TIME-ADAPTIVE KERNEL DENSITY FORECAST: A NEW METHOD FOR WIND POWER UNCERTAINTY MODELING

R.J. Bessa[1], J. Sumaili[1], V. Miranda[1], A. Botterud[2], J. Wang[2], E. Constantinescu[3]
[1]Instituto de Engenharia de Sistemas e Computadores do Porto (INESC Porto), Faculdade de Engenharia, Universidade do Porto
Porto, Portugal
rbessa@inescporto.pt, jean.sumaili@inescporto.pt, vmiranda@inescporto.pt
[2,3]Argonne National Laboratory, [2]Decision and Information Sciences Div., [3]Mathematics and Computer Science Div.
Argonne, IL, USA
abotterud@anl.gov, jianhui.wang@anl.gov, emconsta@mcs.anl.gov

Abstract – This paper reports new contributions to the advancement of wind power uncertainty forecasting beyond the current state-of-the-art. A new kernel density forecast (KDF) method applied to the wind power problem is described. The method is based on the Nadaraya-Watson estimator, and a time-adaptive version of the algorithm is also proposed. Results are presented for different case-studies and compared with linear and splines quantile regression.

Keywords: Wind power forecasting, uncertainty, kernel density estimation, time-adaptive.

1 INTRODUCTION

A wind power forecaster seeks the perfect wind power forecast, but it is common sense that this represents a utopian image. Nevertheless, successful efforts have been made to decrease the wind power forecast error [1]. Furthermore, with the growing penetration of wind power and the economic importance of forecasting errors [2], it is becoming increasingly important to also forecast the uncertainty associated with wind power generation prediction.

An extensive state-of-the-art report on algorithms for wind power uncertainty forecasting can be found in [3]. The most popular statistical algorithms are: splines quantile regression [4], which consists of a linear quantile regression with basis functions formulated as cubic B-splines; adapted resampling [5], which is a process for generating alternative scenarios of power production, and this way, it is possible to change the weights obtained by a fuzzy inference system. Physical approaches for uncertainty forecasting can be found in [6] and [7].

The information provided by probabilistic forecasting algorithms creates additional value (e.g. economic) in several decision-making problems. Botterud *et al.* [8] presented several bidding strategies for wind power in the electricity market, such as expected utility maximization and trade-off between expected value and risk. Matos and Bessa [9] presented a decision making approach for setting the operating reserve requirements using non-parametric probabilistic forecasts (a set of quantiles) as input. Wang *et al.* [10] described a stochastic unit commitment that uses forecasted scenarios of wind power generation as input. Usaola [11] presented a probabilistic power flow that takes into account correlation between wind farms and uses beta distributions for modeling the wind power uncertainty.

Wind power uncertainty can take the form of probabilistic forecasts [4]-[14] or scenarios for short-term wind power generation [12]. Probabilistic forecasting consists of expressing the wind power generation or forecast error in "probabilistic terms", such as: parametric representation (e.g. Gaussian distribution); moments of the distributions (e.g. standard deviation, skewness); a set of quantiles; probability density function (*pdf*). Normally, the uncertainty representation is determined by the algorithm used, e.g. if quantile regression is used, the uncertainty is represented by a set of quantiles.

Models trained in an offline mode (e.g. [4][14]) are unable to cope with (non-stationary) changes in the underlying distributions of the input variables. The trend in the state-of-the-art is to develop algorithms capable of adapting to changes in data [13].

Hence, an algorithm for wind power uncertainty forecasting should ideally have as requisites: i) a high flexibility to represent wind power uncertainty; ii) time-adaptive characteristics.

In this paper we propose a novel Kernel Density Forecast (KDF) algorithm that addresses these two requisites. The output is a *pdf* of the forecasted wind power, and since this representation is generic it can be transformed to several uncertainty forms, such as quantiles, standard deviation, or skewness. A time-adaptive version of the algorithm is described, which means that the model is capable of learning from recent information while discounting older information.

Methods based on Kernel Density Estimation (KDE) are not new in the state-of-the-art, one example can be found in [14]. The authors present an adaptation of the classic Nadaraya-Watson kernel density estimator, where all the kernel functions are biweight functions. The reflection method was used for bounded variables.

Our approach differs in two important ways: 1) our method is based on selecting adequate kernels for modeling the different variables types, 2) our method is time-adaptive.

