6 hours, persistent nature of wind power is already low and pure statistical model, which only relies to the past values of power production, or history data of other explanatory variables, starts to perform poorly. Therefore, the physical world must be included to forecast inputs in the form of meteorological forecasts. These meteorological forecasts can be attained by running NWP models (ECMWF, HIRLAM etc.), which are highly complex and computationally heavy.

Forecast models can be grouped by whether a statistical or physical approach has been used. These models can be further divided into various subcategories.

A physical model tries to refine the wind field in a site by using information from the area of interest, such as local orography, roughness and obstacles. The assumptions and theories are based on the physical behaviour of wind in the atmospheric boundary layer. Thus, a physical model continues refining the wind field with a finer resolution than the NWP forecast can provide. Usually the grid size of NWP forecast can be order of magnitude from couple of square kilometres up to tens of square kilometres. Physical models may require a lot of computational capacity and this can set a limit for a grid size. Since a physical model tries to create wind field in a wind power plant; there must be a wind to power transformation, which needs to be modelled with wind turbine or wind farm power curve. In principle, a physical prediction model would not need any SCADA system, which provides past power value to the prediction model. This means that the model is ready to be used already before some months of data exist from the site. Usually Model Output Statistics (MOS) is used to correct the forecast in case there is a systematic error in the forecast.

A statistical model does not try to reform the wind field inside a NWP grid. However, a statistical model tries to find dependencies between produced wind power (dependent variable) and explanatory variables such as wind and wind direction forecast from NWP. Typically the dependencies are found by minimizing a loss function, which is a function of actual power production and forecasted power values from a statistical model. Statistical models can be also defined by the used modelling techniques, for instance grey-box models are using theoretical knowledge together with measurements. For black-box models such as Neural networks no theoretical knowledge of the process is needed [24].

In practice forecast model's parameters are tuned to correspond past values of power production by ending up with a model, which can make forecasts k-steps ahead, with sufficient inputs. This model type is called as hybrid forecast model, since it relies both on the stochastic and the physical nature of wind power. The main challenge in traditional statistical modelling is forming the structural part of statistical model, including all necessary parameters from the forecasting capability point of view, and excluding parameters that have little value for forecasting accuracy.

The forecast model, which is used in this study, is a statistical model, which takes as an input NWP data and some other explanatory variables.

2.2 Forecast errors

In this section properties of forecast errors will be discussed. The nature of wind is variable, hard to predict accurately, and therefore forecast errors are always present.

The definition of forecast error is the deviation between forecasted and realisation of wind power production and it can be interpreted with the following equation

$$e_t = \hat{p}_t - p_t \tag{1}$$

where p_t is the measured power at time t and $\hat{p}_t$ is corresponding forecasted value.

In wind power forecasting reasons behind forecast error can be separated into two main sources based on their causes: level and phase errors. Level errors are usually caused by biased, systematically erroneous prediction, whereas phase errors are caused by forecasting level of production correctly but failing on predicting timing of changes (Figure 1).

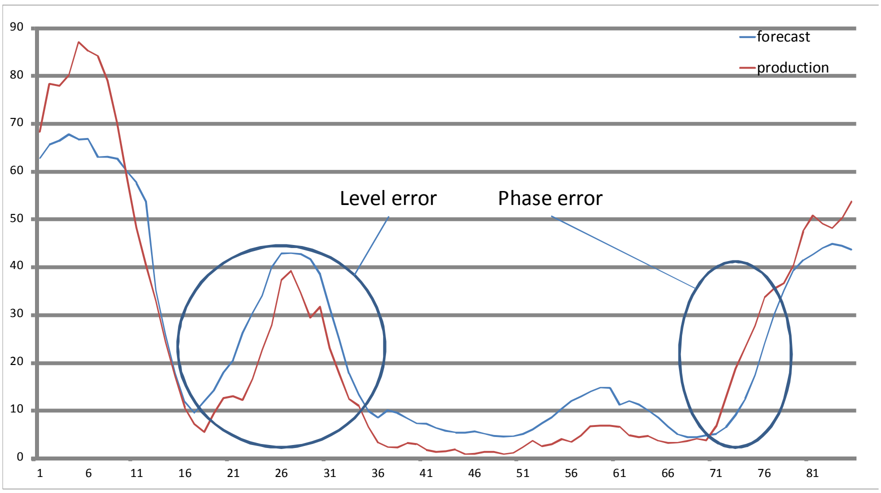


Figure 1. Forecast errors due to predicting the level of production wrong and due to wrongly predicting the timing of wind changes (phase error).

The reason behind level errors might be badly configured forecast model or the source of error might be random. The phase errors are more problematic even when the forecast model accuracy is good. Phase errors are usually caused by larger changes in wind speed i.e. timing of weather front, which could lead to substantial forecast errors. Thus, forecast user must be careful when making decisions that are based on forecasted ramp event since the error of timing can lead to large forecast errors, even if the level is forecasted right.

Usually it is not useful to look at a prediction error at a single hour, but instead calculate the average values of prediction errors in a certain horizon, or average weighted prediction errors. The first basic performance index is called as mean error, or bias, which can written as

$$\text{bias} = \frac{1}{\text{N}}\sum_{i=1}^{N} e_i \tag{2}$$

Bias is an important metric since it shows whether the forecast model systematically under- or overestimates power production. It is usually desirable that forecast model’s bias on different look-ahead hours is zero, which means that the forecasted power values, on average fluctuates around the realised power values. The bias is usually calculated for each forecast horizon separately. It is important to notice that wind to power conversion is a non-linear process, which causes that prediction error distributions are not Gaussian.

The main performance index used to quantify the average error is Mean Absolute Error, MAE. It gives information about performance of prediction model as the positive and negative errors do not cancel each other like in the case of bias. Another similar index describing the performance of a forecasted model is Root-Mean-Square-Error (RMSE), that will give more weight on the larger errors than MAE. The errors are usually presented relative to installed capacity (normalised errors NMAE and NRMSE). Another useful way of presenting errors is relative to total energy produced by wind power, and is useful when looking at the actual amount of imbalances for the production. There are also different ways of presenting the error. A list of some other performance indices can be found in [11].

There are many factors affecting the level of prediction errors. One is that the performance of prediction systems is highly dependent on the location of wind power plant and its surroundings. The main source of uncertainty in wind power predictions lies in NWPs, which means there is only a limited amount of improvement on developing more accurate wind power forecasting models [3].

When looking at forecast errors from a wider geographical area, forecast errors have a tendency to be uncorrelated, which is caused by the spatial smoothing effect [7]. The prediction errors smooth out as the geographical dispersion of wind farms gets larger [8]. The basic idea is that as the number of sites increases, part of the errors from the individual sites will cancel each other and this smoothing effect reduces relative share of total prediction error from the aggregated maximum capacity.

2.3 Using forecasts in electricity market

Short term forecasts are used mainly in electricity markets. This is by setting bids to the electricity market that are based on the best view of participant's power production in near future. In Finland the physical electricity trade is taking place at Nordic power market Nord Pool Spot markets, which consists two separate marketplaces; day-ahead market Elspot and intraday market Elbas. Bids for the Elspot market must include volume and price information, and the bids are set for each delivery hour separately, which are in Finland 01–24, whereas in most of the Nordic countries the delivery hours are one hour lagged behind due to time difference. The bids must be placed before the Elspot-market closes at noon (1 p.m. Finnish time), and therefore there is a 12 hour gap from the closure of Elspot to the first delivery hour. Thus, a forecast tool is needed with a forecast capability at least 36 hours ahead, with an hourly time resolution. The Elspot market is an auction where the production bids are accepted on ascending order from cheapest to most expensive bid to the point where the sum of production bids meet the consumption. Usually wind power producer bids a price 0 because wind power will be generated whatever the price at the market, any price above 0 will benefit the producer.

Intraday market starts after the Elspot has ended. Elbas is a continuous market where for every sell or purchase bid a counterparty is needed. Thus, it can be seen as more traditional commodity or stock market. When making offers to the Elbas, Elspot-price of each delivery hour is public information and usually the prices at Elbas follow the Elspot prices. For wind power more recent forecasts are available from the NWP provider, and closer to the delivery hour also the current wind power production level can be used to forecast production 2–3 hours ahead. It is possible to trade power in Elbas until 30 minutes before a delivery hour. This makes it ideal from the wind power producer point of view since the bidding can be done right before the delivery hour, which decreases forecast error and also lowers balancing costs [5]. Liquidity of Elbas has been quite low, only 0.3% of the electricity consumption is traded in Elbas [6], and Elspot remains as a main market place for trading energy – more than 70% of electricity in the Nordic countries is traded at Elspot.

It is not straightforward to decide whether to correct the forecast errors at intraday market Elbas. When wind power share is still low, the forecast errors are penalised only about 50% of the time. This will only be known after the delivery hour. It was shown in [23] that the revenue for Finnish wind power producers will not necessarily increase although a market participant places bids to the Elbas-market. The balancing prices are not known at the time of Elbas-prices and therefore it is possible that a market participant is correcting imabances, which are not causing any costs in the balancing settlement. However, it was shown in [5] that for large wind penetration levels, like the case of Denmark, intra-day trading can effectively reduce balancing costs. Probably already at lower shares of wind power, correcting the larger forecast errors in the intra-day market would be cost effective for the producer, and this would also reduce the impact of wind power on the balancing markets and system imbalances.

3. Data

In this section data used in the study is presented. The analyses are based on real wind power plant hourly data and forecasts.

Wind production data for 2010–2012 is mainly provided by Finnish Energy industries (Energiateollisuus). This data has Åland wind power plants as one aggregated time series. To make the forecasting to the separate sites, production data from wind turbines located in Åland islands was provided by Allwinds for the years 2010 and 2011. For 2012 the aggregated sum wind power data for whole Åland was used. For the analyses in Chapters 5 and 6 only six and seven months of data, respectively, and less sites were used, according to how wind forecast data was available for several NWP data providers.

3.1 Numerical weather prediction data

The weather forecasts consist of the estimated wind speed and direction valid at a small time interval around each hour at the location and hub height of the wind power plant. Data from three different NWP-providers were used in this study:

- For all sites (Section 4): Foreca ETA for year 2010, Foreca ECMWF for years 2011–12
- For 3 sites (Section 5): Swedish Meteorological and Hydrological Institute (SMHI) and
- For 3 sites (Sections 5–6): Finnish Meteorological Institute FMI.

For 2010, hourly wind speed and direction prediction was provided by Foreca ETA NWP-model. In 2011 it was noticed that Foreca ECMWF NWP outperformed ETA model and the forecasts for wind were changed to ECMWF model. Thus, there is some difference in forecast accuracy when comparing results between year 2010 and years 2011 and 2012. Foreca's predictions are produced twice a day, at 00 and 12 UTC time. However, there are 6 hours of data that cannot be used, due to delays of acquiring the data. When one participates in the Elspot-market, which closes at 1 pm Finnish time, it is possible to notice that forecast made at 12 UTC time cannot be used in day-ahead trading. Thus, the power production predictions are based on the forecast made at 00 UTC. However, when trading in the intraday market the forecasts commenced at 12 UTC until the next 00 UTC forecast are available.

Weather forecasts from FMI and SMHI are both based on the same forecasting system, HIRLAM (HIgh Resolution Limited Area Model), with horizontal resolutions of 7.5 and 5.5 kilometers, respectively. Foreca uses a different model, ETA (2010), with a horizontal resolution of 15 kilometers and ECMWF (2011 and 2012) with 16 kilometers of horizontal resolution. Weather forecasting horizon is reaching 55 hours ahead. Although, forecasting data up to 36 hours ahead are used.

In Section 3.3, data from one NWP provider was used for all sites (24, 25 and 23 sites for years 2010, 2011 and 2012, respectively). In Section 5, data from three NWP providers was used. Six wind turbine sites were used, for which NWP-prediction data from 2–3 providers were available. In Section 6 data from FMI and Foreca ECMWF was used for three sites. Aggregated wind power production data from 30 sites was also included in Chapter 6, which was based on Foreca's ECMWF weather forecasts.

3.2 Sites and regions studied

In Section 3.3 wind power production in Finland has been divided into four different geographical areas. The locations of the turbines sites and aggregation areas used in the study can be seen from the Figure 2. The total length of the area is 695 kilometres, and if the sites located on the Gulf of Finland are neglected, rest of the sites are located on almost linear line, which can be drawn from Åland to Oulu. The four different areas are: one site, whole Åland, Åland and Bothnian Sea (Selkämeri) and Åland, Bothnian Sea and Bothnian Bay (Perämeri). Analysis will also be made for aggregation of all of the sites included in the simulation.

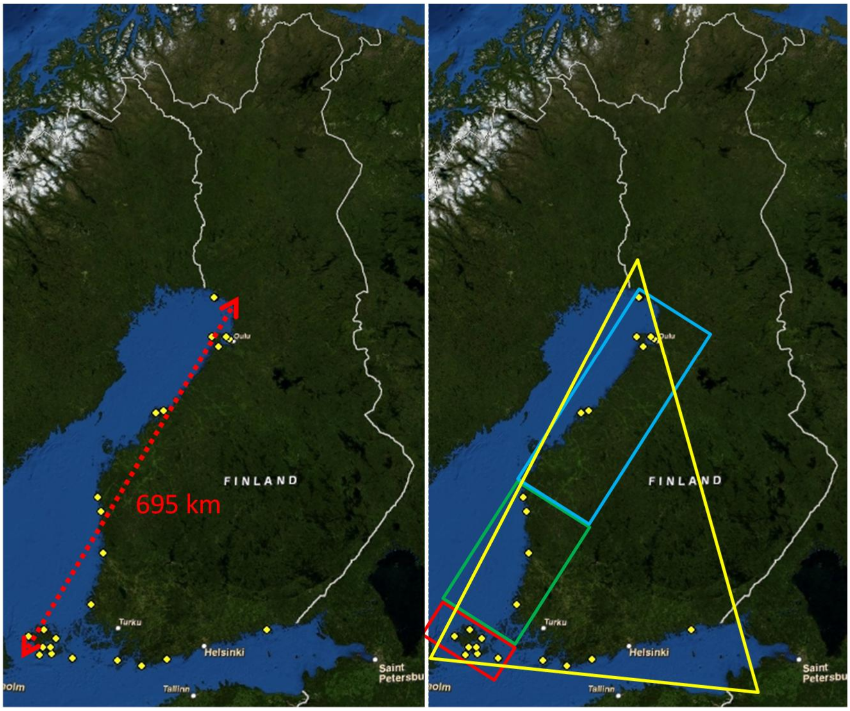


Figure 2. Aggregation areas used in the study. Red, green and blue rectangles are aggregation areas of Åland, Bothnian sea and Bothnian bay, respectively. Yellow triangle is representing the aggregation area of all wind turbine sites included in the study.

Aggregation areas are chosen so that when moving to larger areas there will be approximately 20 MWs more wind power capacity than in the previous area. In Table 1 number of sites in different aggregation areas and their aggregated wind power capacities are presented.

Table 1. Number of sites, installed total capacity and region area for different aggregation areas.

	A [sites]	A+S [sites]	A+S+P [sites]	All [sites]
2010	8 (21.3 MW)	12 (42.9 MW)	18 (62.9 MW)	24 (105.0 MW)
2011	9 (21.3 MW)	13 (42.9 MW)	18 (68.6 MW)	25 (126.0 MW)
2012	9 (21.3 MW)	11 (36.0 MW)	15 (59.7 MW)	23 (130.6 MW)
Area [km^2]	7000	27400	64000	231000

3.3 Market data

Electricity price data is required in Chapter 4 where forecast errors and their induced costs for different sizes of producers are analysed. Actual historical Finnish area prices from Nord Pool Spot and up and down regulation prices from Fingrid are used to analyse balancing costs for a wind power producer.

In Figure 3 development of Finnish area prices and up and down regulation prices for years 2004–2012 are shown. The balancing costs, which are shown as error bars in Figure 3, are the unit prices that a producer must pay extra (or receive less) for having imbalance energy. Thus, unit balancing prices are differences between up and down regulation prices and spot-prices. Balance settlement prices are based on prices in Regulating Power Market and they usually follow spot-prices (Figure 3).

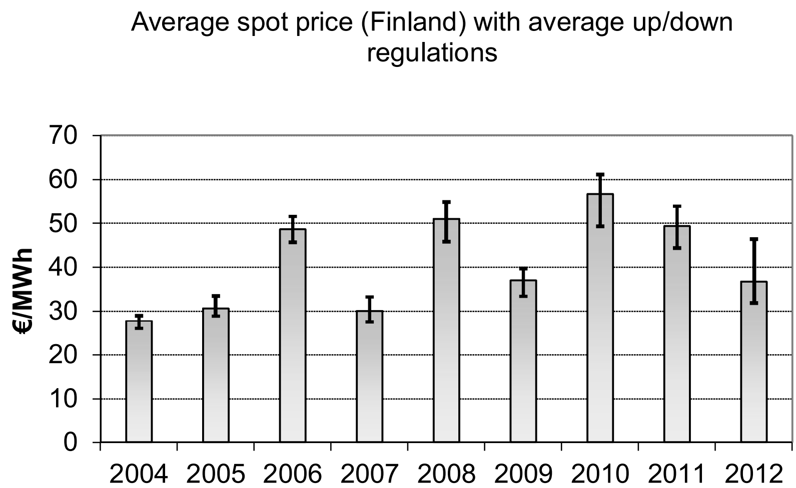


Figure 3. Average Finnish area prices (grey bars) and average balancing costs (black bars) for years 2004–12. This study uses data from 2010–12.

