What performance can be expected by short-term wind power prediction models depending on site characteristics?

Georges Kariniotakis, Ignacio Marti, D. Casas, Pierre Pinson, T. S. Nielsen, H. Madsen, G. Giebel, J. Usaola, I. Sanchez, A.M. Palomares, et al.

To cite this version:

Georges Kariniotakis, Ignacio Marti, D. Casas, Pierre Pinson, T. S. Nielsen, et al.. What performance can be expected by short-term wind power prediction models depending on site characteristics?. European Wind Energy Conference EWEC 2004, European Wind Energy Association, Nov 2004, Londres, United Kingdom. hal-00529266

HAL Id: hal-00529266

https://hal-mines-paristech.archives-ouvertes.fr/hal-00529266

Submitted on 3 Mar 2020

What performance can be expected by short-term wind power prediction models depending on site characteristics?

G. Kariniotakis[1], I. Martí[2], D. Casas[2], P. Pinson[1], T. S. Nielsen[3], H. Madsen[3], G. Giebel[4], J. Usaola[5], I. Sanchez[5],
A.M. Palomares[6], R. Brownsword[7], J. Tambke[8], U. Focken[8], M. Lange[8], P. Louka[9], G. Kallos[9], C. Lac[10],
G. Sideratos[11], G. Descombes[12].

[1]Ecole des Mines de Paris, [2]CENER, [3]IMM-DTU, [4]RISOE, [5]Univ. Carlos III de Madrid, [6]CIEMAT, [7]RAL-CCLRC,
[8]Univ. Oldenburg, [9]IASA, [10]Meteo-France, [11]NTUA,[12]ARIA.
imarti@cener.com & georges.kariniotakis@ensmp.fr

Abstract

The scope of the European R&D project ANEMOS (ENK5-CT-2002-00665) is to improve the accuracy of short-term wind power forecasting technology. In the frame of this project, a number of case-studies were set-up covering onshore wind farms at flat, semi-complex and complex terrain as well as offshore ones. For these cases, several years of measurements and numerical weather predictions from various Numerical Weather Prediction systems were collected. A number of eleven state-of-the-art models were run for the test cases. These are operational models used today in Spain, Denmark, Ireland, Germany, Greece etc. An appropriate protocol was established to enable meaningful and systematic comparisons. The paper presents in detail results from this evaluation as well as guidelines on assessing the performance of a model as a function of the site characteristics.

1. Introduction

Short-term wind power forecasting (WPF) for the next hours up to the next days ahead is recognized as an important requirement by end-users such as transmission system operators, utilities, energy service providers, wind farm managers and energy traders. WPF is expected to be useful in operations such as unit commitment, economic dispatch, reserves allocation, storage management, trading in electricity markets and others. The accuracy of a prediction model is a primary concern since it is directly related to the security of the supply and to operating costs in a deregulated environment.

Operational tools for WPF exist since early 90's while in the last years a large number of models have appeared complicating the choice for end-users. The primary choice criteria are the accuracy and the robustness. These depend on a large number of factors and no systematic approach is available to help in this decision.

In the European R&D project ANEMOS (ENK5-CT-2002-00665) [1] a number of case-studies were set-up covering onshore wind farms at flat, semi-complex and complex terrain as well as offshore ones. Several years of measurements and numerical weather predictions from various systems such as Hirlam, Aladin, Skiron etc. were collected. A number of "base-line" models were run for the test cases, such as Prediktor, WPPT, Previento, Sipreolico, CENER's LocalPred, the Armines AWPPS, RAL's model, ARIA wind, and NTUA's. Most of these systems are operational today and used by system operators in Spain, Germany, Denmark, Ireland and Greece. In fact most of the actual installed capacity in Europe is currently predicted by these systems. The aim of the paper is to evaluate the performance of these state-of–the-art systems and use this performance as reference for evaluation of new models like the models developed in the ANEMOS project. Moreover, an appropriate protocol was established to enable meaningful and systematic comparisons [5]. Further diagnostic criteria permitted to gain insight on the behavior of the models in time and in space.

The paper presents in detail the results of this evaluation procedure providing a clear view on what should be expected in terms of accuracy depending on the case. Guidelines are then proposed for evaluating the level of performance of a prediction model as a function of the characteristics of a site, the size of a wind farm, the geographic location, the quality of data, the type of numerical weather predictions, the type of the model etc. Such guidelines can be used for the selection of a prediction model. The evaluation of the above state-of-the-art models has contributed in identifying new research directions for evolving to a high performance wind power forecasting technology. The paper concludes with the presentation of the improvements that are performed in this area.

2. Case-studies

Within the ANEMOS project six European case studies have been defined to aid the model development. The model comparison presented here is based on these wind farms. The wind farms have been selected to cover a wide range of conditions with respect to climatology and terrain and are located in four different countries:

- the Wusterhusen wind farm in Germany,
- the Alaiz and Sotavento wind farms in Spain,
- the Klim and Tunø wind farms in Denmark,
- and finally the Golagh wind farm in Ireland.

These six sites have had their output predicted off-line by 11 different models from 9 institutes; all based on the same NWP model inputs. For an overview of the differences between the models, see [3].

Different types of sites are well represented among the selected wind farms with three wind farms placed on-shore, two near-shore, and one offshore. One of the Spanish wind farms is placed in complex terrain, the second Spanish and the Irish wind farm are placed in semi-complex terrain and the three remaining wind farms are all situated in simple terrain.

When considering the climatic conditions, the Northern and Western parts of Europe are well represented whereas the Central and Southern parts of Europe are underrepresented. Especially the Mediterranean region is absent from the case study database. However, for the development of new physical models, additional test cases in Corsica and Crete have been defined in the framework of ANEMOS project and results will be reported in future publications.