The paper is organized as follows: section 2 describes the KDF methodology; in section 3 results are presented for different case-studies and compared with linear and splines quantile regression; section 4 presents the conclusions.

2 KERNEL DENSITY FORECAST METHODLOGY

2.1 Kernel Density Estimation

KDE consists of a non-parametric estimator of a probability density function (*pdf*) [15]. 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:

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

where N is the number of samples, K is a kernel function and h the bandwidth parameter. This equation, places a kernel around each sample X_i.

Given i.i.d. multivariate data $X_{1d},\ldots,X_{nd}$ from d different variables drawn from an unknown multivariate density function f, the multivariate KDE is given by the product kernel estimator [16]:

$$\hat{f}(x_1,\ldots,x_d) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d}\frac{1}{h}K_j\left(\frac{x_j - X_{ij}}{h_j}\right) \quad (2)$$

where K_j is the kernel function for variable j with bandwidth h_j.

2.2 Nadaraya-Watson Estimator

Conditional density estimation consists of estimating the density of a random variable Y, knowing that the explanatory random variable X is equal to x. In other words, it consists of estimating the density of Y conditioned to $X=x$, $f(y|X=x)$. The conditional density can be formulated as follows:

$$\hat{f}(y \mid X = x) = \frac{f_{XY}(x,y)}{f_X(x)} \quad (3)$$

where $f_{XY}(x,y)$ is the multivariate density function of X and Y (joint *pdf*) and $f_X(x)$ is the marginal density of X.

It is also possible to have nonparametric conditional density estimation using Eq. 1 and 2. The classic approach is the modified Nadaraya-Watson kernel smoother [17]:

$$\hat{f}(y \mid X = x) = \sum_{i=1}^{N} K_{h_y}(y - Y_i)\cdot w_i(x) \quad (4)$$

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_y controls the smoothness of each conditional density in the y direction, while h_x controls the smoothness between conditional densities in the x direction.

2.3 Formulation for the Wind Power Problem

The wind power density forecast problem can be formulated as: forecast the wind power *pdf* at time step t for each look-ahead time step $t+k$ of a given time-horizon (e.g. up to 72 hours ahead) knowing a set of explanatory variables (numerical weather prediction (NWP) variables, wind power measured values, hour of the day, etc.).

Translating this sentence to an equation, we have:

$$\hat{f}_P(p_{t+k} \mid X = x_{t+k|t}) = \frac{f_{P,X}(p_{t+k}, x_{t+k|t})}{f_X(x_{t+k|t})} \quad (6)$$

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/launched at time step t.

Eq. 6 can be solved using Eq. 4 and 5, where the variable Y is the wind power, and the explanatory variables X are for instance: NWP variables (wind speed, wind direction, pressure), wind power point forecast, measured wind power.

Fig. 1 depicts the joint *pdf* computed using Eq. 2 for data from a real wind farm. This joint *pdf* represents the probability density associated to each joint realization of forecasted wind speed and realized wind power.

Fig. 2 is a discrete representation of the information contained in Eq. 6. It allows seeing the changes in the wind power density function conditioned to different values of forecasted wind (i.e. the explanatory variable).

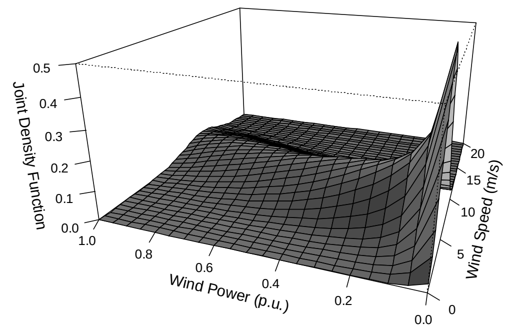

Figure 1: Joint probability density function of forecasted wind speed and measured wind power.

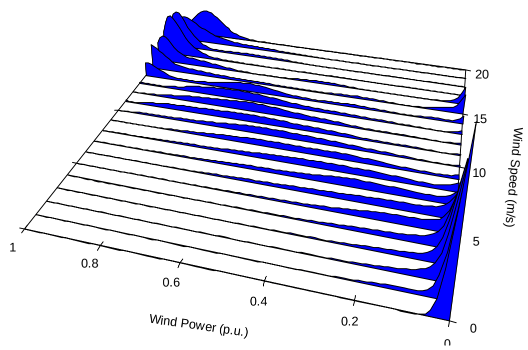

Figure 2: Conditional KDE for forecasted wind speed and wind power generation.

