Evaluation of Advanced Wind Power Forecasting Models – Results of the Anemos Project

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

[1]CENER, [2]Ecole des Mines de Paris, [3]IMM-DTU, [4]RISOE, [5]Univ. Carlos III de Madrid, [6]CIEMAT, [7]RAL-CCLRC, [8]Univ. Oldenburg, [9]EMSYS, [10]NTUA,[11]ARIA.

imarti@cener.com & georges.kariniotakis@ensmp.fr

Abstract

An outstanding question posed today by end-users like power system operators, wind power producers or traders is what performance can be expected by state-of-the-art wind power prediction models. This paper presents results of the first ever intercomparison of a number of advanced prediction systems performed in the frame of the European project Anemos. A framework for error characterization has been developed consisting by a measure- and a distribution-oriented approach. This comparison has given a perspective of the possibilities and limitations of the forecasts in the different test cases that were defined. At a second stage, the homogenous comparison process has permitted to evaluate the possibility of obtaining better performance by exploiting the merits of individual models through model combination. The paper presents the methodology and results from the combination approach.

1. Introduction

The European project Anemos [1] has developed a wide research on several topics related to wind power forecasting such as physical and statistical modeling, uncertainty estimation, upscaling and others. From a very first stage of the project it was recognized by both end-users and modelers the necessity to map the existing wind power forecasting technology both in terms of research approaches and also in terms of performances. Initially an extensive literature review was developed and reported in [4].

Then, a comparison of a number of state of the art prediction models has been carried out in order to know what are the possibilities of the forecasting models under different situations. This comparison has given a perspective of the possibilities and limitations of the forecasts in the different test cases that were defined. This is the first comparison of wind power prediction models that is made at European level; the results are valuable information for the potential users of the prediction models about the typical ranges of error level, and the relation of the accuracy with the wind farm characteristics.

It is shown that the accuracy of power production forecasts as well as wind speed forecasts depends on the features of the wind farm as well as on the prediction model. This intercomparison exercise has been designed to cover different types of wind farms and state of the art forecasting models, therefore the results are a valid reference of the analysed prediction models performance for the final users. The test cases defined include complex terrain and relatively flat areas to take into account the effects of the topography; distance to the shore, different altitudes and climatic conditions.

A database has been developed including wind measurements, power production, and other meteorological data; numerical weather predictions were also included as well as the characteristics of the wind farm for each test case (power curves, digital maps of the terrain and roughness, etc). This database has provided all the necessary data for each model.

From the point of view of the prediction models, this exercise covers a wide variety of technical approaches, from autoregressive models to fuzzy logic neural networks, including MOS systems and boundary layer physical models. 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 or in market trading in Spain, Germany, Denmark, Ireland and Greece.

Apart from the power prediction models, the exercise was also extended to the comparison of numerical weather predictions from different systems. Detailed results of this task are presented in [2].

In order to be able to compare results by different models for the various test cases an appropriate

framework was defined characterized by a measure-oriented approach and a distribution approach for characterizing the deviations of the forecasts in relation to the measurements. The **measure-oriented** approach gathers a set of statistical error measures in the form of an "evaluation protocol" defined in [6]. Using this protocol one can derive conclusions on the performance of prediction methods and on what affects this performance (terrain, season, horizon etc).

The **distribution-oriented** approach aims to focus on the analysis of the joint distributions of predictions and observations. It investigates the influence of certain variables (i.e. level of predicted power, speed, a.o.) on the moments of error distributions (from the bias to kurtosis). This analysis is valuable to characterize prediction errors and relate weaknesses of models to specific causes. In this sense, it is a prerequisite for identifying areas of model improvement. Detailed results are presented in [12].

Finally, the homogenous comparison process has permitted to evaluate the possibility of obtaining better performance by exploiting the merits of individual models through model combination. The paper presents the methodology and results from the combination approach.

2. Selected wind farms

The objective of the benchmark was to study the performance of the prediction models under typical wind farm locations. Six test cases were selected to cover a wide range of conditions with respect to climatology and terrain and are located in four different countries:

- Wusterhusen wind farm in Germany (flat terrain),
- Alaiz (very complex terrain) and Sotavento (complex terrain) wind farms in Spain,
- Klim (flat terrain) and Tunø (offshore) wind farms in Denmark,
- Golagh wind farm in Ireland (complex terrain).

