

Wind Power Forecasting using Fuzzy Neural Networks Enhanced with On-line Prediction Risk Assessment.

P. Pinson and G. N. Kariniotakis, *Member IEEE*.

Abstract—The paper presents an advanced wind forecasting system that uses on-line SCADA measurements, as well as numerical weather predictions (NWP) as input, to predict the power production of wind parks 48 hours ahead. The prediction system integrates models based on adaptive fuzzy-neural networks configured either for short-term (1-10 hours) or long-term (1-48 hours) forecasting. The paper presents detailed one-year evaluation results of the models on the case study of Ireland, where the output of several wind farms is predicted using HIRLAM meteorological forecasts as input. A method for the on-line estimation of confidence intervals of the forecasts is developed together with an appropriate index for assessing on-line the risk due to the inaccuracy of the numerical weather predictions.

Index Terms—Wind power, short-term forecasting, numerical weather predictions, on-line software, adaptive fuzzy-neural networks, confidence intervals, prediction risk.

I. INTRODUCTION

NOWADAYS, wind park installations in Europe exceed 23 GW, while the motivated by the Kyoto Protocol targets of the E.U. for 12% energy demand covered by renewables by year 2010, are translated to 21% electricity generation by renewables. To achieve these targets, wind power in the Member States should arise up to 45-60 GW. Such a large-scale integration of wind power emerges the development of appropriate tools to assist the wind farm operators on their management task.

Short-term forecasts of the wind farms production, up to 48 hours ahead, are necessary for a secure and economic large-scale wind power integration. Wind power prediction tools are useful for end-users such as Independent Power Producers, Transmission and Distribution System Operators (TSO/DSO), Energy Service Providers (ESP) etc. In a liberalised electricity market environment, such tools enhance the competitiveness of wind power, since they reduce the penalties resulting from the wind resource intermittence. Reduced operational and financial risk for the wind farm developers is a motivating factor for undertaking investments on wind farms. Hence, accurate wind power prediction tools contribute indirectly to the increase of the installed wind power capacity.

Wind power forecasting is a far from trivial problem. Wind speed is a non-stationary process both in the mean and variance. Wind power is nonlinear w.r.t. speed with a major difficulty in the area of cut-off speed, where prediction intervals can extend from maximum to zero wind power.

Among the difficulties, one should add the error of numerical weather forecasts, which are often used as input to the models. Often, no adequate information is available online by a data acquisition system (SCADA) to assess the actual operational status of the wind farm (i.e. how many turbines are in operation). The available on-line data can be detailed (i.e. power, speed of each wind turbine) or not (i.e. only total power available). In some situations there is complete lack of data and information from neighbor wind farms has to be assessed.

Research on wind power forecasting is actively pursued by several research centres in Europe. Actually there are two main state-of-the-art approaches; one based on physical or deterministic modelling and a second one based on statistical or timeseries modelling.

The “physical” approach for wind power forecasting is based on a detailed description of the wind park site (orography, roughness, obstacles), the wind turbines (hub height, power curve, thrust curve) and the layout of the wind plant. In [10-12], wind power forecasting platforms based on physical methods are described.

The main input is numerical weather predictions (NWP). Model output statistics are developed to account for systematic errors. Weather predictions are however updated only a limited number of times per day by meteorological services. For this reason, the performance of these models is often satisfactory for rather longer (>6 hours ahead) than short-term horizons.

The alternative “time series”, or statistical, approach includes typical linear models (ARMA, ARX etc) and non-linear ones (i.e. neural networks, conditional parametric models, etc). These models aim to predict the future by “capturing” temporal and spatial dependencies in the data [15]. The input to these models can be on-line SCADA data and numerical weather predictions (NWP). For look-ahead times more than ~10 hours (mentioned hereafter as “long-term”), NWPs are indispensable for an acceptable performance, since they represent weather dynamics that cannot be modelled using only recent on-line data. For shorter horizons, up to ~10 hours ahead (mentioned hereafter as “short-term”), time series models can be based exclusively on recent measurements; however even in this case, NWPs as explanatory input improves results. It is noted that the threshold of 10 hours is mentioned as an example rather than a rule, since it depends on the characteristics of a specific wind profile.

The model presented in this paper belongs to the time series approach. In previous work of the authors, linear autoregressive models, radial basis functions, wavelet networks, feed forward and recurrent neural networks [6], [7], and finally adaptive fuzzy-neural network models were

Part of this work was supported by the European Commission in the frame of the contracts N° ERK5-CT1999-00019 and ENK5-CT-2002-00665.

G. N. Kariniotakis and P. Pinson are with the Center of Energy Studies of Ecole des Mines de Paris (e-mail: kariniotakis@cenerg.cma.fr, pinson@cenerg.cma.fr).

compared for the task of short-term prediction. Fuzzy neural networks, originally used here for wind forecasting, were found to outperform the other approaches in both tasks of short-term and long-term wind prediction [9].

II. THE PREDICTION MODULE.

Adaptive fuzzy-neural networks (F-NN) are applied here for both short-term and long-term wind power prediction.