2.4 Kernel Function Choice

The choice of the kernel function for the wind power forecasting problem is a critical issue. The choice depends on the type of variable.

We have in the wind power problem four different variable types: wind power bounded between 0 and 1 (e.g. rated power); wind speed bounded between 0 and *+Inf*; circular variables like the wind direction; variables such as temperature, between *-Inf* and *+ Inf*. For these four types, different kernels should be considered.

For variables with range [0,1] we use the following beta kernel (used for wind power in Fig. 1) [18]:

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

where $K_{p,q}$ is the density function of a *Beta(p,q)* random variable defined by:

$$K(u;p,q) = \frac{1}{B(p,q)} \cdot u^{p-1} \cdot (1-u)^{q-1}, u \in [0,1] \qquad (8)$$

with *B(.)* denoting the beta function, *p* and *q* are the two positive shape parameters, and *h* being the bandwidth parameter.

For the variables with support [0,+Inf) the gamma kernel (used for wind speed in Fig. 1) [19] was used:

$$\hat{f}_X(x) = \frac{1}{N}\sum_{i=1}^{N} K_{x/h+1,h}(X_i) \qquad (9)$$

where $K_{p,q}$ is the density function of a *Gamma(p,q)* random variable defined by:

$$K(u;p,q) = u^{p-1} \cdot \frac{\exp(-u/q)}{\Gamma(p)\cdot q^p}, u \in [0,+\infty[\qquad (10)$$

with Γ(.) denoting the gamma function, *p* as the shape parameter, *q* as the scale parameter, and *h* is the bandwidth parameter of $K_{p,q}$.

For variables with unbounded support, the natural choices are the Gaussian kernel or the biweight kernel.

For circular variables the approach is to use circular distributions such as the von Mises distribution [20]:

$$g(\theta;\mu,\kappa) = \frac{1}{2\pi \cdot I_0(\kappa)} e^{\kappa \cdot \cos(\theta-\mu)} \qquad (11)$$

where I_0 is the modified Bessel function of the first kind and order 0 and defined by:

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} e^{\kappa \cdot \cos(\theta)} d\theta \qquad (12)$$

The parameter μ is the directional center of the distribution, κ is the concentration parameter and θ belongs to any interval of length 2π. The concentration parameter can be used to control the degree of smoothing in circular KDE, and it is analogous to the bandwidth parameter but larger values lead to less smoothing.

Note that the integrals computed from the beta and gamma kernels may lead to distributions that do not have an integral (area of the distribution) equal to one. Hence, we use the idea of a modified beta kernel estimator [21]:

$$\hat{f}'(x) = \frac{\hat{f}(x)}{\int_0^1 \hat{f}(x)dx} \qquad (13)$$

Since this is only a change of scale, the normalization is employed over the conditional function of Eq. 4.

2.5 Time-adaptive Algorithm

A recursive formula described in the literature [22] can be used for the KDE estimator:

$$\hat{f}_n(x) = \frac{n-1}{n} \cdot \hat{f}_{n-1}(x) + \frac{1}{n \cdot h_i} \cdot K\left(\frac{x-X_i}{h_i}\right) \qquad (14)$$

The extension to the multivariate case (Eq. 2) is straightforward.

Eq. 14 allows updating the density function when new samples are available without the need to entirely recompute the whole density function. However, as the number of t increases, the ratio *(n-1)/n* approaches one (and *1/n* approaches zero), and then the new samples become redundant. Moreover, if there is a change in the general structure of the data (non-stationary data), this recursive estimation is incapable of automatically discard older data.

In order to overcome these problems, the KDE estimator with exponential smoothing [22] can be used:

$$\hat{f}_n(x) = \lambda \cdot \hat{f}_{n-1}(x) + \frac{(1-\lambda)}{h_i} \cdot K\left(\frac{x-X_i}{h_i}\right) \qquad (15)$$

where λ is called *forgetting factor* and controls how quickly or slowly the exponential smoothing adapts to the new data (exponential forgetting). A value of λ close to one means that the exponential smoothing puts more weight on the historical data and little weight on the most recent values, while when λ is closer to zero it means the opposite situation; λ can be represented in terms of *n*, and we have: λ=n/(n+1).