The minimum data set collected for each of the six test cases consists of measured total production and NWP values of wind speed and wind direction interpolated to the site, but more detailed measurements and NWPs are available for most of the cases. For all the wind farms, detailed layout, contour and roughness maps have been collected together with the manufacturer's power and thrust curves of the turbines. In the following each of the wind farms will be presented in more detail.

The Wusterhusen wind farm is placed in the northeastern part of Germany 20 km southeast of the town of Greifswald and 8 km from the shoreline of the Baltic Sea. The wind farm consists of 2 Nordtank NTK500/41 turbines with a total rated capacity of 1.0 MW. The data set covers the period from January 1st 1999 to December 31st 2000. The NWP data has been interpolated to the site and comes from the German 0.15° Lokalmodell. The forecasts are updated once a day with a lead-time of 72 hours. The RIX value [2] is 0 for this wind farm, meaning that no slope is higher that the reference value (30%).

The Sotavento wind farm is placed in Galicia region in the North Western part of Spain approximately 40 km from the coastline of the Atlantic Ocean. The site is located 600-700 m above sea level in semi-complex terrain. The wind farm is a testing site and consists of large number of different turbines with a rated capacity ranging from 600 kW to 1320 kW. The total rated capacity of the Sotavento wind farm is 17.56 MW. The available data covers a period from January 1st 2001 to November 30th 2001. The NWP data consists of interpolated as well as gridded values and originates from the Spanish 0.2º HIRLAM model. The forecasts are updated four times a day with a lead-time of 24 hours. The RIX value for this wind farm is 7.

The Alaiz wind farm is situated 15km south of Pamplona in the Navarra region of Spain in very complex terrain 910 m – 1120 m above sea level. Alaiz is a large wind farm with a rated capacity of 33.09 MW distributed on 49 Gamesa G47-660 wind turbines and one Lagerwey LW750 turbine. The data set covers all of 2001 with NWP data as for the Sotavento wind farm. Alternative NWPs have been analyzed: SKIRON and ALADIN outputs at resolutions of 0.25 degrees (0.1 degrees after January 2003) and 0.1 degrees respectively. The RIX value for this wind farm is 15.

The Klim wind farm is located in the northwestern part of Jutland approximately 8km from the north coast and 50 km west of the city of Aalborg. The wind farm consists of 35 Vestas V44 600 kW turbines with a total rated capacity of 21.0 MW. The data set covers a period from January 1st 1999 to April 30th 2003 with NWP data from the Danish 0.15º HIRLAM model available as interpolated values at site. The forecasts are updated four times a day with a lead-time of 48 hours. The RIX value for this wind farm is 0.

The Tunø Knob wind farm is an offshore one situated 6km of the east coast of Jutland and 10km west of the island of Samsø. The wind is one of the first offshore wind farms in the world and consists of 10 Vestas V39 500 kW turbines with a total rated capacity of 5.0 MW. The measured data only consists of production data and covers a period from March 18th 2002 onto April 30th 2003. NWP data is available from the Danish 0.15º HIRLAM model both as gridded and interpolated values with updating frequency and lead time as for the Klim wind farm. The RIX value for this wind farm is 0.

The **Golagh wind farm** is located in the northwestern part of Ireland (Donegal County) 370 m above sea level. The turbines are 25 Vestas V42 600 kW machines corresponding to a rated capacity of 15.0 MW. Again power production is the only measured data available and data covers a period from August 1st 2002 onto March 31st 2003. NWP data comes from the Irish HIRLAM model as interpolated values at the site. The RIX value for this wind farm is 7.3.

3. Methodology

NWP analysis

The power prediction is based on NWP data, namely wind speed and direction. For this reason, the evaluation of the corresponding NWP models is necessary as a first step to the analysis of the power predictions. Such evaluation is performed for the Alaiz wind farm based on the meteorological observations obtained. The main purpose for selecting Alaiz as test case is the availability of detailed datasets in an area characterized by its topographic complexity.

NWPs for wind power purposes are required to refer to long periods, in order to derive statistically meaningful results. SKIRON system [6] runs operationally since the beginning of 2000 and its forecasts are systematically archived at the premises of Atmospheric Modelling and Weather Forecasting Group of the University of Athens. The forecasting horizon was 72 hours for the period 2000-2002 and 120 hours onwards with corresponding horizontal grid increments of 0.25 and 0.1 degrees. The model output frequency is 1 hour with initialisation at 12UTC. SKIRON datasets are widely used for several model developments and testing from various groups participating in ANEMOS project.

Similar datasets were available also by two widely used meteorological models, namely HIRLAM operated by the National Meteorological Institute of Spain and ALADIN run by Meteo-France for different time periods.

The period 1-31 December 2003 was selected for common meteorological models evaluation before their utilisation for power prediction.

The evaluation of the NWP models was based on:

- model to observation and model to model intercomparisons, and
- derivation of statistical parameters (Bias, Mean Absolute Error - MAE and Root Mean Square Error - RMSE).

Alaiz test case is the one with higher terrain complexity, as shows the RIX value. This complexity is reflected to the NWPs as shown in Figure 1. The monitoring level at the specific location is at 55 m above ground, while the wind predictions from the three NWP models are available at 10 m height. Therefore, direct comparison between model and observed values is difficult, although some useful information can be extracted. Focusing on the evolution (pattern) of the wind with time, which is necessary when comparing at different heights such kind of models with one-point measurements, it is clear that overall SKIRON, ALADIN and HIRLAM forecasts follow the observations. Clearly, as HIRLAM data are of much lower resolution (~50km) than the others, they show smaller variation with time and lower wind speeds. The behaviour exhibited by ALADIN was superior than that of HIRLAM. SKIRON predicts accurately the time variations. Its forecasts show a good agreement with the pattern of the observations even after the first 48 hours.