The average Finnish area prices have varied considerably for the past eight years, from 27.6 €/MWh (2004) up to 56.6 €/MWh (2010). For the years 2010–2012 the average area price has decreased steadily. Duration curves of Finland area prices are shown in Figure 4 for the three years studied. The area prices have been more stable in 2012 than in years 2010–2011.

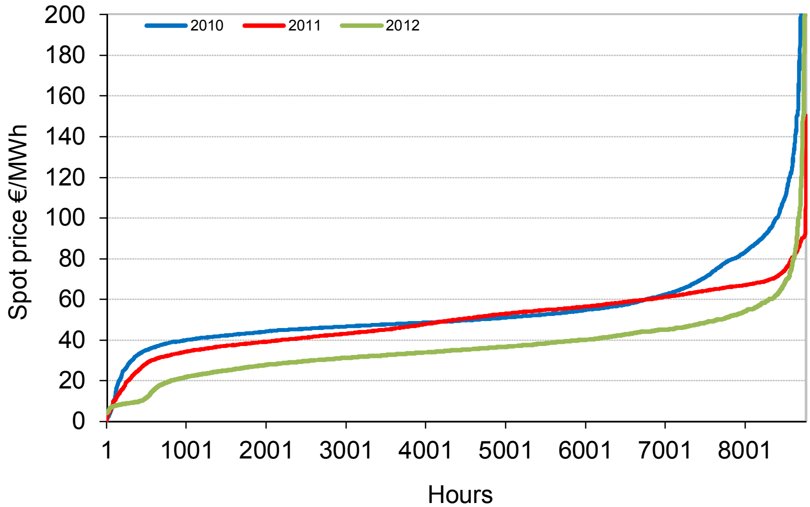


Figure 4. Finland spot duration curves for years 2010–2012.

The unit balancing costs for up and down regulation have been quite symmetric compared to average area price in past eight years (Figure 3). However, the absolute value of unit balancing costs for down regulation have been greater than up regulation for six of those years. The average unit balancing cost for whole 2004–2012 period was 3.9 €/MWh. The variation on different years can still be substantial – for year 2012 the average up regulation cost was 9.6 €/MWh and down regulation cost was -4.9 €/MWh. The duration curves for up and down regulation costs for years 2010–12 are shown in Figure 5. The duration curves are quite similar with an exception that in 2012 up regulation prices have been larger than in previous years.

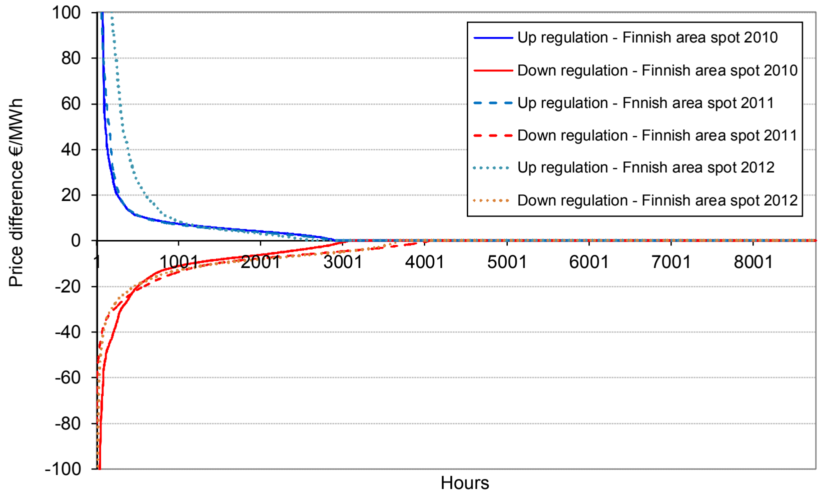


Figure 5. Duration curves for up and down regulation costs in Finland for years 2010–2012.

4. Forecast errors in Finland

In this chapter forecast errors in Finland are analysed for different time horizons and geographical areas. Also, costs for different size market participants are calculated to see what benefits pooling geographically dispersed wind power give to a market participant.

Day-ahead wind power predictions have been made for 24, 25 and 23 sites along the coastline of Finland for the years 2010, 2011 and 2012, respectively. Similar study was performed in [8], however in this study aggregation of geographical areas is carried out differently. The analysis is made for different region sizes:

- one site
- whole Åland
- Åland and Bothnian Sea (Selkämeri) and
- Åland, Bothnian Sea and Bothnian Bay (Perämeri)
- Aggregation of all of the sites included in the simulation.

The main interest is what will happen to the imbalance costs when wind power production is aggregated to larger area. The three areas Åland, Bothnian Sea and Bothnian Bay have approximately 20 MW wind power capacity each. In Table 1 the number of sites in different aggregation areas and their aggregated wind power capacities are presented. The aggregation areas are also shown in Figure 2. Results show costs for years 2010 to 2012. The costs are calculated by using actual market data from Nord Pool Spot and balancing prices from Fingrid. The detailed results can found in Appendix A.

4.1 Description of VTT forecasting model

Wind power forecasts in this study were created using forecast model developed at VTT [9]. The system has been designed to be used as a research model for existing data. It calculates point forecasts for one month at a time, and is given 1–3 months of data (from previous time period) to be used for training. The prediction system is based on a NARX (non-linear autoregressive with exogenous inputs) time-series model that utilizes both the past realized values of the power production and external information. Thus, the model can be characterised as a hybrid statistical model. Meteorological forecasts consisting of the forecasted hourly wind speed and direction for the geographic location and elevation of the wind power plant are used as external information.

In order to make forecasts, two separate time windows are defined: training and forecast window. Forecasts are made on a monthly basis and therefore training window for the forecast model remains the same for the whole month. Training period, where model's monthly parameters are estimated is lagged by a year. Thus, for year 2011 forecasts, model was trained by using data from year 2010, and similarly for year 2010 data from year 2009. Best results were attained by using three months training window, which were centred to the month of interest. Main reasons why data from previous year were used is that seasonal effect of wind power production can be included to the forecast model parameter estimation.

VTT forecast model does not take into account which turbines are out of operation. Thus, there might be outages and the forecasted power production in a site is based on false assumptions. Since the forecast model is a statistical model, it requires a training period where the model's parameters are estimated. Therefore if behaviour of turbines on the training period differs greatly from the behaviour on the forecasting period, it leads to forecast model with badly configured parameters. Some sites were left out of the results because of low technical availability either in training or forecasting period.

4.2 Forecast errors for different time horizons

In Figure 6 average errors for different horizons are presented for single wind power plants for year 2011. The average absolute error NMAE is relative to installed capacity of each wind power plant. All 24 sites are included in the results, one wind power plant showing close to average forecast accuracy, as well as the range of errors are shown.

There are a lot of differences in results for single sites, for example for 24 hours ahead prediction the average absolute error can be between 7–14% of installed capacity. This can be partly due to the different accuracy for different types of sites, like different orography and roughness, and how homogenous the grid cell for the numerical weather prediction model is. Results also depend on the wind resource on the sites – low wind situations are usually easier to predict and thus low wind sites have less error, relative to installed capacity. Prediction errors in low wind situations can still be higher relative to yearly energy.

The forecast error is significantly lowered for the first six hours. The physical reason behind this is that wind itself has an auto-correlated nature and the model takes this into account by last measured power value when making forecasts to the future. As the forecast horizon increases the variation range seems to increase quite linearly.

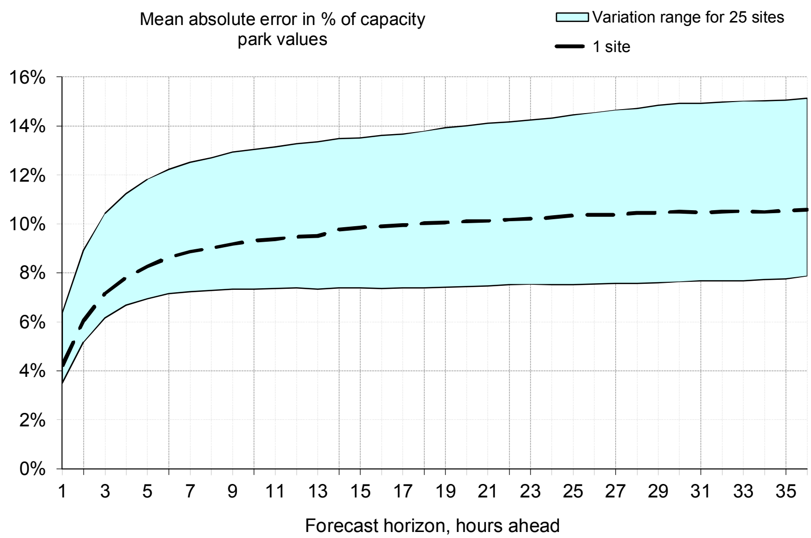


Figure 6. Forecast errors on different horizons for 2011.

Results of calculating the average error for different size areas can be seen from the Figure 7, for all three years. One could notice that MAE of energy is a bit larger for year 2010, which is caused by different wind forecast inputs used. In years 2011 and 2012 same NWP-model were used, whereas 2010 there was a different NWP-model. In 2011 it was noticed that Foreca ECMWF NWP outperformed ETA model and the forecasts for wind were changed to ECMWF model. Therefore, the variation range in the future is more like in years 2011 and 2012.

The MAE, relative to yearly energy production, varies between 52–56% for one site in years 2011–2012. When aggregating the whole Åland's wind power production the MAE values are decreased significantly, when comparing to the single wind farm case. Åland as an aggregation area can easily correspond to a single distribution company in continental Finland. The result of aggregation drops the range of variation to 33%, which is approximately 40% improvement from the single site case.

When including wind power production from Bothnian Sea to the wind power production at Åland, it is possible to further decrease MAE of energy. When calculating the MAE aggregated for all simulated sites it is possible to achieve 20% MAE of energy, which is more than 60% improvement from the single site case.

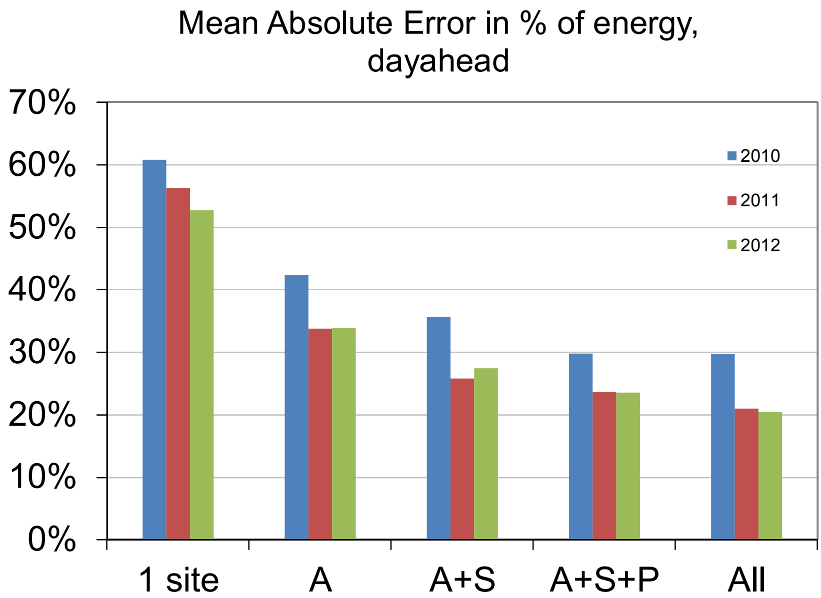


Figure 7. Prediction errors relative to total energy produced. A, S and P stands for Åland, Bothnian Sea and Bothnian Bay, respectively. In "All" column includes every wind farm/turbine, which were included to the simulations on a particular year.

4.3 Forecast errors and share of errors that cause balancing costs from day-ahead bidding

The previous section showed how the aggregation of power production decreases MAE of energy significantly. Decrease of forecast errors will have a large impact on the balancing cost reduction as well.

Figure 8 shows forecast errors and the part of forecast errors that incur balancing costs (wind power produce's balancing need is the same up or down as the system regulation cost). They are presented for 2010, 2011 and 2012, as share of total energy produced by wind power. There are some difference for the results for different years, but the result for aggregation are similar: relative prediction errors are decreasing when aggregation area is increased. The benefits, in terms of forecast error, are clear already when aggregating wind power production in the relatively small region of Åland. In 2011 (2012) for a single wind farm relative share of prediction errors for up- and downwards direction are 33.2% (28.4%) and -23.1% (-24.3%) from the total production, respectively. When the whole Åland is considered as a balancing area, prediction errors are 17.0% (16.3%) and -16.7% (-17.6) for up and down regulation.

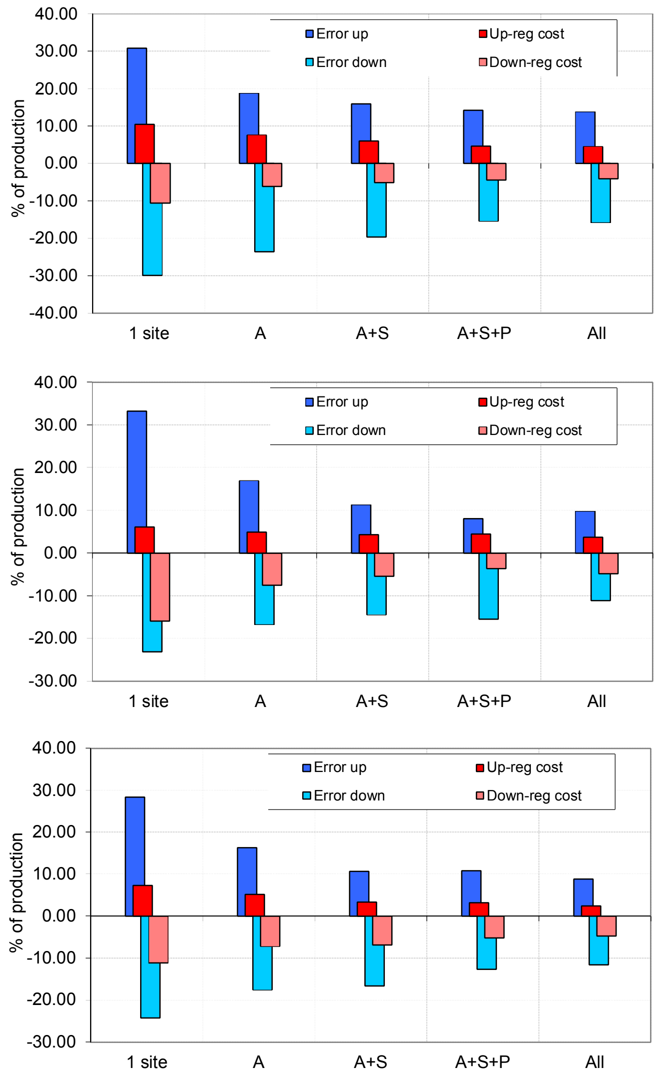


Figure 8. The share of total production that will be prediction errors (blue bars) and share of energy causing penalties in imbalance settlements (red bars), for a single site and clusters of sites in 2010 (upper figure), 2011 (middle) and 2012 (lower). A, S and P stands for Åland, Bothnian Sea and Bothnian Bay, respectively.

In balance settlement, there is no extra penalty if producer has been short and buying up-regulation energy when there has been down-regulation needed in the power system for that hour. The same applies if producer is long and needs to sell surplus energy for down-regulation price and there has been up-regulation need during that hour. In these cases the producer buys missing energy or sells surplus energy at spot price without extra penalty or imbalance cost. As wind power is still very small in Finland and also relatively small in the Nordic electricity market, wind power imbalances do not correlate with system imbalances, and about 50% of time there is no extra cost. This is why in Figure 8 the red bars, showing the amount of energy for which there is imbalance cost, is much smaller than the total forecast error in energy. As the relative share of balancing energy decreases, when the balancing area increases, the relative share of balancing costs decreases. For a single site the share of energy that causes down regulation is -15.9% (11.1%) from the total production and for up regulation 6.1% (7.3%). When aggregating more sites together also the share of energy that causes imbalance costs reduces. For calculating all the sites on both years, the share of total production that causes up regulation and down regulation are 3.70% (2.4%) and -4.8% (4.7%) from the total production, respectively.

The comparison of the years is presented in Figure 9 where the shares of energy that are wrong predicted and shares of energy that result in imbalance costs for the three years 2010 to 2012 are presented for single site and all of the sites. There are some small differences in how the errors are distributed up and down, but similar behaviour can been observed from all of the years. The 2010 prediction errors are a bit larger than prediction errors for years 2011 and 2012, but there is not as much differences in the amount of energy that needs up or down regulation in balance settlement. See for detailed results in Appendix A.

