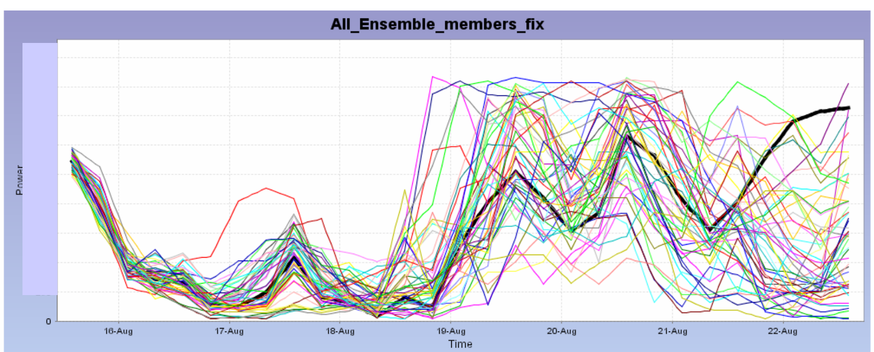
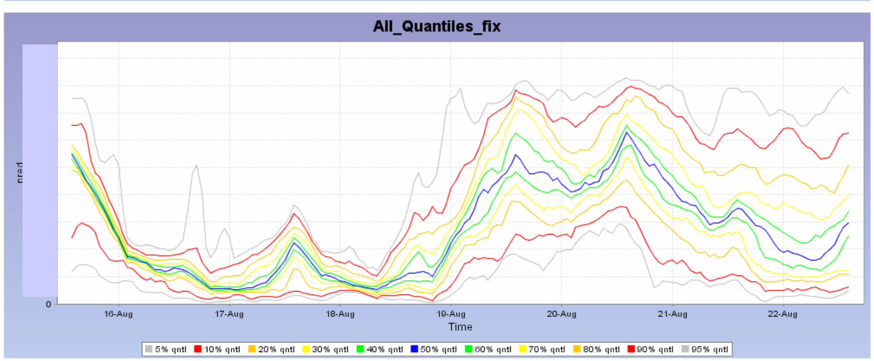
Risø-R-1527(EN)

Wind Power Prediction using Ensembles

Gregor Giebel (ed.), Jake Badger, Lars Landberg, Henrik Aalborg Nielsen, Torben Skov Nielsen, Henrik Madsen, Kai Sattler, Henrik Feddersen, Henrik Vedel, John Tøfting, Lars Kruse, Lars Voulund





Risø National Laboratory
Roskilde
Denmark
September 2005

Author: Gregor Giebel (ed.), Jake Badger, Lars Landberg, Henrik Aalborg Nielsen, Torben Skov Nielsen, Henrik Madsen, Kai Sattler, Henrik Feddersen, Henrik Vedel, John Tøfting, Lars Kruse, Lars Voulund

Title: Wind Power Prediction using Ensembles

Risø-R-1527(EN)
September 2005

Abstract (max. 2000 char.):

The Ensemble project investigated the use of meteorological ensemble fore-casts for the prognosis of uncertainty of the forecasts, and found a good method to make use of ensemble forecasts. This method was then tried based on ensembles from ECMWF in form of a demo application for both the Nysted offshore wind farm and the whole Jutland/Funen area. The utilities used these forecasts for maintenance planning, fuel consumption estimates and over-the-weekend trading on the Leipzig power exchange. Other notable scientific results include the better accuracy of forecasts made up from a simple superposition of two NWP provider (in our case, DMI and DWD), an investigation of the merits of a parameterisation of the turbulent kinetic energy within the delivered wind speed forecasts, and the finding that a "naïve" downscaling of each of the coarse ECMWF ensemble members with higher resolution HIRLAM did not improve the error scores or the result space enough to warrant the computational effort.

ISSN 0106-2840
ISBN 87-550-3464-0

Contract no.:
FU 2101 (Ordrenr. 101 295)

Group's own reg. no.:
1105 068-1

Sponsorship:
Elkraft System under PSO rules

Cover :
All individual ensemble members for western Denmark plotted, with the control forecast in black, and the derived adjusted quantiles.

Pages: 43
Tables: 1
References: 42

Risø National Laboratory
Information Service Department
P.O.Box 49
DK-4000 Roskilde
Denmark
Telephone +45 46774004
bibl@risoe.dk
Fax +45 46774013
www.risoe.dk

Contents

Preface

This report is the final report for the three-year project "Vindkraftforudsigelse med ensemble forecasting" (wind power prediction by ensemble forecasting, FU 2101), which ran from June 2002 to June 2005. It contains an overview of the results achieved in the project, but leaves the details to the specialised reports and papers published during the project. Participants were Risø National Laboratory, Informatics and Mathematical Modelling (IMM) of the Technical University of Denmark (DTU), the Danish Meteorological Institute (DMI), Energi E2 A/S, and Elsam A/S. SEAS A.m.b.A was initially part of the project, but the group participating was bought by E2 and therefore, only E2 continued as partner. Transmission companies Eltra and Elkraft System followed the progress and attended some meetings. The project budget was 5.59 million DKr, of which 3.03 million DKr was a grant under the PSO rules.

The project and its results were presented during the following conferences:

European Wind Energy Conference, Madrid, 2003:

G. Giebel, L. Landberg, J. Badger, K. Sattler, H. Feddersen, T.S. Nielsen, H.Aa. Nielsen, H. Madsen: *Using Ensemble Forecasting for Wind Power*. (Poster)

Global Wind Power Conference, Chicago 2004:

G. Giebel, J. Badger, L. Landberg, H.Aa. Nielsen, H. Madsen, K. Sattler, H. Feddersen: *Wind Power Forecasting Using Ensembles*. (Poster)

H.Aa. Nielsen, H. Madsen, T.S. Nielsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, H. Feddersen: *Wind Power Ensemble Forecasting.* (Poster)

European Wind Energy Conference, London 2004:

G. Giebel, A. Boone: *A Comparison of DMI-Hirlam and DWD-Lokalmodell for Short-Term Forecasting*. (Poster)

In all, 5 reports were written with the main results of the analyses:

H.Aa. Nielsen, T.S. Nielsen, H. Madsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, H. Feddersen: *Comparison of ensemble forecasts with the measurements from the meteorological mast at Risø National Laboratory.* Project report, 2004

H. Feddersen and K. Sattler: *Verification of wind forecasts for a set of experimental DMI-HIRLAM ensemble experiments*. DMI Scientific Report 05-01, 2005

A. Boone, G. Giebel, K. Sattler, H. Feddersen: *Analysis of HIRLAM including Turbulent Kinetic Energy*. Project report, 2005

H.Aa. Nielsen, T.S. Nielsen, H. Madsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, H. Feddersen: *Wind Power Ensemble Forecasting Using Wind Speed and Direction Ensembles from ECMWF or NCEP*. Project report, 2005

H.Aa. Nielsen, D. Yates, H. Madsen, T.S. Nielsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, H. Feddersen: *Analysis of the Results of an On-line Wind Power Quantile Forecasting System*. Project report, 2005

1 Introduction

Short-term forecasting of wind power is an important tool for utilities, to ensure grid stability and a favourable trading performance on the electricity markets. For the last 15 years, our groups have worked in the field and developed short-term prediction models being used operatively at the major Danish utilities since 1996. Short-term forecasting uses Numerical Weather Prediction (NWP) as input, transforms the wind speed predictions to power based on physics, statistics or both, and outputs the power predictions for the next (typically) 48 hours ahead as numbers or graphics.

A wealth of research (see [1] for an exhaustive overview) has shown that the NWP input is the most critical part for the performance of the prediction model. However, only a few cases exist where the quality of different NWP input for the same test case has been investigated. Since the quality of the NWP is outside the control of the short-term prediction providers, research in recent years has focused on the uncertainty of the forecasts. So far, generic uncertainty bands derived from historic accuracy have been used, but the short-term predictions would be much more useful if they could give an estimate of their accuracy at all times.

The next step in the development of the NWP models, apart of ever finer horizontal grid resolution, is to use ensemble forecasts. The ensembles used in this study are produced routinely at US NCEP (National Centers for Environmental Prediction) and at ECMWF (European Centre for Medium Range Weather Forecasts). We collected both since late 2002, and investigated them for their properties in regard to wind speed and power forecasting for Danish wind farms and meteorology masts.

Additional ways of creating ensembles of forecasts are to use ensembles based on lagged initial condition, formed from the overlapping forecasts created at different starting points, and multi-model ensembles, which stem from using the output of different models, preferably with different input models as well. In our project, we use the operational model runs of DMI-HIRLAM and DWD-Lokalmodell.

The idea behind the use of additional input for short-term forecasting is the connection between spread and skill of the forecast. The basic assumption here is that if the different model members are differing widely, then the forecast is very uncertain, while close model tracks mean that this particular weather situation can be forecasted with good accuracy. Landberg et al [6] have shown that this is not necessarily true for Poor Man's Ensembles (defined here as ensembles based on lagged initial condition), when just using the spread for a single point in time, but Pinson and Kariniotakis [2], and Lange and Heinemann [3] showed that using the temporal development of the forecasts, some assertion can be made on the uncertainty.

The high regional penetrations of wind energy can only be integrated successfully using a short-term prediction model, to predict the wind power production for some hours ahead. How many hours look-ahead time one needs, depends on the application. For the scheduling of power plants, the most important time scale is determined by the start-up times of the other power plants in the grid, from 1 hour for gas turbines up to 8 hours or more for the largest coal fired blocks. The markets determine the second time scale of interest. In Denmark, the main market for electricity is the NordPool, followed by the German market. NordPool regulations mandate trading for the next 24-h day at noon the day before. This means that the wind power forecasts have to be accurate for about 37 hours lead-time (1100 hours to 2400 hours next day). A third time scale is involved in maintenance planning, of power plants or the electrical grid. The ideal lead-time for this would be weeks or even months ahead.

The different lead-times can be served with different short-term prediction models. The scheduling of a quick-response grid can be done with time-series analysis models alone, based on past production. For the trading horizon, it is necessary to use a proper Numerical Weather Prediction (NWP) model, of the type used in the large meteorological centres. Actually, short-term prediction models based on NWP outperform time-series models already for 4-6 hours lead-time. However, for a week-ahead prediction, even the best NWP models are not quite good enough, and a month-ahead prediction is also theoretically hardly conceivable due to the inherent chaos in the atmosphere. For an introduction into short-term prediction models, refer to [4].

The typical short-term prediction model uses NWP data from one operational model, run by a meteorological service. *Eg*, the Danish Zephyr/WPPT model uses the four daily runs of the HIRLAM model of the Danish Meteorological Institute (DMI) as input. This is usually the local meteorological institute, since they know the best parameterisations for the local conditions and run the model with the highest resolution for the particular country.

However, using data from more than one met. institute can increase the resilience against errors, and possibly also the accuracy of the forecasts.

Besides the forecasted value for the power, the utilities also would like to have an estimate on the accuracy of the current prediction. It has been established [5, 6] that the errors of the NWP models are not highly dependent on the level of the wind speed. That means that the shape of the power curve has a large influence on the uncertainty of the forecasts: where the power curve is steep, the error is amplified, while in the flat parts of the power curve, the error becomes less relevant. Combining this with the historical performance of the model and a term depending on the horizon is the state-of-the-art in uncertainty prediction these days.

A completely different class of uncertainty could be introduced if we could assess how predictable the current weather situation is. This is where ensemble forecasting comes in. The idea in ensemble forecasting is to cover a larger part of the possible futures through introduction of variation in the initial conditions. This can be used as a sensitivity analysis on the influence of variations in different factors. There are four main possibilities for the calculation of ensembles of forecasts:

Every model run is started with a little variation in the initial conditions. In this way it can be investigated how sensitive the result is against small changes in the initial conditions (a.k.a. the butterfly effect). Even with these small variations, the initial conditions are still compatible with measured data within the likely error in the analysis. Keep in mind that the meteorological observational network has a density over land in the order of 20-50 km. The ECMWF runs this kind of ensemble twice a day with 50 members. NCEP in the US run another one with 11 members.

As a variation of the scheme above, an additional step can be done using a higher-resolution model nested in the ensemble output. At the University of Washington, they run a number of meso-scale models nested in the ensembles of NCEP. In our case, DMI-HIRLAM was run nested in all 51 members of the ensemble of ECMWF.

A multi-scheme ensemble starts with identical data input, but uses different variants of the same model. This can be different data assimilation techniques (optimal interpolation, 3D-Var or 4D-Var), different numerical integration schemes (Eulerian or Lagrangian) or different physical parameterisations. Dependent on the choices made, the model behaviour changes. Note that this does not necessarily mean that the different model set-ups individually do better or worse in the traditional verifications scores.

The ultimate in the above is a multi-model ensemble, using results from the operation models of different institutes, *eg* the Deutscher Wetterdienst or the US National Weather Service.

The easiest possibility is to use a poor man's ensemble. Since for instance DMI delivers a new model run every 6 hours for 48 hours in advance, this means that at every point in time there are up to 7 overlapping forecasts, done at different start times in the past. Note: meteorologists call the multi-model ensemble for poor man's ensemble, and have the name ensemble based on lagged initial condition reserved for the last one of the points above.

A typical way to plot ensemble members is the so-called "spaghetti plot", which shows a single contour for each ensemble member. Where there is agreement in the location of the contour line between members the level of uncertainty for the location of the contour from the ensemble prediction system is low. Conversely, where there is disagreement in the location, like in the centre of Figure 1, the level of uncertainty for the location of the contour is high.

The idea here is that for a hard to predict weather situation, there is large spread between the ensemble members, and the skill of the prediction is low. To which extent there really is a correlation between spread and skill, will be investigated in this project.

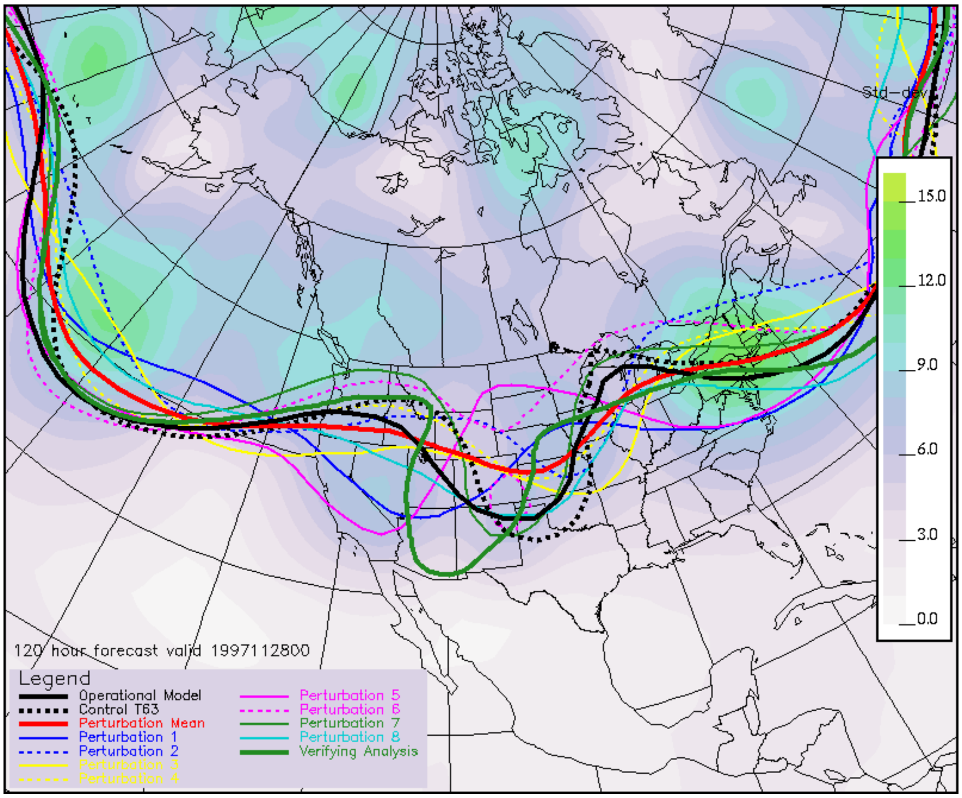


Figure 1: A spaghetti plot of geopotential heights over the continental US. Image Peter Houtekamer, Environment Canada

As just one example for meteorological ensemble forecasting consider e.g. the wind speed ensemble forecast from ECMWF shown in Figure 2, where the initialization corresponds to 12:00 (UTC) at Aug. 14, 2003. The left panel of Figure 2 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 there on, the ensemble members seems to diverge. 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. on 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 paper we will focus on the correctness of the ensemble quantiles after transformation to ensembles of power production.

As described in chapter 2 the spatial resolution of the meteorological ensemble model is approximately 75 km. Hence it cannot be expected that the ensemble quantiles are correct in a probabilistic sense when compared to a wind speed measurement or when ensembles of power production are compared to the actual production of a wind farm.

