

On-line adaptation of confidence intervals based on weather stability for wind power forecasting

Pierre Pinson, Georges Kariniotakis

► **To cite this version:**

Pierre Pinson, Georges Kariniotakis. On-line adaptation of confidence intervals based on weather stability for wind power forecasting. Global WindPower Conference, Mar 2004, Chicago, United States. hal-00529488

HAL Id: hal-00529488

<https://hal-mines-paristech.archives-ouvertes.fr/hal-00529488>

Submitted on 3 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On-line Adaptation of Confidence Intervals based on Weather Stability for Wind Power Forecasting

P. Pinson and G. Kariniotakis

École des Mines de Paris, Center for Energy Studies

BP. 207, F-06904 Sophia-Antipolis, France

pierre.pinson@ensmp.fr, georges.kariniotakis@ensmp.fr

Abstract

Existing literature or tools for wind power forecasting do not consider online estimation of confidence intervals for the output of the forecasting models. Uncertainty is estimated either based on error estimations coming from weather forecasts or on the inappropriate assumption that the error distribution is Gaussian and the intervals symmetrical around the spot predictions. This situation reveals the necessity to develop formal methods for on-line uncertainty estimation adequate for the problem of wind power forecasting. The paper introduces an advanced method for this purpose. The aim is to compute intervals for wind power forecasts with a confidence level defined by the end-user. The intervals are derived after an analysis of wind power prediction error characteristics. The error distribution parameter is estimated in an adaptive way after appropriate exploitation of past errors and using fuzzy set modeling. Then, an index named as MRI, expressing the expected weather stability is used for fine-tuning the intervals. This index reflects the spread of poor man's ensemble weather forecasts. A relation between the MRI and the level of power prediction error is shown: the linear trend is used for narrowing the intervals when weather situation is considered as stable.

Evaluation results of this methodology over a three-year period on the case study of a Danish wind farm and over a one-year period on the case study of nine Irish farms are given. The proposed methodology has an operational nature and can be applied to all kinds of wind power forecasting models.

Keywords: Wind power, short-term forecasting, confidence intervals, weather stability, on-line software, numerical weather predictions, ensemble forecasting, uncertainty.

1 Introduction

NOWADAYS, wind farm installations in Europe exceed 28 GW. Motivated by the Kyoto Protocol, the European Commission has set the target of doubling the share of renewables in gross energy consump-

tion from 6% in 1997 to 12% in 2010 [1]. This directive targets 22,1% indicative share of electricity produced from renewable energy sources in total Community electricity consumption by 2010. To achieve this share, installed wind power capacity in the Member States should increase to 45-60 GW. In 2003, the European Renewable Energy Council (EREC) revised upwards the 2010 target to 75 GW [2]. Such a large-scale integration of wind generation causes several difficulties in the management of a power system. Often, a high level of spinning reserve is allocated to account for the intermittent profile of wind production, thus reducing the benefits from the use of wind energy. Predictions of wind power production up to 48 hours ahead contribute to a secure and economic power system operation. Increasing the value of wind generation through the improvement of prediction systems' performance is one of the priorities in wind energy research needs for the coming years [3].

Apart from spot forecasts of the wind farms output in the next hours, of major importance is to provide tools for assessing on-line the accuracy of these forecasts. Tools for on-line evaluation of the prediction risk are expected to play a major role in trading wind power in a liberalized electricity market since they can prevent or reduce penalties in situations of poor prediction accuracy.

Typical confidence interval methods, developed for models like neural networks [4-6], are based on the assumption that the prediction errors follow a Gaussian distribution. This however is often not the case for wind power predictions, where error distributions exhibit some skewness, while the confidence intervals are not symmetric around the spot prediction due to the shape of the wind turbines power curve. Moreover, the level of predicted wind speed introduces some non-linearity 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 [7], which is based on modeling errors using a β -distribution, the parameters of which have to be estimated by a post-processing algorithm. A similar approach will be described in this work, with reformulated assumptions based on a study of the prediction errors characteristics.

In [8] 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. This method however is limited for application to physical models rather than to statistical ones since it requires local wind speed predictions (at the level of the wind farm) and the use of an explicit power curve, while it does not provide uncertainty as a function of a pre-specified confidence level. The wind speed errors are estimations provided by the Numerical Weather Prediction (NWP) model. As a consequence, this method does not take into account the modeling error itself that might be due to the spatial refinement of weather predictions or to the power curve used. Moreover, wind speed measurements are required, which might not be made available on-line. In a follow-up paper [9], the authors show a relation between specific meteorological patterns (defined from measurements) and various levels of forecasting error: this is a first step in the definition of risk indices in order to quantify the weather predictability and of weather-stability-dependant confidence intervals.

This paper proposes a new approach for assessing the uncertainty of wind power predictions in an adaptive way: the uncertainty estimation is based on the most recent errors. Therefore, it accounts for possible seasonal effects on the level of prediction error (which may be lower in summer than in winter), changes in the quality of input data, etc.

Initially, a generic methodology is developed for the estimation of confidence intervals that can be applied to both "physical" and "statistical" wind power forecasting models. This is due to the fact that the consideration of the power curve is not necessary. The method is based on the power prediction errors and, as a consequence, it accommodates all sources of errors such as modeling error, errors due to the inaccuracy of NWPs and also the stochastic component. It uses past wind power data, which are often available on-line by a Supervisory Control and Data Acquisition (SCADA) system, as well as NWPs, which are nowadays the basic input to all models.

Generally, when confidence higher than 80% is re-

quired, the intervals are quite wide. This can lead to conservative or costly managing strategies of the predicted wind power (i.e. allocation of high spinning reserve). 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 "predict the uncertainty" by considering the expected weather stability. This may allow operators to adjust the risk they are going to undertake when taking decisions related to the predicted wind power. In previous work by the authors [10], a relation between an index reflecting the weather stability and the level of forecasting error of a Fuzzy-Neural Network (Fuzzy-NN) based power prediction model was shown. This relation is used here for fine-tuning the confidence intervals according to the weather stability quantification.