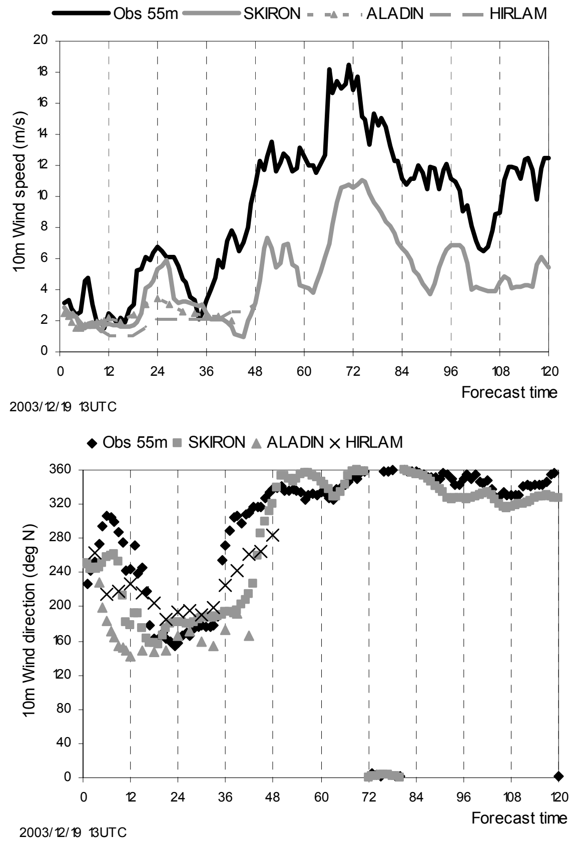

Figure 1: *Wind speed (upper) and direction (lower) time series of measurements and SKIRON, ALADIN and HIRLAM predictions initiated on 19/12/2003 at 12UTC.*

The differences in wind direction between models and observations are within an acceptable range.

In order to illustrate the importance of the model resolution in meteorological predictions and to investigate further the ability/limitations of the models to predict the wind speed at a single location, further analysis of the data was performed. Interpolated SKIRON and HIRLAM forecasts at 55m heights were plotted on the scatter diagram in Figure 2 for different forecasting horizons ranging between 1 to 5 days. It is shown that the forecast skill is considerable even during the 5th day. Both models compare well at low wind speeds (up to 6 m/s). At higher speeds SKIRON follows the observations more consistently (up to approximately 15 m/s), while HIRLAM underestimates the wind systematically. The underestimation in the wind NWP data is partially attributed to smoothing of the topography, especially in lower resolution configurations.

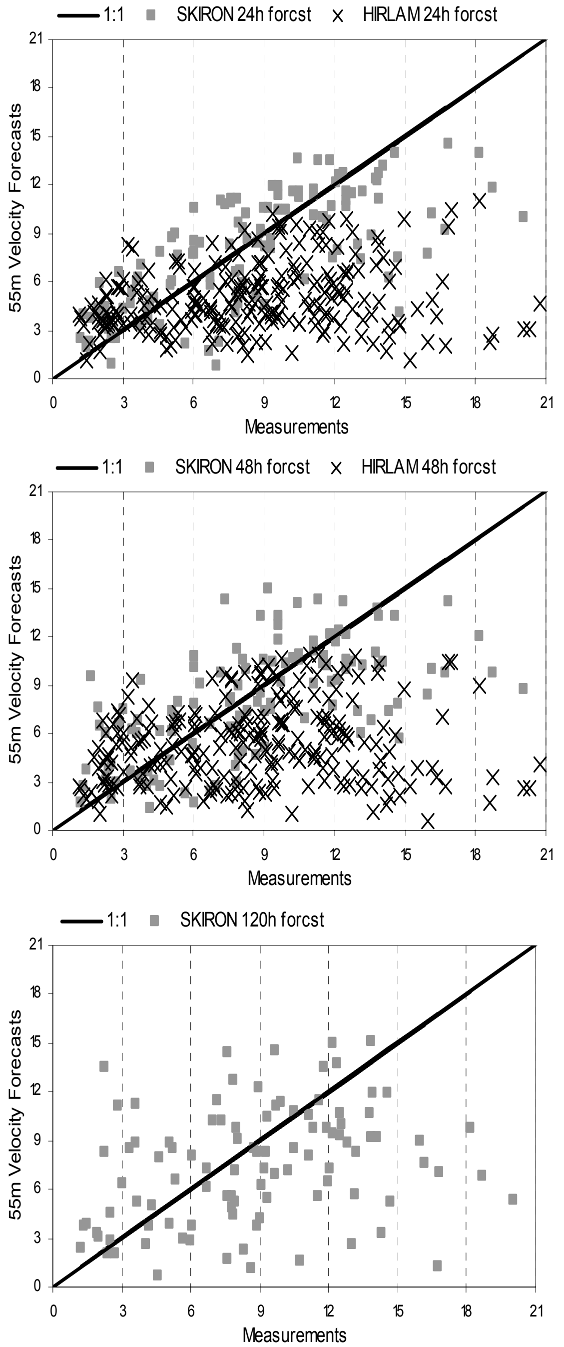

Figure 2: *Scattered plots of SKIRON and HIRLAM predicted wind speed interpolated at 55m vs. the measured values for 24-hour (upper), 48-hour (middle), and 120-hour for SKIRON only (lower) forecasts in December 2003. The thick solid line represents the 1:1 line.*