For the benchmarking 11 state of the art power prediction models [4] have been tested in the selected wind farms. In order to ensure that every prediction model run under the same conditions, a common database was created for each wind farm. Databases include datasets of wind and power measurements, as well NWP and all the necessary information about each wind farm (digital terrain maps with elevation and roughness, wind farm layout, wind turbine power and thrust curves).
The following NWP model outputs were used as inputs for the power prediction models:

- High Resolution Limited Area model (HIRLAM):
  - 0.2º grid resolution in Spain. The forecasts are updated four times a day with a lead-time of 24 hours.
  - 0.15º grid resolution in Denmark and Ireland. The forecasts are updated four times a day with a lead-time of 48 hours.
- Deutschland-Modell (DM) 0.15º grid resolution in Germany. The forecasts are updated once a day with a lead-time of 72 hours.

Test case	Classification	Training period	Validation period	Nominal power [MW]
Tunoknob	Offshore	Mar 2002 Dec 2002	Dec 2002 Apr 2003	5
Klim	Flat	Jan 1999 Feb 2001	Mar 2001 Apr 2003	21
Wusterhusen	Flat	Jan 1999 Jun 2000	Jul 2000 Dec 2000	1
Golagh	Complex	Aug 2002 Jan 2003	Feb 2003 Mar 2003	15
Sotavento	Complex	May 2001 Aug 2001	Sep 2001 Nov 2001	17.56
Alaiz	Very complex	Jan 2001 Aug 2001	Sep 2001 Dec 2001	31.77

Table I: *Characteristics of the wind farms selected as test cases.*

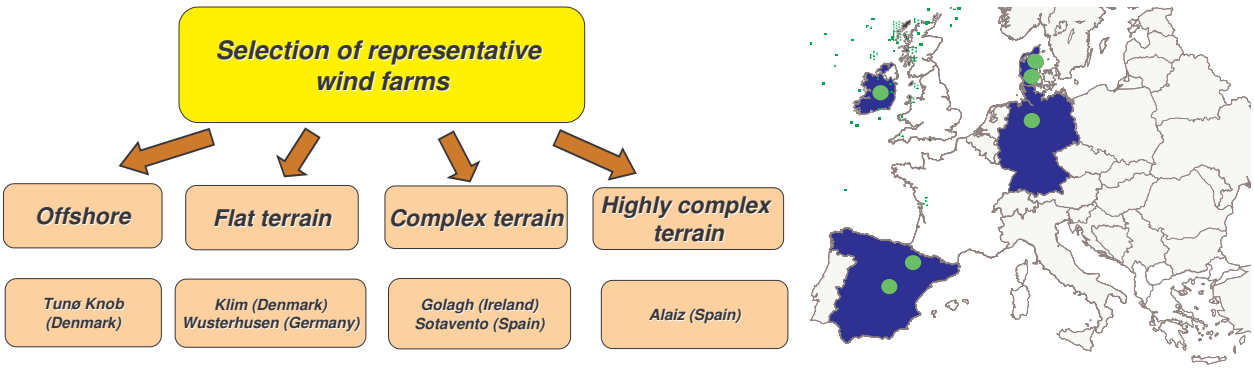

Figure 1: *Test cases in Spain, Ireland, Denmark and Germany.*

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 RIX value is 0 for this wind farm, meaning that no slope is higher that the reference value (30%).

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 RIX value for this wind farm is 7.

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 RIX value for this wind farm is 15.

Klim wind farm is located in the northwestern part of Jutland approximately 8 km 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 RIX value for this wind farm is 0.

Tunø Knob wind farm is situated offshore, 6km of the east coast of Jutland and 10km west of the island of Samsø. This 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 RIX value for this wind farm is 0.

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. The RIX value for this wind farm is 7.3.

3. Design a virtual laboratory for the benchmarking

In order to compare the performance of the power prediction models a benchmarking structure was designed. The objective was to characterise the performance of the models under the same input conditions:

- The different NWP and wind farm data were stored to a common database after conversion to a common format ("Depri") that was defined for this purpose.
- A web secured service was set-up to manage the available files and the results.
- Common NWPs were used for each test case.
- Common wind farm measurements (power production, wind speed and direction in some cases).
- A training period in the data set was defined for each test case in order to train those power prediction 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.
- A forecast error evaluation protocol was developed [6] for evaluating the performance of the prediction modes in a standardised way.