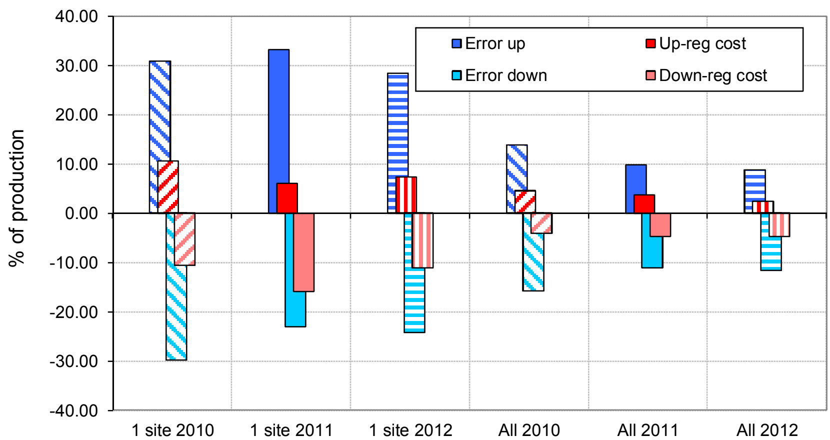


Figure 9. The share of prediction errors from the total production (blue bars) and errors causing penalties in balance settlements (red bars), for a single site and for all of the sites in years 2010–2012.

4.4 Imbalance costs from day-ahead market bidding

Wind power producer's revenue can be formulated by using two terms: revenue from Spot-trade, and cost/revenue of balancing energy for up and down regulation. Balancing energy and regulation need is the same as prediction error for a WPP, and minimizing prediction error is the way to maximize the revenue for the participant

In Figure 10 average unit balancing costs (€/MWh) for different years and areas are shown. There is considerable difference in the balancing costs for the years 2010–12. The forecast errors were reduced in 2011–12, explaining the drop from 2010 to 2011. In 2012 the balancing costs from the market were much higher than in 2011, compared to Elspot prices. This explains the increase in 2012 – as seen from previous graphs, the forecast errors were at a similar level in 2011 and 2012.

The penalty that balancing costs incur for a small producer who has only one site, is on average 3 € less revenue for each produced MWh, than the revenue would have been without forecasting errors. For instance producer who has one 3 MW turbine with 30% capacity factor would have 23 000 € losses from balancing in a year.

Conventional power producers have much less imbalances as they can adjust their output and avoid most of the balancing costs. Wind power plants can also adjust their output by curtailing some production, however this would mean losing the production and thus incurring costs as well, especially as the subsidies are often based on production.

The unit balancing costs decrease considerably as the balancing area increases. When the balancing area is “All”, which means that there are sites all over the coastline of Finland, then the unit balancing costs are only half of that for single wind power plant. Wind power has similarities to electricity consumption, in that there is a strong aggregation benefit as individual forecast errors will partly cancel each other. Acquiring wind power plants from geographically different areas would benefit the producer, as the balancing costs per MWh will be reduced. For the previous example case of small producer, when buying new turbines from other geographical locations and bidding to the market aggregated power production, the balancing cost would even drop down to 1.4 €/MWh. This means imbalances would incur 11 000 € costs in a year, and a saving of almost 13 000 €/year, when comparing to the previous example.

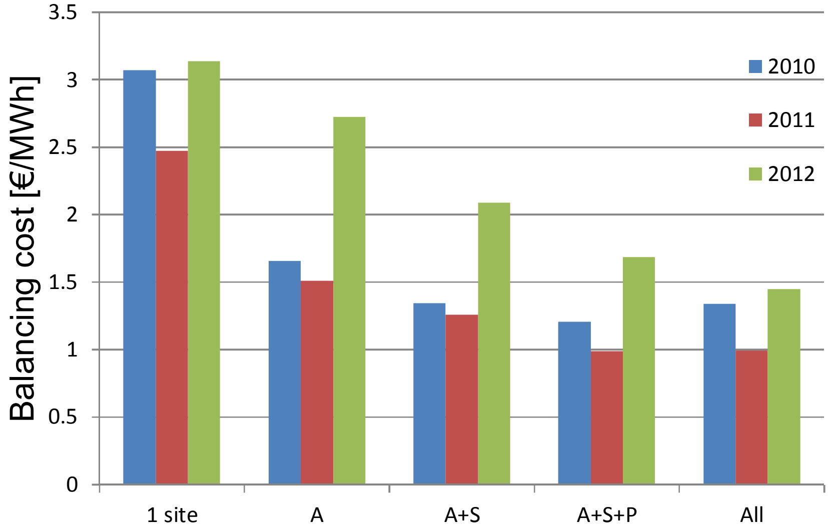


Figure 10. Unit balancing costs. A, S and P stands for Åland, Bothnian Sea and Bothnian Bay, respectively.

The results are calculated for Finland area prices and are only valid for low wind power penetration level (wind power not impacting the Regulating market prices and their direction).

Unit balancing costs are naturally highly dependent on the balancing prices. The level of unit balancing costs in the future depends greatly how much cost effective regulating power will be available in the Nordic Regulating power market – increasing wind power in the system will bring more demand for Regulating power and the prices will probably rise in future. The forecasting model development will improve the accuracy, which may slow down the impact that increasing wind power has on the market prices. We have here looked at day-ahead trading only, the possibility to correct at least part of the largest forecast errors as more accurate forecasts arrive some hours before delivery may lower both the imbalance costs for the producers as well as impacts for the system. This intra-day trade, at Elbas is not analysed in this section. Some analyses have been made before in [23] and show that for small penetration levels of wind power and moderate imbalance prices, a continuous trade in Elbas is not cost effective for the producer.

The results are based on two-price balancing costs used for production. In [8] balancing costs for one-price balancing cost system are calculated for year 2010 in Finland. One price model would be better for a wind power producer since unit balancing costs would be smaller than in two-price balancing system.

5. Improving the accuracy of wind power forecasts by using multiple NWP-models

This section summarises the results of a study [11] on how the accuracy of the forecasts can be improved by using alternative weather forecasts from several providers instead of a single one. In the study, 2–3 weather forecasts were available for each wind power plant for making hourly bids for the electricity markets (day-ahead and 1, 12, and 36 hours ahead). When used as inputs to a prediction system this resulted in several alternative production forecasts that needed to be combined together into a final forecast.

The task was approached by treating it as a model combination problem, where the combined forecast is a weighted linear sum of the alternative forecasts and the challenge lies in selecting the weights so that the long term forecast error is minimized. To take into account changing weather conditions, the combination weights were readjusted online when needed in response to recent forecast errors. This made the combined forecast a linear sum using time-varying weights.

Combination methods previously presented in the literature were implemented and integrated into an existing (VTT) prediction system. How the prediction model functions is explained in more detail in Chapter 4.1. The algorithms were evaluated by generating hourly power forecasts for six wind turbine sites in Finland during a four month forecast period. The results were compared against an average of the alternative production forecasts, where all members of the combination were weighted equally, and to production forecasts based on only a single weather forecast, where model combination was not used.

5.1 Selected combination methods

Previous research on model combination gives some suggestions on how to select a suitable combination method among the various alternatives. Unfortunately, no method has been found to perform well under all conditions and error measures [13]. Results from wind power related applications have also been reported to be very location dependent. For this reason, it was decided to select a number of candidate methods that have been utilized in previous work, and evaluate them using two practical test scenarios:

- *Simple average* (later abbreviated AVG), where all combination members are equally weighted and the final combination is an aritmethic mean of the alternative forecasts.
- *Optimal combination* (OPT) by Bates and Granger [12] that selects the combination weights in order to minimize the variance of the combined errors. Forecast errors may also be assumed to be independent (`OPT-IND`).
- *Regression* (REG-NO) using the realized value (production) as the dependent variable and the alternative predictions as independent variables [14]. Weighting the combination members according to the least squares solution minimizes the squared differences between realized and predicted values. Possible systematical errors of the forecasts may be compensated by using an intercept term in the regression equation (REG).
- *Composite post-processing* (COM) that was developed for combining short-term wind forecasts by weighing the combination members according to their past accuracy [15].
- *Outperformance* (OUT), which weights the competing forecasts proportionally to their relative past performance measured with a squared loss function [16].
- *Fixed share* (SHARE) [17] that readjust the combination weights by repeating the following two steps. First, a combination member is weighted according to its previous error and learning rate of the algorithm. The learning rate controls how fast the algorithm reacts to changing conditions. Second, a specified fraction of the weight is distributed between all combination members.
- *AEC* (AEC), which attempts to improve on a previously presented aggregating algorithm (Aggregated Forecast Combination Through Exponential Reweighting) [19] by allowing the weights of combination members that have been performing poorly in the past to recover along time, and this way make the algorithm more suitable to non-stationary settings [18].
- *Kalman filter* (KALMAN), which is a widely used algorithm within the family of state-space methods [21]. It uses the minimum mean-squared estimate, considering all available information, as an estimate of the current state of the system that is being tracked. In this study Kalman filter was utilized as described in [20].
- *Recursive least squares* that updates the least-squares solution to a regression equation repeatedly when new information becomes available [21]. In this study, the regression equation was formed both with an intercept (RLS) and without (RLS-NO).

More detailed description of the evaluated algorithms and their implementation for the evaluation can be found in [11].

5.2 Methodology and case study

Six wind power plants in Finland (Table 2) were used in testing the new prediction system. The power plants are located along the Finnish coastline and in the Åland islands. Largest distance between the power plants is approximately 600 km.

Table 2. Selected sites for the evaluation of multiple NWP models.

Location	**Capacity (MW)**	**Turbines**	**Hub height (m)**	**Monthly downtime (h)**	**NWP-providers**
Oulunsalo	~ 10	~ 5	~ 60–80	~ 50–100	1,2
Kokkola	< 5	< 5	~ 60	~100–250	1,2
Raahe	> 15	~ 10	< 50	< 30	1,3
Hamina	~ 10	~ 5	~ 100	< 30	1,2,3
Pori	~ 15	~ 10	~ 60–100	~ 50–100	1,2,3
Nyhamn, Åland	~ 15	~ 5	~ 60	n/a	1,2

For all power plants, realized production data and meteorological forecasts from 2–3 providers were available. A single weather forecast was used for all turbines in one wind power plant site. Two months of data was used in parameterizing the combination methods (07/01/2011–08/31/2011), and four months (09/01/2011–12/31/2011) in the evaluation. Depending on the provider, 2–4 updated weather forecasts are delivered per day, reaching up to 55 hours ahead.

Weather forecasts from providers 2 and 3 are both based on the same forecasting system, HIRLAM (HIgh Resolution Limited Area Model), with horizontal resolutions of 7.5 and 5.5 kilometers, respectively. Provider 1 uses a different model, ETA, with a horizontal resolution of 15 kilometers.

When producing updated power forecasts for a particular wind power plant at time *t*, in addition to the weather forecasts also the realized production values are known for the preceding hours *t*-1, …, *t*-n. In most cases, the realized production from all wind turbines belonging to the same wind power plant is aggregated into a single time series. It is worth to note that the average monthly downtime for wind power plant 2 is quite high, which may increase the forecast errors as the detailed data on downtime was not available to be taken into account when providing the forecasts.

As an error measure, RMS (Root Mean Squared) errors are reported separately for each evaluated combination method. The four month forecast period consisted of 2902 hourly time points. Power forecasts were generated separately for each month and variances of the monthly errors were calculated. Considering the error variances and results across different wind power plants gives information on the robustness of the methods. A robust method is less affected by various site specific factors, such as the complexity of the terrain and local weather conditions, and therefore generalizes better to other locations as well. Simple average (AVG) is used as a baseline and the advantage of using more sophisticated combination methods was considered by comparing their results to the average

The combination methods were evaluated using two test scenarios, where wind power forecasts are produced for the needs of the electricity market:

1. *Daily updated forecasts*, where hourly power production forecasts are delivered once per day at 12:00 CET for the upcoming 12–36 hours using same weights for all horizons.
2. *Hourly updated forecasts*, where updated power forecasts are delivered hourly for the next hour in addition to 12 and 36 hours ahead with weights selected separately for each forecasted hour and horizon.

The two scenarios differ in how often updated forecasts are delivered and whether the weights are adjusted separately for each forecasted time point. Both forecasts can be used for day ahead markets, for hourly updated forecasts this means taking the available forecasts at 12:00 CET for all hours next day as basis for bids. Since the production of each hour needs to be forecasted, the delay between consecutive updates determines the required length of the forecast horizon. Daily forecasts pose more problems since the forecast errors increase together with the horizon. When the production forecasts (and bids made according to them) are delivered once per day and for several hours ahead, it is very likely that the weights of the combination would need to be adjusted during this time period, when new information become available. However, this is possible only when new forecasts are delivered, every 24 hours. This means that even if online measurements of the realized production would indicate a need to readjust the combination, the already delivered forecasted production values cannot be changed. Updated weights can therefore be used at the earliest when delivering the next forecast. To what extent this affects the error of the combined forecast is determined by the stationarity of the relative accuracies of the alternative forecasts.

This problem affects the hourly forecasts less, since new forecasts are delivered each hour. When forecasting the next hour the prediction system can react to changing situations faster by readjusting the combination weights at each update and separately for each forecasted hour and horizon, according to forecast errors that occurred during preceding hours.

The prediction system was trained separately for each (monthly) test period using three months of data consisting of past meteorological forecasts and realized power production (in Appendix B). The training phase was performed separately for each wind power plant and forecast horizon.

The combination methods weight the alternative forecasts based on past forecasted and realized values. Of the evaluated methods Fixed share, Kalman filter, Recursive least squares and AEC have been designed for online operation. They continuously monitor the forecast errors and adjust themselves accordingly. Apart from the simple average, the remaining methods (Optimal, regression, Outperformance and COM) need a separate training period to estimate the combination weights. This calibration phase was repeated before each new power forecast delivery, using all available information so far.

The delay between consecutive delivered forecasts restricts how often the combination weights may be modified. The prediction system may internally adjust the weights every hour after knowing the previous forecast error, but updated weights may be used only when combining forecasts for the following delivery. In this study, two cases were studied. First, keeping the weights same for all horizons next day and updating once a day. Second, adjusting the weights separately for each horizon and updating every hour. Weights for the daily forecasts were only allowed to change every 24 hours. This affects the combination methods, since they are unaware of whether the updated weights are in use or not. Since the hourly forecasts are updated each hour, new weights became active immediately when forecasting the following hour.

When producing updated forecasts for the next hour, the error from the previous forecasted hour was available and could be utilized by the prediction system. However, with 12 and 36 hour horizons the realized production, and therefore the forecast error, is known only after several time steps (determined by the length of the horizon) have passed. For example, when forecasting the production for hour t+h at hour t, the realized production for hour t-1 is known, but not for hour t+h-1. In other words, even though the prediction system utilizes the most recent errors available at each time instant, it has no information on the errors that occur after several hours in the future. This affects all combination methods besides averaging, since they all assign the combination weights based on the past accuracy of the members.

Some of the combination methods (regression, Fixed share, Recursive least squares and AEC) have configurable parameters that affect the behavior of the algorithms. The parameters control, for example, how fast the method reacts to changing conditions and how quickly older information gets discarded. Suitable values for these parameters were searched by running the prediction system as previously described using weather forecasts from two providers, combining the power forecasts using various configurations and observing how it affects the forecast error. Two months of data (07/01/2011–08/31/2011) was reserved for this purpose and not used in the actual evaluation. Further details of the parametrization can be found in [11].

On-site measurements of the actual wind speed and direction were not available for this study. The weather conditions during the evaluation period (09/01/2011–12/31/2011) based on weather forecasts shows a relatively high wind period, increasing wind speeds from September to December. At the end of December there were some stormy days. Wind power plant no 6 is a high wind resource site, the others show similar average wind speeds.

An excerpt of the evaluation period showing the hourly forecasted wind conditions together with the realized power production can be seen in Figure 11. Provider 1 frequently forecasts higher wind speeds compared to the other two providers. The differences are larger during more windy conditions. Wind speed histograms (Figure 12) confirm the same observations. Distributions for provider 1 are wider and have more weight assigned to higher wind speeds.

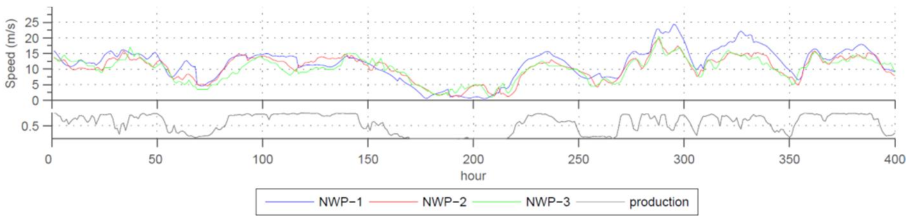


Figure 11. Excerpt from the evaluation period showing the alternative wind speed forecasts used in producing daily forecasts (12–36 hours ahead) for one of the wind power plants. Realized power production has been normalized by the capacity of the site.

