PSO (FU 2101)
Ensemble-forecasts for wind power

Wind Power Ensemble Forecasting Using Wind Speed and Direction Ensembles from ECMWF or NCEP

Henrik Aalborg Nielsen[1], Henrik Madsen[1], Torben Skov Nielsen[1],
Jake Badger[2], Gregor Giebel[2], Lars Landberg[2]
Kai Sattler[3], Henrik Feddersen[3]

[1] Technical University of Denmark, Informatics and Mathematical Modelling, DK-2800 Lyngby
[2] Risø National Laboratory, Wind Energy Department, DK-4000 Roskilde
[3] Danish Meteorological Institute, Research and Development, DK-2100 København Ø

28th September 2005

Contents

1 Summary

Meteorological ensemble forecasts aim at quantifying the uncertainty of a forecast by offering several scenarios of the future development of the weather. Ideally, we would think of the ensembles as samples from a probability distribution function reflecting the uncertainty of the unperturbed forecast. In this report we address the problems of (i) transforming the meteorological ensembles to wind power ensembles and, (ii) correcting the ensemble quantiles to allow a probabilistic interpretation.

The methods are applied to two wind farms using ensembles from the European Centre for Medium-Range Weather Forecasts (ECMWF) or National Center for Environmental Prediction (NCEP) in the U.S.İt is shown that the resulting ensemble quantiles are able to distinguish between situations with low and high uncertainty. Furthermore, often the quantiles indicate an uncertainty which is significantly smaller than the uncertainty which follow from historical (climatological) data. However, quite often the actual wind power production is outside the range of ensemble forecast and therefore it is not possible to obtain information regarding the extreme quantiles. A spread/skill analysis were performed. From this it is concluded that the spread of the quantiles is indeed a good indicator for the actual uncertainty of the forecast.

From the results presented in this report it is not clear if there is any benefit of one ensemble prediction system over the other. However, some strange behaviour is observed for adjusted quantiles based on NCEP ensembles. Hence, we cannot advice to use NCEP ensembles for horizons[1] below 24 hours. Also for one wind farm (Hagesholm) some strange behaviour is observed for a very small range of horizons just over 36 hours. Also, due to the limited number of NCEP-ensembles (11) compared to ECMWF (51) NCEP-ensembles is most suited when only the central part of the probability distribution is required.

2 Introduction

In recent years a growing interest in information about the uncertainty of wind power forecasts in different weather situations has emerged. Based on wind speed measurements and standard meteorological forecasts Bremnes (2002) estimates the power curve of a small wind farm and then models the relation between the actual and forecasted wind speed both with respect to the mean and the variance (including correlation). Pinson and Kariniotakis (2003) use consecutive forecasts as ensembles and based on these a quantity called "meteo-risk" is defined. This quantity measures the agreement between consecutive forecasts and is used to predict the uncertainty of the wind power forecast. Lange and Heinemann (2003) identify relations between typical weather situations and the magnitude

[1] Time from initialization of the meteorological model until the time-point for which the forecast is valid.

of the forecast error. In a research project carried out together with the Eltra (TSO in western Denmark) Nielsen and Madsen (2002) developed a stochastic model describing variance and correlation of the forecast errors when using WPPT, version 2 (Nielsen et al., 2000). In this report we approach the problem of supplying situation specific information about the uncertainty of wind power forecasts by use of meteorological ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF). However, the approach is not limited to this particular ensemble prediction system. Although the approach is applied to wind farms it does not use local measurements of wind speed and therefore it is also applicable to geographical regions at least of sizes comparable to the spatial resolution of the meteorological model.

Meteorological ensemble forecasting attempts to quantify the uncertainty of short and medium range weather forecasts by producing several forecasts generated under slightly different initial conditions or using slightly different meteorological models. As just one example consider e.g. the wind speed ensemble forecast from ECMWF shown in Figure 1, where the initialization corresponds to 12:00 (UTC) at Aug. 14, 2003. The left panel of Figure 1 shows 51 possible scenarios or ensemble members of the wind speed development from the initialization of the model and 7 days ahead. It is seen that up to three days ahead the ensembles are quite similar and from thereon the ensembles seems to diverge (this is of cause dependent on the specific weather situation). Note also that although the spread seems small around day 2 the impact on the power production may be quite high.

From the plot of the individual ensembles it is difficult to deduce quantitative information. Such information can be obtained by plotting quantiles as shown in the right panel of the figure. For operational use such quantiles should be correct in a probabilistic sense, e.g. in the long run the 90% quantile should be exceeded by the actual wind speed in 10% of the cases. The quantiles do not offer information about the autocorrelation. Such correlation may be important e.g. when combining wind and hydro systems or in other energy systems where direct or indirect storage of wind power is possible. However, in this report we will focus on the correctness of the ensemble quantiles after transformation to ensembles of power production.

As described in Section 3 the spatial resolution of the meteorological models is 40-80 km. Hence it cannot be expected that the ensemble quantiles are correct in a probabilistic sense when compared to a wind speed measurement at a point or when ensembles of power production are compared to the actual production of a wind farm.

The outline of the report is as follows. Section 3 describes the data used. Estimation of power curves with special emphasis on issues relating to ensemble forecasting are described in Section 4 and the probabilistic properties of the wind power ensembles are addressed. In Section 5 a method for correcting the ensemble quantiles is presented. The quality of the ensemble forecasts, using ECMWF- or NCEP-ensembles, are addressed in Section 6. Finally in Section 7 we draw conclusions and discuss the results.

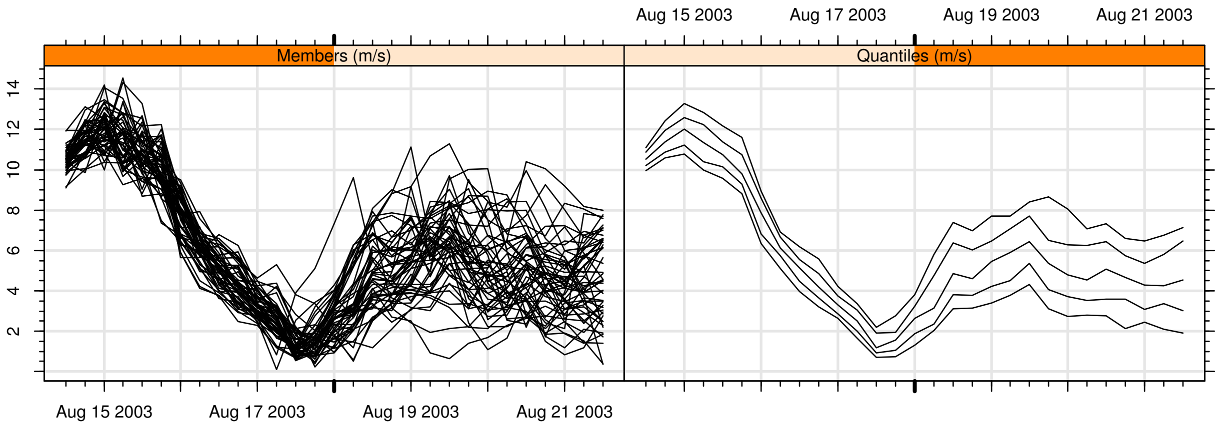


Figure 1: Individual ECMWF wind speed ensemble members and quantiles (10, 25, 50, 75, 90 percent) for the 7 day forecast starting at Aug. 14, 2003, 12:00 (UTC).

3 Data

The data used in this study consists of the wind power production of two wind farms in Denmark together with ensemble forecasts of wind speed and direction. More specifically:

- 15 min. power averages from the Tunø Knob offshore wind farm consisting of 10 Vestas V39 turbines (500kW nominal). Location (UTM Zone 32): 661100E, 6182700N.
- 60 min. power averages from the Hagesholm wind farm consisting of 6 NM-2000 NEG Micron turbines (2000kW nominal). Location (UTM Zone 32): 584700E, 6203700N.
- ECMWF (European Centre for Medium-Range Weather Forecasts) ensembles of wind speed and direction 10m a.g.l. (above ground level) interpolated to the locations of the wind farms. The temporal resolution of the forecasts output is 6 hours. The model is a spectral global model giving a horizontal resolution of approximately 80 km. The calculations are initiated every day at 12:00 (UTC) and the horizon is 10 days. The forecasts are available approximately 17 hours after the initial time. The ensemble consists of one unperturbed forecast and 25 pairs of forecasts for which the initial conditions are perturbed in the positive and negative direction of vectors based on linear combinations of singular vectors (Molteni et al., 1996). Furthermore, for each model run attempts are made to account for sub-grid processes by use of stochastic physics (Buizza et al., 1999). To reduce the storage-requirement it was decided to follow Buizza (2002) and store forecasts for horizons up to 7 days only.
- The NCEP (National Center for Environmental Prediction) ensemble system is based on the NCEP GFS (Global Forecast System) spectral model (Sela, 1980). The system is currently run operationally 4 times daily out to a lead time of 384

hours. The deterministic (unperturbed) forecast is integrated at a spectral truncation of T382 (equivalent to around 40km grid spacing) for the first 180 hours and thereafter at T190 (equivalent to around 80km grid spacing). For all horizons the number for vertical levels is 64. A set of ensemble members is created by adding or subtracting bred modes to the unperturbed analysis (Toth and Kalnay, 1993, 1997). The bred modes are the differences in the deterministic and perturbed forecasts that evolve over the course of the forecast integrations. They are periodically rescaled before being recylced into subsequent forecasts. The perturbed ensemble members are integrated at a spectral truncation of T126 out to 180 hours and thereafter at T62. For all horizons the number of vertical levels is 28. For the 00Z forecast cycle an unperturbed integration is performed at the same resolution as the perturbed members. The operational ensemble system configuration has changed a little during the period of this study. However for format of the data obtained for the purpose of this study has not changed much. Data was collected from the ftp site `ftp://ftpprd.ncep.noaa.gov/pub/data/nccf/com/mrf/prod/` for the 00Z and 12Z forecast cycles at a horizontal resolution of 1x1 degree for leadtimes out to 180 hours at 6 hour intervals. Before April 2004 (as used in this work) the maximum lead time was 84 hours. The fields collected were u (westerly) and v (southerly) wind components at 10 m a.g.l. and 850hPa, temperature at 2m and 850hPa, and mean sea level pressure for the region 12W to 35E and 25N to 65N. For the 00Z forecasts 12 concurrent forecasts are obtained: 2 unperturbed forecasts at T382 and T190 and 5 pairs of perturbed forecasts. For the 12Z forecasts 11 concurrent forecasts are obtained: 1 unperturbed forecast at T382 and 5 pairs of perturbed forecasts. The 00Z cycle forecasts are used in this study because of the availability of the control (unperturbed) forecast made at the same spectal truncation, which is not present for the 12Z cycle forecast.

- The period considered is January 1 – October 31, 2003. Data until June 1 is used for training/estimation of e.g. power curves, while the remaining data are used for testing the methods. For Tunø Knob data are available back to July 1, 1999. Data from before January 1, 2003 is only used in Section 6 to indicate the overall spread of historic power productions.

