Wind power prediction risk indices based on numerical weather prediction ensembles

Erik Holmgren, Siebert Nils, Georges Kariniotakis

Paper presented at the 2010 European Wind Energy Conference, Warsaw, Poland, 20-23 April 2010

Wind power prediction risk indices based on numerical weather prediction ensembles

Erik Holmgren*, Nils Siebert** and George Kariniotakis***
MINES ParisTech, CEP - Center for Energy and Processes
BP 207, 06904 Sophia Antipolis Cedex, France
*erik.holmgren@sserenewables.com **nils.siebert@mines-paristech.fr
***georges.kariniotakis@mines-paristech.fr
Phone: +33 493957501 Fax: +33 493957535

Summary

The large-scale integration of wind generation imposes several difficulties in the management of power systems. Wind power forecasting up to a few days ahead contributes to a secure and economic power system operation. Prediction models of today are mainly focused on spot or probabilistic predictions of wind power. However, in many applications, end-users require additional tools for the on-line estimation of the uncertainty of the predictions. One solution to this is prediction risk indices, computed on wind power forecast ensembles derived from numerical weather prediction ensembles. This paper investigates the usefulness of such risk indices as a complement to usual wind power forecasts for informing on the expected level of uncertainty and the risk for large forecast errors. Results show that risk indices are useful to extract information from power ensembles and can give valuable information about the expected prediction uncertainty.

I. Introduction

Wind power is a rapidly growing renewable energy source increasing its share in electricity systems. Since wind power is variable, accurate forecasting of wind farm production up to a few days ahead is required for reliable large-scale integration. Apart from spot forecasts of the wind farm output for the coming hours, of major importance is the development of tools for the online assessment of the uncertainty of these forecasts and the provision of information about the risk for large forecast errors.

For Wind Power Forecasting (WPF) models based on Numerical Weather Predictions (NWPs) it has been found that the accuracy of the power predictions is highly dependent on the accuracy of the NWPs [1]. Nowadays, weather forecasts can be provided as meteorological ensembles consisting of a set of alternative predictions representing different scenarios. Such ensembles can be useful in the context of WPF, for example to construct wind power prediction ensembles. One way to use these WPF ensembles is then to measure the spread of the ensemble members and relate it to the forecast error. This approach was investigated in [1] where the concept of *Risk Indices* as a measure of the spread was introduced.

In this paper, the usefulness of risk indices as a tool to quantify and communicate the uncertainty of wind power forecasts is further investigated. The concepts and definitions introduced in [1] are used as a starting point and several of the perspectives stated there are further explored herein. This includes a validation of the results on various types of test-cases and a further investigation on other possibilities for estimating the disagreement among ensemble members. It also encompasses the utilization of prediction risk indices by wind power forecast users.

The paper starts with a description of meteorological and wind power forecast ensembles in Section II. Furthermore, the concept and the definition of risk indices are introduced in Section III. From this, the methodology for risk index validation and the objectives of the work are described in Section IV. The case study is then described and characteristics for the forecasting models employed are outlined in Section V. A validation of the risk index is made in Section VI, by comparing this case study with previous results. In [1] the index was defined over a look-ahead time window over 24 hours and the impact on the length of this window is examined here. Furthermore, two alternative definitions of risk indices are proposed and examined. The use of risk indices in an operational context is important and this is investigated in Section VIII. Finally, the conclusions are presented in Section IX.

II. Ensemble forecasts

NWP models are highly sensitive for variations in the initial conditions and the model formulation [2]. Besides increasing the spatial and temporal resolution, effort from meteorological institutions is therefore also placed on providing *meteorological ensembles* which consist of sets of alternative predictions representing different scenarios.

In this paper, the ensembles are provided from the Ensemble Prediction System (EPS) of the European Centre for Medium-range Weather Forecasts (ECMWF). These ensembles consist of 51 members out of whom one member is an unperturbed *control forecast* and the other are alternative predictions. The ensembles are calculated using *singular vectors* which represent the most perturbed states in terms of energy growth in the model during the first two days ahead [2]. The ensemble members are then calculated as linear combinations of singular vectors, which imply that the largest possible deviations are taken into account.

A simple way to produce wind power forecast ensembles is to use NWP ensembles and feed them into a WPF model. An example is shown in Figure 1 including point wind power forecasts derived from the 51 ensemble members as well as the observations for the same period.

In the episode presented here, the ensemble members lie relatively close for the first 30 - 40 hours ahead after which the spread is larger for the rest of the look-ahead times. It can clearly be seen that the forecast errors tend to be larger when the spread among the ensemble members grows. This trend was also found in [3] where it was concluded that the average standard deviation of prediction errors increases as the ensemble spread increases.