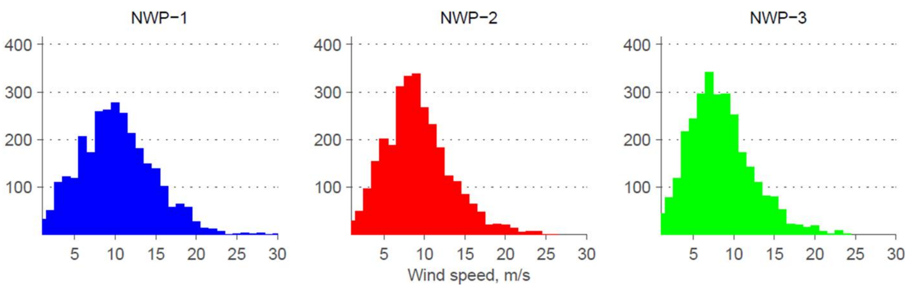


Figure 12. Histograms of forecasted wind speeds used in producing daily updated forecasts (12–36 hours ahead) for one of the wind power plants during the evaluation period.

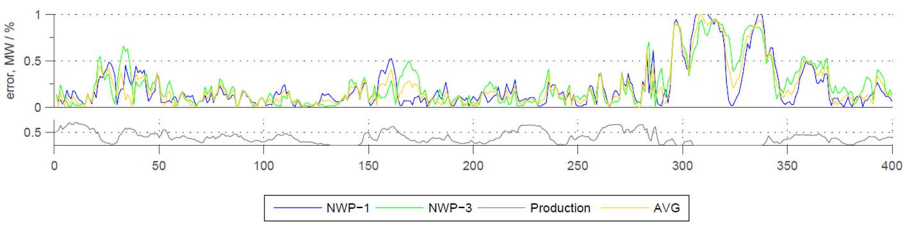
The correlation for the daily updated forecasts (12–36 hours ahead) lies approximately in the range 0.84 to 0.92 for all wind power plants Correlation of forecasts 2 and 3 is larger than the other two pairs, even exceeding 0.9. This is easy to believe since the two forecasts are based on the same HIRLAM system. In general, combining several forecasts usually leads to better results when the forecasts are not too correlated.

Correlation between the hourly updated forecasts (next hour, 12 and 36 hours ahead) decreases with longer horizons. This is explained by the increasing uncertainty, which also increases the differences between the forecasts.

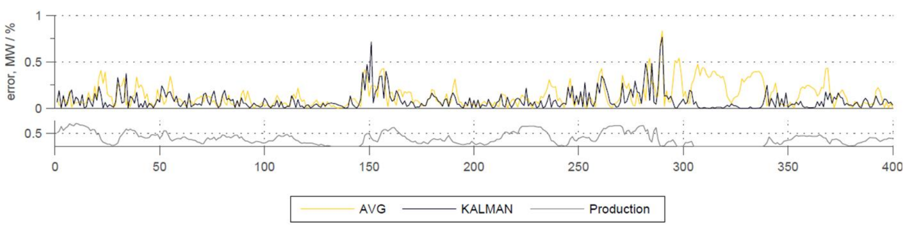
5.3 Results

Detailed resuls are presented in tables of Appendix B. There is a clear benefit of combining that can be seen already from the simplest combination method, averaging. When considering the whole evaluation period it is rare that the average has a larger error than the best performing forecast using a single weather forecast. However, this does happen for example for site 6. For most wind power plants, averaging decreases the errors of the daily forecasts (12–36 hours ahead) by 3–16% compared to power forecasts using a single weather forecast. The average forecast may not be the best performing alternative at each time point, but the advantage is apparent when examining the errors during a longer period.

Comparing the remaining combination methods against averaging for the daily forecasts shows that improvements are very rare and minor. RLS and Kalman filter are clearly unsuitable for this scenario. Also Optimal method without the independence assumption produces remarkably large errors at most sites. When assuming the forecast errors as independent, the errors are decreased. Overall, both variants still perform at most sites worse than averaging.



(a) Daily updated forecast for 12–36 hours ahead. The ordering of the power forecasts based on single weather forecasts varies along time.



(b) Next hour forecasts. The differences between the two combination methods can be clearly seen at times when the production was zero. Kalman filter was able to adapt rapidly to the uncommon conditions, most likely caused by very high wind speeds, by adjusting the combination weights according to recent realized values.

Figure 13. Absolute forecast errors for one of the evaluated sites during the last two weeks of December 2011 normalized by the installed capacity of the site.

5.3.1 Impact of using different weights for different horizons

There is benefit of updating the forecasts hourly and the possibility of adjusting the combination weights after each time step and separately for each forecasted hour and horizon can be clearly seen in the results. Errors of the next hour forecasts are noticeably smaller for all wind power plants when compared to the daily updated forecasts. Here averaging is not the best option, instead Kalman filter and RLS are substantially better than any other method (Appendix B). For both methods, using an intercept leads to improved results and can be recommended for all sites. Errors for Kalman filter with an intercept are 15–32% smaller compared to averaging. A larger fraction of the errors are small, which can be seen from the distributions (Figure 14). Omitting the intercept leads to large errors for sites 1, 2, and 5. This points out that using more sophisticated combination methods may occasionally be risky compared to averaging that has no configurable parameters and is not as condition dependent.

The benefits of Kalman filter over averaging can also be seen by examining the normalized absolute errors of the next hour forecasts during the last two weeks of December that had the storm event (Figure 13). In addition to being overall more accurate, Kalman filter is able to react faster to uncommon circumstances. An example can be seen when the production suddenly dropped to zero, due to very high wind speeds exceeding the cut-off wind speed of the turbines. Kalman filter monitored the realized production values from the recent hours and quickly adapted the forecast accordingly. Compared to the averaged forecast, which reacts only together with the combination members, the forecast errors were significantly smaller.

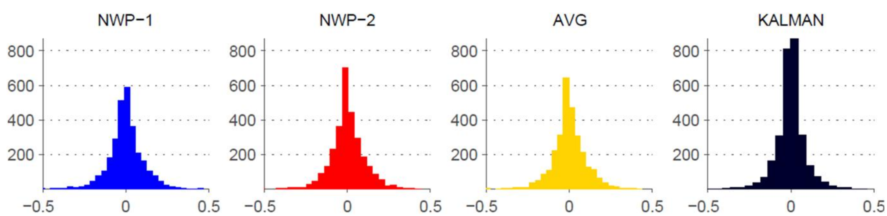


Figure 14. Histograms of normalized errors of next hour forecasts for one of the power plants during the evaluation period.

The distribution for Kalman filter is noticeably sharp, which indicates that a larger fraction of the errors are small. For next hour forecasts RLS produces constantly somewhat larger errors than Kalman filter. Compared to averaging, RLS with an intercept excels in all power plants and decreases the errors by 10–18%. It is also worth to note that in some power plants variances of the monthly errors for hourly forecasts are smaller with Kalman filter and RLS than with averaging. Of the remaining combination methods, Regression with an intercept (0.1–1.4%), Fixed share (1–4%) and AEC (0.4–2.0%) also improve on averaging, but with clearly smaller margins. Once again, Optimal method without independence assumption leads often to large forecast errors.

Updating the forecasts hourly and increasing the forecast horizon to 12 and 36 hours affects Kalman filter and RLS the most. Their forecast errors increase very rapidly which makes the methods unsuitable to these two scenarios. The errors of Fixed share and AEC also increase, and are either larger than with averaging or lead to very minor improvements. The remaining methods occasionally improve slightly over averaging, but the results are site specific and not significant.

Interestingly, improvements resulting from combining using averaging increase when the horizon is longer. When forecasting 36 hours ahead averaging performs better than NWP-2 also for wind power plant 6. This is most likely explained by the fact that the weather forecasts are less correlated with longer horizons, and combining forecasts is usually more beneficial when the members differ more from each other. It is also in line with the general advice that recommends combining forecasts in situations with higher uncertainty.

5.3.2 Selecting the combination members

In addition to selecting a combination method, one also needs to pick the forecasts to be included in the combined forecast. When having more than two forecasts available, it becomes possible to discard some forecast(s) from the combination altogether. In the evaluation, three weather forecasts were available for wind power plants 4 and 5, and therefore it was possible to test the combination methods using all four possible combinations of the alternative power forecasts (Appendix C).

When comparing the forecast errors of using various combination members one notices that they are in line with the previous analysis: averaging is still very favorable for daily updated forecasts, and Kalman filter together with RLS for next hour forecasts. For all scenarios, the best combination varies according to the selected combination method and forecasted wind power plant. Combining forecasts 2 and 3 is frequently the worst option. This is most likely explained by the fact that these two are the most correlated among all forecast pairs, which usually decreases the expected improvements obtained by combining. Overall, differences between different combinations are often quite minor, but seem to increase for the hourly updated forecasts when the horizon is longer.

An interesting question is whether a combination member should be discarded if its forecast error is known to be larger compared to the other members. Based on the results, the answer once again varies according to the wind power plant, forecast update delay and the selected combination method. For example, with next hour forecasts discarding the worst forecast decreases the error, compared to averaging all three forecasts at both wind power plants. However, contrary results (where averaging all forecasts is the best option) are also common. All things considered, the results show that a forecast that would not perform so well on its own may despite this be useful when combined with other forecasts.

5.4 Discussion and recommendations

In this section the benefits of using weather forecasts from several providers and combining the resulting alternative power forecasts were evaluated. Previously used algorithms for forming the combined forecast (Optimal combination by Bates and Granger, linear regression, Outperformance, COM, Fixed share, AEC, Kalman filter and Recursive least squares) were selected and compared against the simplest combination method, averaging. The evaluation was carried out by generating power forecasts for six wind power plants in Finland during a four month forecast period. Two test scenarios were tested: *First* by keeping the combination weights the same for all hours next day (12–36 hours ahead) and updating the forecasts only once per day. *Second* by adjusting the weights separately for each forecasted horizon (next hour, 12 and 36 hours ahead) and updating hourly. The suitability of the selected methods for these purposes was evaluated using accuracy and robustness as criteria.

Using several weather forecasts resulted in decreased forecast errors for most sites and for both test scenarios. Averaging was very close to the best performing combination method for daily updated forecasts with same weights for horizons 12–36 hours ahead. For most tested wind power plants it decreased the errors by 3–16% compared to power forecasts based on a single weather forecast. For next hour forecasts, Kalman filter decreased the errors by 15–32% depending on the site, compared to averaging. AEC and Fixed share were also more accurate than averaging, but one needs to consider whether the improvements are large enough to overweight the additional effort of implementing and configuring a more complex algorithm.

For next hour forecasts the selection of an appropriate combination method is a compromise between improved accuracy and implementation efforts. Averaging is recommended when a simple and parameter free combination method is preferred. However, the best performing combination methods are the ones originally designed for online operation (Kalman filter, RLS, Fixed share and AEC). Adaptive algorithms are able to react to changing conditions by gradually dismissing older information and adjusting the combination weights to take into account the forecast errors during the previous time points. Especially using Kalman filter and RLS with intercepts lead to significant improvements compared to averaging. The benefits of using a Kalman filter are in line with previous research in power forecasting, where the method has been reported to be useful in debiasing weather forecasts [11, 22]. Fixed share and AEC also outperform averaging for the next hour forecasts. However, especially with AEC the improvements might not be significant enough to be worth the extra effort of implementing and configuring a more complex algorithm. Similar to RLS, some improvements could perhaps be obtained by learning their configurable parameters (learning rate etc.) separately for each site.

For hourly updated forecasts 12 and 36 hours ahead the errors of the adaptive methods increase rapidly. This applies especially to Kalman filter, but for most sites neither RLS, Fixed share nor AEC offer meaningful improvements over averaging. Most likely this is caused by changes in the wind conditions during the time interval of producing the forecast and receiving the forecast error. If the conditions vary greatly during the interval, the combination weights assigned several hours ago will most likely be far from optimal. By considering that wind conditions and power production may vary on hourly basis, it is understandable that the situation may also vary during 12 or 36 hours. For this reason, especially Kalman filter and RLS are in most cases applicable only to next hour forecast. In contrast, the longer horizon would most likely affect the results less if the wind conditions would be more stationary. This could explain why RLS without an intercept improves on averaging in wind power plant 5 also for hourly forecasts 12 and 36 hours ahead. More data would be needed to determine whether this is an exception. It is worth to note that the varying wind conditions further explain the behavior of the adaptive methods also when producing daily forecasts 12–36 hours ahead. For daily updated forecasts with same weights for all horizons, simple averaging is very close to the best performing combination method for most sites. Averaging is reliable, but may still perform slightly worse than the best combination member, as can be seen from the results for wind power plant 6. The accuracies of the adaptive methods depend heavily on the delay between forecast updates. The longer delay between consecutive daily updated forecasts puts the algorithms into an unexpected situation by postponing the time point when updated weights become active. This leads to large errors, especially for Kalman filter and RLS. For daily updated forecasts the remaining methods (regression, Optimal, COM and Outperformance) seldom bring notable improvements compared to averaging. This may be caused by the choice of using all available information when determining the combination weights. Better results could perhaps be obtained by restricting the time window, even when weighting the data within the time window equally.

Averaging is more beneficial with longer horizons, compared to using a single weather forecast. Since the combination members are weighted equally, their past performance is of no interest when producing the forecast. The larger benefits for longer horizons is most likely explained by considering that the weather forecasts are less correlated when made further ahead, and that combining forecasts is generally recommended in more uncertain conditions. Neither here do the remaining methods, apart from a few exceptions, offer meaningful improvements over averaging.

Varying the combination of included weather forecasts had a quite minor effect on the forecast error. If the weather forecasts were less correlated, the differences would most likely be larger. Also, having a larger pool of forecasts to choose from might emphasize the importance of selecting the “correct” members. Not surprisingly, combining power forecasts based on the most correlated pair of weather forecasts was often the worst option. Therefore, correlation seems also here to be a good measure when selecting forecasts to be left out from the combination. The best performing combination members vary also according to the combination method. Combining all three forecasts did not always lead to the best results. Then again, including the worst forecast was occasionally more beneficial than discarding it from the combination altogether. Due to the small differences, including all forecasts may nevertheless be appropriate in practise. Another option would be to first select the combination method, and then test all the possible forecast combinations on archived data, and select the best performing combination for final use.

Table 3. Recommendations based on the results of the evaluation.

	updated:	**Daily**	**Hourly**		
Method	hours ahead:	**12–36**	**next hour**	**12**	**36**
AVG	Simple average	**yes**	**yes**	**yes**	**yes**
KALMAN	Kalman filter	no	**yes**	no	no
KALMAN-NO	Kalman filter, no intercept	no	no [1]	no	no
RLS	Recursive least squares	no	**yes**	no	no
RLS	Recursive least squares, no intercept	no	no [1]	no	no
REG	Regression	no	no	no	no
REG-NO	Regression, no intercept	no	no	no	no
OPT	Optimal combination	no	no	no	no
OPT-IND	Optimal with independence assumption	no	no	no	no
OUT	Outperformance	no	no	no	no
COM	COM	no	no	no	no
SHARE	Fixed share	no	possibly [2]	no	no
AEC	AEC	no	possibly [2]	no	no

[1]Including an intercept improves the results and increases the robustness of Kalman filter.
[2]Depending on the site, improvements over averaging might not be significant enough.

6. Probabilistic wind power forecasts using multiple NWP-models

Compared to deterministic point forecasts that consist of a single value for each forecasted time point, a probabilistic forecast provides additional information on the uncertainty of the forecasted values. This is used, for example, when producing interval forecasts, where the outcome is in a range within which the realized value will fall with a specified probability.

In the following, two methods for producing probabilistic wind power forecasts are compared. The methods produce quantile forecasts by using deterministic point forecast as input data. A quantile forecast consists of point values such that a specified proportion of realized values will fall below them. Quantile forecasts were used to derive interval forecasts. Both evaluated methods are based on an empirical approach where the error distribution is estimated using a sample of most recently occurred errors.

Another aim of the evaluation was to examine whether the accuracy of probabilistic forecasts may be increased by using multiple weather forecasts. Point forecasts that were used as input data to the evaluated methods were produced using weather forecasts obtained from two NWP-providers. Combined interval forecasts that utilize all weather forecast data were produced by averaging the alternative forecasts in two different ways. First, by using averaged point forecasts as input to the interval methods. Second, by producing separate interval forecasts for each alternative point forecast and averaging the resulting interval forecasts.

6.1 Probability forecasts

In Figure 15 an example of probability quantiles is shown. The different lines in the figure represent probability quantiles from 5–95%, which mean that the 5% quantile is the line closest to zero and 95% quantile is line nearest to 100%. These probability quantiles formulates probability intervals, which are illustrated with different colours. 90% probability interval consist an area, which is limited by 5% and 95% probability quantiles. Whereas, 10% probability interval is an area bounded by 45% and 55% quantiles. The median, or the centre of probability mass, is located at 50% probability quantile. It is the most likely outcome of forecast, which is the *point-forecast* or deterministic forecast of a forecast model. The VTT forecast model is providing this point-forecast and also NWP is giving its wind forecasts as point forecasts.