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.

Figure 2 summarizes the structure of the benchmarking. This structure has acted as a **"virtual laboratory"** in the frame of the project.

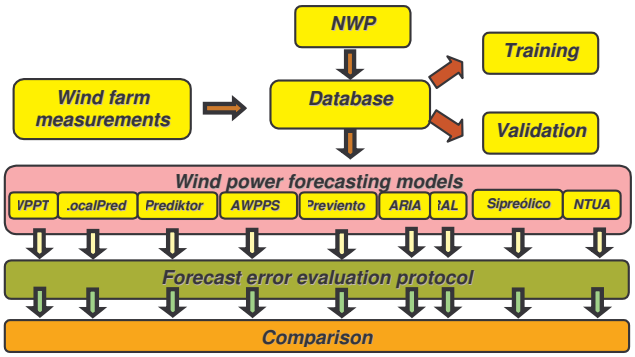

Figure 2: *Design a virtual laboratory for the benchmarking.*

4. Evaluation results

This Section presents representative results of the benchmarking exercise from the Alaiz and Golagh test cases characterized by very complex and complex terrains respectively. Complete results are given in [2].

The Alaiz test case is the one with higher terrain complexity, as indicated by the RIX value (15). This has been proven the most difficult wind farm to predict, with high values of the NMAE (Normalised Mean Square Error) criterion and high dispersion among the performances of the prediction models

(Figure 3). The scale of the errors is higher than common cases of complex terrain; it ranges from 20% to 35% for the different models and horizon 24. The determination coefficient R2 also presents a high dispersion and relatively low values for some of the prediction models.

