Time Adaptive Conditional Kernel Density Estimation for Wind Power Forecasting

R.J. Bessa, V. Miranda *Fellow IEEE*, A. Botterud, *Member IEEE*, J. Wang, *Member IEEE* and Emil M. Constantinescu

Abstract — **This paper reports the application of a new kernel density estimation model based on the Nadaraya-Watson estimator, for the problem of wind power uncertainty forecasting. The new model is described, including the use of kernels specific to the wind power problem. A novel time-adaptive approach is presented. The quality of the new model is benchmarked against a splines quantile regression model currently in use in the industry. The case studies refer to two distinct wind farms in the United States and show that the new model produces better results, evaluated with suitable quality metrics such as calibration, sharpness and skill score.**

Index Terms—**Wind power forecasting, uncertainty, kernel, density estimation, time-adaptive, decision-making.**

I. Introduction

Short term wind power forecasting is an exercise in handling uncertain information – and this paper presents a novel way to represent such information, preserving its diversity in a meaningful way. The so-called point forecast models cannot achieve this, as they provide deterministic predictions.

One may suggest that the attractiveness of a point forecast is still a consequence of the desire to have a "perfect forecast", in the sense that predicted values match exactly the actual ones. Reducing the prediction error has been the subject of a consistent effort in many models, and it has led to the discussion on the actual meaning of forecasting errors and on the information content of the error distribution [1]. But the growing economic importance of prediction errors, amplified by the growing penetration of wind power (c.f. [2]), is moving the perception of the forecasting exercise from accuracy to economic value. Therefore, in order to enable risk analysis, it becomes mandatory that the uncertainty associated with wind power generation is also estimated [3][4].

There are a number of statistical algorithms that may be used for such a purpose. The most important physical and statistical algorithms for wind power forecasting are mentioned in the state-of-the-art report [5]. The most popular statistical algorithms for wind power probabilistic forecasting are: splines quantile regression, which consists of a additive quantile regression with basis functions formulated as natural cubic B-splines [6]; adapted resampling, which uses a fuzzy inference model and resampling to determine the distributions of forecast errors associated with the power output forecast [7]; an adaption of the classic Nadaraya-Watson kernel density estimator (KDE) for modeling the wind power generation [8].

The usefulness of a wind power uncertainty forecast depends on how results are presented, because this links directly to the application of the forecast. A good model, therefore, should be flexible and provide information in several forms. Examples:

- The wind power bidding problem described in [4] requires the full probability density function (*pdf*) when computing the expected utility.
- The unit commitment problem requires at least temporal correlation of forecast errors [9].

Conditional kernel density estimation provides the tool to build models [8][10] that have as output a *pdf* of the forecasted wind power. This can be transformed into several uncertainty representations, such as quantiles, standard deviation, or skewness. This approach therefore provides a complete and robust representation of wind power probabilistic forecasts.

Other directions are currently being proposed and explored, such as in the method described in [11]: here, the forecasted densities take the form of discrete continuous mixtures consisting of a generalized logit-normal distribution with probability masses at the bounds.

Besides flexibility, a good model calls for time-adaptiveness. It has been observed time and again that models trained in an offline mode are unable to capture concept drift associated to non-stationary changes in the underlying distributions of the data variables. However, in the literature one mainly finds KDE-based models trained offline [8][10]. As for non-KDE time-adaptive models, one may refer to [11][12], as examples of the current trend.

This paper describes a novel Kernel Density Forecast (KDF) algorithm that is robust and time-adaptive. A basic proof of concept was initially presented in [13]. This paper now extends the analysis, establishes the fundamentals of the model, and confirms the results for two wind farms in the United States with distinct wind generation behavior.

The model introduces two novelties in the state of the art in wind power forecasting, in which it differs from previous attempts such as [8] or [10]: 1) it is a fully time adaptive model; and 2) it adopts distinct kernels for the several types of

The authors R.J. Bessa and V. Miranda are with INESC TEC - INESC Technology and Science (formerly INESC Porto) and FEUP - Faculty of Engineering, University of Porto, Portugal (emails: rbessa@inescporto.pt and vmiranda@inescporto.pt). The authors A. Botterud, J. Wang and E. Constantinescu are with Argonne National Laboratory, USA (abotterud@anl.gov, jianhui.wang@anl.gov, emconsta@mcs.anl.gov).
The work of R. J. Bessa was supported by Fundação para a Ciência e Tecnologia (FCT) Ph.D. Scholarship SFRH/BD/33738/2009.

variables, including those defined in circular domains.

II. Kernel Density Forecast Methodology

A. The Nadaraya-Watson Estimator

Conditional density estimation consists of estimating the density (*pdf*) of a random variable Y, knowing that the explanatory random variable X is equal to x. In the wind power prediction problem, this means assuming known a set of given values for explanatory variables x (e.g. numerical weather prediction variables, wind power measured values) and consequently forecasting the *pdf* of the wind power at a time step t for each look-ahead time step $t+k$ in a given time horizon.The expression translating this concept is:

$$\hat{f}_P\left(p_{t+k} \mid X = x_{t+k|t}\right) = \frac{f_{P,X}\left(p_{t+k}, x_{t+k|t}\right)}{f_X\left(x_{t+k|t}\right)} \quad (1)$$

where:

- p_{t+k} is the wind power forecasted for look-ahead time $t+k$
- $x_{t+k|t}$ are the explanatory variables forecasted for look-ahead time step $t+k$ and available at time step t,
- $f_{P,X}(p_{t+k},x_{t+k|t})$ is the multivariate density function
- and $f_X(x_{t+k|t})$ is the marginal density of X.

Because the actual *pdf* expressions are unknown and only discrete samples are available, the *pdf* must be estimated and this can be done with a non-parametric estimator by a Kernel Density Estimation (KDE) method [14]. Given independent and identically distributed data (i.i.d.) $X_1,\ldots,X_n$ drawn from an unknown density function f, the univariate KDE is given by:

$$\hat{f}_X(x) = \frac{1}{N \cdot h} \sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right) \quad (2)$$

where N is the number of samples, K is a kernel function and h the bandwidth parameter. This equation translates the following procedure: a kernel is placed around each sample X_i and the estimated density function for the set of data is calculated from dividing by N the sum of the N kernels, each centered on a sample.

With i.i.d. multivariate data $X_{1d},\ldots,X_{Nd}$ corresponding to d different variables, drawn from an unknown multivariate density function f, the multivariate KDE is given by the product kernel estimator [15]:

$$\hat{f}(x_1,\ldots,x_d) = \frac{1}{N \cdot h_1 \cdot \ldots \cdot h_d} \sum_{i=1}^{N} \prod_{j=1}^{d} K_j\left(\frac{x_j - X_{ij}}{h_j}\right) \quad (3)$$

where K is a multivariate kernel function and $h_1,\ldots,h_d$ a bandwidth vector.