The paper presents detailed results from the application of the methodology on the case studies of Ireland and Denmark, where the aim is to predict the output of several wind farms for 48 hours ahead using on-line measurements and predictions from Hirlam NWP system. This evaluation is based on several years of data.

2 On-line estimation of the uncertainty of wind power forecasts

Usually in the literature when off-line performance evaluation results of a model are presented, they refer to statistics calculated over data covering a long period. (a review of several evaluation criteria is given in [11]). Some references illustrate the differences in performance over time i.e. by estimating performance per month of the year. When it comes to on-line operation an even higher variability in performance is expected. For an operational environment it is thus needed to provide end-users with a more short-term assessment of the prediction model accuracy. A common way to provide that information is to design confidence intervals around the spot prediction. These intervals have to be built in a clever way in order to reflect the non-linearities of the power curve. Moreover, as it was shown in [10, 12], the level of prediction error is highly variable with the weather stability. Hence, methods for the design of confidence intervals should take this aspect into account.

The uncertainty assessment corresponds to a visualization of the expected error distribution, which is obtained after analysis of the errors the model made in the previous hours or days, in similar forecasting conditions. The methods developed here estimate uncertainty

as a function of the prediction horizon, the level of predicted power, the cut-off risk, and not on the basis of a global error distribution. For this, two approaches are proposed: the first is based on an adapted resampling method for designing the confidence bounds, while the second is based on fitting a β -distribution whose parameters are estimated by adapted resampling.

2.1 Prediction error distribution characteristics

The first aspect to be considered is that prediction errors depend on the look-ahead time. Indeed, models have their level of error increasing as the lead time augments. Consequently, error distributions are sharper for short look-ahead times and more flat for the long-term. These distributions are almost perfectly symmetric: this means that in the long term, there is as much chance to over- or under-predict the wind generation. When a power prediction model is evaluated, one of the first things that is checked is its bias. This value corresponds to the average difference between predictions and measures — i.e. a kind of systematic error the model makes, and is the mean of the prediction error distribution. This is why we will denote this bias as μ . For a given lead time k , it can be calculated as follows

$$\mu_k = \bar{e}_{t+k/t} = \bar{p}_{t+k} - \bar{\hat{p}}_{t+k/t}, \quad (1)$$

where $\bar{\cdot}$ stands for the arithmetic mean, p_{t+k} is the measured value at time $t+k$ and $\hat{p}_{t+k/t}$ is the forecast made at time t for time $t+k$.

A nice property that is wanted for a prediction tool is that it has no systematic error. But, even if its global bias is null, this does not mean that there is no bias whatever the level of predicted power. Figure 1 illustrates that comment: the solid curve with squares gives the bias of a F-NN prediction model depending on the predicted power level. The range of power generation is splitted into ten equal classes, each representing 10% of the possible predicted power values. Errors are collected, and sorted depending on the predicted power. This way, error distributions related to every power class are obtained. In this Figure, all the lead times are considered.

When the predicted power is close to zero, the bias is significantly positive: the trend is to under-predict. At the inverse, when the predicted power is close to the installed capacity, the trend is clearly to over-predict. Two causes for this phenomenon: the superior bound of the power curve and the cut-off effect. Indeed, if a cut-off is not forecast, the prediction is significantly

greater than the actual power, and this raises the bias of the error distribution.

The second parameter of interest for the error distributions is the standard deviation σ , which reflects the dispersion of the errors around the mean:

$$\sigma_k = \left(\frac{1}{N_p - 1} \sum_{t=1}^{N_p} (e_{t+k/t} - \mu_k)^2 \right)^{\frac{1}{2}}, \quad (2)$$

where k is the look-ahead time and N_p the total number of predictions considered for the evaluation.

In [8, 13], it was shown that the power prediction errors depend on the errors involved in the prediction of wind speed by the NWP system. Due to its shape, the wind park power curve is able to amplify (between cut-in and rated speed) or to reduce (below cut-in speed or between rated and cut-off speed) the uncertainty introduced by the NWPs. Hence, as one can notice that the standard deviation is very similar to the NRMSE that is used for evaluating the performance of wind prediction tools, it is expected that the standard deviation of forecast-power-dependant error distributions will vary in a similar way. In Figure 1, the curve with circles gives the σ for each power class. Indeed, it appears that σ is higher for medium forecast powers (corresponding to the high slope part of the power curve) and lower for low and high forecast power (corresponding to the flat parts of the power curve). This is coherent with similar conclusions in [8, 13].

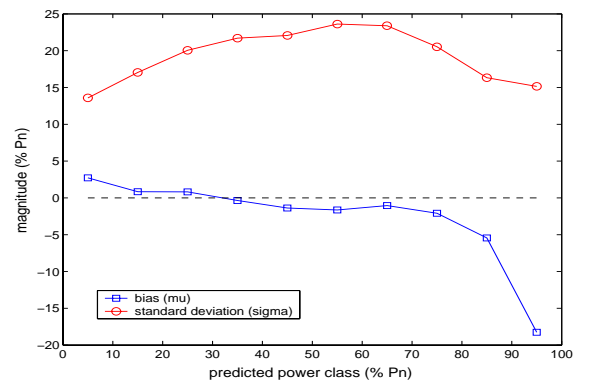


FIGURE 1: Bias and standard deviation of the forecasting error distribution depending on the predicted power class. Results are for a wind farm in Ireland.

Summarizing, error distributions (and its parameters μ and σ) depend on the lead time, and on the level of predicted power. Roughly, There are three zones of the power curve that lead to three different characteristics:

- low predicted power (below cut-in speed): the

trend is to under-predict ($\mu > 0$) and the error dispersion is significantly lower,

- medium predicted power (high slope part of the power curve): the bias can be considered as null, but the error dispersion is higher,
- high predicted power (close to rated power): the trend is clearly to over-predict ($\mu < 0$) and the error dispersion is significantly lower. This bias is amplified by the unpredicted cut-off events, where the prediction error is maximal.

