Conditional Prediction Intervals of Wind Power Generation

Pierre Pinson and George Kariniotakis, *Member, IEEE*

Abstract—A generic method for the providing of prediction intervals of wind power generation is described. Prediction intervals complement the more common wind power point forecasts, by giving a range of potential outcomes for a given probability, their so-called nominal coverage rate. Ideally they inform of the situation-specific uncertainty of point forecasts. In order to avoid a restrictive assumption on the shape of forecast error distributions, focus is given to an empirical and nonparametric approach named adapted resampling. This approach employs a fuzzy inference model that permits to integrate expertise on the characteristics of prediction errors for providing conditional interval forecasts. By simultaneously generating prediction intervals with various nominal coverage rates, one obtains full predictive distributions of wind generation. Adapted resampling is applied here to the case of an onshore Danish wind farm, for which three point forecasting methods are considered as input. The probabilistic forecasts generated are evaluated based on their reliability and sharpness, while compared to forecasts based on quantile regression and the climatology benchmark. The operational application of adapted resampling to the case of a large number of wind farms in Europe and Australia among others is finally discussed.

Index Terms—Forecasting, fuzzy inference, nonlinear, nonstationary, resampling, uncertainty, wind power.

NOMENCLATURE

α	Nominal proportion of quantile forecasts.
$\underline{\alpha}, \bar{\alpha}$	Nominal proportions of the lower and upper bounds of prediction intervals.
$\mathcal{A}$	Fuzzy set.
$(1 - \beta)$	Nominal coverage of interval forecasts.
B	Number of bootstrap replications.
c	Forecast condition.
$\mathcal{C}$	Set of forecast conditions.
$\mathcal{D}$	Support of a fuzzy set.
ϵ	Forecast error.
ξ	Indicator variable for quantile forecasts.
$f, \hat{f}$	Probability density function, density forecast.
$F, \hat{F}$	Cumulative distribution function (cdf), estimated/predictive cdf.
h_f	Fuzzy inference model.
i, j, l, b	Common indices.
$\hat{I}$	Interval forecast.
k	Forecast horizon.
n	Size of the error sample S.
N	Size of dataset used for evaluating probabilistic forecasts.
ν	Function permitting to identify the subsets of forecast conditions.
Ω	Set of forecast errors.
Pn	Wind farm nominal capacity.
$q, \hat{q}$	Quantile, quantile forecast.
S	Sample of forecast errors.
Sc	Scoring rule for probabilistic forecast evaluation.
SSc	Skill score for probabilistic forecast evaluation.
t	Time index.
τ	Membership function of a fuzzy set.
$\hat{u}$	Wind speed forecast.
v	Influential variable.
V	Set of values for an influential variable.
w	Weight in the combination of probability density functions.
$y, \hat{y}$	Wind power measurement, forecast.
$z^{(\alpha)}$	Number of hits when evaluating the quantile forecasts with nominal proportion α.

I. INTRODUCTION

FORECASTS of wind power output are traditionally provided in the form of point forecasts. They have the advantage of being easily understandable because this single number is expected to tell everything about future power generation. In practice though, they just tell about the conditional expectation of the power production process for each look-ahead time. It is known that the accuracy of such forecasts is highly variable, and

fairly low on average. One of the priorities in wind energy research in the short to medium-term relates to the improvement of wind power forecasting methodologies [1]. A large part of research efforts is actually still focused on point forecasting only, with the main objective of increasing forecast accuracy. A review of the state of the art in wind power forecasting is available in [2] and [3]. Even though such efforts may lead to a better understanding and modeling of both the meteorological and power conversion processes, there will always be an inherent and irreducible uncertainty in every prediction. This epistemic uncertainty originates from the incomplete knowledge one has of the processes that influence future events [4].

For the communication of forecast uncertainty, [5] has introduced two complementary approaches, which consist of providing forecast users with skill forecasts based on risk indices [6], or with probabilistic forecasts. The present paper focuses on the latter form of uncertainty estimates, which may be either derived from meteorological ensembles [7], [8], based on physical considerations [9], or finally produced from one of the numerous statistical methods that have appeared in the literature; see [10]–[13] among others. They may take the form of quantile, interval, or density forecasts. If appropriately incorporated in decision-making methods, they permit to significantly increase the value of wind generation. Recent developments in the literature support this claim. A first example relates to dynamic reserve quantification, for which the optimal reserve level to be defined by the system operator is a function of the combined uncertainty of load and wind power forecasts, as well as of the outage rates of the conventional power plants [14], [15]. In a similar manner, information about wind power forecast uncertainty has been shown to be beneficial for the optimal operation of combined wind-hydro power plants [16]. Finally if considering the design of trading strategies in liberalized electricity pools, the optimal bid of wind power producers has been shown to be a specific quantile of probabilistic forecasts, which can be determined through stochastic optimization methods [17]. The method introduced here is of the statistical type, and is referred to as adapted resampling since it is inspired by the general class of resampling (and bootstrapping) methods extensively described in [18].

A primary objective that motivated the development of the adapted resampling method relates to the possibility of complementing any wind power point forecasting method with prediction intervals in an operational environment. In that sense, it has an empirical and nonparametric nature: prediction intervals are directly based on past errors made by the point forecasting method considered, without assuming any shape of error distributions. Point forecasts can consequently be "dressed" with a set of prediction intervals with different nominal coverage rates in order to obtain full predictive distributions of wind power. Such an approach also permits to accommodate two important characteristics of the wind power generation process that are its nonstationarity and nonlinearity. More particularly for the last point, the adapted resampling method integrates a fuzzy inference model allowing to additionally rely on expert knowledge of forecast uncertainty characteristics, consequently making the issued prediction intervals conditional to forecast conditions. The background idea of adapted resampling has originally been introduced in [5]. Since then, this method has been fully developed, evaluated on a number of offline and online test cases, and used operationally as an integrated module of the ANEMOS wind power prediction platform, installed in various locations worldwide, including some European countries and Australia among others.

The empirical and nonparametric approach to the estimation of prediction intervals, which comprises the core of the adapted resampling method, is introduced in Section II. The way that expert knowledge is integrated via classification of forecast conditions, and through the design of a fuzzy inference model, is described in Section III. This inference model defines conditional prediction intervals based on a combination of empirical error distributions for various sets of forecast conditions. The method is subsequently applied in Section IV to the test case of an onshore wind farm in Denmark, for which point forecasts are available from three different state-of-the-art methods. Evaluation results include a comparative analysis with a common benchmark (climatology), as well as with another state-of-the-art method, time-adaptive quantile regression (described in [12]). Our experience with application of the method to a number of wind farms worldwide allows us to discuss certain operational aspects in Section V. Concluding remarks end the paper in Section VI.

II. EMPIRICAL AND NONPARAMETRIC APPROACH TO THE ESTIMATION OF PREDICTION INTERVALS

The nonlinearity of the wind power generation process, mainly due to the power curve [9], makes that assumptions that prediction errors follow a known parametric family of distributions appear as weak assumptions. It has especially been shown that such conditional distributions cannot be considered Gaussian [19]. This then motivates the development of nonparametric approaches. In parallel, if using an empirical approach relying on past forecast errors of a point forecasting method of interest only, no assumption is made about the underlying model employed for issuing point forecasts. A nonparametric and empirical approach to forecast uncertainty estimation can consequently be suitable for application to point forecasting methods either of the statistical or physical type, or even if being the result of some combination procedure [12], [13], [21].

An important shortcoming of nonparametric approaches based on quantile regression (as described in [10], [12], [13], and [21]), however, is that a specific model needs to be set-up and trained for each quantile of the predictive distribution to be issued. This may lead to a large number of models for building the whole predictive distributions, thus raising computational costs. In addition, since models are independently trained, this may also yield inconsistent results in certain situations, i.e., crossing quantiles, which is not desirable from both theoretical and practical points of view. In contrast here, the proposed adapted resampling method permits modeling the whole distribution at once, thus avoiding the potential problem of crossing quantiles. Important definitions related to nonparametric probabilistic forecasting are given first, followed by a description of the basis empirical approach.

A. Basics of Nonparametric Probabilistic Forecasting

Denote by y_t the power production measured at time t which corresponds to a realization of the random variable Y_t. Then write f_t and F_t the probability density function and related cumulative distribution function of Y_t, respectively. Formally, provided that F_t is a strictly increasing function, the quantile $q_t^{(\alpha)}$ with proportion $\alpha \in [0, 1]$ of the random variable Y_t is uniquely defined as the minimum value of x such that

$$\mathrm{P}(Y_t < x) = \alpha \tag{1}$$

or equivalently as

$$q_t^{(\alpha)} = F_t^{-1}(\alpha). \tag{2}$$

Then, a quantile forecast $\hat{q}_{t+k|t}^{(\alpha)}$ with nominal proportion α is an estimate of $q_{t+k}^{(\alpha)}$ produced at time t for lead time $t + k$, given the information set Ω_t up to time t.