The nonparametric conditional density estimation for Eq. 2 in the form referred to as the Nadaraya-Watson (NW) Estimator is as follows [16]:

$$\hat{f}(p \mid X = x) = \sum_{i=1}^{N} K_{h_p}(p - P_i) \cdot w_i(x) \quad (4)$$

having

$$w_i(x) = \frac{K_{h_x}(x - X_i)}{\sum_{i=1}^{N} K_{h_x}(x - X_i)} \quad (5)$$

where the bandwidth h_p controls the smoothness of each conditional density in the p direction, while h_x controls the smoothness between conditional densities in the x direction.

The application of Eq. 3 over real data from a wind farm produced the estimated joint *pdf* depicted in Fig. 1. This joint *pdf* represents the probability density associated to each joint realization of forecasted wind speed and wind power values. In this plot each contour line perpendicular to the "power" axis represents a conditional density (i.e. *pdf* conditioned to the forecasted wind speed value). Fig. 2 presents an example of a conditional *pdf* (i.e. ouput of KDF from Eq. 1 and 4) for the forecasted wind power conditioned to a forecasted wind speed of 5 m/s. This corresponds to construct a univariate *pdf* for a given x (5 m/s for Fig. 2) by weighting ($w_i(x)$ in Eq. 4 and 5) each wind power value (p in Eq. 4) proportionally to the closeness of the corresponding x_i (wind speed in this case).

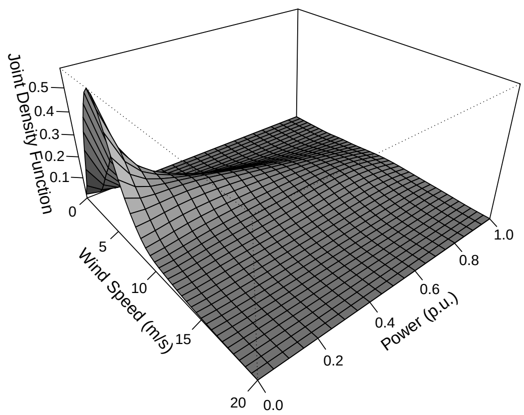

Fig. 1. Joint density function between wind power and forecasted wind speed.

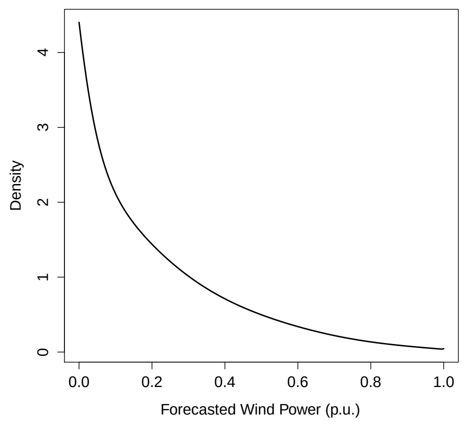

Fig. 2. Example of wind power forecast *pdf* for a given wind speed forecast.

B. A choice of kernels

Gaussian kernels are a common resource in KDE, but are not the most adequate choice in the wind power uncertainty forecasting problem. There are mainly three different types of variables in this problem:

- interval bounded such as wind power between 0 and 1;
- one-sided bounded, such as wind speed between 0 and +Inf;
- and circular variables like the wind direction or the hour of the day.

Variables with range [0,1] will be dealt with a beta kernel [17]:

$$\hat{f}_X(x) = \frac{1}{N}\sum_{i=1}^{N} K_{x/h+1,(1-x)/h+1}(X_i) \quad (6)$$

where $K_{p,q}$ is the density function of a *Beta(p,q)* random variable with p and q as the two positive shape parameters, and h being the bandwidth parameter.

Fig. 3 depicts the beta kernel shape for five different points. As shown, the beta kernel presents a varying shape according to the values of *x*, in fact the varying shape changes the amount of smoothing applied to the kernel estimator. Moreover, the kernel is non-negative, and presents a good capability for modeling near the boundaries 0 and 1.

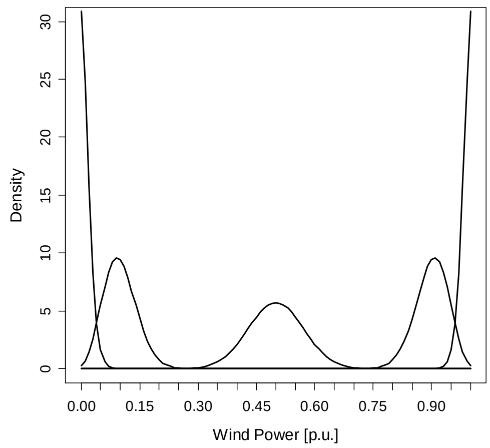

Fig. 3. Beta kernel of Eq. 6 for h=0.02 (x=0.01, x=0.1, x=0.5, x=0.9, x=0.99).

Variables with range [0,+Inf) will be associated with a gamma kernel [18]:

$$\hat{f}_X(x) = \frac{1}{N}\sum_{i=1}^{N} K_{x/h+1,h}(X_i) \quad (7)$$

where h is the bandwidth parameter of $K_{p,q}$, which is the density function of a *Gamma(p,q)* random variable with p as the shape parameter and q as the scale parameter.

Fig. 4 depicts the gamma kernel shape for five different points. Such as the beta kernel, this kernel also presents a varying shape according to the values of *x* and it is non-negative, but in this case it only saturates near the boundary 0.

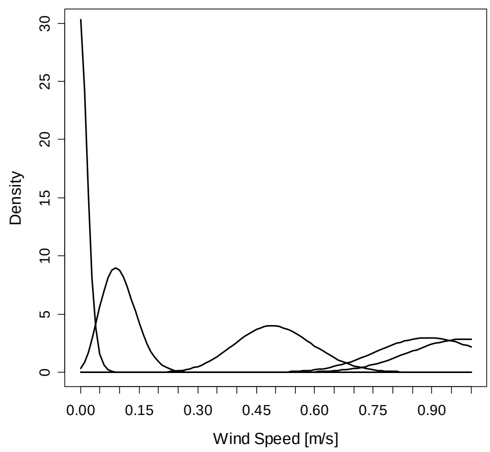

Fig. 4. Gamma kernel of Eq. 7 for h=0.02 (x=0.01, x=0.1, x=0.5, x=0.9, x=0.99).

Circular variables required the use of circular distributions such as the von Mises distribution [19]:

$$g(\theta;\mu,\kappa) = \frac{1}{2\pi \cdot I_0(\kappa)} e^{\kappa \cdot \cos(\theta-\mu)} \quad (8)$$

where

- I_0 is the modified Bessel function of the first kind and order 0,
- μ is the directional center of the distribution,
- κ is the concentration parameter
- θ belongs to any interval of length 2π.

