EPREV - A WIND POWER FORECASTING TOOL FOR PORTUGAL

A. Rodrigues[1,4], J. A. Lopes[1,2], P. Miranda[3], J. Palma[1], C. Monteiro[1,2], J. N. Sousa[2], R. J. Bessa[2], C. Rodrigues[1], J. Matos[4]

[1]FEUP – Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias 4200-465 Porto - PORTUGAL
tel.: +351 225 081 763 email: ahr@fe.up.pt

[2]INESC-Porto – Instituto de Engenharia de Sistemas e Computadores do Porto
Rua Dr. Roberto Frias 4200-465 Porto – PORTUGAL
tel.: +351 222 094 000 email: jpl@fe.up.pt

[3]Centro de Geofísica da Faculdade de Ciências da Universidade de Lisboa
Campo Grande, Edifício C5 1149-016 Lisboa – PORTUGAL
tel.: +351 217 500 885 email: pmmiranda@fc.ul.pt

[4]INEGI – Instituto de Engenharia Mecânica e Gestão Industrial
Rua do Barroco, n.º 174 4465-591 Leça do Balio - PORTUGAL
tel.: +351 229 578 716 email: jmatos@inegi.up.pt

Abstract

Wind energy experiences in Portugal an increasing interest. Slightly more than 1700 MW were operating by the end of 2006, in a system with a global capacity of about 12 GW (8,5 GW peak demand). Several new wind farms are under construction and a considerable amount of connection points are or will be granted in the coming years. More than 5000 MW are expected to be connected to the grid around 2012, the global generating capacity being then about 16 GW. Clearly, a wind power forecasting system must be implemented that will help to deal with the significant penetration of the technology in the electrical system.

A group of wind farm promoters, owning the majority of the capacity installed so far, ordered to a consortium of universities and research institutes the development of a forecasting tool, giving rise of the EPREV project, wholly financed by them.

The system will have the following main characteristics:
- Wind speed and active power forecasting up to 72 hours;
- Evaluation of the forecasting uncertainty;
- Possibility of using the predictions of physical models and the information from the wind farm Supervisory Control And Data Acquisition (SCADA);
- Capacity of predicting only with SCADA information for very short term.

The main components of the system are:
- A human-machine-interface, allowing the control of the system, the selection and aggregation of forecasting models and the visualization of results;
- A power forecasting model for individual wind turbines and for wind farms. A cascade of models is used, starting in the mesoscale simulation, with up to 2 km resolution. The outputs of the mesoscale models are corrected and statistically adapted to the fine scale conditions. Two models and different boundary conditions are run, in three nested domains (54x54, 18x18 and 6x6 km). The advantage of using a 2x2 km resolution is also tested.

The statistical models are fed with recent information from the wind farms, after a learning process that made use of the historical information of its operation. Three different types of statistical models are employed: Power Curve Model (PCM), Auto Regressive (AR) and Neural Network Assembling Model (NNAM).

The wind simulation at the wind farm scale is done both by linearized physical models and Computational Fluid Dynamics (CFD) models, namely using VENTOS®, a code developed at the University of Porto.

The duration of the project is planned to be 1 year, including off-line tests of the complete system for 3 wind farms, for performance evaluation purposes.

Keywords: wind energy, wind power forecast, wind forecast, statistical models, physical models

1. Introduction

Portugal had, by the end of 2006, an installed wind energy capacity of about 1700 MW, and a total of more than 5000 MW is planned by 2012. The importance of short-term forecast of energy production from such an important wind energy penetration will grow as the energy grid systems will require more interaction with wind energy plants. Forecasts are an essential tool to support negotiations in the electricity daily market and an instrument for maintenance planning. Figure 1 illustrates the tasks a wind power forecasting tool may perform.