Interval forecasts (equivalently referred to as prediction intervals) give a range of possible values within which the true effect y_t is expected to lie with a certain probability, its nominal coverage rate $(1-\beta)$, $\beta \in [0, 1]$. A prediction interval $\hat{I}_{t+k|t}^{(\beta)}$ issued at time t for time $t + k$ is defined by its lower and upper bounds, which indeed are quantile forecasts

$$\hat{I}_{t+k|t}^{(\beta)} = \left[\hat{q}_{t+k|t}^{(\underline{\alpha})}, \ \hat{q}_{t+k|t}^{(\bar{\alpha})} \right] \tag{3}$$

whose nominal proportions $\underline{\alpha}$ and $\bar{\alpha}$ are such that

$$\bar{\alpha} - \underline{\alpha} = 1 - \beta. \tag{4}$$

This general definition of prediction intervals makes that a prediction interval is not uniquely defined by its nominal coverage rate. It is thus also necessary to decide on the way they should be centered on the probability density function. Commonly, it is chosen to center (in probability) the intervals on the median, so that there is the same probability that an uncovered true effect p_{t+k} lies below or above the estimated interval. This translates to

$$\underline{\alpha} = 1 - \bar{\alpha} = \frac{\beta}{2}. \tag{5}$$

Such prediction intervals are commonly referred to as central prediction intervals. A discussion on the other types of prediction intervals whose bounds can be defined from (3) and (4) is given in [19].

For a wide range of decision-making problems related to wind power management, a single quantile forecast is not sufficient for making optimal decisions. It is instead necessary to have the whole information about the random variable Y_{t+k} for each look-ahead time. A nonparametric forecast $\hat{f}_{t+k|t}$ of the density function of the variable of interest (i.e., wind power production) can be generated by gathering a set of m quantile forecasts

$$\hat{f}_{t+k|t} = \left\{ \hat{q}_{t+k|t}^{(\alpha_i)} \mid 0 \le \alpha_1 < \cdots < \alpha_i < \cdots < \alpha_m \le 1 \right\} \tag{6}$$

that is, with chosen nominal proportions spread over the unit interval. Such probabilistic forecasts are hereafter referred to as predictive distributions.

B. Description of the Basis Empirical Approach

The development and successful application of empirical-type approaches to prediction interval estimation can be traced back to [22]. Variants of this method have been applied to some other forecasting exercises for which prediction errors proved to be non-Gaussian. Especially, one of those variants has been employed for estimating prediction intervals associated to point forecasts of electricity load produced with a neural-network-based method [23].

The first step consists of collecting forecast errors that the point forecasting method under consideration has made in the recent past, thus allowing to rely on the most recent information about the method's performance. For that purpose, a window in the past (a certain number of hours) is defined and used as a sliding window for storing the errors. When a power measurement is received, it is compared with all the past predictions made for that time. Note that in the following, hourly predictions are considered for the sake of example, without limiting the applicability of the method to different temporal resolutions. The size n of the sliding window determines the size of the samples of errors. At time t, a separate sample $S_{t,k}$ is defined for each prediction horizon k (i.e., for 1-h ahead, 2-h ahead, and so on) since prediction uncertainty significantly varies depending upon look-ahead time. Write $\Omega_{t,k}$ the set of past prediction errors associated to k-step ahead point predictions up to current time t

$$\Omega_{t,k} = \{\epsilon_{t-i+k|t-i}, \ i \in \mathbb{N}, \ i \ge k\} \tag{7}$$

where $\epsilon_{t-i+k|t-i}$ is the normalized prediction error related to the point forecast $\hat{y}_{t-i+k|t-i}$. Since the wind generation process is bounded, we will hereafter only deal with normalized errors and predicted values (both normalized by nominal capacity Pn). Straightforwardly, by renumbering the elements of $\Omega_{t,k}$, an error sample $S_{t,k}$ containing the last n k-step ahead point prediction errors at time t is given as

$$S_{t,k} = \{\epsilon_i \in \Omega_{t,k}, \ i = 1, \ldots, n\}. \tag{8}$$

If a given sample is full when aiming at adding a new error value, the most aged error value is discarded while replaced by the most recent one. This idea is motivated by the nonstationary aspect of wind power prediction errors.

The empirical distribution function $\hat{f}_{t,k}^{\epsilon}$ of errors, at time t and for horizon k, is defined as the discrete distribution that puts probability $1/n$ on each element of $S_{t,k}$. It can be shown that $\hat{f}_{t,k}^{\epsilon}$ is the nonparametric maximum likelihood estimate of the true distribution function of errors $f_{t,k}^{\epsilon}$ (see the definition of nonparametric maximum likelihood in [18, p. 310], as well as the proof of the above), meaning that the likelihood is maximized, but without making any parametric assumption about the error process. Consequently, any parameter $\hat{\theta}(\hat{f}_{t,k}^{\epsilon})$ estimated from $\hat{f}_{t,k}^{\epsilon}$ is the nonparametric maximum likelihood estimate of the

parameter $\theta(f_{t,k}^\epsilon)$. For practical use, we introduce the cumulative distribution function $\hat{F}_{t,k}^\epsilon(x)$, which gives the fraction of errors less than or equal to x

$$\hat{F}_{t,k}^\epsilon(x) = \frac{1}{n}\#\{\epsilon_i \in S_{t,k} \mid \epsilon_i \le x\}. \tag{9}$$

In the above, "#" is the cardinality mathematical operator, which returns the number of elements of the set to which it is applied.

An underlying assumption of the empirical approach is that future uncertainty can be expressed from the recently witnessed behavior of the point prediction method. This translates to saying that the empirical distribution function of errors $\hat{f}_{t,k}^\epsilon$ can be seen as an estimate of the distribution of errors associated to the point forecast $\hat{y}_{t+k|t}$. Therefore, an empirical predictive distribution $\hat{f}_{t+k|t}$ of wind power output at lead time $t+k$ can be constructed as follows:

$$\hat{f}_{t+k|t} \rightarrow \{\hat{y}_{t+k|t} + \epsilon_i, \epsilon_i \in S_{t,k}\} \tag{10}$$

with an equal probability $1/n$ associated to each element of $\hat{f}_{t+k|t}$.

Since the bounds of the central prediction interval $\hat{I}_{t+k|t}^{(\beta)}$ with nominal coverage rate $(1-\beta)$ are defined as the quantiles with proportions $\underline{\alpha}$ and $\bar{\alpha}$ of the predictive distribution $\hat{f}_{t+k|t}$, as defined in (5), they are given by

$$\hat{q}_{t+k|t}^{(\underline{\alpha})} = \hat{y}_{t+k|t} + \hat{F}_{t,k}^\epsilon{}^{-1}(\underline{\alpha}) \tag{11}$$

$$\hat{q}_{t+k|t}^{(\bar{\alpha})} = \hat{y}_{t+k|t} + \hat{F}_{t,k}^\epsilon{}^{-1}(\bar{\alpha}). \tag{12}$$

Such a construction of the predictive distribution $\hat{f}_{t+k|t}$ of wind generation from recent performance implicitly assumes the representativeness of the sample data. This type of assumption is actually made by any type of forecasting procedure. This means that issued prediction intervals may only provide a lower bound on real forecast uncertainty. This will be illustrated in Section IV.

At this stage, one may consider building prediction intervals directly based on empirical distributions themselves, or alternatively by employing kernel smoothing techniques. The former option is preferred here, since for the application considered, empirical distributions can be defined by a large number of data points, thus diminishing the interest of kernel smoothing. In addition, employing kernel smoothing techniques would require the introduction and estimation of an additional parameter, namely the kernel bandwidth. This is not desirable in view of our main objective, which is to propose a model-free and empirical method. Note though that further research may envisage this possibility.

III. CONDITIONAL PREDICTION INTERVALS WITH A FUZZY INFERENCE MODEL

For the specific case of wind power forecasting, a number of variables may influence the characteristics of forecast error distributions, which will be referred to as influential variables. They obviously include predicted power [9], but they may also include predicted wind speed and direction, etc.