The concentration parameter κ can be used to control the degree of smoothing in circular KDE. It is analogous to the bandwidth parameter but larger values lead to less smoothing.

Fig. 5 depicts an example (created with R package circular [20].) of circular KDE for the wind direction data. The points are represented in polar coordinates and placed in the circle line, while the three lines represent the density estimation for these sample points with different κ values.

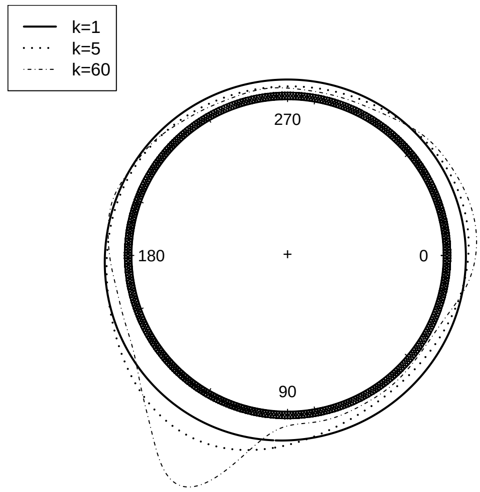

Fig. 5. Circular kernel density estimation for the wind direction data (κ=1,5, 60).

Because integrals computed from the beta and gamma kernels may lead to functions that do not have an integral equal to one, a modified kernel estimator [21] is used:

$$\hat{f}'(x) = \frac{\hat{f}(x)}{\int_0^1 \hat{f}(x)dx} \quad (9)$$

As this means a simple change of scale, the normalization is employed over the conditional function of Eq. 4.

Additional kernel functions were also tested; however, the kernels from Eq. 6-9 achieved the best performance. The comparisons can be found in [22].

C. Time-adaptive Estimator

A *pdf* time-adaptive estimator must not only learn new features from newly arriving samples but also forget old data whose information no longer contributes to build an accurate model in the present. Furthermore, a time-adaptive estimator must allow for the updating of the density function without the need to re-compute the whole *pdf* from scratch.

A recursive formula for the KDE estimator was proposed in [23]:

$$\hat{f}_n(x) = \frac{n-1}{n} \cdot \hat{f}_{n-1}(x) + \frac{1}{n \cdot h} \cdot K\left(\frac{x - X_i}{h}\right) \tag{10}$$

This formula can hardly be seen as time-adaptive because as n increases, the ratio *(n-1)/n* approaches unity (and *1/n* approaches zero), and the new samples become redundant – the model ceases to learn. Also this recursive expression does not guarantee an "automatic" discarding of old data in the eventuality of a serious concept drift, i.e., a change in the generating structure of the (non-stationary) data.

In order to build a suitable time-adaptive estimator that accounts for nonstationarity, one useful strategy is to adopt a forgetting factor λ that may progressively reduce the importance of old samples and give importance to new data. This is called a KDE estimator with exponential smoothing [24]:

$$\hat{f}_n(x) = \lambda \cdot \hat{f}_{n-1}(x) + \frac{(1-\lambda)}{h} \cdot K\left(\frac{x - X_i}{h}\right) \tag{11}$$

where λ controls how quickly or slowly the exponential smoothing adapts to the new data (exponential forgetting). Small values of λ will give much importance to newly arrived data; values of λ close to one give heavy weight to the historical data. The λ value can be related with an equivalent sliding window of size n by the following equation: $\lambda = n/(n+1)$; however, we should stress that λ is not a function of n. This parameter may also be adjusted during the use of the model, but this task requires a more complex approach capable of detecting data samples from different concepts.

The extension to the multivariate case is straightforward.

This paper introduces the conversion of the static Nadaraya-Watson (NW) estimator to a recursive time-adaptive estimator using Eq. 11. Its expression is:

$$\hat{f}(p \mid X = x)_t = \frac{\lambda \cdot \hat{f}_{t-1}(x,p) + (1-\lambda) \cdot \left[K\left(\frac{x - X_i}{h_x}\right) \cdot K\left(\frac{p - P_i}{h_p}\right)\right]}{\lambda \cdot \hat{f}_{t-1}(x) + (1-\lambda) \cdot K\left(\frac{x - X_i}{h_x}\right)} \tag{12}$$

where $f(p|X=x)_t$ means the knowledge of the model at time instant t, which is updated using recent values of P and X.

III. CASE STUDIES AND EVALUATION FRAMEWORK

In order to verify the quality of the results provided by the new model, the results for two different existing wind farms (designated WF1 and WF2) located in flat terrain in the U.S. Midwest are presented in this Section.

The complete dataset (SCADA and NWP) for both wind farms corresponds to the period between January 2009 and February 2010. The NWP data were generated with the Weather Research and Forecasting (WRF) model [26] by Argonne National Laboratory and consist of several weather variables (e.g. wind speed, direction, temperature) for one reference point inside each wind farm.

The exercises adopted the following schedule: wind power is forecasted at 6 AM for the horizon of t+6 up to t+48 hours. The resolution of the forecast data is one hour. The training dataset was defined from 1 January 2009 to 21 November 2009, and the testing set from 22 November 2009 to 20 February 2010 (a 70/30% ratio).

In order to obtain an independent quality assessment, the results obtained with the Nadaraya-Watson estimator will be compared with the splines quantile regression [6], which is one of the tools more often used. The analysis will be focused on this comparison and not on prediction results by themselves. The following section details the steps taken and the quality indicators adopted.

A. Evaluation Framework

In [27] one finds a systematic framework to evaluate probabilistic forecasts, which was adopted in this work. The evaluation relied on three metrics: calibration, sharpness and skill score. For reasons of comparison, the probabilistic forecast is represented through a set of quantiles ranging from 1% to 5% in 1% steps, them from 5% to 95% in 5% steps, and finally from 95% to 99% in 1% steps.

1) Calibration.

It is a measure of the agreement between nominal proportions (forecasted probabilities) and the ones computed from the evaluation sample (empirical). For instance, a 65% quantile should contain 65% of the observed values lower or equal to its value.

An indicator variable for a quantile forecast $\hat{q}^{\alpha}_{t+k|t}$ with nominal proportion α is:

$$\xi_k^{\alpha_i} = \begin{cases} 1 \; if \;\; p_{t+k} \le \hat{q}^{\alpha_i}_{t+k|t} \\ 0 \; otherwise \end{cases} \tag{13}$$

The indicator variable refers to the actual outcome of p_{t+k} at time $t+k$. Furthermore, these indicators are defined as follows:

$$n^{\alpha}_{k,1} = \#\left\{\xi^{\alpha}_{i,k} = 1\right\} = \sum_{i=1}^{N} \xi^{\alpha}_{i,k} \tag{14}$$

$$n^{\alpha}_{k,0} = \#\left\{\xi^{\alpha}_{i,k} = 0\right\} = N - n^{\alpha}_{k,1} \tag{15}$$

i.e., as sums of hits and misses, respectively, for a given horizon k over N realizations.