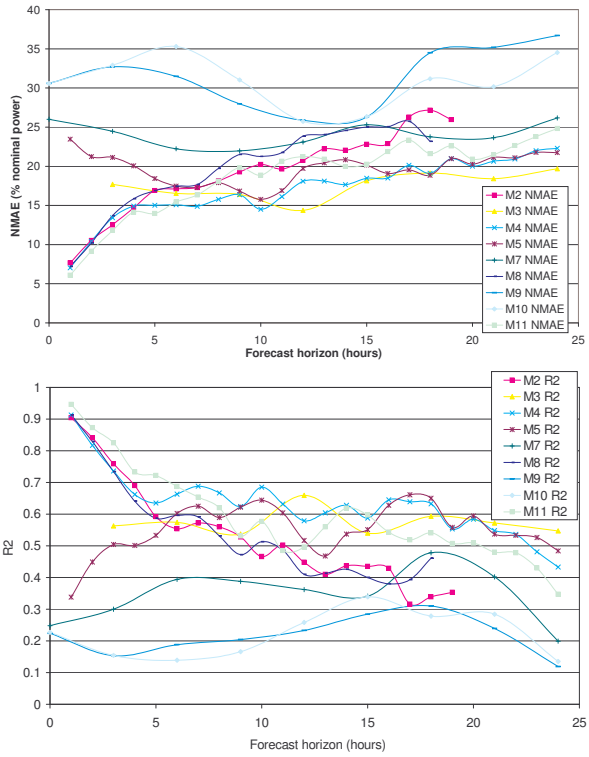

Figure 3: *NMAE and R2 for Alaiz test case.*

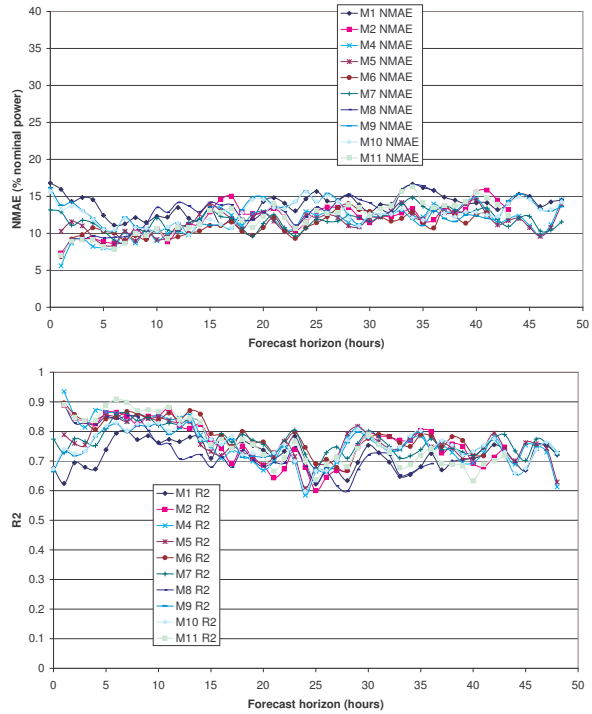

Figure 4: *NMAE and R2 for Golagh test case.*

The NMAE values for Golagh wind farm are less dependent on the forecast horizon than for the other sites. 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 4).

In general, it can be seen in the 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.

This study revealed both in a qualitative and quantitative way how performance of the prediction models is related to the complexity of the terrain. Figure 5 represents the variation of the average value of the NMAE for the 12 hours forecast horizon, for each test case. Figure 5 represents the performance of the studied power prediction models, showing the best, the worse and the average performance at 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.

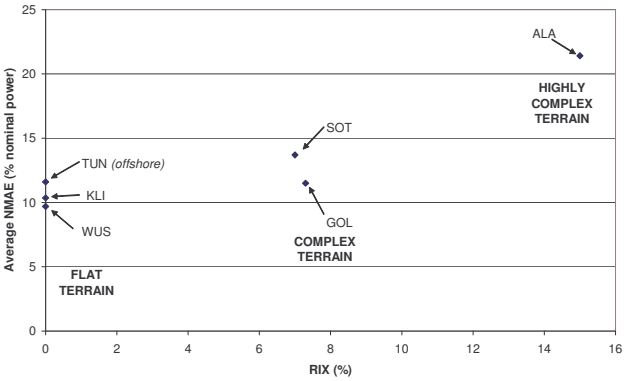

Figure 5: *Average NMAE for 12 hours forecast horizon vs RIX at each test case. Qualitative comparison.*

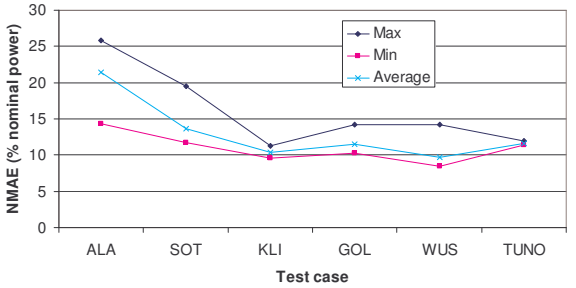

Figure 6: *Average NMAE for 12 hours forecast horizon vs RIX at each test case ordered by RIX value. Qualitative comparison.*

5. Distribution-oriented evaluation.

In a second stage a distribution-oriented approach for forecast verification was developed for highlighting the characteristics of forecast uncertainty. This approach is based on the notion that it is the joint distribution of forecasts $\hat{p}$ and observations p, $q(\hat{p},p)$ which contains all the non-time-dependent information about a prediction method's quality [8]. Such a distribution-oriented approach is also known as the Murphy-Winkler verification framework. While it is rather hard to directly study this joint distribution, one can instead focus on the various conditional and marginal distributions for deriving the necessary conclusions on the joint distribution properties. These various distributions include the conditional distributions of the observations given the forecasts $q(p|\hat{p})$, the conditional distributions of the forecasts given the observations $q(\hat{p}|p)$, the marginal distribution of the observations $q(p)$ and finally the marginal distribution of the forecasts $q(\hat{p})$. For all the various aspects of forecast quality and the way they can be assessed from the analysis of these distributions, we refer to [9]. Some of these aspects will be mentioned throughout the present paper.