Prediction intervals issued with the empirical approach described above would be the same whatever the level of influen-

tial variables: they actually are unconditional interval forecasts. It is unlikely that samples of prediction errors would be representative of the current—and thus conditional—uncertainty. Consider the following illustrative example. Over the previous n hours, all forecasts and measurements were in the low power range, therefore translating to the samples $S_{t,k}$ containing small forecast errors only. At the present time, however, wind power forecasts are in the medium power range, where forecast uncertainty is much higher. Prediction intervals derived from the samples $S_{t,k}$ would therefore be too narrow and not reflect the current forecast uncertainty. As a consequence, it is necessary to propose a more dynamic approach that would be appropriate for issuing conditional prediction intervals.

The idea of classification of forecast conditions is introduced in the first stage, followed by the description of a fuzzy inference model permitting to issue conditional prediction intervals. The fuzzy inference model conceptually defines distributions of prediction errors as a weighted combination of the empirical error distributions obtained for various subsets of forecast conditions. Such combination is finally performed via the use of a multi-sample resampling scheme.

A. Classification of Forecast Conditions

Let us define a forecast condition $c_{t,k}$ at time t and horizon k as a set of values of the considered influential variables. Denote by v^l the lth influential variable (say that we consider L different variables, hence $l = 1, \ldots, L$), and $v_{t,k}^l$ its value at time t and for horizon k. For instance, one can define $v_{t,k}^1 = \hat{y}_{t+k|t}$ and $v_{t,k}^2 = \hat{u}_{t+k|t}$, with $\hat{u}_{t+k|t}$ the wind speed forecast used as input to the point forecasting method. These will be the two influential variables considered in the illustrative example of the mapping of the power curve presented hereafter. It is assumed that influential variables are bounded, since values of meteorological variables or power production must stay in a certain physical range, and can thus be normalized

$$v_{t,k}^l \in V_l = [0,1] \quad \forall l, t, k. \tag{13}$$

Note that forecast errors are also normalized and bounded, though they lie in the range $[-1,1]$.

A forecast condition at time t for lead time $t+k$ is uniquely defined by the association of the values of each of the L influential variables

$$c_{t,k} = \{v_{t,k}^1, v_{t,k}^2, \ldots, v_{t,k}^L\}, \quad c_{t,k} \in \mathcal{C} = V_1 \times \cdots \times V_L \tag{14}$$

where $\mathcal{C}$ is the set of possible forecast conditions at any time t and look-ahead time k.

Then, $\mathcal{C}$ is mapped with a finite number of subsets to which are associated different kinds of characteristics of prediction error distributions. For that purpose, consider J_l ranges of possible values for each of the influential variables v^l ($l = 1, \ldots, L$). Consequently, define $V_l^{j_l}$ the subset of V_l that contains the variable values in the j_lth range. By construction, V_l is the union of all of its subsets

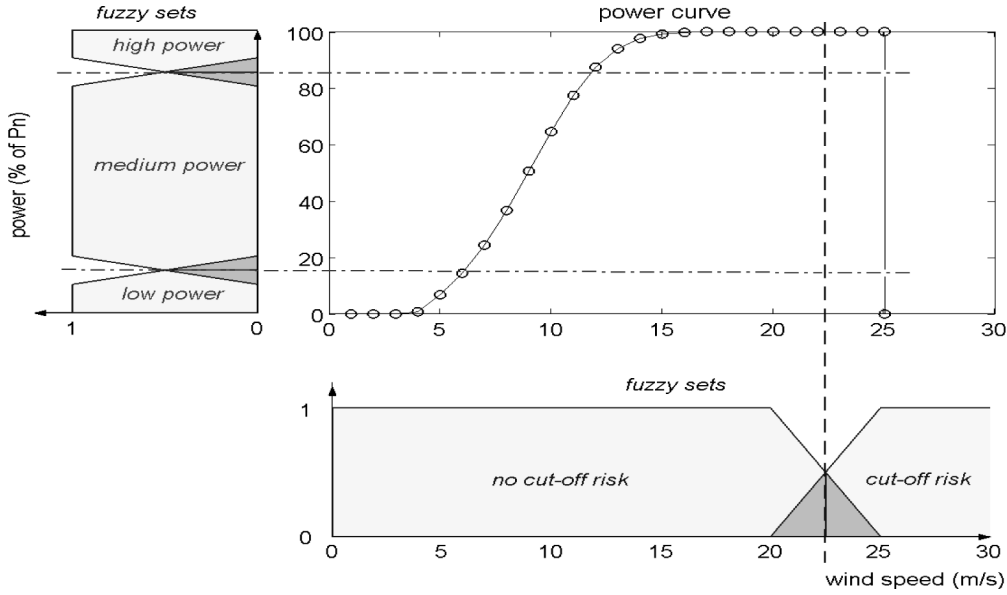$$V_l = V_l^1 \cup V_l^2 \cup \cdots \cup V_l^{J_l}, \quad \forall l \tag{15}$$

Fig. 1. Example mapping of the forecast uncertainty induced by the power curve (normalized by the nominal capacity Pn). The range of possible predicted power values is divided into three ranges ("low", "medium", and "high"), to which are associated three trapezoidal fuzzy sets, in order to account for the nonlinearity induced by the power variable. Similarly, the range of possible forecast wind speed values is divided into two ranges ("no cut-off risk" and "cut-off risk"), owing to the nonlinearity induced by the cut-off, to which are associated two trapezoidal fuzzy sets.

such that none of these subsets are overlapping

$$V_l^i \cap V_l^j, = \emptyset, \quad \forall l, i, j, \ i \neq j. \qquad (16)$$

Now that the sets of possible values for the various influential variables are split into subsets accounting for different characteristics of prediction error distributions, $\mathcal{C}$ can also be split into all possible associations of the subsets for the various influential variables. Write

$$\begin{aligned} \mathcal{C}(\{(l, j_l)\}) &= \mathcal{C}((1, j_1), \dots, (L, j_L)) \\ &= V_1^{j_1} \times V_2^{j_2} \times \cdots \times V_L^{j_L}, \quad \forall j_l \end{aligned} \qquad (17)$$

these subsets corresponding to the j_lth range of values for each of the L different influential variables. This hence yields N_s subsets, where

$$N_s = \prod_{l=1}^{L} J_l. \qquad (18)$$

As an illustrative example, let us again consider forecast wind power and forecast wind speed as influential variables. Three subsets are defined on the range of forecast wind power, while two are only defined on the range of forecast wind speed. Such a situation is represented in Fig. 1, where this classification reflects the influence of both the nonlinear and bounded nature of the power curve and its cut-off. The example subset of forecast conditions $\mathcal{C}((1,1),(2,1))$ corresponds to the case for which both predicted wind speed and power lie in their first subset, and for which forecast uncertainty is expected and known to be low. In parallel, $\mathcal{C}((1,2),(2,1))$ relates to the case where predicted wind speed lies in its first subset, while predicted power is in the medium range. In contrast, forecast uncertainty is known to be significantly high there. A classification of forecast conditions with different related characteristics of prediction error distributions, such as the example one discussed here,

can only be the result of a thorough analysis of the error-generating process. Analyses of forecasting errors are often very informative (see, e.g., [19]), and allow the analyst to gain expertise about the prediction problem at hand. Note that for the case study of Section IV, the level of predicted power will be seen as the unique influential variable, following our analysis of the forecast uncertainty characteristics, and supported by works on forecast error analysis [19] or wind power probabilistic forecasting model building [12], [13].