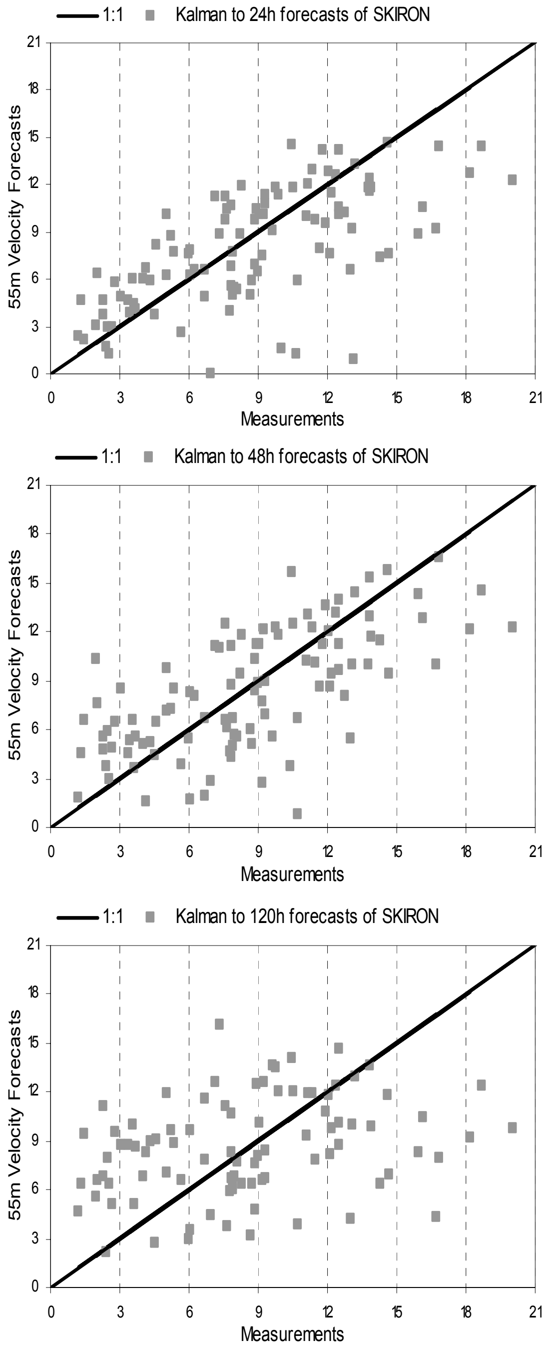

Figure 3: *Scattered plots of SKIRON predicted wind speed after Kalman filtering vs. the measured values for 24-hour (upper), 48-hour (middle), and 120-hour (lower) forecasts in December 2003. The thick solid line represents the 1:1 line.*

Kalman filtering is a methodology that has been widely applied for removing systematic errors in predicted meteorological parameters that follow linear behaviour, mainly air temperature. Within the framework of ANEMOS, a series of experiments was carried out for non-linear meteorological parameters, such as wind speed predictions, by utilising Kalman filtering techniques. Indicative results for Kalman filtering application to SKIRON data are illustrated in Figure 3.

Comparing Figure 2 and 3, it is clear that the application of such computationally inexpensive method for improving NWPs at wind farms is worthwhile. The results of the NWP model evaluation for Alaiz wind farm suggested that the existing regional weather prediction models can be considered as mature for wind potential forecasting applications, while their ability is also associated with the resolution used by each model. On the other hand, the disagreement in the behavior between models and point-measurements, is due to the fact that model grid-cell covers a large area, which for the particular models is approximately 10×10 km^2 for SKIRON and ALADIN and 50×50 km^2 for HIRLAM, leading to smoothing the topography of that area and hence averaging the characteristics of the flow in the area. In addition, discrepancies between models and measurements may be attributed to errors in the initial and lateral boundary conditions provided by the global model, and/or possible errors in the observations, e.g. instrumental errors, influence of the wind generators due to their location with respect to the measurement mast, etc.

4. Power production forecasts analysis

In order to compare the performance or the different power prediction models, the following procedure has been developed:

- A common format was defined to manage the different data (NWP and wind farm data).
- Common NWP for each test case.
- Common wind farm measurements (power production, wind speed and direction in some cases).
- A training period was defined for each test case in order to train those models that need it.
- An independent testing period was defined for each test case. The results presented in this paper correspond to the testing period of the test cases.

With this procedure every prediction model had the same inputs to generate the forecasts, each model using the information that needed contained in the common database.

In order to present homogeneous results, the following forecasts have been analyzed:

- Predictions calculated at 00 UTC.
- +12 hours forecasts horizons for the comparison of model performance.

The error parameters that have been calculated are the following:

- Mean absolute error normalized with the nominal power of the wind farm (NMAE).
- Determination coefficient (R^2).
- Percentage of errors lower than 10% of the nominal power.

The prediction models are identified by a number, without any reference to their names.

5. Results

Wusterhusen test case

This test case showed relative good performance with low values of NMAE, below 16% for all prediction models except for model 8 (see Figure 4).