Figure 2 shows the average forecasted wind speed versus horizon for NCEP and ECMWF forecasts. The diurnal variation is clearly seen in the forecasts. Furthermore, there is a large difference between the the NCEP analysis (i.e. the 0-hour forecast) and the remaining horizons.

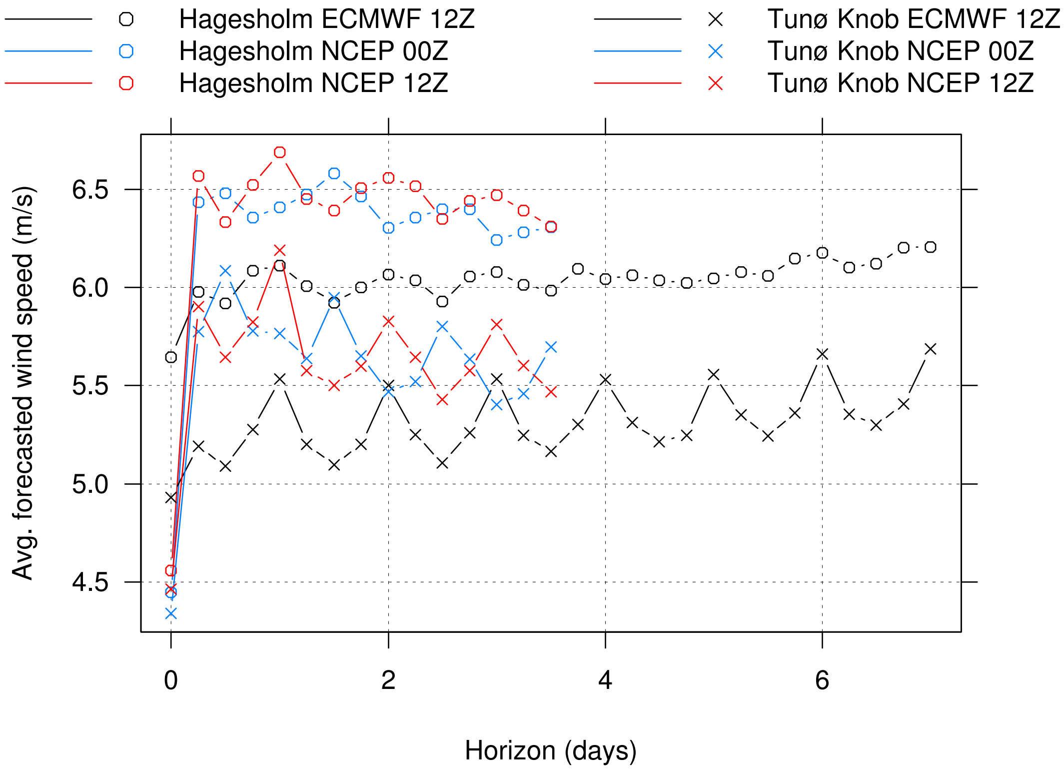


Figure 2: Average forecasted wind speed versus horizon for NCEP and ECMWF forecasts.

4 Power Curves

Since a wind speed measurement is rarely representative for a whole wind farm and since the availability and reliability of such measurements often is low in reality it was decided that the method to be developed should not use wind speed measurements. Consequently, power curves must be estimated using the power output and the unperturbed forecast. One of the challenges implied by this approach is that the unperturbed forecast of course is associated with some uncertainty when compared to the wind speed and direction experienced by the wind farm. Bias depending on the speed and direction can be corrected for by the statistical methods used, but random variation will result in biased estimates of the power curve, e.g. if a power curve model is fitted specifically for 48 hours forecasts then this model will never predict the maximum production.

With the aim of obtaining e.g. a small root mean squared error of a forecast the use of biased estimates obtained as just outlined is appropriate (Jonsson, 1994; Nielsen et al., 2002a,b). However, with the aim of producing ensembles of the wind power production the bias of the estimated power curve should be small. The approach used here is characterized by (i) a simple transformation of the power to force the power curve estimate to span the full range of possible power productions and (ii) estimation of a power curve not depending on the horizon. It is also important to use all observations available. The power measurements are available as averages over either 15 or 60 minute intervals, while the forecasts are available every sixth hour and can be interpreted as an average over a relatively small interval of time (approximately 10 minutes). To align a forecast with every observation additional forecasts are created using linear interpolation. Since the forecasts are updated daily we must consider horizons from 0 to 24 hours in order to use every observation once. Due to the strange behaviour of the NCEP-analysis (cf. Figure 2 on page 6) the NCEP analysis is not used and hence for NCEP horizons 6 to 30 hours are used.

With respect to the transformation let P denote the power output of the wind farm, the transformed power y is found using

$$y = c_0 + c_1 \log \frac{P - \underline{P}}{\overline{P} - P} \ , \tag{1}$$

where c_0, c_1, $\underline{P}$, and $\overline{P}$ are coefficients to be determined from data. To ensure that the total range of the power is covered $\underline{P}$ is determined as the largest multiple of 10 kW which is smaller than all observations and $\overline{P}$ as the smallest multiple of 10 kW which is larger than all observations. The remaining constants which determine the placement and the slope of the transformation is found by nonlinear regression in the inverse of (1) when y is replaced by the unperturbed forecast of the wind speed. This amounts to estimating a simple logistic-shaped power curve with a fixed span of power output. For the data used cut-out do not seem to occur.

To account for the wind direction, deviations from the logistic shape, and to adjust for bias originating from the uncertainty of the forecast the transformed power output is modelled as

$$y = f(u, v) + g(u, v)\tau + e \ , \tag{2}$$

where the wind velocity (u, v) is a vector representation of the unperturbed forecast of speed and direction, τ is the forecast horizon (lead time) of the meteorological model, f and g are smooth functions, and e is the error term. The term $g(u, v)\tau$ adjust for an increasing uncertainty with horizon which may also depend on the wind speed and direction. In this way the bias of the estimate of $f(u, v)$ is reduced and after transformation to the original scale this is used as the power curve. The non-parametric approach to modelling and estimation is similar to the approach used in WPPT, version 4 (Nielsen et al., 2002c). However, here the adaptive version of the estimation method is not used.

When no parametric assumptions are placed on the functions f and g the model (2) is a conditional parametric model (Cleveland, 1994; Hastie and Tibshirani, 1993). The functions are estimated by local regression with bandwidths chosen using the nearest neighbour principle and a tricube weight function. A local linear approximation is used for f and a local constant approximation is used for g. It is possible to fit models of this kind using the standard S-PLUS function `loess`, but we have used the S-PLUS library LFLM (Nielsen, 1997) because it is more flexible. The nearest neighbour bandwidth is selected by considering the actual resulting bandwidth and by requiring the fit not to exhibit excess variability. It is found that a bandwidth of 10% is required for the fit to be sufficiently smooth. Since this bandwidth results in an actual bandwidth of 4-7 m/s when the forecasted wind speed is 5 m/s it is not attempted to increase the bandwidth further.

Figure 3 displays the resulting estimate of the power curve of Tunø Knob. The difference between ECMWF and NCEP forecasts is relatively large and the dependence on wind direction is rather different. This can partly be explained by the difference in mean forecasted values as displayed on Figure 2. Within both of the forecast systems (ECMWF/NCEP) a rather large dependence on the forecasted wind direction is evident. Furthermore, the right panel of Figure 3 displays also the estimate obtained when excluding $g(u, v)\tau$ from the model. It is seen that this in most cases shifts the power curve to the right, while the transformation preserves the full span. The decision whether to include the forecast horizon (lead time) of the meteorological model in the model (2) used for estimation of the power curve is based on the probabilistic properties of the wind power ensembles obtained by filtering all ensembles trough the power curves obtained. Let r be the rank of the observed power as compared to the ensembles obtained as just outlined and let N be the number of ensembles, i.e. $N = 51$ for ECMWF ensembles. If the ensemble forecast is correct in a probabilistic sense then, except for rounding due the finite number of ensembles, $(r - 1)/N$ will be uniformly distributed on $[0, 1]$. This observation is the basis of rank histograms, also called Talagrand diagrams (Toth et al., 2003). The uniformity of $(r-1)/N$ can be judged from data given that these are grouped according to some criteria. Here the data are grouped according to site, exclusion/inclusion of horizon in the power curve model, and horizon in steps of six hours. Instead of histograms the uniformity of $(r - 1)/N$ is judged from Quantile-Quantile plots or QQ-plots which is a

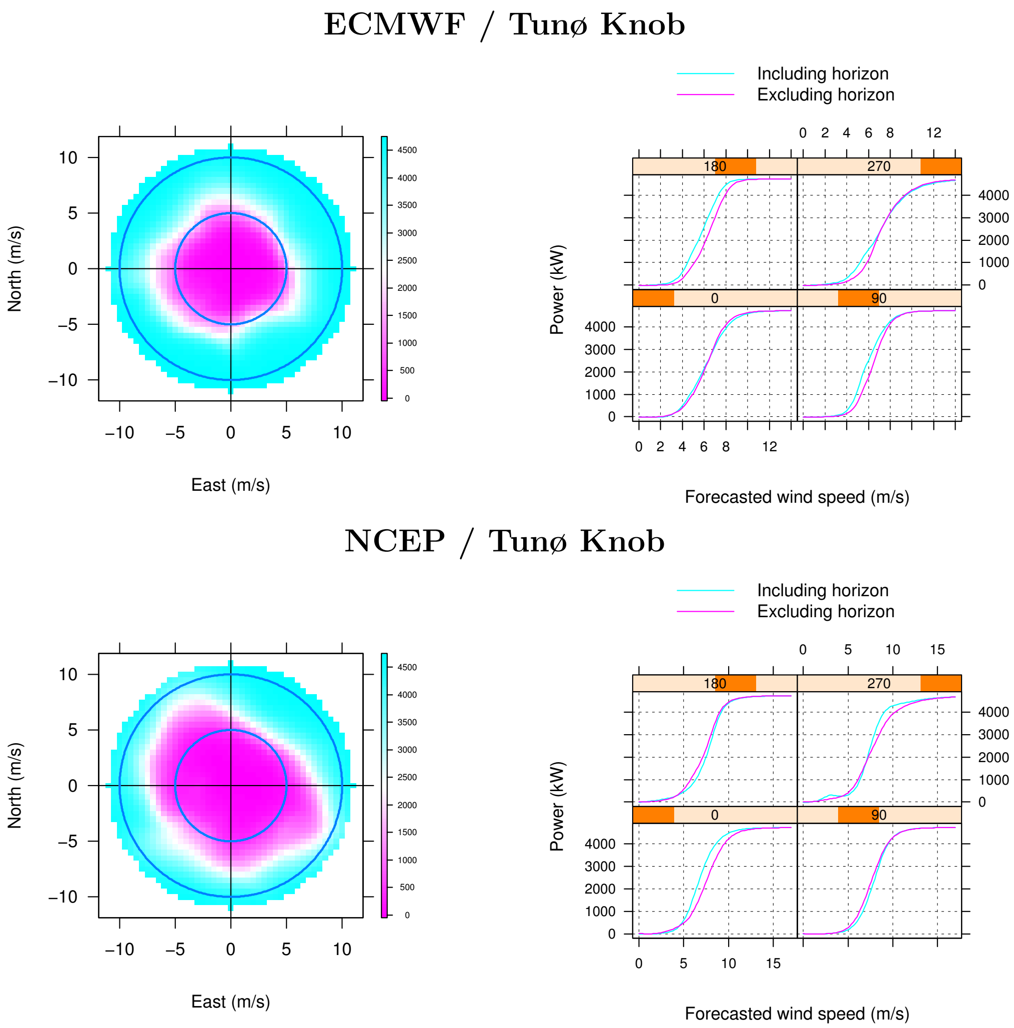


Figure 3: Left panel: Level plot of the direction dependent power curve for Tunø Knob. The circles mark 5 and 10 m/s forecasted wind speed. Right panel: Power curves for selected wind directions (degrees) for the model corresponding to the left panel and for a model where $g(u, v)\tau$, i.e. the horizon, is excluded from the model.

standard tool for comparing distributions (Chambers et al., 1983). Ideally, these plots should be a straight line between $(0, 0)$ and $(1, 1)$.