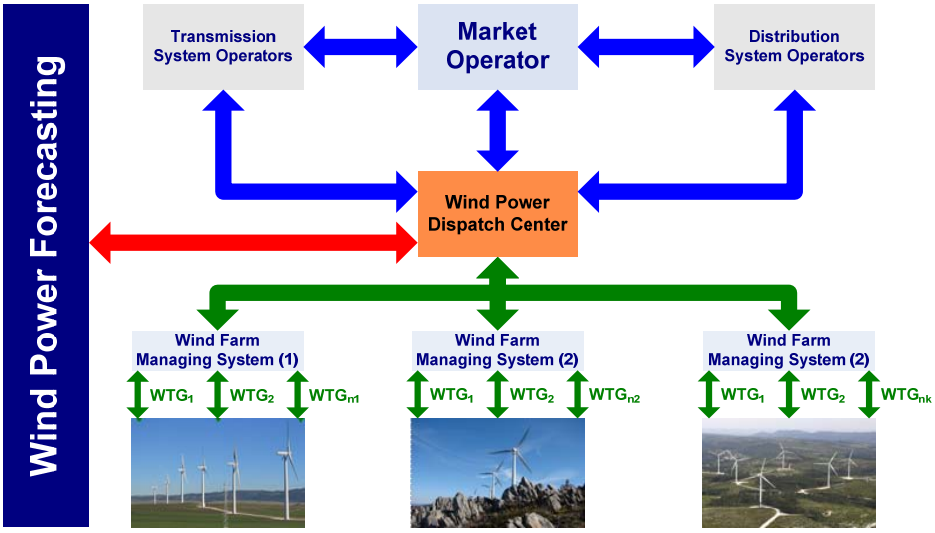


Fig. 1: Use of wind power forecast

A forecasting tool for wind power – EPREV - is under development, by a consortium of universities and research institutes, financed by a group of wind farm investors. The goal is to forecast the wind power output up to a 72 hours horizon with a sampling period of 30 minute, using historical and meteorological information. The duration of the project is planned to be 1 year, including offline tests of the complete system for 3 wind farms, for performance evaluation purposes.

2. Methodology

The proposed forecasting methodology is based on a chain of models, beginning at mesoscale models which give the wind conditions forecast in the 72 hours subsequent period for a characteristic site of each wind farm. These forecasts are converted to power either using physical or statistical models.

Statistical models require training which is done using historic data from the wind farm. However, in more recent wind farms the data length may not be sufficient to train the statistical models. In those cases, physical models shall assist in building an artificial power series to be used to do this task. Physical models can also be used to convert meteorological forecasts into power forecast as a complementary tool to the statistical models.

Figure 2 illustrates the adopted methodology.

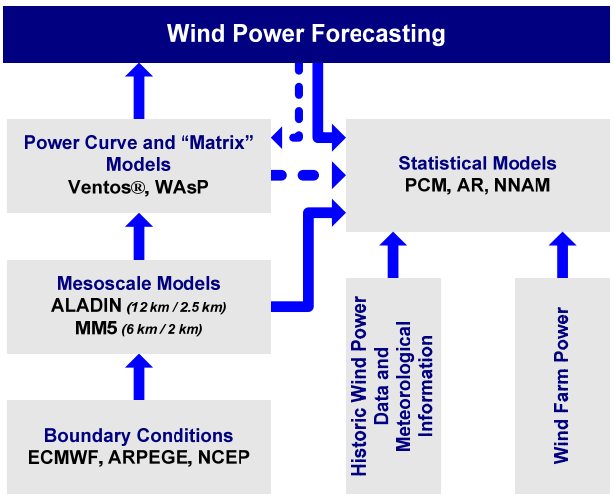


Fig. 2: General methodology

2.1. Setup of meteorological forecasts

For medium latitudes and time scales up to few hours, the variability of wind resource is strongly dependent of synoptic variability, associated to the movement and development of meteorological circulation mechanisms such as depressions, anticyclones and frontal perturbations. The diagnosis and behaviour forecast of such mechanisms is at the centre of meteorological services operation. Weather forecast global models perform successfully this analysis and forecast task.

In a more local scale, there are circulation mechanisms with impact on the regional wind climate which can be misinterpreted by the meteorological stations network or by the global forecast models. These mechanisms are classified as mesoscale orographic circulations, with purely mechanic or thermal-mechanic forcing, mesoscale storms associated circulations and breeze circulations due to strong roughness changes in the surface. Part of these circulation mechanisms is well represented in mesoscale meteorological models, however the quality of the forecasts is strongly dependent on the quality of the supplied initial and boundary conditions.

Finally, some circulation mechanisms are of much reduced scale, wind farm scale, generically belonging to a turbulent circulations scale. They are poorly represented by mesoscale models in operationally feasible spatial resolutions or are too dependent on the boundary conditions details, normally inaccessible in an operational context.

The three referred circulation mechanisms suggest wind forecast methodology able to integrate the three layers of forecast models: a global model, available in the National Weather Services; a mesoscale model whose operation can be performed in several ways; a forecast model at the wind farm's scale. Three forecast models are under test, using three global models made available by ECMWF (European Centre for Medium-range Weather Forecasts), by NCEP (National Centre for Environmental Prediction, USA) and by Météo-France, and two mesoscale models, MM5 and ALADIN, in two spatial resolutions. 6x6 and 2x2 km.

Presently, work is more focused on the MM5 model. MM5 [1, 2] is used in many weather forecasts such as the University of Lisbon [3]. It is a non-hydrostatic model, of limited area, which requires initial conditions supplied a meteorological analysis system and boundary conditions supplied by a global forecast system.