The empirical proportions are computed with the following expression:

$$\hat{\alpha}^{\alpha}_k = \frac{n^{\alpha}_{k,1}}{n^{\alpha}_{k,1} + n^{\alpha}_{k,0}} = \frac{n^{\alpha}_{k,1}}{N} \tag{16}$$

The same indicators can be obtained from the predictions instead of the observed values, leading to predicted proportions α. The deviation from the "perfect calibration" (where empirical proportions match nominal or forecasted proportions) or the bias is given by:

$$b^{\alpha}_k = \alpha - \hat{\alpha}^{\alpha}_k \tag{17}$$

2) Sharpness

This measure describes to what extent probabilistic forecasts tend to look like discrete forecasts. Quantiles are gathered by pairs in order to obtain intervals with different nominal coverage rates. Let $\delta^{\alpha}_{t+k|t} = \hat{q}^{1-\alpha/2}_{t+k|t} - \hat{q}^{\alpha/2}_{t+k|t}$ be the size of the interval forecast with nominal coverage rate *1-α* estimated at time t for lead time $t+k$. For most forecast users, narrow intervals are desired because they mean a more compact representation of the pdf. In this work, sharpness is evaluated as the mean size of the distance between quantiles:

$$\bar{\delta}_k^{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \delta_{i,k}^{\alpha} \qquad (17)$$

3) Skill score

Skill scores try to condense in a single number a variety of aspects in the evaluation of probabilistic forecasts. They have the advantage of compact information but the relative contributions of each feature are obscured. For this work, a skill score was calculated for a set of m quantiles as:

$$S_c\left(\hat{f}_{t+k}, p_{t+k}\right) = \sum_{i=1}^{m}\left(\xi_{t+k}^{\alpha_i} - \alpha_i\right)\cdot\left(p_{t+k} - \hat{q}_{t+k}^{\alpha_i}\right) \qquad (18)$$

where p_{t+k} is the realized wind power, α_i is the quantile proportion, q_{t+k} is the forecasted quantile and ξ is the indicator variable in Eq. 13. The higher the scoring, the better: the maximum value is 0 for "perfect" probabilistic forecasts.

B. Which metric to use

Each metric conveys information of a different type. In wind power forecasting, and because predictions may be used for business purposes (offer of power in the market, reserve setting, etc.), *calibration* and *sharpness* provide meaningful information for assessing the potential impact of forecast. *Calibration* is important because the quantiles are related with the energy and power generated and wind power producers will tend to favor models that provide forecasting quantiles close to the observed ones. Sharpness provides information about the “degree” of uncertainty in each forecast; forecasts with broad intervals are not very informative for decision-making. The relation between these two metrics and decision-making problems will be discussed in section V. Skill scores, while giving a holistic assessment including calibration and sharpness, fail to highlight relevant individual, aspects but may nevertheless provide a global evaluation.

IV. CONFIRMATION OF THE QUALITY OF THE NEW TIME ADAPTIVE NARADAYA-WATSON KDF

A. Feature selection analysis and choice of model

Different explanatory variables were considered to build several NW KDF models: wind speed (M0); wind speed + direction (M1); wind speed + hour of the day (M2); wind speed + look-ahead time step (M3); wind speed + direction + hour of the day (M4); wind speed + direction + look-ahead time step (M5).

The determination of the kernel bandwidth is an important step in the KDF method. In this paper we used a 10-fold cross validation for computing the bandwidths. However, this should be a topic of future research, in particular for the time-adaptive version. Sensitivity analysis on the impact of bandwidth’s size in this method and problem can be found in [22].

The 10-fold cross validation process was conducted in the training dataset for a different bandwidth value ranging from 0.001 to 0.5 with 0.001 increments. The selection rule is the bandwidth with the lowest average skill score.

The following kernel functions, determined by 10-fold cross validation, were used:

- beta kernels for wind power generation (bandwidth of 0.008 in WF1 and 0.004 in WF2);
- gamma kernels for wind speed forecast (bandwidth of 0.05 in WF1 and 0.1 in WF2);
- von Mises distribution for wind direction and hour of the day (bandwidth equal to 2.5);
- beta kernel for the look-ahead time step (bandwidth equal to 0.1).

Using the study data, statistical models were developed and probabilistic forecasts generated.

Fig. 6 and Fig. 7 display the skill score calculated for the five models for WF1 and WF2. The differences between the two cases are not significant: the performance of all models is rather similar, with a slight advantage for models M2 and M3, whose curves dominate the other.

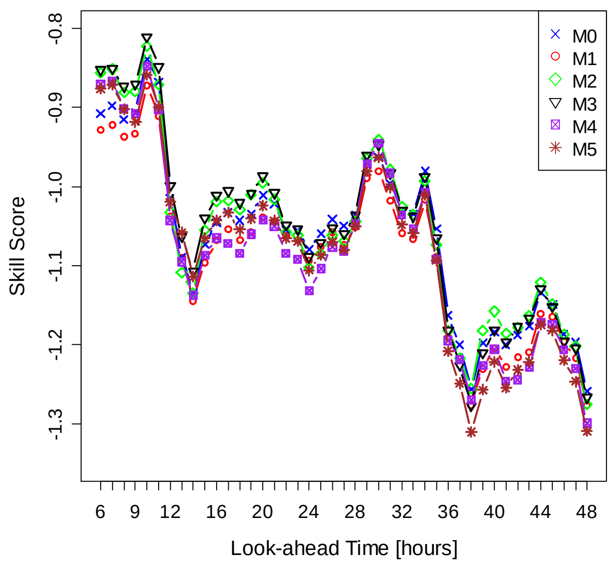

Fig. 6. Skill score diagram for Wind Farm 1 and NW models M0-M5.

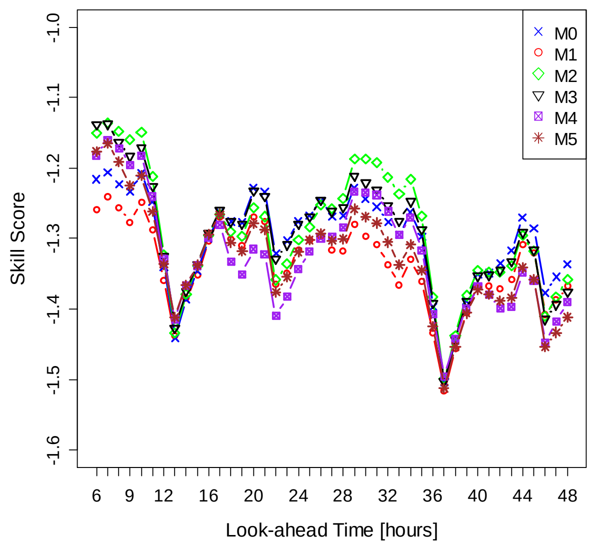

Fig. 7. Skill score diagram for Wind Farm 2 and NW models M0-M5.