The adaptivity property stands for the capacity of the model to fine-tune its parameters during on-line operation. This is an important requirement for a non-stationary process like wind speed or power. Adaptivity of the model compensates changes in the environment of the application that may happen during the lifetime of a wind farm. Such changes can be changes in the number of wind turbines (extension of the wind farm, maintenance or availability of the machines that is usually not available through SCADA), in the performance of the wind turbines due to aging, changes in the surrounding of the wind park (i.e. vegetation), or changes in the configuration of the model used to produce the NWP.

The core F-NN model is generic and can be trained on appropriate input depending on the final use, which can be either short-term or long-term prediction.

A. Short-term models.

Short-term models receive historic values of wind power as input, as well as explanatory data, such as wind speed and direction, to predict wind power. The general form of a simple model with input only past values of power is:

$$\hat{P}(t+1) = f(P(t), P(t-1), \dots, P(t-m))$$

The generic fuzzy-neural function $f(\cdot)$ is described in Section III. Multi-step ahead forecasts are generated using the model in an iterative way. I.e., in order to produce a forecast for $t+2$, the forecast for $t+1$ is fed back as input to the model. This approach presents the drawback that does not permit to iterate explanatory input, since no forecasts can be available for such quantities. To handle this problem, models using the look-ahead time k as an input variable can be considered.

An alternative approach is to develop multi-output models, or to tune a different model for each time-step. The implementation of this approach is complex and requires high development effort, which can be prohibitive in case of a large number of wind farms.

The short-term models based on fuzzy-neural networks can be useful for horizons up to ~10 hours. They are found to outperform Persistence up to 20% according to the time-step [7-9]. Persistence is a simple approach used as reference to evaluate the performance of advanced models. It assumes that the "wind in the future will be the same as the wind now".

Short-term predictions are adequate for small applications, for which NWP are not available, e.g. in the case of islands. In larger systems, timeseries models based on meteorological information, as the one presented below, outperform short-term models (improvement up to 40% w.r.t. persistence for horizons up to 10 hours).

B. Models based on meteorological information.

For "long-term" horizons up to 24-48 hours ahead, it is necessary to include numerical weather predictions (NWP) as explanatory input to the model in order to have an acceptable performance. NWPs include usually wind speed, direction and temperature at 10 m, as well as at several altitude levels defined by atmospheric pressure. NWPs can be provided for the geographical coordinates of the wind farm or for a grid of four points surrounding the farm. In the second case, the spatial resolution of the NWP model is of primary importance. Meteorological models with high resolution are often more accurate but require high computation time to produce forecasts, and as a consequence, they do not update frequently their output (i.e. 1-4 times per day). In contrast, forecasts from low-resolution NWP models are more frequently available.

The developed forecasting tool is able to operate with input from different NWP systems. In the frame of this work it was tested and gave satisfactory results with input from the SKIRON system for the case of Crete, and also from HIRLAM for the case of Ireland. SKIRON forecasts were provided for a grid of 15x15 km (System B in Fig. 1), while HIRLAM predictions were provided at the level of the wind farm (System A in Fig. 1).

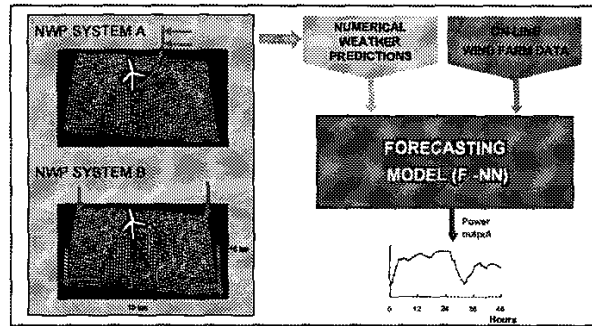


Fig. 1. General scheme of the "long-term" prediction model with examples of two configurations of NWP systems used as input (SKIRON, HIRLAM).

The developed model receives on-line data as well as NWPs as input to predict the wind farms production for the next 48 hours. These forecasts are updated every hour based on the most recent wind power measurements. Wind power data are necessary for the on-line updating procedure, independently if they are used or not as input variables to the model. The updating procedure permits mainly a good performance of the model for the first hours (i.e. 1-6 hours) of the considered horizon. Model configurations that do not update their forecasts based on recent wind power data were found to perform worse than persistence in look-ahead times up to 6 hours ahead. Finally, the consideration of on-line information, other than wind power (i.e. wind speed or direction), was not found to contribute in the accuracy of the results. The general scheme of the model is shown in Fig. 1.

The aim of the prediction model is to capture the relations between input (meteorological information, on-line data) and output (future total wind park power). Such mapping includes the following implicit relations:

- Temporal correlations between past and future data of the process (autoregressive aspect of the model).
- Conversion of wind speed (meteorological predictions) from the height or the atmospheric level they are given to the hub height of the wind turbines.
- Spatial projection of the meteorological wind speed forecasts from the NWP grid points (e.g. 15x15 km) to the level of the wind farm ("downscaling").
- Correction of the wind park output for factors affecting the total production (i.e. array effects, effect of wind direction etc).

The advantage of a model such as the fuzzy neural network model, compared to models based on the "physical" approach, is that it permits to avoid all the above intermediate modeling steps. Moreover, its adaptive mode can compensate situations like the ones explained in the previous Section.

