

Skill forecasting from ensemble predictions of wind power

P. Pinson^{*,a}, H.Aa. Nielsen^a, H. Madsen^a, G. Kariniotakis^b

^a*DTU Informatics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

^b*Centre for Energy and Processes, Ecole des Mines de Paris, 06904 Sophia Antipolis, France*

Abstract

Optimal management and trading of wind generation calls for the providing of uncertainty estimates along with the commonly provided short-term wind power point predictions. Alternative approaches for the use of probabilistic forecasting are introduced. More precisely, focus is given to prediction risk indices aiming to give a comprehensive signal on the expected level of forecast uncertainty. Ensemble predictions of wind generation are used as input. A proposal for the definition of prediction risk indices is given. Such skill forecasts are based on the spread of ensemble forecasts (i.e. a set of alternative scenarios for the coming period) for a single prediction horizon or over a look-ahead period. It is shown on the test case of a Danish offshore wind farm how these prediction risk indices may be related to several levels of forecast uncertainty (and potential energy imbalances). Wind power ensemble predictions are derived from the conversion of ECMWF and NCEP ensemble forecasts of meteorological variables to wind power ensemble forecasts, as well as by a lagged average approach alternative. The ability of prediction risk indices calculated from the various types of ensembles forecasts to resolve among situations with different levels of uncertainty is discussed.

Key words: wind power, forecasting, uncertainty, skill forecasting, prediction risk indices, ensemble forecasting.

*Corresponding author

Email addresses: pp@imm.dtu.dk (P. Pinson), han@imm.dtu.dk (H.Aa. Nielsen), hm@imm.dtu.dk (H. Madsen), georges.kariniotakis@mines-paristech.fr (G. Kariniotakis)

Preprint submitted to Elsevier

October 12, 2008

1. Introduction

The large scale integration of wind generation capacities induces difficulties in the management of a power system. An additional challenge is to conciliate this deployment with the on-going deregulation of the European electricity markets. 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 [1]. Relevant prediction horizons are up to 72-hours ahead. Focus is given to this forecast length in the present paper. A complete state of the art on wind power forecasting has been published in [2], while a more recent one may be found in [3]. A large part of the recent research works in wind power forecasting has concentrated on associating uncertainty estimates to point forecasts. Pinson and Kariniotakis [4] have described two complementary approaches that consist of providing forecast users with skill forecasts (commonly in the form of prediction risk indices) or alternatively with probabilistic forecasts. Probabilistic forecasts describe the conditional distribution of wind generation for a given look-ahead time, by quoting some of its quantiles, or with prediction intervals, or alternatively by giving the whole predictive density. For a thorough discussion on the probabilistic forecasting approach, we refer to [5] and references therein. In contrast, focus is given here to the former type of uncertainty indicators. In practice, a skill forecast translates to a comprehensive signal on the confidence one may have in provided forecasts for the coming period, which can be take the form of a colour code (like green-yellow-red for instance) or alternatively by some risk index values (say, from 1 to 5). With that in mind, the present paper aims to have both an academic and a practical value. On the academic side, it shows that it is indeed possible to define some type of skill forecasts in the case of wind power generation, and how this may be done. On the practical side, it shows how such a skill forecasting methodology may be applied to a particular test case, and the type of information forecast users may be provided with. Even if results are presented and discussed for a single test case only, qualitatively similar results would be obtained for other wind farms with different nominal capacities (e.g. future large offshore wind farms for which forecast errors may have significant consequences), or for different wind power forecasting methodologies employed.

It appears that low quality forecasts of wind generation are partly due to the power prediction model, and partly to the Numerical Weather Prediction (NWP) systems. Indeed, during some periods weather dynamics can be relatively more predictable, while at some other point in time they may prove to be unpredictable, and this regardless of the forecasting method employed. Since power predictions are derived from nonlinear transformations of wind forecasts, the level of uncertainty in meteorological predictions may be amplified or dampened through this transformation. Providing forecast users with an *a priori* warning of expected level of prediction uncertainty may allow them to develop alternative (and more or less risk averse) strategies. In an operational context, a skill forecast associated to a given point prediction may be more easily understood than probabilistic forecasts. Also, skill forecasts are not directly related to a given prediction method since they relate to an assessment of the inherent predictability of weather dynamics. For deriving skill forecasts, ensemble forecasts are used as input. An alternative to their use would relate to the works of Lange and Focken [6] towards the definition of weather dynamics indicators. More precisely, they utilize methods from synoptic climatology to classify the local weather conditions based on measurements of wind speed and direction, as well as pressure, and consequently relate them to different levels of forecast uncertainty. Ensemble forecasts of meteorological variables consist of a set of alternative forecast scenarios for the coming period, obtained by stochastic perturbation of the initial conditions of NWP models and possibly stochastic parameterization of these models [7]. Different types of meteorological ensemble predictions are considered here, provided either by the European Centre for Medium-range Weather Forecasts (ECMWF) [8, 9] or by the National Centre for Environmental Prediction (NCEP) [10]. They are converted to ensemble predictions of wind power following the method described by Nielsen *et al.* [11]. Lagged average ensembles, which consist of a set of forecasts with common lead time, but issued at different basetimes, are used as a benchmark.

The paper is structured as following. The various types of ensemble predictions of wind power, which are used as input to the proposed skill forecasting methodology, are introduced in Section 2. The methodology for skill forecasting is then described in Section 3, with focus given to the definition of prediction risk indices and to the way such risk indices will be related to the level of prediction error in a probabilistic manner. The ability of the various risk indices to inform of expected uncertainty is subsequently evaluated and discussed

in Section 4, for the test case of a Danish offshore wind farm over a period of 10 months. The paper ends with concluding remarks in Section 5.

2. Ensemble predictions of wind power

2.1. The meteorological ensemble predictions from ECMWF and NCEP

The ECMWF ensembles are produced from a T255 model, which is spectral model with a truncation at wave number 255. The horizontal resolution of this model is of approximately 80km. 50 ensemble members are associated with the control forecast. They actually go by pairs: an analysis based on singular vectors is used for perturbing the initial conditions in both positive and negative directions, leading to 25 pairs of perturbed members. Singular vectors are the perturbations with the largest energy growth during the first two days [8]. This growth is assumed to be linear in time. The perturbations of the initial conditions are given by linear combination of the singular vectors: because singular vectors tend to be localized in space, they are combined such that they have a more uniform spatial distribution over the considered area. Note that in addition to this representation of the uncertainties in the initial conditions, the ECMWF ensemble prediction system also integrates a stochastic representation of model uncertainties [9]. The ECMWF ensemble predictions for the present study have a maximum look-ahead time of 168-hours ahead (7 days), with a temporal resolution of 6 hours. Calculations are initiated everyday at 12:00. In an operational context, forecasts are so far available 17 hours after initial time.