The wind direction does not seem to improve the models. A possible explanation is that the wind farms are located in flat terrain – wind direction would possibly be more important in wind farms located in mountainous or costal terrain. In the following sections, the analysis will be focused on model M3, one of the models with best skill score performance. The negative spikes observed in the skill score may be due to a loss in NWP wind accuracy at day-night transition, which present relatively higher forecasting difficulties.

B. *Results for offline models*

In this section the NW model M3 is compared with a splines quantile regression (SplinesQR) model with 6 degrees of freedom, using the same input and for both wind farms.

The calibration diagrams obtained for the two wind farms are represented in Fig. 8 and Fig. 9. In fact, these figures display the deviation between the predicted and observed proportions for the quantile range.

In both cases, the best calibration performance is obtained with the NW model for almost all quantiles. SplinesQR presents a slightly better calibration for the left tail but underestimates its quantiles, while the NW model overestimates; the NW models present the best calibration for the right tail.

In practical terms, this means that the NW model generates a *pdf* whose shape leads to proportions following the observed ones in a much closer way than the SplinesQR model.

The Splines QR model presents, however, a slightly better sharpness performance in both wind farms. This is illustrated in Fig. 10 for WF1, the result for WF2 being virtually identical.

As noted in [8], there seems to be a trade-off between calibration and sharpness, meaning that improved calibration will usually come with degraded sharpness.

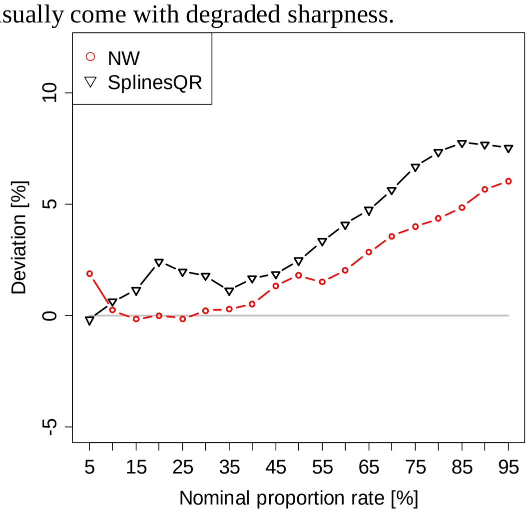

Fig. 8. Calibration diagram for WF1 in offline NW and QR models.

Finally, Fig. 11 depicts the skill score computed for each look-ahead time step. SplinesQR is marginally better than NW for some look-ahead steps, but it is also worse in others. There is no visible advantage for any of the methods. Note that, as mentioned in [27], the skill score of Eq. 18 is a generalization of the loss function considered in quantile regression, hence it could justify why quantile regression presents a good performance in this criterion.

Everything considered, comparing two models with equivalent skill scores, we argue that a model displaying better calibration should be preferred if the difference loss in sharpness is not significant, and this is the case of the new NW model.

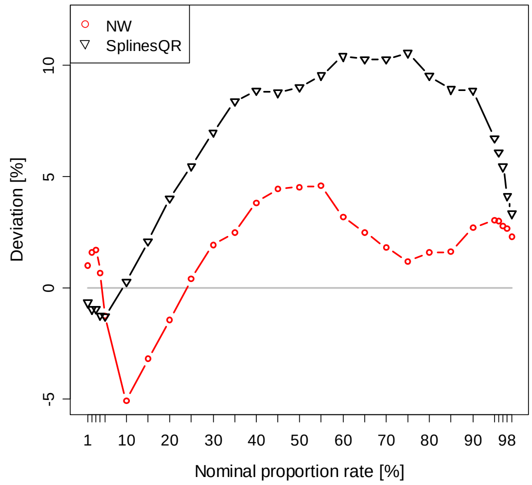

Fig. 9. Calibration diagram for WF2 in offline NW and QR models.

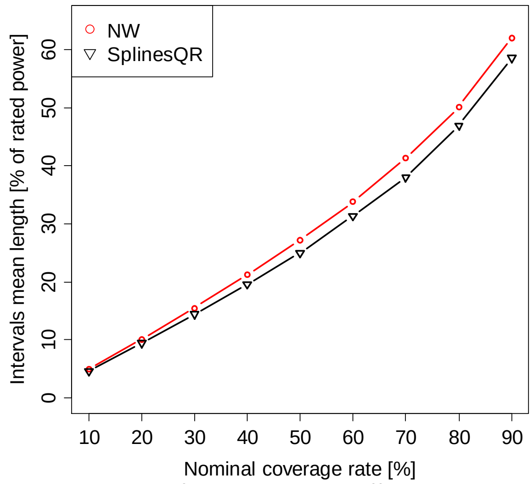

Fig. 10. Sharpness diagram for WF1 obtained in offline NW and QR models. The result for WF2 is similar.

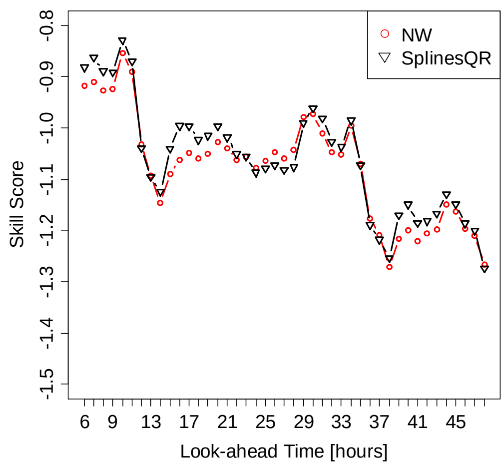

Fig. 11. Skill score diagram for wind farm 1 obtained with offline NW and QR models.

C. Results for the time-adaptive model

In order to assess the properties of the new time-adaptive NW model, its results are now compared with the offline NW model for different values of the forgetting factor (λ).

For a better understanding of the meaning associated with different λ values, λ was represented by the corresponding size n of an equivalent sliding window, which is given by $\lambda=n/(n+1)$. So, three values for λ were considered: 0.99963477 (corresponds to n=2738 points), 0.999 (corresponds to n=1000 points) and 0.995 (corresponds to n=200 points). The same kernels and bandwidths used for the offline version were adopted for the time-adaptive models.

Fig. 12 and Fig. 13 display the calibration results for both wind farms, and there are differences to notice. While in WF1 a high value of λ produces in general a better calibration result, in WF2 a smaller value (fewer points in the sliding window) leads to better results in almost all quantiles, meaning that the model for WF2 is more sensitive to newly arrived data and needs to adapt and forget old data.

The offline model presented a slightly better calibration in a few quantiles but is surpassed by the adaptive models in a large quantile range, especially in WF2. Only between the 5% and 30% quantiles the offline model has a better value than the time-adaptive, and the adaptive models present better calibration at the tails.