Following a distribution-oriented approach, we have applied in the frame of the benchmarking exercise of the Anemos project a methodology consisting in studying the influence of a given variable (e.g. predicted power) on the moments of prediction error distributions (from the first to fourth order). Denote by $e_{t+k/t}$ the prediction error related to the power prediction $\hat{p}_{t+k/t}$ made at time t for lead time $t+k$. This is because these moments correspond to different characteristics of prediction errors:

- The mean μ_k^e locates the 'center of gravity' of a distribution and provides information on the systematic part of the error. It is given by the bias, as defined in [3].
- The standard deviation σ_k^e reflects the dispersion of a distribution, thus telling on the level of prediction uncertainty. It is given by the Normalized Standard Deviation of the Errors (NSDE) as defined in [10].
- The skewness ν_k^e describes the lack of symmetry of a distribution. It gives the most likely direction of expected prediction errors. A distribution with an asymmetric tail extending out to the right is referred to as positively skewed. The skewness is often estimated following Fisher's formula:

$$\nu_k^e = \frac{N_T}{(N_T-1)(N_T-2)} \sum_{t=1}^{N_T} \left(\frac{e_{t+k/t}-\mu_k^e}{\sigma_k^e} \right)^3 ,$$

  where N_T is the number of available prediction series in the evaluation set.
- the excess kurtosis κ_k^e informs on the shape of a given distribution, compared to the shape of normal distributions. The excess kurtosis for a normal distribution is equal to zero. Then, a positive excess kurtosis translates to a sharper peak and heavier tails. This moment is estimated by:

$$\kappa_k^e = \frac{N_T(N_T+1)}{(N_T-1)(N_T-2)(N_T-3)} \sum_{t=1}^{N_T} \left(\frac{e_{t+k/t}-\mu_k^e}{\sigma_k^e} \right)^4 - \frac{3(N_T-1)^2}{(N_T-2)(N_T-3)}$$

Application of the distribution-oriented approach for highlighting the effect of the power curve.

It is known that the contribution of the power curve to the power forecasting errors is to amplify or dampen wind speed prediction errors depending on the level of predicted wind speed. The power curve thus alters the shape of the wind speed error distributions. While previous studies [11] have only focused on the effect of the power curve on the general level of prediction error (expressed with measures), we want to go further here by basing our study on the distribution-oriented approach introduced above for better showing how the level of predicted power impacts error characteristics. We concentrate on the Tunø Knob test case which consists in fact an illustrative example of the conclusions that were drawn from the whole evaluation study. The analysis is based on 536 forecasts over a period of four and a half months. Wind power predictions are provided by 5 state-of-the-art methods (denoted by M1, M2,…, M5), with HIRLAM meteorological forecasts as input. M1, M2 and M3 are statistical prediction methods, while M4 and M5 belong to the family of physical prediction approaches.

A first investigation consists in studying the conditional distributions of the measures given the forecasts $q(p|\hat{p})$. This permits to assess the reliability of wind power forecasts [8]. Reliability is defined as the correspondence between the mean of the observations associated to a particular forecast and that forecast. It therefore translates to studying the dependence of the systematic error to the level of the predictand.

For this purpose, the range of forecast power values is split in 10% sub-ranges. In Figure 7, the normalized biases for the various prediction methods are given for each of the 10%-ranges of forecast values, for 18-hour ahead prediction. This horizon is chosen randomly since we have not observed significant differences as a function of the horizon. We will also concentrate on that particular prediction horizon in the remaining of the paragraph, keeping in mind that derived conclusions can be generalized over the whole range of look-ahead times. Bias values exhibit significant variations over the range of possible predicted outcomes. These values are comprised between -7% and 9% of P_n, and seem to have a general trend to be positive in the lower part of the power curve and negative in its higher part.

However, it does not appear possible to establish a clear relation between level of predicted power and prediction bias. Such behaviour has also been noticed for the other case-studies considered in the Anemos project with higher bias values for the wind farms located in semi-complex and complex terrain.

In a general way, even if methods' reliability is not perfect, we cannot identify a systematic lack of reliability in certain zones of the power curve or for a given method, apart from the trend we have expressed above.

The next step is to evaluate what are the variations in the shape of error distributions depending on the predictand value. First, the nonlinear and bounded nature of the energy conversion process makes that the skewness of error distributions evolves with the level of predicted power. Figure 8 depicts this evolution. Distributions are positively skewed for low predicted values and then negatively skewed when these values are in the high part of the power curve. Moreover, the nonlinear process acts on both the spread and peakedness of error distributions.

The spread dependence to the level of predicted power is shown in Figure 9, in which the spread is quantified by the standard deviation. Remember that the uncertainty of a given process is usually seen as the variability of its related observations. Then, studying the evolution of the spread of conditional distributions of the measures given the forecasts $q(p|\hat{p})$ relates to evaluating the predictand-dependent uncertainty. In parallel, excess kurtosis, as a function of predicted power, is depicted in Figure 10.