These conclusions have been confirmed in several case-studies. They form the basis for defining in the following Sections a method for designing confidence intervals that is appropriate to the wind power prediction problem.

2.2 Error pre-processing based on fuzzy set modeling

The first step before computing confidence intervals is to collect the prediction errors the model made in the past. The intervals to be computed will rely on the most recent information on the model's performance. For this, a window in the past (a certain number of hours) is defined and used as a sliding window for storing the errors. The size of this window defines the size of the samples of errors. A separate sample is developed for each prediction horizon k (i.e. for 1-hour ahead, 2-hour ahead, and so on). 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 ago).

To account for the power curve effect detailed in the previous paragraph, the power curve is divided into three ranges of power: low, medium and high, which are characterized by fuzzy sets. The prediction errors are then classified in samples $S_{k,p}^1$, $S_{k,p}^2$ and $S_{k,p}^3$, depending on the predicted power range (Figure 2). Hence, the confidence interval estimation is carried out using the error samples corresponding to the power class of the forecast power.

In a similar way, in order to deal with the risk due to the cut-off event, the range of wind speed values is divided into two ranges corresponding to a "no cut-off risk zone" for low wind speeds, and to a "cut-off risk zone" for wind speeds close or higher than cut-off. Like for the predicted power, errors are stored in samples $S_{k,ws}^1$ and $S_{k,ws}^2$, depending on the cut-off risk. An

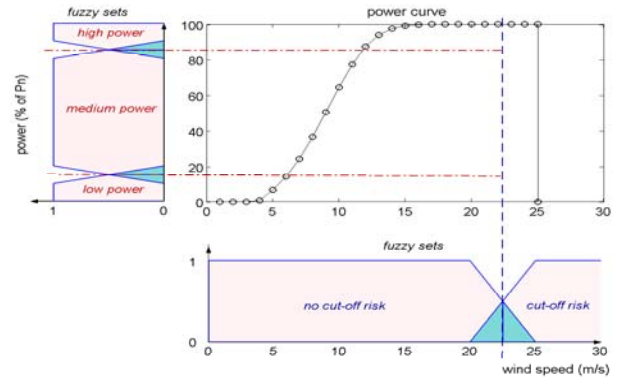


FIGURE 2: Splitting the power curve into three power class fuzzy sets and two cut-off risk zones.

appropriate trapezoidal fuzzy set is associated to each zone as shown in Figure 2.

2.3 Distribution parameter estimation by an adapted 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 a high number of times. This evaluation of the population distribution can serve to estimate a mean, a variance, or even confidence intervals. No assumption is made concerning the distribution.

Let us consider a sample containing N observations of a stochastic variable Y for a given stochastic process. The procedure to compute from this sample a distribution parameter λ will be based on the following steps, which are to be repeated a sufficiently large number of times (N_{loop}): firstly create a new sample by selecting randomly and with replacement N observations out of the original sample, and secondly estimate λ_i ($i \in 1, \dots, N_{loop}$) for this generated sample. Then, by computing the average of every λ_i , good estimate of the λ parameter can be obtained [14, 15].

In the case of wind power forecasting, the resampling method is applied by considering error samples defined as a function of the look-ahead time, the power range and the wind speed range. For a given horizon k , Figure 2 represents the splitting of the predicted power range of values into three fuzzy sets $A_{k,p}^1$, $A_{k,p}^2$ and $A_{k,p}^3$, accounting respectively for low, medium and high predicted power. In a similar way, the forecast wind speed values allows one to define two fuzzy sets $A_{k,ws}^1$ and $A_{k,ws}^2$ for situations without or with a risk of cut-off event.

In order to account for the specific shape of the power curve, the first step of the Resampling method is adapted by using fuzzy rules, in order to create a new sample that reflects the current conditions (for an introduction to the fuzzy logic theory we refer to [16]). A fuzzy rule will be of the form:

$$\begin{aligned} \text{IF } \hat{p}_{t+k/t} \in \mathcal{D}(\mathcal{A}_{k,p}^1) \text{ and } \hat{w}_{s_{t+k/t}} \in \mathcal{D}(\mathcal{A}_{k,ws}^1) \\ \text{THEN } X \in \mathbb{R}^N, \quad X \subset (S_{k,p}^1 \cap S_{k,ws}^1), \end{aligned} \quad (3)$$

where $\mathcal{D}(\mathcal{A})$ stands for the support of the fuzzy set \mathcal{A} .

This rule means that if the predicted power is in the low range and the forecast wind speed in the "no cut-off risk" range, then the new generated sample $X \in \mathbb{R}^N$ will be composed by values picked in the intersection of the error samples accounting for these specific situations: X will be generated by selecting randomly and with replacement N values out of the intersection of $S_{k,p}^1$ and $S_{k,ws}^1$

Since there are three fuzzy sets defined for the predicted power and two for the forecast wind speed, six rules of that kind can be formulated. However, only N values have to be picked out to create a new sample. Therefore, the fuzzy set membership functions are used to define the share of each rule in the final sample. This consideration leads to a new form of the fuzzy rule:

$$\begin{aligned} \text{IF } \hat{p}_{t+k/t} \in \mathcal{D}(\mathcal{A}_{k,p}^i) \text{ and } \hat{w}_{s_{t+k/t}} \in \mathcal{D}(\mathcal{A}_{k,ws}^j) \\ \text{THEN } X^{ij} \in \mathbb{R}^{N^{ij}}, \quad X^{ij} \subset (S_{k,p}^i \cap S_{k,ws}^j), \end{aligned} \quad (4)$$

with

$$N^{ij} = \frac{m_{k,p}^i(\hat{p}_{t+k/t}) \cdot m_{k,ws}^j(\hat{w}_{s_{t+k/t}})}{\sum_{l=1}^3 \sum_{n=1}^2 m_{k,p}^l(\hat{p}_{t+k/t}) \cdot m_{k,ws}^n(\hat{w}_{s_{t+k/t}})} N, \quad (5)$$

and $i \in \{1, 2, 3\}$, $j \in \{1, 2\}$. In this expression, $m_{k,p}^i(\cdot)$ and $m_{k,ws}^j(\cdot)$ are the membership functions of respectively the i^{th} fuzzy set associated to power and the j^{th} fuzzy set associated to wind speed.