III. MODEL DEVELOPMENT AND GENERALIZATION.

A. General description of the fuzzy-neural network model.

The fuzzy model can be expressed in the form of rules of the type:

"IF \underline{x} is A THEN y is B "

where \underline{x} , y are linguistic variables and A , B are fuzzy sets. In the case of time-series prediction rules may have the form:

R : IF x_1 is A_1 , and ..., and x_n is A_n THEN $y = g(x_1, \dots, x_n)$

where:

x_1, \dots, x_n are real-valued variables representing input variables of the system defined in the universes of discourse X_1, \dots, X_n respectively.

A_1, \dots, A_n are fuzzy sets.

y is variable of the consequence whose value is inferred. In the specific problem it represents future wind power $(\hat{P}(t+1), \hat{P}(t+2), \dots)$.

$g(\cdot)$ is a function that implies the value of y when x_1, \dots, x_n satisfy the premise. The function $g(\cdot)$ in the consequent part of the rules may be a linear or a non-linear one or even a constant. In the case of a linear function the fuzzy rule-base takes the form:

R^1 : IF x_1 is A_1^1, \dots , and x_n is A_n^1 THEN $y^1 = p_0^1 + p_1^1 x_1 + \dots + p_n^1 x_n$
 \vdots
 R^m : IF x_1 is A_1^m, \dots , and x_n is A_n^m THEN $y^m = p_0^m + p_1^m x_1 + \dots + p_n^m x_n$

Each rule gives an estimation of the output y_i according to the conditions defined by the fuzzy sets in the premises. In the context of timeseries prediction, each variable x_i in the premise corresponds to a past value of the process (i.e. power: $P(t)$, $P(t-1)$, ...) or past values of explanatory input (i.e. wind speed: $WS(t)$, $WS(t-1)$, ...) or meteorological forecasts ($WS_m(t+1)$, $WS_m(t+2)$, ...).

A linear function in the consequence is indeed an ARX (autoregressive with exogenous variables) model. It is clear that with the above definitions, the rule-base consists of an ensemble of "local" models. Local modeling is a desired

property of the model, especially in the case of a non-stationary process such as wind generation.

Fuzzy sets in the premises are modeled here using Gaussian functions:

$$\mu_{A_j^i}(x_j) = \exp\left(-\frac{(x_j - a_j^i)^2}{b_j^i}\right)$$

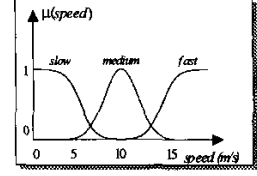


Fig. 2. Representation of fuzzy wind speeds. "Speed" is a linguistic variable with three terms "slow", "medium", and "fast" represented as fuzzy sets with the membership functions shown in the Figure.

In the case of a linear function in the consequence, the model may be written analytically as following:

$$\hat{y} = \frac{\sum_{i=1}^m \left(p_0^i + \sum_{j=1}^n p_j^i x_j \right) \prod_{j=1}^n \mu_{A_j^i}(x_j)}{\sum_{i=1}^m \prod_{j=1}^n \mu_{A_j^i}(x_j)} \triangleq \sum_{i=1}^m w^i \hat{y}^i$$

B. Learning and Generalization.

Model building is characterized by two phases: (i) optimization of the model architecture and (ii) tuning of the model internal parameters (learning).

These two phases are driven by the requirement for good "generalization". Generalization is the capacity of the model to perform well when it predicts new data (data not used during the two phases of model development). It is a primary requirement for the on-line use of a model.

The tuning of the model parameters is performed taking into account [6]:

- *Learning rules* based on stochastic gradient for tuning the parameters a , b , p of the model.
- Learning rules are appropriately developed to minimize simultaneously prediction error and the Information content of the model (max entropy). This acts as a *self-regularization* process that permits to avoid overfitting of the data.
- *Simulated annealing* is performed for controlling the evolution of the learning process through appropriate adaptation of the learning rate.
- *Early-stopping* is applied to the learning process is early-to avoid overfitting.
- *Cross-validation* is applied to terminate learning. For this purpose, a subset of the data (validation set) is reserved.
- The cross-validation criterion is expressed as a weighted function of the performance of the model over the whole prediction horizon. By this way, generalization is optimized for multi-step ahead prediction.

The above process permits to tune optimally a model with a specific architecture. The architecture of a model is defined by the types of input variables and the number of fuzzy sets associated to each one. For each type of measured data it is needed to decide the number of past values to be used as

input. When NWP's are considered ("past values" have no sense), it is necessary to select the relevant information (forecasts of wind speed, direction, etc) for the model.

This selection procedure, which is also similar to other types of models like neural networks, is a time consuming one due to the infinite number of combinations that can be tested. Often it is performed by trial-and-error, where several candidate configurations are tested. It is noted that the evaluation of each candidate model requires carrying out the above-described learning process.