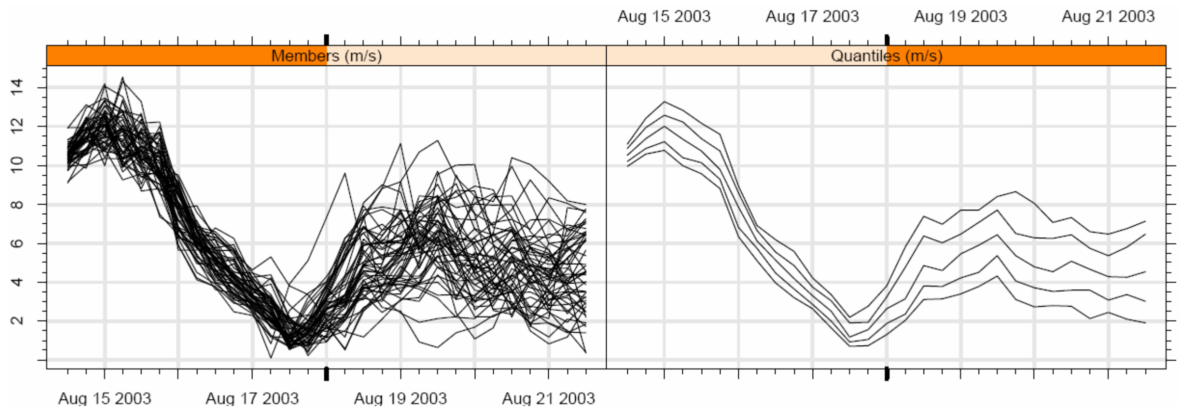


Figure 2: Individual ECMWF wind speed ensemble members and quantiles (10, 25, 50, 75, 90 percent) for the 7-day forecast starting at Aug. 14, 2003, 12:00 UTC

A number of groups in the field are currently investigating the benefits of ensemble forecasts.

Giebel *et al* [7] and Waldl and Giebel [8,9] investigated the relative merits of the Danish HIRLAM model, the older Deutschlandmodell of the DWD and a combination of both for a wind farm in Germany. There, the RMSE of the Deutschlandmodell was slightly better than the one of the Danish model, while a simple arithmetic mean of both models yields an even lower RMSE.

Moehrlen *et al* [10] use a multi-model ensemble of different parameterisation schemes within HIRLAM. They make the point that, with the spacing of the observational network being 30-40 km, it might be a better use of resources to run the NWP model not in the highest possible resolution (in the study 1.4 km), but use the computer cycles instead for calculating ensembles. A doubling of resolution means a factor 8 in running time (since one has to double the number of points in all four dimensions). The same effort could therefore be used to generate 8 ensemble members. The effects of lower resolution might not be so bad, since effects well below the spacing of the observational grid are mainly invented by the model anyway, and could be taken care of by using direction dependent roughness instead. This is only valid if the resolution is already good enough to properly represent fronts and meso-scale developments. Their group is also the leader of an EU-funded project called Honeymoon. One part of the project is to reduce the large-scale phase errors using ensemble prediction.

Landberg *et al* [6] used a poor man's ensemble to estimate the error of the forecast for one wind farm. The assumption is that when the forecasts change from one NWP run to the next, then the weather is hard to forecast and the error is large. However, this uncertainty forecast fared no better than an uncertainty derived from the wind speed level.

Roulston *et al* [11] evaluated the value of ECMWF forecasts for the power markets. Using a rather simple market model, they found that the best way to use the ensemble was what they called climatology conditioned on EPS (the ECMWF Ensemble Prediction System). The algorithm was to find 10 days in a reference set of historical forecasts for which the wind speed forecast at the site was closest to the current forecast. This set was then used to sample the probability distribution of the forecast. This was done for the 10^{th}, 50^{th} and 90^{th} percentile of the ensemble forecasts.

This report contains the main conclusions of 5 separate reports, plus the user experience of the participating utilities. In chapter 2, the different NWP models used are described. In chapter 3, the models used in conjunction with the ensemble forecasts are explained. In chapter 4, the scientific results are deliberated, while in chapter 5 the operational results of the demo application are in the foreground.

2 The meteorological (ensemble) systems

A general overview of ensemble systems and numerical weather prediction has been given already in the introduction. In this chapter, an overview is given over the various systems employed in this project, namely the NWP systems DMI-HIRLAM and Lokalmodell, as well as the dedicated ensemble systems from NCEP and ECMWF. Also in this chapter is a description of the set-up of the nested HIRLAM in every member of the ECMWF EPS, and a description of the dedicated parameterisations of the HIRLAM wind speed using turbulent kinetic energy.

2.1 DMI-HIRLAM

The HIgh Resolution Limited Area Model HIRLAM is a collaboration of the following meteorological institutes:

- Danish Meteorological Institute (DMI) (Denmark)
- Finnish Meteorological Institute (FMI) (Finland)
- Icelandic Meteorological Office (VI) (Iceland)
- Irish Meteorological Service (IMS) (Ireland)
- Royal Netherlands Meteorological Institute (KNMI) (The Netherlands)
- The Norwegian Meteorological Institute (DNMI) (Norway)
- Spanish Meteorological Institute (INM) (Spain)
- Swedish Meteorological and Hydrological Institute (SMHI) (Sweden)

There is a research cooperation with Météo-France (France).

The aim of the project is to develop and maintain a numerical short-range weather forecasting system for operational use by the participating institutes. The project has started in 1985. Since 1 January 2003 the project is in its sixth phase, HIRLAM-6. The HIRLAM-4 system, which to a large extent was built on the research results of the fourth phase of HIRLAM, is now used in routine weather forecasting by DMI, FMI, IMS, KNMI, met.no, INM, and SMHI.

A reference version of HIRLAM is maintained at the European Centre for Medium range Weather Forecasts (ECMWF), and all changes to HIRLAM are introduced via the reference system. Each HIRLAM member can obtain new versions via their links to ECMWF, or through the system manager.

The Danish set-up used in this study consists of four submodels, consisting of the identical mathematical core, each covering a part of the total domain in various resolutions [12]. Each model makes its own analysis every 3 hours in order to create an adequate initial state for that model. The nesting strategy including the analysis cycles for each model is described in [13]. The furthest out is HIRLAM-G, which covers an area with cornerpoints in Siberia, California, the Caribbean and Egypt, hence a good share of the Northern Hemisphere. This model has the lowest resolution, with a horizontal resolution of 48 km (0.45°) and a time step of 240 s. The boundary conditions for this model stem from the global ECMWF model [14], which is run twice a day and gives boundary conditions with a 6 hour time step. This model, like the HIRLAM models, has 40 vertical levels. The HIRLAM-G hands over the boundary conditions to two models with a 16km (0.15°) horizontal resolution and 90s time step, one (N) covering Greenland, the other (E) covering Europe. HIRLAM-E is then used to provide the boundary conditions to the model used here, HIRLAM-D, which covers Denmark and parts of northern Germany, western Sweden and southern Norway with a resolution of 5.5 km (0.05°), with a 30 s time step. The DMI provided us every 6 hours with forecasts in three-hourly time steps for up to 48 hours ahead. The grid points nearest to the farms were used.

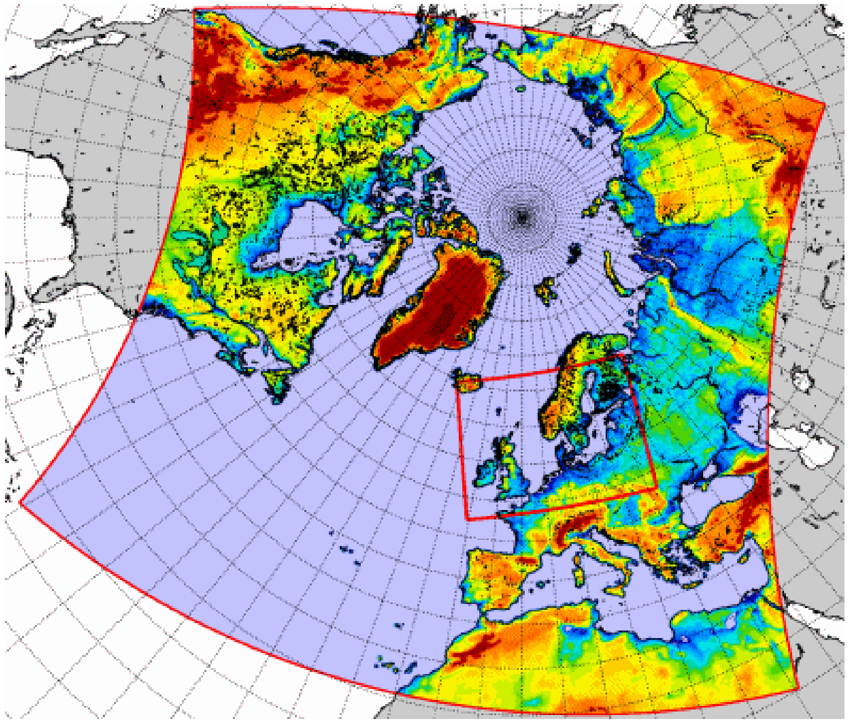


Figure 3: The actual set-up of DMI-HIRLAM since June 2004.

Please note that the DMI changed their operational set-up starting June 14, 2004, essentially switching to a new version of the HIRLAM reference system [17] and upping the resolution of the "G" model to 0.15° and expanding the area of the "D" model significantly, while leaving out the "N" and "E" steps (Figure 3). However, for the comparison of this study, only data from before the change was used.

2.2 DWD-Lokalmodell

The DWD had a major model change in December 1999. Before, the DWD used a model suite consisting of a global DWD model driving the Europamodell, which then contained nested the Deutschlandmodell. This model was used in the previous study. Now [15], only two models are employed: The GME as a global model on a icosahedral-hexagonal grid with 60km horizontal resolution and 31 vertical levels.

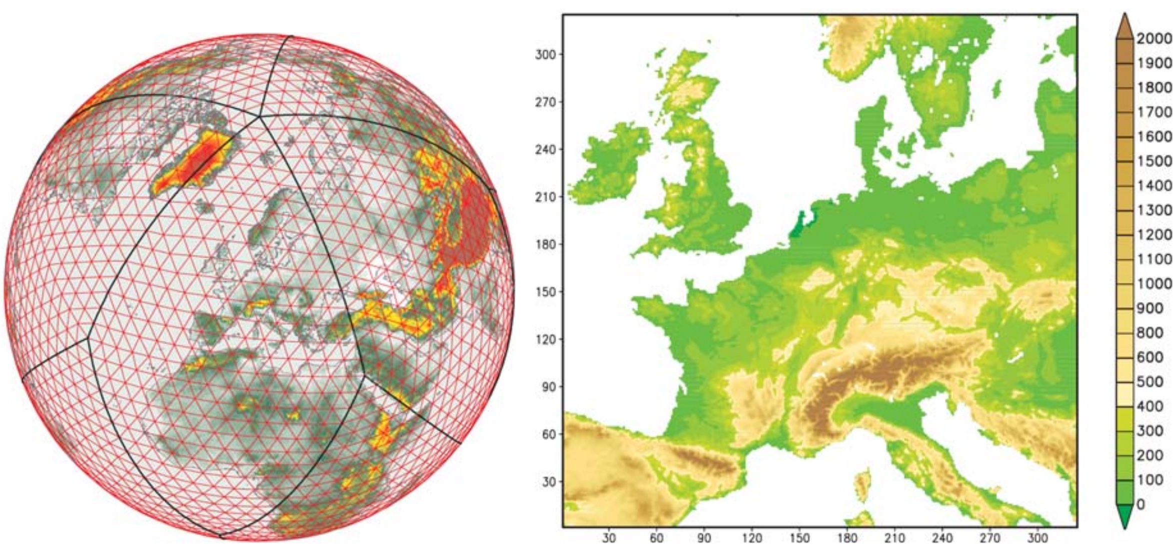


Figure 4: The grid of the Globalmodell (left) and the domain of the Lokalmodell (right). Source: [15]

The Lokalmodell is nested, with a 7km rotated spherical Arakawa-C grid. It calculates 35 vertical layers, 10 within first 1500m from the surface. A grid encompassing all of Denmark and a bit of the surroundings was delivered twice a day with forecasts starting a 00Z and 12Z, going out to 48 hours. The data period ran from 1 December 2002 to 30 November 2003, *ie* one year.

2.3 NCEP ensembles

NCEP (National Centers for Environmental Prediction, http://www.emc.ncep.noaa.gov/) in the US operates a twice daily ensemble of global weather forecasts. At 00Z, 12 forecasts are made comprising of 1 control (unperturbed forecast) at high resolution (T170 truncation from 0 to 168 hours and then T126 to 384 hours), 1 control at the ensemble resolution (T126 from 0 to 84 hours ahead and then T62 out to 384 hours), and 10 perturbed forecasts at the ensemble resolution. At 12Z, 1 control and 10 perturbed forecasts are made all at ensemble resolution. The model used is the same as used for the Global Forecast System GFS, which also gives deterministic forecasts for the entire globe.

The perturbations are constructed by a so-called “breeding cycle”. The breeding of one such perturbation is given as an example. It is started by creating an alternative initial state of the atmosphere by “seeding” the best guess state of atmosphere with a random perturbation. The two slightly different states of the atmosphere are integrated forward in time using the forecast model. One day later the difference between the two forecasts is used, with a rescaling, as the initial perturbation in the next forecast. The rescaling is done to limit the amplitude of the perturbation to a scale similar to that of errors in the initial best guess state of the atmosphere. The structures that emerge from this “breeding cycle” are the fastest growing non-linear perturbations and are thought to resemble the important errors in the initial best guess state of the atmosphere.

The data is collected automatically twice daily from ftp://ftpprd.ncep.noaa.gov/ in the form of GRIB files. These are then locally unpacked and for a region covering Europe and part of North Africa (12°W-25°E, 25°N-65°N) data from a selection of fields is extracted and stored as netCDF files. The meteorological fields stored are: the u-component and v-component of the wind at 10m and 850hPa, the temperature at 2m and 850hPa, and mean sea level pressure. The forecast data has a look-ahead time increment of 6 hours out to 7.5 days (3.5 days before 01/04/2004). The spatial resolution is 1° x 1° which is approximately 100 km.s.

2.4 ECMWF ensemble system

The ECWMF (European Centre for Medium-Range Weather Forecasts, located in Reading/UK) ensemble comprises one global 10-day control forecast and 50 similar forecasts that are based on small, initial perturbations to the control initial condition. The perturbations are based on so-called singular vectors, calculated so as to maximise the linear growth of the perturbation kinetic energy after 48 hours. Perturbations are calculated separately for the Northern and Southern Hemispheres and the Tropics. In addition, the effect of model errors is addressed by the use of ‘stochastic physics’ in the NWP model. For each perturbed ensemble member the combined effect of the physical parameterisations is randomly perturbed by up to 50% every six hours. The model output comes as six-hourly values for the wind speed in 10 m a.g.l.

Data storage from ECMWF is with the same meteorological fields and the same region as for NCEP. The forecast cycles begin at 12Z each day. The spatial resolution is approximately 75 km and the time horizon is 7 days.

2.5 HIRLAM ensembles based on ECMWF EPS

In our project we use both the raw ECMWF ensemble forecasts every six hours (horizontal resolution ~ 75 km), and we use the DMI-HIRLAM model to downscale the ECMWF ensemble forecasts to a horizontal resolution of approximately 20 km for the first 72 hours with output every hour. This is actually done with two different approaches: DMI-HIRLAM is nested in every of the 50 perturbed ensemble members plus the control forecast, and HIRLAM is run nested in the control forecast with 4 alternative convection and condensation schemes.

The version of HIRLAM used in this study is the one that has been developed for operational use at the Danish Meteorological Institute (DMI) [13], except that we have adapted the model domain and the horizontal resolution so they fit the areas of interest and the available computer resources in the present study.

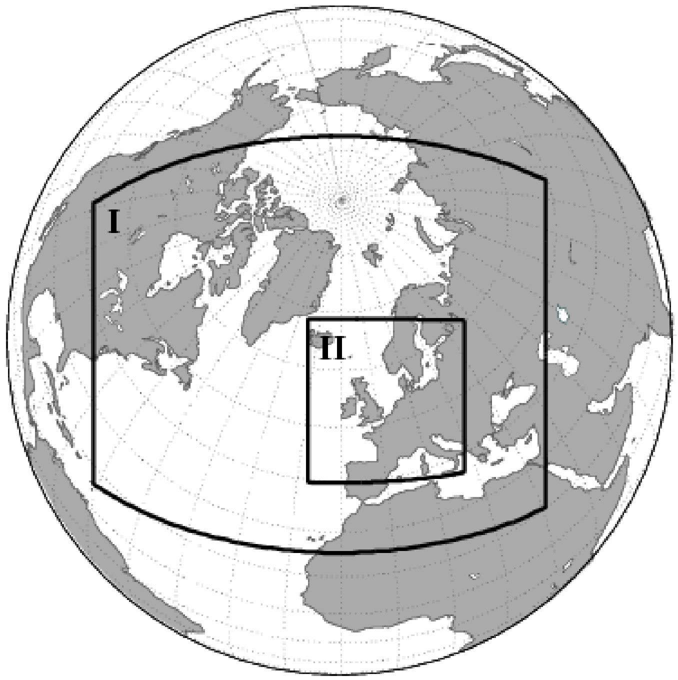


Figure 5: HIRLAM nesting. Horizontal resolution for outer and inner models is 0.6° and 0.2°, respectively. Boundary relaxation zones are included in the displayed domains.