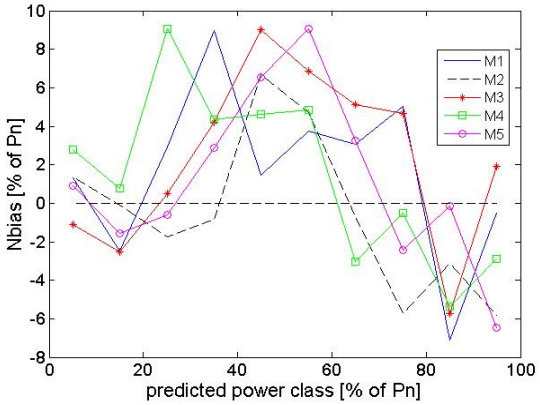

Figure 7: *Normalized bias of the forecasting error distributions depending on the predicted power range.*

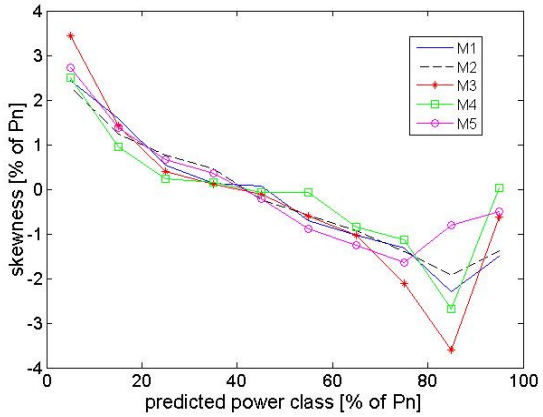

Figure 8*: Skewness of the forecasting error distributions depending on the predicted power range*

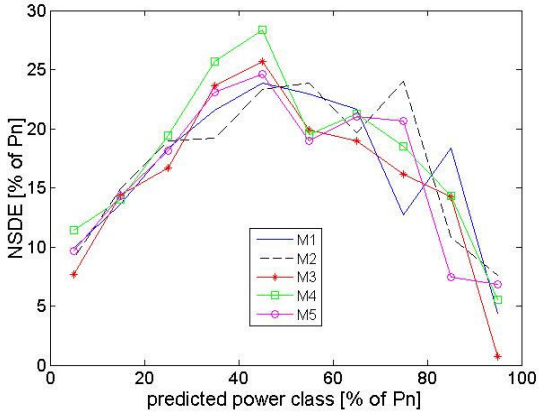

Figure 9*: Normalized standard deviation (quantified by the NSDE) of the forecasting error distributions depending on the predicted power range.*

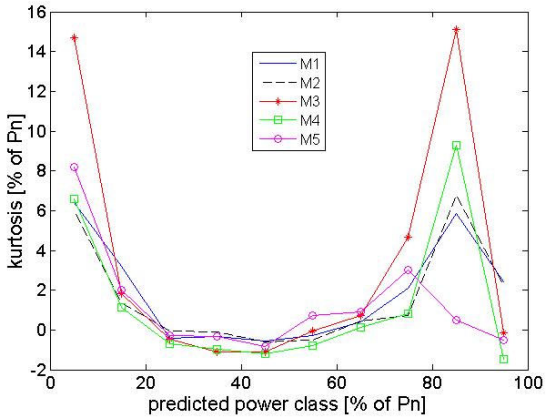

Figure 10*: Excess kurtosis of the forecasting error distributions depending on the predicted power range.*

High excess kurtosis values correspond to predicted power values close to minimum and maximum wind generation. Error distributions are highly peaked in these zones of the power curve. And, in the medium power range, slightly negative excess kurtosis values indicate that distributions are more flat than Gaussian distributions. At the same time, NSDE curves are almost symmetric with respect to the 50% power value. In the range of values related to the steep part of the power curve, the standard deviation is larger than for power values close to the power curve plateaus (say two or three times larger). Also, it can be seen that the standard deviations for these two plateaus are rather similar. The ratio between the uncertainty in the steep part of the power curve and the one in the low and high parts is approximately the same for all the prediction methods and the test cases considered in the full evaluation study, even if the shape of the standard deviation curves slightly differs from one test case to another. This tells us that the variations of the wind power forecasting uncertainty are similar whatever the wind farm and are in fact due to the wind-speed-to-power conversion process. Uncertainty levels may be higher when it is harder to predict wind speed (e.g. for complex terrain or offshore), but the way forecast uncertainty will vary as a function of the level of predicted power will be similar.