Figure 3 represents the 2005 annual average wind speed in continental Portugal using MM5 with a horizontal resolution in the inner domain of 6 km with the ECMWF operational analysis and the correspondent forecasts. These were performed in the calculus cluster of the University of Lisbon Geophysical Centre, CGUL.

An increase of resolution in the domains has an impact on the average speed figures, especially on mainland, where orography and roughness play a major role.

Simulations with limited area models require initial and boundary conditions in the lateral walls of the volume throughout the simulation process.

Initial conditions are given from a global network of observations, doing an ensemble of meteorological stations data and a great deal of remote observations, mainly based on meteorological satellites.

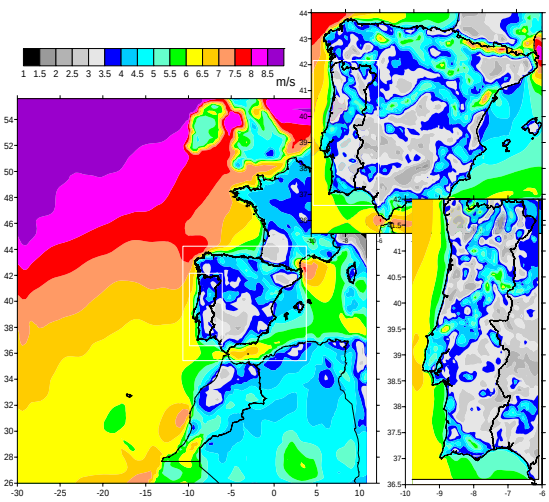


Fig. 3: Nested domains in the MM5 forecasts at resolutions of: 54 km (larger domain), 18 km and 6 km. Also shown the 2005 mean wind speed in the three domains

From this data, a three-dimensional state of the atmosphere map is daily built every twenty four hours, at 6 hours GMT. Hence, forecasts with a time length up to 72 hours are built four times per day. In order to test some combinations with the available atmospheric models, a set of 365 simulations to 72 hours based on the 00 h GMT were obtained with MM5 using boundary conditions supplied by NCEP and ECMWF. It was observed that at this 6-30 h time scale, both simulations are remarkably coherent with each other. However, small local differences are found, but always inferior to 10 %.

Forecasts are performed in a 30 minute sampling period and the local forecasts are performed by interpolation.

2.2. Physical models

The here designated Wind Farm Power Curve – WFPC – or Matrix models are based on the computational simulation of atmospheric flow in the wind farm area seeking, based on wind speed and direction at a certain site inside the domain, to estimate wind characteristics at the points where turbines are installed. With the turbine's power and axial thrust curves it is possible to estimate the global output of the wind farm, including wake losses.

Computational simulations consist on the resolution of flow equations, considering the characteristics of the site and its surroundings, such as topography, roughness and obstacles. Building the wind farm's power curve requires a great number of simulations in order to predict the energy output of the wind farm for the whole range of wind velocities in which the wind turbines operate, and for all wind rose sectors.

The WFPC shall be used to produce aliases wind farm's power series able to training of the statistical models, but can also be used for the wind farm's output forecasting after the corresponding wind forecast for the sampling 30 minute period.

In a first approximation, one of the most widely spread models in the wind resource assessment, WAsP [4], which makes used of the European Wind Atlas methodology [5], is used to build the WFPC. Being this a model that uses linearized versions of the fundamental equations, among other simplifications, it has several limitations. However, corrections can be performed either by cross-predictions between stations analysis using all of the data recorded in them, either by the analysis of the terrain's complexity [6].

The WFPC are also generated using the Computer Fluid Dynamics based model, VENTOS® [7, 8, 9], where the fluid flow equations are solved in their original non-linear form. The use of CFD to this purpose is somewhat new, but important improvements are expected in the construction of the wind farm matrixes, specially in cases os high terrain complexity.

Table 1 shows the results of this approach comparing the power recorded in the SCADA, P_{SCADA}, of a wind farm used as test case and the forecasted power, $P_{Forecast}$,using data from local stations and the generated WFPC. The predicted average power is higher than the measured one. The standard deviation presents a figure slightly larger than the average figure, indicating the existence of fluctuations of about 120 % for both time series.

Table 1: Information on the measured and predicted power

	$\overline{P}[\%]$	$\sigma(P)$	$100\times\sigma(P)/\overline{P}$
P_{SCADA}	455.8	553.0	121
$P_{Forecast}$	511.7	612.3	119

2.3. Statistical models

Meteorological forecasts can be converted into power using statistical models. The wind farm yield data is obtained through remote connection to each wind farm. This connection can be *dial-up*, *internet* or *GSM*, depending on the wind farm's infrastructures. The information collected from each wind farm is based on the generation of each wind turbine and wind farm generation.