NCEP meteorological ensembles are obtained by applying the breeding method, which is an alternative to singular vectors for estimating the subspace of fastest growing perturbations [10]. This method simulates the development of growing errors in the analysis cycle. A set of ensemble members is created by adding or subtracting bred modes to the unperturbed analysis. Note that the NCEP ensembles only attempt to sample the analysis error and do not account for model deficiencies. In our case, this set is composed of the unperturbed forecasts plus 10 perturbed members. The horizontal resolution is of approximately 100km (T126 truncation). NCEP ensembles with such a resolution are issued once a day at 00:00. The forecasts have a 6-hour temporal resolution up to 84-hours ahead. An operational advantage of NCEP ensembles with respect to those from ECMWF relates to the fact the methodology employed for generating NCEP ensembles is less computationally demanding, consequently yielding a shorter delay between initial time of computation and actual forecast delivery. More details on both NCEP and ECMWF approaches are given in [13].

2.2. Conversion to ensembles of wind power

Either from ECMWF or NCEP, the meteorological variables of interest are some of the near-surface weather variables, i.e. wind speed and direction at 10m above ground level. Both dataset consist of 300 series of ensemble predictions, for a period covering the first 10 months of 2003. Power predictions are produced with the statistical method introduced, discussed and evaluated in [11]. This method implicitly accounts for the conversion of wind forecasts to hub height and for the conversion of wind as seen by the wind farm to power generation. This is done with semi-parametric statistical models in a data-driven fashion, thus allowing to model such conversion from available data instead of aiming at describing it in a purely physical manner. For a discussion on the differences between statistical and physical approaches to wind power forecasting, we refer to [2, 3]. Because the NWP temporal resolution is of 6 hours, and power measurements are averages over 15 minutes, the meteorological forecasts are linearly interpolated in order to have them corresponding to the available observations. Note that since a model for the conversion of wind forecasts to power forecasts is defined for each horizon individually, the potentially highly correlated nature of consecutive pairs of wind forecasts and power measurements does not affect the training of the model. However, it is obvious that such an interpolation cannot increase the variability of power predictions to the level one could expect from the wind generation process on a 15-minute basis. In addition, this low variability is also due to the coarse spatial resolution of the NWP systems. It may have been a better compromise to convert both series of forecasts and measurements to an hourly resolution, which is so far a more common temporal resolution for day-ahead wind power management or trading. The 15-minute resolution is motivated by the

end-user requirements for the case-study. This difference in temporal resolution however does not affect the results qualitatively, see for instance the study performed in [14]. The wind farm considered is the Tunø Knob offshore wind farm, located few kilometres off the east coast of Jutland in Denmark. It is composed by 10 wind turbines of 500kW (with a hub height of 40.5m), thus yielding a wind farm nominal capacity P_n of 5MW.

Regarding the wind-to-power conversion part of the statistical method employed, it is characterized first by a logarithmic transformation of estimated power values to force these estimates to span the whole range of possible values (i.e. between 0 and P_n). In addition to being a function of wind speed and direction forecasts, the power curve model accounts for the prediction horizon. Accounting for prediction horizon is important owing to the fact that the noise level – in other words, the potential magnitude of forecast errors – in the wind forecasts increase with the forecast horizon, consequently leading to different optimal power curve models. This particular point is further discussed in e.g. [11, 12]. The nonparametric approach for power curve modelling and coefficient estimation is described in [15]. Note that ensemble predictions of wind power are not recalibrated for making them probabilistically reliable. Reliability is related to the probabilistic bias of ensemble forecasts, as a type of systematic error that can be corrected by employing statistical post-processing techniques. Such recalibration however would not change their resolution, i.e. their ability to resolve among situations with different levels of uncertainty [16], which also derives from the inherent value of the ensemble forecasting system. For readers interested in reliability evaluation of those particular wind power ensemble forecasts and methods for their recalibration, we refer to [11, 12]. Hereafter, the ECMWF-based and NCEP-based ensemble predictions of wind generation will be referred to as ECMWF-EPW and NCEP-EPW, respectively. Both power measurements and forecasts are normalized by the wind farm nominal capacity P_n .

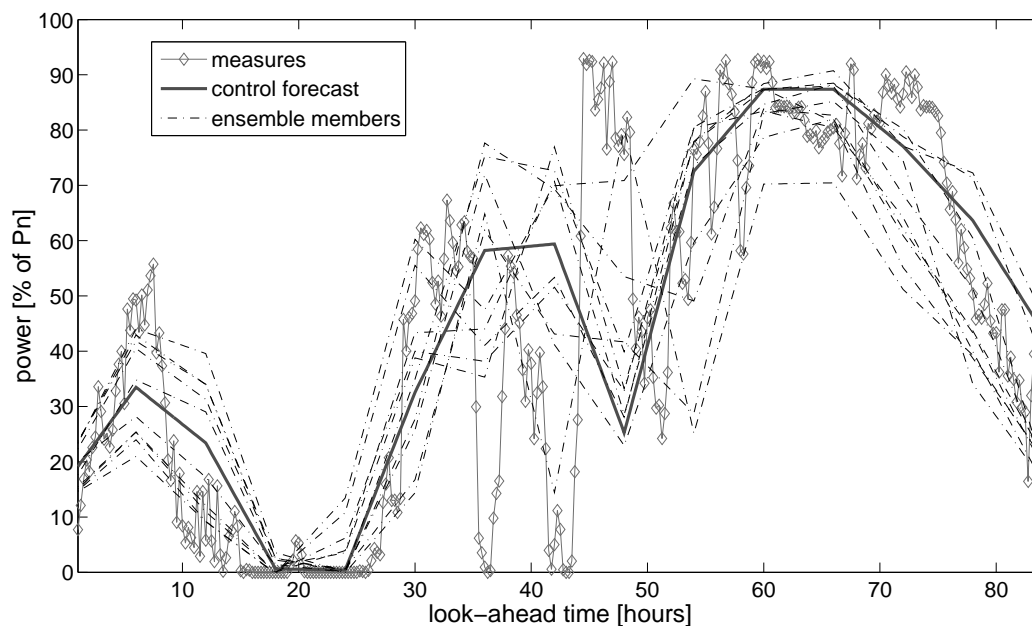


Figure 1: Example of ensemble predictions of wind generation based on NCEP meteorological ensembles.

Fig. 1 displays an episode with point predictions of wind generation produced from a NCEP control forecast (with a maximum look-ahead time of 84-hours ahead), compared with power measurements. The Figure also depicts the related 10 NCEP-EPW ensemble members. The general agreement between the control predictions and measurements is rather good over the whole forecast length, except for prediction

horizons from 35 to 50-hours ahead. In parallel, one notices that the agreement between the ensemble members varies throughout this episode. Both the meteorological ensemble prediction system and the nonlinear model that is applied for converting the forecasts of meteorological variables to power predictions contribute to the variability of the discrepancy among ensemble members. The former reflects the growth of the uncertainty in the estimation of the initial state for the NWP model while the latter amplifies or dampens that growth depending on the level of predicted power, thus accounting for the nonlinear nature of the energy conversion process. For instance, when predicted wind generation is low the envelope containing the alternative predictions is much tighter (horizons between 16 and 26-hours ahead in Fig. 1). And, for look-ahead times from 35 to 55-hours ahead, the disagreement between members is more pronounced, with a criss-crossing of some of these members, while some others stay at a similar level.