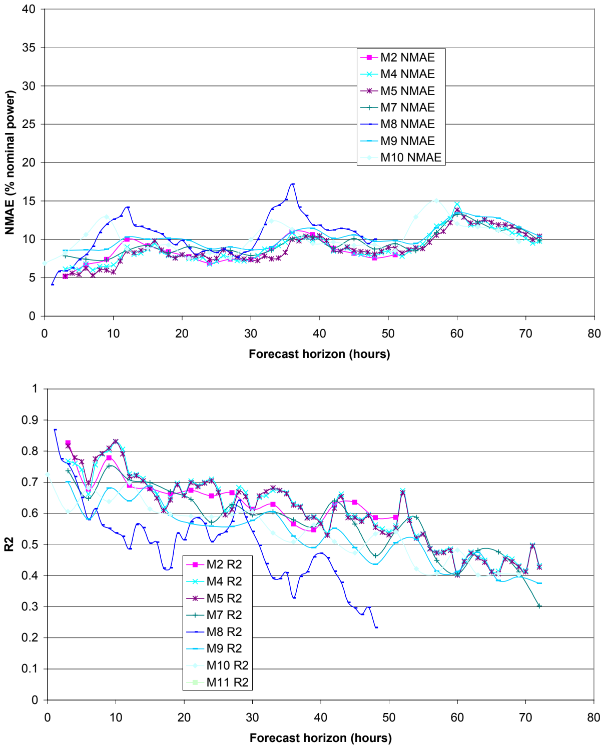

Figure 4 : *NMAE and R2 for Wusterhusen test case.*

A significant diurnal variation of the NMAE has been detected, having a first maximum for forecast horizons around 12 hours, a second maximum for horizons around 36 hours and a third one around 60 hours. The NMAE fluctuation is more evident for some of the models (8 and 10), while the others seem to be less affected. With respect to the determination coefficient R2, all the models (except model 8) have similar values, being the R2 values lower for the higher forecasts horizons.

Alaiz test case

Alaiz test case is the one with higher terrain complexity, as shows the RIX value. This is the most difficult wind farm to predict, with high values of the NMAE and high dispersion of the values for the different prediction modes (Figure 5). The scale of the errors is higher, ranging from 20% to 35% for the different modes and horizon 24. The determination coefficient R2 also presents a high dispersion and relatively low values for some of the prediction models.