Then, the generated sample will be composed by the subsamples created by all the rules. However, we make the assumption that when the forecast wind speed is very high, showing a risk of cut-off event, the value of the predicted power will not have a significant influence on that risk. Thus, the three rules corresponding to the cut-off risk are gathered to form only one rule.

Finally, the created sample consists in

$$X = \begin{pmatrix} X^{11} \\ X^{21} \\ X^{31} \\ X^{.2} \end{pmatrix}, \quad X \in \mathbb{R}^N, \quad (6)$$

where $X^{.2}$ denotes the subsample obtained with the unique cut-off risk rule defined above.

The basic loop of the adapted Resampling algorithm can be reformulated as in Algorithm 1. Then, the error distribution parameters are estimated by using this adapted Resampling method for every lead time k .

Algorithm 1 Calculate distribution parameters μ and σ by an adapted Resampling approach

$$\mu \leftarrow 0$$

$$\sigma \leftarrow 0$$

for $i = 0$ to N_{loop} **do**

create a new sample Y_i by selecting randomly and with replacement N values out of the original sample and following (4), (5) and (6)

$$\mu_i \leftarrow E[Y_i]$$

$$\sigma_i \leftarrow (E[(Y_i - E[Y_i])^2])^{\frac{1}{2}}$$

end for

$$\mu \leftarrow \frac{1}{N_{loop}} \sum_{i=1}^{N_{loop}} \mu_i$$

$$\sigma \leftarrow \frac{1}{N_{loop}} \sum_{i=1}^{N_{loop}} \sigma_i$$

2.4 Design of appropriate confidence intervals

Here is a formal definition of confidence intervals: the interval $\mathcal{I}(X)$ computed from the sample data X which, were the study repeated multiple times, would contain $(1 - \alpha)$ % of the time the true effect a , $(1 - \alpha)$ being the confidence level:

$$P(a \in \mathcal{I}(X)) = P(a \in]z_{\alpha/2}, z_{1-\alpha/2}[) = 1 - \alpha. \quad (7)$$

The following paragraphs expose two alternative methods for the building of confidence bounds: by applying the adapted resampling method described above, or by fitting a β -distribution to the error distributions.

2.4.1 Main approach based on adapted Resampling

Since Resampling permits to estimate parameters of a distribution like its mean or its variance, it can also be directly used for the estimation of confidence bounds.

For this, one has to pick the $\alpha/2$ lowest $z_{\alpha/2}^i$ and $(1 - \alpha/2)$ highest $z_{1-\alpha/2}^i$ values of the i^{th} created sample and to compute their mean to obtain the interval limits:

$$\begin{aligned} z_{\alpha/2} &= \frac{1}{N_{loop}} \sum_{i=1}^{N_{loop}} z_{\alpha/2}^i, \\ z_{1-\alpha/2} &= \frac{1}{N_{loop}} \sum_{i=1}^{N_{loop}} z_{1-\alpha/2}^i. \end{aligned} \quad (8)$$

The main advantage of this method is that intervals are built without fitting a known distribution on the available data.

2.4.2 Alternative approach based on the β -distribution

The β -distribution has nice properties that correspond to the characteristics of the wind power prediction problem: it is bounded between 0 and 1, while predicted power values lie between these two values if they are normalized by the wind farm rated capacity P_n , and its modes and moments can be controlled by two shape parameters α and β . The density function for that distribution has the following form:

$$f_{\beta}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0, 1], \quad (9)$$

where $\Gamma(\cdot)$ is the Gamma function. The definition of the Gamma function leads to an important constraint on the shape parameters:

$$\alpha > 0 \quad \text{and} \quad \beta > 0 \quad (10)$$

The fitting of prediction errors to that particular distribution has already been proposed in [7], with pre-defined standard deviation (depending on 4 classes of predicted power) and the assumption that the distribution mean is given by the predicted power value. This assumption is the result of another assumption, which states that the prediction model bias is removed completely. However, even if this assumption can be true in the long-term, we have shown previously that this bias is significantly positive for low predicted power, and clearly negative for forecast power close to the rated power. This is why we propose instead to base the methodology on the following assumption: since the error distribution bias can be estimated by the Resampling approach, the β -distribution mean μ_k^{β} (for a given lead time k) can be considered to be the predicted power plus the estimated bias. The standard deviation of the distribution is then estimated with the adapted Resampling approach:

$$\mu_k^{\beta} = \hat{p}_{t+k/t} + \mu_k, \quad (11)$$

$$\sigma_k^{\beta} = \sigma_k. \quad (12)$$

The β -distribution is completely characterized by its mean and its variance. Indeed, there is a direct relation between these first moments and the shape parameters α_k and β_k :

$$\alpha_k = \mu_k^{\beta} \left[\frac{\mu_k^{\beta}(1 - \mu_k^{\beta})}{(\sigma_k^{\beta})^2} - 1 \right], \quad (13)$$

$$\beta_k = (1 - \mu_k^{\beta}) \left[\frac{\mu_k^{\beta}(1 - \mu_k^{\beta})}{(\sigma_k^{\beta})^2} - 1 \right]. \quad (14)$$

To compute the β -distribution quantiles for a given confidence level, we refer to the approximations using continuous fractions that are given in [17].