In order to associate specific characteristics of prediction error distributions to each subset of $\mathcal{C}$, the empirical approach described above is extended by associating a collection of recent prediction errors to each of these subsets. As introduced in (7), $\Omega_{t,k}$ is the set of all the past k-step ahead prediction errors up to time t. Define now $\Omega_{t,k}(\{(l, j_l)\})$ the subset of past prediction errors corresponding to the subset of forecast conditions $\mathcal{C}(\{(l, j_l)\})$

$$\Omega_{t,k}(\{(l, j_l)\}) = \{\epsilon_{t-i+k/t-i} \in \Omega_{t,k} \mid c_{t-i,k} \in \mathcal{C}(\{(l, j_l)\})\}. \qquad (19)$$

Finally, as was done in (8), we can extract from each subset $\Omega_{t,k}(\{(l, j_l)\})$ a sample $S_{t,k}(\{(l, j_l)\})$ of size n containing the last n forecasting errors, but in similar forecast conditions

$$S_{t,k}(\{(l, j_l)\}) = \{\epsilon_i \in \Omega_{t,k}(\{(l, j_l)\}), \ i = 1, \dots, n\}. \qquad (20)$$

Therefore, each of the subsets $\mathcal{C}(\{(l, j_l)\})$ is characterized by its own empirical distribution function $\hat{f}_{t,k}^\epsilon(\{(l, j_l)\})$, drawn from a different sample of past errors. $\hat{f}_{t,k}^\epsilon(\{(l, j_l)\})$ is thus a conditional distribution function since it is an estimate of the distribution function of prediction errors given that $c_{t,k}$ is an item of $\mathcal{C}(\{(l, j_l)\})$. This empirical distribution function puts probability $1/n$ on each element of $S_{t,k}(\{(l, j_l)\})$:

$$\hat{f}_{t,k}^\epsilon(\{(l, j_l)\}) \rightarrow \{\epsilon_i, \ \epsilon_i \in S_{t,k}(\{(l, j_l)\})\}. \qquad (21)$$

B. Fuzzy Inference Model

The previously described classification is the basis for deriving an empirical and distribution-free method that provides conditional prediction intervals, given particular forecast conditions. The choice of the influential variables, as well as the splitting of the sets of possible values into various subsets with different characteristics of related prediction error distributions, are the result of the expertise one has of the process of interest. A fuzzy inference model $h_f(c_{t,k})$ is introduced here, which conceptually defines conditional distributions of prediction errors $\hat{f}_{t,k}^{\epsilon}(c_{t,k})$ given the forecast condition $c_{t,k}$.

Fuzzy logic is an alternative paradigm to that of binary logic. It considers that to each event, a degree of truth being a continuous function between 0 and 1 can be associated. A nice introduction to fuzzy logic theory can be found in [24]. In Section III-A, the set $\mathcal{C}$ of possible forecast conditions has been mapped with several subsets $\mathcal{C}(\{(l, j_l)\})$ related to different forecast uncertainty characteristics. Particularly, a given subset $\mathcal{C}(\{(l, j_l)\})$ is defined as the association of the subsets $V_l^{j_l}$ $(l = 1, \ldots, L)$ for the various influential variables considered; see (17). Here, a fuzzy set $\mathcal{A}_l^{j_l}$ is associated to each of these V-subsets, as is done in the illustrative example of Fig. 1 where the mapping of forecast conditions is based on the known nonlinearities induced by the power curve. A fuzzy set is characterized by a membership function $\tau_l^{j_l}$, which defines the degree of truth of $v_{t,k}^l$ being an element of $V_l^{j_l}$

$$\tau_l^{j_l} : v_{t,k}^l \to \tau_l^{j_l}\left(v_{t,k}^l\right) \in [0, 1]. \tag{22}$$

The subset of forecast conditions $\mathcal{C}(\{(l, j_l)\})$ is defined as the association of the L subsets $V_l^{j_l}$. Therefore, the degree of truth of a given forecast condition $c_{t,k} = \{v_{t,k}^l\}_{l=1,\ldots,L}$ being an element of $\mathcal{C}(\{(l, j_l)\})$ is given by the product of the membership values for every influential variable

$$\tau(c_{t,k}, \{(l, j_l)\}) = \tau\left(c_{t,k} \in \mathcal{C}(\{(l, j_l)\})\right) = \prod_{l=1}^{L} \tau_l^{j_l}\left(v_{t,k}^l\right). \tag{23}$$

The fuzzy inference model introduced here relies on fuzzy rules of the type

"**IF** $v_{t,k}^1 \in \mathcal{D}\left(\mathcal{A}_1^{j_1}\right)$ and $\ldots$ and $v_{t,k}^L \in \mathcal{D}\left(\mathcal{A}_L^{j_L}\right)$
THEN $\epsilon_{t+k|t} \sim F_{t,k}^{\epsilon}(\{(l, j_l)\})$" $\tag{24}$

where $\mathcal{D}\left(\mathcal{A}_l^{j_l}\right)$ stands for the support of the fuzzy set $\mathcal{A}_l^{j_l}$. The "IF" part is referred to as the premise of the rule, whereas the "THEN" part is called the conclusion. Note that the above rule is equivalent to

"**IF** $c_{t,k} \in \mathcal{D}(\mathcal{A}_{\mathcal{C}}(\{(l, j_l)\}))$
THEN $\epsilon_{t+k|t} \sim \hat{F}_{t,k}^{\epsilon}(\{(l, j_l)\})$" $\tag{25}$

where

$$\mathcal{D}(\mathcal{A}_{\mathcal{C}}(\{(l, j_l)\})) = \mathcal{D}\left(\mathcal{A}_1^{j_1}\right) \times \cdots \times \mathcal{D}\left(\mathcal{A}_L^{j_L}\right). \tag{26}$$

Actually, the rule of (25) states that if the forecast condition $c_{t,k}$ can be considered as being an element of a given subset

$\mathcal{C}(\{(l, j_l)\})$ of $\mathcal{C}$, then the prediction error $\epsilon_{t+k|t}$ follows the distribution $F_{t,k}^{\epsilon}(\{(l, j_l)\})$.

Subsequently, a rule base is composed by rules similar to that given by (25), which span all the possible subsets of $\mathcal{C}$. The number of fuzzy rules is hence given by the number of subsets N_s used to map the set of possible forecast conditions. For convenience, an index i is associated to each of the N_s subsets, and we introduce the function $\eta(i)$ that returns the $\{(l, j_l(i))\}_{l=1,\ldots,L}$ pairs that serve to identify the corresponding subset

$$\eta : i \in \{1, \ldots, N_s\} \to \{(l, j_l(i))\}_{l=1,\ldots,L} \tag{27}$$

such that each of the $\{(l, j_l(i))\}_{l=1,\ldots,L}$ pairs is given by a unique value of i. The ith rule of the fuzzy rule base is of the form

"**IF** $c_{t,k} \in \mathcal{D}(\mathcal{A}_{\mathcal{C}}(\eta(i)))$ **THEN** $\epsilon_{t+k|t} \sim F_{t,k}^{\epsilon}(\eta(i))$" $\tag{28}$

where

$$\mathcal{D}(\mathcal{A}_{\mathcal{C}}(\eta(i))) = \mathcal{D}\left(\mathcal{A}_1^{j_1(i)}\right) \times \cdots \times \mathcal{D}\left(\mathcal{A}_L^{j_L(i)}\right). \tag{29}$$

The inference procedure based on this fuzzy logic model consists of applying the rule-base to the forecast condition $c_{t,k}$ in order to provide the overall conclusion as the weighted average of the conclusion of each rule. The weight w_i for each rule is given by the degree of truth of the related premise, normalized by the sum of the weights for each rule

$$w_i(c_{t,k}) = \frac{\tau(c_{t,k}, \eta(i))}{\sum_{i=1}^{N_s} \tau(c_{t,k}, \eta(i))}, \quad i = 1, \ldots, N_s \tag{30}$$

with $\tau(c_{t,k}, \eta(i))$ defined by (23).

By doing so, the fuzzy inference model tells what the contribution of each of the error distributions $f_{t,k}^{\epsilon}(\eta(i))$ in the error distribution $f_{t,k}^{\epsilon}$ is, given the current forecast condition $c_{t,k}$. The fuzzy inference model can finally be written as

$$h_f : c_{t,k} \to \epsilon_{t+k|t} \sim f_{t,k}^{\epsilon} = \sum_{i=1}^{N_s} w_i(c_{t,k}) f_{t,k}^{\epsilon}(\eta(i)). \tag{31}$$

C. Combining Error Distributions With a Multisample Resampling Scheme

Given a specific forecast condition $c_{t,k}$, the inference model h_f defines the distribution $f_{t,k}^{\epsilon}$ of prediction errors $\epsilon_{t+k|t}$ as a combination of several empirical distributions corresponding to the various subsets of forecast conditions. One should then propose a method for such a combination, similarly to the issue of summarizing expert opinions in a single combined distribution; see review of that problem in [25]. With that combination objective in mind, we hereafter employ a method based on a multisample resampling scheme. This method has been shown to be superior to the more classical linear opinion pool for the wind power forecasting application [19]. A description of the linear opinion pool is available in [26].

Write $S = \{\epsilon_j\}_{j=1,\ldots,n}$ a random sample from a probability distribution f. The observations ϵ_j are assumed to be independent and identically distributed (i.i.d.). Following the terminology of the bootstrapping literature (see, e.g., [18]), the plug-in estimate of a parameter $\theta = h(f)$ is defined to be $\hat{\theta} = h(\hat{f})$. This means that we estimate the true parameter of f by applying the same function to the empirical distribution function $\hat{f}$. This is what is performed in (11) and (12) for estimating the lower and upper bounds of the prediction intervals. The elements of S are used for setting up an estimate $\hat{F}$ of the cumulative distribution function associated with f.