From these figures we conclude that an important trait of the time-adaptive model is that it changes the bias of the prediction, compared to the offline model, and that the bias change is influenced by the forgetting factor λ.

As for sharpness, the results show that there is no significant impact in this metric when moving from offline to time-adaptive models.

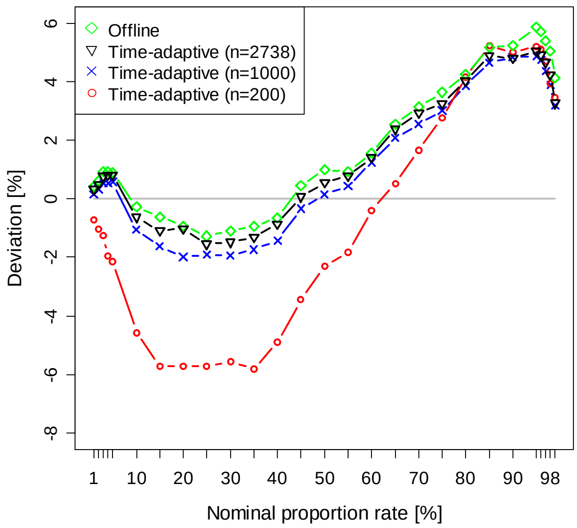

Fig. 12. Calibration diagram for WF1: offline vs. time-adaptive NW.

Fig. 14 presents the sharpness diagram for WF1, where one notices that for the adaptive models with best calibration there is no impact on sharpness (a slight improvement is only observable for the model with lower λ). For WF2, the sharpness curves were virtually identical for all models.

Because there seems to be superiority in calibration without degrading sharpness, it is expected that the time-adaptive NW model may perform better in terms of skill score than the offline model. In fact, this can be observed in Fig. 15 and Fig. 16: in the first 20 look-ahead time steps for WF1, and in the first 28 look-ahead time steps for WF2, the best adaptive models outperform the offline model. In the rest of the look-ahead period, the results merge.

The results lead to the following conclusions: the time-adaptive model exhibits better characteristics than the offline version, i.e. a better calibration with same sharpness, resulting in more favorable global skill score.

V. Goodness in Probabilistic Forecasts: A Discussion

This paper, while proposing a new approach to wind power uncertainty forecasting, based on the Nadaraya-Watson Estimator, coupled with the use of specific kernels and evolved into a time-adaptive model, adopted the classical point of view of statistical analysis: i.e., the quality of the result is solely measured against the desired outcome. The evaluation of this quality is made through the calculation of several indices (calibration, sharpness and skill score), all having one trait in common: the basic objective is to reduce, under some metric, the distance between predicted and observed values.

This has been coined as the "forecaster paradigm" in [2], where an extensive discussion is presented, especially focused on point forecasting but with evident extensions. The issue with this paradigm is that it does not take in account in any way the impact of errors and uncertainty in the decision making processes that make use of the forecasts produced.

In Section III, when discussing which metric to use, some hints were already given on this matter. Forecasts will be used by decision-makers and it is useful to distinguish between wind power generation companies (WGENCO) and system operators (SO). Each group has its own set of criteria that associates value or penalty to the forecast (and its errors or uncertainty). It may happen, as shown in [2], that these sets are not only disjoint but that convenient solutions for a WGENCO are deemed as undesired by the SO and vice versa.

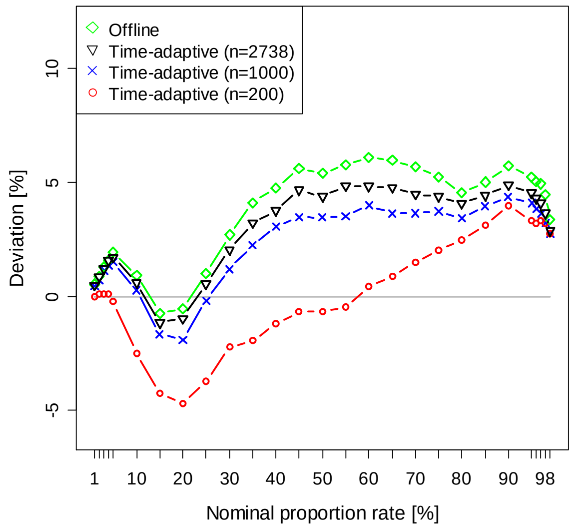

Fig. 13. Calibration diagram for WF2: offline vs. time-adaptive NW.

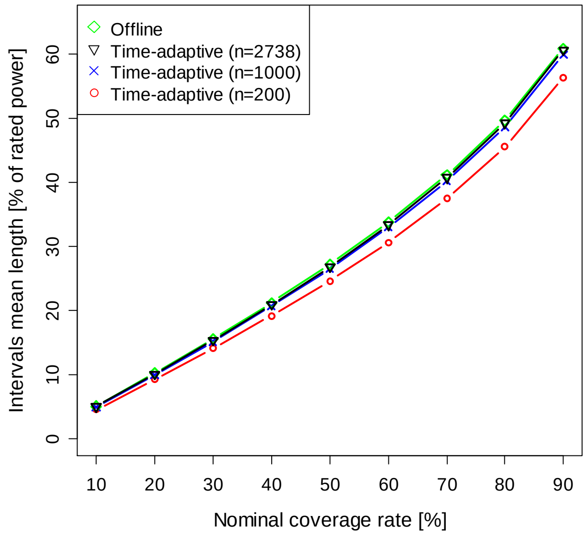

Fig. 14. Sharpness diagram for WF1 and time-adaptive NW.

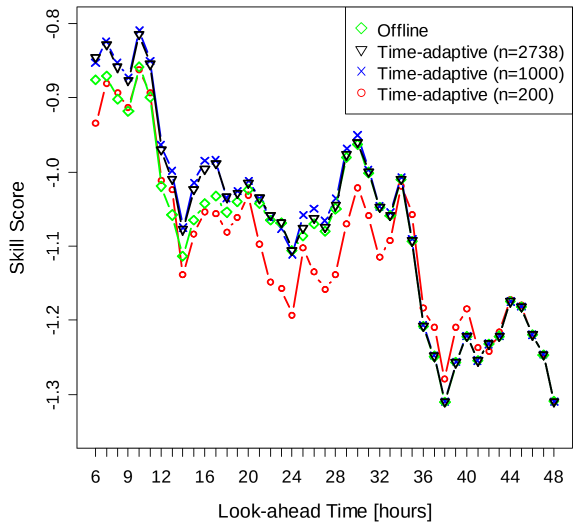

Fig. 15. Skill score diagram for WF1: offline vs. time-adaptive NW.