The system is developed in JAVA language, which makes it possible to function properly regardless of the operating system of the machine where it is installed. It also enables future expansions of part of the application for mobile devices. Within that application, there are connections to R software, which is a computational statistical application enabling the use of tools like neural networks, least squares, and quantile regression, among others. The neural networks used are the feed-forward type; tests were carried out this model being the one with better results. Figure 4 shows the architecture of the statistical model.

The Power Curve Model (PCM) converts the meteorological forecasts into power to be generated by the wind turbine or the wind farm. Besides modelling the characteristic power curve of each wind turbine or wind farm, this model establishes a correspondence between wind speed and power as well as a correction in the production of the wind turbine or farm, due to the error associated with weather forecasts. Several approaches in the implementation of this model are under test, from which: instance-based learning (degree 0, 1 and 2 functions), polynomial function (3rd degree function), least squares, to adjust to a sigmoid function, and neural network. New approaches are to be experimented, with the purpose of improving the model's performance.

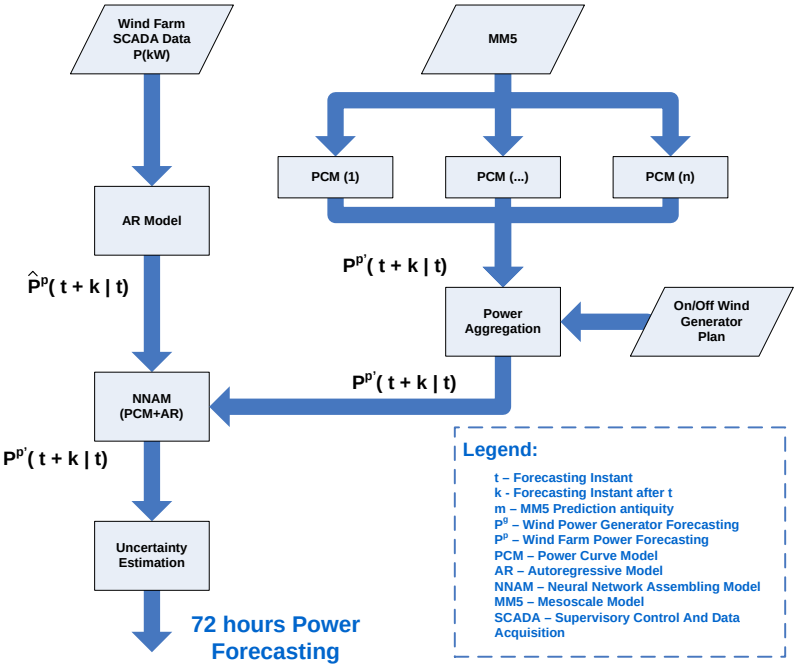


Fig. 4: Architecture of the statistical model

The PCM can be applied to each wind turbine or to the wind farm. In the case of application of the PCM to each wind turbine, individual predictions will be obtained. The input data, the wind forecasts, can be obtained for each wind turbine or for a reference point inside the wind farm area. Based on the results obtained in one of the test cases, the approach by wind turbine, with the predictions for each wind turbine, using neural networks, was found to be the best one. Figure 5 is a scatter with the predicted wind farm's power curve and the real power curve of a wind farm used as a test case.