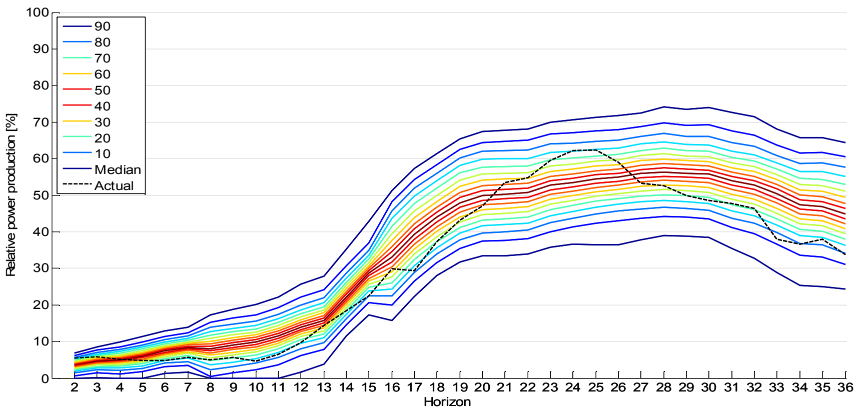


Figure 15. Example of a probability forecast.

6.2 Methods

Quantile forecasts were derived from deterministic point forecasts by adding the estimated error of each time point to the point forecast. The error of each time point was estimated using past errors during similar forecasted production levels. For this purpose, forecast errors were stored and binned according to the normalized forecasted production level when realized values became available. Breakpoints were assigned at values 0, 0.2, 0.5 and 0.7. Each forecast error was placed into the bin corresponding to its closest breakpoint. This was done separately for each wind power plant and time horizon, since they are known to affect the forecast errors. To make the system adaptive to changing conditions, the oldest point within the bin was discarded if the total number of points in the bin exceeded a threshold of 350 points.

Samples belonging to each bin were used to estimate the distribution of the most recent forecast errors during a certain forecasted production level. With the first method (abbreviated later `CDF`), the empirical cumulative distribution function of past *N* errors within a bin was calculated as

$$F(x) = \frac{1}{N}\sum_{i=1}^{N} I(x, x_i), \tag{3}$$

where I is an indicator function

$$I(x, y) = 1 \; if \; x \geq y, otherwise \; 0 \tag{4}$$

The empirical inverse cumulative distribution of past errors was calculated as

$$F^{-1}(y) = \max x \; s.t. F(x) \leq y \tag{5}$$

and evaluated at each time point and quantile proportion to get estimates of the forecast errors at specified probabilities.

With the other evaluated method (abbreviated later `KDE`) the distribution of binned forecast errors was estimated by using a kernel density estimate of the error distribution

$$F(x) = \frac{1}{Nh} \sum_{i=1}^{N} \frac{K(x - x_i)}{h}, \tag{6}$$

where x_i are past errors within the bin, K the kernel function, and h the bandwidth, which controls the level of smoothing. For the implementation, `ksdensity` - method of Matlab was used with a Gaussian kernel and automatic bandwidth selection to evaluate the inverse of F at each time point and quantile.

The resulting quantile points may be used to derive intervals with specified coverages. For example, the range of the 90% interval was formed by using as endpoints estimated error values corresponding to the 0.05% (lower limit) and 0.95% (upper) quantiles, and adding these values to the deterministic point forecast.

6.3 Study cases

Hourly deterministic wind power point forecasts for three wind power plants in Finland from two different NWP-provider was used in this study. In addition, the two forecasts were averaged hourly to get a simple, combined point forecast.

Deterministic forecasts were produced using a statistical model that uses production data and wind forecasts as input (VTT model, [9]). Wind forecasts were obtained from two providers, and were based on different NWP-models, ECMWF (16 km horizontal resolution) and Hirlam (7.5 km). The wind power prediction model was trained and run separately using wind forecast data from both providers to get two alternative point power forecasts with 1–36 hour lead times covering the period 1.9.2011–31.3.2012 (5110 time points).

Interval forecasts were produced using the two evaluated methods, `CDF` and `KDE`. Three alternative point forecasts were used as input data: deterministic forecasts of two different NWP-providers (Foreca and FMI) and their average. Finally, interval forecasts obtained by using both methods were combined hourly by averaging the upper and lower limits for all forecasted time points. Therefore, interval forecasts were produced using `KDE` and `CDF` methods for each site for four different evaluation methods shown in Table 4.

Table 4. Types of interval forecasts evaluated.

Interval forecast	Input to the interval method
NWP-1	Point forecasts using weather forecasts from provider 1
NWP-2	Point forecasts using weather forecasts from provider 2
AVG-1	Averaged point forecasts NWP-1 and NWP-2
AVG-2	Both point forecasts (NWP-1 and NWP-2). The two resulting interval forecasts were combined by averaging the upper and lower limits for all forecasted time points.

Table 5 gives some information about the wind farms that were used in the modelling. Wind farm at the Pori is a bit larger than the other two sites and it consists 10 turbines, whereas in Oulunsalo and Hamina there are only five turbines in each site.

Table 5. Wind power plants used in the evaluation.

Location	Capacity (MW)	Turbines	Hub height (m)
Oulunsalo	~ 10	~ 5	~ 60–80
Hamina	~ 10	~ 5	~ 100
Pori	~ 15	~ 10	~ 60–100

6.4 Evaluation metrics

Accuracy was considered as the most important property when evaluating the interval forecasts. This was measured by calculating the portion of realized values that fell within the interval. The resulting empirical coverage value should be as close as possible to the nominal coverage of the interval, for example 70% interval should have values 70% of the time. The difference between the values constitutes the bias of the interval forecast, and was calculated by subtracting the empirical coverage from the nominal coverage. Therefore, a negative bias indicates that the empirical coverage was too high, i.e. the realized values fell within the interval too often compared with the nominal coverage.

Sharpness of an interval forecast was measured by the average length of the hourly intervals. When two interval forecasts have similar biases, the one with a smaller average interval length (which is sharper) should be preferred. Standard deviation of the interval lengths indicates to what level the interval forecast varies based on the current conditions. A higher value is considered favorable, since it indicates that the forecast has been adjusted to changing conditions by varying the lengths of the interval. For example, it is intuitively clear that the uncertainty of the forecast is higher during high winds, and that this increased uncertainty should lead to a change in the interval length.

Finally, a skill score [10] for quantile with proportion α was calculated as

$$\sum_{i=1}^{N}(I-\alpha)(p_{i-}\hat{p}_i^{\alpha}), \tag{7}$$

where p_i is the realized power production of hour *i*, $\hat{p}_i^{\alpha}$ the forecasted production corresponding to quantile α of hour *i* and I an indicator variable

$$I = 1\ if\ p_i \leq \hat{p}_i^{\alpha}, otherwise\ 0 \tag{8}$$

A perfect forecast would receive a skill score of 0. The score is an attempt to summarise the performance of an interval forecast by considering both the accuracy and interval lengths.

When evaluating the results, 750 hours (points) were discarded from the beginning of the test period to take into account that both methods require past forecast errors for estimating the error distributions. Therefore the actual test period consisted of 4360 points.

6.5 Results

The following analysis is based on the results of evaluating interval forecasts with 90% coverage. In other words 90% ($\alpha = 0.9\%$) of realized values should fall within the interval during the test period. Correspondingly, the skill scores have been calculated for quantile forecast corresponding to the 90% quantile.

Comparing the biases of the `CDF`-method when using different point forecasts as input shows that AVG-2 leads to smallest biases for most horizons (Figure 16). The bias for AVG-2 interval forecasts are often positive and lie in the range -0.5–2.5% for all sites and horizons. This means that the empirical coverages were too low. Differences to AVG-1 forecasts are quite small, approximately up to 0.5 percentage points. Differences between NWP-1 and NWP-2 are approximately 0.5–1.0 percentage points in favour of NWP-2.

Results for interval forecasts produced using the `KDE`-method show that the biases lie in the range -2.0–3.0%. With `KDE`, using point forecasts AVG-1 leads to smaller biases compared to using AVG-2. In Oulunsalo and Hamina, AVG-1 is the most accurate interval forecast for most horizons. In Pori, NWP-2 leads to slightly smaller biases. In most cases NWP-2 is more accurate than NWP-1.

Overall, it can be seen that utilizing several weather forecast providers improves the accuracy of interval forecasts. Next, `CDF` and `KDE` forecasts were compared side by side. Both weather forecasts providers were used based on the approach that were previously found to lead to smaller biases: AVG-2 (i.e. averaging the two interval forecasts) was used with `CDF` and AVG-1 (averaging the two point forecasts used as input) with `KDE`. The results for the two methods are in most cases quite close to each other (Figure 18). This is also reflected in the skill scores that are very similar (Figure 19). There clearly is no significant difference between the two methods. The increasing uncertainty when forecasting with longer horizons is apparent when examining the length of the intervals (Figure 20): both methods produce longer intervals when the forecast horizon is further ahead. Average length increases from 22% at to almost 45%, relative to the production level. In other words, the methods compensate the higher uncertainty by making the intervals longer. This explains why the biases do not seem to depend on the horizon. On average, intervals of `CDF` are slightly longer than with `KDE`, especially above horizons of 12 hours, although the differences are only up to a couple of percentage points. There is no big difference on average length of interval forecasts (Figure 20), other than that the curvature of Hamina’s standard deviation differs from the other two sites.

Probabilistic wind power forecasts may be derived from deterministic point forecasts by collecting past forecast errors, evaluating the error distribution at specified points and adding the result to the forecasted values for each time point. The evaluated two methods resulted in biases that lay the range -1.0–2.5% for the 90% interval forecast.

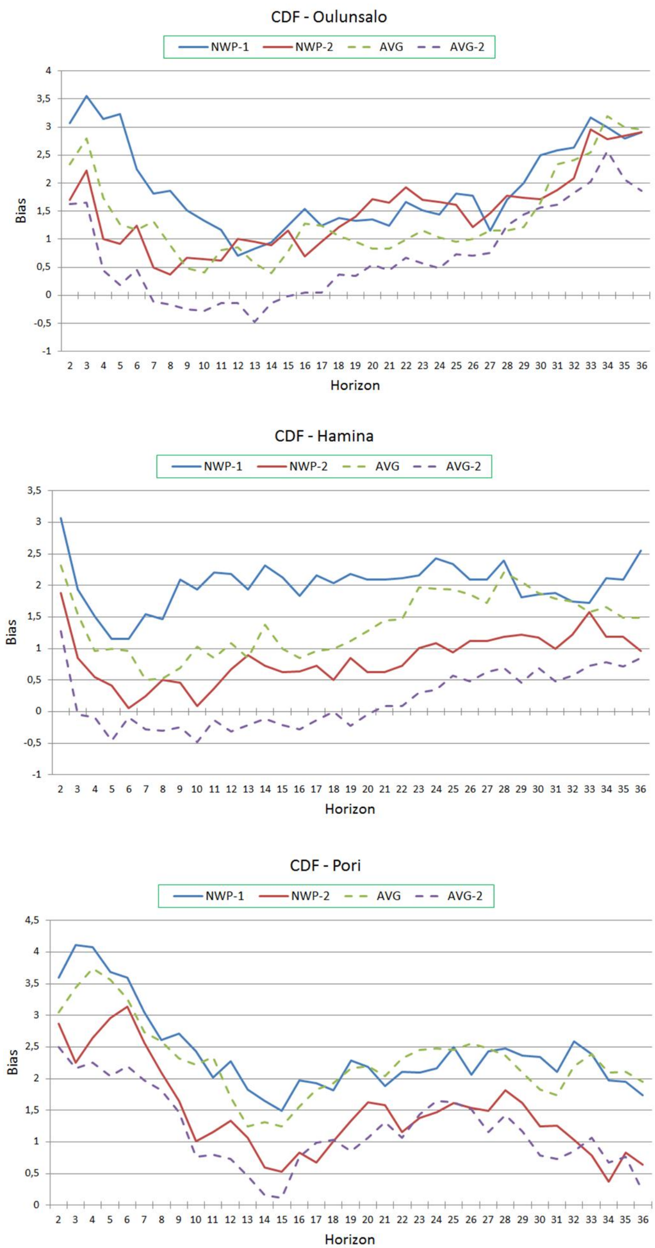


Figure 16. Biases of interval forecasts produced using the `CDF`-method for Oulunsalo (top), Hamina (middle) and Pori (bottom).

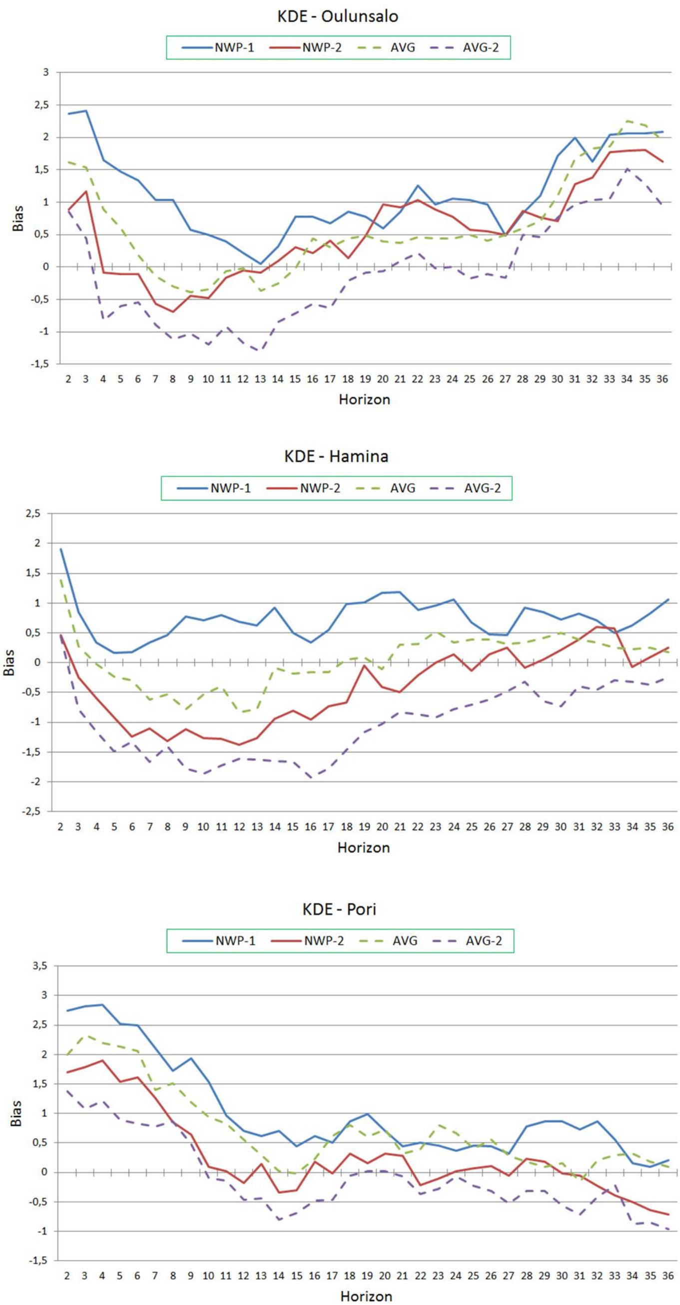


Figure 17. Biases of interval forecasts produced using the `KDE`-method for Oulunsalo (top), Hamina (middle) and Pori (bottom).

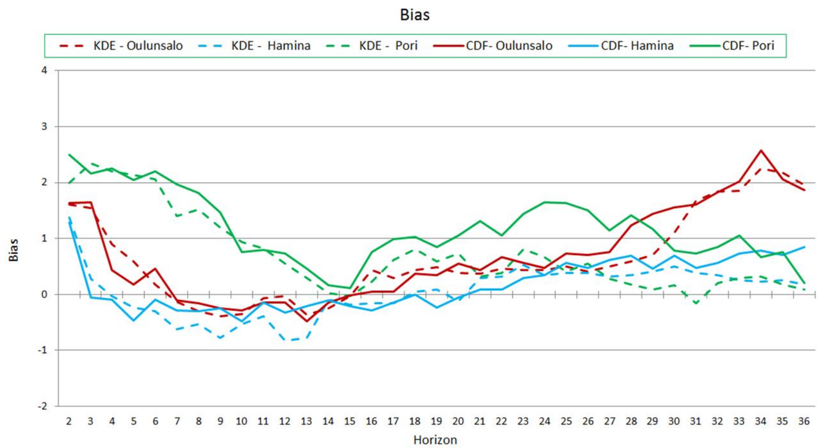


Figure 18. Biases of `CDF` and `KDE` interval forecasts.