2.3. Lagged Average ensembles of wind power

The so-called lagged average ensembles are considered as a benchmark, since they comprise a free alternative to the use of NCEP-based and ECMWF-based ensemble predictions of wind power. Lagged average ensembles consist of a set of forecasts with common lead time, but for different basetimes – the basetime of a forecast corresponding to its time of origin. In other words, these forecasts are obtained from different initial conditions, but by employing the unperturbed prediction model used for issuing control forecasts. Here, lagged average ensembles are made up by lagging the power prediction series estimated from the ECMWF unperturbed forecasts. The maximum look-ahead time for these ensembles is set to 3-day ahead, which is relevant with the current needs for the management or trading of wind generation. Fig. 2 gives an illustrative example of such lagged average ensembles.

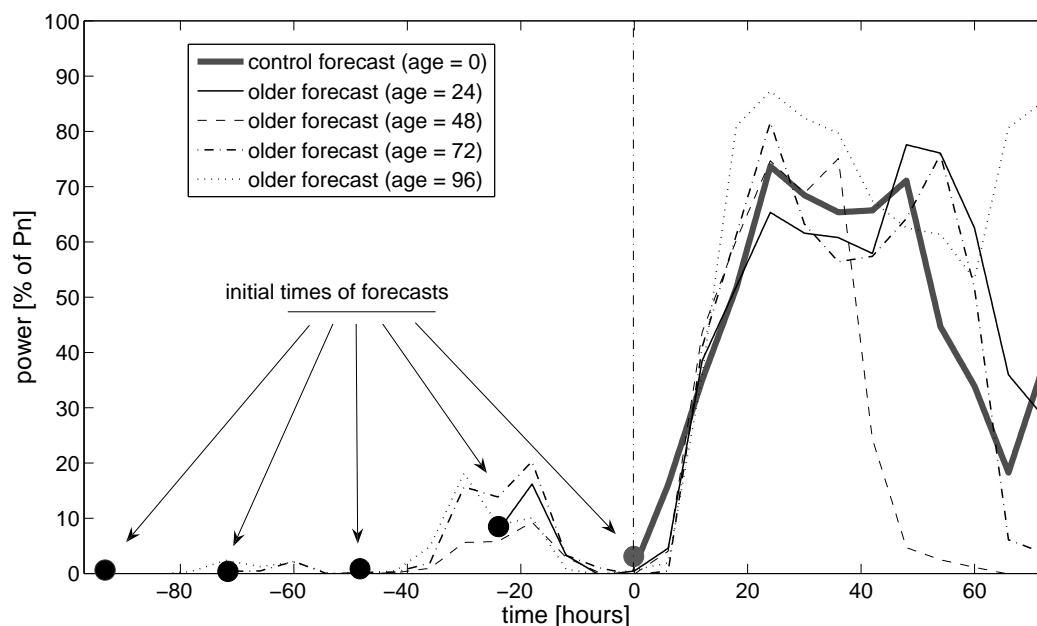


Figure 2: Example of lagged average ensemble predictions of wind generation based on ECMWF 7-day ahead control forecasts.

Since power predictions estimated from the ECMWF control forecasts have a maximum horizon of 7-day ahead and are issued every 24 hours, there are always 5 forecast series overlapping over the following 72 hours. Therefore, the 5 members composing the lagged average ensembles include the last available forecast, plus the forecasts issued 24, 48, 72 and 96 hours before. For the specific case of lagged average ensembles, the term ‘control forecast’ will be used for referring to the most recent power prediction.

3. Skill forecasts based on wind power ensembles

A methodology for skill forecasting based on ensemble forecasts of wind power production is introduced here. First, we propose a definition for prediction risk indices, based on the dispersion of wind power ensembles over a single prediction horizon, or over a set of successive look-ahead times. Then, it is explained how such prediction risk indices may be used as skill forecasts, i.e. forecasts of the distributions of expected prediction errors. The relation between prediction risk indices and the level of prediction error is described with conditional probability diagrams.

3.1. Definition of prediction risk indices

As a result of the observed relation between the spread of ensemble members and the standard deviation of the errors described in e.g. [17], it is proposed hereafter to define prediction risk indices as a measure of the ensemble spread. This measure is a continuous one, in contrast to some categorical measures¹ introduced in the meteorological literature, such as the mode population [18] or the ensemble statistical entropy [19]. Our choice is motivated by the conclusions from Grimit [20], stating that continuous measures of ensemble spread are more appropriate if forecast users have a continuous utility function². We assume that this is the case for users of wind power predictions, either for the management or trading of wind generation.

If the ensemble forecasts of wind power are issued at time t , there are then J alternative predictions $\hat{p}_{t+k|t}^{(j)}$ ($j = 1, \dots, J$) for any lead time $t+k$. The weighted standard deviation $\tilde{\sigma}_{t,k}$ of the set of alternative predictions is used as a measure of spread for that look-ahead time. $\tilde{\sigma}_{t,k}$ is given by

$$\tilde{\sigma}_{t,k} = \left[\frac{J}{J-1} \sum_{j=1}^J w_j \left(\hat{p}_{t+k|t}^{(j)} - \bar{p}_{t+k|t} \right)^2 \right]^{\frac{1}{2}} \quad (1)$$

such that the sum of the weights w_j totals 1, and with $\bar{p}_{t+k|t}$ the weighted mean of the J alternative predictions for that lead time, that is

$$\bar{p}_{t+k|t} = \sum_{j=1}^J w_j \hat{p}_{t+k|t}^{(j)} \quad (2)$$

The weights in Eq. (1) may be interpreted as the ability of the various alternative predictions to inform on predictability. If considering for instance an algorithm that derives a best-guess forecast as a weighted average of the ensemble members, these weights can be directly used in the calculation of $\tilde{\sigma}_{t,k}$. A similar remark is valid if considering lagged average ensembles, for which the weights in the optimal combination of the alternative predictions are a function of their age. Note that if all weights are set to $1/J$ one retrieves the usual definition of the sample standard deviation.

Ensembles forecasts and measurements do not have the same temporal resolution. Even if power predictions are interpolated in order to have a correspondence between these predictions and related measurements, the actual temporal resolution of ECMWF and NCEP meteorological ensembles is of six hours only. In addition, weather predictability does not have an instantaneous nature: it is very unlikely that wind generation would be easily predictable for a given look-ahead time, and then highly unpredictable for the next forecast step. This is the reason why it is envisaged here to estimate predictability over a look-ahead period. Therefore, we generalize the use of the weighted standard deviation by computing the average of $\tilde{\sigma}_{t,k}$ over a set of consecutive horizons, from look-ahead time k_1 to k_2 . This average weighted standard deviation defines a *Normalized Prediction Risk Index*, abbreviated NPRI, calculated as

$$\text{NPRI}(k_1, k_2) := \frac{1}{k_2 - k_1 + 1} \sum_{i=k_1}^{k_2} \tilde{\sigma}_{t,i} \quad (3)$$

¹The basic idea of categorical measures of ensemble spread consists in dividing the range of possible forecast values in several bins, and to count the numbers of ensemble members falling in each bin.