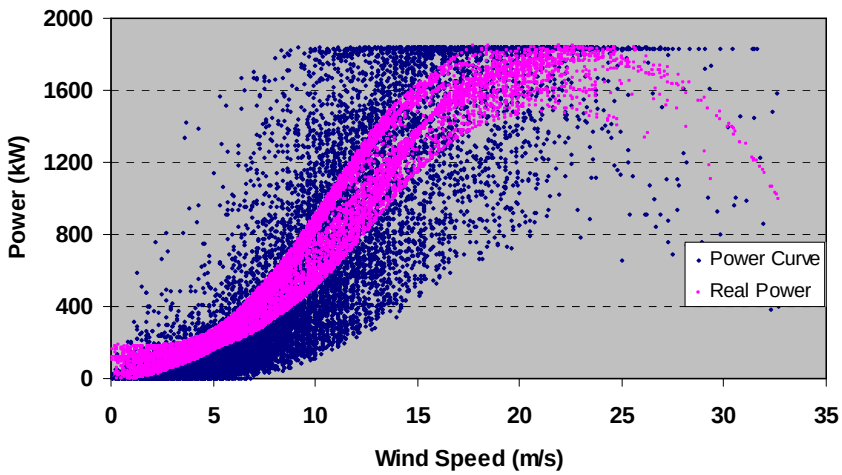
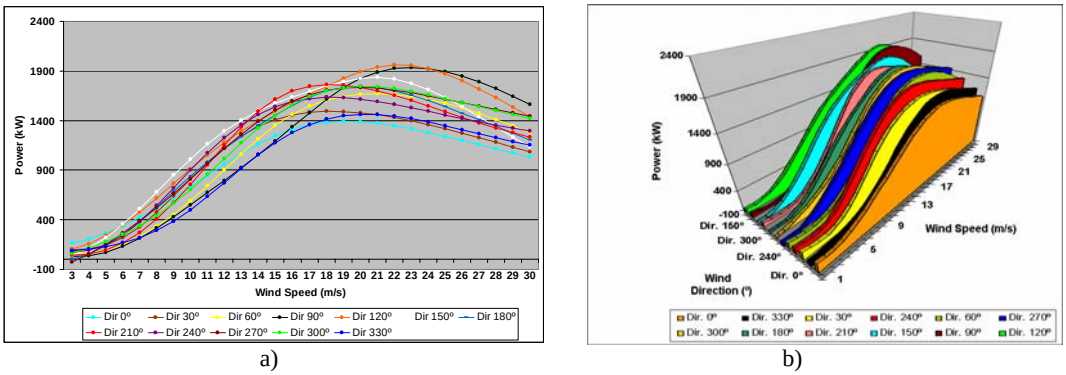


Fig. 5: Graphic with the Result of the PCM with Neural Network for a wind generator

In the case of applying the PCM to the wind farm, the input data are the wind forecasts for a reference point and the wind farm production. In the training of this model, there will be a filter to identify the *On/Off* plans of each wind turbine and wind farm. This information is of great importance in the model's functional mode, as it allows the identification of the available wind turbines, hence defining the wind farm production factor. Figures 6.a and 6.b illustrate the response of the PCM depending on the direction and wind speed for the case of Pinheiro Wind Farm, one of the wind farms being used to test the methodology. As shown, the model's response varies according to the sector of the wind direction, as a result of the local sector wise speed-ups.



a) b)

Fig. 6.a: Response of the PCM depending on the direction and wind speed.
b: Response of the PCM depending on the direction and wind speed (3D)

Given that this model is dependent on the meteorological forecasts and that there may be problems associated with their availability like communication problems, unavailability of data by the satellites, among others, there was the need to couple a prediction model that could complete the model described above. The autoregressive model (AR) was born in this context. This model presents better results in the first six hours of the time horizon as it is a dynamic model, depending only on the monitoring of the wind farm's production. In an AR model, the present value of the time series X_t is linearly expressed in terms of the past values of the series and the random noise at relative to instant t. The order of this process depends on the oldest value on which the regression is based. In an autoregressive process of p order, the model can be written

$$X_y = \delta + \phi_1 \cdot X_{t-1} + \phi_2 \cdot X_{t-2} + \ldots + \phi_p \cdot X_{t-p} + a_t \qquad (eq.1)$$

in which the several ϕ_1 are real constants and the a_t series presents normal distribution of independent values. Figure 7 illustrates the model's dynamic response. We can see that the model adapts to variations that correspond to the seasonal pattern of the resource.

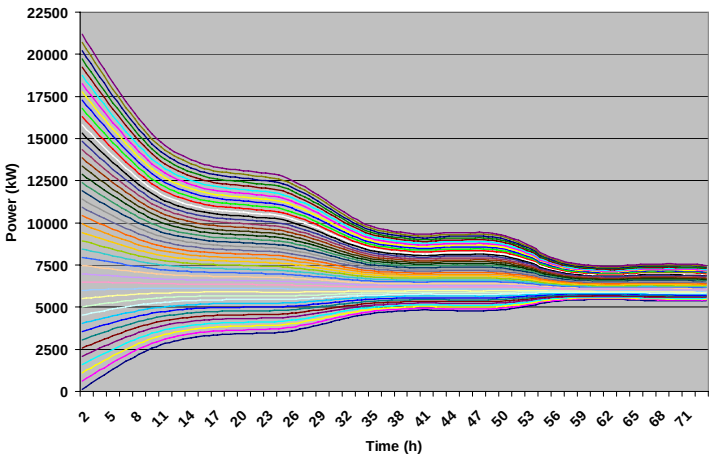


Fig. 7: AR model's dynamic response

For any wind velocity series, the model is purely autoregressive, of first or second order [10, 11]. If something abnormal occurs in the acquisition of the wind farm production data (for example, a flaw in the generation records), the system uses the latest generation values held in the database, changing the order of the model according to the flaw's instant.