In this work, the trial-and-error has been replaced by a fully automated process for model architecture optimization. The constrained nonlinear simplex ("Complex") optimization algorithm is used for this purpose. The algorithm has been modified for handling both discrete and continuous decision variables. The optimization process is based on the evaluation of the surface of the generalization function (defined as the performance of a model on the validation set) using a complex of points. Each point corresponds to a candidate model. The computational cost is high due to the necessity of the algorithm to tune each candidate model. However, in global, the automatic nature of the process permits to save considerable engineering time compared to the trial-and-error.

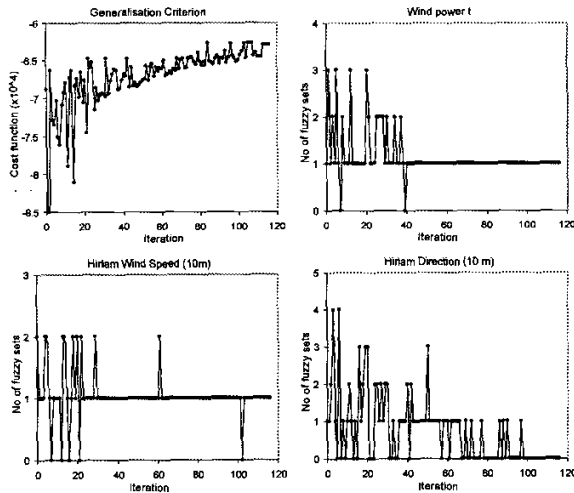


Fig. 3. Evolution of the algorithm for the model architecture optimization.

An alternative genetic algorithm approach did not present any advantages with respect to the simpler "Complex" algorithm. Genetic algorithms appeared to be less parsimonious w.r.t the number of models they need to test in order to converge compared to the Complex algorithm

Each decision variable in Complex represents the number of fuzzy sets associated to each type of input data. In the special case, when the algorithm converges to zero-number of fuzzy sets for a specific type of data, then this input is excluded from the model as non-significant. By this way the algorithm performs input selection. When the number of fuzzy sets is converging to one, then the variable does not participate in the premises, but appears only in the function of the

consequent part. Parsimony in the selection of input is critical to avoid overfitting from overparametrized models.

Fig. 3 shows an example of a run of the Complex algorithm. 115 candidate models are totally examined. The input selection is performed among past values of wind power and HIRLAM wind speed, direction and temperature forecasts. The upper left figure shows the evolution of the Complex objective function. Each point in the figure corresponds to the "generalization" performance of a candidate model on the validation set. The rest of Figures show the number of fuzzy sets associated by the algorithm to each input type of data.

When the number of fuzzy sets for all variables is either one or zero then a single "rule" is obtained. The premise has no significance and the model corresponds to a simple linear function of the input variables. This limit case corresponds to the ARX class of models. Consequently, the optimization process can indeed exclude the use of a nonlinear fuzzy model and lead to a classical linear one. In this way, a selection between linear and nonlinear models is performed.

IV. UNCERTAINTY OF THE WIND POWER PREDICTIONS.

Spot predictions of the wind farm production for the next 48 hours is a primary requirement for end-users. However, for an optimal management of the wind power production it is necessary to provide end-users with appropriate tools for on-line assessment of the associated prediction risk.

Confidence intervals are a response to that need since they give an estimation of the error linked to power predictions. Given that confidence intervals are estimations of the uncertainty based on the past performance of the model, the second objective of this work is to propose additional tools able to assess the prediction risk based on the most recent information available, i.e. the one that can be extracted from numerical weather predictions.

Typical confidence interval methods, developed for models like neural networks [1]-[5], are based on the assumption that the prediction errors follow a Gaussian distribution. This however is often not the case for wind power prediction where error distributions appear some skewness, while the confidence intervals are not symmetric around the spot prediction due to the bounding of the wind park power. On the other hand, the level of predicted wind speed introduces some nonlinearity to the estimation of the intervals; i.e. at the cut-off speed, the lower power interval may switch to zero due to the cut-off effect. The limits introduced by the wind farm power curve (min, max power) are taken into account by the method proposed in [13], which is based on modelling errors using a β -distribution, the parameters of which, have to be estimated by a post-processing algorithm. This approach however is applicable only to "physical" models since such models estimate power using a wind park power curve, which is not the case for statistical models as the ones considered here.

In [11], [12] wind speed errors are classified as a function of look ahead time and then they are transformed to power prediction errors using the wind turbine power curve vs wind speed. This method however is also limited for application to physical models rather than statistical ones since it requires

local wind speed predictions (at the level of the wind farm), while it does not provide uncertainty as a function of a specified confidence level. On the other hand, this method requires wind speed measurements, which, in general, might not be available on-line.

The methodology proposed here for the estimation of confidence intervals is generic and can be applied to both physical and statistical wind power forecasting models. This is due to the fact that it is based on the past wind power data, which are often available on-line by a SCADA system, as well as on the numerical weather predictions, which are the basic input to all models.