²The utility function for a forecast user is introduced and further discussed by Pinson *et al.* [21].

with $\bar{\sigma}_{t,i}$ introduced in Eq. (1). In the following, NPRI_h denotes the prediction risk index calculated on a per-horizon basis (i.e. such that $k_1 = k_2$) while NPRI_d stands for look-ahead periods of 24 hours. For instance NPRI_d for day 1 is calculated with $k_1 = 1$, meaning the first looking-ahead time available (15-minutes ahead), and $k_2 = 96$, corresponding to the 96th look-ahead time (24-hours ahead). NPRI_d for day 2 and 3 are defined in a similar manner, that is, with $(k_1, k_2) = (97, 192)$ for day 2, and with $(k_1, k_2) = (193, 288)$ for the case of day 3. In contrast, NPRI_h is calculated for each of the 288 look-ahead times available.

3.2. Relating NPRI and prediction errors

3.2.1. Considering prediction errors as energy imbalances

Prediction errors are expressed in the form of energy imbalances since it is aimed to show that NPRI can be used for informing on the level of expected prediction error over a certain period of time. Energy imbalances are defined as the sum of prediction errors in absolute values times the temporal resolution t_r of forecasts and measures (i.e. 15 minutes in our case). The energy imbalance $d(t; k_1, k_2)$ between lead time $t + k_1$ and lead time $t + k_2$ is then given by

$$d(t; k_1, k_2) = t_r \sum_{i=k_1}^{k_2} |p_{t+i} - \bar{p}_{t+i|t}| \quad (4)$$

where $\bar{p}_{t+i|t}$ is the forecast issued at time t for time $t + i$ derived from the set of ensemble members, i.e. the ensemble weighted mean as defined in (2), while p_{t+i} is the corresponding measure.

Both measured and predicted amounts of energy are normalized quantities, since power forecasts and measurements are normalized values (by P_n). By calculating energy imbalances as a total volume of absolute deviations between forecasts and measurements, production surplus and shortage are similarly accounted for. The energy imbalance is not equal to the difference between measured and predicted amounts of energy. Prediction risk indices are meant for estimating the expected level of uncertainty, but cannot give the sign of forecast errors. If considering a single look-ahead time, the normalized imbalance equals the prediction error in absolute value. And, for successive horizons, it is equivalent to the area between the forecast and measurement curves (integrated absolute forecast error).

Prediction risk indices should give information on the expected level of forecast uncertainty whatever the considered point prediction is. Therefore, we do not concentrate hereafter on the use of the best available point forecast of wind generation that can be derived from ensembles, i.e. given by the ensemble mean, or alternatively by the weighted mean for lagged average ensembles [17]. Instead, the considered point predictions are computed in the usual way, i.e. applying a statistical power curve model to the control forecast provided by meteorological centres (here ECMWF and NCEP).

3.2.2. Conditional probability diagrams for relating NPRI to the level of expected prediction error

Following the works by Gritmit [20], the relationship between prediction risk indices and the level of prediction error of the considered point prediction method is derived from a probabilistic perspective. This proposal significantly differs from the more common approach consisting of fitting a linear regressor between the measures quantifying the ensemble spread and the predictor's skill, associated with a correlation coefficient that assesses the strength of this relation (see [22, 23, 24] among others). The inconsistency of using the correlation coefficient for that purpose has been discussed by Gritmit and Mass [23]: considering it for measuring the strength of the relationship between the ensemble spread and the predictor's skill implicitly assumes a linear relation between the spread estimator and the evaluation criterion for the prediction error, which is not true in practice³.

A possibility for expressing the relation between NPRI and the related prediction error in a probabilistic manner is the usage of contingency tables, which give the probabilities of events defined by the occurrence of

³Actually, it has been shown for an ideal ensemble of infinite size that the spread-error correlation can be written analytically, as a function of the temporal variability of the ensemble spread [25]. In this model, the prediction error is in absolute value. For an infinite spread variability, this spread-error correlation has an asymptote of value 0.8 [24].

NPRI-range/error-range pairs. Such an idea has been proposed first by Houtekamer [25] and consequently applied by Whitaker and Loughe [24]. Though, our choice goes for conditional probability diagrams similar to those used by Moore and Kleeman [26], which easily give a visual information on the relation between prediction risk indices and the level of forecast uncertainty.

Conditional probability diagrams summarize the distribution of energy imbalances given NPRI values. The range of NPRI values is divided in categories defined as equally populated classes. This follows from the idea that it is not the value of NPRI by itself that tells if the situation is more or less uncertain, but more where this value is located in the climatological distribution of NPRI values [19, 24]. Also, considering equally populated classes of NPRI values will permit to compare skill forecasts made from ECMWF-EPW, NCEP-EPW or the lagged average ensembles as input, independent of the range of their ensemble spread. A similar reasoning applies for energy imbalances, which are normalized by their climatological value depending on the look-ahead period. This climatological value corresponds to the average imbalance over the 10-month evaluation period for each look-ahead period. When mentioning imbalance levels, they will indeed be relative and expressed in percentage of their climatological value. Thus, we will study how NPRI has the ability to tell if these imbalances are lower or higher than usual, independent of the global performance of the considered point prediction method.

4. Results

The ability of prediction risk indices to differentiate between situations with low and high uncertainty is evaluated here, depending on the use of ECMWF-EPW, NCEP-EPW, or lagged average ensembles as input. This study is for the test case of the Tunø Knob wind farm, over a period covering the first 10 months of 2003. A comparison is made between prediction risk indices defined on a per look-ahead time basis (NPRI_h) or over periods of 24 hours (NPRI_d).

4.1. Pointwise estimation of expected uncertainty

In a first stage, the ability of NPRI to inform on the level of prediction uncertainty when calculated for each look-ahead time is evaluated. This is done for the three sets of wind power ensemble predictions individually. As explained in paragraph 3.1, NPRI_h corresponds to the weighted standard deviation of the ensemble members for a given look-ahead time. The weights for its calculation are set to $1/J$ for ECMWF-EPW and NCEP-EPW (where J the number of ensemble members equals 11 and 51, respectively). Alternatively, the weights given in table 1 are used for the case of the lagged average ensembles, following [17].

Since the amount of data available is fairly limited (only 300 series of wind power predictions, over a 10-month evaluation period), and also for comparison with results of the second part of the study, NPRI and energy imbalance values are gathered for each day ahead. This means for the example of day 1 and for describing the probabilistic relationship between energy imbalances and NPRI_h for the example of day 1, all 96 (energy imbalance)-(NPRI_h) pairs for h between 0 and 24 are put in the same sample dataset. This sample dataset is then used for generating the conditional probability diagram for ‘day 1’.