The fusion of the models mentioned before is rather complex, given that the precise moment in which a model is worse than the other is unknown. It is also impossible to control the problems arising in the input information for which the fusion is applied to the whole forecasting horizon. The approach adopted was a neural network,

$$NNF_{t+i|t} = f\left(MA_{t+i|t}, MB_{t+i|(t+i-n)}, \varepsilon_{MA}, \varepsilon_{MB}\right) \qquad (eq.2)$$

where *MA* and *MB* correspond to the forecasts of the models that are meant to be fused, and ε_{MA} and ε_{MB} correspond to the error in the last 4 iterations by forecasting models *A* and *B*. Another fusion approach – LogAR - was tested

$$LogAR_{t+i|t} = (a \times e \times \ln(n)) \times MA_{t+i|t} + (1 - a \times e \times \ln(n)) \times MB_{t+i|(t+i-n)} \qquad (eq.3)$$

where *a* is estimated by the least squares method and $MAt+i|t$ is the autoregressive forecast, for prediction horizon $t+i|t$; *i* is the forecast time step and *t* is the instant of prediction (time horizon origin). $MBt+i|(t+i-n)$ is the Numerical Weather Prediction (NWP) for the forecast time step *i* made at time origin $t+i-n$ (received before the prediction instant $i-n<t$). This approach presented results that were less satisfactory than *NNF,* because it permits, in advantage to the *LogAR* model, a control of the forecasting models by measuring the error in each iteration.

As for the uncertainty, the calculation is based on an error prediction model. The purpose is to predict the ratio between the power measured by SCADA and the predicted power which, in effect, represents the error itself. For the purpose, the correlation between $P_{SCADA}/P_{predict}$ and other explanatory variables were analysed. The variable that better correlates with $P_{SCADA}/P_{predict}$ is the predicted value itself.

The construction of the model consists of a regression from quartiles to quartiles 25 %, 50 % and 75 %. The quartile regression makes it possible to analyse the impact of the explanatory variables over the dependent variable for a certain conditioned interval, estimating an interval for the prediction instead of an average expected value as with a least square-based regression.

3. Test cases

The development and implementation of a forecast tool can only be done if proper validation accompanies the process. To do that, the several modules are under test in three Portuguese wind farms: Pinheiro Wind Farm, in the northern part of the country, with 21.6 MW installed power, Cabeço da Rainha Wind Farm, located in central interior of the country, with 16.2 MW, and Chão Falcão Wind Farm, with 34.5 MW and also in central Portugal, but only about 25 km from the sea. Being sites with different climatologically regimes, the quality of the methodology validation exercise should be reinforced.

So far, tests were performed only on Pinheiro Wind Farm. The wind farm started operating the end of 2002. It is constituted of 12 wind turbines with a nominal power of 1.8 MW, rotor diameter of 70 m and hub height of 65 m. The wind conditions at the site are monitored by stations PORT071 and PORT214, both having wind vanes and anemometer installed at hub height. Figure 8 shows the layout of the wind farm and the location of the monitoring stations.

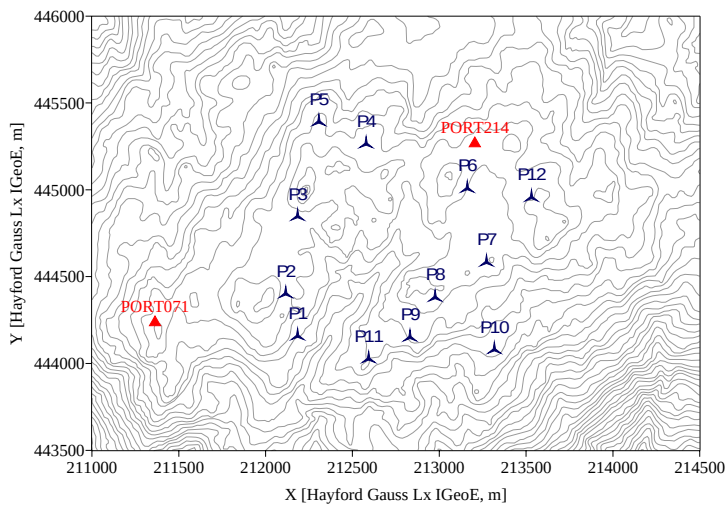


Fig. 8 – Pinheiro Wind Farm

Meteorological forecasts in several temporal windows, for the year 2005, were obtained using MM5 with ECMWF and NCEP boundary conditions. These forecasts were later compared with the actual measured figures from the monitoring stations. Figure 9 shows a scatter of such comparison with data collected at station PORT071.