It has however also been found in previous studies that the spread of the WPF ensemble members does not reflect the whole range of possible outcomes [4]. This trend can be observed in Figure 1 where the observations are in many cases outside the ensemble set. This has been explained to be due to a misrepresentation of the uncertainty in the initial state of the atmosphere or to error growth in the model. The implication is that ensemble forecasts of wind power are not reliable in a probabilistic sense, where reliability is referred to as the probabilistic correctness of the predicted distributions.

In order to produce wind power forecast ensembles with as good probabilistic correctness of the members as possible, it is therefore necessary to use WPF models that preserve the ensemble spread well [5]. The model used in [3] to generate wind power ensembles from wind speed and direction ensembles uses this idea. The model is tuned to obtain the most accurate description of the power curve in terms of low bias and is therefore not optimized to produce as accurate predictions as possible in terms of Root Mean Square Error (RMSE) criterion, which is generally used for model fitting. This causes larger forecast errors compared to state-of-the-art models, but the spread of the ensemble members is better preserved implying that more information can be extracted in the form of risk indices.

III. Risk indices

The idea behind risk indices is to measure the ensemble spread, capture this information in the form of an index and use this to give an indication of the expected level of forecast error. An existing definition is the Normalized Prediction Risk Index (NPRI), introduced in [3]. It is based on the weighted standard deviation, $\widetilde{\sigma}_{t,k}$, of the J ensemble members, $\widehat{P}^{(j)}_{t+k|t}$, at prediction time t for look-ahead time $t+k$:

$$\widetilde{\sigma}_{t,k} = \left[\frac{J}{J-1} \sum_{j=1}^{J} w_j \left(\widehat{P}^{(j)}_{t+k|t} - \overline{\widehat{P}_{t+k|t}} \right)^2 \right]^{1/2} \tag{1}$$

where the sum of the weights, w_j, equals one:

$$\sum_{j=1}^{J} w_j = 1$$

and $\overline{\widehat{P}_{t+k|t}}$ is the mean of the ensemble members for look-ahead time k:

$$\overline{\widehat{P}_{t+k|t}} = \frac{1}{J} \sum_{j=1}^{J} \widehat{P}^{(j)}_{t+k|t} \tag{2}$$

For ensembles where all members are expected to be equally likely, as in the case of ensembles obtained from ECMWF NWPs [6], w_j is set to $1/J$ for all ensemble members. In other cases it can be advantageous to give different weights to different members.

The NPRI is then defined as:

$$NPRI(k_1, k_2) = \frac{1}{k_2 - k_1 + 1} \sum_{i=k_1}^{k_2} \widetilde{\sigma}_{t,i} \tag{3}$$

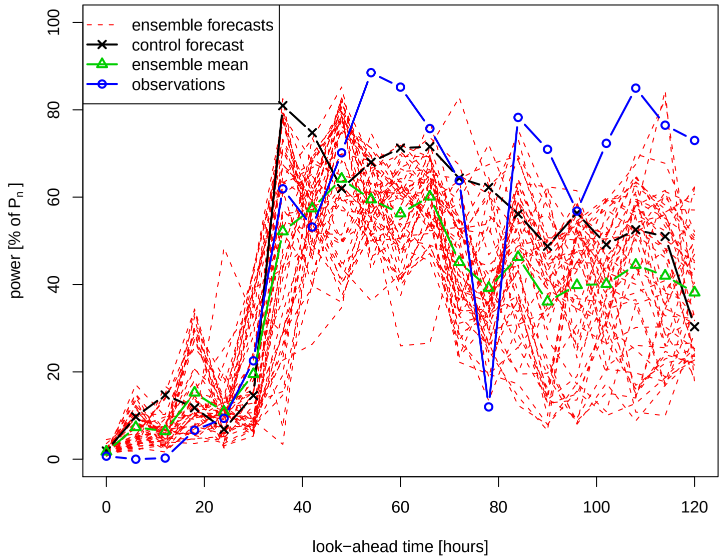


Fig. 1: *Example of power forecast ensembles based on ECMWF meteorological ensembles with corresponding observations. Forecasts are normalized using the nominal power of the wind farm, P_n.*

where k_1 and k_2 are two look-ahead times, with $k_1 \leq k_2$.

This paper focuses on the NPRI given over look-ahead time windows $[k_1, k_2]$. Following the notations in [1], an NPRI given over one day, $k_2 - k_1 = 24$, is noted $NPRI_d$. The idea behind considering look-ahead time windows is that weather, and thereby power predictability, seldom changes significantly within short time periods due to the relatively slow changes of atmospheric processes. Another motivation is that tools for assessing the uncertainty for specific look-ahead times already exist, while tools for look-ahead time windows have not received as much attention.