The QQ-plots for the training period are displayed in Figure 4 for horizons 36 to 60 hours since these, considering the calculation time, are the most relevant horizons from a Danish perspective. It is seen that the ensemble forecasts are not correct in a probabilistic sense, but using a power curve model where the horizon is included during estimation results in curves closer to the line of identity, especially for the ECMWF-ensembles. For this reason model (2) including the horizon is preferred.

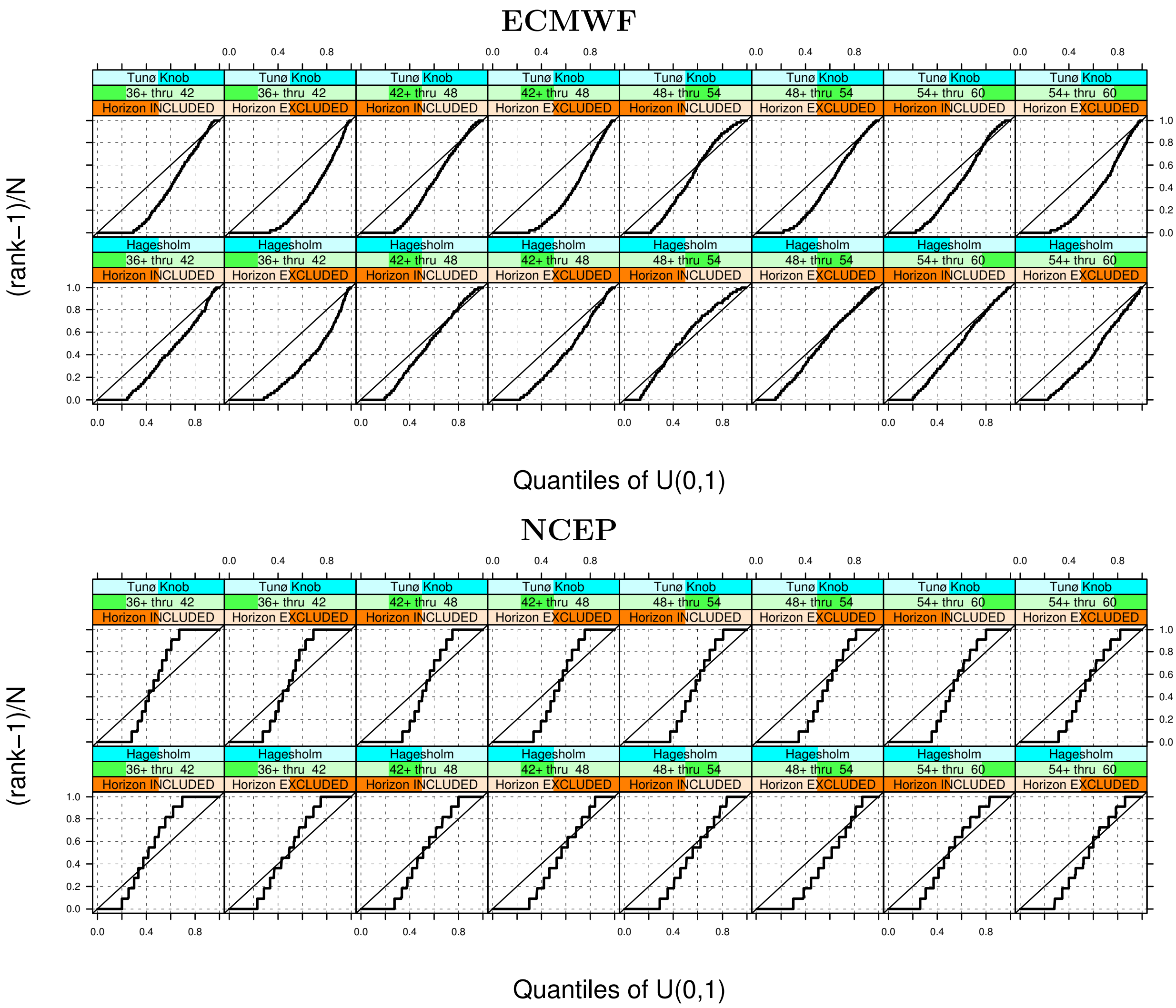


Figure 4: QQ-plots of $(r-1)/N$ for the two sites for horizons ranging from 36 to 60 hours and when including or excluding the horizon when estimating the power curve using model (2). The plots are based on the training data.

5 Correction of Quantiles

As seen from Figure 4 quantiles based on the power ensembles obtained as described in Section 4 can not be interpreted in a strict probabilistic sense. However, based on plots like Figure 4 the raw ensemble quantiles can be adjusted by looking up e.g. the raw 20% quantile on the 2nd axis and reading off the correct probability on the 1st axis. For instance from the plot in the lower left corner it can be seen that the raw 20% quantile in reality is closer to a 40% quantile. Also, for this particular plot the actual probability is approximately 25% when the raw probability reaches zero; for this reason we can not get information about true probabilities below 25%.

With the aim of adjusting the raw quantiles derived from the power ensembles a model which estimates the true probability from the raw probability and the horizon is derived. For the training period the raw probabilities are calculated as $(r-1)/N$, see Section 4. Hereafter, the data are grouped by each individual horizon, i.e. 15 min. steps for Tunø Knob and 60 min. steps for Hagesholm. For each horizon the data are then sorted by the raw probability and correct probabilities are obtained by generating equidistantly spaced values between 0 and 1. This amounts to generating the data used in QQ-plots for each horizon separately.

Figure 5 displays some common properties of the data. It is seen that quite strong fluctuations with the horizon or time of day are present. Furthermore, given the horizon the adjusted probability can be approximated by a smooth function of the raw probability. However, saturation occurs for raw probabilities equal to 0 and 1. This is because observations occur relatively frequently outside the range of the ensemble forecast and for the example displayed in the left panel of Figure 5 it is not possible to obtain probabilistic correct quantiles for probabilities below approximately 30% or above approximately 90%. Furthermore, the steep slopes are difficult to model and are likely to induce local bias of the estimates. For this reason it is decided to exclude data with raw probabilities equal to 0 or 1 from the analysis. Furthermore, the adjusted probabilities are logit-transformed (see (3) below) in order to ensure that the model obtained returns adjusted probabilities between 0 and 1.

Due to the structure of the data a conditional parametric model (Cleveland, 1994; Hastie and Tibshirani, 1993) is used where the dependence on the raw probability p_r is modelled as a cubic spline with boundary knots at 0 and 1 and two equidistantly placed internal knots (de Boor, 1978). The coefficients of the spline are estimated non-parametrically as smooth functions of the horizon τ. The model is formally written

$$\log \frac{p_a}{1-p_a} = \mathbf{B}(p_r)\boldsymbol{\theta}(\tau) + \epsilon \tag{3}$$

where p_a is the adjusted probability, $\mathbf{B}(p_r)$ is a matrix representing a spline basis expansion of the raw probability p_r, $\boldsymbol{\theta}(\tau)$ is the vector of smooth functions, and ϵ is the error. Estimation is performed using local regression with a fixed bandwidth and a tricube

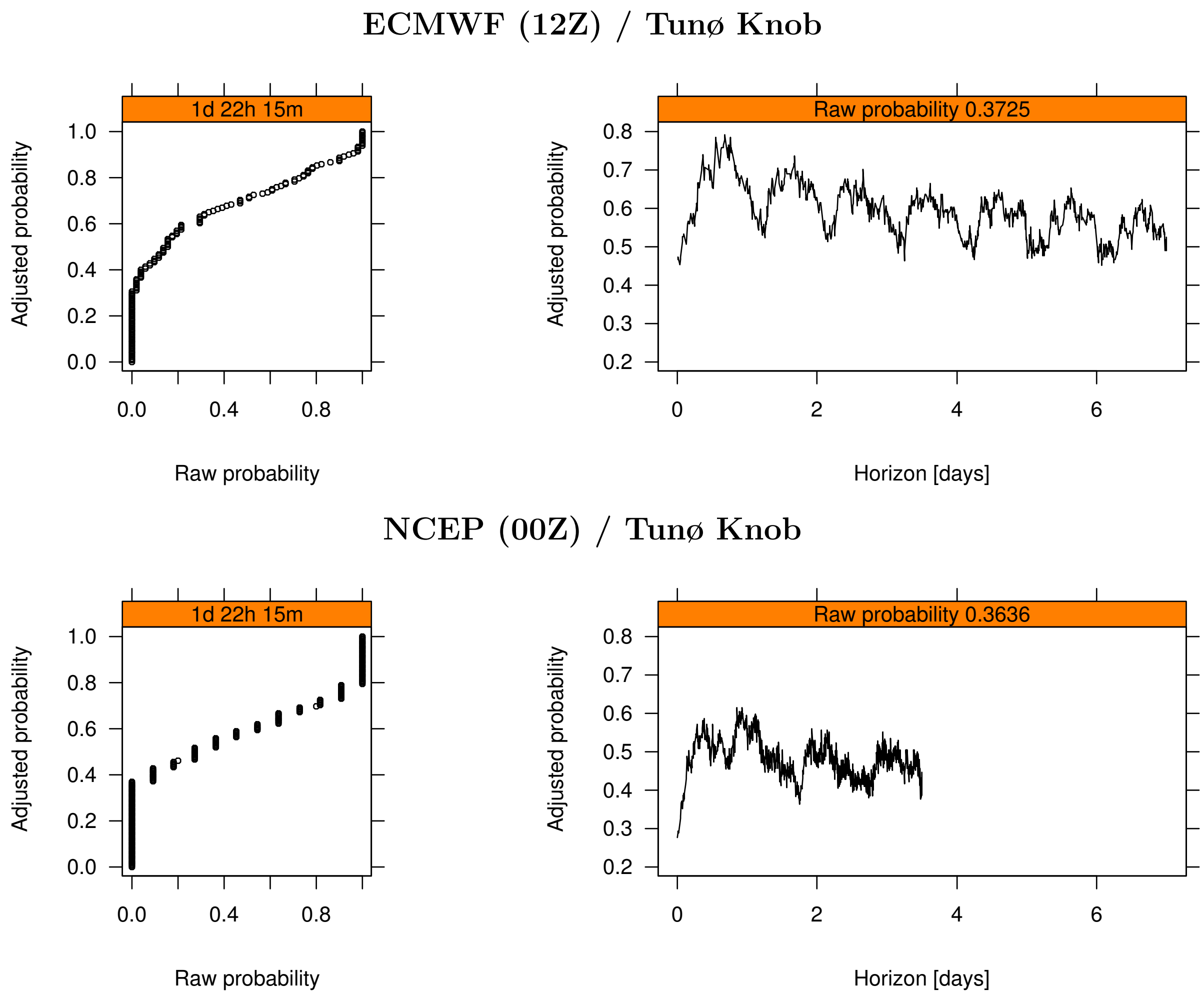


Figure 5: Examples of the dependence of the adjusted probability on the raw probability given a specific horizon (left) and on the horizon given a specific raw probability (right).

weight function. Due to the presence of peaks (Figure 5, right), the functions $\boldsymbol{\theta}(\tau)$ are locally approximated by 2nd order polynomials. The software LFLM (Nielsen, 1997) is used[2]. The bandwidth is chosen by inspecting the residuals of the fit for Tunø Knob; to eliminate clear systematic variation with the horizon a bandwidth of two hours is chosen.