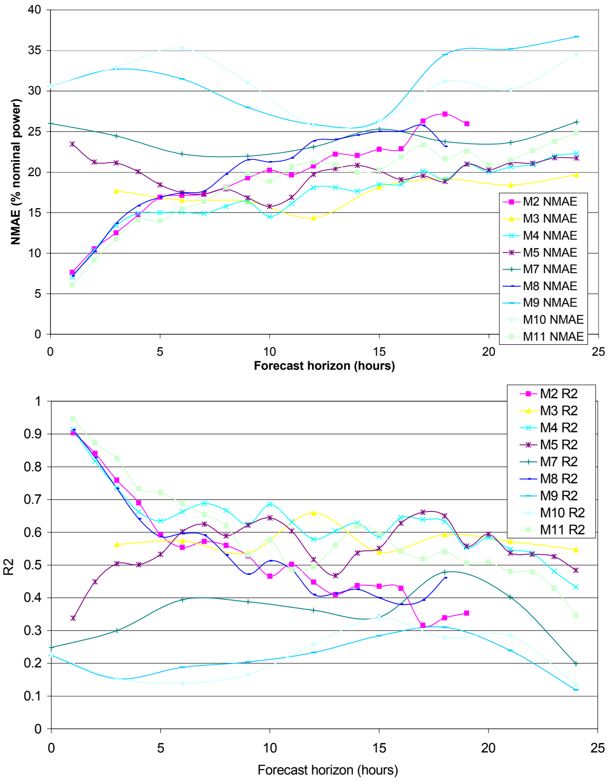

Figure 5: *NMAE and R2 for Alaiz test case.*

Sotavento test case

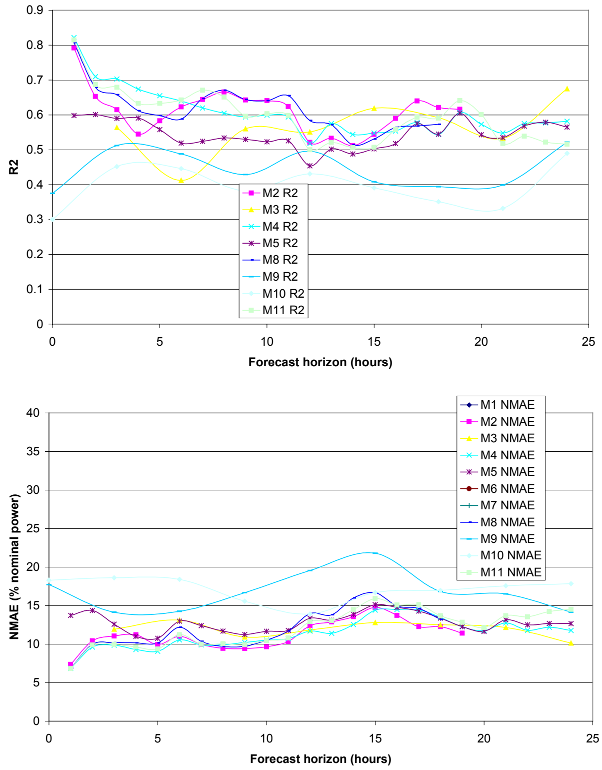

Figure 6: *NMAE and R2 for Sotavento test case.*

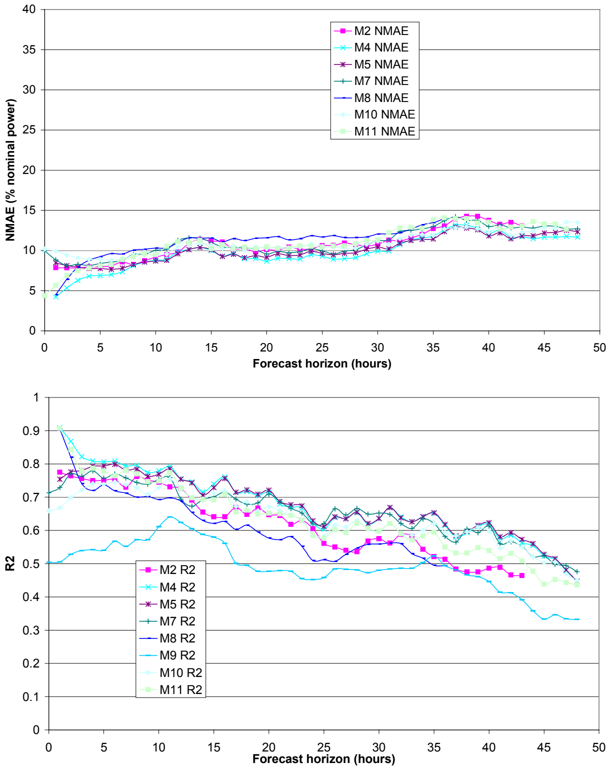

Figure 7: *NMAE and R2 for Klim test case.*

The results for Sotavento wind farm show a stable behavior of the NMAE, without a marked increase of the errors with the forecast horizon. Although not at the same degree as for Alaiz, there is a significant dispersion of the error values for the different prediction models (Figure 6).

Klim test case

Klim wind farm is the test case with lower values of NMAE, below 15% for all forecast horizons; and with very low dispersion of the values for the different prediction models.

The determination coefficient R2 also shows low dispersion except for model 9 and a progressive reduction of R2 with the forecast horizon.

Tunø Knob test case

Tunø Knob is an offshore wind farm. Low values of the NMAE are presented in Figure 8 with very low dispersion of the error values for the different prediction models. A progressive increase of the NMAE with the forecast horizon has been found.

The determination coefficient R2 ranges from 0.6 to 0.7 for the 24 hours forecast horizon (except for model 8 with a value below 0.6), while for 48 hours forecasts R2 values are around 0.5.

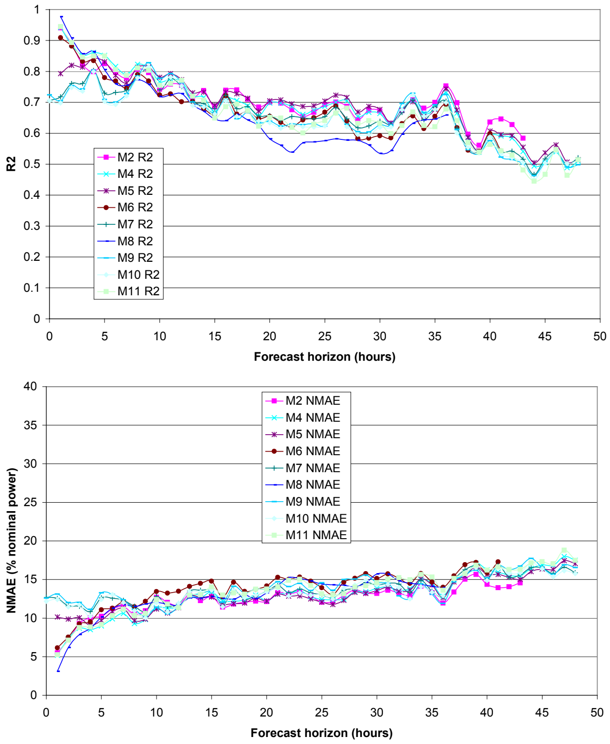

Figure 8: *NMAE and R2 for Tuno test case.*

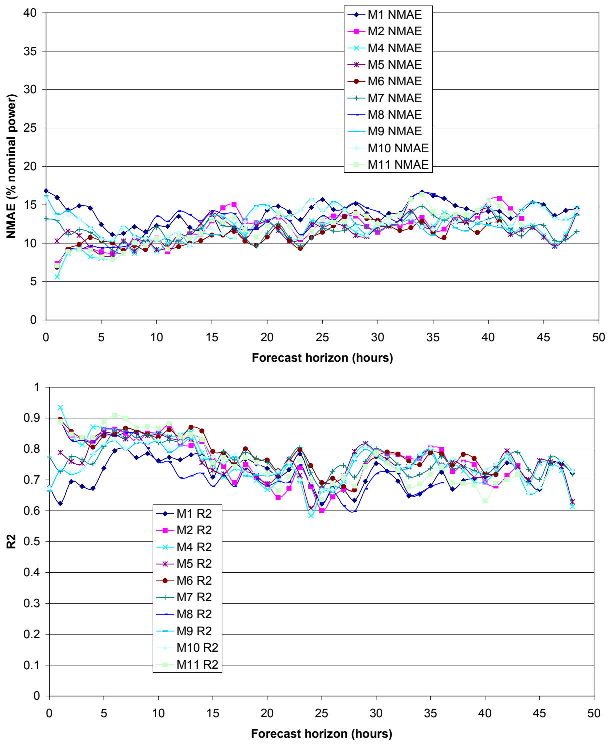

Figure 9: *NMAE and R2 for Golagh test case.*

Golagh test case

The NMAE values for Golagh wind farm are less dependent on the forecast horizon than for the other wind farms. The range of variation of NMAE for 24 hours horizon is 10% - 16%, being comparable for longer forecast horizons. Similar behavior can be seen for R2 values (Figure 9).

In general it can be seen in the previous figures that for the first forecast horizons, those models with autoadaptivity capabilities show better results (lower NMAE and higher R2 values). This improvement is more evident in the first 6 hours.