6. Combination of forecasts

It is not infrequent in wind energy to have access to more than one predictions of the wind farm production for the next hours. This has been the case in the Anemos project. In those cases, the adaptive combination of forecasts might be a useful methodology to generate an efficient single forecast.

In this work, a new adaptive combination method, called AEC, is proposed. The method is called Adaptive Exponential Combination (AEC) and is similar to always using the best individual predictor.

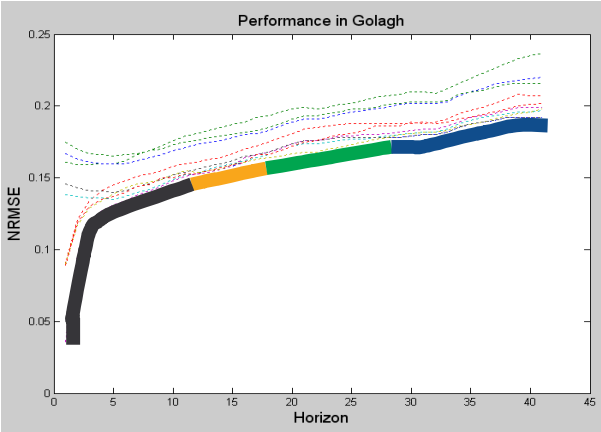

Figure 11*: Example of individual prediction models (dotted lines) and the expected effect of error reduction by prediction combination in Golagh wind farm. Combination depends on forecast horizon.*

It is based on a two step combination methodology to combine a set of alternative predictions. This two step procedure aims to take the advantage of the different approaches of forecast combination. In the first step, several combination methods are used, being the AEC one of them. In the second step, the AEC method is used to combine the alternative combinations of the first step. The application to a real data set illustrates the usefulness of the proposed methods to obtain the best output from a set of alternative predictions.

Two different types of combination approaches in a unified method

There are many approaches in the literature to perform combination of forecasts. For convenience, here we will classify the alternative approaches into two main classes depending on the goal of the combination.

The first class of combination methods will be denoted as *combination for improvement*. In this class, we target the best (constrained) linear combination of a set of forecasts. Methods to perform this combination for improvement can be based on the regression methodology, aimed at minimizing the residual variance of the linear combination. Ideally, the optimal linear combination would outperform the individual forecasts. In a practical situation, it is unclear how far we are of the ideal performance that can be obtained by combination. It is then possible that such combination is worse than some of the individual forecasts.

The second class of combination methods is called *combination for adaptation*. In this class we look for the combination that performs as well as the best individual procedure. This second class can then be interpreted as similar to a dynamic model selection, where the combination tends to put all the weigh to the best predictor, whenever it is clear that one of the predictors is the best one.

Two-steps combination of forecasts

In a practical situation we will not know in advance whether it would be better to use a combination for improvement method or, conversely, a combination for adaptation method. In order to benefit from both types of approaches we will apply them for wind energy forecast in a two step procedure. In a first step, we will apply alternative combination procedures based on the two mentioned approaches: one or more methods of adaptive combination for improvement, and also the AEC method of combining for adaptation.

In a second step, we will treat these alternative combinations of the first stage as a new combination problem, and will combine them to obtain the final combination. In this second step we will not expect to improve over the combined predictions of the first step, but only to assure that the final combination is as good as the best of the competing combined predictions. Then, it is a combination for adaptation problem and only the AEC method will be used. We will refer to this practice as a two-steps combination of forecasts.

Application to wind energy predictions

We will compare the performance of different combination schemes applied to a set of alternative forecasts of hourly wind energy production. We have 9 alternative series of forecasts produced with the prediction models tested in Anemos project, denoted as P1 to P9 as well as the time series of real observations for Golagh wind farm. The time span is three months, with a total of 2118 time periods. The 9 forecasters have had the opportunity to build and train their respective methods using a large enough portion of older data from the wind farm.

For the exercise, each hour the 9 forecasters had to supply predictions for the next 48 hours.