A. Error margin.

Let us define the prediction error for the look-ahead time h as following:

$$e_h = p_h^{\text{predicted}} - p_h^{\text{measured}}$$

This error can vary between -100% and 100% of the nominal wind park power. For a non-bounded model it can take even values outside that range. The possible error of the prediction model, defined as “error margin”, depends on the level of measured wind power. Fig. 4 represents graphically the error margin as a function of the wind park characteristic curve. For wind speeds below cut-in speeds, the error margin is maximal since the model can predict a production up to the nominal wind park power. On the contrary, for higher wind speeds the model will show a negative error margin, i.e. the generated power is likely to be greater than the one proposed by the prediction model. Close to the cut-off wind speed the uncertainty is again maximal since the model can switch from a negative error margin to a positive one, or the inverse.

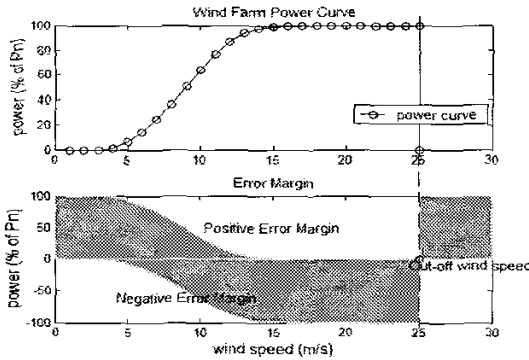


Fig. 4. The error margin as a function of the wind park power curve.

B. Classification of prediction errors.

Before computing standard deviations or confidence intervals, an important point is to collect the prediction errors the model made in the past. For that purpose, a sample size has to be defined. Based on this, the prediction errors are stored in samples and updated hour by hour. There are several samples because we consider separately 1-hour ahead, 2-hour ahead, and so on up to 48-hour ahead prediction errors. It is clear from Fig. 5 that error distributions, depending on the

prediction horizon, are significantly different. Obviously, the uncertainty for these various horizons must be different.

The collected errors are the most recent ones at a given time: when the actual measured wind power is known, that value is compared with all the past predictions made for that time (from 48 hours to 1 hour before). Hence, we will use the most recent information to compute the prediction uncertainty or confidence intervals.

In case that on-line data are not available (i.e. the wind farm is not connected to a SCADA system), an alternative is to carry out an off-line study to design the intervals based on the prediction errors of the model.

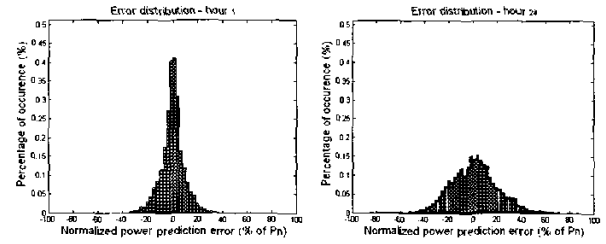


Fig. 5. Prediction error distribution varies depending on the prediction horizon (left picture: 1-hour ahead prediction error distribution, right picture: 24-hour ahead prediction error distribution).

The power prediction errors depend on the predicted wind speed [10-12]. That link is due to the slope of the power curve. We can translate this argument in terms of predicted power, i.e. the prediction error is dependent on the predicted wind power. The uncertainty for low and high power output is considerably smaller than for medium power output. This is due to the high slope of the power curve between cut-in and rated speed. To account for this, the wind power curve is divided into three ranges of power: low, medium and high. The prediction errors are classified then as a function of these three ranges. Hence, the confidence interval estimation is carried out using the error sample corresponding to the power class of the predicted power.

C. Confidence interval estimation

Here is a formal definition of confidence intervals: the interval computed from the sample data which, were the study repeated multiple times, would contain the true effect CL% of the time, CL being the confidence level.

1) The simple method

The simplest approach one can use to estimate confidence intervals for the prediction error is to assume that this error follows a Gaussian distribution and is centred on the prediction. This is for instance the basis of the Delta method, as described in [4]. Then, intervals are given by

$$P_h \pm z_{0,025} \sigma_h$$

where $z_{0,025}$ is the critical point of the standard Normal distribution, P_h the forecasted wind generation and σ_h the standard deviation of the error sample at the prediction horizon h .

2) The Resampling approach

A given set of observations (the sample) is a part of a whole population and can be seen as representative. The aim of methods like Resampling is to have a better idea of the population distribution by going through the sample hundreds or thousands of times. This evaluation of the population distribution can serve to estimate a mean, a variance, etc. No assumption is made concerning the distribution.

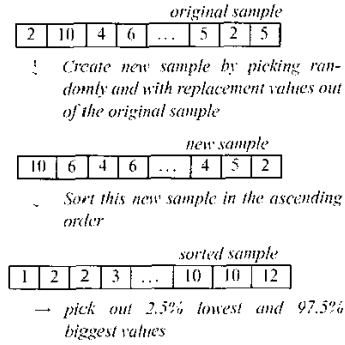


Fig. 6. One step of the Resampling process.

Let us consider a sample containing N_{sample} observations of a mean m_{sample} for a given phenomenon. The procedure to compute from this sample the 95% confidence interval for that mean is as following: One has to pick randomly and with replacement N_{sample} values out of the original sample to create a new sample. Then, one sorts the new sample in the ascending order, and determinates what are the 2.5% lowest and 97.5% biggest value of that set.