Table 1: Weights used for calculating the NPRI when considering lagged average ensembles. These weights are those that have been used in [17] for combination of ensemble members in order to obtain an optimal single-valued prediction.

Forecast age [hours]	24	48	72	96
Weight	0.36	0.26	0.21	0.17

4.1.1. The NPRI_h ability to inform on the expected imbalance level

Fig. 3 gives the example of a conditional probability diagram for ECMWF-EPW for day 3 (i.e. for look-ahead times between 48 and 72-hours ahead). It takes the form of boxplots, summarizing the distribution of energy imbalances depending on the class of NPRI_h values. Boxplots are centred on the average NPRI_h

values for each class. Five equally-populated NPRI classes are considered. Considering the number of forecast series in the dataset (300), the number of look-ahead times, the sorting in 3 sample datasets corresponding to days 1, 2 and 3, and finally the division into 5 equally populated classes of NPRI_h values, one obtains for each NPRI_h class an empirical distribution made of 5760 items (in other words, 5760 (energy imbalance)-(NPRI_h) pairs).

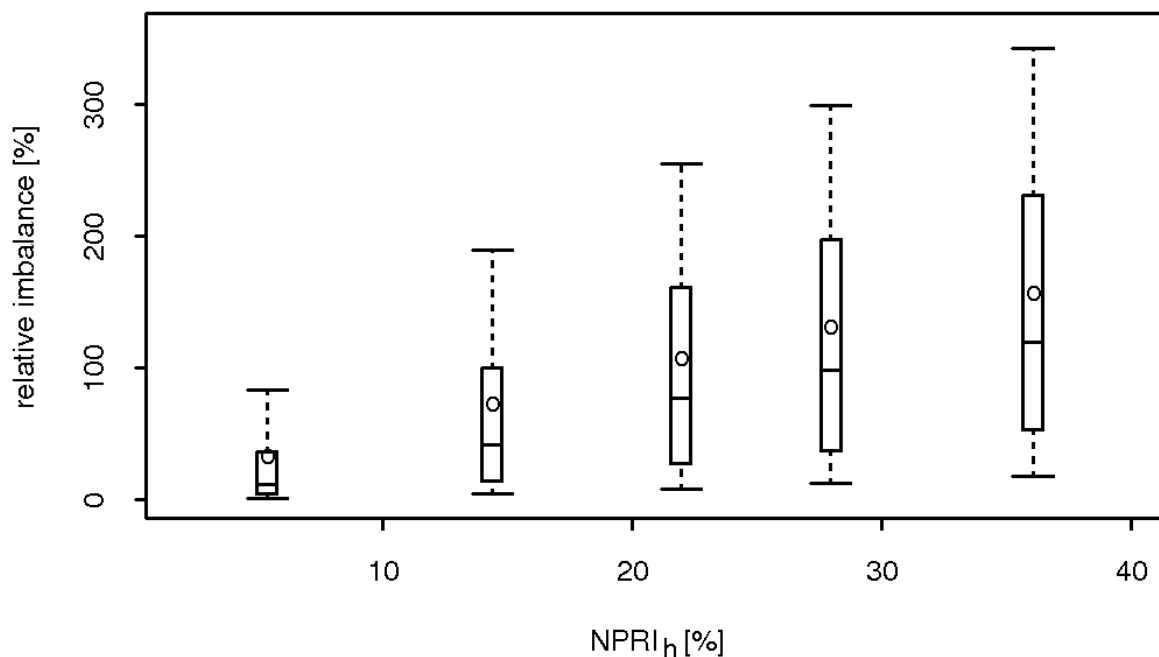


Figure 3: Conditional probability diagram for NPRI_h (from ECMWF-based ensemble forecasts) and for prediction horizons from 48 to 72-hours ahead. Boxplots give the 10% and 90% quantiles (lower and upper tips), the lower and upper quartiles (box bounds), the median (central line) and the mean (o).

The mean values give the general trend between NPRI_h and the level of prediction error. There is a steady (and quasi linear) increase in the mean imbalance level when going from the lowest to the highest NPRI_h class. When NPRI_h values belong to the NPRI_h class 1, the average imbalance level equals 30% of the climatological imbalance level. Though, for class 5, this average imbalance is more than 5 times larger, reaching 155% of the climatological value. Using NPRI_h with ECMWF-EPW proves to be a possibility for resolving between situations with various levels of expected imbalances.

The most interesting information comes from the quoted quantiles of the conditional probability distributions given the NPRI_h class, since it informs of a lower and an upper bound for expected imbalances. In Fig. 3, one sees for instance that if NPRI_h lies in the first class, then 90% of imbalances are below 90% of the climatological imbalance level (for the considered look-ahead time). On the other hand there is still a 10% probability that the level of imbalance exceeds 340% of the climatological imbalance value if the NPRI_h value is in class 5. The imbalance distributions become much wider when NPRI_h values are larger: the 10% quantiles are still pretty close to zero, but the 90% ones get much higher. This upper bound on the expected imbalance level informs on the risk of relying on the provided wind power point prediction. From a risk

aversion point of view, it would be preferable to make conservative decisions if NPRI_h values lie in class 5. Note that here, imbalances are relative to their climatological level in order to better illustrate the fact that prediction uncertainty is lower or higher than usual. Though, in an operational context, expected levels of imbalance can be expressed with physical units (e.g. in MWh or in percentage of the maximum possible generation over the time range).

The example of day 3 has been chosen for describing the relationship between the introduced NPRI applied to ECMWF-EPW and the level of imbalance in a probabilistic manner. The same kind of relation can be observed for the other days, or for the other types of ensembles. Though, that relationship may exhibit slightly different characteristics, which are studied in the following paragraph.

4.1.2. Comparison between ECMWF-EPW, NCEP-EPW and the lagged average ensembles

The relation between NPRI_h (for ECMWF-EPW) and imbalance on a per-look-ahead time basis has been described above. Here, a comparison is made between the information content of the three types of ensemble predictions of wind generation considered. This comparison is possible for the first 3 days ahead. The specific case of day 1 is irrelevant since ECMWF-EPW are only available at the end of this first day in an operational context. Focus is given to day 2 (look-ahead times between 24 and 48-hours ahead) for highlighting the differences between the various wind power ensembles. Similar analyses were carried out for day 3, and the following comments are representative for the whole study.

Even if the NPRI_h values are not scattered in the same manner if considering NCEP-EPW, ECMWF-EPW, or the lagged average ensembles, using classes of NPRI_h values enable the study of the inherent ability of the various ensemble approaches to resolve between situations with low and high uncertainty. These categories of NPRI_h values allow us to leave aside the problem of their distributions and to see how their variations may have an indicative value for estimating the relative level of imbalance. Therefore, when comparing the various approaches, we do not mention ranges of NPRI_h values, but only the NPRI_h class, numbered from 1 to 5. Table 2 gathers some of the quantiles ($r^{(\alpha)}$, with α the proportion) and the mean μ of the conditional probability distributions of imbalances, given the NPRI_h class, for the three types of ensemble predictions.