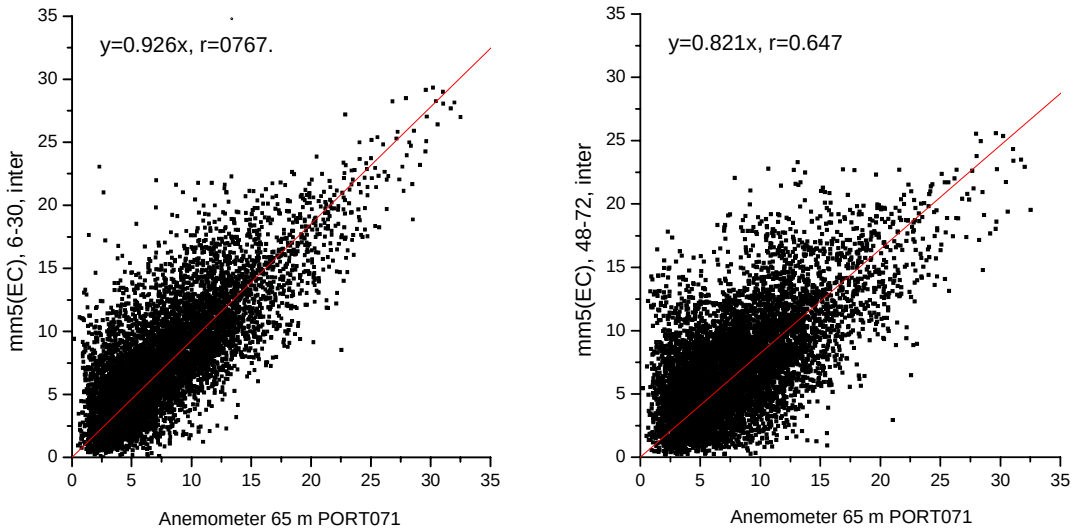


Fig. 9: Scatter plot of simulations *vs* observations in the Pinheiro site, PORT071 station, all 2005 forecasts: (i) in the time window 6-30 h, (ii) 48-72 h., MM5, with ECMWF boundary conditions, 6 km resolution.

In this particular case, a systematic under-estimation is perceived. It is noticed how dispersion grows the farther the time window is. Also in this particular case, the magnitude of the underestimation grows with the increase of the time window and its distance to the instant when forecasts are performed. This last feature implies that transfer functions between forecast predictions and local wind shall be established when using physical modules to create historic data or to forecast power.

The WFPC was reasonably well reproduced by the employed models, WAsP and Ventos®, being the average power BIAS of – 78 kW and – 323 kW. It is referred that in case of WAsP, two WFPC were employed, each one using as reference sites the sites of the two monitoring stations. Some corrections were also performed to the results, based on the results of local wind measurements carried out along many several years. In the case of Ventos® only station PORT071 was used. Nevertheless, the yearly BIAS was found to be only 1.5 % of the wind farm nominal power.

This quantification of the error still lays an open issue among the groups involved in the EPREV project. Not only the average BIAS but also its dispersion can light up the quality of the models and their limitation. Even so, the magnitude of BIAS using WFPCs built with both models is encouraging.

As for the use of statistical models, around 70 % of SCADA data of Pinheiro Wind Farm in 2005 was used for the training of the models, keeping the remaining 30 % for test and validation of the results. The obtained global results are considered to be satisfactory as the improvement in the global model, given the persistence, is equivalent to the reference projects in this area [12].

Comparison with the persistence forecast model is presented on figure 10, based on [13], the improvement being calculated using equation 4. As seen, the improvement of the quality on the prediction averaged for the 72 hours period is about 66 %.

$$improvement(k) = \frac{\left(RMSE_{Persist}(k) - RMSE_{EPREV}(k)\right)}{RMSE_{Persist}(k)} \qquad \textit{(eq. 4)}$$