This procedure represents one step of the Resampling process (Fig. 6). Indeed, these three actions are to be repeated a large number of times to re-create the population again and again. One always gets a new sample that is close to the original one, and the whole population distribution is not really simulated by this way. But by calculating the mean of respectively the 2.5% lowest and the 97.5% biggest value of these randomly created samples, good estimates of the confidence limits for the m_{sample} value can be computed [1], [14], [17].

Such a method can be applied directly to the wind power forecasting problem by considering separately the 48 error samples corresponding to the various look-ahead times. These samples of errors are treated one after the other using the procedure described above, assuming that the prediction error the model makes is the mean of a distribution; and that we would like to compute 95% confidence intervals for that mean.

V. INFLUENCE OF WEATHER STABILITY

Not only the power prediction model can be responsible for bad forecasts, but also the numerical weather prediction system (i.e. due to low weather stability). Indeed, an unstable atmospheric situation can lead to very poor numerical weather predictions and thus to worthless wind energy ones. At the opposite, when the atmospheric situation is very stable, one

can expect more accurate wind power predictions from the model.

A. Assessment of weather stability.

Meteorological Centres are able to produce different scenarios by perturbing the initial conditions of the forecasting model or by using different NWP models. These scenarios are called ensemble forecasts and permit to evaluate the stability of the weather regime [17]. However, for wind power applications only one forecast for the next 48 hours is often made available (or purchased) at a given time - for instance, HIRLAM gives a unique 48-hour ahead forecast every 6 hours. However, for a given hour, several predictions can be available produced at different time origins in the past (-6, -12, -18... hours). In a stable and well-predicted weather situation it is expected that these predictions will not differ significantly. Weather stability can be assessed by comparing all the available forecasts for the considered period.

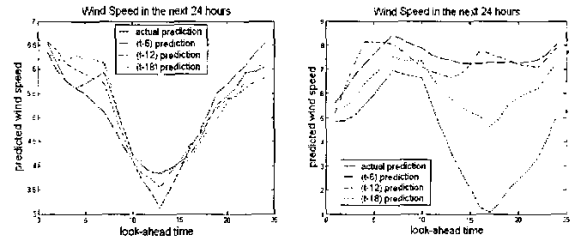


Fig. 7. Stable (left picture) and unstable (right picture) weather situations.

Because we want to have a general evaluation of that stability, 4 predictions of various ages (0, 6, 12 & 18 hours) for the following 24 hours are compared. Fig. 7 gives the examples of a stable atmospheric situation (left picture, the forecasts are quite close) and of an unstable one (right picture, spread forecasts).

B. Development of a norm to assess the weather stability

There are several possibilities to measure the spread of the various weather forecasts. In [10] the standard deviation of the forecasts for each time-step is mentioned as an example.

Our aim here is to evaluate the global atmospheric situation. This is why a unique representative index is defined for the following N_h hours instead of indexes for every look-ahead time.

Thus, in order to calculate the distance between two forecasts, we propose a 2-norm between the N_h -valued vectors containing the predicted wind speed for the N_h following hours. Let us denote the N_f available wind speed forecasts (NWPs) vectors by ws_i with $i=1, \dots, N_f$, ws_1 being the most recent forecast and ws_{N_f} the oldest one. The distance between the i^{th} and the j^{th} vectors is given by:

$$d(ws_i, ws_j) = \left(\frac{1}{N_h} \sum_{k=1}^{N_h} (ws_i[k] - ws_j[k])^2 \right)^{\frac{1}{2}}$$

Then, the index I_n , called hereafter "meteo-risk" MR-Index, is defined to measure the spread of the weather forecasts at a given time:

$$I_i \triangleq \sum_{k=1}^{N_f-1} p_k \cdot d(ws_k, ws_{k+1}),$$

with $p_k, k=1, \dots, N_f-1$ appropriate weights defines so that:

$$\begin{cases} p_k > p_{k+1}, & k=1, \dots, N_f-1, \\ \sum_{k=1}^{N_f} p_k = 1. \end{cases}$$

The use of weights p_k permits to give more importance to the recent information we get from the weather predictions.

C. Relation between weather stability and wind power prediction error.

In this work, the period for the calculation of the meteorological index is a 24-hour period ($N_h=24$). Since HIRLAM forecasts are provided every 6 hours, there are four wind speed predictions covering the period ($N_f=4$). However, the same methodology can be applied to seven available HIRLAM forecasts on a 6-hour period for instance.

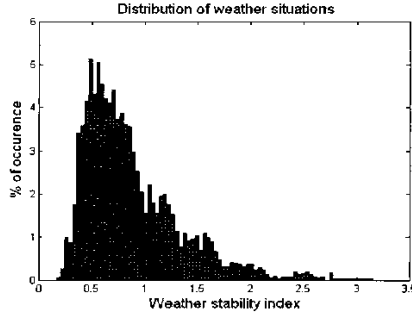


Fig. 8. The distribution of weather situations as expressed by the MR-Index.

Regarding the distribution of weather situations (Fig. 8), one can notice that most of the times the weather regime is quite stable (low MR-Index values) and that there are only few occurrences of really spread forecasts.