Denote by $X = \{x_j\}_{j=1,\ldots,n}$ a random sample that is i.i.d. $U[0,1]$. Probability theory tells us that the sample $\{F^{-1}(x_j)\}_{j=1,\ldots,n}$ is i.i.d. f. Then, the idea of resampling states that since $\hat{F}$ is an estimate of the true cumulative distribution function, one can use it for drawing alternative samples that would lead to other empirical distribution functions of the true distribution f. In practice, this alternative sample $S^{(b)}$ $(b = 1,\ldots,B)$, called bootstrap sample, is obtained by picking randomly and with replacement n values out of the original sample S. $\hat{\theta}^{(b)}$ is a bootstrap replication of the θ statistic. The obtained set of bootstrap replications may then be employed for statistical inference on the true parameter θ.

The basic idea of resampling is employed here for estimating a given parameter θ of a combined probability distribution, by having an interpretation of the combination given by the fuzzy inference model of (31) slightly different from that of the linear opinion pool. Remember that the fuzzy inference model assigns a weight to each of the N_s distributions $f_{t,k}^\epsilon(\eta(i))$. The distributions $f_{t,k}^\epsilon(\eta(i))$ can be approximated by the empirical distributions $\hat{f}_{t,k}^\epsilon(\eta(i))$. The linear opinion pool approach states that these weights can be seen as probabilities and that one can construct a combined distribution by associating these probabilities to each sample. The difference introduced here is that these weights w_i $(i = 1,\ldots,N_s)$ are to be used for defining the share of each of the representative samples of errors $S_{t,k}(\eta(i))$ in a representative sample drawn from the combined distribution. This interpretation leads to generating B bootstrap sample $S_{t,k}^{(b)}$ and compute a bootstrap replication $\hat{\theta}^{(b)}$ for each of them. Given n the size of the error samples, a bootstrap sample $S_{t,k}^{(b)}$ (also of size n) is constructed as follows:

$$S_{t,k}^{(b)} = \left\{ S_{t,k}^{(b)}(\eta(i)) \right\}_{i=1,\ldots,N_s} \tag{32}$$

such that, for $i = 1,\ldots,N_s$

$$S_{t,k}^{(b)}(\eta(i)) = \{\epsilon_j \mid \epsilon_j \in S_{t,k}(\eta(i))\}_{j=1,\ldots,w_i n} \tag{33}$$

where the items of $S_{t,k}^{(b)}(\eta(i))$ are picked randomly and with replacement from $S_{t,k}(\eta(i))$. Note that the number of elements in the subsample $S_{t,k}^{(b)}(\eta(i))$ is $w_i n$, where w_i is the weight given by the fuzzy inference model of (31).

The parameters of interest are the quantiles of the combined probability distribution $\hat{f}_{t,k}^\epsilon$. Write $\hat{F}_{t,k}^{\epsilon,(b)}$ the related cumulative distribution function [following the definition of (9)]. The boot-

strap replications of the lower and upper bounds of the interval forecast $\hat{I}_{t+k|t}^{(\beta)}$ with nominal coverage rate $(1-\beta)$ are given by

$$\hat{q}_{t+k|t}^{(\underline{\alpha})(b)} = \hat{y}_{t+k|t} + \hat{F}_{t,k}^{\epsilon,(b)-1}(\underline{\alpha}) \tag{34}$$

$$\hat{q}_{t+k|t}^{(\bar{\alpha})(b)} = \hat{y}_{t+k|t} + \hat{F}_{t,k}^{\epsilon,(b)-1}(\bar{\alpha}). \tag{35}$$

The bootstrap expectation is subsequently obtained by taking the mean of all the bootstrap replications, and yields an estimate of the interval bounds

$$\hat{q}_{t+k|t}^{(\underline{\alpha})} = \frac{1}{B} \sum_{b=1}^{B} \hat{q}_{t+k|t}^{(\underline{\alpha})(b)} \tag{36}$$

$$\hat{q}_{t+k|t}^{(\bar{\alpha})} = \frac{1}{B} \sum_{b=1}^{B} \hat{q}_{t+k|t}^{(\bar{\alpha})(b)}. \tag{37}$$

By constituting these B bootstrap samples, we actually use all the information included in the individual samples by drawing alternative scenarios. Also, while it is explained in [18, pp. 124–126] that the bootstrap expectation serves for calculating the bias associated to the original estimate of a distribution parameter from a single sample, it has a completely different meaning here, since we apply that form of resampling for a multisample problem. The overall method for issuing conditional prediction intervals of wind power is referred to as adapted resampling owing to its similarities with the original resampling approach.

IV. APPLICATION AND RESULTS

A. Description of Case Studies

The test case of an onshore Danish wind farm located in Klim, North Jutland, is considered. Its nominal capacity is defined as Pn = 21 MW. Series of wind power point forecasts are available from three statistical forecasting methods: Sipreólico [27], Armines Wind Power Prediction System (AWPPS, [28]), and Wind Power Prediction Tool (WPPT, [29]). They use the same meteorological forecasts of wind speed and direction from High Resolution Limited Area Model (HIRLAM) as input. Corresponding wind power measurements are also available. The period covered by these point forecasts and corresponding measurements is from March 1, 2001 until May 2, 2003. For confidentiality reasons, the results presented in the following are not attached to any specific point prediction method. They are related to point prediction methods M1, M2, and M3 instead. This should not be seen as an issue, since what we aim for is to evaluate the set of probabilistic forecasts generated from these point predictions, not the point predictions themselves. Both forecasts and measurements have an hourly temporal resolution. Forecasts are up to 43 and 48 h ahead (depending on the point forecast method), and updated hourly. For consistency, we only perform a comparative evaluation for horizons up to 43 h ahead. Since the first 2000 forecast series are used for initialization of the probabilistic forecasting method, we are left with an evaluation set composed of 16 937 series of wind power point forecasts and associated prediction intervals. Extensive description of this dataset, related results, as well as results from other datasets, are available in [19]. Note that we do not consider any
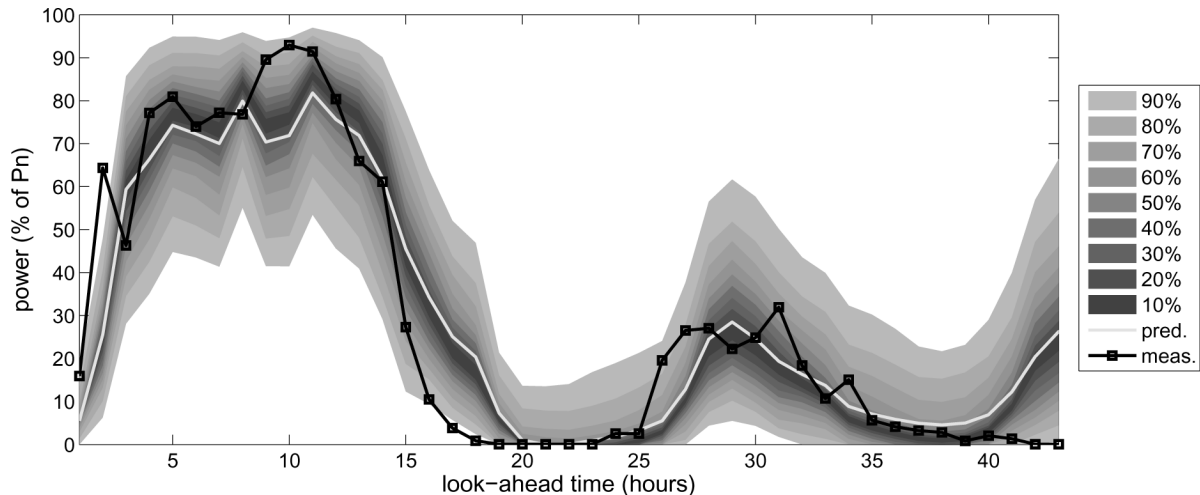
Fig. 2. Example of wind power point predictions associated with a set of interval forecasts. The point predictions are given by M2 and the central interval forecasts are estimated consequently with the adapted resampling method. Nominal coverage rates range from 10% to 90%. These sets of predictions and intervals were issued on January 27, 2001 at 18:00, for the Klim wind farm in North Jutland, Denmark.

benchmark intervals based on a parametric assumption about the shape of error distributions. The superiority of adapted resampling against the assumption of Gaussian and β error distributions has already been discussed; see [19] and references therein. An example comparison of adapted resampling with time-adaptive quantile regression [12], which is another non-parametric method, will be given here. An extensive comparison of these two state-of-the-art methods can be found in [32].