The NPRI is compared with the total energy imbalance D in the look-ahead time window, $[k_1, k_2]$, given by:

$$D_{t+k_1}^{t+k_2} = t_r \sum_{i=k_1}^{k_2} |P_{t+i} - \widehat{P}_{t+i|t}| \qquad (4)$$

where t_r is the temporal resolution of the wind power predictions.

The energy imbalances are normalized by their climatological mean, $\overline{D}$, given by the average energy imbalance over a longer time period:

$$\overline{D}_{k_1}^{k_2} = \frac{1}{N} \sum_{t=1}^{N} D_{t+k_1}^{t+k_2} \qquad (5)$$

It is however not the value of a risk index in itself that gives information about the uncertainty of the situation but rather its location in the climatological distribution of risk index values [1]. Sorting and dividing risk index values into a number of classes gives the possibility to compute statistics of energy imbalance distributions for each class. For a computed NPRI value, the mean and quantiles of the energy imbalance distributions for the corresponding class can then be used to give information about the probability for different levels of energy imbalances. This can for example be used to give the probability, or risk, for an energy imbalance larger than 150 % of the average imbalance. Such information can be important for end-users of wind power forecasts, for example to develop alternative electricity trading strategies depending on the risk for large energy imbalances.

IV. Methodology for Risk Index Concept Validation

When using risk indices, some criteria for evaluating the quality of the indices are needed. These are presented here followed by a detailed description of the objevtives of the work.

A. Evaluation criteria

In order to evaluate the performance of the NPRI, two criteria are considered; the indices ability to differentiate between high and low levels of energy imbalances and the sharpness of the energy imbalance distributions. In [1], differentiation ability was defined as the Ratio of Mean Imbalance, here noted as RMI, in the last and first class, giving the following expression using 5 classes:

$$RMI = \frac{MI_5}{MI_1} \tag{6}$$

A higher RMI indicates that the index can distinguish better between more and less predictable situations.

Information about the sharpness of the distributions of energy imbalances in each class are given using the quantiles. A robust quantile measure is the Inter-Quartile Range (IQR) which is the difference between the upper and lower quartiles, $Q_{0.75}$ and $Q_{0.25}$:

$$IQR = Q_{0.75} - Q_{0.25} \tag{7}$$

A lower IQR means that the NPRI gives a sharper indication of the expected energy imbalance while a higher value means that the uncertainty of the expected energy imbalance is larger. How the sharpness varies between the 5 classes is given here by IQR_{1-5}. An increase of the IQR with class number indicates that the uncertainty about the energy imbalances increases with risk index value, thus with increasing ensemble spread.

B. Exploration of risk indices

After validating the NPRI approach, a number of explorations can be made.

1) Combination of WPF models: Besides using the same model for producing wind power forecast ensembles to compute risk indices and to compute energy imbalances it is also of interest to use two different models for these two purposes.

2) Other temporal scales: In addition to computing risk indices on a look-ahead time window length of 24 hours, lengths of 12 and 48 hours are also examined.

3) Alternative risk indices: Even if the NPRI is found to work well also for the test cases in this study, it is of interest to examine whether other definitions of risk indices could be useful. Two such alternatives are presented here.

The *MaxMin* index is a simple index that uses the difference between the maximum and minimum values of the J ensemble members, including the control member, for each prediction time t and each look-ahead time k as the measure of the ensemble spread. The index is then used in a similar way as the NPRI by calculating the mean of the differences between maximum and minimum of the ensemble members over a look-ahead time window $[k_1, k_2]$:

$$MaxMin(t, k_1, k_2) = \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \left(\max_j \widehat{P}^{(j)}_{t+k|t} - \min_j \widehat{P}^{(j)}_{t+k|t} \right) \tag{8}$$

A modification of the *MaxMin* index is to calculate the difference between the maximum and minimum value of ensemble predictions over a whole look-ahead time window. This index is defined as the *MaxMinMax* index:

$$MaxMinMax(t, k_1, k_2) = \max_{[k_1,k_2]} \left(\max_j \widehat{P}^{(j)}_{t+k|t} \right) - \min_{[k_1,k_2]} \left(\min_j \widehat{P}^{(j)}_{t+k|t} \right) \tag{9}$$

It can be expected to be useful in order to inform on the risk for large variations in the power output over a look-ahead time window.

C. Using risk indices in decision making

One of the most important aspects of risk indices is their use in an operational context. A proposal of a method to use risk indices in decision making processes is to give warnings or alerts when the probability for a given energy imbalance is larger than a certain value. According to previous findings it is reasonable to give alerts for large energy imbalances when the $NPRI_d$ value is large.