Using models (2) and (3), estimated using the training period data, adjusted quantiles are calculated for the test period. Given the observations the corresponding probabilities p_o can then be calculated. If the adjusted quantiles are correct in a probabilistic sense these should be uniformly distributed between 0 and 1. However, due to saturation as described above, no information about probabilities outside the range of the estimated p_a is available. Hence, p_o should be uniformly distributed on the range of the estimated p_a, considering each horizon separately.

[2]With two internal knots the basis $\mathbf{B}(p_r)$ has 6 columns; the resulting model can not be fitted using standard S-PLUS.

In Figure 6 results based on the test period are depicted in terms of the QQ-plots considering the range of estimated p_a described above. It is seen that the adjusted quantiles are close to being probabilistically correct; the maximum deviation seems to be below 10%. In general the results are better when using ECMWF-ensembles than when using NCEP-ensembles, especially for Hagesholm.

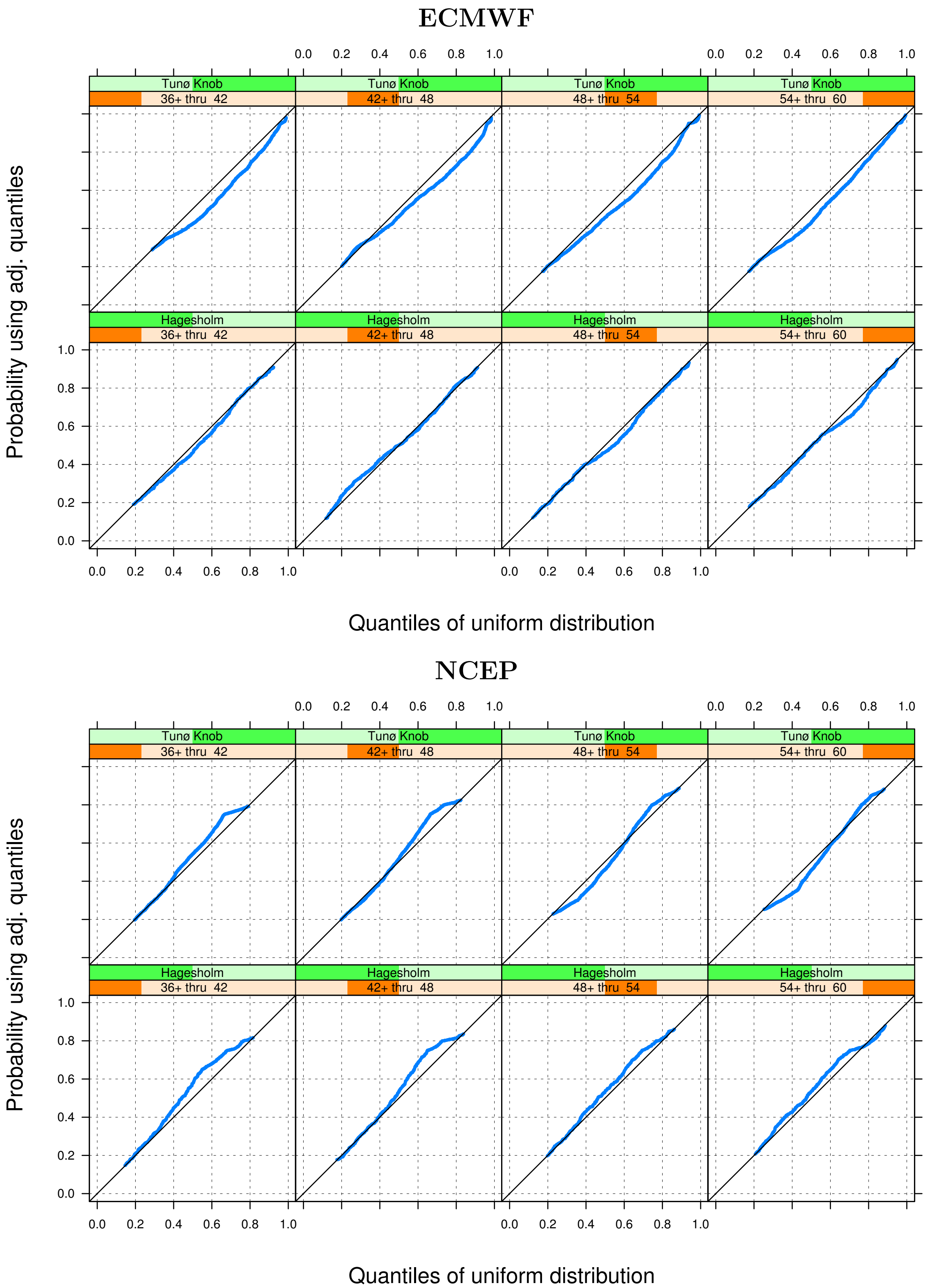


Figure 6: QQ-plots for adjusted quantiles (test period) the two wind farms and for horizons ranging from 36 to 60 hours in steps of 6 hours. Plots covering all horizons can be found in Figures 16 and 15.

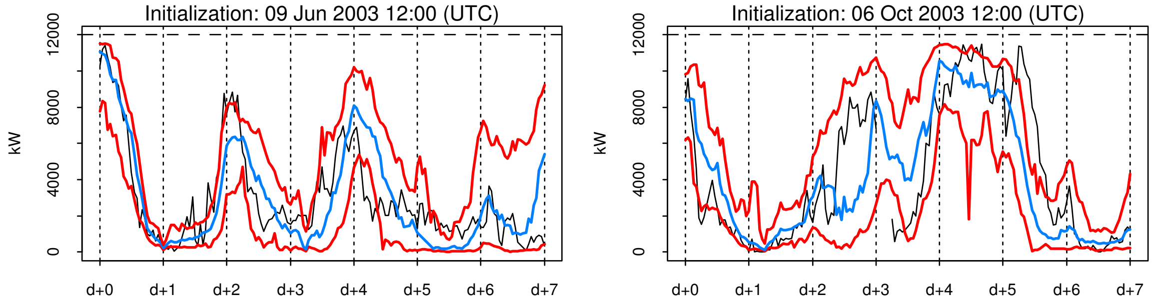


Figure 7: Two examples of adjusted ensemble quantiles for the wind farm at Hagesholm. The red lines indicate the 25% and 75% quantiles, the median is indicated by the blue line, and the actual production is indicated by the black line.

6 Sharpness, resolution, and spread/skill of the ensemble forecasts.

Although it is not the main focus of the report the quality of the obtained ensemble quantiles of wind power is briefly addressed in this section. Examples of forecasted quantiles (25%, 50%, and 75%) are displayed in Figure 7. For the forecast with initialization time 12:00 Oct. 6, 2003 (UTC) it is seen that the Inter Quartile Range (IQR), i.e. the power (vertical) difference between the 1st and the 3rd quartile (the 25% and the 75% quantile), is relatively high between day 2 and 3. The same forecast has markedly lower uncertainty between day 5 and 6 although the level of the power production is similar.

Given ensemble quantiles which are correct in a probabilistic sense the quality of these depends on

(i) the ability to distinguish between situations with low and high uncertainty and on

(ii) the sharpness of the distributions.

Here the sharpness is measured as the IQR. For Tunø Knob and horizons 36-42 hours the 25% quantile is just below the estimated range of p_a, cf. Figure 6. To obtain values for the 25% quantile linear interpolation to the natural 0% quantile $\underline{P}$ is used. For Tunø Knob $\underline{P}$ is set to $-80kW$.

Qualitatively (i) is fulfilled if both low and high values of the IQR occur and with respect to (ii) the IQR should be smaller than the IQR obtained from historic production data. These aspects are addressed for Tunø Knob in Figure 8 and for Hagesholm in Figure 9 for horizons ranging from 36 to 60 hours. The figures show the 5%, 50% and 95% quantiles of the distribution of all IQRs in our test set, in six plots ordered by the actually produced power. Especially for small and large power productions is the IQR usually low, and typically much lower than the historical IQR derived from the measurement, as shown by the dashed lines in the plots. The 5% and 95% quantiles of the ensemble IQR indicate high variability and for probabilistic correct quantiles this can be interpreted as the fulfillment of (i). Furthermore, it is seen that in many situations the ensemble IQR is significantly smaller than the IQR of the historic power productions, i.e. the ensemble forecast is sharp compared to historic data. For the steep part of the power curve the IQR is smaller than the IQR of the historic data in approximately 50% of the cases.

Figure 10 on page 20 shows the IQR plotted against horizon for the full test period. Especially for NCEP some strange behavior is observed for some horizons. The diurnal variation is again seen clearly in the ECMWF plots. To a large extent this can be explained by small ranges of valid quantiles as can be seen on Figure 15 on page 28. The plots indicate that for a large fraction of the time small IQR-values are observed. This is confirmed by the histograms of IQR in Figure 17 on page 30. Due the marked non-normality of the IQR-values both standard and robust measures of sharpness (location) and resolution (scale) are addressed:

- For sharpness (location) the mean is used with the median being the robust alternative. See Figure 11 on page 21.
- For resolution (scale) the standard deviation (SD) is used with the median absolute deviation (MAD) being the robust alternative. The MAD is scaled so that *if* the sample is Gaussian expected value of the MAD is equal to that of SD. See Figure 12 on page 22.

The plots do not indicate any large difference between quantile forecasts from based on either NCEP or ECMWF. However:

- With respect to sharpness the NCEP-based quantiles vary rapidly with horizon, especially for Hagesholm. There is a small tendency for the NCEP-based quantiles to have smaller IQR than the ECMWF-based quantiles.
- With respect to resolution the ECMWF-based quantiles seems to have slightly higher variation in IQR-values than the NCEP-based quantiles.