The variability of the mean imbalance can be seen as a criterion for evaluating the ability of the different approaches to discriminate between several levels of forecast uncertainty. We quantify that variability by the ratio between the mean imbalances for NPRI_h class 5 and 1. This ratio is equal to 5.4, 5.1 and 3.2 for ECMWF-EPW, NCEP-EPW and the lagged average ensembles, respectively. ECMWF-EPW and NCEP-EPW have a higher discrimination ability (with a slight advantage for ECMWF-EPW) compared to that of the lagged average ensembles.

Then, focus is given to the quantiles of conditional imbalance distributions. The increase in the spread⁴ of these distributions when going from the first to the fifth class is more significant for ECMWF-EPW, followed by NCEP-EPW and the lagged average ensembles. This can also be seen as another criterion for stating that ECMWF-based ensembles better resolve among situations, since the variations in the range of expected imbalances are more pronounced. If looking separately at lower ($r^{(0.1)}$ and $r^{(0.25)}$) and upper ($r^{(0.75)}$ and $r^{(0.9)}$) quantiles, one sees that lower quantiles are more variable for NCEP-EPW (from one class to the next) while upper quantiles are more variable for ECMWF-EPW. The first one better resolves situations for which low levels of prediction errors may be expected, while the second better discriminates situations with high forecast uncertainty. Therefore, when having a risk aversion point of view, NPRI_h used with ECMWF-EPW gives a more valuable information on the risk one may face when relying on the provided point prediction.

The increase in the mean imbalance depending on the NPRI_h class is not as steady for the lagged average ensembles than for the others. In addition, the four lowest quantiles decrease between class 4 and 5. If the mean imbalance is higher for NPRI class 5, it is only because this NPRI class contains very large prediction errors. But, it also contains more low prediction errors than the fourth class. Note that this lagged averaging

⁴Here, the spread of the imbalance distributions can be quantified by the inter quartile range, or alternatively by the distance between the quantiles with proportion 0.1 and 0.9. In general, the inter quartile range is preferred, since it comprises a more robust measure of the spread of an empirical distribution.

Table 2: Characteristics of the conditional imbalance distributions given the NPRI_h class for forecast horizons between 48 and 72 hour. $r^{(\alpha)}$ denotes the quantile with proportion α , while μ relates to the mean. Results are for ECMWF-EPW, NCEP-EPW, and the lagged average ensembles respectively. Imbalance values correspond to relative imbalances (in % of their climatological level).

(a) - ECMWF-EPW

NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.3	2.9	10.2	28.7	68.9	27.7
2	4.2	15.2	42.7	89.6	176.7	70.5
3	10.8	30.1	78.4	157.7	277.7	113.7
4	15.2	44.3	113.5	202.7	298.3	137.4
5	16.0	47.5	116.4	223.5	333.6	150.7

(b) - NCEP-EPW

NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.4	4.7	10.8	29.3	84.8	30.4
2	10.5	22.4	40.3	91.3	192.6	75.9
3	17.2	39.8	76.2	146.1	243.4	108.7
4	24.9	56.5	103.3	176.5	275.7	130.93
5	26.8	69.4	135.2	218.7	305.1	154.0

(c) - Lagged Average ensembles

NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.4	3.7	13.0	40.2	99.8	40.0
2	5.1	20.0	57.3	117.0	223.5	89.1
3	9.2	28.9	88.4	171.6	264.4	114.9
4	11.2	35.3	101.8	192.6	290.2	127.6
5	8.7	27.1	76.7	187.8	323.5	128.3

approach, even if less skilful for skill forecasting on a per-step ahead basis, has the great advantage of being a gratis and easily applicable alternative to the use of ECMWF-based and NCEP-based ensemble predictions.

4.2. Estimation of the uncertainty for a look-ahead period

In a second stage, the possibility of providing an estimation of expected uncertainty for a look-ahead period is considered, by calculating NPRI over a set of successive look-ahead times. The benefits of this temporal integration of uncertainty estimation is assessed by looking at the relation between NPRI_d classes and energy imbalances. Both quantities are calculated over 24 hours (thus for 96 consecutive look-ahead times). While Möhrlen [27] discussed the benefits of considering a larger area when assessing the spread-skill relationship of wind power ensembles, the aim here is to show how skill forecasting can benefit from temporal averaging of spread and skill. Also, adding this temporal component would be relevant for real-world applications, since NPRI_d would be related to levels of energy imbalance over the considered time range. Note that current methods for uncertainty estimation of wind power predictions only focus on providing pointwise uncertainty estimates (cf. discussion in [5]).

The 300 series of ensemble predictions over the 10-month evaluation period are considered. Both NPRI_d values and energy imbalances are calculated for look-ahead times between 0 and 24-hours ahead (day 1), 24 and 48-hours ahead (day 2), etc. NPRI_d values are sorted in 5 equally populated classes. In contrast to NPRI_h , there is only one (energy imbalance)-(NPRI_d) pair available per day considered. To each of these classes are associated the empirical distributions of related energy imbalances, which contain 60 items each.

The same quantities than in the above are used for summarizing the characteristics of conditional probability distributions (i.e. mean, median, quartiles and 10 and 90% quantiles).

4.2.1. The NPRI_d ability to inform on the expected imbalance level

Focus is given first to the same example discussed in paragraph 4.1, which relates to the use of NPRI_d with ECMWF-EPW for horizons between 48 and 72-hours ahead. Related conditional probability diagrams are depicted in Fig. 4. As an interpretation of these conditional distributions, one sees for instance that the relative imbalance over day 3 is between 55 and 295% of its climatological level when NPRI_d lies in its fifth class. This same relative imbalance ranges from 5 to 85% of the climatological value only when NPRI_d belongs to the first class.