Table II shows the root mean squared error (RMSE) normalised with nominal power of each forecaster for selected horizons. This table displays the minimum RMSE across the forecasters, and then the difference between the RMSE of each forecaster and the minimum.

We can see from this table that the best predictor (bold numbers) is different at each horizon. As mentioned above, we perform a two-steps combination of forecasts. In a first step we combine the 9 competing forecast using the combination for improvement method (denoted as C1) and the combination for adaptation (AEC). In the second step, we combine these two combinations of forecasts of the first step using the AEC method. We will treat each prediction horizon as an independent combination problem. The results are displayed in Table III. The column of Minimum RMSE is the same as in the previous table.

	Minimum Difference from minimum RMSE									
H	**RMSE**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
1	0.036	0.125	0.054	0.102	**0.000**	0.054	0.110	0.131	0.139	0.053
6	0.134	0.028	0.014	0.003	**0.000**	0.009	0.007	0.026	0.032	0.006
12	0.153	0.028	0.008	**0.000**	**0.000**	0.001	0.003	0.020	0.023	**0.000**
18	0.164	0.029	0.011	**0.000**	0.004	**0.000**	0.004	0.019	0.021	**0.000**
24	0.171	0.030	0.016	0.005	0.008	**0.000**	0.007	0.023	0.025	0.009
30	0.179	0.031	0.009	0.002	0.004	0.002	**0.000**	0.023	0.024	0.010
36	0.189	0.034	0.005	0.003	0.004	**0.000**	**0.000**	0.023	0.022	0.010
41	0.192	0.045	0.010	0.005	0.007	0.006	**0.000**	0.028	0.024	0.015

Table II: *Minimum RMSE along the alternative predictors at each horizon, and difference of the RMSE of each predictor to that minimum.*

		Difference from minimum RMSE		
		First Step		Second Step
H	**Min. RMSE**	**C1**	**AEC**	**AEC**
1	0.036	0.019	0.000	-0.030
6	0.134	-0.001	0.003	0.000
12	0.153	-0.003	-0.001	-0.006
18	0.164	-0.007	-0.001	0.000
24	0.171	-0.003	-0.001	-0.003
30	0.179	-0.001	0.001	-0.005
36	0.189	-0.002	-0.001	-0.010
41	0.192	-0.004	-0.003	-0.015

Table III: *Minimum RMSE along the alternative predictors at each horizon, and difference of the RMSE of each combination with respect to that minimum.*

We see in this table that the proposed two steps combination methodology, together with the use of the proposed AEC method, helps to take the best

performance of a set of competing predictions for this wind farm.

7. Conclusions

Nine state of the art wind power prediction models have been compared in six wind farms. This is the first comparison that is carried out at European level on short term prediction. A framework has been developed for the benchmarking of the models, including a protocol for error analysis, common databases for each test case and the definition of a standard format for data and predictions exchange.

The results showed a dependency of the prediction errors on the complexity of the terrain as well as on the forecast horizon. The distribution of errors was studied and also the relation of the errors with the power curve.

Finally an algorithm to combine power prediction forecasts was developed and analysed, being possible to optimise the performance of a set of forecasts for a given wind farm.

8. 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.

9. References

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

[2] **Kariniotakis, G.**, **Marti, I.,** et al What performance can be expected by short-term wind power prediction models depending on site characteristics?. Proceedings of the 2004 European Union Wind Energy Conference and Exhibition, London, UK, November 22-25, 2004.

[3] **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

[4] **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

[5] **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.

[6] **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.

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

[8] **A. H.Murphy and R. L.Winkler.** A general framework for forecast verification. Monthly Weather Review, 115:1330–1338, 1987.

[9] **A. H. Murphy**. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather Forecasting, 8(2):281–293, 1993.

[10] **H. Madsen, P. Pinson, T.S. Nielsen, H.Aa. Nielsen, G. Kariniotakis**. Standardizing the performance evaluation of short-term wind power prediction models. Wind Engineering, 2006. In Press.

[11] **M. Lange.** On the uncertainty of wind power predictions - Analysis of the forecast accuracy and statistical distribution of errors. Trans. of the ASME, J. Solar Energy Eng., 127(2):177–194, May 2005

[12] **P. Pinson.** Estimation of the uncertainty in wind power foerecasting, PhD thesis, Ecole des Mines de Paris, Center for Energy & Processes, 23 March 2006.