However, as indicated on Figure 6 on page 14 (and Figures 15 (page 28) and 16 (page 29)) the ECMWF-based quantiles are also (slightly) more reliable than the NCEP-based quantiles. Note that the small IQR-values observed relatively frequently for horizons up to five days indicate that even on longer horizons situations with low uncertainty occur. Figure 13 on page 22 shows such an example.

On a side note, we looked at the potential of the IQR to explain the actual accuracy of the ensemble forecast, taken here as the median (50% quantile) of all ensemble forecast members. The idea is that there should be a low ensemble spread correlating with a high achieved accuracy of the forecast and vice versa. In other words, there should be a connection between forecast spread and forecast skill. Figure 14 shows the spread/skill relationship for the median (50% quantile) taken as a point forecast. It is seen that the spread/skill relationship is not very dependent on the horizon considered. This is taken as an indication of the IQR actually reflecting the underlying uncertainty.

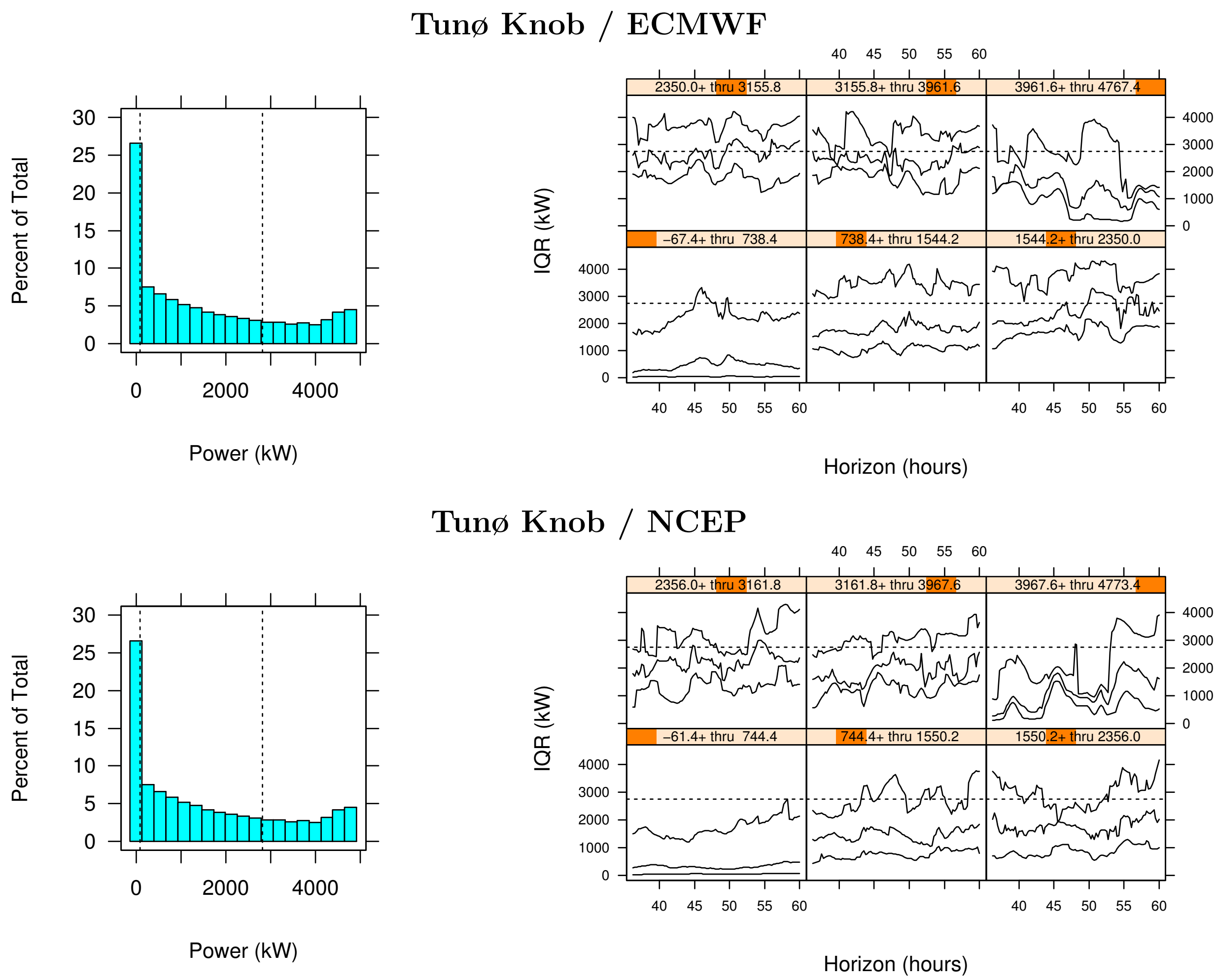
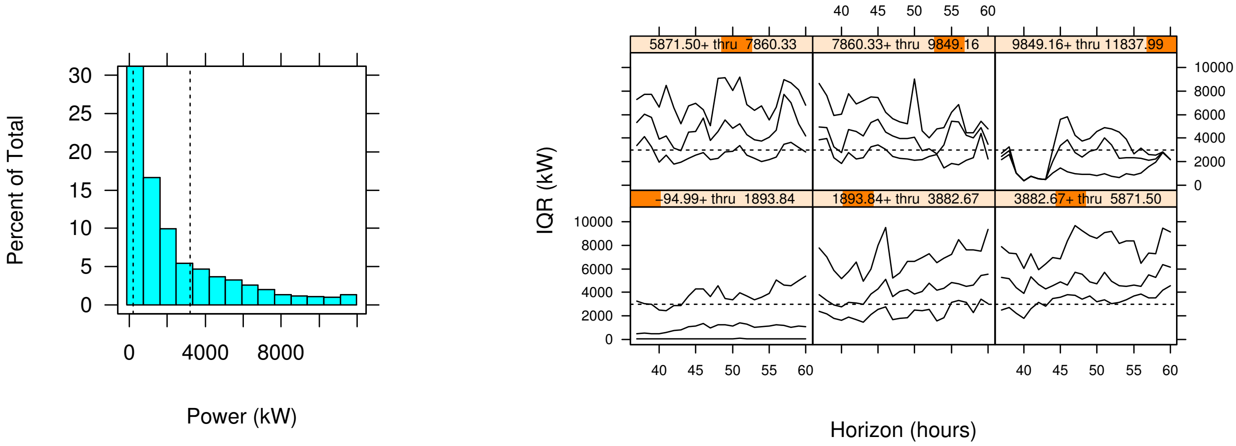


Figure 8: Tunø Knob: Histogram of power production during the period 30/06/1999 23:15 – 01/06/2003 00:00 (UTC) with 25% and 75% quantiles indicated by vertical lines (left). Quantiles (5%, 50%, 95%) of the IQR of ensemble quantiles adjusted as described in Section 5 (right). The horizontal lines indicate the IQR of the historic data, i.e. the difference between the two vertical lines on the left. The grouping variable is the forecasted power production in terms of the 50% ensemble quantile.

Hagesholm / ECMWF



Hagesholm / NCEP

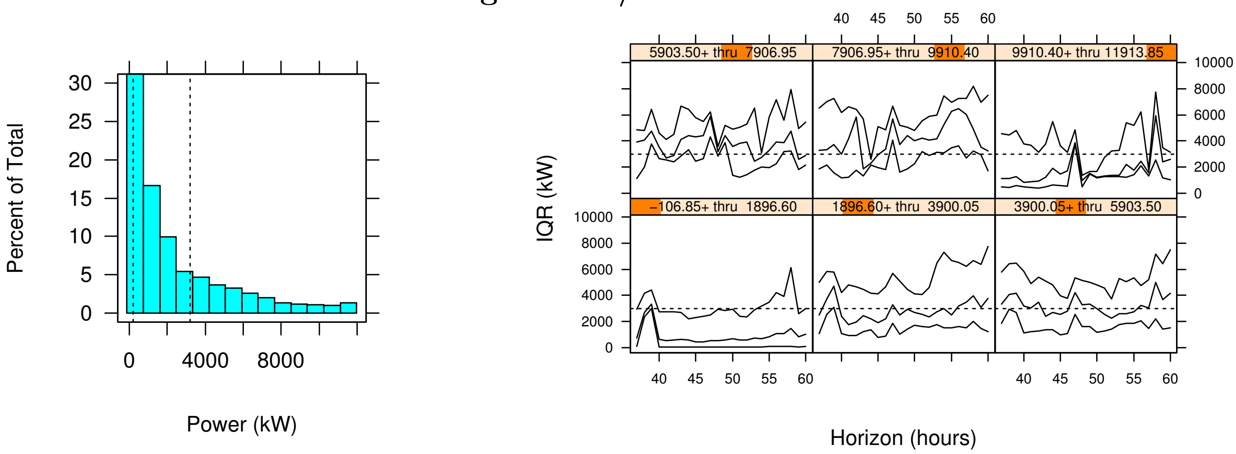


Figure 9: Hagesholm: Histogram of power production during the period 1/12/2002 00:00 – 01/06/2003 00:00 (UTC) with 25% and 75% quantiles indicated by vertical lines (left). Quantiles (5%, 50%, 95%) of the IQR of ensemble quantiles adjusted as described in Section 5 (right). The horizontal lines indicate the IQR of the historic data, i.e. the difference between the two vertical lines on the left. The grouping variable is the forecasted power production in terms of the 50% ensemble quantile.