In order to use this approach a criterion for when an alert should be given is needed. The proposal here is to give alerts when there is a probability equal or larger to y for an energy imbalance $D^{t+k_2}_{t+k_1}$ that is x times larger than usual. Mathematically this can be formulated as:

$$f\left(D^{t+k_2}_{t+k_1}\right) = \begin{cases} 1, & p\left(D^{t+k_2}_{t+k_1} > x \cdot \overline{D}^{k_2}_{k_1}\right) > y \\ 0, & p\left(D^{t+k_2}_{t+k_1} > x \cdot \overline{D}^{k_2}_{k_1}\right) \leq y \end{cases} \tag{10}$$

where $\overline{D}^{k_2}_{k_1}$ is the long-term average energy imbalance. The parameters x and y can then be varied depending on the user's sensitivity for large energy imbalances and risk aversion policy.

To investigate whether an alert was correct, or whether it was a wrong decision to not give an alert, a table with the frequencies of correct alerts and correct false alerts can be made. Such tables, sometimes referred to as *confusion matrices* [7], are often used in statistical decision making processes and are organized as shown in Table I.

TABLE I: *A confusion matrix presenting possible errors in statistical decision making processes.*

	Alert needed	**Alert not needed**
Alert made	True alert (TP)	False alert (FP)
Alert not made	False non alert (FN)	True non alert (TN)

The first row of the matrix presents the two cases that can occur when a decision to make an alert is taken. If this decision is found to be justified, this is referred to as a *True Positive* (TP). On the other hand, if this decision turns out to be false, a so called *False Positive* (FP) is generated. When a decision to not make an alert is taken, two cases can also occur. Either an alert was needed, generating a *False Negative* (FN), or the decision to not make an alert is found to be correct, giving a *True Negative* (TN).

Such matrices can in this context be used to compare risk indices from different WPF models, different definitions of risk indices and so forth when giving alerts. The option with largest frequencies of correct decisions, TP and TN, would then be preferable. For end-users of a wind power forecast, different possible decisions incorporate different strategies and the two types of errors could also be associated to different costs. Whether the chosen method gives few FP or few FN can therefore be considered when determining the best approach.

V. Case study

The WPF models, wind farms and data used in the case study are here described.

A. WPF models

The findings presented in Section II motivate a comparison between results obtained by using two different WPF models, one optimized to produce as accurate forecasts as possible, and one that preserves the ensemble spread well. Predictions are therefore made with both an advanced statistical model, *Random Forest* (RF), and a simple physical model, the *Power Curve* (PC) model.

RF is a regression method that uses classification trees to establish the relation between the observations and a set of explanatory variables in order to make predictions. The model is tuned to minimize the forecast error in terms of RMSE [8]. Predictions of wind speed and direction at 10 meter above ground level are used as explanatory variables here together with the last available measurement of wind generation. The control member of the NWPs is used for model learning and power predictions are then derived for all the 51 ensemble members. Since the relations between the explanatory variables vary as a function of look-ahead time, the model is trained and predictions are made for each look-ahead time separately.

The PC model uses the relationship between wind speed and power in the theoretical power curves of the turbines to predict the power output from wind speed predictions. The advantage is that no statistical training is made and the PC model can therefore be directly applied on all the NWP ensemble members. The theoretical power curves of the individual turbines are aggregated to get an approximation of the total power curve for a wind farm. The power curves are provided at the turbine hub height and the predicted wind speeds are therefore first converted to hub height using a logarithmic relation between wind speed u at heights z_1 and z_2:

$$u(z_2) = \frac{ln(z_2/z_0)}{ln(z_1/z_0)} u(z_1) \qquad (11)$$

where z_0 is the roughness length of the site [3].

B. Wind farms and data

In [1] risk indices were investigated on a Danish offshore wind farm. This study is carried out on three French onshore wind farms located in different areas of France with different terrain complexities and local meteorological conditions. Table II presents basic characteristics of these farms.

TABLE II: *Characteristics of the wind farms.*

Farm	**Terrain type**	**# of turbines**	P_n **[MW]**
WF-1	flat	5	12.2
WF-2	complex	9	8.1
WF-3	flat	4	10.12

The data used covers the 18 months from July 2004 to December 2005 with NWPs issued twice a day, at noon and at midnight, giving a total of 1080 records. When using statistical WPF models such as RF that need model learning, it is important to perform training and testing of the model on different and independent data sets. Data containing NWP and power measurements are therefore split into two sets, with a portion of 2/3 of the available time-series used for the learning part and the remaining for testing. This gives 360 wind power ensemble predictions for model evaluation and computation of risk indices.