6. Comparison of model performance

For each wind farm different prediction models were available:

- Wusterhusen (WUS) 8 prediction models (M2, M4, M5, M7, M8, M9, M10).
- Alaiz (ALA) 9 prediction models (M2, M3, M4, M5, M7, M8, M9, M10, M11).
- Sotavento (SOT) 8 prediction models (M2, M3, M4, M5, M8, M9, M10, M11).
- Klim (KLI) 8 prediction models (M2, M4, M5, M7, M8, M9, M10, M11).
- Tuno (TUN) 9 prediction models (M2, M4, M5, M6, M7, M8, M9, M10, M11).
- Golagh (GOL) 10 prediction models (M1, M2, M4, M5, M6, M7, M8, M9, M10, M11).

Although not the same number of prediction models has been analyzed for every test case and different validation periods were studied, a qualitative comparison can be performed.

In order to compare the performance of the different prediction models a common forecast horizon for all the test cases has been selected: 12 hours. For this horizon and for each test case the following values were calculated, between the prediction models available for each test case:

- Average value of NMAE.
- Maximum value of NMAE.
- Minimum value of NMAE.

It can be seen that the average value of the NMAE varies between 10% (Wusterhusen wind farm) to 21% (Alaiz wind farm). The lowest dispersion corresponds to Tunø Knob wind farm being the difference between the best and the worst prediction model less than 0.6%; for Alaiz wind farm the dispersion in NMAE was the highest, 11.5%.

The performance of the prediction models is related to the complexity of the terrain. Figure 8 represents the variation of the average value of the NMAE for the 12 hours forecast horizon, for each test case. It can be seen that there is a significant increase in the NMAE values as the complexity of the terrain increases (higher RIX values). The offshore wind farm (Tunø) has slightly higher values of NMAE but similar to the ones obtained for the flat terrain wind farms.

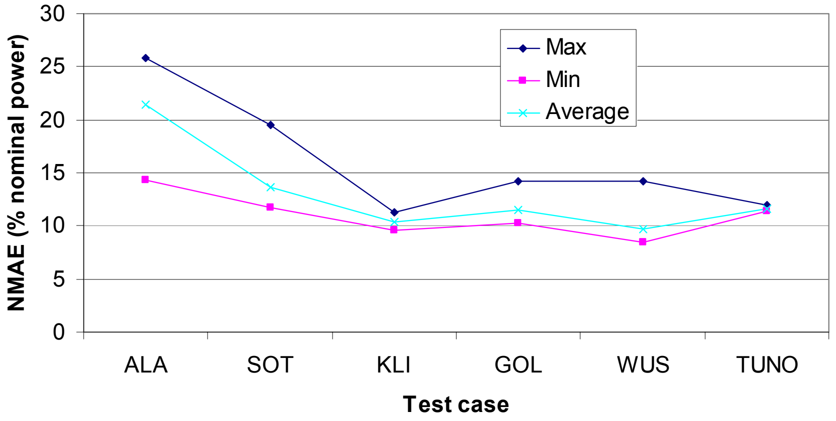

Figure 10: *NMAE variation for each test case. 12 hours forecast horizon. Qualitative comparison.*

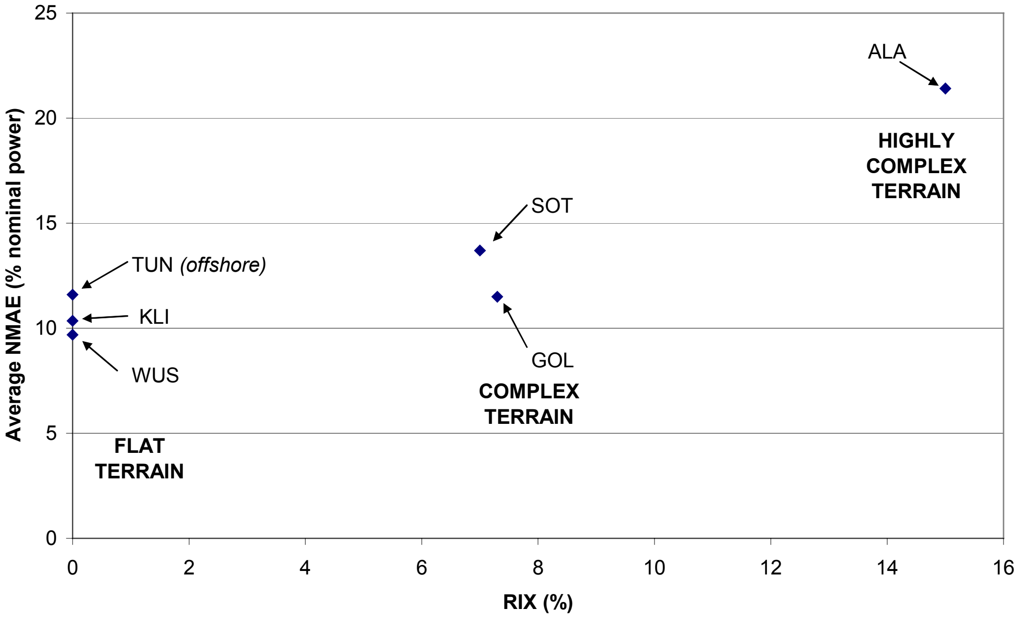

Figure 11:*Average NMAE for 12 hours forecast horizon vs RIX at each test case. Qualitative comparison.*