One has to notice the importance of the constraint formulated by (10): if this constraint is not satisfied, then it is not possible to model the prediction error distribution with a β -distribution. Reformulating this constraint in terms of estimated mean and standard deviation with Equations (13) and (14) yields:

$$\mu_k^{\beta} > 0, \quad (15)$$

$$\mu_k^{\beta} < 1, \quad (16)$$

$$\mu_k^{\beta}(1 - \mu_k^{\beta}) - (\sigma_k^{\beta})^2 > 0. \quad (17)$$

The last constraint signifies that as the distribution mean is closer to the limits the distribution variance has to be lower: this is a similar behavior we observed in the study of the prediction errors depending on the level of predicted power. Relation (17) strictly defines an upper bound for σ_k^{β} for a given μ_k^{β} .

3 Considering weather stability for fine-tuning the width of the intervals

The methods for the design of confidence intervals developed in the first part of this paper are based on the past performance of the model. Although the most recent information is used and constantly updated, they do not consider the current meteorological situation and thus a crucial information: *weather predictability*.

Low quality forecasts are due partly to the power prediction model, and partly to the numerical weather prediction system (due in turn to low weather stability). Indeed, an unstable atmospheric situation can lead to very poor numerical weather predictions and thus to worthless wind energy ones. In contrast, when the atmospheric situation is stable, one can expect more accurate wind power predictions from the model.

In [9], methods from synoptic climatology are utilized to classify the local weather conditions thanks to measurements of wind speed and direction, as well as pressure. This classification, through principal component analysis and cluster analysis methods, allows one to reveal characteristic meteorological patterns that can be associated to various levels of prediction error. Indeed, when low-pressure systems are dominant, the level of error is higher, and at the inverse, when high-pressure systems govern, the wind prediction error is much lower. However, no link with wind power prediction errors is shown, and this study is based on meteorological measurements, which are often not available online. This means that it is not sure the derived method can be used afterwards in a preventive way for detecting situations for which high level of forecasting error is expected. That is the reason why the ideas developed in the following paragraphs exploit the information included in the NWP (and not in the measurements) in order to develop tools for on-line estimation of the meteorological risk in power predictions.

3.1 Wind speed ensemble forecasts for the assessment of weather stability

Meteorological Centers are able to produce different scenarios of Numerical Weather Predictions by perturbing the initial conditions of the forecasting model or by using different NWP models. These scenarios are called ensemble forecasts and permit one to evaluate the stability of the weather regime as well as the meteorological predictability [18]. Both the U.S. National Center for Environmental Prediction (NCEP) and the European Center for Medium-Range Weather Forecasts (ECMWF) have produced operational ensembles for more than ten years.

The NWP wind speed prediction error is composed by a part that is independent of the lead time and by an error that has a linear growth with the prediction horizon. The first includes the effects of weather disturbances that are smaller than the NWP resolution, while the second is due partly to the NWP model errors and partly to an error in the estimation of the initial state [19]. Ensemble forecasts permit one to see the influence of this misestimation of the initial state in the weather forecasting evolution, and thus to preventively quantify the prediction uncertainty [20].

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. Nevertheless, for

a given hour, several predictions can be available from different time origins in the past (-6 hours, -12 hours, -18 hours, etc.). This kind of ensembles is known as poor man's ensemble forecasts. In a stable and well-predicted weather situation it is expected that these predictions will not differ significantly. Comparing all the available forecasts for the considered period can assess weather stability and predictability.

Because we want to have a general evaluation of that stability, 4 sets of predictions of various ages (0, 6, 12 and 18 hours) for the following 24 hours are compared. Figure 3 gives the examples of a stable atmospheric situation (left picture, the forecasts are quite close) and of an unstable one (right picture, spread forecasts).

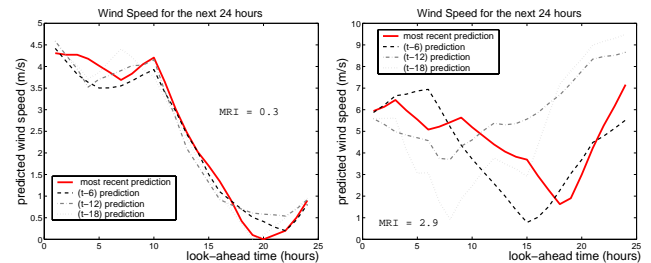


FIGURE 3: Stable (left picture) and unstable (right picture) weather situations.

3.2 Relation between weather stability and wind power prediction error

There are several possibilities to measure the spread of the various weather forecasts, which appear to be equivalent. In [13] 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. In order to calculate the distance between two sets of forecasts, the authors have proposed in [10] to use a kind of euclidian distance between the N_h -valued vectors containing the predicted wind speed for the N_h following hours. Focus is given to the spread of wind speed forecasts because this variable is the main and most sensible input of wind power prediction models. Then, an index, called hereafter "meteo-risk" MRI-index, is defined in N_h to measure the spread of the weather forecasts at a given time. This index uses the most recent forecast as a reference and reflects the variability of the older forecasts.

In the frame of the case studies of the paper, the horizon N_h for the calculation of the MRI-index is set to 24 hours. Since Hirlam forecasts are provided every

6 hours, there are four sets of wind speed predictions covering the period.

The MRI-index can be used to describe the distribution of weather situations as shown in [10]. In order to have a simple and clear illustration of the link between weather stability and the level of prediction error, we give the average performance of the Fuzzy-Neural Network (Fuzzy-NN) prediction model described in [21] for a wind farm in Ireland, which we compare to its performance for weather situations considered as stable or unstable, as described by the MRI-index. One can see that there is a great variation of the model performance depending on the weather conditions.