The temporal and spatial resolutions for ECMWF meteorological ensembles are coarser than for single forecasts. The temporal resolution is 6 hours which is far larger than the resolution of the power measurements of 10 minutes. Some compromise is therefore required. One option is to linearly interpolate the weather forecasts to match the resolution of the power measurements. This is however not preferable since it is rather unlikely that this transformation is representative of the true evolution of the meteorological variables. The opposite option is to keep the 6 hour resolution from the NWP and adapt the resolution of the power measurements to that. The average power output for each hour is therefore computed and the hours corresponding to the NWP data are extracted. This option is therefore mainly considered and an investigation with a temporal resolution of 1 hour, which is more suitable in an operational context, show that the results are in general the same.

The spatial resolution of the NWP data is 1^o in both longitude and latitude corresponding to about 75 - 80 km in East-West direction and 110 km in North-South direction. The NWP are taken from the point in the grid situated closest to the wind farm. This is a relatively simple option but results show that using weighted means of a number of grid points surrounding the farm gives qualitatively similar results.

VI. Results

Before presenting the results in terms of risk indices, the evaluations of the models in terms of average error and preservation of ensemble spread, are presented.

A. Model performance

The difference in model performance in terms of prediction error is shown in Figure 2 where the Normalized Mean Absolute Error (NMAE) is presented for the RF control member, the PC control member, the PC ensemble mean and climatology for 0 to 120 hours ahead for WF-3.

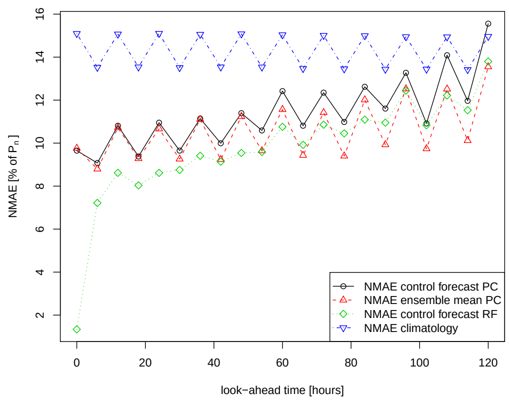
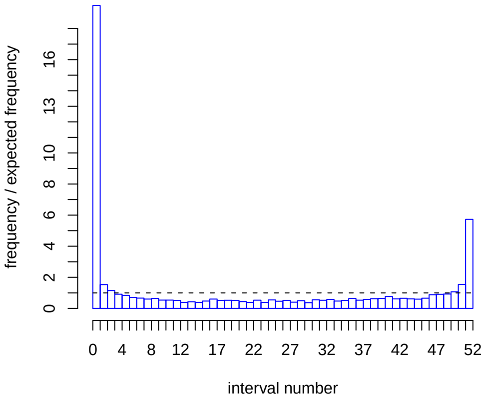


Fig. 2: *NMAE for WF-3 using the Power Curve (PC) model, Random Forest (RF) and climatology.*

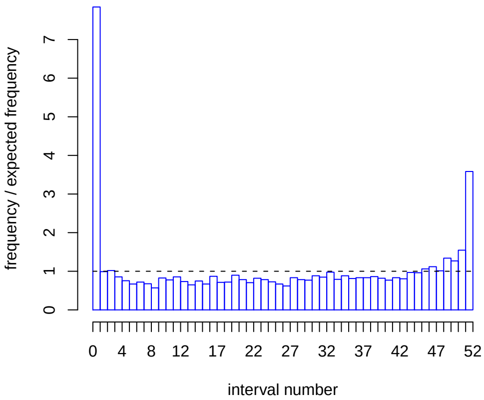
The RF model is found to give lower forecast errors than the PC model, especially up to 60 hours ahead. An increasing improvement with look-ahead time using the ensemble mean for the PC model can also be observed. Similar trends are found for the other two farms.

How well the models preserve the ensemble spread is investigated by studying whether the observations fall with the same rate over the whole set of ensemble members. This information is displayed in *Talagrand diagrams* [2]. In order to perform this, the n ensemble members are sorted according to their value for each run and each look-ahead time and the position of the observation among the ensemble members predictions is determined. For n ensemble members, this gives $n+1$ possible positions for the observation. If the ensemble members where correct in a probabilistic sense and the members represented all possible outcomes well, the observations would be found as many times in each of the $n+1$ positions.

Talagrand diagrams for the RF model and the PC model ensembles are shown for WF-3 in Figure 3. Counts are shown in relative frequencies with a line included to indicate the ideal case with equal number of observations in all positions.