For the case study described in the Next Section, we collect wind power prediction errors as obtained by a fuzzy-NN model configuration for a period covering 5000-hours. For the same period the MR-Index is estimated. The scatter plot of the left part of Fig. 9 shows the relation between prediction error and MR-Index. By binning the data, calculating the average error for each bin, and comparing these averages to the global prediction error of the model (13.78% of nominal power), the representative points in Fig. 9 (right picture) are obtained. The prediction error increases linearly with the MR-Index: the tighter the HIRLAM predictions are, the more accurate is the fuzzy-NN model. A linear fitting gives the solid curve.

D. Fine-tuning of the confidence intervals

The linear model drawn above is used here to define a scale factor for confidence intervals depending on the value of the MR-Index. This scale factor is then applied to adapt the interval width in the following N_h hours. For instance, when the meteorological index equals 0.5, the size of the confidence intervals for the following 24 hours is diminished by more

than 20%. Moreover, our aim is only to narrow the intervals; so, only scale factors smaller than 1 will be considered. It can be seen from Fig. 8 that most of the times (around 70%) the weather situation permits to decrease the interval size.

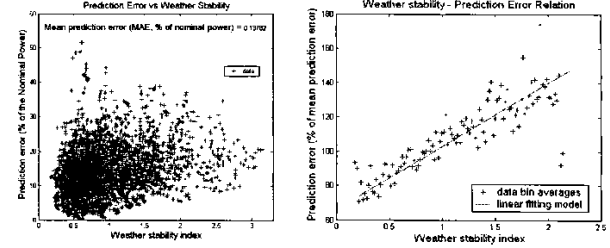


Fig. 9. Prediction errors vs. MR-Index over a 5000-hour data set (left picture) and the linear relation between the prediction error and the MR-Index (right picture).

VI. RESULTS

Results are presented for three real wind farms in Ireland (WF-A to WF-C) with a total installed power of a few tenths of MW. The available time series cover a period of 12600 hours from which 6600 were used for training (learning set), 1000 for cross-validation and 5000 (210 days) for testing the performance of the model. The results presented here are on the testing set. Concerning the computation of confidence intervals, 12 days of prediction errors are stored in the samples. The desired confidence level is set to 95%. This is indeed a quite strict confidence level since usually in practice confidence of around 85% is requested for the problem of wind prediction.

Figure 10 shows the improvement w.r.t. Persistence as obtained for each one of the three wind farms and as a function of the prediction horizon. In all cases the advanced model outperforms persistence up to 45%. Figure 11 depicts an episode with the wind power predictions for the next 42 hours compared to the real values. The confidence intervals as estimated by the two approaches are shown. The benefit from the application of the resampling based approach is evident especially for the first 24 hours.

Finally, for the 3 wind farms, the MR-Index has been applied to narrow the intervals for the first 24 hours ahead. An example of the obtained decrease is depicted in Fig. 12. Based on the 5000 hours testing set, the interval width is globally reduced by more than 10% for horizons up to 24-hours ahead (Tab. I.). The corresponding real (observed) confidence of the intervals is not really affected. For instance, the 95% Resampling confidence intervals for WF-A have their size reduced by 10.27% in average, while the real confidence is lowered by 1.11%.

Wind farm	Average width reduction of the confidence intervals (%)	Confidence loss (%)
WF-A	10.27	1.11
WF-B	10.80	1.34
WF-C	9.06	1.00

Tab. I. Width reduction and the resulting confidence loss with the dynamic meteorological-index-dependant scaling for three wind farms in Ireland (5000 hours of simulation).

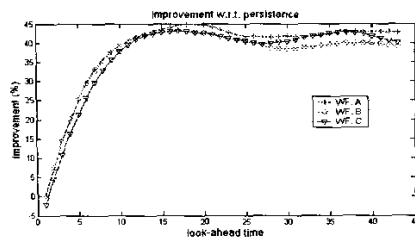


Fig. 10. Performance of the model: improvement w.r.t. Persistence for 3 wind farms in Ireland.

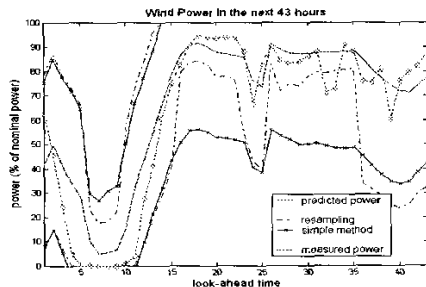


Fig. 11. Wind power prediction with simple and Resampling confidence intervals (WF-A).

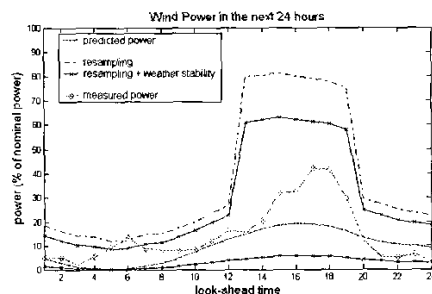


Fig. 12. Wind power prediction with Resampling confidence intervals and the ones after taking into account the weather stability (WF-A).