The model setup for the present study is a doubly 1-way nested setup, illustrated in Figure 5. The outer HIRLAM model is nested to the global ECMWF ensemble prediction model [16] which is run at a horizontal resolution of T255 (corresponding to approximately 80km) with 31 vertical levels. A relaxation zone of 10 HIRLAM grid points in width surrounds the domains at the lateral boundaries. The outer HIRLAM model has a horizontal resolution of 0.6° (~66km) with 31 vertical levels. It is initialized at 12 UTC by the ECMWF T255L31 analysis data or one of its 50 perturbations. The lateral boundaries are updated by the respective control or perturbed forecast with a 6 hour frequency. An interpolation in time is applied between the boundary data updates in order to assure smooth transitions. The outer HIRLAM model utilizes semi-Lagrangian advection. This permits a time step of 600s in the dynamic part of the model as well as for the physical parameterizations.

The inner HIRLAM model has a horizontal resolution of 0.2° (22km) and also 31 vertical levels, the latter having the same structure as for the outer HIRLAM model. The inner model is initialized by the 0.6° outer HIRLAM model and updated hourly at its lateral boundaries by the outer HIRLAM model. The grid topologies of the two models differ only in the horizontal grid spacing. Thus, the boundary fields are treated in a much more consistent way than between the global model and the outer HIRLAM model, where the horizontal grids differ in projection type, and vertical interpolation is involved. The inner 0.2° HIRLAM model utilizes Eulerian advection and runs with a 90s time step for the dynamic part of the model. The time step for the physical parameterizations has been chosen to 540s.

2.6 HIRLAM with TKE

The HIRLAM model employs a number of prognostic variables to describe the state of the atmosphere, e.g. pressure, temperature and wind. As NWP models are being developed further and with growing availability of computing power, more prognostic variables are added. One of those has been the turbulent kinetic energy, which describes the energy content of turbulence on the sub-grid scale of the model [17]. The TKE represents some properties of air motion, which the wind variables of the model may not cover, because these act on the resolved scales of the model grid. As a wind turbine is affected also by winds on smaller scales than what the model resolves, it seems worth investigating the potential impact of the kinetic energy connected to TKE on wind power predictions.

The kinetic energy of air motion per volume unit reads

$E_{kin} = ½ \rho v^2$.

In HIRLAM, the air density ρ and the wind velocity v are related to the scales resolved by the model grid. In the case of the HIRLAM-T model at DMI, the model grid has about 15km mesh size, which means that wind velocity in the model and also air density refer to an area of about 225 square kilometres. The following ansatz includes contributions from smaller scales on the kinetic energy flux in the model by making use of TKE:

$F_{kin} = \rho |v| (½ v^2 + \alpha E_{TK})$.

E_{TK} is the TKE, and α is an effect factor, which was chosen to be 0.33 in this work.

In order to utilize TKE for the wind power prediction in Zephyr, the kinetic energy flux is calculated for 70m height and a ``proxy'' wind speed determined by

$$w_{70e} = \left[\frac{2F_{kin}}{\rho_n}\right]^{\frac{1}{3}},$$

where ρ_n is the normalization density ρ_n=1.226kg/m^3.

Two other approaches are considered, too. They are based on wind gust diagnosis, and the first one is a physical approach employing TKE [18]. It determines wind gust on basis of the large scale wind, turbulence and thermodynamic stability of the atmosphere. The method considers the amount of energy available from turbulent eddies to transport an air parcel from within the atmospheric boundary layer (ABL) down to the surface (see Figure 6).

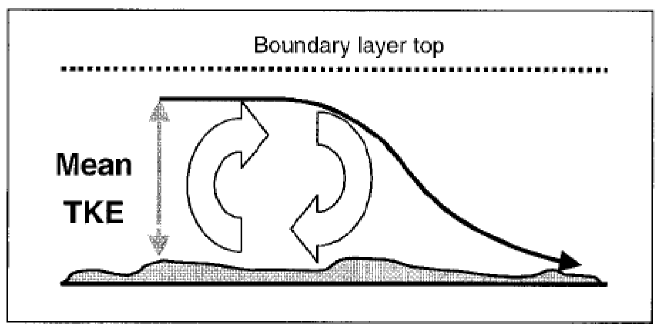


Figure 6: Principle of the approach of Brasseur (2001) for determination of wind gust near the surface (Figure from [18]).

This is compared to the energy necessary to overcome the energy barrier due to atmospheric stability, which suppresses the vertical downward movement. The comparison demands that the amount of energy available for the downward transport of an air parcel is at least as large as the potential energy of the air parcel before its displacement:

$$\frac{1}{z_p}\int_0^{z_p} E_{\mathrm{TK}}(z)\mathrm{d}z \geqslant \int_0^{z_p} g\frac{\Delta\theta_v(z)}{\theta_v(z)}\mathrm{d}z,$$

where θ_v denotes virtual potential temperature, g is the gravity constant, and z_p denotes the vertical level under consideration. The wind gust at the surface is then determined by the maximum wind speed at one of the levels z_p:

$$w_{10g,b} = \max\left\lfloor\sqrt{u^2(z_p)+v^2(z_p)}\right\rfloor.$$

The second approach for determining wind gust is based on similarity theory [19]. This method assumes that wind gust is a function of surface stress, height and the depth of the Atmospheric Boundary Layer (ABL) under stable atmospheric conditions. If the atmosphere is unstable there is an additional dependency on the buoyant production of TKE. The wind gust is thus expressed as

$$W_{10g} = w_{10}\left(1 + c_n u_{*0} + \gamma_u c_b{}^3\sqrt{B_p h}\right),$$

where w_{10} denotes the 10m average wind speed. The gust factor in parenthesis includes the surface stress expressed by the friction velocity u_{*0}, and the buoyancy term to the right depends on the ABL height h and the buoyant production of TKE

$$B_p = \left(\frac{g}{\theta_0}\right)\frac{H_{v,0}}{\rho_0 c_p},$$

which depends on the potential temperature θ_0, air density ρ_0 and the virtual sensible heat flux $H_{v,0}$ at the surface, c_p is the specific heat capacity of air at constant pressure. The other constants in the gust factor term are:

$$\gamma_u = \{^{0\ \text{stable}}_{1\ \text{unstable}}, \quad c_n = 5.2, \quad c_b = 1.44.$$

In cases where the atmosphere is thermodynamically stable the wind gust only depends on the surface friction and not on buoyancy (γ_u=0). The latter can increase the gust factor significantly under unstable conditions (Figure 7).

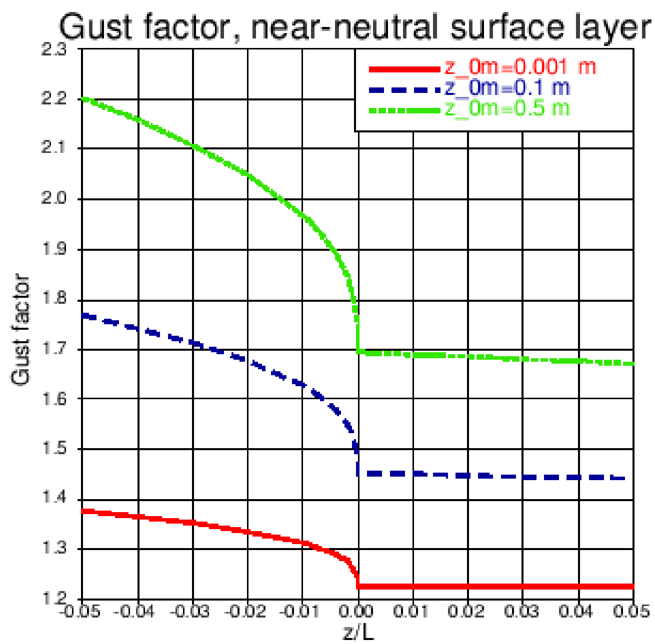


Figure 7: Gust factor w10g/w10 as function of atmospheric stability (z/L), plotted for different amounts of surface roughness. Under stable conditions (z/L>0), the gust factor remains constant for a certain surface stress.

3 The models used

Initially, we hoped to integrate the uncertainty forecasts in Zephyr/WPPT. This was soon changed to a distribution of more science and less implementation within the project. Therefore, both Risø and IMM used their own tools (Zephyr/Prediktor and a scientific implementation of the model structure of Zephyr/WPPT in the programming language S-Plus, respectively) for the analysis. Both models have been described elsewhere, therefore we only want to give a brief overview of both model structures.

3.1 Zephyr/Prediktor

The Prediktor model (nowadays part of the model suite of the Zephyr collaboration of Risø and DTU) has been around for quite some time. It started when Landberg [20] developed a short-term prediction model based on physical reasoning similar to the methodology developed for the European Wind Atlas [21]. The idea is to use the wind speed and direction from a NWP, then transform this wind to the local site, then to use the power curve and finally to modify this with the park efficiency. This general idea is shown in Figure 8. Note that the statistical improvement module MOS can either set in before the transformation to the local wind, or before the transformation to power, or at the end of the model chain trying to change the power. A combination of all these is also possible. Landberg used the Danish or Risø version for all the parts in the model: the HIRLAM model of the DMI as NWP input, the WAsP model from Risø to convert the wind to the local conditions and the Risø PARK model (now integrated into WAsP) to account for the lower output in a wind park due to wake effects. Two general possibilities for the transformation of the HIRLAM wind to the local conditions exist: the wind could be from one of the higher levels in the atmosphere, and hence be treated as a geostrophic wind, or the wind could be the NWPs offering for the wind in 10m a.g.l. Usually this wind will not be very accurately tailored to the local conditions, but will be a rather general wind over an average roughness representative for the area modelled at the grid point. In the NWP, even orography on a scale smaller than the spatial resolution of the model is frequently parameterised as roughness. If the wind from the upper level is used, the procedure is as follows: from the geostrophic wind and the local roughness, the friction velocity u_* is calculated using the geostrophic drag law. This is then used in the logarithmic height profile, again together with the local roughness. If the wind is already the 10m-wind, then the logarithmic profile can be used directly.

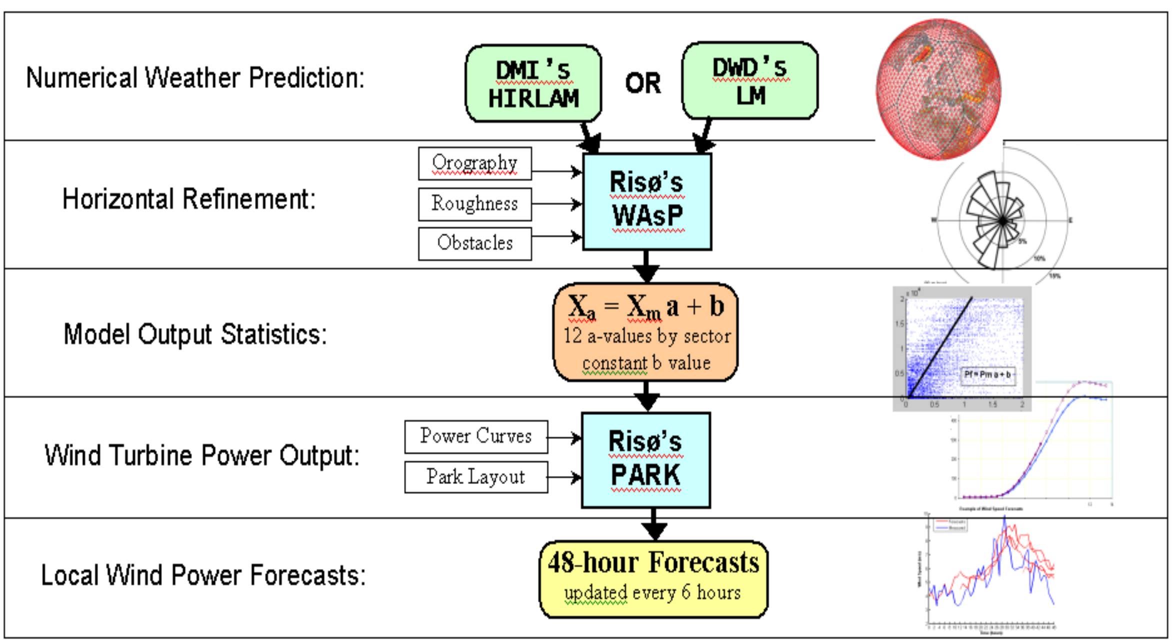


Figure 8: The set-up of the Risø Zephyr/Prediktor model.

The site assessment regarding roughness is done as input for WAsP. There, either a roughness rose or a roughness map is needed. From this, WAsP determines an average roughness at hub height. This is the roughness used in the geostrophic drag law or the logarithmic profile. In his original work, Landberg [20] determined the ideal HIRLAM level to be modelling level 27, since this gave the best results. However, the DMI changed the operational HIRLAM model in June 1998, and Joensen *et al* [22] found that after the change the 10m-wind was much better than the winds from the higher levels. So in the last iterations of the Risø model, the 10m-wind is used. Of course, this analysis has to be done for each new weather model.

The model ran operatively in the dispatch centre of SEAS and Elkraft, the utilities for Sjælland (Eastern Denmark). However, it lacks an upscaling model to relate the wind farm predictions to the output of the whole region.

Landberg actually showed later [23] that the mathematics of the wind power prediction are close to linear, except for the wind direction (when derived from the geostrophic wind) and of course, the power curve. The benefits of a Kalman filter as MOS for Prediktor has been researched by Giebel [24]. Since generally the error in the NWP varies randomly and not so much in a correlated fashion, the use for an adaptive system such as the Kalman filter is limited. While it is possible to tune the filter to yield better results than a static MOS system, the improvements are pronounced in only few cases. Part of this is the need for the filter to be relatively "stiff" (non-varying between two data points) to discourage runaway behaviour. However, even with a standard (static) MOS, the results of Prediktor can be improved, in some cases drastically so [7]. See also the paper on the Anemos model comparison elsewhere in these proceedings, where for some cases, the improvement of a MOS system over the pure physical approach is very marked.

3.2 The Power Curve model

Since a wind speed measurement is rarely representative for a whole wind farm and since the availability/reliability of such measurements often is low in reality it is decided to develop a method not using measurements of wind speed. 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 [25, 26, 27]. 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.

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

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

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 fraction of 10 kW which is smaller than all observations and $\overline{P}$ as the smallest fraction 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 and deviations from the logistic shape, and to adjust for bias originating from the uncertainty of the forecast the transformed power output is modelled as

$$y = f(u, v) + g(u, v)\,\tau + e, \qquad (2)$$

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 [28]. 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 [29, 30]. 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 [31] 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 9 displays the resulting estimate of the power curve of Tunø Knob. A rather large dependence on the forecasted wind direction is evident. Furthermore, the right panel of the figure 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.

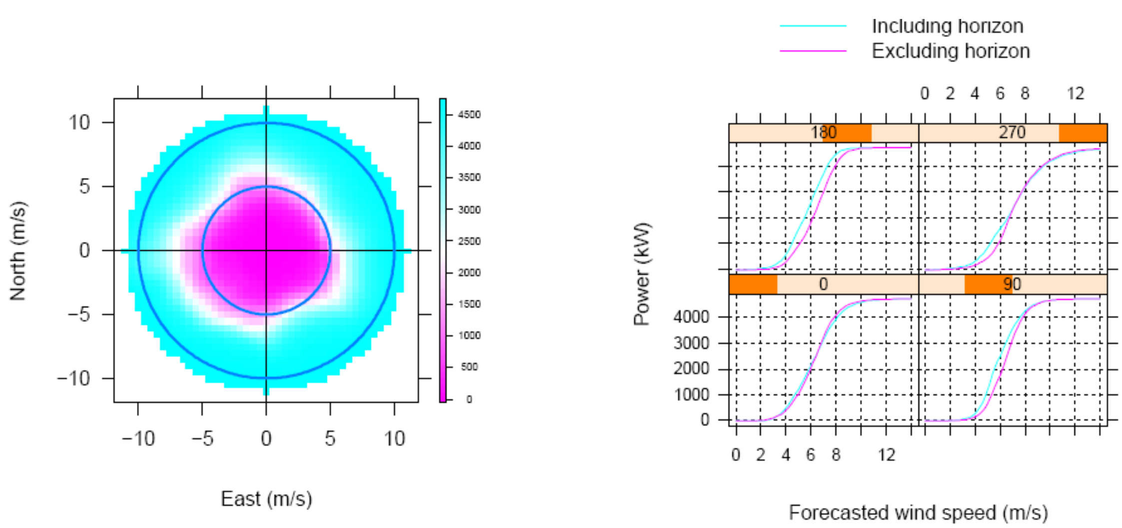


Figure 9: Left panel: Level plot of the direction dependent power curve for Tunø Knob. The circles mark 5 and 10 m/s forecasted wind speed. Right panel: Power curves for selected wind directions (degrees) for the model corresponding to the left panel and for a model where the influence from the horizon, i.e. g(u,v) τ, is excluded from (2)..