(a) Random Forest



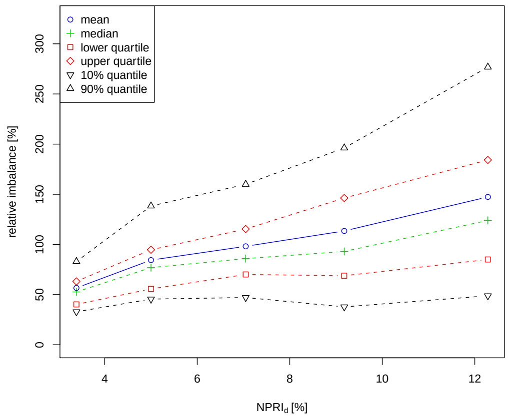
(b) Power Curve model

Fig. 3: *Talagrand diagrams for the predictions made by Random Forest and the Power Curve model for WF-3.*

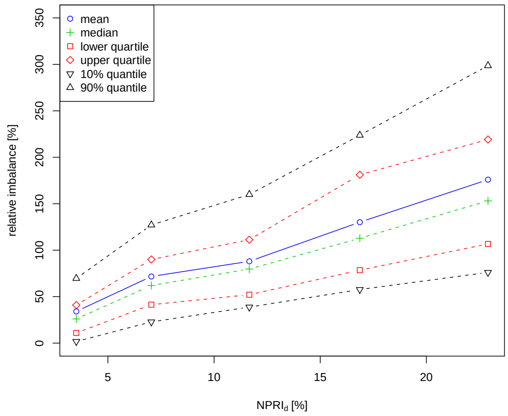
Both diagrams are U-shaped which shows that the observations are found more often outside the range of ensemble members than they should. This trend, which is also often observed for meteorological ensembles [2], confirms that the wind power ensemble predictions are not correct in a probabilistic sense. The behaviour is much stronger for the predictions made by the RF model, illustrating that the ensembles generated by RF have more difficulty predicting low and high power than the PC model. This strengthens the idea that the PC model is a better option in order to get more dispersed ensemble members.

B. Validation of the NPRI

Figure 4 show plots of the $NPRI_d$ for day 3 ahead when using either RF or the PC model. The energy imbalances are computed on the control member of the ensembles.



(a) Random Forest



(b) Power Curve model

Fig. 4: $NPRI_d$ *for day 3 for WF-3 from predictions made with Random Forest and the Power Curve model.*

As observed, the average relative energy imbalance increases almost linearly with class nr, thus with increasing NPRI and increasing ensemble spread. The RMI, which measures the rate of increase, is larger in the PC case than in the RF case, equal to 5.15 compared to 2.50. This shows that risk indices based on the PC model ensembles better distinguishes between low and high energy imbalances. The energy imbalances are however larger, as it was observed in Figure 2 where the forecast errors using the PC model were larger than those using the RF model for the considered look-ahead time window.

The next remark to be made is that the NPRI values are significantly larger using the PC model compared with the RF model. This confirms that the PC model generates more dispersed ensembles.

In terms of sharpness of the distributions, the IQR_{1-5} are slightly lower for RF, equal to 23 - 97 % compared to 30 - 112 % for the PC model. There is a relatively large difference between the classes, with decreasing sharpness for higher classes, which gives a useful distinction in the uncertainty of the expected energy imbalance for different $NPRI_d$ values. It can for example be observed in Figure 4a that the energy imbalance is never above average, 100 %, when the $NPRI_d$ is in the first class while it can be expected to be larger than average when the $NPRI_d$ is in the last class.

The values of the upper quantiles of the energy imbalance distributions are also of interest since they provide information about the probability or risk for large energy imbalances. It can for example be seen in Figure 4a that there is about 10 % risk for an energy imbalance twice as large as normal when the $NPRI_d$ is found in class 4. The risk for such a large energy imbalance is very low for lower classes while it is higher when the $NPRI_d$ is found in class 5.

The results for the other two farms are similar but with slightly lower RMIs for WF-1 and even lower RMIs for WF-2. The IQR_{1-5} are larger for those farms than for WF-3, especially for the first classes. The lower ratios and wider imbalance distributions can be explained by the fact that the level of predictability for WF-1 and WF-2 are lower than for WF-3, with larger average errors for those farms.

The RMIs obtained here can be compared to the ratio of 4.2 obtained in [1]. In terms of sharpness of the distributions, the results are similar to the ones observed in that study. This validates the NPRI approach and shows that the impacts from the model on the average error and the ensemble spread are crucial.

VII. Exploration of risk indices

Even though the NPRI approach can be considered as validated it is interesting to explore it into the different directions outlined in Section IV.