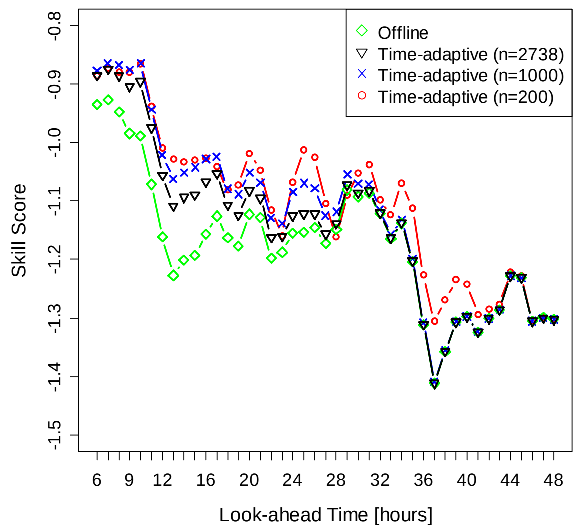

Fig. 16. Skill score diagram for WF2: offline vs. time-adaptive NW.

Ideally, then, the training or selection of a prediction method should have as cost function the economic value of the forecasts, translated into a specific model. These models are usually very complex and there is rarely a practical way to represent them in full during the process of tuning a prediction model. Therefore, one should resort to evaluation metrics that may in a sense approach the decision-making consequences.

Two relevant decision-making problems may be recalled: i) optimal wind power bidding in the electricity market (for a WGENCO); ii) stochastic unit commitment (for a SO).

In the wind power bidding problem, some models [28][29] try to find the "optimal quantile" – a bid should be equal to a quantile corresponding to the maximum expected value of the successful bids. Thus the bias in the probabilistic forecast influences strongly what should be considered "bad" and "good" bids. A metric that is associated with bias will therefore be an appropriate indicator for the quality of a uncertainty prediction method to be used for bidding, and this metric may be calibration.

The same can be said for problems requiring an evaluation of a trade-off between expected value and risk [4]. If the risk measure adopted is the value at risk (which is a quantile of the *pdf*), a large bias may lead to risk over- or underestimation. So, calibration as a measure of quality is again a good indicator. Furthermore, a high accuracy in the computation of the expected value can only be achieved if the model exhibits good calibration. Sharpness is not very relevant in this context and the skill score may be misleading.

The NW KDE, namely the time adaptive model, produces results with better calibration than SplinesQR and this suggests that, in the context of the bidding problem, it should be preferred.

In the stochastic Unit Commitment (UC) problem, the decision-making must take into account time dependency in wind power. Stochastic programming models must deal with a discrete set of wind power scenarios that should be generated coherently from the probabilistic forecasts [9].

The quality of commitment decisions is highly dependent on the quality of the scenarios generated and these, in turn, on the quality of the probabilistic forecasts. Models exhibiting higher calibration and sharpness may lead to a better discrete scenario generation. Therefore, a model that reaches a good compromise between sharpness and calibration is a model convenient for adoption upstream of the UC problem.

This was evident in [30], where the consequences from scenarios generated after probabilistic forecasts produced either by the NW or SplinesQR methods were assessed, with the conclusion that the superior calibration of the NW model resulted in decisions leading to lower operational costs.

The proposed NW models, with good calibration and sharpness indices represented in the form of good skill scores (superior to the SplinesQR model), are therefore serious candidates to be used to produce predictions useful for the stochastic UC problem.

VI. Conclusions

This paper presents a new model to estimate the uncertainty in short term wind power forecasts and the confirmation of its value from testing with real data from two large wind farms in the United States. Novelties introduced in the Nadaraya-Watson kernel density estimator for the wind power uncertainty forecasting problem, include the adoption of specific kernels for the explanatory variables used and the development of a time-adaptive model. The validation of the new models was done against splines quantile regression, a method being currently used.

The case studies of two wind farms showed that the probabilistic models in the two cases had somewhat different characteristics – in one of the wind farms the concept drift was more evident, i.e., in a time-adaptive version it became necessary to give more importance to newly arriving information. Nevertheless, in both cases the new models revealed superiority over the benchmarking model. The results demonstrated that the NW model leads to better calibration, while the quantile regression methods have a tendency to present a better sharpness performance. The importance of having confirmed this in two wind farms with different behavior should not be disregarded.

Having argued that calibration is of major importance for the wind power industry and for power system operators, the results show that the new NW model is an alternative with some competitive advantage regarding the model previously proposed and used, based on quantile regression.

Still, from a theoretical perspective, density estimation models offer an important advantage over other models. The NW model produces a full description of the forecasted *pdf*, which allows the derivation of many forms of representation of uncertainty. Moreover, it offers no difficulty in representing multimodal distributions and computing modes instead of just the expected value. In summary, density estimation models, such as the NW model presented here, offer significant advantages for the wind power industry when uncertainty estimation is required.

VII. Acknowledgements

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (‘‘Argonne’’). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up non-exclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

The authors acknowledge Horizon Wind Energy for providing the wind farm data used in the analysis.

VIII. References

[1] R.J. Bessa, V. Miranda, and J. Gama, “Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting,” *IEEE Trans. on Power Sys.*, vol. 24 (4), pp. 1657-1666, 2009.

[2] R.J. Bessa, V. Miranda, A. Botterud, and J. Wang, “‘Good’ or ‘bad’ wind power forecasts: a relative concept,” *Wind Energy*, vol. 14 (5), pp. 625-636, Jul. 2011.

[3] J. Wang, A. Botterud, R.J. Bessa, H. Keko, L. Carvalho, D. Issicaba, J. Sumaili, and V. Miranda, “Representing wind power forecasting uncertainty in unit commitment,” *Applied Energy*, vol. 88, pp. 4014-4023, Nov. 2011.

[4] A. Botterud, Z. Zhou, J. Wang, R.J. Bessa, H. Keko, J. Sumaili and V. Miranda, “Wind power trading under uncertainty in LMP markets,” *IEEE Transactions on Power Systems*, vol. 27(2), pp. 894-903, May 2012.

[5] C. Monteiro, R.J. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, “Wind power forecasting: state-of-the-art 2009,” Report ANL/DIS-10-1, Argonne National Laboratory, 2009. [Online] http://www.dis.anl.gov/projects/windpowerforecasting.html

[6] 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(1-2), pp. 95-108, 2006.

[7] P. Pinson and G. Kariniotakis, “Conditional prediction intervals of wind power generation,” *IEEE Trans. on Power Sys*., vol. 25(4), pp. 1845-1856, 2010.

[8] J. Juban, N. Siebert, and G. Kariniotakis, “Probabilistic short-term wind power forecasting for the optimal management of wind generation,” in *Proc. of the IEEE PowerTech Conference*, Switzerland, July 2007.

[9] P. Pinson, G. Papaefthymiou, B. Klockl, H.Aa. Nielsen, and H. Madsen, “From probabilistic forecasts to statistical scenarios of short-term wind power production,” *Wind Energy*, vol. 12(1), pp. 51-62, 2009.