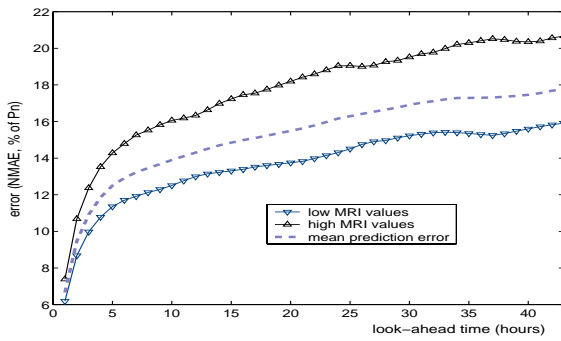


FIGURE 4: The level of power prediction error depending on the weather stability as this is described by the MRI-index for a wind farm in Ireland.

The link of prediction error to weather stability is shown for the case study of a Danish wind farm. The power prediction errors, as obtained by the Fuzzy-Neural Network (F-NN) prediction model described in [21], are collected for a period covering 3 years. For the same period the MRI-index is estimated. The prediction error e_t^{24} for the next 24 hours, corresponding to the power forecast made at time t , is calculated as follows:

$$e_t^{24} = \frac{1}{24 \cdot P_n} \sum_{k=1}^{24} |e_{t+k/t}|. \quad (18)$$

Then, these errors are binned by NPRI-index values, and the average error $e_{t,j}^{24}$ ($j = 1, 2, \dots, N_{bin}$) for the next 24 hours for each bin is computed. Finally, by comparing these averages to the global prediction error $\overline{e_t^{24}}$ of the model

$$\overline{e_t^{24}} = \frac{1}{N_p} \sum_t e_t^{24}, \quad (19)$$

where N_p is the total number of predictions made in the testing set, the representative points in Figure 5 are obtained.

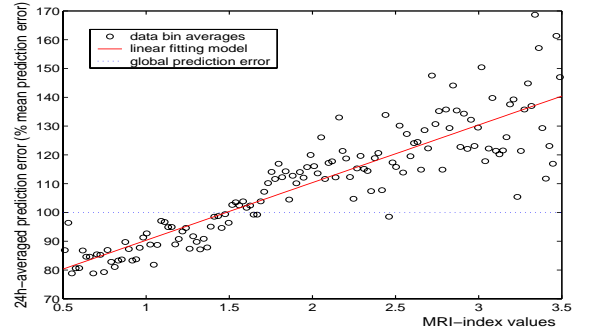


FIGURE 5: Prediction errors vs MRI-index over a three year dataset for a wind farm in Denmark: there is a roughly linear relation between the prediction error and MRI-index values.

This figure exhibits a roughly linear trend: the prediction error tends to increase linearly with the MRI-index: the tighter the Hirlam predictions are, the more accurate the wind power prediction model is. A linear fitting gives the solid curve. However, the information we can extract from this figure highlights a trend and not a direct relation between the meteorological risk and the prediction error, because it is not possible to link directly a MRI-index value to an error value, though we can say that for low or high MRI values, there are respectively less and more chances for high prediction errors. Making this assumption would mean that the prediction error the model makes, follows an affine empiric relation:

$$e_t^{24} = e_0 + s \cdot \text{MRI}, \quad (20)$$

which is composed by a basic part of the error e_0 and by a NWP-dependent error, the latest being a direct consequence of the prediction model sensibility to the weather stability. The slope s of the linear fitting model represents that sensibility.

3.3 Fine-tuning of the confidence intervals

The relation (20) indicates that when the MRI-index is low, the model is expected to be more accurate. In that case one would be ready to accept tighter confidence intervals for the predictions. The aim here is to use Equation (20) to define a scale factor for the confidence intervals depending on the value of the MRI-index. This scale factor can be applied to either enlarge or narrow the intervals width in the following N_h hours. For instance, when the meteorological index equals 0.5, the size of the intervals for the following 24 hours is reduced by around 20% (for the example given by Figure 5). The strategy chosen here is to only narrow

the intervals when the MRI-index allows one to do so. This can be done most of the times (around 65% of the times) [10].

The interest of those results is that they are obtained from spot numerical weather predictions (that are used to generate poor man's ensemble forecasts) and not from 'real' meteorological ensembles. This is a real operational advantage: almost all of the wind power prediction systems use single NWP as input and not ensemble NWP, which are harder to purchase, and more expensive.

4 Discussion

An important question concerning the intervals arises: how to choose an optimal confidence level? When required confidence is higher than 90%, intervals can be embarrassingly wide, because they will contain extreme prediction errors (or even outliers). However, if one requires low confidence level (50% for instance), intervals will be much more narrow and thus more robust with respect to extreme prediction errors, but this will mean that actual future values are equally likely to lie inside or outside these bounds. In both cases, confidence intervals appear to be hard to handle and that is why an intermediate confidence level (80-85%) seems a good compromise.

Moreover, the fact that confidence intervals are designed for multi-step ahead forecasts imposes to define what is the real required confidence. As a matter of fact, there is a difference between a confidence for each predicted value and a confidence required over the whole prediction horizon. For instance, if 85% confidence is required for one-day ahead hourly predictions, the former corresponds to "each of the 24 intervals will contain the true value 85% of the times", while the latter means that "the 24 intervals will contain all the 24 true values 85% of the times". The second way of thinking is obviously much more restrictive and seems less applicable in our case. As we explained in previous sections, the method for bound estimation is applied separately for every look-ahead time, because we consider that confidence should be required for each predicted value and not for a set of predictions. Therefore, this is the way the observed confidence will be checked in order to assess their performance.

5 Results

Results are presented for nine real wind farms in Ireland (WF_a to WF_i) and for a wind farm in Denmark (WF_{dk})

with a total installed power of several tens of MW. The prediction model is the Adaptive Fuzzy-NN model described in [21], which provides two-days ahead predictions with on-line production data and NWP as input.