VII. CONCLUSIONS

A fuzzy-neural network model for wind power prediction enhanced with methods for uncertainty estimation is developed. The proposed confidence intervals are conditioned by the past errors that the model made depending on the prediction horizon and the power class, as well as the current weather stability described by a new meteorological risk index. Results from the prediction of three wind farms in Ireland show a considerable improvement w.r.t. Persistence. The developed methods have been implemented in operational software and installed for on line operation at Crete, Ireland and Madeira in the frame of the European Project ENK-CT1999-00119.

VIII. REFERENCES

- [1] Alves da Silva A. P., Moulin L. S., "Confidence intervals for neural network based short-term load forecasting", *IEEE Trans. on Power Systems*, Vol. 15, No. 4, Nov. 2000.
- [2] Carney J. G., Cunningham P., Bhagwan U., "Confidence and prediction intervals for neural network ensembles", *Proc. of the International Joint Conf. on Neural Networks 1999*, paper 2090, 1999

- [3] De Veaux R., Schumi J., Scheinsberg J., Shellington D., Hungar L. H., "Prediction intervals for neural networks via nonlinear regression", *American Statistical Association*, vol. 40, n°4, Nov. 1998, pp. 273-282
- [4] Dybowski R., Gant V., "Confidence intervals and prediction intervals for feed-forward neural networks", in *Clinical Applications of Artificial Neural Networks*, Ed.: Dybowski-Gant, Cambridge University Press, 2000.
- [5] Heskes T., "Practical confidence and prediction intervals", in *Advances in Neural Information Processing Systems 9*, Ed.: Mozer-Jordan-Petsche, Cambridge, 1997, MIT Press, pp. 176-182.
- [6] Kariniotakis G. N., *Contribution to the development of an advanced control system for the optimal management of autonomous wind-diesel systems*, PhD thesis, Ecole Nationale Supérieure des Mines de Paris, Centre d'Energétique, December 1996.
- [7] Kariniotakis G. N., Stavrakakis G. S., Nogaret E., "Wind power forecasting using advanced neural network models", *IEEE Trans. on Energy Conversion*, Vol. 11, No. 4, Dec. 1996, pp. 762-767.
- [8] Kariniotakis, G. N., Nogaret, E., Stavrakakis, G., "Advanced Short-Term Forecasting of Wind Power Production", in *Proc. 1997 European Wind Energy Conference*, Dublin, 1997, pp. 751-754.
- [9] Dutton A. G., Kariniotakis G. N., Halliday J. A., Nogaret E., "Load and Wind Power Forecasting Methods for the Optimal Management of Isolated Power Systems with High Wind Penetration", *Wind Engineering*, Vol. 23, No 2, 2000.
- [10] Landberg L., "Poor-man's ensemble forecasting for error estimation", *Proc. of the 2002 AWEA Portland Conf.*, Portland, 2-5 June 2002.
- [11] Lange M., Waldl H.-P., "Assessing the uncertainty of wind power predictions with regard to specific weather situations", in *Proc. of the 2001 European Wind Energy Association Conference, EWEC'01*, Copenhagen (Denmark), pp.695-698, 2-6 July 2001.
- [12] Lange M., Waldl H.-P., "Accuracy of short term wind power predictions depending on meteorological conditions", in *Proc. of the 2002 Global Windpower Conference*, Paris (France), 2-5 April 2002.
- [13] Luig A., Bofinger S., Beyer H. G., "Analysis of confidence intervals for the prediction of regional wind power output", in *Proc. of the 2001 European Wind Energy Association Conference, EWEC'01*, Copenhagen (Denmark), pp.725-728, 2-6 July 2001.
- [14] Masters M., *Neural, Novel & Hybrid Algorithms for Time-Series Prediction*, J. Wiley & sons, 1995.
- [15] Nielsen T. S., Madsen H., Nielsen H. Aa., Landberg L., Giebel G., "Prediction of regional wind power", in *Proc. of the 2002 Global Windpower Conference*, Paris (France), 2-5 April 2002.
- [16] Roulston M. S., Kaplan, D. T., Hardenberg J., Smith L. A., "Value of the ensemble prediction system for forecasting wind energy production", in *Proc. of the 2001 European Wind Energy Association Conference, EWEC'01*, Copenhagen, pp.717-720, 2-6 July 2001.
- [17] Simon J. L., *Resampling : The new statistics*, 2nd edition, Oct. 1997, [Online] Available: <http://www.resample.com/content/text/index.html>

IX. ACKNOWLEDGMENT

The authors gratefully acknowledge ESB National Grid for providing data for the realization of this work.

X. BIOGRAPHIES

P. Pinson was born in Poitiers, France. He received his applied mathematics MSc degree from the National Institute for Applied Sciences (INSA Toulouse). He is currently a PhD student at the Center of Energy Studies of Ecole des Mines de Paris. His research interests include artificial intelligence and renewable energies.

G. N. Kariniotakis was born in Athens, Greece. He received his production and management engineering and MSc degrees from the Technical University of Crete, Greece and his PhD degree from Ecole des Mines de Paris in 1996. He is currently with the Center of Energy Studies of Ecole des Mines de Paris as a scientific project manager. He is a member of IEEE. His research interests include among others renewable energies, distributed generation and artificial intelligence.