4 Scientific results

4.1 Quantiles for beginners

At the beginning of the project, we just knew the general principle of meteorological ensembles. You take plenty of data, initialise your NWP model with it in some clever way, and let it run forward from that point in time. This is the so-called deterministic forecast, which is the best bet of the coming weather. You then introduce some variation to that initial guess of the weather, so far that it still is consistent with the measurements. Since the measurements are not everywhere, some assimilation has to be done to acquire the current state of the atmosphere. This data assimilation can usually be done in a few somewhat different ways, thereby changing the atmospheric fields used as input for the NWP slightly. The idea with the meteorological ensembles is hence to try to transfer the uncertainty in the measurements and assimilation by running the same meteorological model with different starting conditions into an uncertainty of the outcomes. Different institutes use different techniques: the ECMWF uses singular vectors, NCEP bred modes (see chapter 2 for a discussion). The result is a certain spread in the ensembles. Since a NWP model has roughly 10^7 degrees of freedom, not all initial variations in the data can be reasonably tried out. Therefore the meteorological institutes have developed means of making sure that the variation brought into the model input is yielding a large variation in output. ECMWF uses a target horizon of 2 days, after which the variation should show a sizeable effect. However, since the main aim of ensemble forecasts is disaster prevention, it is not necessarily aimed at presenting a probabilistically correct distribution of the different verification scores – the correct finding of extreme weather is more important in this respect. This had been a central assumption we had when we started the project.

The way from the initial 51 members towards a probabilistic distribution is shown in Figure 2. From the spaghetti plot, a section is done at each horizon, and a frequency distribution is found. From this, the quantiles are estimated. In the long run, if these quantiles are probabilistically correct, then one would expect that the 80% quantile (or fractile) is exceeded in 20% of all cases. The way to establish whether that is true is through rank histograms. 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 [32]. 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 standard tool for comparing distributions [33]. 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 10 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. For this reason model (2) including the horizon is preferred.

Since we can see that the quantiles in Figure 10 are not very close to the line of identity, the quantiles as calculated straight from the ensemble distributions cannot be used directly to predict the uncertainty of a forecast. Therefore, we had to find a more clever way to get to the information contained in the ensemble spread.

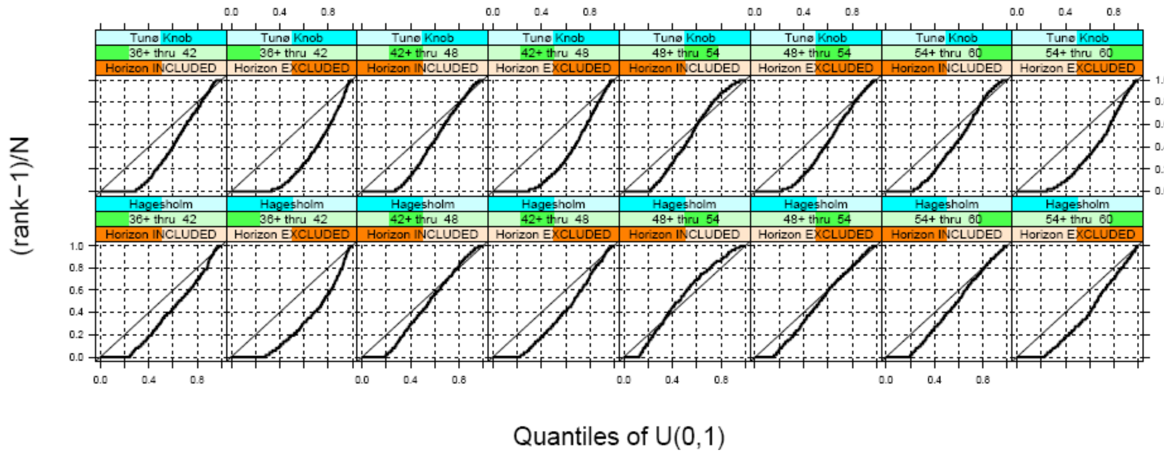


Figure 10: QQ-plots of (r-1)/N for the two sites for horizons ranging from 36 to 60 hours and when including or excluding the horizon when estimating the power curve using model (2). The plots are based on training data, and use the ECMWF ensemble.

4.2 Advanced quantiles

As seen from Figure 10 quantiles based on the power ensembles obtained as described in chapter 4.1 can not be interpreted in a strict probabilistic sense. However, based on plots like Figure 10 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 reach zero; for this reason we cannot 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 chapter 4.1. 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 probabilistic 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 11 displays some common properties of the data. It is seen that quite strong fluctuations as a function of the horizon 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 4 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 [29, 30] 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 [34]. The coefficients of the spline are estimated non-parametrically as smooth functions of the horizon τ.

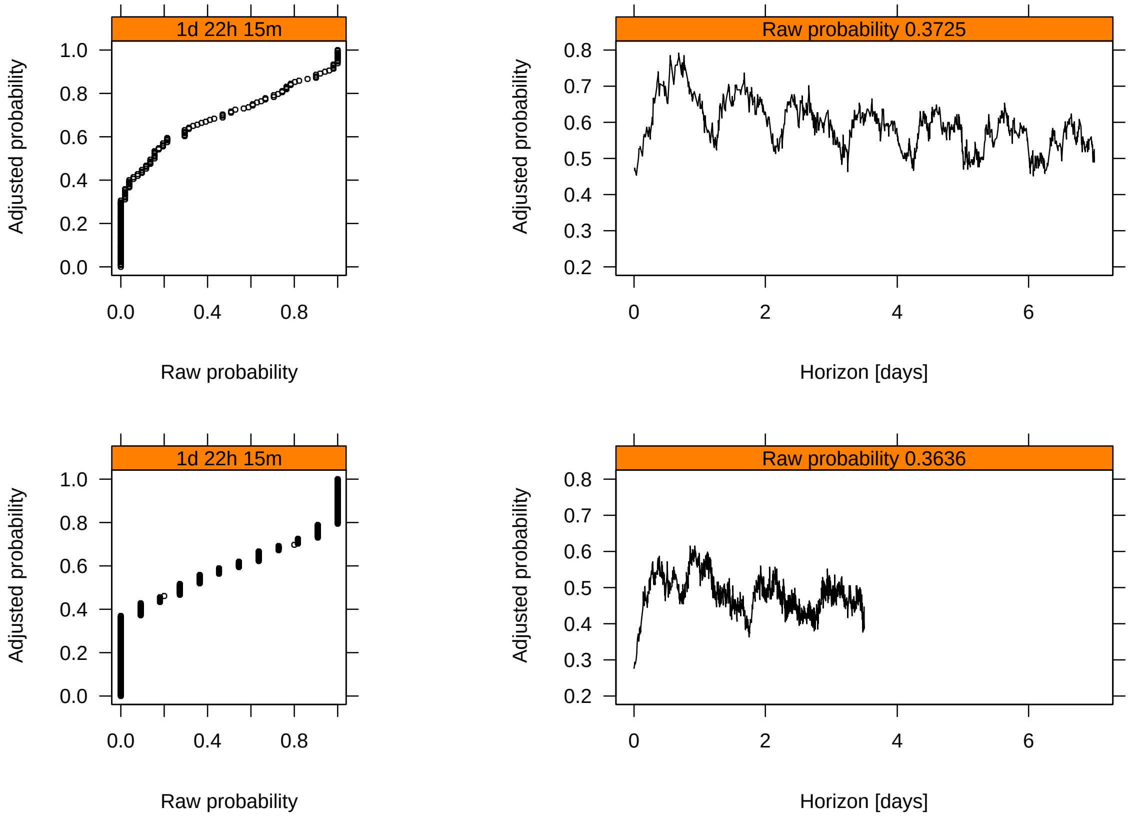


Figure 11: Examples of the dependence of the adjusted probability on the raw probability given a specific horizon (left) and on the horizon given a specific raw probability (right), for the ECMWF ensembles (top) and NCEP (bottom).

The model is formally written

$$\log\left(\frac{p_a}{1-p_a}\right) = B(p_r)\theta(\tau) + e, \qquad (3)$$

where p_a is the adjusted probability, $B(p_r)$ is a matrix representing a spline basis expansion of the raw probability p_r, $\theta(\tau)$ is the vector of smooth functions, and e is the error. Estimation is performed using local regression with a fixed bandwidth and a tricube weight function. Due to the presence of peaks (Figure 11, right), the functions $\theta(\tau)$ are locally approximated by 2nd order polynomials. The software LFLM [31] is used. 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.

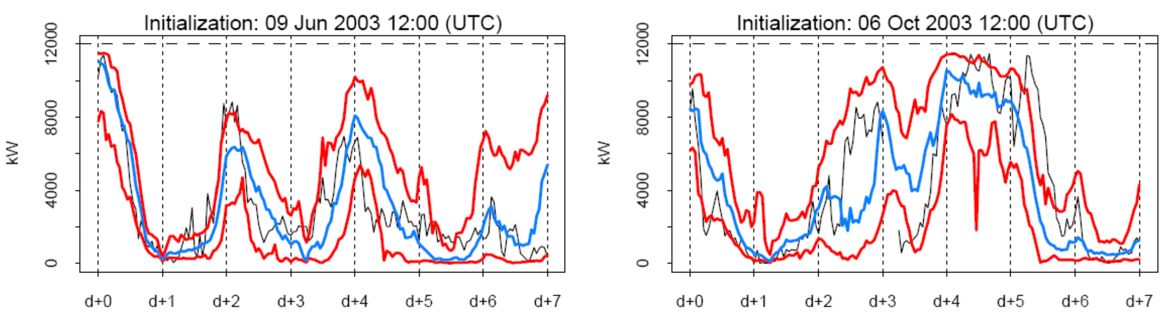


Figure 12: Examples of quantile forecasts for Hagesholm. The blue line indicates the median while the two red lines indicate the 25% and 75% quantiles.

Figure 12 shows two examples of the resulting quantile forecasts. In the following section, the performance of an actual system based on the results above is described.

4.3 Scientific results of the demo application

Section 5 and a larger report [36] describe a demo-application developed based on the models and methods described above. The application uses ECMWF-ensemble forecasts. Based on a period of operation from July 28, 2004 until March 31, 2005 an analysis of the performance was completed. Two instances of the application were run:

- For the Elsam-setup the total production in Jylland/Fynen, with the exception of Horns Rev, is forecasted on an hourly basis.
- For the E2-setup the production at Nysted Offshore is forecasted at on a 15 minute basis.

Results from these two very different setups were analyzed. Originally it was planned to re-calibrate the models on a monthly basis. However, this did not actually occur; the E2-setup were re-calibrated four times and the Elsam-setup only twice. Therefore we both consider the original runs and so-called reruns where the system were run in retrospect with monthly re-calibration. Reruns were also performed using NCEP-ensemble forecasts.

The quality of the quantile forecasts are evaluated using four different criteria:

Reliability addresses whether quantile forecasts are indeed quantiles by comparing the nominal value of the quantile to the number of times the observation is actually below the quantile considered. Repeating this procedure for a range of quantiles allows a plot of the actual versus nominal values to be constructed. Ideally, the resulting line should be the line of identity. In order to perform this analysis the data must be grouped. In this report we group the data by ranges of horizons.

Sharpness measures the average uncertainty indicated by the quantile forecast system by addressing the difference between forecasted quantiles which are symmetric about the 50% quantile. The sharpness is calculated individually for each horizon and quantified both in terms of the mean and the median of the difference in quantiles.

Resolution measures the variation in uncertainty indicated by the quantile forecast system by addressing the same basic quantities as for sharpness, but by calculating measures of variation instead of e.g. the mean. The resolution is calculated individually for each horizon and quantified both in terms of the standard deviation (SD) and the median absolute deviation (MAD).

The spread / skill relationship address the relation between the magnitude of the error of a point forecast, i.e. a forecast consisting of a single value for each horizon / forecast time, and the uncertainty indicated by the quantile forecast system. For this analysis there should be a *tendency* of larger errors when the uncertainty indicated by the quantile forecast system is high. Note that it is the *spread* of the errors which should be reflected by the quantile forecast system.

4.3.1 Reliability

The ECMWF ensemble system is initialized daily at 12: 00 (UTC). Taking into account the calculation time, but disregarding daylights savings and the one hour difference between UTC and the time zone in the NordPool area, the relevant horizons for bidding at the NordPool market are 36 to 60 hours. Figure 13 shows reliability diagrams for these horizons. These diagrams and tables are constructed by counting actual (relative) number of times the observation is below a quantile forecast with a given nominal value. Ideally, when the period under consideration is very long, there should be a complete agreement between the actual and nominal values. Unfortunately, for finite sized periods, it is not simple to quantify the deviation which can occur by sheer chance [35].

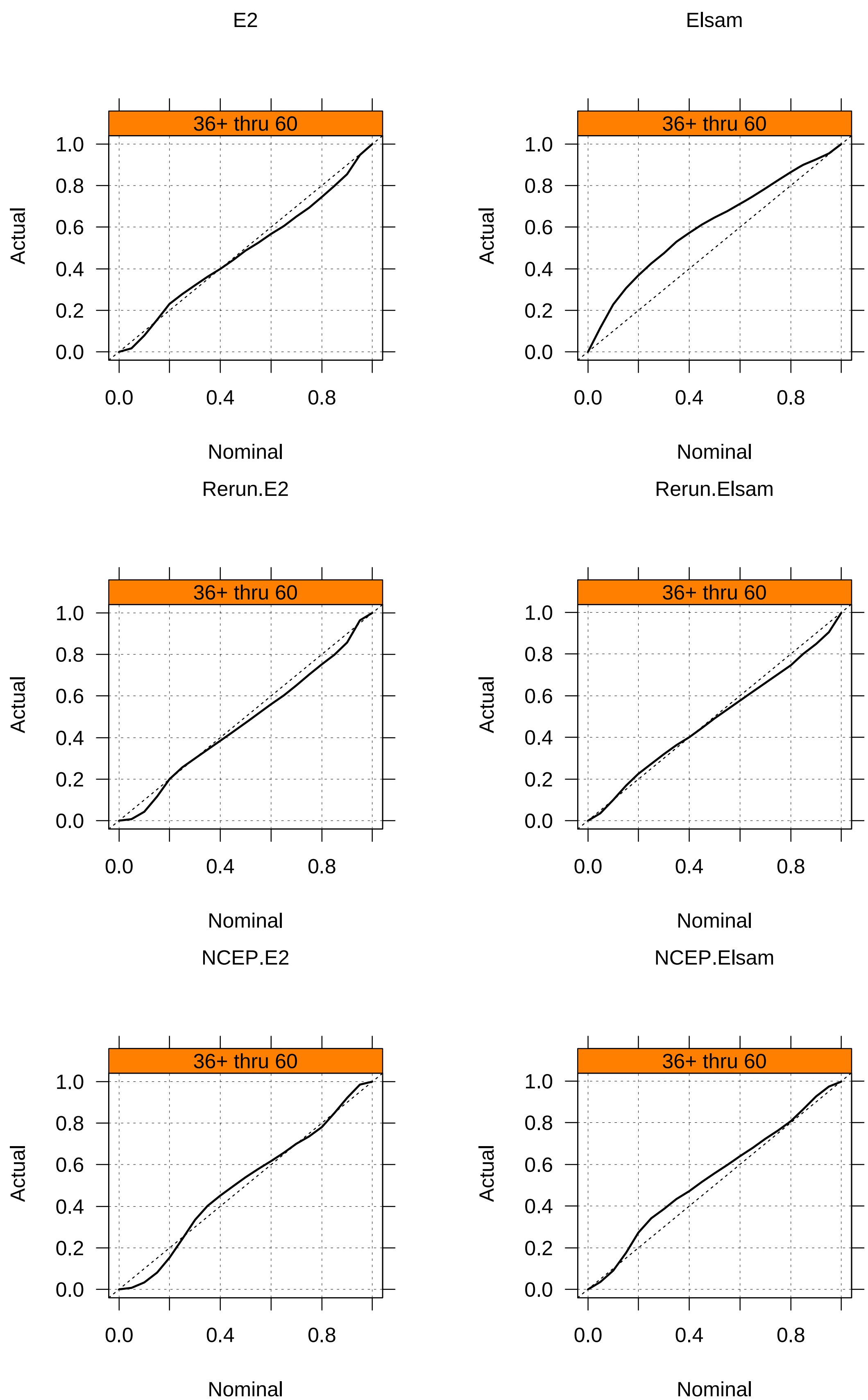


Figure 13: Reliability for horizon most relevant for trading on NordPool.

As noted above the re-calibration or training has been performed on a more regular basis for E2 than for Elsam.

This fact seem to be reflected in the reliability diagrams of the original runs; for E2 the line is much closer to the line of identity than for Elsam. For the rerun, i.e. the regular calibration, nearly the same curve is obtained for E2 and for Elsam the reliability is markedly improved when calibrated at regular basis. This highlights the importance of adaptive methods or at least regular re-calibration[1]. The rerun using the NCEP ensembles show a somewhat similar performance, but especially for Elsam the deviation from the line of identity is not quite satisfactory.