The available time-series for the Irish wind farms cover a period of almost two years, from which 6600 hours were used for training (learning set), 1000 hours for cross-validation and one year for testing the performance of the model and of the uncertainty assessment methods. Regarding WF_{dk} , the time-series cover a period of almost two years (learning: 12000 hours, cross-validation: 2000 hours, testing: three years). The results presented here are on the testing set. Concerning the computation of error distribution parameters (and thus confidence intervals), 12 days of prediction errors are stored in the samples. The desired confidence level is set to 85%. Finally, the MRI-index is computed each hour for the following 24 hours by using the last four Hirlam wind speed forecasts. A different scaling of the intervals with the weather predictability estimation is defined for each wind farm, based on a statistical analysis of the relation between MRI and model prediction errors.

Figure 6 depicts an episode with the wind power predictions for the next 43 hours compared to the real values for WF_{dk} . The 85% confidence intervals are built with the resampling method described above.

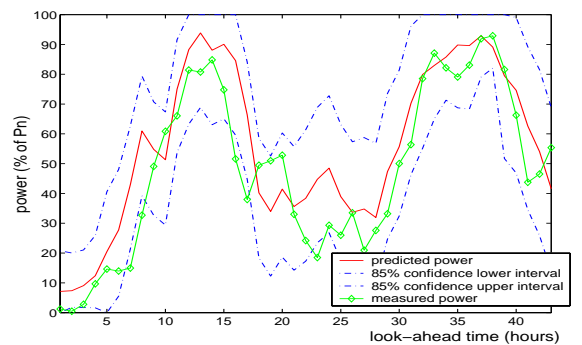


FIGURE 6: Wind power prediction with 85% confidence intervals for WF_{dk} .

In order to illustrate the fine-tuning of the intervals, Figure 7 gives the example of a weather situation classified as stable with respect to the "meteo-risk" index, for the Irish wind farm denoted as WF_h . For the first 24 look-ahead times the resampling on past errors produces quite broad intervals, but their size is reduced by more than 20% afterwards thanks to the consideration of the weather situation.

The resampling approach that is used to estimate the error distribution parameters and to design the confidence intervals displays a dynamic behavior thanks to

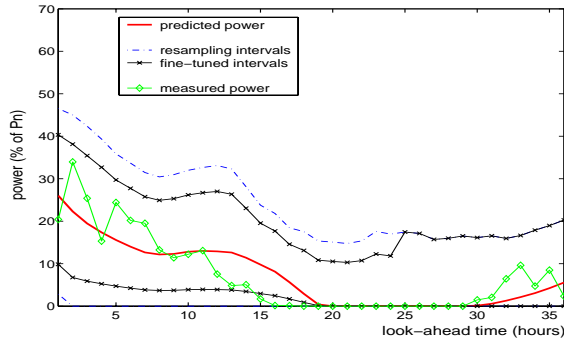


FIGURE 7: Spot wind power predictions together with the resampling and fine-tuned confidence intervals for WF_{dk} . The intervals are narrowed for the first 24 hours due to a low MRI-index value.

the updating of the sample of errors and also thanks to the fuzzy set modeling of the power curve. An interesting point is that these intervals are non-symmetric: for instance the prediction error is more likely to be positive than negative when forecasting low power output. This adaptivity property of the method was already commented in previous work [10].

The performance of the confidence interval estimation for the nine wind farms in Ireland is summarized in Table 1, i.e. the observed confidence at the end of the evaluation period (one year) for both resampling and β intervals. One can notice that the formers are slightly too wide, since the observed confidence is a bit higher than requested, while the latter are relatively narrow. This is due to both the assumption on the shape of error distributions and to a slight underestimation of the error standard deviation by the adapted resampling.

TABLE 1: Evaluation over the testing sets of the observed confidence for a preset confidence level of 85% for the two types of intervals

wind farm	β -int. (%)	resampling int. (%)
WF_a	77.76	87.79
WF_b	84.01	86.94
WF_c	80.57	86.99
WF_d	83.27	85.64
WF_e	79.48	86.29
WF_f	82.52	86.60
WF_g	75.31	87.46
WF_h	79.89	86.20
WF_i	80.97	85.84

The results given by Table 1 correspond to intervals obtained without considering the weather stability estimated by the MRI-index. Indeed, the reduction of the interval size for stable weather situation is shown in Ta-

ble 2. This Table gives the number of times the intervals are narrowed, the average size reduction over the whole evaluation period, as well as the confidence loss that may be due to the linear scaling factor assumption or to a too optimistic estimation of the weather predictability.

TABLE 2: Observed confidence for the two types of intervals at the end of the testing set and effects of the MRI-index on the interval reduction. (*ocri*: observed confidence of resampling intervals, *ocfi*: observed confidence of fine-tune intervals, *ntir*: number of times intervals are reduced, *awr*: average width reduction)

wind farm	<i>ocri</i> (%)	<i>ocfi</i> (%)	<i>ntir</i> (%)	<i>awr</i> (%)
WF_a	87.89	86.13	64.60	8.83
WF_b	86.95	84.94	70.70	10.53
WF_c	86.99	86.08	69.54	7.70
WF_d	85.64	84.19	59.64	8.02
WF_e	86.29	84.89	63.62	8.54
WF_f	86.60	84.98	61.90	8.12
WF_g	87.46	85.98	62.74	8.49
WF_h	86.20	84.23	68.54	10.03
WF_i	85.84	84.15	68.38	9.59

One can see that the consideration of the weather stability allows one to narrow the intervals most of the times (around 65%), and the average reduction is up to almost 11% of their initial size (WF_b). The corresponding confidence loss is not significant.

6 Conclusions

A generic methodology for assessing on-line the uncertainty of short-term wind power forecasts has been presented. The developed methods were introduced after a study of the main prediction error distribution characteristics. For the computation of the error distribution parameters, adapted resampling over up-to-date collections of past errors is performed. This dynamic approach has been proven well adapted to the prediction problem of nonstationary wind power production. This is due to the consideration of the prediction horizon, the level of predicted power and the cut-off risk when sorting and resampling past errors. Then, a comparison has been made between confidence bounds produced either directly with the adapted resampling method or by fitting a β distribution.