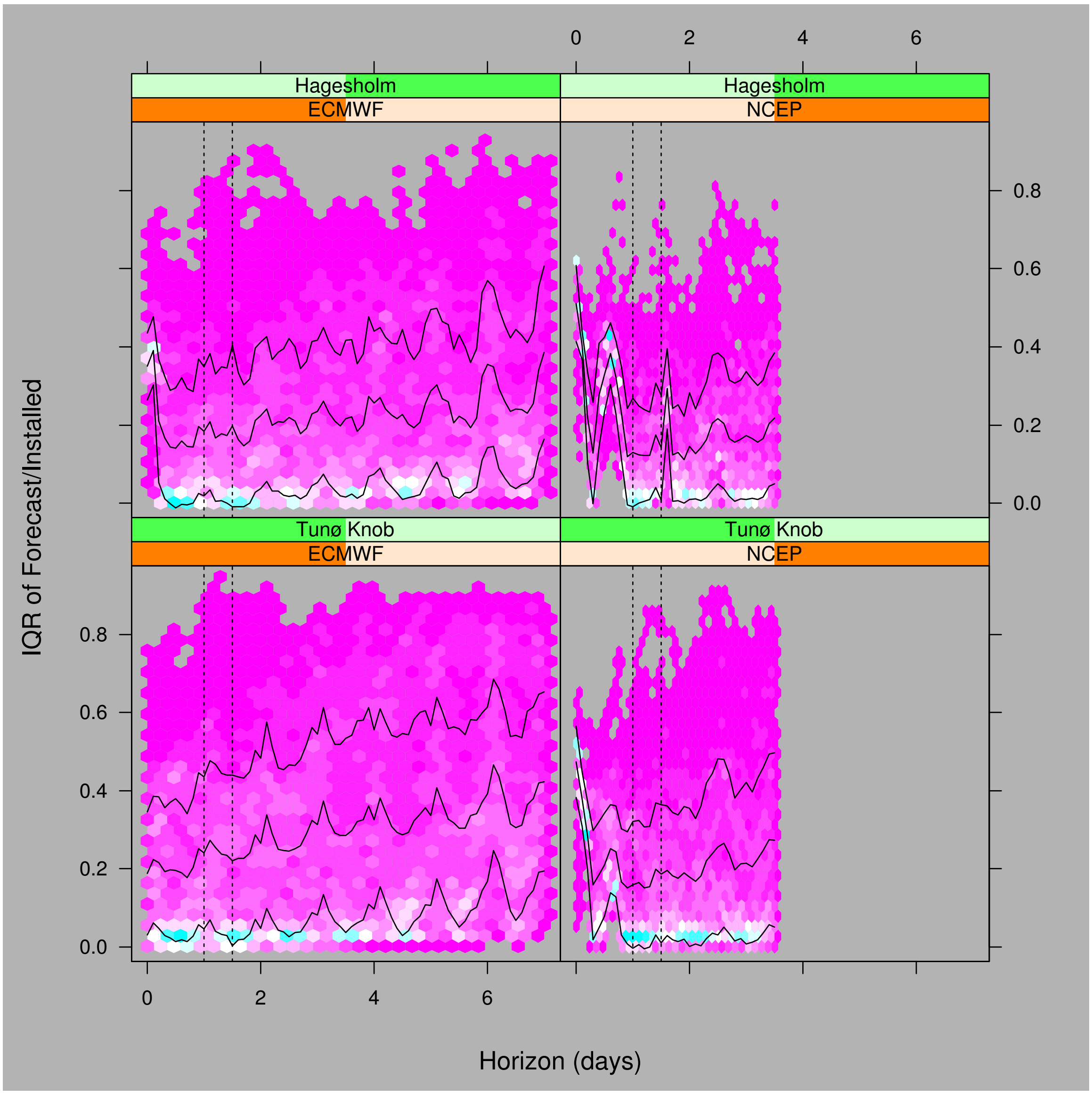


Figure 10: Color coded scatter plot of horizon and Inter Quartile Range (IQR) normalized by installed capacity (test period). The color code is relative to each subplot (cyan indicates the maximum count). The mean IQR $\pm$ one SD is overlayed as lines on the plots. Horizons (lead times) of 24 and 36 hours are shown by vertical lines.

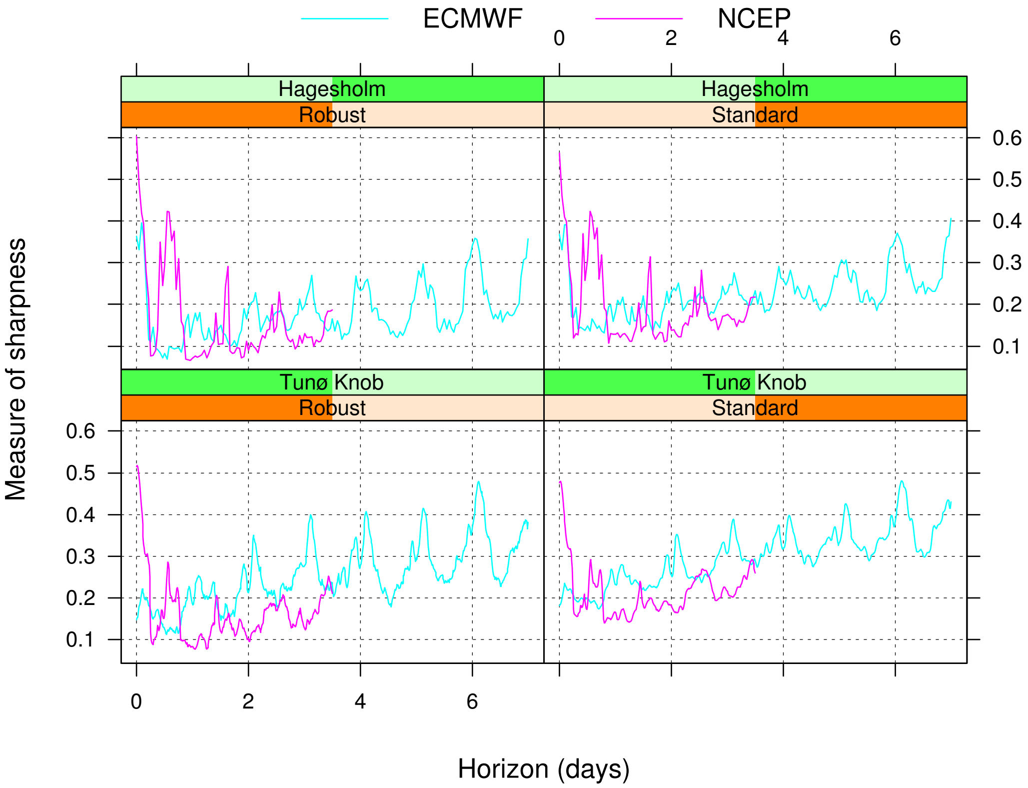


Figure 11: Measure of sharpness (standard; mean, robust; median) against horizon (test period).

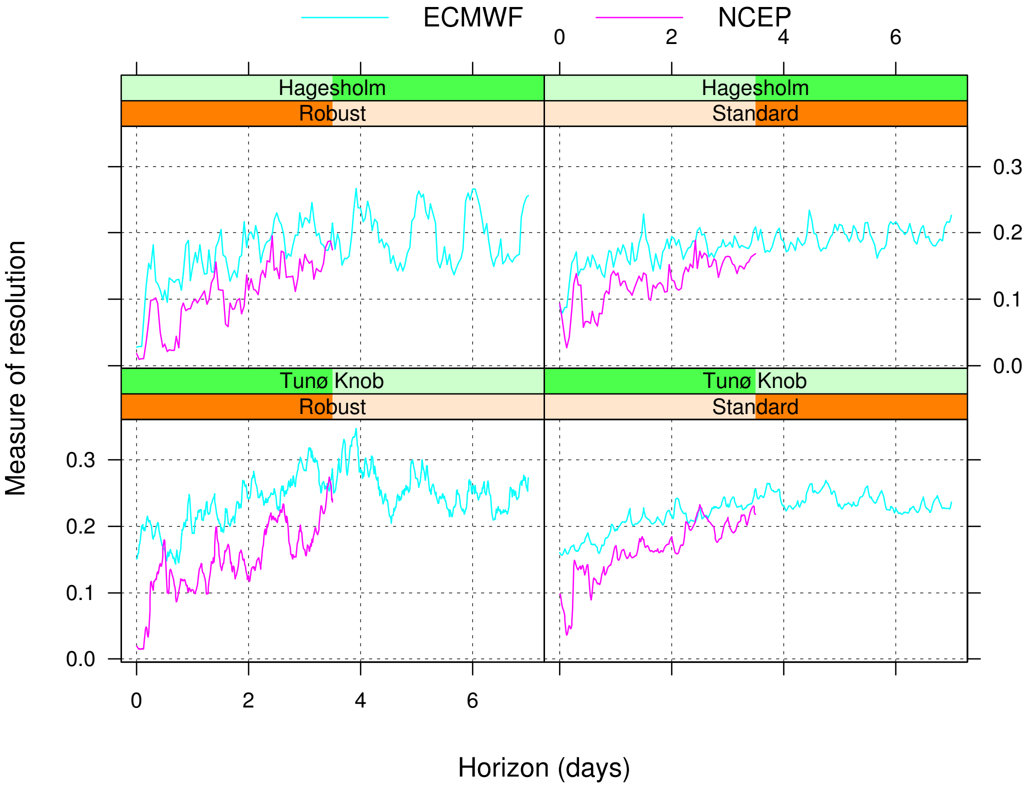


Figure 12: Measure of resolution (standard; SD (standard deviation), robust; MAD (Median Absolute Deviation)) against horizon (test period). The MAD is scaled so that equality with SD is obtained if the distribution is Gaussian.

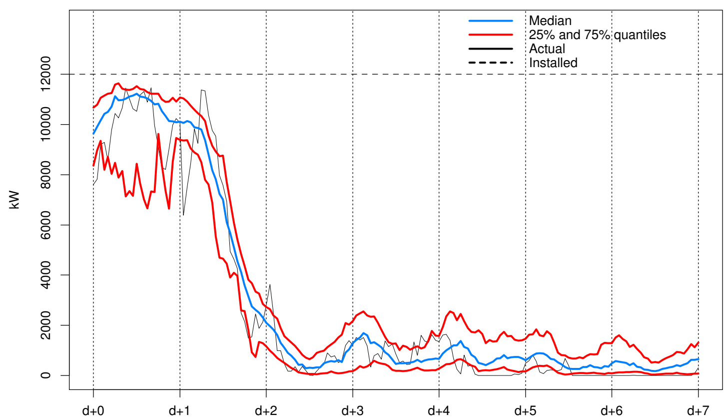


Figure 13: Adjusted ensemble quantiles for the wind farm at Hagesholm for the forecast with ECMWF initialization time Oct. 10, 2003 12:00 (UTC). The red lines indicate the 25% and 75% quantiles, the median is indicated by the blue line, and the actual production is indicated by the black line.