A detailed analysis [36, Section 4] indicates that for the Elsam setup there seems to be a benefit in using the ECMWF ensembles rather than the NCEP ensembles. For the E2 setup there seem to be a very small benefit of NCEP over ECMWF.

4.3.2 Sharpness

Here sharpness of a quantile forecast is defined as the average or median size of the Inter Quartile Range (IQR), i.e. the difference between the 1st and the 3rd quartile (the 25% and 75% quantiles). Figure 14 shows these numbers, based on values of IQR normalized using the installed capacity, for each horizon. Due to the issues regarding reliability the original runs will not be commented further here.

Comparing the E2 and Elsam setup it is seen that the Elsam setup results in sharper forecasts than the E2 setup. This is expected since the Elsam setup covers a large geographic region where errors tend to cancel each other. Also, for the larger region, the typical power output changes much more smoothly than for the single farm at Nysted, and the extremes (full power or zero power) are reached much less frequently. Comparing the NCEP and the ECMWF ensembles these seem to have approximately the same sharpness for the Elsam setup. However for the horizons mainly relevant for NordPool the ECMWF ensembles seems to result in somewhat sharper forecasts, especially for the horizons 36 to 48 hours. For the E2 setup this effect is even more pronounced.

In [36, App B] similar plots for the difference of other quantiles symmetric about the median is displayed. It is notable that, considering the horizons relevant for NordPool, the difference between the 90% and 10% quantile forecast is on average not more than 30% of the installed capacity when considering the Elsam setup. For the E2 setup the corresponding value is approximately 70%. A possible explanation is the difference in geographical regions considered in the two setups. Note also that even for the longest horizons the measures sharpness does not exceed 60% of the installed capacity for the Elsam setup. For the E2 setup the corresponding number is 90%.

4.3.3 Resolution

Here resolution of a quantile forecast is defined as the SD (Standard Deviation) and MAD (Median Absolute Deviation)[2] of the IQR. High values of SD and MAD indicate that the forecasting system is able to separate situations with high and low uncertainty. Given reliable systems with similar sharpness, the system with high resolution (high SD and MAD) may be preferable (assuming the resolution to be a real phenomenon and not originating from random noise on the estimates).

Figure 15 shows the values of SD and MAD, based on values of IQR normalized using the installed capacity, for each horizon. Due to the issues regarding reliability the original runs will not be commented further here.

[1] Given daily updates of the power data this can easily be done since the re-calibration of the statistical models takes less than 10 minutes, when using six months of data.

[2] The median absolute deviation is multiplied by 1.4826, whereby it is approximately equal to the standard deviation for large Gaussian samples.

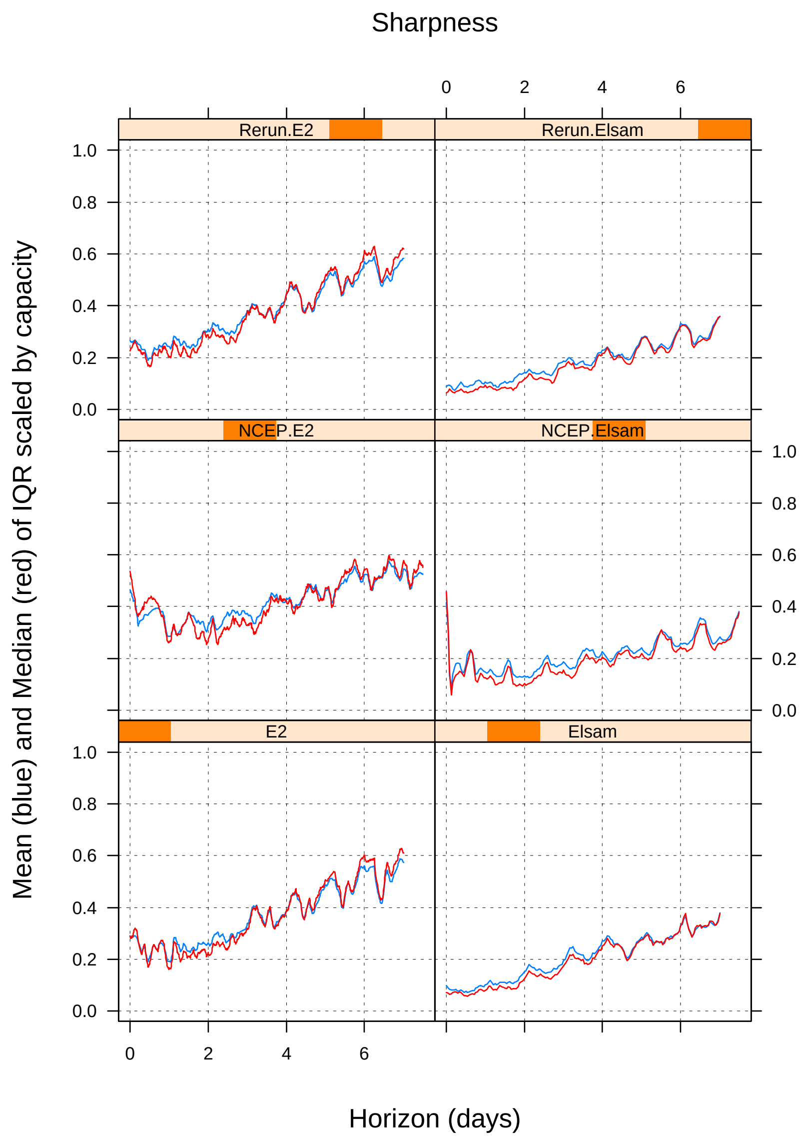


Figure 14: Sharpness compared to installed capacity.

First the Elsam setup is considered. Disregarding approximately the first 36 hours, where especially the NCEP-based quantile forecasts are not very reliable, it is seen that the systems have approximately the same resolution. The NCEP-based quantile forecasts seem to have slightly higher resolution than the ECMWF-based forecasts.

Considering the E2 setup the NCEP-based quantile forecasts seem to have slightly higher resolution than the ECMWF-based setup. Note that for horizons longer than four days the resolution for the E2-setup drops as the horizon increases. Comparing with the sharpness in Figure 14 it is seen that this happens when the average IQR is higher than approximately half of the installed capacity. It is natural that in this case the variation in IQR cannot continue to grow and it means that the uncertainty is often high for these horizons.

Comparing the Elsam and E2 setups w.r.t. resolution it is seen that the E2 setup has higher resolution than the Elsam setup. However, this is expected since the Elsam setup produce markedly sharper forecasts than the E2 setup, cf. Figure 14, and in a strict sense the comparison is not appropriate.

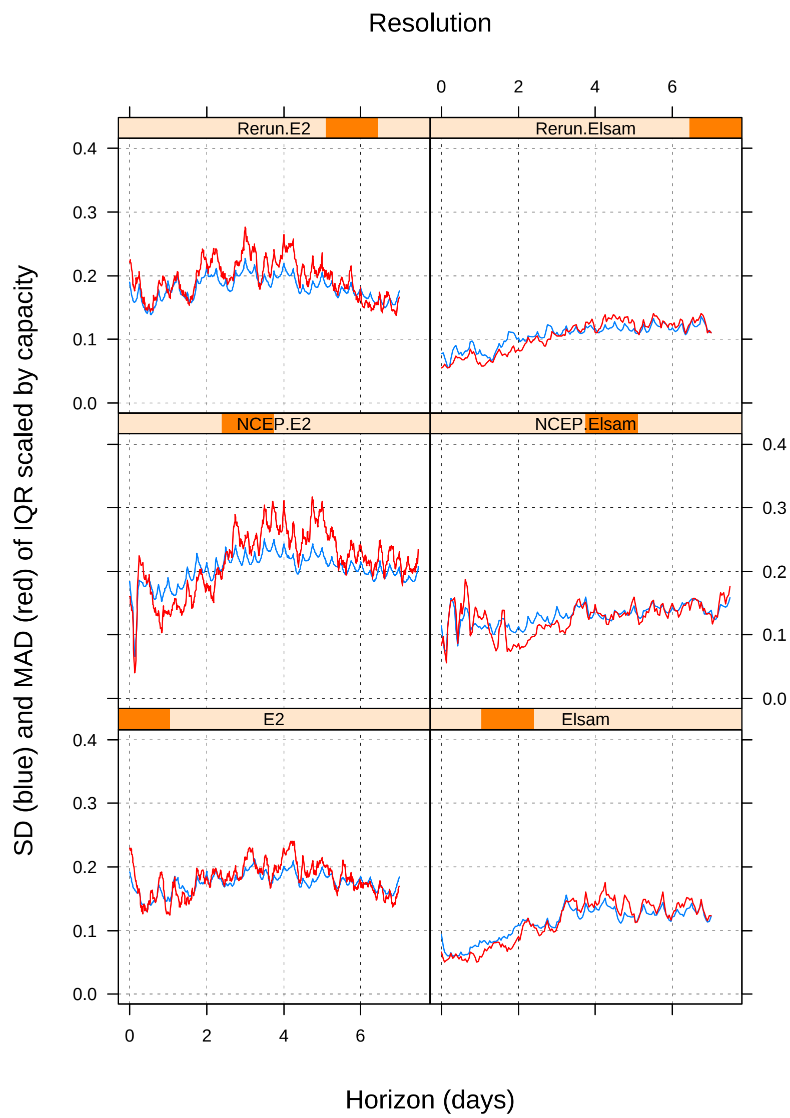


Figure 15: Resolution compared to installed capacity.

4.3.4 Spread / skill relationship

The spread / skill relationship refers to the relation between a point forecast (i.e. a single value) and the uncertainty indicated by the ensemble forecasting system. Here we will let the IQR mentioned in the previous sections represent the uncertainty indicated by the forecasting system. Several possibilities exists w.r.t. the point forecast; here the median forecast is selected as the reference. For details see [36].

Figure 16 shows the resulting spread / skill plots grouped by horizon. A clear relation between IQR and the magnitude of the absolute forecast error is evident from the plots. This clearly confirms that the quantile forecast systems indeed contain relevant information regarding the uncertainty. Note also that when disregarding the original runs and the horizons below 24 hours the plots are similar for a board range of horizons. This further confirms that the information regarding the uncertainty is contained in the values of IQR.

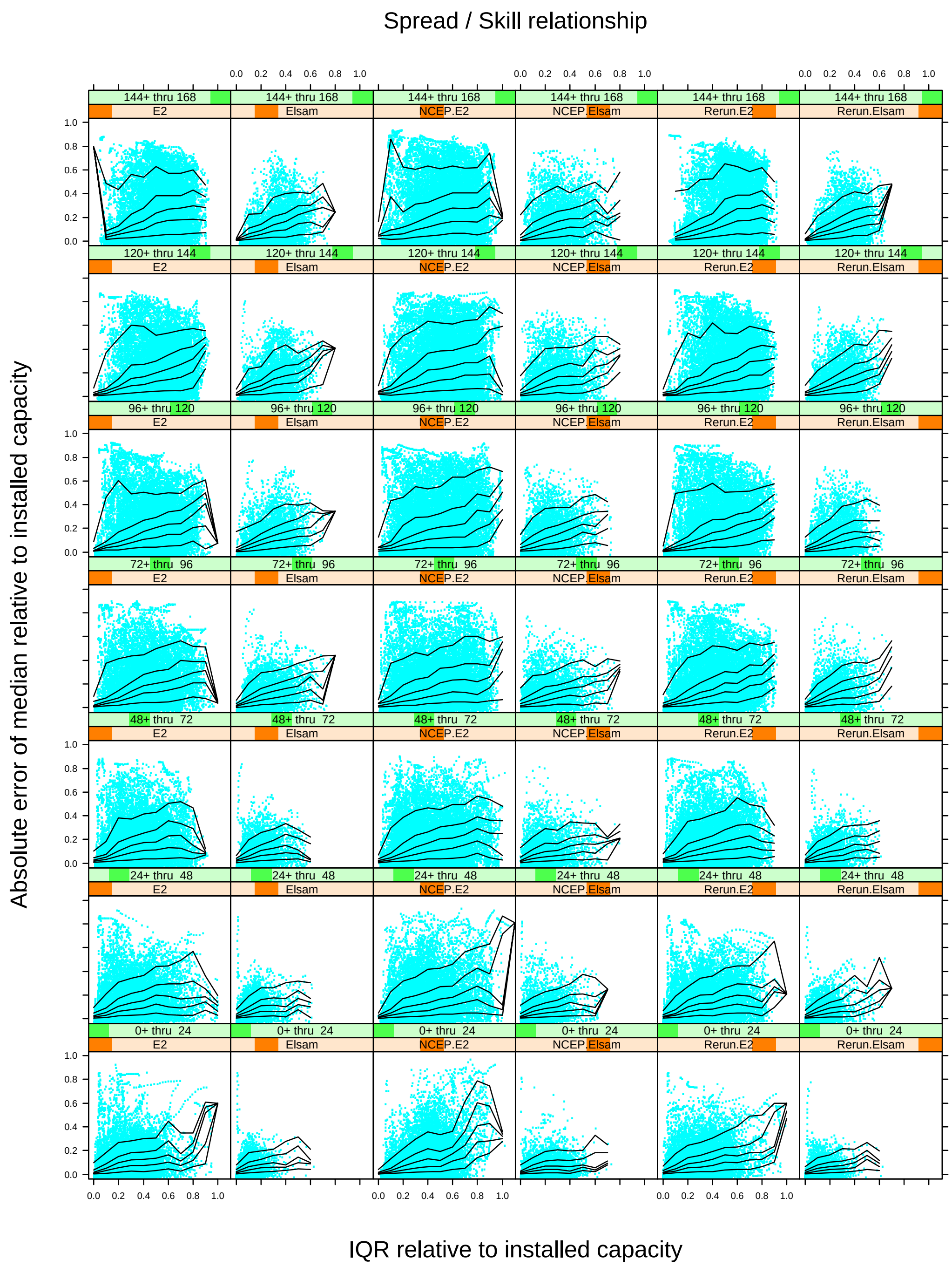


Figure 16: Spread / skill relationship when the 50% quantile forecast is used as the point forecast. The black lines indicate the lines for which 10%, 30%, 50%, 70%, and 90% of the errors are below, when grouping by the value on the 1st axis rounded to one decimal point.

4.4 The benefits of turbulence

This chapter is an excerpt of a larger report [37]. Wind speed measurements for three heights at Risø (44m, 76m, and 125m) and for six heights at Høvsøre (10m, 40m, 60m, 80m, 100m, and 116m) were collected for the first seven months of 2005. Wind speed predictions for these heights and locations for this time period were made using Prediktor. These predictions were based on both the standard forecasts of the HIRLAM model and five different wind speed data types that include parameterizations of the Turbulent Kinetic Energy.

See chapter 2.6 for an explanation. Predictions were also made using the five NWP points nearest to the meteorological mast at Høvsøre.

The effect on wind speed prediction errors based on the six different HIRLAM versions was quantified using six different measures of error. The effect on wind speed prediction errors based on two of the six HIRLAM versions using the five NWP points nearest to the meteorological mast at Høvsøre was also quantified.

For a summary of results, the Mean Absolute Error was chosen as the most significant measure of error. The average percent changes in MAE from the standard HIRLAM predictions when using predictions based on one of the other HIRLAM versions are shown in Table 1. The average percent change in MAE is the mean of the percent changes in MAE at each forecast length from 0 to 48 hours. Each number in the table indicates by what magnitude the MAE changes when a certain version of HIRLAM is used for wind speed predictions instead of the standard version. Positive numbers indicate increases of the MAE, while negative numbers indicate decreases of the MAE. The best performing HIRLAM version (greatest negative percent change) is highlighted in yellow for each case.

	W10	**W10g**	**W10gb**	**W70**	**W70e**
Risø 44m	-7,4	7,8	7,7	-4,3	-4,9
Risø 76m	-6,2	11,2	14,0	-14,0	-13,8
Risø 125m	-7,0	12,8	8,6	-21,3	-21,4
Høvsøre 10m	3,0	6,8	20,2	3,9	4,4
Høvsøre 40m	-3,7	3,5	13,0	-5,9	-5,6
Høvsøre 60m	-4,5	4,2	8,1	-9,6	-9,2
Høvsøre 80m	-4,8	4,7	3,2	-12,0	-11,7
Høvsøre 100m	-4,7	5,2	0,8	-13,1	-12,9
Høvsøre 116m	-4,3	6,2	-1,6	-13,4	-13,2
Høvsøre (offshore) 10m	-4,9	-6,6	29,7	3,0	3,1
Høvsøre (offshore) 40m	-11,1	-9,4	21,0	-6,4	-6,4
Høvsøre (offshore) 60m	-10,0	-5,5	16,4	-9,2	-9,2
Høvsøre (offshore) 80m	-8,3	-1,7	9,9	-11,0	-10,8
Høvsøre (offshore) 100m	-7,0	0,8	6,2	-11,7	-11,5
Høvsøre (offshore) 116m	-5,3	3,6	1,8	-11,6	-11,5

Table 1: Average percent change in wind speed Mean Absolute Error from standard HIRLAM for the five new HIRLAM wind speed parameterisations.