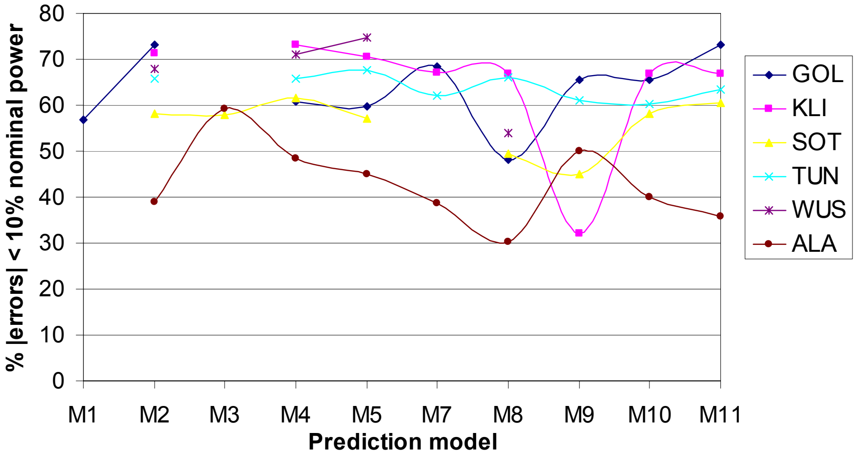

Figure 12: *Prediction model performance at each test case. Qualitative comparison.*

Power production forecasts have been calculated using as input to the prediction models the NWP provided by the meteorological services of the different countries (Germany, Spain, Denmark and Ireland), part of the performance variation of the forecasts is due to the several NWP used. However if we concentrate only in the Spanish cases, Alaiz and Sotavento, both had the NWP obtained with the same model and the increase of NMAE with RIX appears. The same applies for the Danish test cases (Klim and Tuno), both NWP were calculated with the same model.

Figure 12 shows the relative performance of each prediction model for every test case, in terms of % of absolute errors lower than 10% of nominal power.

7. Conclusions

In this paper the results of the first power prediction model competition in Europe are presented. This comparison has been made possible after a procedure followed for standardizing the available data, the definition of an evaluation protocol and last but not least the set-up on a web-based electronic infrastructure for handling data files and results respecting confidentiality issues.

The analysis of the different NWPs showed that the existing weather prediction models are able to provide useful wind predictions for the power production forecasts. Spatial resolution is of major importance especially in complex terrain. Low spatial resolution of NWPs may lead to an increase of errors in wind predictions. The benefits from post-treatment of NWPs with statistical correction techniques such as Kalman filtering are shown.

The performance of the power prediction models is related to the complexity of the terrain. For complex terrain higher NMAE values, above 35% for 24 hours horizon, and higher dispersion of the prediction model errors were obtained. In flat terrain NMAE values were below 10% for 24 hours horizon.

The offshore wind farm showed similar prediction errors to the flat terrain wind farms. Especially in these situations the results of the models are relatively close. This means that the models considered here are on the good side. This is neither trivial to obtain nor granted for an arbitrary model. This is due to the fact that the models considered here are in their majority "expert" ones in the sense that they are able to extract the maximum of the available information contained in the data for their calibration, while they are tuned for optimal generalization.

From the above results it can be concluded that for every test case, the optimum performance is obtained with different prediction models. In other words, no one of the studied forecasting models had the best performance for every horizon or for every case-study. This finding opens the possibility to improvements of performance by model combination.

This work continues in the ANEMOS project where following tasks include the study of the possibilities of error reduction by the combination of different predictions. Recent studies in this direction [4] show that it is possible to reduce errors in the power production forecasts by combining different NWP.

Further reporting of the results will include more criteria such as the RMSE, which represents better the impact of high errors. This criterion is expected to be sensitive for several end-users. Benchmarking is extended to upscaling and offshore cases and to more complex terrain cases (i.e. Crete and Corsica). Furthermore, comparisons with new models developed in the project are carried out (i.e. high-resolution models, CFD modeling for complex terrain, new statistical modeling with robust parameter estimation, etc), studies on offshore wind profiles. Ongoing research in the project focuses also on the development and evaluation of approaches for on-line uncertainty assessment.

Acknowledgements

This work was performed in the frame of the ANEMOS Project (ENK5-CT-2002-00665) funded in part by the European Commission. Acknowledgments are due to EDF, EHN, ELSAM, ESB, EWE, IDAE & PPC for providing the data for this work. Special thanks to the National Meteorological Institute of Spain for providing HIRLAM NWP of Spain.

References

[1] http://anemos.cma.fr

[2] **Bowen, A.J., Mortensen, N.G.**, Exploring the limits of WAsP: the Wind Atlas Analysis and Application Program. Proceedings of the 1996 European Union Wind Energy Conference and Exhibition, Göteborg, Sweden, May 20-24, 1996, pp. 584-587

[3] **Giebel, G., G. Kariniotakis, R. Brownsword:** *The State-of-the-Art in Short-Term Prediction of Wind Power – A Literature Overview.* Position paper for the ANEMOS project, download from http://anemos.cma.fr

[4] **Giebel, G., Boone, B.,** *A Comparison of the DMI-Hirlam and DWD-Lokalmodell for Short-Term Forecasting.* Proceedings of the EWEC, London (UK), 22-25 November 2004.

[5] **Madsen, H., Kariniotakis, G., Nielsen, H.Aa., Nielsen, T.S., Pinson, P.,** *"A Protocol for Standardising the Performance Evaluation of Short-Term Wind Power Prediction Models",* CD-rom Procd. of the Global WindPower 2004 Conference, Chicago, USA, Mar. 28-31, 2004.

[6] **Papadopoulos, A., Katsafados, P., Kallos, G.,** Regional weather forecasting for marine application. *Global Atmos. Ocean Syst.*, *8*, No 2-3, 2001, 219-237.