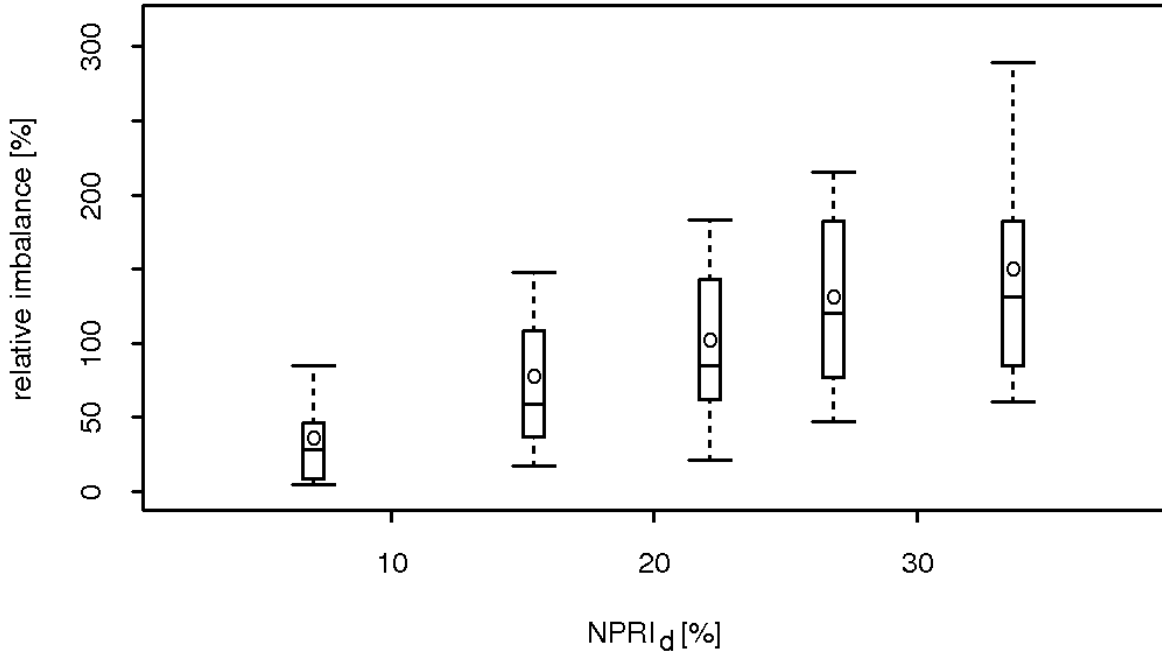


Figure 4: Conditional probability diagram for NPRI_d (from ECMWF-based ensemble forecasts) and for day 3 (look-ahead times between 48 and 72-hours ahead). Boxplots give the 10% and 90% quantiles (lower and upper tips), the lower and upper quartiles (box bounds), the median (central line) and the mean (o).

Similarly to the analysis carried out in section 4.1, the increase of the mean energy imbalance with the NPRI_d class is steady and quasi linear, with mean imbalance levels ranging from 35 to 150% of their climatological value. Therefore, the ability of NPRI_d to be an indicator of the level of expected uncertainty is still valid when considering temporal averaging. In addition, imbalance distributions for every NPRI_d class appear to be sharper than those obtained when using NPRI_h as an uncertainty indicator (cf. Fig. 3). For instance, the inter quartile range for the NPRI class 4 equals 165% in the latter case, while it is only of 105% for the former one. These distributions are sharper first because the upper quartiles are at lower level and also because the lower quartiles are at a higher level (the same remark is valid for the 10% and 90% quantiles). Temporal averaging of skill smoothes the differences between low and large prediction errors.

From a skill forecasting point of view, sharper distributions of expected imbalance are beneficial, since they give more confidence in the estimation of forecast uncertainty.

4.2.2. Comparison between ECMWF-EPW, NCEP-EPW and the lagged average ensembles

The comparison between the three types of ensemble predictions is again carried out for day 2. Table 3 gathers some of the quantiles and the mean of condition imbalance distributions given the NPRI_d class. If no particular mention, the following remarks are also valid for day 1 and day 3.

Table 3: Characteristics of the conditional imbalance distributions given the NPRI_d class for day 2. $r^{(\alpha)}$ denotes the quantile with proportion α , while μ relates to the mean. Results are for ECMWF-EPW, NCEP-EPW, and the lagged average ensembles respectively. Imbalance values correspond to relative imbalances (in % of their climatological level).

(a) - ECMWF-EPW

NPRI_d class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	3.6	8.6	24.4	48.8	88.3	34.9
2	23.5	41.2	61.4	100.8	171.1	76.9
3	29.7	60.9	94.7	140.3	192.7	112.3
4	67.3	91.1	115.0	164.1	203.0	130.1
5	60.1	102.5	125.6	176.4	222.6	144.6

(b) - NCEP-EPW

NPRI_d class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	3.5	10.9	24.0	36.9	71.9	35.8
2	23.8	41.0	80.6	110.4	140.7	81.6
3	53.6	70.9	92.4	140.3	189.7	108.9
4	70.5	87.2	118.8	158.0	198.1	129.1
5	61.8	95.6	134.3	181.0	234.5	143.5

(c) - Lagged Average ensembles

NPRI_d class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	7.1	12.9	30.4	49.1	97.9	41.9
2	8.1	44.3	75.8	118.4	170.1	88.9
3	47.4	77.9	96.8	153.2	229.1	116.3
4	29.9	77.4	115.6	154.8	194.9	117.8
5	55.1	88.9	120.9	176.3	195.0	135.5

The ratios between the mean imbalance for NPRI_d class 5 and NPRI_d class 1 equal 4.2, 4 and 3.2 for ECMWF-EPW, NCEP-EPW and lagged average ensembles, respectively. This ratio is similar to that given when focusing on NPRI_h for lagged average ensembles, while it is significantly lower for the two others. This decrease is mainly due to the smoothing of skill, not to a diminution in the ensembles' ability to resolve among situations. In a general manner, ECMWF-EPW and NCEP-EPW are still more skilful for indicating the expected level of prediction uncertainty. Though, lagged average ensembles gain from the consideration of a temporal component for uncertainty estimation.

Imbalance distributions when considering NPRI_d are much sharper than when considering NPRI_h for the three types of ensembles. The inter quartile range is here between 26 (for NCEP-EPW) and 40% (for ECMWF-EPW) for the first class of NPRI_d values. These values are slightly higher than those in table 3. But for the other NPRI_d classes, it is actually the inverse: the inter quartile range is much lower when imbalance values are sorted depending on NPRI_d values. The reduction of the inter quartile range is up to

50%. Therefore, in terms of skill forecasting, NPRI_d appears to be a better indicator, owing to these sharper distributions.

5. Conclusions

An investigation on the use of ensemble predictions of wind generation revealed its potential for associating skill forecasts to point predictions of wind generation. Focus has been given to the possibility of providing a different information than that provided by probabilistic forecasts. It consists of prediction risk indices, which may be seen as indicators on the predictability of wind generation, and therefore as a comprehensive signal on the confidence forecast users may have on the forecasts provided regardless of the prediction method employed. The prediction risk index NPRI that has been introduced reflects the spread of ensemble members for a single or a set of successive look-ahead times. Various types of wind power ensemble forecasts have been considered: lagged average ensembles (obtained by lagging the ECMWF control forecasts, 5 members), as well as ensembles derived from ECMWF (51 members) and NCEP (11 members) ensemble predictions of meteorological variables. The investigation has been carried out on the case-study of the Tunø Knob wind farm, over a period of 10 months.