A. Combination of WPF models

The PC model is found to give more useful information on the size of the expected energy imbalance depending on the ensemble spread than the RF model. However, since the energy imbalances are larger, the question is then whether it is preferable to produce forecast that gives more information of the expected errors with the drawback of producing less accurate forecasts. The ideal case would of course be to have a model or a combination of models giving accurate forecasts while still distinguishing well between more and less predictable situations. Since the two models used here have different characteristics it is interesting to study whether they could be combined using energy imbalances from the RF model predictions and compute the NPRI form the PC model predictions.

The results of this investigation are shown in Table III where statistics for $NPRI_d$ for day 2 ahead for the three wind farms using either the RF model or the PC model to compute the $NPRI_d$ are presented. The energy imbalances are computed from the RF model control forecast in both cases.

TABLE III: *$NPRI_d$ using either Random Forest (RF) or the Power Curve (PC) model to generate the ensembles. The energy imbalances are calculated on the standard RF control forecast in both cases. Results are for day 2 ahead.*

	Wind farm	$NPRI_d$ RF	$NPRI_d$ PC
RMI	WF-1	2.67	2.57
	WF-2	2.01	1.76
	WF-3	2.49	2.39
$IQR_{1-5}[\%]$	WF-1	23 - 81	24 - 94
	WF-2	27 - 65	30 - 62
	WF-3	25 - 91	28 - 85

Even though some minor differences between the two options are observed, the performances are very similar with slightly lower RMIs using the PC model for the NPRI and similar IQR_{1-5}s. There seem thus not to be any gain in using the PC model predictions as a complement to the RF predictions. In the remainder of this article, all predictions and NPRI estimations are therefore made using the RF model. It could however still be interesting to investigate other combinations of WPF models in the future.

B. Other temporal scales

When comparing the results in Table III, which shows results for day 2 ahead, with Figure 4, where day 3 ahead were displayed for WF-3, it can be noted that the results for day 2 ahead and day 3 ahead are similar. This shows that the $NPRI_d$ approach is valid for both these windows.

In order to investigate the impact of the length of the look-ahead time window a comparison is made with window sizes of 48, 24 and 12 hours with windows centred around 48 hours ahead. The NPRI is calculated on predictions made with a 1 hour resolution for the three farms and the results are shown in Table IV.

TABLE IV: *Results of NPRI for three different window lengths, all centred around 48 hours ahead.*

	Wind farm	24 - 72 h	36 - 60 h	42 - 54 h
RMI	WF-1	2.12	2.43	2.62
	WF-2	1.65	2.00	2.22
	WF-3	2.17	2.23	2.47
$IQR_{1-5}[\%]$	WF-1	27 - 63	27 - 72	26 - 99
	WF-2	30 - 42	35 - 61	34 - 81
	WF-3	21 - 42	22 - 74	29 - 77

The RMI decreases with longer look-ahead time window giving a smaller distinction between energy imbalances in different classes, a trend that is apparent for all three farms. The distributions also get narrower for longer look-ahead time windows, especially for the higher classes. These results are due to the fact that the energy imbalance and the average spread of ensemble members are evened out over longer time windows. The results show also that using shorter look-ahead time windows than 24 hours could be an interesting option motivating further attention.

C. Alternative risk indices

Statistics of the performances of the *MaxMin* index and the *MaxMinMax* index introduced in Section IV are shown in Table V together with the $NPRI_d$ for day 3 ahead for the three farms.

TABLE V: *Results for $NPRI_d$, the MaxMin index and the MaxMinMax index for day 3 ahead for the three farms.*

	Wind farm	$NPRI_d$	*MaxMin*	*MaxMinMax*
RMI	WF-1	2.08	2.32	1.93
	WF-2	2.01	1.83	1.64
	WF-3	2.50	2.44	2.00
$IQR_{1-5}[\%]$	WF-1	33 - 69	36 - 69	43 - 68
	WF-2	35 - 71	40 - 67	48 - 74
	WF-3	23 - 97	24 - 95	32 - 87

The performances of the two alternative indices are found to be relatively good compared to the $NPRI_d$, especially for the *MaxMin* index with similar performance both in terms of RMI and IQR_{1-5}. The *MaxMinMax* index performs worse than the other two in terms of RMI and has also wider distributions for lower classes.

Since the $NPRI_d$ and the *MaxMin* index show similar performance, it is important to examine whether they are correlated or not. Studying the correlation shows that the two indices are almost equally correlated with the energy imbalance and the relation between the two indices is strong. This implies that the indices measure roughly the same thing and confirms that the distance between the maximum and minimum of the ensemble members gives almost the same information as the standard deviation in informing on the level of energy imbalance. Since the distance between the maximum and minimum value is easier to interpret for an observer than the standard deviation, the *MaxMin* index can be a very useful alternative to $NPRI_d$.