Regarding the mapping of forecast uncertainty, since the dataset available only exhibits very few occurrences of cut-off events (14 occurrences over the whole dataset), it has not appeared appropriate to consider that nonlinearity. This could be done in the future when treating other datasets with more cut-off events. This would only imply a different mapping of the power curve, though the overall methodology would still be the same. However, as for any existing statistical method for wind power probabilistic forecasting, the relative infrequent occurrence of cut-off events may directly impact the quality of resulting probabilistic forecasts. It would actually be difficult to evaluate them anyway, as evaluation would also be based on a few occurrences of such events. In parallel after extensive analysis of the dataset available, we have made the choice to disregard the potential effect of other meteorological variables, like wind direction for instance, on forecast uncertainty. This is because their influence has not been found significant. Therefore, focus is only given to the influence of the level of predicted power on forecast uncertainty. Note though that point forecasts of wind power generation are a function of these meteorological variables. Their effect on forecast uncertainty is thus indirectly considered by relying on wind power point forecasts as influential variable. This is actually in line with some other investigations on probabilistic forecasting of wind power generation, which also considered the wind power point forecast as the main explanatory variable for their probabilistic forecasting models [12], [13].

Following the guidelines given in [19], the range of power values is divided in five equal ranges, associated with trapezoidal fuzzy sets of the form of those depicted in Fig. 1. In

parallel, the error sample size is set to 300 elements, and the number of bootstrap replications to 50. Fig. 2 depicts an episode consisting of a set of point forecasts provided by M2, issued on January 27, 2003 at 18:00 for the Klim wind farm, along with corresponding power measurements. A set of interval forecasts generated with the adapted resampling method is associated to the point predictions, in the form of a fan chart, thus yielding predictive distributions. The nominal coverage rates for these intervals were set to 10, 20, ..., 90%, as will be the case for all prediction intervals evaluated in the following. Following (5) and (6), the resulting predictive densities are then defined by a set of 18 quantile forecasts, whose nominal proportions range from 5% to 95%, with 5% increments, except for the median. Note that the resulting predictive densities are censored densities, meaning that if the forecasted probabilities of events outside of the range of possible generation are not null, such probabilities are then transformed into a probability mass on the related bound, as discussed in [30] and [31] for instance.

It has been explained in Section II-A that emphasis is here on nonparametric central prediction intervals. From the example of Fig. 2, one clearly sees that interval forecasts are not symmetric around the point forecasts, and that for low nominal coverage rates, they may not even cover them. Point forecasts relate to the mean of predictive distributions, while interval forecasts are centered on the median. When the asymmetry of predictive distributions is more pronounced (for low and high predicted power values for instance), the difference between mean and median becomes larger. This is clear here for horizons 3 to 15, or 33 to 43 h ahead. Moreover, the effects of both the lead time and the level of predicted power can be seen from the figure. Prediction intervals are fairly tight for the very first horizons, due to the low level of predicted power and also due to higher forecast accuracy of statistical methods for short-range horizons. Then, they become rather wide when predicted power is in the medium-range: forecast uncertainty is higher in such a case. They then become narrower again for horizons between 20 and 25 h ahead, since predicted power is again at a low level. They finally are slightly wider as forecast horizons get closer to two days ahead.
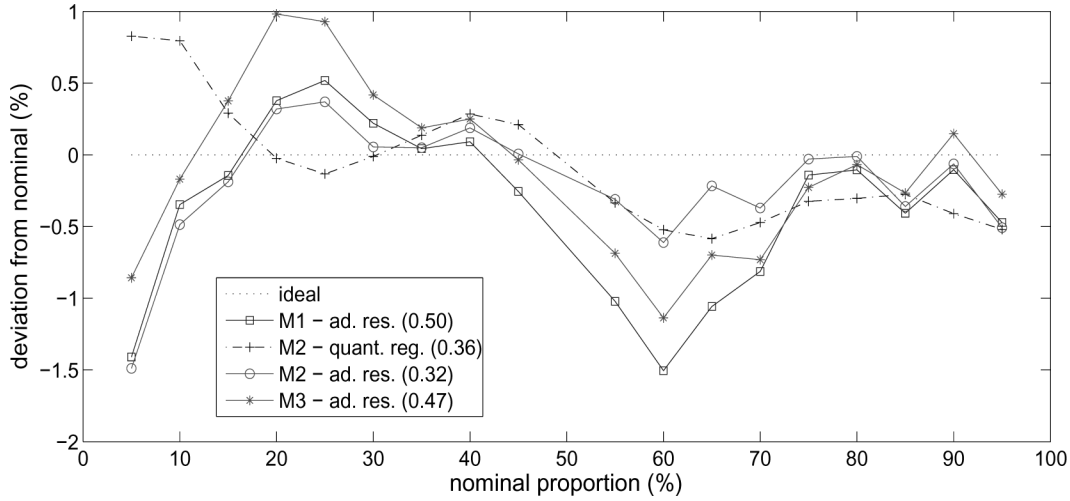
Fig. 3. Reliability diagrams evaluating the reliability of probabilistic forecasts obtained when adapted resampling is applied to the three different point prediction methods M1, M2, and M3. Figures in the legend correspond to the average absolute deviation from the ideal case. Example comparison is made with the case where probabilistic forecasts originate from the application of time-adaptive quantile regression to the point forecasts of M2.

The evaluation framework presented in [32] is employed here for assessing the quality of the predictive distributions obtained with adapted resampling. This framework is somehow based on the paradigm of "maximizing sharpness of predictive distributions subject to reliability" introduced in [33]. Reliability corresponds to the probabilistic correctness of the forecasts, while sharpness relates to the concentration of probabilities, or alternatively to their ability to provide a situation-specific assessment of forecast uncertainty. Focus is given first to reliability, followed by an assessment of the sharpness of predictive distributions, through evaluation of their overall skill.

B. Reliability Evaluation

Following [32], reliability is evaluated as a form of probabilistic correctness with reliability diagrams giving the deviation between observed ($\hat{a}^{(\alpha)}$) and nominal (α) proportions of quantile forecasts—the closest to zero, the better. Remember that these nominal proportions range from 5% to 95%, with 5% increments, except for the median. In practice, if focusing on the quantile forecasts with nominal proportion α, this deviation is defined as

$$\hat{a}^{(\alpha)} - \alpha = \frac{z^{(\alpha)}}{N} - \alpha \qquad (38)$$

where N is the total number of observations, while $z^{(\alpha)}$ is the number of hits, that is, the number of times wind power observations actually lie below the quantile forecasts with nominal proportion α. The reliability diagrams employed here summarize these deviations for the various nominal proportions of the quantile forecasts defining predictive distributions.

Fig. 3 depicts the reliability evaluation results for predictive distributions obtained after application of adapted resampling to point forecasts issued by M1, M2, and M3, and after application of time-adaptive quantile regression to point forecasts issued by M2. These reliability diagrams are for the whole forecast length. This means that all probabilistic forecasts for the various look-ahead times are indifferently used for calculation of

deviations between nominal and observed proportions. In Fig. 3, the x-axis gives the nominal probabilities of quantile forecasts, and the various curves display the deviation (in percent) from "perfect reliability", as expressed by (38). This ideal situation of perfect reliability is represented by the dash-dot straight line. Then, a 1%-deviation for the quantile with nominal proportion 20%, for the example case of M3 in Fig. 3, actually shows that the observed proportion is equal to 21%. Figures in the legend correspond to the average absolute deviation from the ideal case, over the range of nominal proportions (and also over the forecast length). They can be seen as the overall probabilistic bias of these probabilistic forecasts.