In most cases, W70 produces the smallest wind speed errors, improving upon the standard HIRLAM version in every case except Høvsøre 10m. For both locations (Risø and Høvsøre), the improvement resulting from using W70 instead of the standard HIRLAM version increases with hub height. W10 also improves upon standard HIRLAM in every case except Høvsøre 10m, but usually not by as much as W70. For the lowest measurement heights, W10 usually produces the smallest wind speed errors. In almost every case, W70e produces slightly greater wind speed errors than W70.

W10g and W10gb, fulfilling our expectations, usually produce greater wind speed errors than the standard HIRLAM version, especially at lower hub heights. W10g performs worse than standard HIRLAM in every case except for the four lowest hub heights at Høvsøre (10m, 40m, 60m, and 80m) when only wind directions coming from offshore (210^o to 330^o) are considered. W10gb performs worse than the standard HIRLAM version in every case except Høvsøre 116m.

Høvsøre is located very close to the western coast of Jutland. When only wind speeds from the offshore sectors are considered, instead of all wind directions, W10 and W10g perform better, W10gb perform worse, and W70 and W70e perform about the same. However, data for wind speeds coming from offshore was available for just 88 days, and therefore any conclusions should be treated with caution.

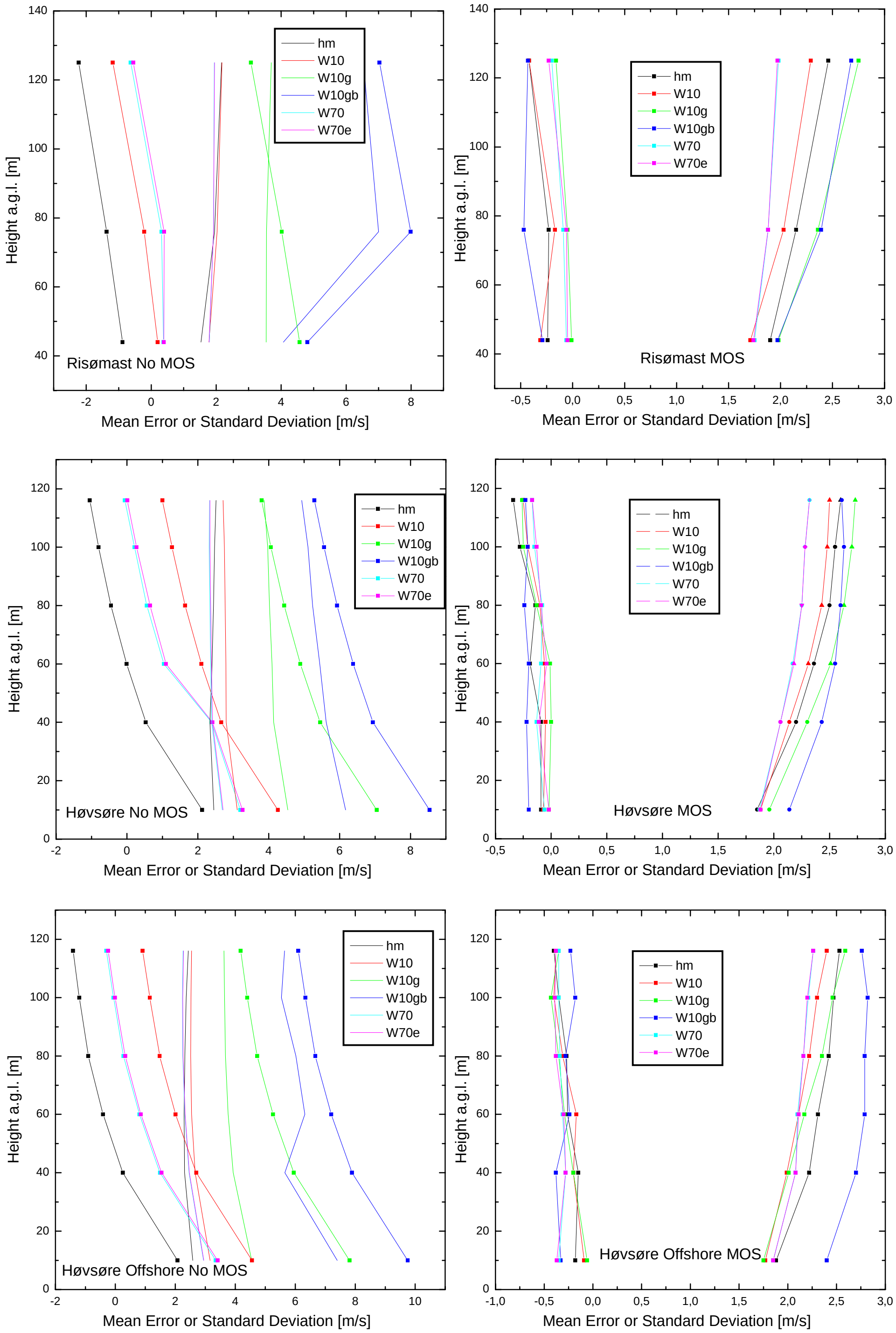


Figure 17: The dependence on height of the mean errors and standard deviations with (left) and without MOS (right) for the three cases of Risø, Høvsøre and Høvsøre offshore. In the left coloumn, the lines with symbols show the mean error, or bias.

Figure 17 summarizes the mean error and standard deviation of the plots on the previous pages, in dependence of the height along the mast. The effect of the MOS can be seen very clearly: the bias is nearly totally removed for all heights. Note that the MOS scheme used in Prediktor does not remove the bias completely, but aims at finding the lowest possible MAE. Here also, the large bias for the gust schemes (W10g and W10gb, in blue and green) is obvious. It is not surprising that these two schemes even after the MOS step are the worst. The large curvature in the Høvsøre NoMOS biases seems to indicate that there are issues with the height adjustment, possibly with the roughness rose used for the surroundings of the mast. Regarding the standard deviation, the 70-m level wind speeds are the clear winner for all but the lowest available levels, even nearly beating the dedicated 10-m winds at Høvsøre for that height. It is not clear why the standard deviations rise with height for all the MOS cases, though this probably has to do with the higher average wind speeds at higher height.

The most significant findings are summarized as follows:

- All wind speed predictions show an increasingly negative bias as the height increases.
- For both locations, the W70 HIRLAM version usually produces the most accurate wind speed predictions. However, at lower heights, the W10 version sometimes produces the most accurate predictions.
- The improvement in the accuracy of wind speed predictions when using the W70 HIRLAM version increases with increasing height. When using the W10 HIRLAM version, the improvement in accuracy does not change significantly with height.
- Both the W70 and W10 HIRLAM versions outperform the standard HIRLAM version in every case except one: Høvsøre 10m.
- W70e produces wind speed predictions only slightly less accurate than W70 in most cases. This indicates that the inclusion of TKE in the current form does not have the desired positive effect.
- The wind gust predictions W10g and W10gb produce wind speeds that are, expectedly, less accurate than the standard HIRLAM predictions.
- For Risø, the accuracy of the wind speed predictions does not depend on the time of year. For Høvsøre, the wind speed predictions are less accurate during the months of January and February when the wind speeds are very high.
- The difference in accuracy between various versions of HIRLAM does not depend on the forecast length.
- There is no significant difference in the improvement in the accuracy of wind speed predictions when only winds from offshore directions at Høvsøre are considered.
- The choice of an NWP point located further from the meteorological mast at Høvsøre in the direction of offshore winds improved the accuracy of the wind speed predictions slightly.

4.5 Nesting HIRLAM in the ECMWF ensembles

One of the main reasons for the improvements in forecast skill is undoubtedly the availability of increasingly more powerful computers that allow the NWP models to be run at higher and higher spatial resolution and with more comprehensive parameterization schemes. The increased computer power also allows a quantification of the forecast uncertainty through the use of ensemble forecasts, and there is an ongoing debate as to whether extra computer time is best spent on increasing the model resolution of a deterministic forecast system or increasing the ensemble size of a probabilistic forecast system [38].

This chapter investigates a simple ensemble approach based on dynamical downscaling of all the ensemble members from an ECMWF-EPS simulation. It shortly summarizes the findings of a larger report [39]. It tries to address the question on the benefit to gain in 10m wind forecasts when integrating all members of a large scale EPS using a high resolution NWP model. For this purpose the DMI-HIRLAM model (version 2003) was nested daily into the individual ECMWF-EPS members over Europe during winter 2002/03, and a verification against observations of 10m wind speed was performed for both the DMI-HIRLAM ensemble and the ECMWF-EPS. For the details of the nesting, see chapter 2.5. The verification was carried out against station observations from about 50 stations in the inner HIRLAM domain (see Figure 5).

Given that some sort of post-processing including a bias removal of the ensemble forecasts is required, the most important aspect of the forecast is the ability to predict variations in time. That is, we would like to see a high correlation in time between forecasts and observations. Figure 18 shows the average correlation in time between the ensemble mean and the verifying observations, as well as the average correlation stratified according to ensemble spread as described above. The correlations vary remarkably little as a function of lead time. The HIRLAM forecasts for which the ensemble spread falls in the 50% category, are very poor for 0 lead time, but this result suffers from sampling problems as the spread almost always falls in the low spread category for 0 lead time. For the HIRLAM forecasts with low spread the ensemble mean generally correlates better with observations than is the case for forecasts with larger spread, and for the ECMWF forecasts the lowest spread forecasts are also better than the ensemble forecasts with larger spread for the longer lead times, indicating a relation between spread and skill.

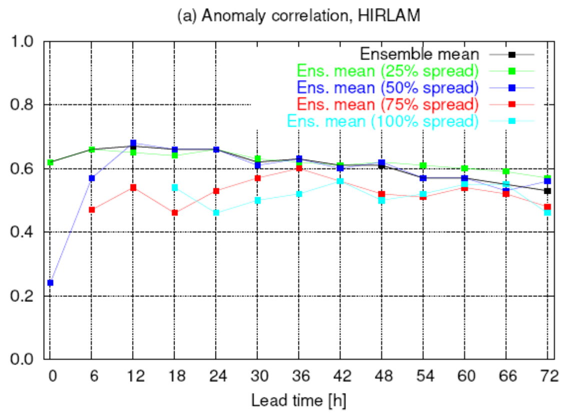


Figure 18: Average correlation in time between ensemble mean and observations for all forecasts (black curve) and stratified according to ensemble spread (coloured curves) for (a) HIRLAM and (b) ECMWF ensemble forecasts.

The conclusions of the report were as follows: The use of a simple dynamical downscaling of all members from the global ECMWF ensemble prediction system using the higher resolution of the HIRLAM model does not improve the 10m wind speed forecast skill scores over those of the global ECMWF ensemble prediction system. Possibly, the benefit of a simple dynamical downscaling running a limited-area model is only seen in complex terrain and in extreme weather.

The ECMWF ensemble prediction system is designed to give realistic spread in the medium-range. The results in the present report document that the ensemble spread in the short-range is too small, and downscaling with HIRLAM is not the solution to this problem for the investigated period.

There are several other possible ensemble approaches that may lead to more realistic ensemble spread. The simplest approach is probably to apply some sort of statistical post-processing, such as quantile estimation [40] or ensemble inflation [41], whereby the post-processed ensembles can be made more reliable. In the present setup perturbations are only applied to the initial condition in the ECMWF model; perturbations could also be applied to the initial condition(s) in HIRLAM. However, a better approach might be to include an ensemble data assimilation system to the HIRLAM model, e.g. based on ensemble Kalman filtering. A likely explanation for the larger spread in the ECMWF model is the inclusion of stochastic physics; stochastic physics could also be implemented in HIRLAM.

Thus, there are plenty of opportunities to develop a short-range ensemble prediction system, which would not only be of interest to the wind power industry, but could provide a general tool for estimating short-range weather forecast uncertainty.

4.6 Using multiple models

One of the points we investigated in this project is the benefit of multi-model ensembles, where the NWP input to the short-term forecasting model comes from different meteorological institutes. To this aim, we bought one year of data for Denmark from the DWD (Deutscher Wetterdienst, German Meteorological Service), to be used concurrently with the DMI-HIRLAM data we already get. We chose DWD, since the two NWP models even are initialized by different global models: DMI-HIRLAM is driven by the European ECMWF model, while the DWD runs its own global model. Both NWP inputs were then fed into Prediktor (see chapter 3.1), and the MOS step was used on every model output separately. The thin lines in Figure 19 refer to the results without MOS, while the thick lines are the results with MOS. These results were then simply averaged for every forecast time and horizon to yield the red line.

As to our knowledge, this attempt at using two models concurrently has been done only once before: Giebel *et al* [7] and Waldl and Giebel [8,9] investigated the relative merits of the Danish HIRLAM model, the older Deutschlandmodell of the DWD and a combination of both for a wind farm in Germany. There, the Root Mean Square Error (RMSE) of the Deutschlandmodell was slightly better than the one of the Danish model, while a simple arithmetic mean of both models yields an even lower RMSE. Whether this holds true for the latest incarnations of the models for a number of test cases in Denmark, is shown shortly here. This chapter is actually a shorter version of a paper we presented at the EWEC conference in London, November 2004 [42].

For the six wind farms presented in this chapter (Fjaldene, Hagesholm, Klim, Middelgrunden, Syltholm and Tunø Knob), we had between 5-min data and 15-min data. No interpolation or averaging has been performed.

The results for the six wind farms are shown in Figure 19. There are various lessons to be learned from these results:

As is usual, the error increases linearly (plus scatter) with the forecast horizon. In most cases, the error increases for the single NWP inputs from about 6-12 % for the shortest horizons to about 9-15 % of the installed capacity. Note that Zephyr/Prediktor does not take the measured time series into account, therefore the error at zero hours forecast is not zero.

In most cases, HIRLAM is marginally better – however, the forecasts are surprisingly similar in quality. In the previous study, the German model performed marginally better for wind farms in Germany. This suggests that the local model usually performs best for a given area.

The MOS step increases the performance in nearly all cases, in some cases even dramatically (*eg* Syltholm).

The simple combination of both models works even better than any of the models themselves. This could probably be improved upon with a weighting of the models according to their accuracy. This weighting could be horizon dependent, or dependent on the synoptic scale weather situation.

In earlier studies, there was a clear minimum of error at about 6-9 hours forecast length. This seems to have disappeared. Seemingly, the data assimilation of the NWP models has improved a lot, so that the models are no longer as drastically “forced” into an unnatural state by the data acquisition step.

One additional advantage of a combination of two NWP models is that in the (according to our experience quite rare) event that one NWP model should not arrive on time for whatever reason, the other one would be there. Since the quality of the forecasts is heavily dependent on the horizon length, just substituting a 18-hour forecast with a new 6-hour forecast increases the accuracy by a fair margin.

Note that the combination of DMI+DWD did use both MOS’ed results, and just averaged them. It might have been even better to combine the two models as input before calculating the MOS parameters, although the effect due to the linear MOS system is quite probably small.

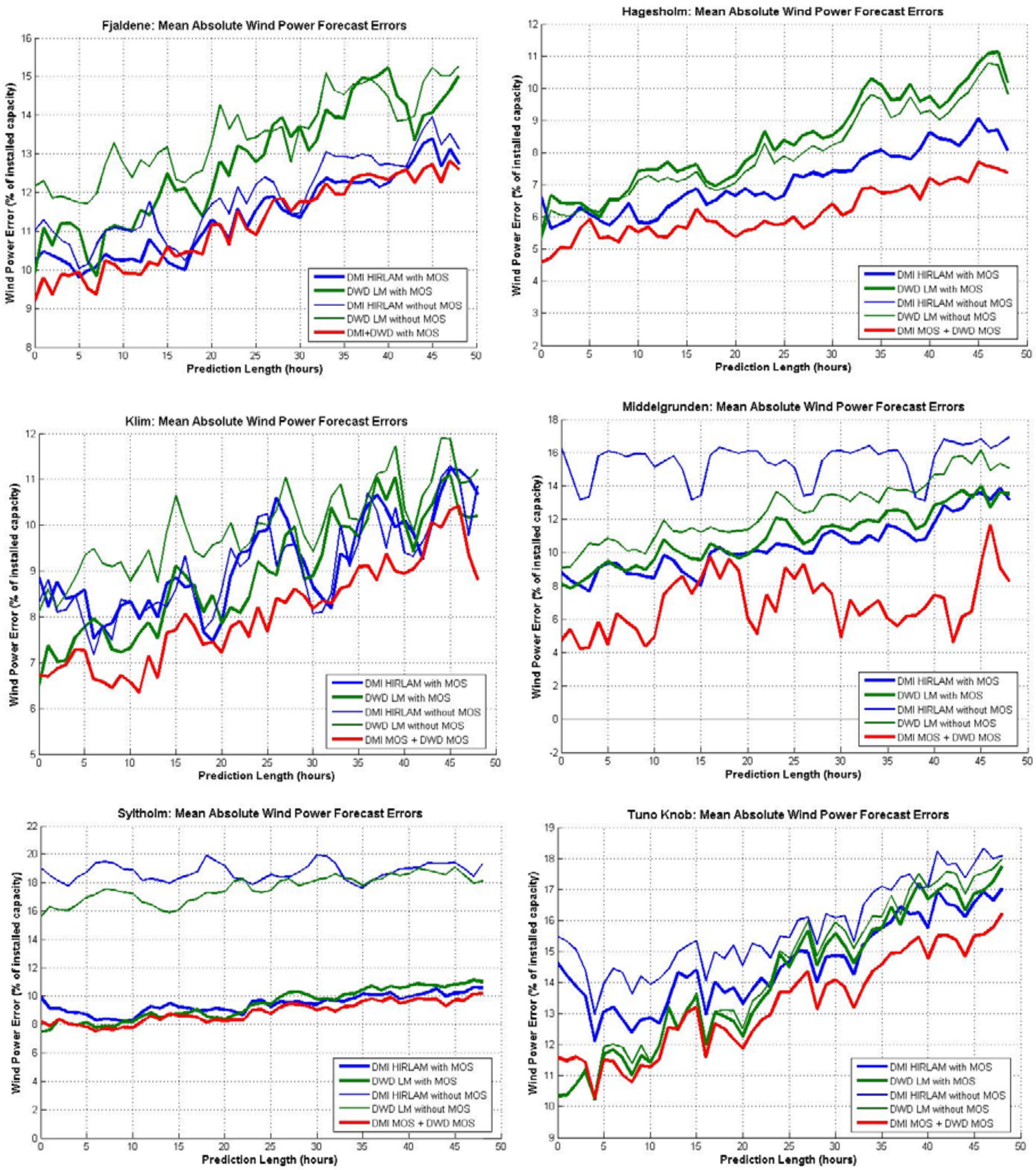


Figure 19: The Mean Absolute Errors for five different configurations for six wind farms in Denmark. The results are scaled as percent of installed capacity. Note that the scaling varies from plot to plot.