VIII. Using risk indices in decision making

The results of the method for using risk indices in decision making processes, outlined in Section IV, are presented here.

As observed in Figure 4a, it can be motivated to give warnings or alerts for large energy imbalances when the $NPRI_d$ is large. Using the approach presented in Section IV by giving alerts when the probability for an energy imbalance larger than 1.5 times the usual is larger than 0.2 is evaluated on day 2 ahead for WF-3. The results are shown in Table VI with the number of the four possible outcomes. Three fourths (3/4) of the $NPRI_d$ - energy imbalance pairs have been used to build up the imbalance distributions and the approach is evaluated on the remaining quarter, corresponding to 90 cases.

TABLE VI: *Results from issuing alerts when the probability for an energy imbalance larger than 1.5 times usual is higher than 0.2. The results are given for day 2 ahead for WF-3.*

	Alert needed	No alert needed	$\sum$
Alert made	4	1	5
No alert made	15	70	85
$\sum$	19	71	90

Alerts were made in 5 of the 90 cases and this decision was found to be correct in 4 of these. When it was decided to not give an alert, this was correct in 70 of the 85 cases. The results are thus relatively good with particularly a low number of false positives.

Using this decision option on the other farms results in larger shares of alerts and larger portions of FP. The approach of giving alerts is interesting and requires more attention, for example by examining other risk strategies. It is then vital to evaluate the results on longer data sets.

IX. Conclusions and perspectives

This work has been an extension of the work made in [1]. The concept and use of risk indices has been validated and based on the findings presented in this paper, further research in the field can be carried out. Several interesting contributions have been made and many of the ideas proposed and evaluated here merit further investigation.

The main purpose of this work has been to analyze how risk indices, based on wind power forecast ensembles, can be used to give information about the uncertainty of wind power forecasts. The risk indices have been found not only able to give an indication of the expected energy imbalance but also to be useful in giving information about the related uncertainty. The indices are especially found to be useful to give an indication about the risk for large forecast errors.

Wind power forecast ensembles are found not to capture the whole range of possible outcomes. The choice of prediction model, particularly how the ensemble spread is preserved when wind power ensembles are derived from NWP ensembles, is therefore of major importance. Further investigation of the use of more advanced prediction models that still preserves the ensemble spread relatively well is necessary. Also of interest is whether other combinations of models than the one investigated here could give better results.

The extension of the previously used definition of a risk index into other temporal scales show that increasing the scales averages out the risk indices values and makes them less informative. The energy imbalances are however also evened out over larger time windows.

Concerning alternative risk indices, it is found that the more simple *MaxMin* index could be an interesting alternative to the $NPRI_d$.

Finally, it has been found that confusion matrices can be useful to evaluate the ability of risk indices when giving alerts in cases when there is a high risk for large forecast errors. This approach of using risk indices in decision making is promising and merit more attention.

Acknowledgments

The work in this paper has been performed in the frame of the Anemos.plus (FP6, N 038692) and SafeWind (FP7, N 213740) projects funded in part by the European Commission which is hereby greatfully acknowledged. The authors also thank EDF for providing the power data and ECMWF for providing the meteorological ensemble data.

References

[1] P. Pinson, H. Nielsen, H. Madsen, and G. Kariniotakis, "Skill forecasting from ensemble predictions of wind power," *Applied Energy*, vol. 86, pp. 1326–1334, 2008.

[2] M. Ehrendorfer, "Predicting the uncertainty of numerical weather forecasts: a review," *Meteorol. Zeitschrift*, vol. N.F. 6, pp. 147–183, 1997.

[3] P. Pinson, "Estimation of the uncertainty in wind power forecasting," Ph.D. dissertation, Ecole des Mines de Paris, 2006.

[4] H. A. Nielsen, T. S. Nielsen, H. Madsen, G. Giebel, J. Badger, L. Landberg, K. Sattler, L. Voulund, and J. Tøfting, "From wind ensembles to probabilistic information about future wind power production - results from an actual application," Risø National Laboratory, Wind Energy Department Danish Meteorological Institution, Meteorological Research Division Energi E2 A/S Elsam Elkraft A/S, Tech. Rep., 2006.

[5] P. Pinson and H. Madsen, "Ensemble-based forecasting at Horns Rev: Ensembe conversion and kernel dressing," DTU Informatics, Technical University of Denmark, Tech. Rep., 2009.

[6] ——, "Ensemble-based probabilistic forecasting at Horns Rev," DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark, Tech. Rep., 2008.

[7] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[8] N. Meinshausen, "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, pp. 983–999, 2006.