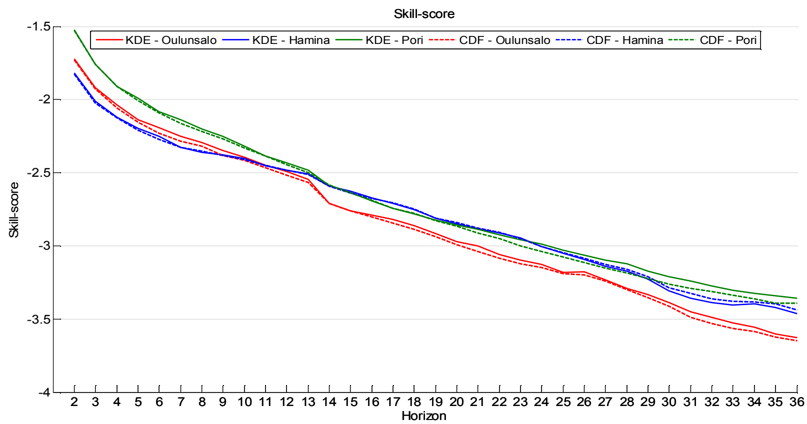


Figure 19. Skill scores (Equation 5) of `CDF` and `KDE` interval forecasts.

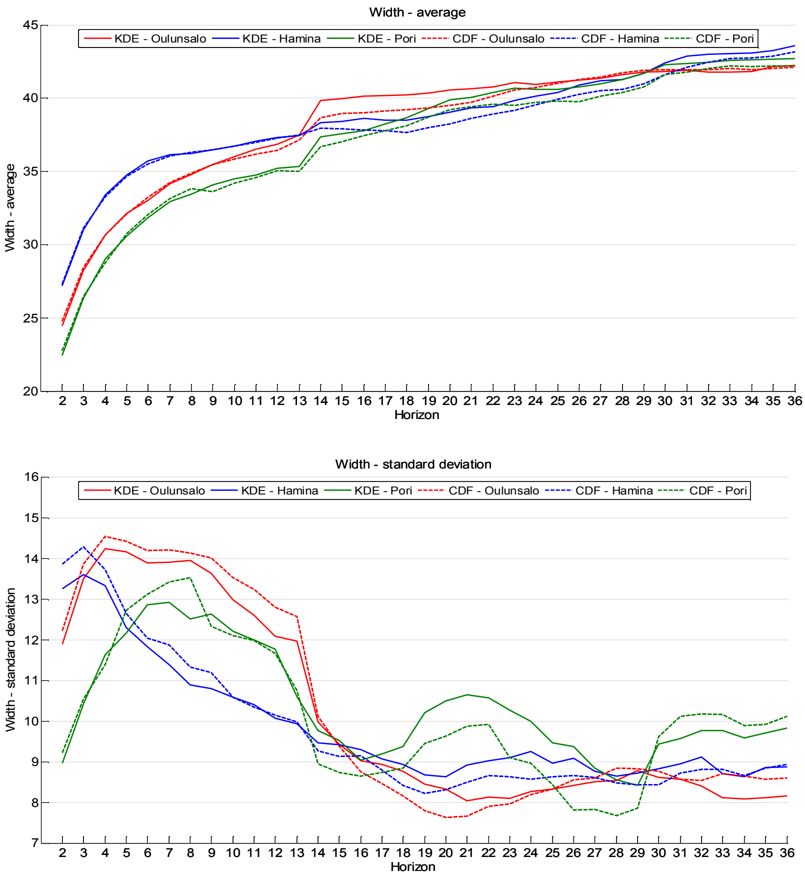


Figure 20. Average length (top) and standard deviation of hourly lengths (bottom) of `CDF` and `KDE` interval forecasts.

6.6 Probability forecasting for aggregated production

In the previous chapter methodology how to draw probabilistic forecasts was introduced. Also, tools to analyse probability intervals were shown. In this chapter similar study will be carried out, however whereas in the previous chapter the weight was given to probability intervals of single wind power plants in this chapter aggregated wind power production is studied. There will be two cases where effects of aggregation are studied. First, aggregated values of three wind power plants, which are the same as in Chapter 6. Second, aggregated values of 30 wind power plants in Finland. However, ensemble forecast is possible to perform only for the three wind farms since the NWP data from FMI is only received from those three points. NWPs from Foreca-ECMWF are available for 30 wind farms. Thus, the averaged point-forecast, AVG and averaged quantile forecast, AVG2 are not possible to calculate when examining probability intervals from 30 wind farms.

The results of aggregation can be seen in Figure 21 to Figure 26. From Figure 21 one can see that the skill-scores are clearly improving (increasing) as a wider area is used for aggregating wind power production. Skill score tries to combine both the accuracy and length of the interval into one metric. Thus, a clear improvement in skill-score will definitely will lead to more accurate forecasting intervals.

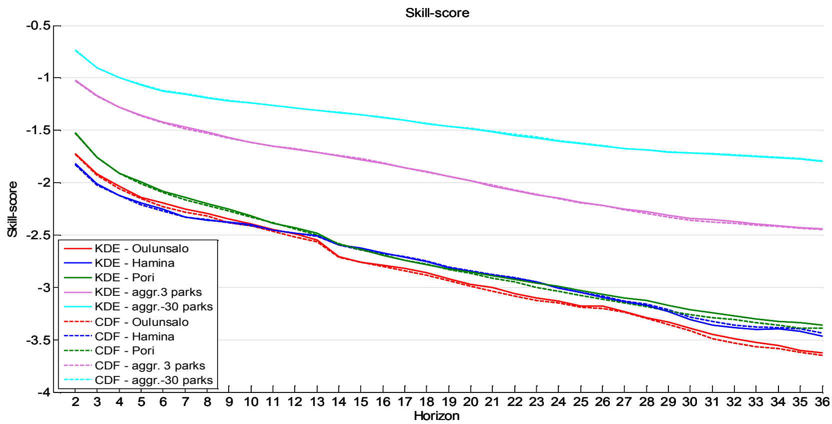


Figure 21. Skill-score for different look ahead hours. KDE stands for Kernel Density Estimation and CDF Cumulative Distribution Function. Aggregated forecasts have higher skill-scores than single turbine sites.

Figure 22 shows how the uncertainty decreases as more wind farms are aggregated together. The average 90% interval width varies between 16–30% of nominal capacity for three aggregated wind farms, and between 11–24% for 30 aggregated wind farms.

It is natural that the interval widths will increase as forecasted further to the future since the uncertainty of wind forecast provided by NWP forecast will increase with increasing forecasting horizon.

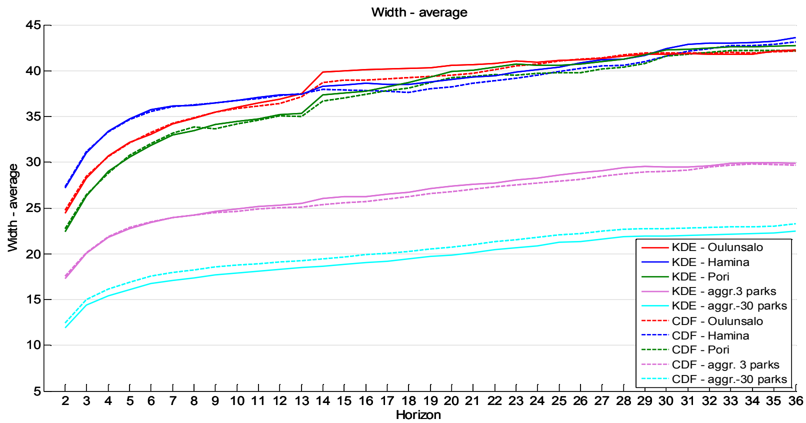


Figure 22. Width of 90% probability interval for different look ahead hours. Aggregated forecasts have smaller intervals than single turbine sites.

Aggregation has also effects on the standard deviation as it is possible to see from Figure 23. Aggregation significantly lowers standard deviation, as interval width is not varying as much as it is for single wind power plants. The main reason for reduced variability of width is spatial smoothing effect, which levels out wind power production. This is quite understandable since for one site power production can vary a lot from time to time and therefore the interval widths must be flexible in order to consider this variability. However, when aggregating more wind turbine sites, the power production is levelled out and thus the interval widths can be more static. The same study would have been done to other than 90% probability interval, but wide probability intervals are usually providing more information than the narrow probability intervals.

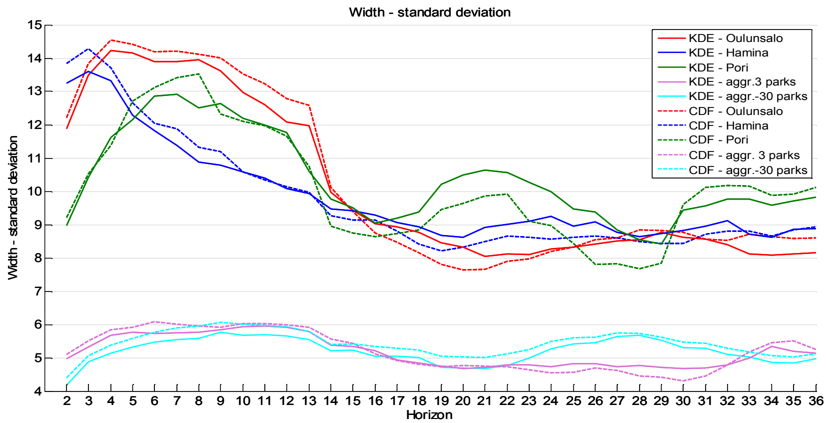


Figure 23. Standard deviation of 90% probability interval. Aggregated forecast forecasts do not have as variable uncertainty interval widths for different weather situation than the single turbine sites.

In Figure 24 to Figure 26 aggregated data from three wind power plant s are used. Average interval widths have a tendency to grow as the forecast horizon increases, see Figure 24. This can be explained so that uncertainty of 2 hours ahead forecast is much lower than the uncertainty of 24 or 36 hours ahead forecast.

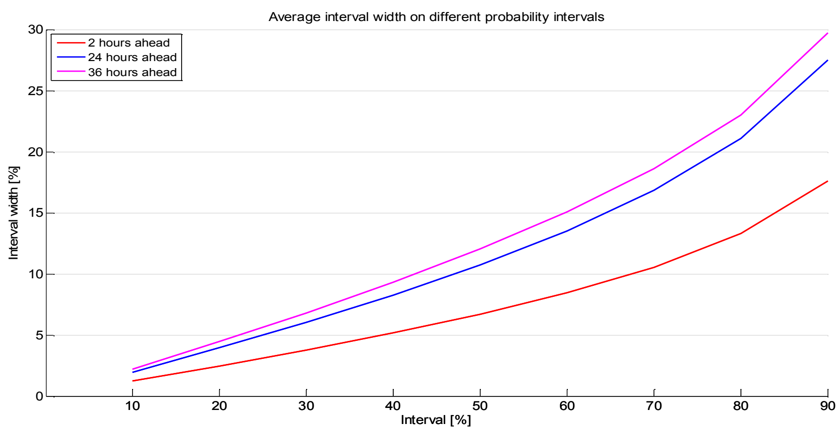


Figure 24. Average interval width on different probability intervals for forecasts 2, 24 and 36 hours ahead for three aggregated sites.

The development of different probability intervals, as the forecast horizon increases, can be seen in Figure 25. Since, the forecast model considers last measured power when making the 36-hour forecast ahead; there is some non-linearity in the couple first hours of forecasting horizon. However, as the horizon increases, its dependency to the last measured power starts to diminish, which happens for forecast horizons 5–6 hours ahead, and after that the interval width increases quasi-linearly. The same phenomena can be also seen from Figure 22.

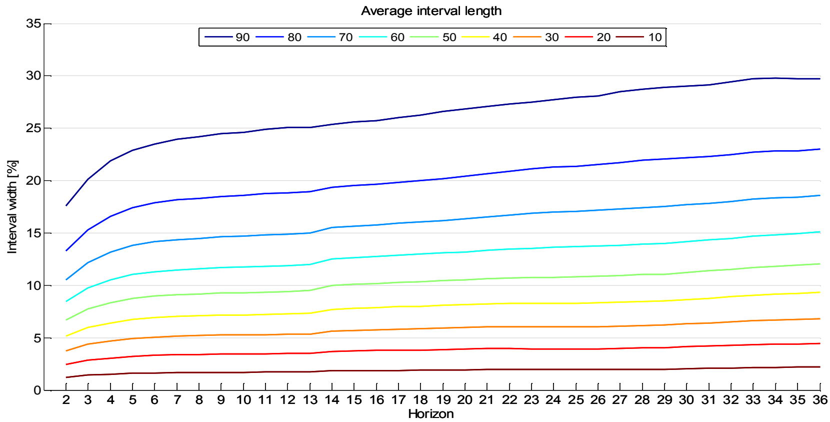


Figure 25. Development of probability interval widths for different probability intervals for three aggregated sites.

In Figure 15 and Figure 26 a probability forecast of wind power is shown. There are many properties, which are important to understand. The following explanation will be made for single turbine, and the same analogy will work for aggregated wind production data but different wind speeds. First of all, since the wind power curve is a strongly a nonlinear process, which has properties that on the middle of the power curve small deviation in wind speed will cause a large deviation in wind turbine's output. Thus, if the wind speed will fluctuate approximately between 4–10 m/s, the output will change in relation of cube of wind speed. However, if the wind speed is lower than 4 m/s or higher than 10 m/s, the wind to power conversion is much flatter in these areas and thus small deviations in wind speed will only cause small deviations in the power output of a wind turbine. When looking at the Figure 15 one can see that the probability distribution is really compact on the first seven look-ahead hours. This is due to two different aspects, first the previous mentioned small variability of wind power plant output in low wind speeds, and secondly, intervals should be smaller on average when forecasting to near future, as it is possible to see from Figure 22. As the wind power forecast increases to 50% of the nominal capacity, the uncertainty increases notably as different probability quantiles starts to spread from the median quantile, which is a sign of increased uncertainty. One could notice from Figure 15 how well the forecasted power production corresponds to actual power production. This is not necessarily always the case. Thus, in Figure 26 another example of interval forecast is shown where forecast has a level error and therefore the actual power is quite frequently outside of the 90% probability interval. Although, the actual power is outside of the 90% probability interval it does not mean that the intervals are necessarily badly adjusted. By the probability theory there should be 10% of the time when the realization of power production should be outside of the 90% probability interval. Therefore, one could ask why not to use 99.9% probability intervals. The answer is simple, by the probability theory if a high reliability level is chosen, upper and lower quantiles are reaching the extreme values most of the time and thus there won't be much information on those intervals

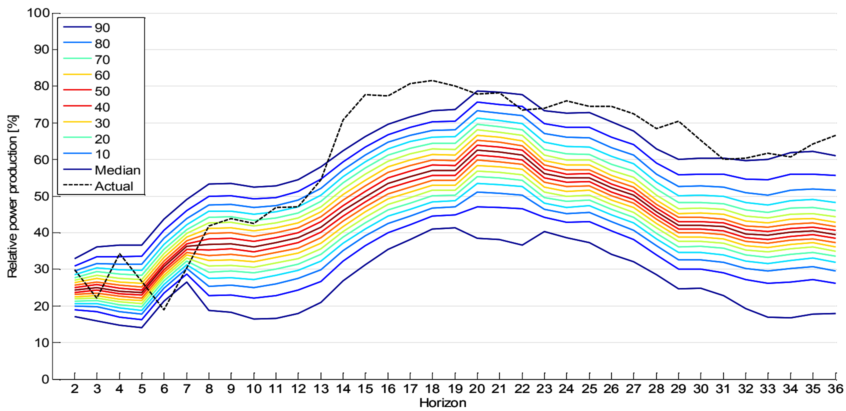


Figure 26. Example of probability forecast for three aggregated sites where there is some level error.

7. Summary and discussion

In this publication short term forecasting of wind power is studied mainly from a wind power producer point of view. The forecast errors and imbalance costs from the day-ahead Nordic electricity markets are calculated based on real data from distributed wind power plants. Improvements to forecasting accuracy are presented using several wind forecast providers, and measures for uncertainty of the forecast are presented.

Imbalance cost calculations have been made for years 2010–2012 for over 20 wind turbine sites in Finland. Predictions are calculated by VTT's wind power prediction model, and the prediction errors were calculated for each site with one hour resolution for day-ahead trading purposes (12–36 hours ahead). The predictions errors were calculated for different sizes of geographical areas, starting from one site and ending with calculating prediction errors for the aggregation of all sites. Aggregation of sites lowers relative share of prediction errors considerably, up to 60%. The balancing costs were also reduced up to 60%, from 3 €/MWh for one site to 1–1.4 €/MWh to aggregate 24 sites. Pooling wind power production for balance settlement will be very beneficial, and larger producers who can have sites from larger geographical area will benefit in lower imbalance costs. The aggregation benefits were already significant for smaller areas, resulting in 30–40% decrease in forecast errors and 13–36% decrease in unit balancing costs, depending on the year. The resulting costs are strongly dependent on Regulating Market prices that determine the prices for the imbalances. Similar level of forecast errors resulted in 40% higher imbalance costs for 2012 compared with 2011. Thus, increasing prices in the Regulating Power Market will impact the imbalance costs in future. The forecasting model development will improve the accuracy, which may slow down the impact that increasing wind power has on the balancing market prices.