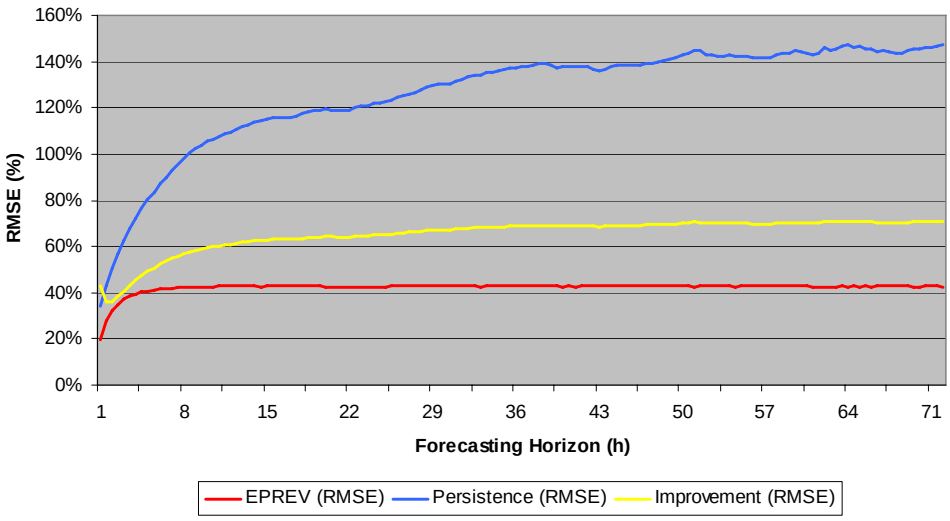


Fig. 10 –EPREV system improvement

4. Conclusions and future work

A wind power forecast model, consisting in several modules, is being built by a consortium of Portuguese universities and research institutes, financed by a group of wind farm investors. Work is being done in the meteorological forecast up to 72 hours using mesoscale models and different boundary conditions. Statistical models are being designed with the same purpose, using forecasts and training from historical data. Also physical models are being tested in order to convert meteorological forecasts into wind power forecasts.

Three Portuguese wind farms are used as test cases to validate the developed tools. So far the model has been applied to two wind farms but only in one case the tests are complete. Despite the need of further developments, namely for the combined use of the different models and for the uncertainty forecasting, among others, the results obtained with the models used individually are very promising and in line with the state of the art.

The models shall now be implemented and tested in other wind farms. The definition of transfer functions between the meteorological forecasts and the corresponding meteorological conditions, at the reference sites for which the WFPCs, are to be built and applied to train the statistical models for wind farms without historical data. Guidelines for the use of the forecasting tool as a whole, combining different approaches and results, and for the appropriate choice in each wind farm or zone, are also part of the work to be done. This will be a significant contribution for the use of the tool in operational mode, the step that shell follow the conclusion of the present project.

Acknowledgments

The authors sincerely acknowledge the companies that form the EPREV consortium for the support given to this project, for their trust and permanent collaboration.

References

[1] Dudhia, J., Gill, D., Mannig, K., Wang, W., Bruyere, C., Kelly, S., Lackley, K., *PSU/NCAR Mesoscale Modelling System Tutorial Class Notes and User's Guide: MM5 Modelling System Version 3*, 2003, NCAR Mesoscale and Microscale Meteorology Division, 362 pp.

[2] Grell, G.A., J. Dudhia, and D.R. Stauffer, *A description of the fifth-generation Penn State/NCAR mesoscale model (MM5)*, 1994, NCAR technical Note, NCAR/Tn-398+STR, 117 pp.

[3] Tomé, R., *Previsão do tempo com modelos de mesoscala: Casos de estudo com o modelo MM5 nos Açores, Relatório de Estágio*, 2004, FCUL.

[4] *Wind Atlas Application Program*, Risø National Laboratory, Denmark

[5] Troen, I., Peterson, E. L., *European Wind Atlas*, Risø National Laboratory, Denmark 1998

[6] Bowen, A. J. and Mortensen, N. G., *WAsP prediction errors due to site orography*, Risø National Laboratory, December 2004

[7] Castro, F.A., *Métodos Numéricos para a Simulação de Escoamentos Atmosféricos sobre Topografia Complexa*, PhD thesis, Universidade do Porto, 1997

[8] Castro, F.A., Palma, J.M.L.M. e Silva Lopes, A.: *Simulation of the Askervein flow. Part 1: Reynolds averaged Navier-Stokes equations (k-ε turbulence model)*, 2003, Boundary-Layer Meteorology, 107:501-530.

[9] Lopes da Costa, J.C., Castro, F.A., Palma, J.M.L.M. e Stuart, P.: *Computer Simulation of Atmospheric Flows Over Real Forests for Wind Energy Resource Evaluation*, 2006, Journal of Wind Engineering and Industrial Aerodynamics, 94:603-620

[10] Box, GEP, Jenkins, GM, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970

[11] Jim Elkins, NOAA, 1988, *Engineering Statistics Handbook*, NIST, 2003

[12] Kariniotakis G, *Short-term Forecasting of Wind Power - Uncertainty & Risk*, EES-UETP Course, 2006, Madrid

[13] Madsen H, Kariniotakis G, Nielsen H, *A Protocol for Standardizing the Performance Evaluation of Short-Term Wind Power Prediction Models*, 2004