The developed methodology considers equally populated classes of NPRI values (more precisely 5 classes), and establishes with a probabilistic approach their relation with distributions of energy imbalances. It has been shown that for all the three different types of wind power ensembles the NPRI could provide useful information on the expected level of forecast uncertainty. An important point addresses the possibility and interest of defining prediction risk indices for a look-ahead period. They permit to inform on the level of expected energy imbalance over the period considered. This is in contrast to the common providing of pointwise uncertainty estimates. Moreover, an important conclusion is also that the gratis alternative of making up wind power ensembles by lagging available point predictions proved to be valuable for estimating the level of expected prediction uncertainty. Considering NCEP-based or ECMWF-based ensembles of wind generation is justified by their better ability to resolve between low and high predictability situations. For the test case of the paper, it appears that the ECMWF-based ensembles of wind generation have higher informative value owing to their higher discrimination ability. This conclusion may not be general though, as the discrimination ability of various types of ensemble forecasts may depend upon e.g. climatic conditions, forecast range, or even the definition of the prediction risk employed.

Perspectives regarding follow-up studies include: *(i)* a validation of the results on various types of test-cases located in areas with different meteorological characteristics (for which predictability may be more or less easily estimated); *(ii)* a further investigation on other possibilities for estimating the disagreement among ensemble members e.g. with categorical measures (mode population and ensemble entropy); *(iii)* a study of other ensemble prediction systems, which may be more appropriate for shorter-range applications than those considered here; *(iv)* the use of such prediction risk indices in forecast combination or regime-switching methods in order to dampen the risk of large prediction errors.

Finally, a last perspective concerns the real-world utilization of prediction risk indices by end-users of wind power forecasts. Such risk indices may be used for instance in maintenance operation planning or for choosing among a set of more or less conservative participation strategies in electricity markets. As a first step, prediction risk indices can be communicated as a complement to point predictions. This way, forecast users will get used to that information, as a signal on the confidence they may have on the provided forecasts. Then, a second step will be to define how to make alternative decisions (more or less conservative depending on the risk aversion of end-users) depending on the value of the prediction risk index, and to demonstrate the resulting operational benefits.

Acknowledgments

The work presented has been performed in the frame of two projects: ‘Wind Power Ensemble Forecasting’ supported by Danish PSO funds (ORDRE-101295/FU2101), and ‘Anemos’ partly funded by the European Commission (ENK5-CT2002-00665), which are hereby greatly acknowledged. The authors also gratefully

acknowledge Elsam Engineering (now part of DONG Energy) for providing the power data, ECMWF and NCEP for providing the meteorological ensemble data. Acknowledgments are finally due to two anonymous reviewers for their helpful comments and suggestions.

References

- [1] Thor S-E, Weis-Taylor P. Long-term research and development needs for wind energy for the time frame 2000-2020. *Wind Energ* 2002;5:73-5.
- [2] Giebel G, Kariniotakis G, Brownsword R. State of the art on short-term wind power prediction. Technical Report, Anemos Project deliverable report D1.1 (available online: <http://anemos.cma.fr>); 2003.
- [3] Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitona E. A review on the young history of wind power short-term prediction. *Renewable Sustainable Energy Rev* 2008; 12: 1725-44.
- [4] Pinson P, Kariniotakis G. On-line assessment of prediction risk for wind power production forecasts. *Wind Energ* 2004;7:119-32.
- [5] Pinson P, Nielsen HAa, Møller JK, Madsen H, Kariniotakis G 2007 Nonparametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energ* 2007;10:497-516.
- [6] Lange M, Focken U. Physical approach to short-term wind power prediction. Berlin: Springer; 2006.
- [7] Palmer TN. Predicting uncertainty in forecasts of weather and climate. *Rep Prog Phys* 2000;63:71-116.
- [8] Molteni F, Buizza R, Palmer TN, Petroliagis T. The ECMWF ensemble prediction system: methodology and validation. *Quat J Royal Met Soc* 1996;122:73-119.
- [9] Buizza R, Miller M, Palmer TN. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quat J Royal Met Soc* 1999;125:2887-908.
- [10] Toth Z, Kalnay E. Ensemble forecasting at NCEP and the breeding method. *Mon Wea Rev* 1997;125:3297-319.
- [11] Nielsen HAa, Nielsen TS, Madsen H, Giebel G, Badger J. From wind ensembles to probabilistic information about future wind power production - Results from an actual application. *Proc IEEE PMAPS Conference, Stockholm, Sweden*; 2006.
- [12] Giebel G, Badger J, Landberg L, Nielsen HAa, Nielsen TS, Madsen H, Sattler K, Feddersen H, Vedel H, Tøfting J, Kruse L, Voulund L. Wind power prediction using ensembles. Technical Report, Risø National Laboratory, Risø-R-1527(EN), Roskilde, Denmark, 2005.
- [13] Buizza R, Houtekamer PL, Toth Z, Pellerin G, Wei M, Zhu Y. A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Mon Wea Rev* 2005;133:1076-97.
- [14] Kariniotakis G, Pinson P. Uncertainty of short-term wind power forecasts - A methodology for online assessment. *Proc PMAPS'04 Conference, IEEE, 'Probabilistic Methods Applied to power Systems', Ames, Iowa (USA)*; 2004.
- [15] Nielsen TS, Nielsen HAa, Madsen H. Prediction of wind power using time-varying coefficient functions. *Proc 15th IFAC World Congress, Barcelona, Spain*; 2002.
- [16] Toth Z, Tallagrand O, Candille G, Zhu Y. Probability and ensemble forecasts. In: Jolliffe IT, Stephenson DB, editors. *Forecast verification: a practitioner's guide in atmospheric science*, New York: John Wiley & Sons; 2003, p. 137-163.
- [17] Pinson P. Estimation of the uncertainty in wind power forecasting. PhD Thesis, Ecole des Mines de Paris, France; 2006.
- [18] Toth Z, Zhu Y, Marchock T. The use of ensembles to identify forecasts with small and large uncertainty. *Wea Forecasting* 2001;16:463-77.
- [19] Ziehmann C. Skill prediction of local weather forecasts based on the ECMWF ensemble. *Nonlin Proc Geophys* 2001;8:419-28.
- [20] Gritmit EP. Redefining the ensemble spread-skill relationship from a probabilistic perspective. NCEP invited presentation, Camp Spring, Maryland (USA); 2004.
- [21] Pinson P, Chevallier C, Kariniotakis G. Trading wind generation with short-term probabilistic forecasts of wind power. *IEEE Trans Power Syst* 2007; 22:1148-56.
- [22] Barker T. The relationship between spread and error in extended range forecasts. *J Climate* 1991;4:733-42.
- [23] Gritmit EP, Mass CF. Measuring the ensemble spread-skill relationship from a probabilistic perspective: stochastic ensemble results. *Mon Wea Rev* 2007;135:203-21.
- [24] Whitaker JS, Loughe AF. The relationship between ensemble spread and ensemble mean skill. *Mon Wea Rev* 1998;126:3292-302.
- [25] Houtekamer PL. Global and local skill forecasts. *Mon Wea Rev* 1993;121:1834-46.
- [26] Moore A, Kleeman R. Skill assessment for ENSO using ensemble prediction. *Quat J Royal Met Soc* 1998;124:557-84.
- [27] Möhrlen C. Uncertainty in wind energy forecasting. PhD thesis, University College Cork, Ireland; 2004.