The analyses presented will give some insight also from the system operator (TSO) point of view of. It is beneficial to have dispersed wind power production. Also improving the accuracy, and correcting errors before delivery will result in less demand for balancing power from the Regulating power markets.

Improving the accuracy of forecast models is one way to minimize the prediction errors. Using more elaborate forecasting models will also incur a cost to the producer, however, when larger amounts of wind power are predicted, the costs for forecasting can usually be gained by reduced imbalance costs. Another measure to reduce the errors is by dispersing wind power plant sites on wider geographical area to decrease relative prediction errors. Another way to reduce the imbalances from day-ahead forecasts is to use intra-day markets. However, it is not straightforward to decide whether to correct the forecast errors at intraday market Elbas. When wind power share is still low, the forecast errors are penalised only about 50% of the time. This will only be known after the delivery hour and therefore it is possible that a market participant is correcting imbalances, which are not causing any costs in the balancing settlement. The revenue for Finnish wind power producers will not necessarily increase although a market participant places bids to the Elbas-market. For large wind penetration levels, like the case of Denmark, intra-day trading can effectively reduce balancing costs. Probably already at lower shares of wind power, correcting the larger forecast errors in the intra-day market would be cost effective for the producer, and this would also reduce the impact of wind power on the balancing markets and system imbalances

Combining wind forecasts from different Numerical Weather Prediction providers was studied with different combination methods for 6 sites. Averaging different providers' forecasts will lower the forecast errors by 6% for day-ahead purposes where this robust method gave best results. When combining forecasts for short horizons like the following hour, more advanced combining techniques such as Kalmar filtering or recursive least squares provided better results. Of all combination methods, Kalman filter and RLS are most dependent on the delay between consecutive forecasts and the length of the forecast horizon. They work best when the combination weights are allowed to be updated frequently.

Two different uncertainty quantification methods, based on empirical cumulative density function and kernel densities, were analysed for 3 sites. Combination of forecasts was done by averaging separate forecasts. There was no big difference between the methods, but the bias of 90% probability interval was a bit smaller for method that uses kernel densities. Probability densities were created for each prediction horizon separately and it was based on the past performance of deterministic forecasts or prediction errors. Aggregated wind prediction data was also studied, and it lowered 90% probability interval width significantly. Therefore, aggregation of wind power production will not only decrease relative prediction errors, but also decreases the variation and uncertainty of prediction errors.

Acknowledgements

The wind power production data was received from Finnish Energy Industry and Allwinds for sites in Åland. The permission to use more detailed results for the six sites were obtained from wind power producers.

The authors greatly acknowledge the use of the data as this work would not be possible without relying on real production data.

References

1. Turkia, V., Holttinen, H. 2011. Wind energy statistics of Finland. Yearly report 2011. (In Finnish.) VTT Technology 74. ISBN 978-951-38-7909-9. VTT, Espoo, Finland.

2. Ministry of Employment and the Economy (TEM). 2011. EU energy co-operation (http://www.tem.fi). TEM, Helsinki, Finland.

3. Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., Draxl, C. 2011. The State-Of-The-Art in Short-Term Prediction of Wind Power A Literature Overview, 2nd Edition. Project report for the Anemos.plus and SafeWind projects. 110 p. Risø, Roskilde, Denmark.

4. Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., Conzelmann, G. 2009. A Quick Guide to Wind Power Forecasting: State-of-the-Art 2009. Argonne National Laboratory, Argonne, IL.

5. Holttinen, H. 2005. Optimal electricity market for wind power. Energy Policy, Elsevier, 33(16), pp. 2052–2063.

6. Weber, C. 2010. Adequate intraday market design to enable the integration of wind energy into the European power systems. Energy Policy, Vol. 38, Issue 7, pp. 3155–3163. ISSN 0301-4215, http://dx.doi.org/10.1016/j.enpol.2009.07.040. (http://www.sciencedirect.com/science/article/pii/S0301421509005564)

7. Giebel, G., Sørensen, P., Holttinen, H. 2007. Forecast error of aggregated wind power. European Wind Energy Association, Brussels, Belgium.

8. Holttinen, H., Stenberg, A. 2007. Wind power balancing costs for different size actors in the Nordic electricity market. European Wind Energy Association, Brussels, Belgium.

9. Holttinen, H., Saarikivi, P., Repo, S., Ikäheimo, J., Koreneff, G. 2006. Prediction errors and balancing costs for wind power production in Finland. Proceedings of the 6th International Workshop on Large-scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms. Delft, NL, 26–28 Oct. 2006. Pp. 113–120.

10. Bessa, R.J., Miranda, V., Botterud, A., Zhou, Z., Wang, J. 2012. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renewable Energy, 40(1), pp. 29–39. doi:10.1016/j.renene.2011.08.015.

11. Sillanpää, S. 2012. Wind power prediction using multiple weather forecasts. Master’s thesis. Department of Computer Science, University of Helsinki.

12. Bates, J.M., Granger, C.W.J. 1969. The combination of forecasts, Operational. Research Quarterly, 20, pp. 451–468.

13. de Menezes, L., Bunn, D., Taylor, J.W. 2000. Review of guidelines for the use of combined forecasts. European Journal of Operational Research, 120(1), pp. 190–204.

14. Granger, C.W.J., Ramanathan, R. 1984. Improved methods of combining forecasts. Journal of Forecasting, (3)2, pp. 197–204.

15. Sweeney, C., Lynch, P. 2011. Adaptive post-processing of short-term wind forecasts for energy applications. Wind Energy, (14)3, pp. 317–325.

16. Bunn, D. 1975. A Bayesian approach to the linear combination of forecasts. Operational Research Quarterly, (26)2, pp. 325–329.

17. Herbster, M., Warmuth, M. 1998. Tracking the best expert. Machine Learning, (32)2, pp. 151–178.

18. Sanchez, I. 2008. Adaptive combination of forecasts with application to wind energy. International Journal of Forecasting, 24(4), pp. 679–693.

19. Yang, Y. 2004. Combining forecasting procedures: some theoretical results. Econometric Theory, 20, pp. 176–222.

20. Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., Kallos, G. 2008. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. Journal of Wind Engineering and Industrial Aerodynamics, (96)12, pp. 2348–2362.

21. Haykin, S. 1995. Adaptive Filter Theory. Prentice Hall, Upper Saddle River, NJ. 936 p.

22. Crochet, P. 2004. Adaptive Kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in Iceland. Meteorological Applications, 11(2), pp. 173–187.

23. Holttinen, H., Koreneff, G. 2012. Imbalance Cost of Wind Power for a Hydro Power Producer in Finland. Wind Engineering, Vol. 35, No. 1, pp. 53–68. Doi: 10.1260/0309-524X.36.1.53

24. Nielsen, A.H., Madsen H. 2000. Predicting the Heat Consumption in District Heating Systems using Meteorological Forecasts. IMM, DTU, Kgs. Lyngby, Denmark.

Appendix A: More detailed results for forecast errors and balancing costs for different size market players

Table A1. Detailed results of forecast error and costs for different sizes of producers, 2010.

	1 site	8 sites (A)	12 sites (A+S)	18 sites (A+S+P)	24 sites (All)
Pred. Error:					
% prod.	60.7%	42.3%	35.6%	29.7%	29.6%
% time	100.0%	100.0%	100.0%	100.0%	100.0%
Error up:					
% prod.	30.9%	18.8%	15.9%	14.3%	13.8%
% time	64.5%	55.6%	56.0%	57.1%	55.8%
Error down:					
% prod.	29.8%	23.5%	19.7%	15.4%	15.8%
% time	35.5%	44.4%	44.0%	42.9%	44.2%
Up-reg cost:					
% prod.	10.6%	7.7%	6.0%	4.7%	4.6%
% time	11.7%	14.2%	13.5%	13.0%	13.1%
€/MWh, reg	11.07	10.59	10.03	10.51	13.19
Down-reg cost:					
% prod.	10.6%	6.2%	5.1%	4.4%	4.1%
% time	24.1%	20.0%	19.1%	19.3%	17.9%
€/MWh, reg	17.95	13.70	14.64	16.29	18.07
Spot price no					
error €/MWh	55.05	56.24	55.76	55.73	55.81
Balancing cost €/MWh, prod two-price system	**3.07**	**1.66**	**1.34**	**1.21**	**1.34**
Net income €/MWh	51.97	54.59	54.41	54.52	54.47
MAE of energy	**0.61**	**0.42**	**0.36**	**0.30**	**0.30**

Table A2. Detailed results of forecast error and costs for different sizes of producers, 2011.

	1 site	9 sites (A)	13 sites (A+S)	18 sites (A+S+P)	25 sites (All)
Pred. Error:					
% prod.	56.3%	33.7%	25.8%	23.6%	20.9%
% time	100.0%	100.0%	100.0%	100.0%	100.0%
Error up:					
% prod.	33.2%	17.0%	11.3%	8.1%	9.8%
% time	65.2%	56.1%	51.6%	45.1%	52.1%
Error down:					
% prod.	23.1%	16.8%	14.5%	15.5%	11.2%
% time	34.8%	43.9%	48.4%	54.9%	47.9%
Up-reg cost:					
% prod.	6.1%	5.0%	4.3%	4.5%	3.7%
% time	10.2%	13.6%	15.5%	17.4%	15.9%
€/MWh, reg	11.33	14.25	15.44	13.10	13.78
Down-reg cost:					
% prod.	15.9%	7.5%	5.5%	3.6%	4.8%
% time	29.2%	24.9%	23.2%	19.1%	23.5%
€/MWh, reg	11.23	10.62	10.86	10.78	10.18
Spot price no					
error €/MWh	47.31	48.94	48.70	47.97	48.42
Balancing cost €/MWh, prod two-price system	**2.47**	**1.51**	**1.26**	**0.99**	**1.00**
Net income €/MWh	44.84	47.44	47.44	46.98	47.43
MAE of energy	**0.56**	**0.34**	**0.26**	**0.24**	**0.21**

Table A3. Detailed results of forecast error and costs for different sizes of producers, 2012.

	1 site	9 sites (A)	11 sites (A+S)	15 sites (A+S+P)	23 sites (All)
Pred. Error:					
% prod.	52.7%	33.9%	27.4%	23.5%	20.5%
% time	100.0%	100.0%	100.0%	100.0%	100.0%
Error down:					
% prod.	24.3%	17.6%	16.7%	12.7%	11.7%
% time	59.9%	56.3%	61.0%	57.9%	59.3%
Error up:					
% prod.	28.4%	16.3%	10.7%	10.8%	8.8%
% time	40.1%	43.7%	39.0%	42.1%	40.7%
Up-reg cost:					
% prod.	7.3%	5.2%	3.3%	3.2%	2.4%
% time	11.1%	13.2%	11.7%	12.3%	11.5%
€/MWh. reg	25.38	37.44	38.87	34.65	37.01
Down-reg cost:					
% prod.	11.1%	7.3%	6.9%	5.2%	4.7%
% time	24.6%	23.0%	24.9%	23.4%	23.8%
€/MWh. reg	11.48	10.84	11.56	11.32	12.21
Spot price no					
error €/MWh	37.13	37.34	37.26	37.09	36.99
Balancing cost €/MWh, prod two-price system	**3.14**	**2.72**	**2.09**	**1.69**	**1.45**
Net income €/MWh	33.99	34.62	35.18	35.41	35.54
MAE of energy	0.53	0.34	0.27	0.24	0.20

Appendix B: More detailed results for forecast combinations

Table B1. Monthly power forecasts were produced by training the prediction system separately for each month using three months of data.

Month	Forecast period	Training period
	(month/day/year)	
9/2011	09/01/2011 – 09/30/2011	10/01/2011 – 12/31/2011
10/2011	10/01/2011 – 10/31/2011	09/01/2011 – 09/30/2011 & 11/01/2011 – 12/31/2011
11/2011	11/01/2011 – 11/30/2011	09/01/2011 – 10/31/2011 & 12/01/2011 – 12/31/2011
12/2011	12/01/2011 – 12/31/2011	09/01/2011 - 11/30/2011

Table B2. Normalized root mean squared errors (NRMSE) for all wind power plants during the evaluation period. Monthly error variances in parentheses. Results for Kalman filter are not included when the errors are very large.

:: 12-36 hours ahead:

Method	1	2	3	4	5	6	mean
NWP-1	12,51 (3,57)	16,60	18,89	18,01	20,45	21,59	18,01
NWP-2	11,81	16,42	- (-)	- (-)	- (-)	18,09	15,44
NWP-3	- (-)	- (-)	18,77	17,43	20,48	- (-)	18,89
AVG	11,41 (2,59)	15,90	17,36	16,52	19,27	18,19	16,44
RLS	14,11 (5,57)	18,97	20,69	19,17	22,05	19,71	19,12
RLS-NO	12,45 (2,55)	17,52	19,63	19,41	20,32	19,60	18,15
REG	11,61 (2,13)	18,81	17,70	16,60	19,62	17,91	17,04
REG-NO	11,56 (2,34)	16,99	17,68	16,68	19,63	18,05	16,76
OPT	16,97 (4,64)	16,13	20,00	18,25	19,29	21,19	18,64
OPT-IND	11,68 (3,24)	15,96	17,77	16,71	19,28	18,00	16,57
OUT	11,51 (2,46)	15,90	17,35	16,60	19,37	18,30	16,50
COM	11,54 (2,98)	15,87	17,59	16,54	19,23	17,86	16,44
SHARE	11,37 (2,50)	15,90	17,40	16,51	19,25	18,20	16,44
AEC	11,40 (2,29)	15,90	17,43	16,58	19,51	18,36	16,53

:: Next hour:

Method	1	2	3	4	5	6	mean
NWP-1	10,57	13,74	15,34	15,86	14,87	17,25	14,60
NWP-2	10,51	13,89	- (-)	- (-)	- (-)	16,30	13,57
NWP-3	- (-)	- (-)	15,85	16,32	15,41	- (-)	15,86
AVG	10,42 (1,19)	13,68	15,27	15,73	14,94	16,31	14,39
KALMAN	8,86 (1,89)	11,23	11,48	10,67	11,12	12,62	11,00
KALMAN-	11,93	27,45	11,95	12,39	16,68	13,44	15,64
RLS	9,08 (0,75)	12,35	12,87	14,12	12,25	14,04	12,45
RLS-NO	9,78 (0,57)	13,35	13,63	15,05	13,19	14,65	13,27
REG	10,38	13,67	15,2Z3	15,69	14,73	16,14	14,31
REG-NO	10,43 (1,29)	13,78	15,25	15,74	14,80	16,21	14,37
OPT	12,12 (4,07)	14,56	17,40	17,07	15,18	16,84	15,53
OPT-IND	10,49 (1,40)	13,77	15,38	16,02	15,03	16,27	14,50
OUT	10,46 (1,21)	13,69	15,29	15,80	14,95	16,39	14,43
COM	10,48 (1,36)	13,74	15,34	15,91	14,98	16,25	14,45
SHARE	10,14 (1,06)	13,53	14,72	15,19	14,50	15,64	13,95
AEC	10,34 (1,19)	13,62	15,06	15,49	14,80	15,96	14,21

:: 12 hours ahead:

Method	1	2	3	4	5	6	mean
NWP-1	11,82 (2,43)	16,17	16,60	17,31	19,10	20,14	16,85
NWP-2	11,57	15,92	- (-)	- (-)	- (-)	17,10	14,86
NWP-3	- (-)	- (-)	17,31	17,34	19,71	- (-)	18,12
AVG	11,12 (1,77)	15,61	15,76	16,40	18,40	17,21	15,75
RLS	13,41 (5,18)	17,45	16,44	18,16	19,58	21,20	17,71
RLS-NO	11,92	16,50	15,93	17,53	17,97	20,16	16,67
REG	11,37 (1,41)	17,63	15,88	16,63	18,69	17,23	16,24
REG-NO	11,32	16,30	15,86	16,63	18,68	17,07	15,98
OPT	15,80	18,31	17,95	16,98	19,85	18,39	17,88
OPT-IND	11,46 (2,33)	15,81	16,52	16,79	18,40	17,10	16,01
OUT	11,23 (1,61)	15,73	15,89	16,51	18,46	17,27	15,85
COM	11,30 (2,11)	15,72	16,19	16,60	18,37	17,04	15,87
SHARE	11,09	15,61	15,84	16,42	18,56	17,25	15,79
AEC	11,08 (1,57)	15,63	16,08	16,55	18,89	17,65	15,98

Method	1	2	3	4	5	6	mean
NWP-1	13,10	17,90	19,89	19,38	21,88	22,90	19,18
NWP-2	11,77	17,32	- (-)	- (-)	- (-)	18,91	16,00
NWP-3	- (-)	- (-)	19,02	17,40	20,86	- (-)	19,09
AVG	11,61 (2,52)	16,96	17,97	16,88	20,03	18,84	17,05
RLS	14,24	22,14	24,28	20,58	20,25	25,34	21,14
RLS-NO	12,17 (3,32)	20,12	18,91	20,25	19,34	22,05	18,81
REG	11,68 (2,35)	19,48	18,24	16,89	20,62	18,58	17,58
REG-NO	11,64 (2,55)	18,22	18,16	17,14	20,65	18,78	17,43
OPT	18,38	17,70	21,33	18,16	20,50	22,12	19,70
OPT-IND	12,41 (1,70)	16,80	18,93	16,75	20,49	19,05	17,40
OUT	11,75	17,14	18,03	17,03	20,09	19,08	17,19
COM	12,16 (1,63)	16,84	18,60	16,67	20,27	18,88	17,24
SHARE	11,57	16,98	17,90	16,85	20,08	18,68	17,01
AEC	11,61 (2,24)	17,27	18,03	17,10	20,33	18,90	17,21

Table B3. Percentage changes in forecast errors (NRMSE). Power forecasts using a single weather forecast (NWP-1, NWP-2 and NWP-3) and the remaining combination methods are compared against averaging. Results for Kalman filter are not included when the errors are very large.