In a second stage, a previously defined meteorological risk (MRI) index has been introduced to evaluate the weather stability, by reflecting the spread of poor man's ensemble forecasts of wind speed. A linear trend between this index and the level of prediction error for the

following 24 hours has been shown and has permitted to act on the confidence interval width afterwards depending on the weather predictability. We have shown over long testing periods that the size reduction of the intervals was significant without altering the observed bound confidence. This fine-tuning is a second element that characterizes the dynamic approach developed in this paper.

The advantage of the described methodology for the uncertainty management is its operational nature. Indeed, the techniques for the basic design of intervals are suitable for every kinds of deterministic wind power forecasting methods and the fine-tuning is based on single NWP (used to generate poor-man's ensemble forecasts of wind speed) that are nowadays a common input to most of the prediction models. In the future, other ways of integrating the weather predictability in the uncertainty assessment methodology will be studied, like the utilization of the MRI-index for the error classification for instance.

The developed tool for on-line uncertainty evaluation is integrated in the AWPPS (Armines Wind Power Prediction System) as well as in the next generation wind power forecasting platform developed in the frame of the ANEMOS project. On-line installations to various onshore and offshore wind farms and on-line evaluation are planned for 2005.

Finally, as the value of wind power forecasting models has to be assessed, a study is on-going aiming to evaluate and quantify the benefits of the proposed methodology for the uncertainty and prediction risk management. Emphasis is given to the contribution of the methodology to estimate reserves, storage requirements as well as developing strategies for trading wind power in electricity markets.

Acknowledgment

This work was performed in the frame of the ANEMOS Project (ENK5-CT-2002-00665) funded in part by the European Commission. The authors gratefully acknowledge ESB National Grid and ELSAM for providing data for the realization of the study.

References

[1] Official Journal of the European Commission, "Directive 2001/77/EC of the European Parliament and of the Council, on the promotion of electricity produced from renewable energy sources in the internal electricity market." Sept. 2001.

- [2] A. Zervos, "Developing wind energy to meet the Kyoto targets in the European Union," *Wind Energy*, vol. 6, pp. 309–319, July–September 2003.
- [3] S.-E. Thor and P. Weis-Taylor, "Long-term research and development needs for wind energy for the time frame 2000-2020," *Wind Energy*, vol. 5, pp. 73–75, April–June 2003.
- [4] J. G. Carney, P. Cunningham, and U. Bhagwan, "Confidence and prediction intervals for neural network ensembles," in *Proc. of the International Joint Conference on Neural Networks 1999*, 1999. paper 2090.
- [5] T. Heskes, "Practical confidence and prediction intervals," in *Advances in Neural Information Processing Systems 9* (M. J. M. Mozer and T. Petsche, eds.), MIT Press, Cambridge, 1997.
- [6] R. De Veaux, J. Schumi, J. Schweinberg, D. Shellington, and L. H. Ungar, "Prediction intervals for neural networks via nonlinear regression," *American Statistical Association*, vol. 40, Nov. 1998.
- [7] A. Luig, Bofinger, and H. G. Beyer, "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, June 2001.
- [8] M. Lange and H.-P. Waldl, "Accuracy of short-term wind power prediction depending on meteorological conditions," in *Proc. of the 2002 Global Windpower Conference, Paris, France*, April 2002.
- [9] M. Lange and D. Heinemann, "Relating the uncertainty of short-term wind speed prediction to meteorological situations with methods from synoptic climatology," in *Proc. of the 2003 European Wind Energy Association Conference, EWEC'03, Madrid, Spain*, June 2003.
- [10] P. Pinson and G. Kariniotakis, "On-line assessment of prediction risk for wind power production forecasts," *Wind Energy*, to appear in 2004.
- [11] H. Madsen, H. A. Nielsen, T. S. Nielsen, G. Kariniotakis, and P. Pinson, "A protocol for standardizing the performance evaluation of short-term wind power prediction models," in *Proc. of the 2004 Global Windpower Conference, Chicago, U.S.*, March 2004.
- [12] P. Pinson and G. Kariniotakis, "Uncertainty and prediction risk assessment of short-term wind power forecasts," in *Proc. of the 2004 EAWE Conference, 'The science of making torque from wind', Delft, The Netherlands*, April 2004.
- [13] L. Landberg, G. Giebel, L. Myllerup, J. Badger, H. Madsen, and T. S. Nielsen, "Poor-man's ensemble forecasting for error estimation," in *Proc. of the 2002 American Wind Energy Association conference, Portland, Oregon*, June 2002.

- [14] T. Masters, *Neural, Novel & Hybrid Algorithms for Time Series Prediction*. John Wiley, 1995.
- [15] J. L. Simon, *Resampling: The new statistics*. Available online, 2nd ed., 1997.
- [16] L.-W. Wang, *Adaptive fuzzy systems and control*. Prentice Hall, 1994.
- [17] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with formulas, graphs, and mathematical tables*. Dover, 1991.
- [18] J. M. Murphy, "The impact of ensemble forecasts on predictability," *Q. J. R. Meteorol. Soc.*, vol. 114, pp. 463–493, 1988.
- [19] J. U. Joergensen and C. Moehrlen, "Verification of ensemble prediction systems for a new market: wind energy." Special project interim report, available online, June 2003.
- [20] Z. Toth, Y. Zhu, and T. Marchock, "The use of ensembles to identify forecasts with small and large uncertainty," *Weather and Forecasting*, vol. 16, no. 4, pp. 463–477, 2001.
- [21] G. Kariniotakis and D. Mayer, "An advanced on-line wind resource prediction system for the optimal management of wind parks," in *Proc. of the 2002 Med-Power Conference, Athens, Greece*, Nov. 2002.