5 The demo application

After a few months into the project, we realised that the analysis of the possibilities within the ensembles proper would require clearly more time than we had allocated for that task. Therefore, it was decided to not include a full implementation in the online version of Zephyr/WPPT installed at the utilities sites, but to try out the ideas developed in the data analysis in the form of a demo application. The demo was installed during the summer of 2004, and has run continuously since (except for a few days when forecasts were missing either due to problems getting the ECMWF forecasts or due to disk space (quota) problems on IMMs webserver). Note that the scientific performance of the demo application has been reported already in section 4.3.

5.1 The system

The demo was based on the 51 ensemble members available to us from ECMWF via DMI. The ECMWF ensemble simulations where started daily at 12 UTC. The results from these simulations were available at around midnight local time. A data extraction process was then launched to transform the selected variables onto the desired grid and to copy the data into a temporary archive, from where it was transferred to the server at DMI. From the server at DMI, Risø retrieved a part covering Europe every day at around 05:45 local time. From that, two forecasts were cut out on the server at Risø: one covering all of western Denmark (the Eltra area, ie Jutland and Fynen), and another one for the offshore grid point nearest the Nysted offshore wind farm owned by E2, and feeding into the Elkraft System electrical grid. These forecasts were then transferred to a server at IMM, where various Perl, Jython and S-Plus scripts were run to finally put the 51 members, the raw quantiles and the transformed quantiles onto a web page, hosted on IMMs web server. The output graphics were adjusted a few times during the run of the model, after input from the utilities.

The same set-up has only been used offline for the NCEP ensembles, since the quality of the NCEP ensembles was deemed not to be adequate. However, in the last month of the project, the NCEP ensembles have been operationalised as an additional source of forecasts (the GUI is still missing though).

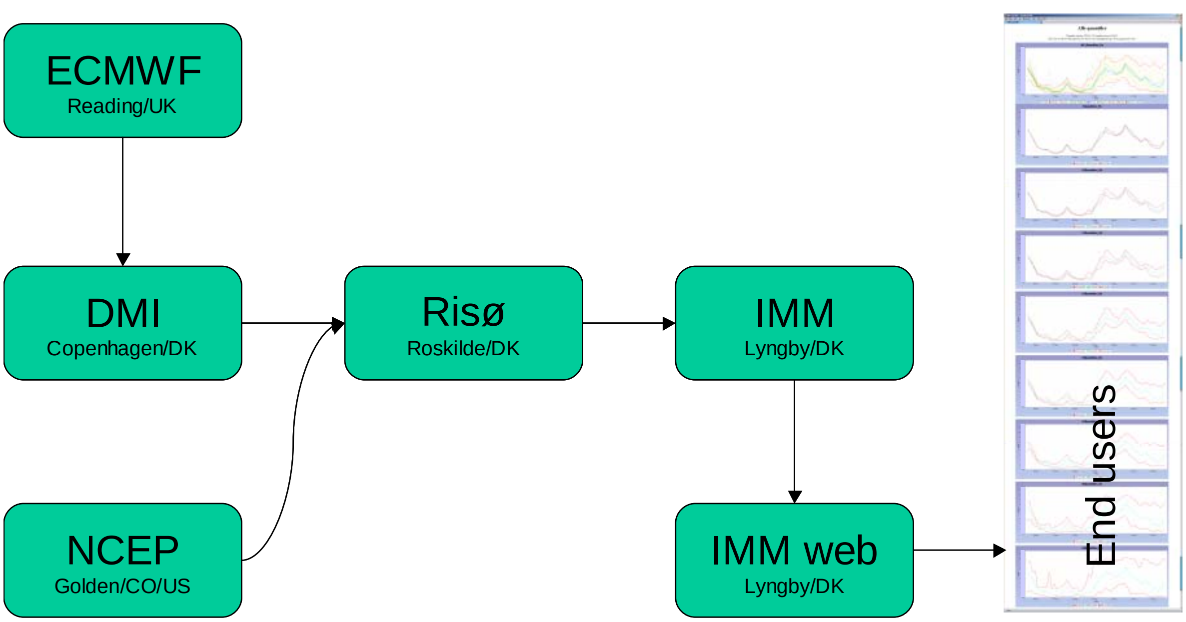


Figure 20: The data flow of the demo application. Note that the string with the data from NCEP has not been installed yet on an operational basis. Note also that in the continuation of the demo application, the IMM servers would not be used, only web servers at Risø.

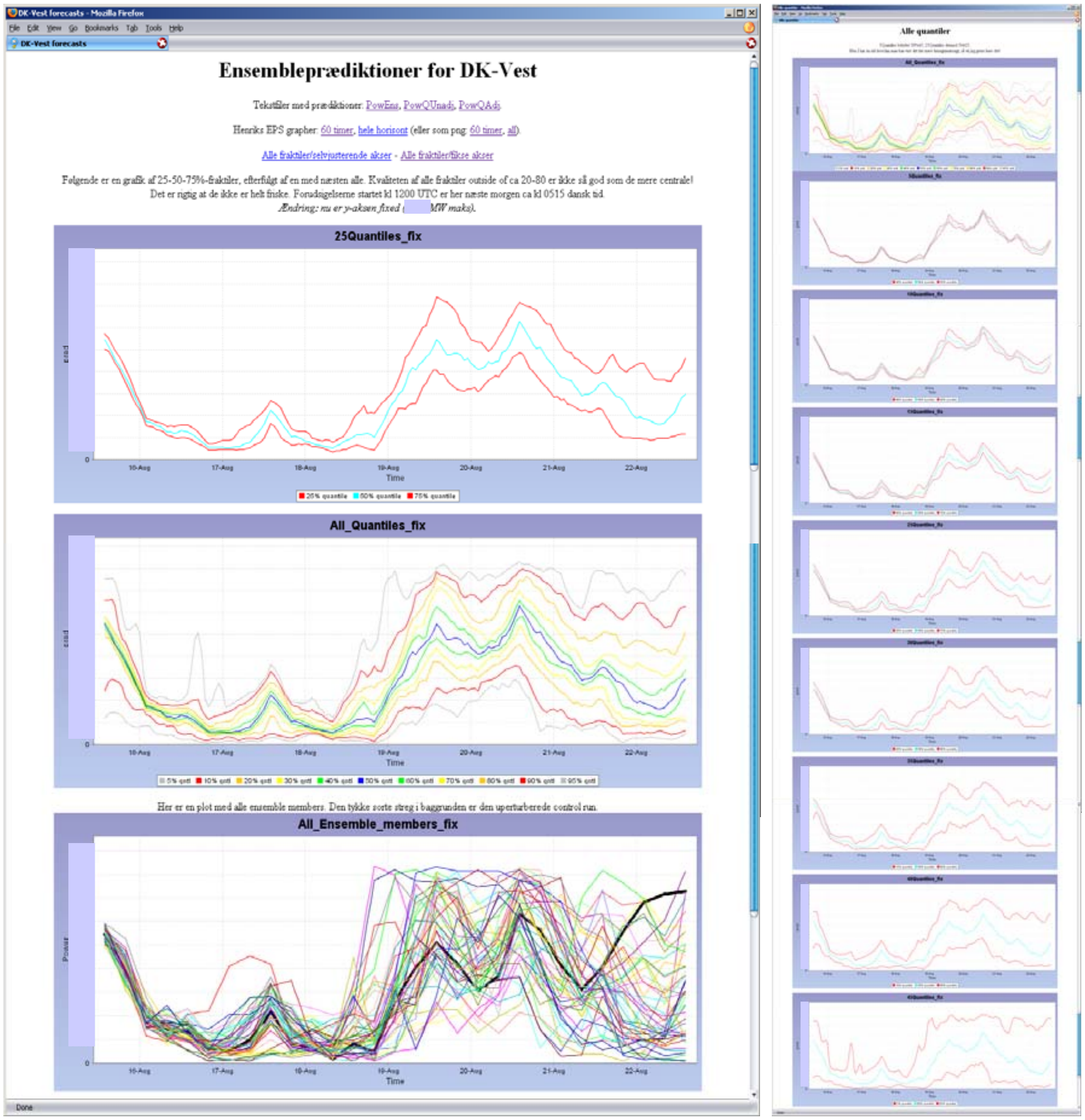


Figure 21: Two different views of the demo application. Left: the main page, with the 25, 50 and 75% quantile on top, all calculated quantiles in the middle, and all single ensemble runs converted to power at the bottom. Right: a cascading plot of all quantiles at the top, and the 50% quantiles plus continuously widening sets of symmetric quantiles (45 and 55% for the second plot, 40 and 60% for the third, and so on). The results are for western Denmark, but the actual magnitude has been marked unreadable on request of Elsam.

The plots shown were twofold, as can be seen in Figure 21: one "opening page" with a few links, amongst other things to the page seen on the right, and to the forecasts as text files, to use directly in some decision support tool for trading or maintenance planning. The opening page also showed three views of the ensembles: a so-called spaghetti plot (the lowest one), which contained the deterministic forecast (the unperturbed forecast) as a thick line and the 50 perturbed ensemble members in different colours, already converted to power. Additionally, two plots are shown with the adjusted quantiles, one with nearly all quantiles between 5% and 95%, and one just with the 25%, 50% and 75% quantiles. Keep in mind that the outermost quantiles are less accurate due to the cut-off effects than the more central quantiles. On the right is a row of plots with growing quantiles around the central quantile.

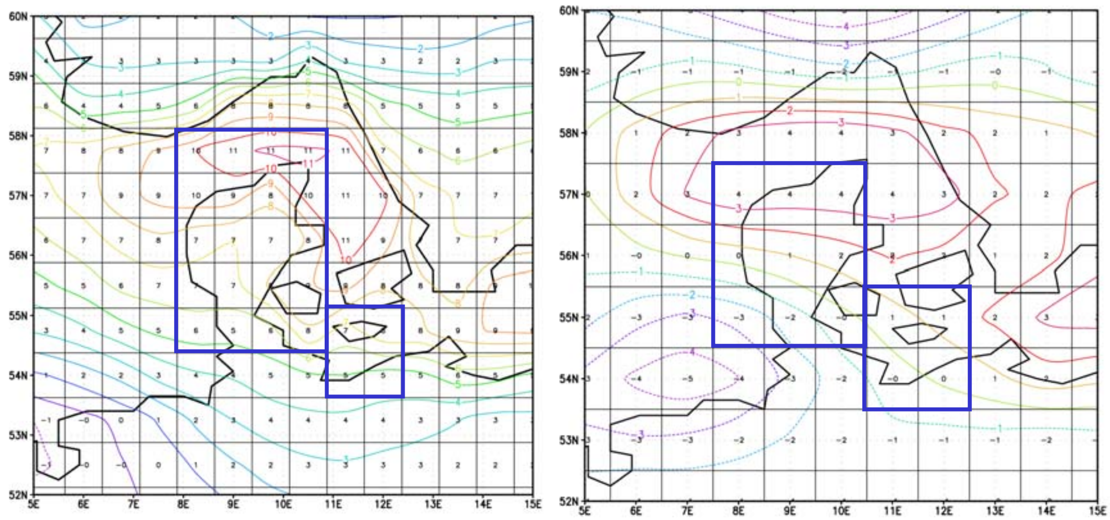


Figure 22: The grid points used for Jutland/Fyn and Nysted from ECMWF (left) and NCEP (right).

5.2 Results from the operational use - Elsam

Elsam already uses the Zephyr/WPPT software for their scheduling and trading. The longer forecasts available through the demo system also enabled new uses for wind power forecasts.

5.2.1 Control room functions

In Elsams central control room the production of electricity and district heat is optimised. A broad distribution of production facilities ranging from small wind farms over minor decentralised CHP plants to large central power stations are in the centre of the optimisation process. In the control room, the short-term optimisation is performed with a horizon of up to one week.

The optimisation process consists of a planning phase and a operational phase. In both these phases, wind power forecasts partake of the necessary data basis.

Planning phase

The task of the planning phase is to dispatch the operational production facilities for the next day in order to obtain maximum income. The task, in short, is to sell the possible electricity production at the maximum price. On the basis of the obtained sales volume it is decided which production facilities should be in service, and how to distribute the load between them. This happens mostly with regard to the production unit cost of the individual power plants.

Operational phase

In the operational phase, which always is the current day, the power plants that are to run are already dispatched, and the total production plan is fixed. The task is now to stick to this plan.

Deviations from the plan, f.x. due to breakdown of production or transmission facilities or when the wind power production of Elsam’s wind turbines deviates from the predictions, have to be compensated by up- or down-regulation of the large central CHP plants. By correcting the wind power forecasts continuously, it is possible to distribute the load optimally. In this phase, the operational 36-hour wind power short-term prediction is used to update the forecasts for Elsam’s own wind farms. The updated values then go into Elsam’s load distribution system.

5.2.2 Use of wind power predictions at Elsam

The primary focus is on the next day, since the trading of electricity on the spot market (in NordPool language, that is the market for the next full day, not to be confused with the balance market, where trading happens hourly) happens once daily. That also defines the most important horizon for short-term forecasts: from 1200 hours to midnight and the day after, that is 12-36 hours.

For the start and stop decision of large central power stations, 36 hours is not long enough. This, amongst other things, is because the cost of starting a power plant is quite significant. Often enough, it is not worth it to start a power station for just one day of operation. Therefore, predictions of wind power and the related uncertainty some days ahead are essential. In the following, a few situations are described where 7-days forecasts are an important tool for decision support.

Electricity trade over the weekend

The German electricity exchange EEX is closed over the weekend. That means that Elsam has to register its supply and demand on the EEX already on Friday for Saturday through Monday. Bilateral trade also is performed on Friday. A large part of the planning for Saturday through Monday therefore happens on Friday morning. The largest part of the wind power production is traded on the NordPool exchange, which is open 7 days a week. Since wind power can represent a sizeable part of all electricity traded on the market, and thereby influence the market price, it is of great economic importance to have a good estimate of the coming wind power.

Often there will be a possibility to sell a certain amount of Elsams production on the EEX or bilaterally for the entire weekend. Alternatively, that amount of power can be sold on the NordPool electricity exchange. The wind power forecasts for Jutland/Fynen can therefore indicate whether it is better to sell already on Friday on the EEX or bilaterally, or whether one should wait and sell on the NordPool.

The same considerations are also valid for periods with consecutive holidays, like Easter, Pentecost or Christmas.

Power station failure

When a power station fails, Elsams central control room is contacted by the local control room. It is now the team responsible for the optimisation of Elsams production, which in concert with the technicians of the power station decide when the failure should be fixed. In some cases it is possible to continue with the operation of the power station for a few days, until it fits best into the overall dispatch to take the power station down for repair. If it runs longer than that, the damage can worsen. If the repair as such takes 1 day, then including the cool-off period the power station is out for 36 hours.

Therefore, a decision has to be taken as to when the power station best can be spared. The market price follows supply and demand. Everything else being equal, a high wind power prediction has the promise of lower market prices. Since there often is need for extra personnel like external specialists for the repair, the decision about the repair time has to be taken several days in advance. In such situations it is of major importance that the wind power and uncertainty predictions 4-5 days ahead are accurate.

Fuel demand predictions

Once a week, Elsams load distribution team establishes fuel demand predictions for the coming 10 days. The predictions are used in various departments of the company.

The predictions show the expected daily demand of coal for the individual power stations. Coal firing of power stations leads to different types of by-products, for example fly ash. If the quality of the ash is good, then it can be sold to the cement industry. The quality depends on the coal type and by which power station it was produced. The fuel demand predictions are therefore used to determine the right mix of coal types.