[10] R.J. Bessa, J. Mendes, V. Miranda, A. Botterud, J. Wang, and Z. Zhou, “Quantile-copula density forecast for wind power uncertainty modeling,” in *Proc. of the IEEE PowerTech Conference*, Norway, June 2011.

[11] P. Pinson, “Very short-term probabilistic forecasting of wind power with generalized logit-Normal distributions,” *Journal of the Royal Statistical Society, Series C*, in press, 2012.

[12] J.K. Møller, H.A. Nielsen, and H. Madsen, “Time-adaptive quantile regression,” *Comp. Stat. & Data Anal.*, vol. 52(3), pp. 1292-1303, Jan. 2008.

[13] R.J. Bessa, J. Sumaili, V. Miranda, A. Botterud, J. Wang, and E. Constantinescu, “Time-adaptive kernel density forecast: a new method for wind power uncertainty modeling,” in *Proc. of the 17th PSCC Conference*, Stockholm, Sweden, Aug. 2011.

[14] M. Rosenblatt, “Remarks on some nonparametric estimates of a density function,” *The Annals of Math. Stat.*, vol. 27(3), pp. 832-837, 1956.

[15] M.P. Wand and M.C. Jones, “Multivariate plug-in bandwidth selection,” *Comp. Stat.*, vol. 9, pp. 97-116, 1994.

[16] R.J. Hyndman, D.M. Bashtannyk and G.K. Grunwald, “Estimating and visualizing conditional densities,” *Journal of Comp. and Graph. Stat*., vol. 5(4), pp. 315-336, Dec. 1996.

[17] S.X. Chen, “Beta kernel estimators for density functions,” *Comp. Stat. & Data Anal.*, vol. 31(2), pp. 131-145, 1999.

[18] S.X. Chen, “Probability density function estimation using gamma kernels,” *Annals of the Inst. of Stat. Math.*, vol. 52(3), pp. 471-480, Sept. 2000.

[19] K. V. Mardia and P. E. Jupp, *Directional Statistics*, New York: Wiley, Nov. 1999.

[20] C. Agostinelli and U. Lund, R package 'circular': Circular Statistics (version 0.4-3). [Online] https://r-forge.r-project.org/projects/circular/

[21] C. Gourieroux and A. Monfort, “(Non) Consistency of the beta kernel estimator for recovery rate distribution,” Working Paper Nº2006-31, Istitut National de la Statistique et des Etudes Economiques, Dec. 2006.

[22] J. Mendes, R.J. Bessa, H. Keko, J.S. Akilimali, V. Miranda, C. Ferreira, *et al.*, “Development and testing of improved statistical methods for wind power forecasting,” Technical Report ANL/DIS-11-7, Argonne National Laboratory, Sept. 2011. [Online] http://www.dis.anl.gov/projects/windpowerforecasting.html

[23] E.J. Wegman and H.I. Davies, “Remarks on recursive estimators of a probability density,” *The Annals of Stat*., vol. 7(2), pp. 316-327, 1979.

[24] E.J. Wegman and D.J. Marchette, “On some techniques for streaming data: a case study of internet packet headers,” *Journal of Comp. and Graph. Stat.*, vol. 12(4), pp. 893-914, 2003.

[25] Eastern Wind Integration and Transmission Study (EWITS), National Renewable Energy Laboratory (NREL). Information at: http://www.nrel.gov/wind/systemsintegration/ewits.html

[26] W. Skamarock, J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, W. Wang, and J.G. Powers, “A Description of the advanced research WRF version 2,” NCAR/TN–468+STR Technical note, Jun. 2005.

[27] P. Pinson, H.A. Nielsen, J.K. Moller, H. Madsen and G. Kariniotakis, “Nonparametric probabilistic forecasts of wind power: required properties and evaluation,” *Wind Energy*, vol. 10(6), pp. 497-516, Nov. 2007.

[28] J.B. Bremnes, “Probabilistic wind power forecasts using local quantile regression,” *Wind Energy*, vol. 7(1), pp. 47-54, Mar. 2004.
[29] V.R. Jose and R.L. Winkler, “Evaluating quantile assessments,” *Operations Research*, vol. 57(5), pp. 1287-1297, Sept. 2009.
[30] A. Botterud, Z. Zhou, J. Wang, J. Valenzuela, R.J. Bessa, J. Sumaili, H. Keko, and V. Miranda, “Unit commitment and operating reserves with probabilistic wind Power forecasts,” in *Proc. of the IEEE PES Trondheim PowerTech 2011*, Norway, June 19-23 2011.

IX. Biographies

Ricardo Bessa received his Licenciado (five years) degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 2006 in Electrical and Computer Engineering. In 2008 he received his Master of Science degree in Data Analysis and Decision Support Systems on the Faculty of Economics of the University of Porto (FEP). Currently, he is a researcher at INESC TEC in its Power Systems Unit and a PhD student of the Doctoral program in Sustainable Energy Systems (MIT Portugal) at FEUP. His research interests include wind power forecasting, electric vehicles, data mining and decision-aid methods.

Vladimiro Miranda received his Licenciado, Ph.D., and Agregado degrees from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 1977, 1982, and 1991, all in electrical engineering. In 1981, he joined FEUP and currently holds the position of Professor Catedrático. He is also currently a Director of INESC TEC, an advanced research institute in Portugal. He has authored many papers and been responsible for many projects in areas related to the application of computational intelligence to power systems.

Audun Botterud received his M.Sc. in industrial engineering (1997) and a Ph.D. in electrical power engineering (2003), both from the Norwegian University of Science and Technology. He is an energy systems engineer in the Center for Energy, Environmental, and Economic Systems Analysis (CEEESA) at Argonne National Laboratory. He was previously with SINTEF Energy Research in Trondheim, Norway. His research interests include electricity markets, power systems, renewable energy, wind power integration, stochastic optimization, and agent-based modeling.

Jianhui Wang received his B.S. degree in management science and engineering (2001) and an M.S. degree in technical economics and management (2004), both from North China Electric Power University, China, and his Ph.D. in electrical engineering from Illinois Institute of Technology, USA (2007). Presently, he is a computational engineer - energy systems with CEEESA at Argonne National Laboratory. His research interests include energy economics and policy, agent-based modeling and simulation, and electric power systems optimization and economics. He is chair of the IEEE PES power system operation methods subcommittee and co-chair of the task force on integration of wind and solar power into power system operations.

Emil M. Constantinescu received the B.Sc. and M.S. degrees from the College of Automatic Controls and Computers, Bucharest Polytechnic University, Bucharest, Romania, in 2001 and 2002, respectively, and the Ph.D. degree in computer science from Virginia Tech, Blacksburg, in 2008. He is currently an assistant computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory, Argonne, IL. His research interests include uncertainty quantification in weather and climate models and their applications to energy systems.