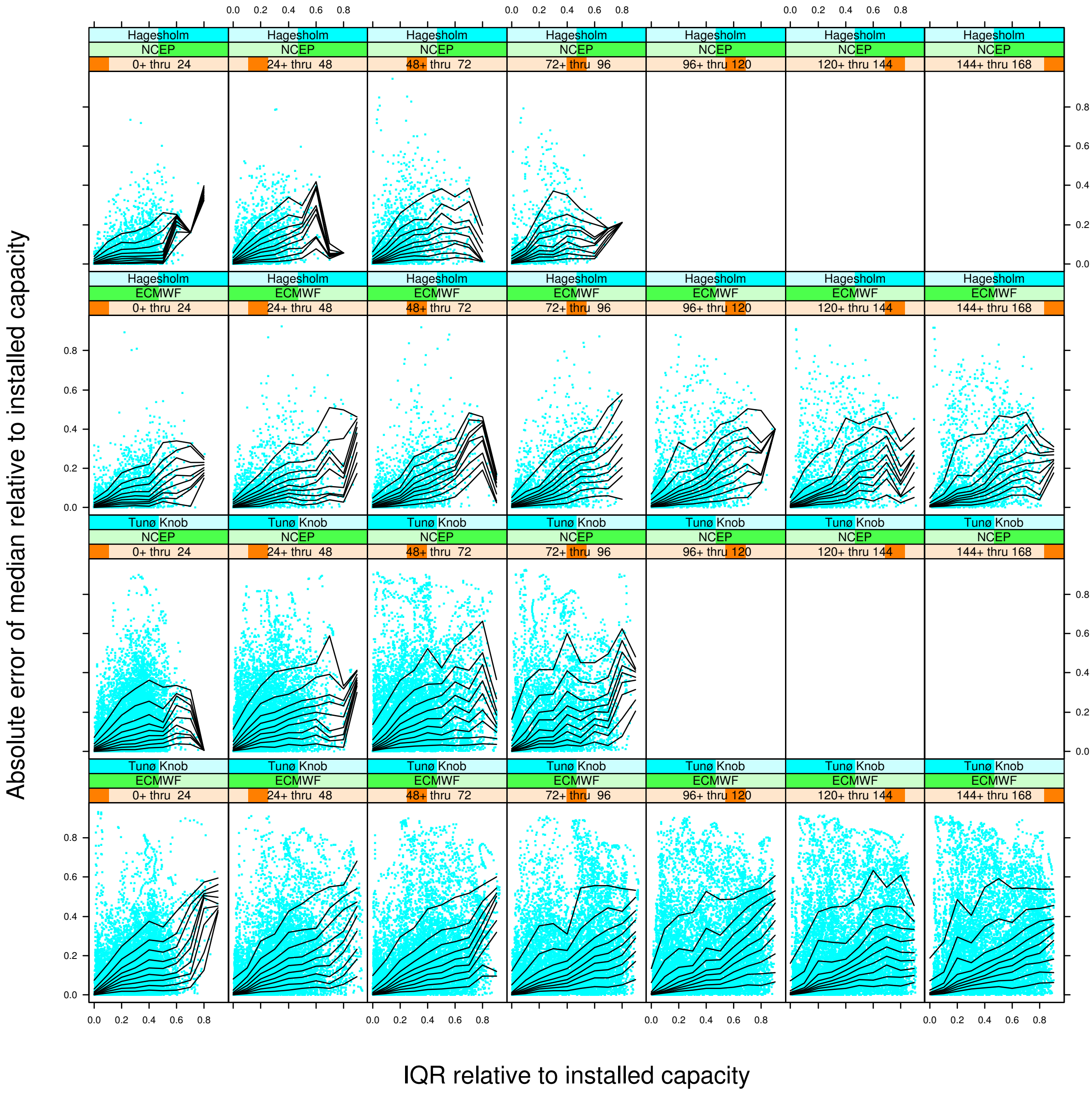


Figure 14: Absolute error of the median considered as a point forecast versus the IQR, both relative to installed capacity. The lines indicates the 10%, 20%, . . . , 90% quantiles of the error when splitting the IQR (1st axis) in ten groups. The plots are shown for horizons in groups of 24 hour.

7 Conclusion and Discussion

The main focus of this report is the development of a method, which is able to (i) transform meteorological ensembles of wind speed and direction to ensembles of power production, and (ii) adjust the ensemble quantiles so that these are probabilistic correct on the long run when evaluated on a forecast horizon basis. Ensembles of the wind speed and direction 10m a.g.l. from the European Centre for Medium-Range Weather Forecasts (ECMWF) and from the National Center for Environmental Prediction (NCEP) in the U.S. are used in the investigation.

Opposed to point forecasts, it is found to be important to estimate the power curve in a way that reduces the bias of the estimate. Ideally, an unbiased estimate should be used. Such an estimate could be obtained if very adequate wind speed measurements were available at each farm. However, ensemble forecasts corresponding to such measurements are not available because ensemble forecasting systems use low spatial resolution. Consequently, a method must be developed that transforms the meteorological ensembles to ensembles corresponding to the local measurements. Basically, this poses the same issues regarding biased estimates as for the power curve estimation.

To ensure that the full range of possible power productions can be forecasted by the power curve the power curve model works on appropriately transformed power output. Hereafter, the power curve is estimated using a conditional parametric model based on the unperturbed forecast where a conditional linear dependence on the horizon adjusts for uncertainty of the meteorological forecast. It is shown that in terms of the probabilistic properties of the power ensembles the chosen approach is preferable to just excluding the forecast horizon from the model.

It is also shown that the quantiles derived from the power ensemble forecasts are not correct in a probabilistic sense. However, except for the extreme quantiles, these can be transformed to probabilistic more adequate quantiles. As a tool for such an transformation a conditional parametric model is used. The quality of the transformation is verified by testing it on independent data. Interestingly, the coefficients of the transformation varies rapidly with the horizon (a bandwidth of two hours is used for estimating in model (3) on page 11). One reason could be that the meteorological model does not predict the diurnal variation adequately. Also, note that the forecasts are available only for 00:00, 06:00, 12:00, and 18:00; the interpolation used may also induce some of the more high-frequent variations in the coefficients. Due to the observations just mentioned it seems appropriate to address the issue of fine-tune the power-curve modelling. Other approaches for estimation of the power curve should also be investigated. One such example is inverse regression; if the dependence of wind direction can be neglected, e.g. for the power curve of a region, and if the power production measurements can be adjusted to represent the full capacity of the region then uncertainty is only associated with the unperturbed forecast of the wind speed. It can easily be seen that this leads to a (non-parametric) regression model where the independent variable is the power production and the response is the unperturbed forecast of the wind speed. The inverse of the estimated curve is then an unbiased estimate of the power curve.

While not the main focus of the report the ability of the resulting power quantiles to distinguish between situations with low or high uncertainty is addressed. It is found that the power quantiles have this ability in that the distance between the 25% and the 75% quantile varies markedly. As a consequence we conclude that the meteorological ensembles possess this ability also. Taking the median as a point forecast the absolute error of this forecast is compared to the Inter Quartile Range (IQR) of the forecasted quantiles. This analysis show a clear spread/skill relationship which is only marginally influenced by the horizon. Hence, the spread of the quantiles is indeed a good indicator for the actual uncertainty of the forecast.

From the results presented in this report it is not clear if there is any benefit of one ensemble prediction system over the other. However, some strange behaviour is observed for adjusted quantiles based on NCEP ensembles. Hence, we cannot advice to use NCEP ensembles for horizons below 24 hours. Also for one wind farm (Hagesholm) some strange behaviour is observed for a very small range of horizons just over 36 hours. Also, due to the limited number of NCEP-ensembles (11) compared to ECMWF (51), NCEP-ensembles are most suited when only the central part of the probability distribution is required.

In some operational settings the production of individual farms is not of primary interest. If quantiles for a limited number of farms are of interest one solution is to generate power ensembles for each farm. However, the model for adjusting the quantiles must be based on the total production of the farms in question. In case of many farms it might be more practical to work with the total production of geographical regions. In this case a power curve of the region must be estimated, but the quantiles can be adjusted as described in this report. If the total production is not available then up-scaling becomes an issue too.

8 Acknowledgements

The project under which the work described here is carried out is sponsored by the Danish utilities PSO fund (ORDRE-101295 / FU 2101) which is hereby greatly acknowledged. Furthermore, the authors wish to thank Niels Emsholm (E2), John Tøfting (Elsam), the European Centre for Medium-Range Weather Forecasts, and the National Center for Environmental Prediction in the U.S. for supplying the data used in this study.

References

Bremnes J. Probabilistic wind power forecasts by means of statistical model. In *Proceedings of the IEA R&D Wind Annex XI Joint Action Symposium on Wind Forecasting*

Techniques, pp. 49–58. Norrköping, Sweden, 2002.

Buizza R. Trends in ensemble performance. *ECMWF Newsletter*, (94):2–5, 2002. `http://www.ecmwf.int/publications/newsletters`.

Buizza R., Miller M., and Palmer T. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, **125**(560):2887–2908, 1999.

Chambers J., Cleveland W., Kleiner B., and Tukey P. *Graphical methods for data analysis*. Wadsworth Publishing Co Inc, 1983.

Cleveland W. *Multivariate Analysis and Its Applications*, chapter Coplots, Nonparametric Regression, and conditionally Parametric Fits, pp. 21–36. Institute of Mathematical Statistics, Hayward, 1994.

de Boor C. *A Practical Guide to Splines*. Springer Verlag, Berlin, 1978.

Hastie T. and Tibshirani R. Varying-coefficient models. *Journal of the Royal Statistical Society, Series B, Methodological*, **55**:757–796, 1993.

Jonsson B. Prediction with a linear regression model and errors in a regressor. *International Journal of Forecasting*, **10**(4):549–555, 1994.

Lange M. and Heinemann D. Relating the uncertainty of short-term wind speed predictions to meteorological situations with methods from synoptic climatology. In *Proceedings of the European Wind Energy Conference & Exhibition*. Madrid, Spain, 2003. `http://www.ewea.org`.

Molteni F., Buizza R., Palmer T., and Petroliagis T. The ecmwf ensemble prediction system: methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, **122**(529):73–119, 1996.

Nielsen H. LFLM version 1.0, an S-PLUS / R library for locally weighted fitting of linear models. Technical Report 22, Dept. of Mathematical Modelling, Tech. Univ. of Denmark, DK-2800 Lyngby, Denmark, 1997. `http://www.imm.dtu.dk/~han/software.html`.

Nielsen H. and Madsen H. *Analysis and simulation of prediction errors for wind power productions reported to NordPool* (in danish). Technical report, Informatics and Mathematical Modelling, Tech. Univ. of Denmark, Lyngby, 2002.

Nielsen H., Nielsen T., and Madsen H. On on-line systems for short-term forecasting for energy systems. In *Proceedings of the OR 2002 conference*, pp. 265–271. Springer, Klagenfurt, Austria, 2002a.

Nielsen H., Nielsen T., and Madsen H. Using meteorological forecasts for short term wind power forecasting. In *Proceedings of the IEA R&D Wind Annex XI Joint Action Symposium on Wind Forecasting Techniques*, pp. 49–58. Norrköping, Sweden, 2002b.

Nielsen T., Madsen H., and Christensen H. WPPT – a tool for wind power prediction. In *Proceedings of the Wind Power for the 21st Century Conference*. Kassel, Germany, 2000.

Nielsen T., Nielsen H., and Madsen H. Prediction of wind power using time-varying coefficient-functions. In *Proceedings of the XV IFAC World Congress*. Barcelona, 2002c.

Pinson P. and Kariniotakis G. On-line assessment of prediction risk for wind power production forecasts. In *Proceedings of the European Wind Energy Conference & Exhibition*. Madrid, Spain, 2003. `http://www.ewea.org`.

Sela J. Spectral modeling at the national meteorological center. *Mon. Wea. Rev.*, **108**:1279–1292, 1980.

Toth Z. and Kalnay E. Ensemble forecasting at nmc: the generation of perturbations. *Bull. Am. Met. Soc.*, **74**:2317–2330, 1993.

Toth Z. and Kalnay E. Ensemble forecasting at ncep: the breeding method. *Mon. Wea. Rev.*, **125**:3297–3318, 1997.

Toth Z., Talagrand O., Candille G., and Zhu Y. In I. Jolliffe and D. Stephenson (editors), *Forecast verification – a practitioner's guide in atmospheric science*, chapter Probability and ensemble forecasts. Wiley, 2003.