There is a clear connection between wind power production and coal demand. The electricity demand is more or less known, so the more wind power there is accessible, the less coal is needed. Therefore, the wind power predictions can give an indication of the share of electricity to be produced by coal.

5.2.3 General user experiences

One should be clear about that this is a diverse user group. Therefore, not all employees were using the predictions to the same extent. A survey showed that some used it quite frequently, while others more on an ad-hoc basis. There also were different attitudes to the importance of the prediction.

There was general satisfaction with the user interface, but it was foremost the 50% quantile people were using, usually disregarding the other quantiles. All users agreed that it should be easy to pull out the data into a spreadsheet in a simple way.

In comparison to the 36-hour predictions, the ensemble predictions were used not as systematically, but rather in special situations like described above, or to give an overview of the expected market situation the next few days. The ensemble predictions were generally used on Fridays to get a market assessment for the weekend and Monday.

A systematic estimation of the economic benefits of the predictions has not been performed. The users themselves do not immediately have a possibility to judge the quality of the predictions. That would need a statistical analysis of a large body of material. There is clearly a wish to get such an analysis. Furthermore, some people would have liked to be able to switch between multiple NWP providers.

5.2.4 Future plans

Elsam is currently developing a new production planning system. The system consists of various modules, each covering its own planning horizon. There are three modules:

- Day ahead planning (short term)
- Weekly planning (medium term)
- Yearly planning (long term)

The 7-day wind predictions could therefore be especially interesting for the weekly planning module. The aim of the weekly planning is, amongst other things, to calculate and optimize the unit commitment. Here the optimal schedule is developed for which power stations should be running during which period in the next 4-5 days. In addition to the power plants specific running cost and the predictions for the prices on the electricity markets, the wind power predictions are an important parameter. If the wind power production is large, then it will drive production from the large central coal fired power stations out of the market. The market share of the station will thus become so small that not even the fixed production cost (idling cost) are covered. The station has therefore to be stopped.

Furthermore, it could be interesting with some calculation of the consequences of certain parts of the wind power distribution. In this case there should be a use for the quantiles. Think for example the situation where it would be economically viable to run with one coal fired station if the wind power reaches 500 MW, but where the power station should be shut off if the wind power production reaches 750 MW.

5.3 Results from the operational use – E2

The operational use of the PSO-Ensemble demo application at ENERGI E2 A/S was initiated in the middle of November 2004 and has been used on a daily basis for production-planning since then.

From the middle of January 2005 data has been collected and stored in a database.

At present and throughout the whole period the demo has been used each day for planning the number of power-plants to be used for the coming 3-5 day period, expected fuel-plan for the coming 1-2 weeks and expected amounts of power to be traded the following days. The possibility for planning the required number of running power-plants for the comings days (taking also other market-parameters into account) has given ENERGI E2 A/S a valid basis for determining an economically sound start-up strategy. The use of the demo provides ENERGI E2 A/S with an additional tool in order to plan supplies of different kinds of fuels, especially biomass-fuels. The storage-capability of biomass-fuels is typically smaller than that of coal and oil (gas being supplied directly when needed). This means that the expected consumption for the coming 1-2 weeks should be met by an equivalent delivery. In cooperation with our fuel-purchase department this planning is discussed on a weekly basis. Finally the demo gives input to a discussion with our trade-department regarding the expected trade for the coming days.

The demo has run very stable, only few days with problems accessing the system. The lay-out is logical and can be easily read and understood. ENERGI E2 A/S expects to continue the use of the demo in our daily planning.

5.4 The future of the demo

The utilities were happy enough with the demo system to leave it running also after the end of the project, even though this will cost them a small service fee. This is in part due to the fact that DMI only could provide the ECMWF ensembles for scientific purposes within the framework of this project.

6 Summary / conclusions

The use of the ensemble prediction methodology in numerical weather prediction is well established for global models like the ECMWF EPS. It provides useful information on the synoptic scale uncertainty in the weather prediction. An enhancement by limited area models like HIRLAM necessitates more than simple downscaling, unless extreme events are the only subject of interest.

A major result is a method to transform the ensemble percentiles to percentiles of the observed distribution. This is necessary since the ensemble spread itself is not probabilistically correct. Due to the insufficient spread in the investigated ensembles especially for short horizons, not all confidence intervals can be transformed. However, in all cases we have seen, the 25% and the 75% quantile were available.

This method is able to transform meteorological ensembles of wind speed and direction to ensembles of power production, and 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 Centers for Environmental Prediction (NCEP) in the U.S. were 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. However, this is not possible without further theoretical and practical developments. It is 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. Based on these investigations and developments a demo-application were developed and implemented, giving the utilities involved in the project access to probabilistic forecasts for 1 week ahead.

Note that by forecasting only quantiles the information about autocorrelation in the future power production is not available to the user as it would have been had the power-ensembles been reliable in a probabilistic sense. Fore instance with the aim of obtaining a realistic uncertainty of the fuel demand forecasts mentioned in section 5.2 the information regarding correlation is important. However, further developments are needed in order to supply such information.

From the analysis of the results of the demo-application it is clear that regular re-calibration is extremely important. Given regular re-calibration it is shown that approximately reliable quantile forecasts can be produced. This also confirms that the use of adaptive methods is important for wind power forecasting in general. For a final software implementation the re-calibration should be performed automatically.

The sharpness and resolution of the forecasts was analyzed. Comparing the results obtained using the ECMWF and the NCEP ensemble forecasts it is seen that approximately the same sharpness is obtained. However, especially for horizons between 36 and 48 hour the quantile forecasts based on the ECMWF ensembles are sharper than the quantile forecasts based on the NCEP ensembles. Considering the calculation time of the models these horizons are very relevant from a market point of view. Note, however that the initialization times of the ECMWF and the NCEP models differ (12: 00 and 00: 00 (UTC), respectively) and therefore the comparisons outlined are mostly relevant from a scientific point of view. Also, w.r.t. resolution the difference in results when using the ECMWF or the NCEP ensembles is small. There is however a tendency for the quantile forecasts based on the NCEP ensembles to have a higher resolution than those based on the ECMWF ensembles.

When considering the difference between the 10% and 90% quantile forecasts it is notable that for the Elsam setup, which covers a large geographical area, the average difference between the quantiles is not more than 60% of the installed capacity for any horizon. For the horizons most relevant for bidding on NordPool the corresponding number is 30%. This confirms that on the large scale it is possible to forecast the wind power production to a large extent.

The spread / skill relation is also investigated. It is shown that there is a good relation between the error of the 50% quantile (median) forecast taken as a point forecast and the Inter Quartile Range (difference between the

25% and 75% quantile forecasts). Furthermore, this relation do not seem to be affected by the horizon considered. This indeed confirms that the actual uncertainty is reflected in the quantile forecasts.

The demo was well accepted by the utilities, and led to new uses of wind power predictions, such as fuel demand forecasts, trading of power over the weekend, or unit commitment optimisation. The demo application continues to run and to be used also after the project.

Additionally, it could be shown that combining forecasts from two different meteorological institutes (in our case, DMI and DWD) is beneficial for the quality of the forecasts, even if they are just averaged. Further research into adaptive methods for choosing the optimal model for each situation is under way.

A novel approach for the inclusion of Turbulent Kinetic Energy into the delivered HIRLAM wind speeds was developed. However, more research is needed to find the optimal use of this parameter in the wind power forecasts. As a side result of the analysis could be shown that the use of a wind closer to hub height has clear advantages, as the NWP model includes much more atmospheric knowledge than is used for the usual 10-m a.g.l. wind speed predictions.

Concluding can be said that the project developed a method to use the wealth of information contained in meteorological ensembles. A demo application brought the results of the analyses (longer forecasts, and uncertainty forecasts) to the end users in the group. The utilities as end users have used the data in their day-to-day operation. Additionally, we looked into ensemble downscaling, the parameterisation of TKE and multi-model ensembles, with interesting and/or encouraging results.

Glossary

CHP	Combined Heat and Power
DMI	Danmarks Meteorologiske Institut (Danish Meteorological Institute, Copenhagen, DK)
DTU	Danmarks Tekniske Universitet (Technical University of Denmark, Lynby, DK)
DWD	Deutscher Wetterdienst (German Weather Service, Offenbach, DE)
ECMWF	European Centre for Medium Range Weather Forecasts (Reading, UK)
EPS	Ensemble Prediction System (used generically or for the one of ECMWF)
IMM	Informatik og Matematisk Modellering (an institute at the DTU)
NCEP	National Centers for Environmental Prediction (Camp Springs, Maryland, US)
NWP	Numerical Weather Prediction
PSO	Public Service Obligation (a user financed research grant scheme in Denmark)
TKE	Turbulent Kinetic Energy
Point forecast	A wind power forecast for one specific site and point in time
Uncertainty forecast	The forecast of the uncertainty of the point forecast

Acknowledgements

The project under which the work described here was carried out was 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), and the European Centre for Medium-Range Weather Forecasts for supplying the data used in this study. Devon Yates is acknowledged for her work on the demo result analysis.

References

1 Giebel, G., G. Kariniotakis, R. Brownsword: *The State-of-the-Art in Short-Term Forecasting of Wind Power - A Literature Overview*. Position paper for the Anemos project, to be downloaded from anemos.cma.fr. 2003

2 Pinson, P., G. Kariniotakis: *On-line Assessment of Prediction Risk for Wind Power Production Forecasts*. Proceedings of the European Wind Energy Conference and Exhibition, Madrid, Spain, 16-19 June 2003.

3 Lange, M., D. Heinemann: *Accuracy of short term wind power predictions depending on meteorological conditions*. Poster P_GWP091 on the Global Windpower Conference and Exhibition, Paris, France, 2-5 April 2002

4 Landberg, L., G. Giebel, H.Aa. Nielsen, T.S. Nielsen, H. Madsen: *Short-term Prediction – An Overview*. Wind Energy **6**(3), pp. 273-280, June 2003. DOI 10.1002/we.96

5 Lange, M., H.-P. Waldl: *Assessing the Uncertainty of Wind Power Predictions with Regard to Specific Weather Situations*. Proceedings of the European Wind Energy Conference, Copenhagen, Denmark, 2-6 June 2001, pp. 695-698, ISBN 3-936338-09-4. *(Note: the paper is misprinted in the proceedings, better follow the link provided to their university homepage.)*

6 Landberg, L., G. Giebel, L. Myllerup, J. Badger, T.S. Nielsen, H. Madsen: *Poor man's ensemble forecasting for error estimation.* AWEA, Portland/Oregon (US), 2-5 June 2002

7 Giebel, G., L. Landberg, K. Mönnich, H.-P. Waldl: *Relative Performance of different Numerical Weather Prediction Models for Short Term Prediction of Wind Energy*. Proceedings of the European Wind Energy Conference, Nice, France, 1-5 March 1999, pp. 1078-1081, ISBN 1 902916 00 X

8 Waldl, H.-P., and G. Giebel: *The Quality of a 48-Hours Wind Power Forecast Using the German and Danish Weather Prediction Model.* Wind Power for the 21st Century, EUWEC Special Topic Conference, Kassel (DE), 25-27 Sept 2000

9 Waldl, H.-P., and G. Giebel: *Einfluss des dänischen und des deutschen Wettervorhersagemodells auf die Qualität einer 48-Stunden-Windleistungsprognose*. 5. Deutsche Windenergiekonferenz DEWEK 2000, Wilhelmshaven (DE), 7-8 Jun 2000, pp. 145-148

10 Moehrlen, C., J. Jørgensen, K. Sattler, E. McKeogh: *Power Predictions in Complex Terrain With an Operational Numerical Weather Prediction Model in Ireland Including Ensemble Forecasting*. Poster on the World Wind Energy Conference in Berlin, Germany, June 2002

11 Roulston, M.S., D.T. Kaplan, J. Hardenberg, L.A. Smith: *Value of the ECMWF Ensemble Prediction System for Forecasting Wind Energy Production*. Proceedings of the European Wind Energy Conference, Copenhagen, Denmark, 2-6 June 2001, pp. 699-702, ISBN 3-936338-09-4

12 Machenhauer, B. (Ed), 1988, *HIRLAM final report*, HIRLAM Technical Report 5, Danish Meteorological Institute, Copenhagen, Denmark. For an introduction, see also *http://www.dmi.dk/eng/f+u/* .

13 Sass, B.H., N.W. Nielsen, J.U. Jørgensen, B. Amstrup, M. Kmit, and K.S. Mogensen: *The operational DMI-HIRLAM system 2002-version*. DMI Tech. Rep. 02-05

14 See http://www.ecmwf.int/

15 *Quarterly Report of the Operational NWP-Models of the Deutscher Wetterdienst.* No 28, downloaded from www.dwd.de. August 2001

16 Molteni, F., R. Buizza, T.N. Palmer, and T. Petroliagis: *The ECMWF ensemble prediction system: Methodology and validation*. Q. J. R. Meteor. Soc.. **122**, 1996, pp. 73–119

17 Undén, P., L. Rontu, J. Calvo, G. Cats, J. Cuxart, K. Eerola, C. Fortelius, J.A. Garcia-Moya, N. Gustafsson, C. Jones, S. Järvenoja, H. Järvinen, P. Lynch, A. McDonald, R. MsGrath, B. Nvascues, V. Odegaard, E. Rodriguez, M. Rummukainen, R. Rööm, K. Sattler, H. Savijärvi, B.H. Sass, B.W. Schreur, H. The, and S. Tijm: *The HIRLAM-5 Scientific Documentation*. HIRLAM-5 project, SMHI Norrköping, Sweden, 2002

18 Brasseur, O., 2001: *Development and Application of a Physical Approach to Estimating Wind Gusts*. Mon. Wea. Rev., Vol. 129, 5-25.

19 Nielsen, N.W. and Petersen, C., 2001: *Calculation of wind gusts in DMI-HIRLAM*. DMI Scientific Report, 01-03, 32pp.

20 Landberg, L: *Short-term Prediction of Local Wind Conditions.* PhD-Thesis, Risø-R-702(EN), Risø National Laboratory, Roskilde, Denmark 1994, ISBN 87-550-1916-1

21 See http://www.windatlas.dk/

22 Joensen, A., G. Giebel, L. Landberg, H. Madsen, and H. Aa. Nielsen: *Model Output Statistics Applied to Wind Power Prediction*. Proceedings of the European Wind Energy Conference, Nice, France, 1-5 March 1999, pp. 1177-1180, ISBN 1 902916 00 X

23 Landberg, L: *A Mathematical Look at a Physical Power Prediction Model.* Wind Energy **1**, pp. 23-28 (1998). DOI: 10.1002/(SICI)1099-1824(199809)1:1<23::AID-WE9>3.0.CO;2-9

24 Giebel, G.: *On the Benefits of Distributed Generation of Wind Energy in Europe.* PhD thesis from the Carl von Ossietzky Universität Oldenburg. VDI-Verlag, Schriftenreihe Energietechnik, 2001. ISBN 3-18-344406-2

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

26 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, 2002.

27 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. Norrköing, Sweden, 2002.

28 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.

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

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

31 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.

32 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.

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

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

35 Pinson, P., G. Kariniotakis, H.Aa. Nielsen, T.S. Nielsen, and H. Madsen: *Properties of quantile and interval forecasts of wind generation and their evaluation*. In Proceedings of the European Wind Energy Conference & Exhibition, Athens, March 2006. EWEA (to appear).

36 Nielsen, H.Aa., D. Yates, H. Madsen, T.S. Nielsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, and H. Feddersen: *Analysis of the results of an on-line wind power quantile forecasting system.* Technical report, Informatics and Mathematical Modelling, 2005. Prepared under PSO-project FU2101.

37 A. Boone, G. Giebel, K. Sattler, H. Feddersen: *Analysis of HIRLAM including Turbulent Kinetic Energy*. Project report, 2005

38 Zhu, Y., Toth, Z., Wobus, R., Richardson, D., and Mylne, K.. 2002. The economic value of ensemble-based weather forecasts. Bull. Amer. Met. Soc.. 83, 73–83.

39 Feddersen, H., and K. Sattler: *Verification of wind forecasts for a set of experimental DMI-HIRLAM ensemble experiments*. DMI Scientific Report 05-01, 2005

40 Bremnes, J. B.. 2004. *Probabilistic forecasts of precipitation in terms of quantiles using NWP model output*. Mon. Wea. Rev.. 132, 338–347.

41 Feddersen, H. and U. Andersen: *A method for statistical downscaling of seasonal ensemble predictions*. Tellus A 57, pp. 398-408, 2005.

42 Giebel, G., and A. Boone: *A Comparison of DMI-Hirlam and DWD-Lokalmodell for Short-Term Forecasting*. Poster on the European Wind Power Conference and Exhibition, London(UK), 22-25 November 2004

Mission

To promote an innovative and environmentally sustainable technological development within the areas of energy, industrial technology and bioproduction through research, innovation and advisory services.

Vision

Risø's research **shall extend the boundaries** for the understanding of nature's processes and interactions right down to the molecular nanoscale.

The results obtained shall **set new trends** for the development of sustainable technologies within the fields of energy, industrial technology and biotechnology.

The efforts made **shall benefit** Danish society and lead to the development of new multi-billion industries.