Daily updated forecasts
(same weights for all horizons)

:: 12-36 hours ahead:

			site:			
Method	**1**	**2**	**3**	**4**	**5**	**6**
NWP-1	8,79	4,25	8,11	8,26	5,78	15,75
NWP-2	3,4	3,2	-	-	-	-0,54
NWP-3	-	-	7,53	5,2	5,89	-
AVG	0	0	0	0	0	0
RLS	23,69	19,36	19,21	16,03	14,42	8,36
RLS-NO	9,08	10,18	13,09	17,47	5,46	7,76
REG	1,74	18,31	1,99	0,45	1,8	-1,55
REG-NO	1,27	6,91	1,86	0,98	1,84	-0,78
OPT	48,76	1,49	15,25	10,48	0,12	16,51
OPT-IND	2,33	0,41	2,38	1,16	0,05	-1,01
OUT	0,84	0,01	-0,06	0,49	0,49	0,64
COM	1,16	-0,18	1,37	0,12	-0,19	-1,8
SHARE	-0,4	0,03	0,27	-0,04	-0,11	0,06
AEC	-0,13	-0,01	0,39	0,38	1,23	0,93

Hourly updated forecasts
(separate weights for all hours and horizons)

:: Next hour:

Method	**1**	**2**	**3**	**4**	**5**	**6**
NWP-1	1,37	0,5	0,45	0,78	-0,52	5,45
NWP-2	0,87	1,53	-	-	-	-0,05
NWP-3	-	-	3,64	3,59	3,01	-
AVG	0	0	0	0	0	0
KALMAN	-14,97	-17,85	-24,83	-32,17	-25,58	-22,63
KALMAN-NO	14,47	100,73	-21,77	-21,23	11,64	-17,57
RLS	-12,85	-9,68	-15,73	-10,24	-18,01	-13,91
RLS-NO	-6,18	-2,41	-10,78	-4,38	-11,71	-10,19
REG	-0,39	-0,08	-0,3	-0,29	-1,41	-1,04
REG-NO	0,06	0,79	-0,14	0,02	-0,96	-0,61
OPT	16,34	6,43	13,89	8,48	1,56	3,3
OPT-IND	0,7	0,72	0,7	1,82	0,6	-0,22
OUT	0,34	0,1	0,1	0,42	0,03	0,53
COM	0,52	0,47	0,45	1,08	0,28	-0,32
SHARE	-2,72	-1,04	-3,66	-3,46	-2,99	-4,11
AEC	-0,77	-0,43	-1,38	-1,53	-0,98	-2,11

:: 12 hours ahead:

Method	**1**	**2**	**3**	**4**	**5**	**6**
NWP-1	5,96	3,44	5,03	5,24	3,63	14,52
NWP-2	3,92	1,92	-	-	-	-0,68
NWP-3	-	-	8,92	5,41	6,63	-
AVG	0	0	0	0	0	0
RLS	20,6	11,79	4,29	10,72	6,38	23,17
RLS-NO	7,21	5,72	1,04	6,9	-2,34	17,1
REG	2,3	12,93	0,72	1,39	1,54	0,12
REG-NO	1,84	4,43	0,58	1,4	1,49	-0,85
OPT	42,12	17,26	13,9	3,57	7,86	6,82
OPT-IND	3,1	1,26	4,79	2,37	-0,01	-0,65

OUT	0,98	0,75	0,82	0,65	0,31	0,32
COM	1,69	0,73	2,68	1,2	-0,16	-0,98
SHARE	-0,22	-0,01	0,51	0,12	0,83	0,2
AEC	-0,36	0,14	1,99	0,92	2,64	2,52

:: 36 hours ahead:

Method	1	2	3	4	5	6
NWP-1	11,31	5,24	9,65	12,92	8,46	17,75
NWP-2	1,33	2,05	-	-	-	0,42
NWP-3	-	-	5,53	3,01	3,95	-
AVG	0	0	0	0	0	0
RLS	22,62	30,52	35,1	21,95	1,08	34,54
RLS-NO	4,77	18,6	5,21	20,01	-3,45	17,09
REG	0,58	14,85	1,48	0,09	2,93	-1,36
REG-NO	0,19	7,43	1,05	1,57	3,09	-0,28
OPT	58,26	4,34	18,67	7,58	2,35	17,46
OPT-IND	6,87	-0,96	5,32	-0,78	2,29	1,15
OUT	1,2	1,06	0,31	0,88	0,28	1,29
COM	4,68	-0,69	3,51	-1,26	1,17	0,26
SHARE	-0,35	0,12	-0,39	-0,17	0,22	-0,83
AEC	-0,05	1,82	0,33	1,33	1,48	0,32

Table B4. Forecast errors (NRMSE) when varying the combination members (C1, C2, etc). Results for Kalman filter are not included when the errors are very large.

Daily updated forecasts (same weights for all horizons)

:: 12-36 hours ahead:

			site 4:					site 5:		
Method	**C1**	**C2**	**C3**	**C4**	**mean**	**C1**	**C2**	**C3**	**C4**	**mean**
NWP-1	18,01	18,01	18,01	-	18,01	20,45	20,45	20,45	-	20,45
NWP-2	17,16	-	17,16	17,16	17,16	20,38	-	20,38	20,38	20,38
NWP-3	17,43	17,43	-	17,43	17,43	20,48	20,48	-	20,48	20,48
AVG	16,08	16,52	16,34	16,51	16,36	19,06	19,27	19,28	19,68	19,32
RLS	19,79	19,17	19,5	18,33	19,2	21,8	22,05	20,03	21,64	21,38
RLS-NO	21,69	19,41	19,99	17,2	19,57	20,13	20,32	18,76	19,45	19,67
REG	16,34	16,6	16,54	16,7	16,54	19,48	19,62	19,61	20,07	19,69
REG-NO	16,4	16,68	16,59	16,77	16,61	19,43	19,63	19,59	19,99	19,66
OPT	19,36	18,25	18,95	19,54	19,03	19,03	19,29	19,34	19,7	19,34
OPT-IND	16,59	16,71	17,02	16,69	16,75	19,05	19,28	19,34	19,75	19,36
OUT	16,17	16,6	16,46	16,52	16,44	19,08	19,37	19,27	19,69	19,35
COM	16,33	16,54	16,79	16,56	16,56	19,02	19,23	19,28	19,7	19,31
SHARE	16,13	16,51	16,4	16,49	16,38	18,97	19,25	19,29	19,64	19,29
AEC	16,12	16,58	16,52	16,46	16,42	19,25	19,51	19,53	19,8	19,52

Hourly updated forecasts (separate weights for all hours and horizons)

:: Next hour:

			site 4:					site 5:		
Method	**C1**	**C2**	**C3**	**C4**	**mean**	**C1**	**C2**	**C3**	**C4**	**mean**
NWP-1	15,86	15,86	15,86	-	15,86	14,87	14,87	14,87	-	14,87
NWP-2	15,84	-	15,84	15,84	15,84	14,92	-	14,92	14,92	14,92
NWP-3	16,32	16,32	-	16,32	16,32	15,41	15,41	-	15,41	15,41
AVG	15,6	15,73	15,52	15,83	15,67	14,81	14,94	14,71	14,97	14,86
KALMAN	11,02	10,67	10,67	10,25	10,65	11,32	11,12	10,98	11,55	11,24
KALMAN-NO	12,59	12,39	12,3	12,94	12,56	11,68	16,68	11,15	11,6	12,78
RLS	14,79	14,12	14,17	14,54	14,41	12,99	12,25	12,25	12,53	12,51
RLS-NO	15,37	15,05	14,74	15,12	15,07	13,72	13,19	13,17	13,52	13,4

REG	15,54	15,69	15,52	15,73	15,62	14,62	14,73	14,61	14,77	14,68
REG-NO	15,59	15,74	15,57	15,79	15,67	14,68	14,8	14,67	14,83	14,75
OPT	16,92	17,07	18,16	16,14	17,07	14,89	15,18	14,81	14,99	14,97
OPT-IND	15,78	16,02	15,79	15,82	15,85	14,85	15,03	14,71	14,99	14,9
OUT	15,6	15,8	15,53	15,94	15,72	14,81	14,95	14,8	14,97	14,88
COM	15,72	15,91	15,69	15,82	15,78	14,83	14,98	14,7	14,98	14,87
SHARE	14,89	15,19	15,04	15,47	15,15	14,25	14,5	14,37	14,73	14,46
AEC	15,32	15,49	15,3	15,72	15,46	14,67	14,8	14,62	14,89	14,74

:: 12 hours ahead:

Method	C1	C2	C3	C4	mean	C1	C2	C3	C4	mean
NWP-1	17,31	17,31	17,31	-	17,31	19,1	19,1	19,1	-	19,1
NWP-2	16,27	-	16,27	16,27	16,27	19,15	-	19,15	19,15	19,15
NWP-3	17,34	17,34	-	17,34	17,34	19,71	19,71	-	19,71	19,71
AVG	15,79	16,4	15,75	16,12	16,01	18,18	18,4	18,22	18,8	18,4
RLS	21,84	18,16	22,86	17,74	20,15	19,53	19,58	19,68	20,39	19,8
RLS-NO	21,56	17,53	22,64	17,04	19,69	17,93	17,97	17,87	18,29	18,02
REG	15,93	16,63	15,93	16,28	16,19	18,47	18,69	18,46	18,95	18,64
REG-NO	15,95	16,63	15,94	16,25	16,19	18,45	18,68	18,42	18,93	18,62
OPT	17,96	16,98	18,72	17,93	17,9	18,35	19,85	18,83	18,92	18,99
OPT-IND	16,2	16,79	16,2	16,18	16,34	18,21	18,4	18,21	18,9	18,43
OUT	16,06	16,51	15,86	16,2	16,16	18,24	18,46	18,26	18,87	18,46
COM	16,02	16,6	16,02	16,13	16,19	18,17	18,37	18,18	18,85	18,39
SHARE	15,81	16,42	15,71	16,14	16,02	18,31	18,56	18,23	18,82	18,48
AEC	15,95	16,55	15,86	16,3	16,16	18,51	18,89	18,44	18,94	18,69

:: 36 hours ahead:

Method	C1	C2	C3	C4	mean	C1	C2	C3	C4	mean
NWP-1	19,38	19,38	19,38	-	19,38	21,88	21,88	21,88	-	21,88
NWP-2	17,77	-	17,77	17,77	17,77	21,44	-	21,44	21,44	21,44
NWP-3	17,4	17,4	-	17,4	17,4	20,86	20,86	-	20,86	20,86
AVG	16,43	16,88	17,06	16,74	16,78	19,85	20,03	20,36	20,35	20,15
RLS	19,51	20,58	17,87	19,92	19,47	32,39	20,25	24,1	25,55	25,57
RLS-NO	19,35	20,25	17,6	19,58	19,2	24,03	19,34	24,65	22,41	22,61
REG	16,6	16,89	17,16	17,01	16,92	20,45	20,62	20,82	20,81	20,68
REG-NO	16,8	17,14	17,36	17,09	17,1	20,43	20,65	20,81	20,79	20,67
OPT	24,12	18,16	20,93	22,37	21,39	20,84	20,5	20,51	20,77	20,66
OPT-IND	16,58	16,75	17,12	17,01	16,86	20,2	20,49	20,38	20,71	20,44
OUT	16,69	17,03	17,34	16,76	16,95	19,92	20,09	20,43	20,43	20,22
COM	16,39	16,67	16,96	16,85	16,72	20,02	20,27	20,36	20,56	20,3
SHARE	16,43	16,85	17,06	16,75	16,77	19,88	20,08	20,37	20,38	20,18
AEC	16,81	17,1	17,42	17,05	17,1	20,06	20,33	20,5	20,52	20,35

Series title and number

VTT Technology 95

Title	**Wind power forecasting accuracy and uncertainty in Finland**
Author(s)	Hannele Holttinen, Jari Miettinen & Samuli Sillanpää
Abstract	Wind power cannot be dispatched so the production levels need to be forecasted for electricity market trading. Lower prediction errors mean lower regulation balancing costs, since relatively less energy needs to go through balance settlement. From the power system operator point of view, wind power forecast errors will impact the system net imbalances when the share of wind power increases, and more accurate forecasts mean less regulating capacity will be activated from the real time Regulating Power Market. In this publication short term forecasting of wind power is studied mainly from a wind power producer point of view. The forecast errors and imbalance costs from the day-ahead Nordic electricity markets are calculated based on real data from distributed wind power plants. Improvements to forecasting accuracy are presented using several wind forecast providers, and measures for uncertainty of the forecast are presented. Aggregation of sites lowers relative share of prediction errors considerably, up to 60%. The balancing costs were also reduced up to 60%, from 3 €/MWh for one site to 1–1.4 €/MWh to aggregate 24 sites. Pooling wind power production for balance settlement will be very beneficial, and larger producers who can have sites from larger geographical area will benefit in lower imbalance costs. The aggregation benefits were already significant for smaller areas, resulting in 30–40% decrease in forecast errors and 13–36% decrease in unit balancing costs, depending on the year. The resulting costs are strongly dependent on Regulating Market prices that determine the prices for the imbalances. Similar level of forecast errors resulted in 40% higher imbalance costs for 2012 compared with 2011. Combining wind forecasts from different Numerical Weather Prediction providers was studied with different combination methods for 6 sites. Averaging different providers' forecasts will lower the forecast errors by 6% for day-ahead purposes. When combining forecasts for short horizons like the following hour, more advanced combining techniques than simple average, such as Kalmar filtering or recursive least squares provided better results. Two different uncertainty quantification methods, based on empirical cumulative density function and kernel densities, were analysed for 3 sites. Aggregation of wind power production will not only decrease relative prediction errors, but also decreases the variation and uncertainty of prediction errors.
ISBN, ISSN	ISBN 978-951-38-7985-3 (URL: http://www.vtt.fi/publications/index.jsp) ISSN-L 2242-1211 ISSN 2242-122X (Online)
Date	April 2013
Language	English
Pages	60 p. + app. 8 p.
Keywords	Wind power, wind energy, forecasting, uncertainty, electricity market, imbalance costs
Publisher	VTT Technical Research Centre of Finland P.O. Box 1000, FI-02044 VTT, Finland, Tel. +358 20 722 111

VTT Technical Research Centre of Finland is a globally networked multitechnological contract research organization. VTT provides high-end technology solutions, research and innovation services. We enhance our customers' competitiveness, thereby creating prerequisites for society's sustainable development, employment, and wellbeing.

Turnover: EUR 300 million
Personnel: 3,200

VTT publications

VTT employees publish their research results in Finnish and foreign scientific journals, trade periodicals and publication series, in books, in conference papers, in patents and in VTT's own publication series. The VTT publication series are VTT Visions, VTT Science, VTT Technology and VTT Research Highlights. About 100 high-quality scientific and professional publications are released in these series each year. All the publications are released in electronic format and most of them also in print.

VTT Visions

This series contains future visions and foresights on technological, societal and business topics that VTT considers important. It is aimed primarily at decision-makers and experts in companies and in public administration.

VTT Science

This series showcases VTT's scientific expertise and features doctoral dissertations and other peer-reviewed publications. It is aimed primarily at researchers and the scientific community.

VTT Technology

This series features the outcomes of public research projects, technology and market reviews, literature reviews, manuals and papers from conferences organised by VTT. It is aimed at professionals, developers and practical users.

VTT Research Highlights

This series presents summaries of recent research results, solutions and impacts in selected VTT research areas. Its target group consists of customers, decision-makers and collaborators.

VTT TECHNOLOGY 95 Wind power forecasting accuracy and uncertainty in Finland

Wind power forecasting accuracy and uncertainty in Finland

In this publication short term forecasting of wind power is studied mainly from a wind power producer point of view. The forecast errors and imbalance costs from the day-ahead Nordic electricity markets are calculated based on real data from distributed wind power plants. Improvements to forecasting accuracy are presented using several wind forecast providers, and measures for uncertainty of the forecast are presented.

ISBN 978-951-38-7985-3 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN-L 2242-1211
ISSN 2242-122X (Online)