A Additional plots

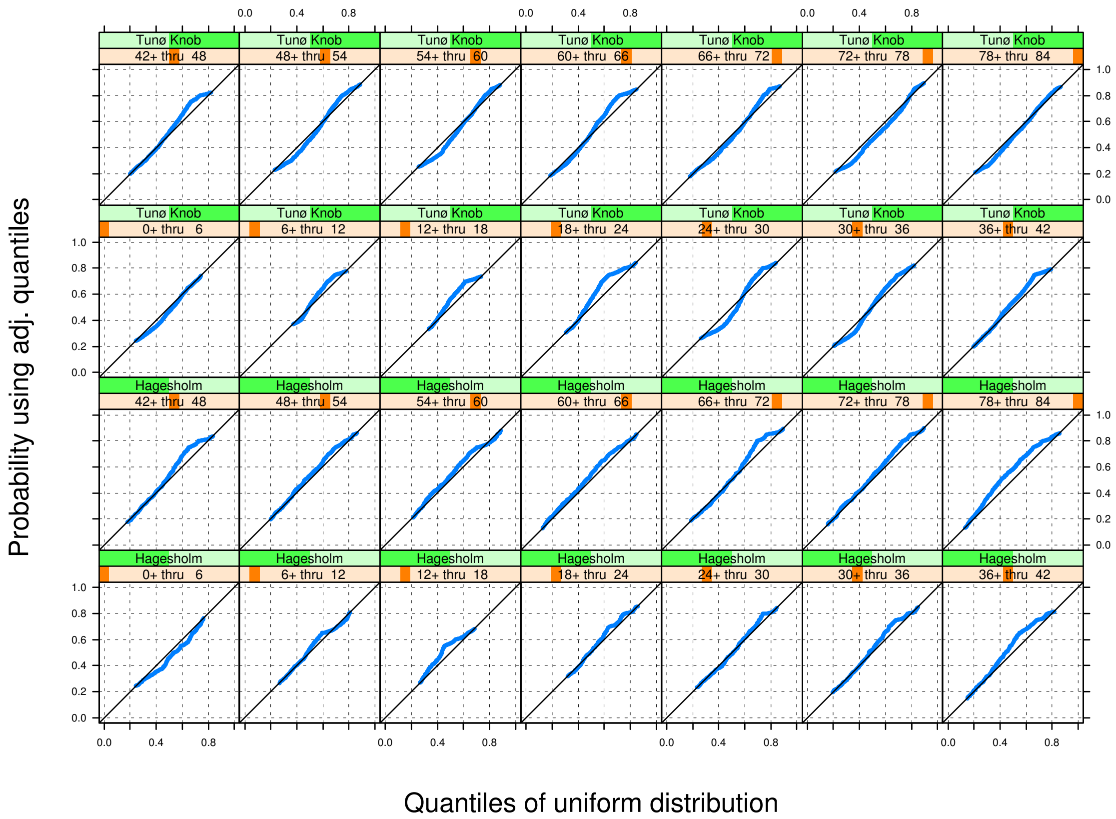


Figure 15: QQ-plots for adjusted quantiles (NCEP, test period) for the two wind farms and for horizons in steps of 6 hours.

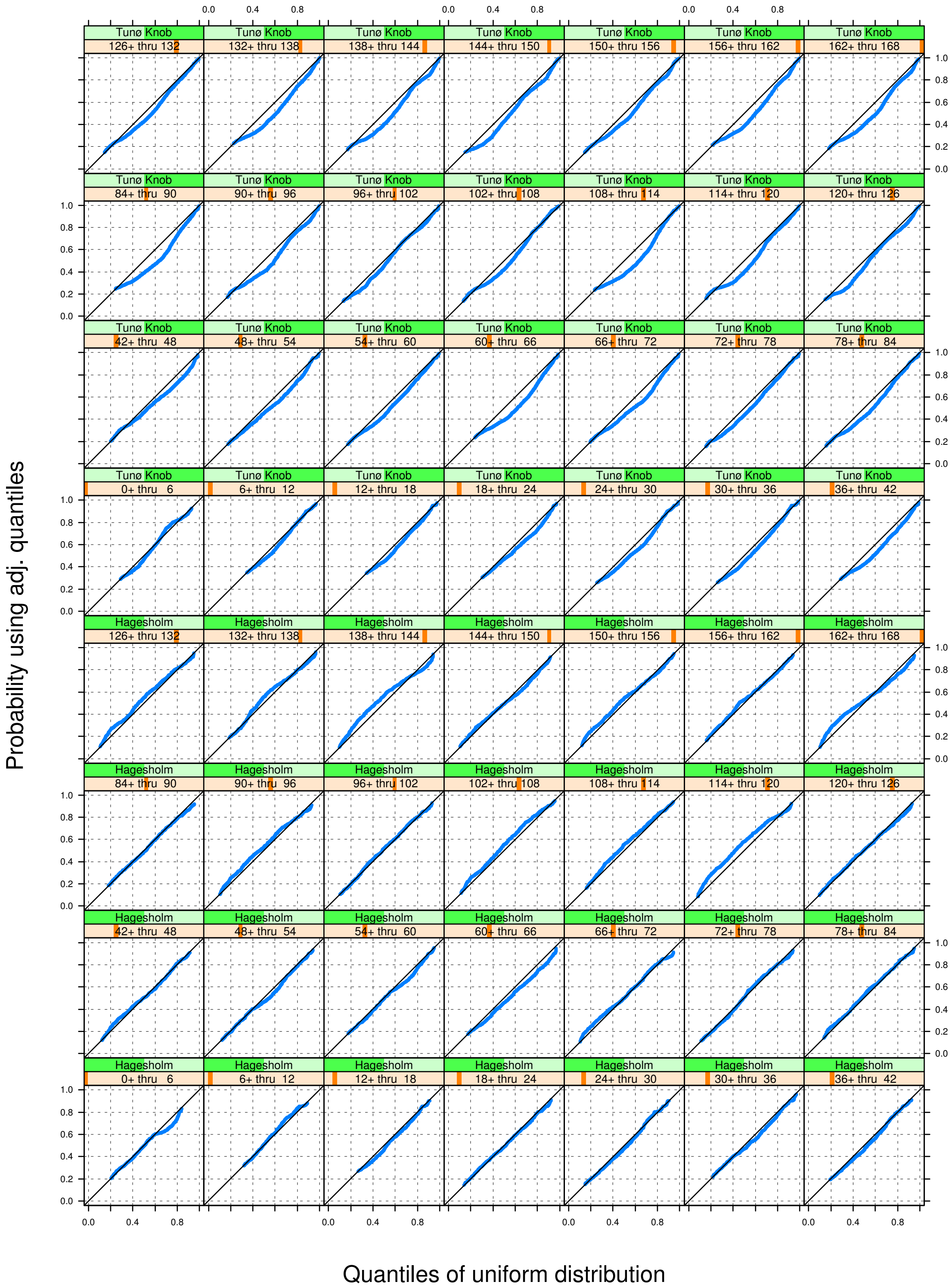


Figure 16: QQ-plots for adjusted quantiles (ECMWF, test period) for the two wind farms and for horizons in steps of 6 hours.

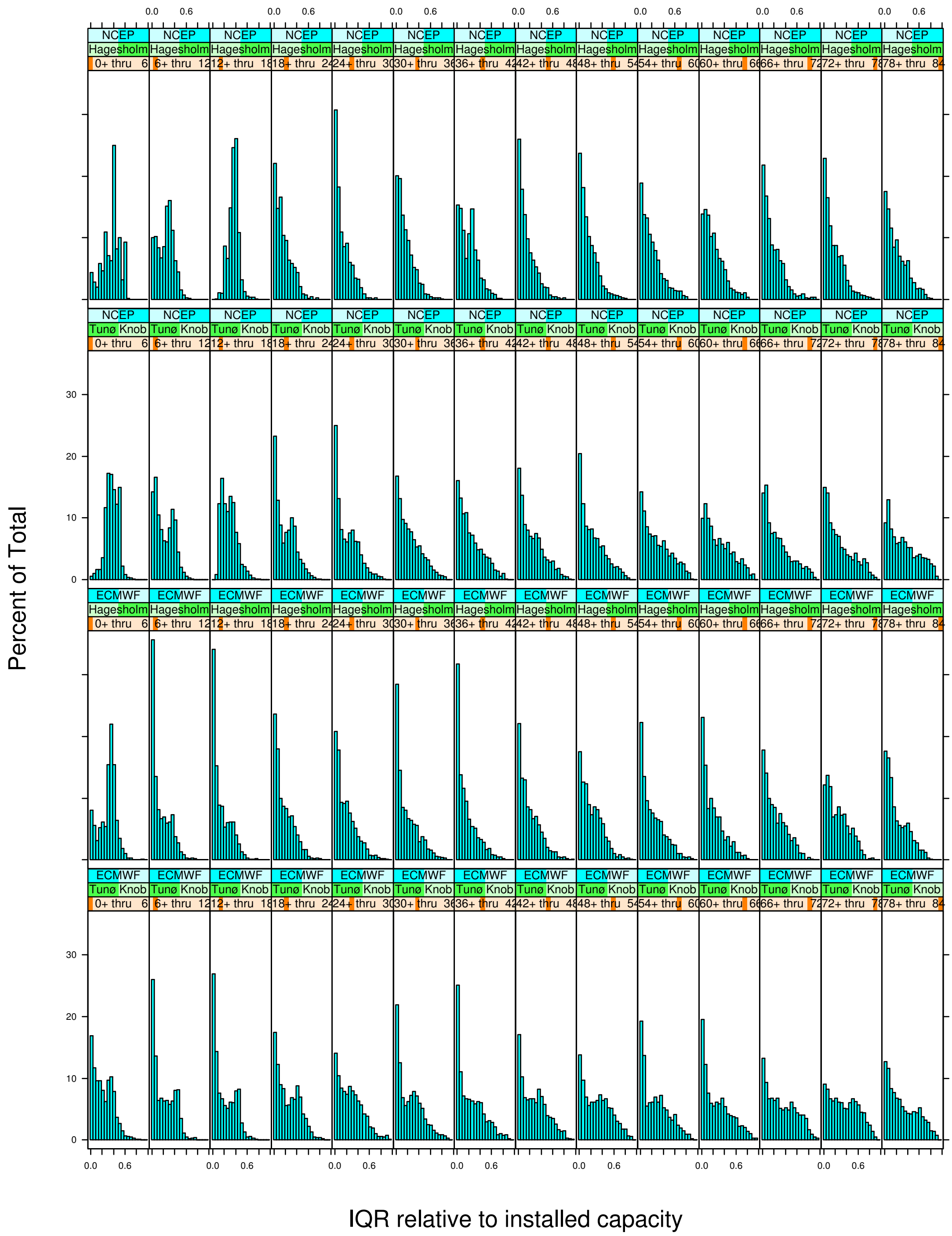


Figure 17: Histograms of IQR (test period) of horizons in steps of 6 hours up to 84 hours.