One observes from Fig. 3 that the deviations from perfect reliability are contained in a $\pm 1.5\%$ envelope whatever the considered point prediction method, and subsequent probabilistic forecasting method are. In parallel, the average absolute deviation values are contained between 0.32% and 0.5%. Such small differences among deviations may not obviously be due to lower and higher reliability of probabilistic forecasts, but may be due to sampling and serial correlation effects instead [34]. In general, the reliability of the intervals could be expected to be lower for low and high nominal proportions since it is harder to model the very tails of error distributions. This is not the case here. One notices, however, a general trend, which is that quantiles for proportions below 50% tend to be overestimated while those above the median tend to be underestimated. Prediction intervals are slightly too narrow on average. This goes along with our comment in Section II-B such that methods of estimating future uncertainty usually rely on past experience of a given model performance and therefore do not integrate the additional uncertainty of predicting new data. In a global manner, it may still be appraised that evaluated probabilistic forecasts appear sufficiently reliable (and this whatever the point forecasts used as input to adapted resampling), thus allowing us to move on to sharpness evaluation. An important point is that both adapted resampling and quantile regression appear to have a similar, and acceptable, level of reliability.
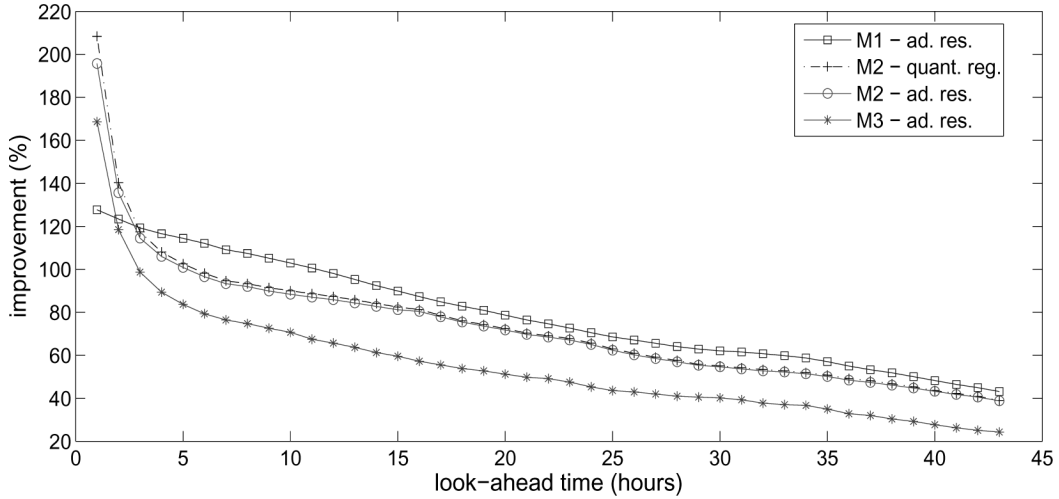
Fig. 4. Skill score as a function of the horizon, for probabilistic forecasts obtained when adapted resampling is applied to the three different point prediction methods M1, M2, and M3. Example comparison is made with the case where probabilistic forecasts originate from the application of time-adaptive quantile regression to the point forecasts of M2.

C. Sharpness Evaluation

When reliability of probabilistic forecasts has been assessed, a sharpness evaluation of these forecasts can be performed through the study of their overall skill. Skill encompasses all aspects of probabilistic forecast quality. Following [32], the skill of predictive distributions is studied based on a proper scoring rule which, for a given predictive distribution $\hat{f}_{t+k|t}$ and corresponding realization y_{t+k}, writes

$$\mathrm{Sc}(\hat{f}_{t+k|t}, y_{t+k}) = \sum_{i=1}^{m} \left(\xi^{(\alpha_i)} - \alpha_i \right) \left(y_{t+k} - \hat{q}_{t+k|t}^{(\alpha_i)} \right) \quad (39)$$

with $\xi^{(\alpha_i)}$ an indicator variable for the quantile forecast $\hat{q}_{t+k|t}^{(\alpha_i)}$, being equal to 1 if the observation is below $\hat{q}_{t+k|t}^{(\alpha_i)}$, and 0 otherwise. Over an evaluation set with N forecast series, the above scoring rule can be applied to evaluate the skill of predictive distributions as a function of the look-ahead time k, with

$$\mathrm{Sc}_k = \frac{1}{N} \sum_{t=1}^{N} \mathrm{Sc}(\hat{f}_{t+k|t}, y_{t+k}). \quad (40)$$

This scoring rule is positively oriented—the higher, the better, and admits a maximum value of 0 for perfect probabilistic predictions.

In order to appraise the actual skill of predictive distributions, they should be evaluated with respect to a relevant benchmark. The most common benchmark for probabilistic forecasts of weather variables and weather-related processes (such as wind power generation) is climatology. The climatology predictive distribution is given by the distribution of all available observations, and is thus a unique and unconditional probabilistic forecast. Note that the climatology probabilistic forecasts are by definition probabilistically reliable. If denoting by Sc_k^0 the scoring rule value for climatology for a given look-ahead time k, one then defines the skill score

$$\mathrm{SSc}_k = 100 \frac{\mathrm{Sc}_k^0 - \mathrm{Sc}_k}{\mathrm{Sc}_k^0} \quad (41)$$

giving the improvement (in percent) of the probabilistic forecasting method considered over the climatology benchmark.

Fig. 4 depicts the evolution of this skill score as a function of the look-ahead time, for predictive distributions obtained with adapted resampling (with input the point forecasts from M1, M2, and M3) and with quantile regression (with input from M2). The skill score steadily decreases with the look-ahead time, meeting the general statement that it is harder to predict for lead times further in the future, as is also the case for point forecasts. The improvements over climatology are positive over the whole forecast length, highly significant for short look-ahead times, and steadily decreasing for further ones. An interesting point is that the choice of a given point prediction method as input has an influence on the quality of the resulting predictive distributions. It appears that considering M1 leads to better probabilistic forecasts for the very first look-ahead times, whereas considering M2 is better for further ones. One also sees that the difference between adapted resampling and quantile regression is negligible (the lines are almost on top of each other). This is while adapted resampling has the advantage of being model-free, and only requiring a single set-up to produce full predictive densities, in contrast to the 18 models that are needed to be proposed for quantile regression (i.e., one for each of the quantiles).

Note that since predictive distributions actually relate to estimates of the error distributions of the point forecasts, a point prediction method with sharper error distributions will yield sharper probabilistic forecasts. This comment is of course valid only if the prediction interval estimation approach has a real ability to reflect the error distribution associated to a given point forecast. For the test case of the present paper, we observe a significant difference in skill among the probabilistic forecasts generated from the various point forecasts used as input. In addition, in view of comparison with other application test cases documented in [19] for instance, it appears that one may also expect significant differences in the skill of probabilistic forecasts depending of the site characteristics, in terms of terrain, size of the wind farm(s), offshore or onshore conditions, etc., which can

be explained by inherently higher or lower predictability at this particular site.

V. DISCUSSION ON OPERATIONAL ASPECTS

The developed approach has been implemented into an operational module and integrated in the ANEMOS wind power forecasting system developed in the frame of an EU-funded project under the same [35]. This module produces prediction intervals for point predictions coming from classical wind power prediction models or for the combination of such predictions. The advantage of this two-step approach is that it permits to maintain the classical deterministic physical or statistical models within the modeling chain. These point forecasting models are especially designed and trained, and are the result of several years of research and improvements. An alternative approach would be a one-step approach where the modeling chain contains only one model that directly generates the whole probability density function of wind production for each look-ahead time. This two-step approach is also straightforwardly applicable to produce prediction intervals for the aggregated production of a group of wind farms which is often an operational requirement. This is done by considering the aggregated wind power prediction as input to the uncertainty module. Such a solution is preferable to that consisting of generating probabilistic forecasts for each individual wind farms and then aggregating the estimates. This is because this aggregation would then correspond to summing nonparametric densities of interdependent random variables (both spatially and temporally), which can be a complex task. Note that some developments towards the understanding and modeling of complex temporal and spatial interdependence structure in wind power forecast uncertainty have been initiated in [36].

The module has been operationally applied to several real world cases including SONI, the TSO of Northern Ireland, U.K., and the Market and System Operator of Australia AEMO (ex Nemmco), as well as others. In these cases, prediction intervals are produced for all wind farms in the area of these operators as well as for several aggregation levels. An important issue that appears in an operational environment is how to handle situations where there is "down-regulation" of power (i.e., controlled reduction of the wind farm output) due to network or other constraints. The consideration of scheduled maintenance of wind turbines within a wind farm is also a challenge. A solution can be to proceed with a simple scaling of the prediction intervals, this scaling being influenced by the operational constraints, or by the number of wind turbines out of order. In practice, this translates to considering that the wind farm (or group of wind farms) has a potentially varying nominal capacity.

VI. CONCLUSIONS

Adapted resampling is a generic method suitable for the on-line assessment of forecast uncertainty of state-of-the-art wind power point forecasting methods in operational conditions. It permits to provide conditional prediction intervals based on recent observed accuracy of a point forecasting method of interest, while accounting for expert knowledge of forecast uncertainty characteristics. The method relies on a classification of recent

forecast errors, a fuzzy inference model, and a multisample resampling scheme for combination of probability distributions. A great advantage of this method is that it permits to "dress" point forecasts with predictive distributions informing of their situation-dependent uncertainty, in a consistent and computationally inexpensive manner.

The quality of the resulting probabilistic forecasting approach has been illustrated based on the test case of an onshore Danish wind farm, while it has been explained that other evaluation results exist for a number of offline case-studies, as well as for a large number of operational ones. This is due to the fact that adapted resampling is integrated in the form of a module in the ANEMOS wind power forecasting platform, installed and operated in a number of European countries and in Australia. Adapted resampling appears to be (probabilistically) reliable. In parallel, the sharpness of predictive distributions issued from this method is bounded by the quality of the point predictions they complement. This follows from the fact that probabilistic forecasts are based on the modeling of the point forecast error distributions. The sharper these distributions, the sharper the resulting probabilistic forecasts. As a consequence, it may be advised that future methods for probabilistic forecasting of wind power generation do not rely on point forecasts as input, but directly consider the modeling of predictive distributions from meteorological conditions.

It is unlikely that in the near future, probabilistic forecasts will be fully integrated into the wide range of decision-making problems related to wind power management, trading, and integration into the electricity network. This is even though it can be shown that for a large class of decision-making problems, decisions based on traditional point forecasts only prove to be suboptimal. Efforts in that direction will have to be concentrated on appropriate research, demonstration of resulting benefits, and transfer of knowledge via training of relevant personnel.

ACKNOWLEDGMENT

REFERENCES

[1] S.-E. Thor and P. Weis-Taylor, "Long-term research and development needs for wind energy for the time frame 2000–2020," *Wind Energy*, vol. 5, no. 1, pp. 73–75, 2002.

[2] G. Giebel, G. Kariniotakis, and R. Brownsword, State of the Art on Short-Term Wind Power Prediction, Tech. Rep., Anemos Project Deliverable Report D1.1, 2003. [Online]. Available: http://anemos.cma.

[3] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitona, "A review on the young history of wind power short-term prediction," *Renew. Sust. Energy Rev.*, vol. 12, pp. 1725–1744, 2008.

[4] T. R. Stewart, "Uncertainty, judgement, and error in prediction," in *Prediction: Science, Decision-Making, and the Future of Nature*, D. Sarewitz, R. A. Pielke, and R. Byerly, Eds.   Washington, DC: Island Press, 2000, pp. 41–57.

[5] P. Pinson and G. Kariniotakis, "On-line assessment of prediction risk for wind power production forecasts," *Wind Energy*, vol. 7, no. 2, pp. 119–132, 2004.

[6] P. Pinson, H. A. Nielsen, H. Madsen, and G. Kariniotakis, "Skill forecasting from ensemble predictions of wind power," *Appl. Energy*, vol. 186, no. 7–8, pp. 1326–1334, 2009.

[7] G. Giebel, Wind Power Prediction Using Ensembles Technical Report, Risø-R-1527(EN), 2005.

[8] P. Pinson and H. Madsen, "Ensemble-based probabilistic forecasting at Horns Rev," *Wind Energy*, vol. 12, no. 2, pp. 137–155, 2009.

[9] M. Lange, "On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors," *J. Solar Energy Eng. (Trans. ASME)*, vol. 127, no. 2, pp. 177–184, 2005.

[10] J. B. Bremnes, "A comparison of a few statistical models for making quantile wind power forecasts," *Wind Energy*, vol. 9, no. 1–2, pp. 3–11, 2006.

[11] T. Gneiting, K. Larson, K. Westrick, M. G. Genton, and E. Aldrich, "Calibrated probabilistic forecasting at the stateline wind energy center—The regime-switching space-time method," *J. Amer. Statist. Assoc.*, vol. 101, pp. 968–979, 2006.

[12] J. K. Møller, H. A. Nielsen, and H. Madsen, "Time-adaptive quantile regression," *Comp. Stat. Data Anal.*, vol. 52, no. 3, pp. 1292–1303, 2008.

[13] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, no. 1–2, pp. 95–108, 2006.

[14] R. Doherty and M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 587–595, May 2005.

[15] M. A. Matos and R. Bessa, "Operating reserve adequacy evaluation using uncertainties of wind power forecast," in *Proc. IEEE Power Tech Conf.*, Bucharest, Romania, 2009.

[16] E. D. Castronuovo and J. A. P. Lopes, "On the optimization of the daily operation of a wind-hydro power plant," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1599–1606, Aug. 2004.

[17] P. Pinson, C. Chevallier, and G. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1148–1156, Aug. 2007.

[18] B. Enfron and R. J. Tibshirani, *An Introduction to the Bootstrap*. London, U.K.: Chapman&Hall/CRC, 2007.

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

[20] C. Chatfield, *The Analysis of Time-Series: An Introduction*, 6th ed. London, U.K.: Chapman&Hall/CRC, 2003.

[21] J. W. Taylor and D. W. Bunn, "Investigating improvements in the accuracy of prediction intervals for combination of forecasts: A simulation study," *Int. J. Forecast.*, vol. 19, no. 1, pp. 57–70, 1999.

[22] W. H. Williams and M. L. Goodman, "A simple method for the construction of empirical confidence limits for economic forecasts," *J. Amer. Statist. Assoc.*, vol. 66, no. 336, pp. 752–754, 1971.

[23] A. P. A. da Silva and L. S. Moulin, "Confidence intervals for neural network based short term load forecasting," *IEEE Trans. Power Syst.*, vol. 15, no. 4, pp. 1191–1196, Nov. 2000.

[24] L.-W. Wang, *Adaptive Fuzzy Systems and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[25] R. T. Clemen and R. L. Winkler, "Combining probability distributions from experts in risk analysis," *Risk Anal.*, vol. 19, no. 2, pp. 187–203, 1999.

[26] C. Genest and K. J. McConway, "Allocating the weights in the linear opinion pool," *J. Forecast.*, vol. 9, no. 1, pp. 53–73, 1990.

[27] I. Sanchez, "Short-term prediction of wind energy production," *Int. J. Forecast.*, vol. 22, no. 1, pp. 43–56, 2006.

[28] G. Kariniotakis and D. Mayer, "An advanced on-line wind resource prediction system for the optimal management of wind parks," in *Proc. IEEE MedPower'02 Conf.*, Athens, Greece, 2002.

[29] T. S. Nielsen, H. A. Nielsen, and H. Madsen, "Prediction of wind power using time-varying coefficient-functions," in *Proc. 15th IFAC World Congr.*, Barcelona, Spain, 2002.

[30] J. Juban, N. Siebert, and G. Kariniotakis, "Probabilistic short-term wind power forecasting for the optimal management of wind generation," in *Proc. IEEE PowerTech 2007*, Lausanne, Switzerland, 2007, pp. 683–688.

[31] T. L. Thorarinsdottir and T. Gneiting, "Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression," *J. R. Statist. Soc. A*, vol. 173, pp. 371–388, 2010.

[32] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. Kariniotakis, "Nonparametric probabilistic forecasts of wind power: Required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.

[33] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *J. R. Statist. Soc. B*, vol. 69, pp. 243–268, 2007.

[34] P. Pinson, P. McSharry, and H. Madsen, "Reliability diagrams for nonparametric density forecasts of continuous variables: Accounting for serial correlation," *Quart. J. R. Meteorol. Soc.*, vol. 136, no. 646, pp. 77–90, 2010.

[35] [Online]. Available: http://www.anemos-project.eu.

[36] P. Pinson, G. Papaefthymiou, B. Klöckl, H. A. Nielsen, and H. Madsen, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, 2009.

Pierre Pinson received the M.Sc. degree in applied mathematics from the National Institute for Applied Sciences (INSA), Toulouse, France, in 2002 and the Ph.D. degree in energetics from the Ecole des Mines de Paris, Paris, France, in 2006.

He is an Associate Professor in stochastic energy systems at the Informatics and Mathematical Modeling Department of the Technical University of Denmark, Kgs. Lyngby, Denmark. He is currently on sabbatical, while acting as consultant for the European Centre for Medium-Range Weather Forecasts (ECMWF). His research interests include among others forecasting, uncertainty estimation, optimization under uncertainty, decision sciences, and renewable energies.

George Kariniotakis (S'95–M'02) was born in Athens, Greece. He received the Production and Management Engineering and M.Sc. degrees from the the Technical University of Crete, Crete, Greece, and the Ph.D. degree from the Ecole des Mines de Paris, Paris, France, in 1996.

He is currently with the Center of Energy and Processes of the Ecole des Mines de Paris as a Senior Scientist and Head of the Renewable Energies Group. His research interests include among others renewable energies, distributed generation, and artificial intelligence.