The Nadaraya-Watson estimator described in section 2.2 can be converted to a time-adaptive estimator using Eq. 15. The estimator becomes:

$$\hat{f}_t(y \mid X=x) = \frac{\lambda \cdot \hat{f}_{t-1}(x,y) + (1-\lambda)\cdot\left[K_{h_x}\left(\frac{x-X_i}{h_x}\right)\cdot K_{h_y}\left(\frac{y-Y_i}{h_y}\right)\right]}{\lambda \cdot \hat{f}_{t-1}(x) + (1-\lambda)\cdot K\left(\frac{x-X_i}{h_x}\right)} \qquad (16)$$

where $f_t(y|X=x)$ means the knowledge of the model at time instant *t*, which is updated using recent values of *Y* and *X*.

The time-adaptive wind power forecast problem consists of the following main steps:

1. $\hat{f}_t(p_{t+k} \mid X = x_{t+k|t})$: KDF model with knowledge at time step *t*;
2. Obtain new values of measured wind power generation and corresponding NWP data for the same period. This recent data is used to update

the knowledge of the model (using Eq. 16), and the model in (1) becomes $\hat{f}_{t-1}\left(p_{t+k} \mid X = x_{t+k|t}\right)$; this process is repeated when new values are available.

3 CASE STUDIES

3.1 Description

Two different sets of data are used as case studies. The first dataset consists of day-ahead wind power forecasted and realized values for 15 hypothetical wind power sites in the state of Illinois, obtained from NREL's Eastern Wind Integration and Transmission Study [23]. We used the wind power data for the period Jan-Aug to train the uncertainty forecast models. The months between September and December are used as a test dataset.

The second dataset is from a large wind farm located in flat terrain in the U.S. Midwest. The complete dataset (SCADA and NWP) correspond to the period between January 1st 2009 and February 20th 2010. The NWP data was generated with the Weather Research and Forecasting (WRF) model [24] by Argonne National Laboratory and consists of several weather variables (e.g. wind speed, direction, temperature) for one reference point inside the wind farm.

The temporal horizon of the NWP predictions was as follows: wind power is forecasted at 6 AM for the temporal horizon of t+6 up to t+48 hours. The temporal resolution of the forecasts is one hour. The training dataset was selected to have 70% of the all available data (30% of for testing): training set from 1 January 2009 to 21 November 2009 (12169 points), and the testing set from 22 November 2009 to 20 February 2010 (5203 points).

3.2 Evaluation Framework

The results obtained with the Nadaraya-Watson estimator were compared with the linear quantile regression model and the splines quantile regression [4].

A framework to evaluate wind power probabilistic forecasts detailed in [25] was followed in this paper. Three metrics were used for evaluation: calibration, sharpness and skill score.

Calibration is a measure of the agreement between nominal proportions (forecasted probabilities) and the ones computed from the evaluation sample. In other words, for a quantile the empirical proportion should equal the nominal, e.g. an 85% quantile should contain 85% of the observed values lower or equal to its value.

In order to evaluate quantile forecasts, it is necessary to define the indicator variable. An indicator variable for a quantile forecast $\hat{q}^{\alpha}_{t+k|t}$ with nominal proportion α is:

$$\xi_k^{\alpha_i} = \begin{cases} 1\, if \;\; p_{t+k} \le \hat{q}^{\alpha_i}_{t+k|t} \\ 0\; otherwise \end{cases} \tag{17}$$

The indicator variable refers to the actual outcome of p_{t+k} at time $t+k$.

Furthermore, these indicators are defined as follows:

$$n^{\alpha}_{k,1} = \#\left\{\xi^{\alpha}_{i,k} = 1\right\} = \sum_{i=1}^{N} \xi^{\alpha}_{i,k} \tag{18}$$

$$n^{\alpha}_{k,0} = \#\left\{\xi^{\alpha}_{i,k} = 0\right\} = n^{\alpha}_{k,1} - N \tag{19}$$

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

The empirical proportions are computed with the Eq. 17 and 18 as follows:

$$\hat{\alpha}^{\alpha}_{k} = \frac{n^{\alpha}_{k,1}}{n^{\alpha}_{k,1} + n^{\alpha}_{k,0}} \tag{20}$$

The difference between empirical and nominal proportions is considered the bias of the probabilistic forecasting method.

Sharpness is the tendency of probability forecasts towards 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 $t+k$. In this paper, sharpness is measured by the mean size of the distance between quantiles:

$$\bar{\delta}^{\alpha}_{k} = \frac{1}{N}\sum_{i=1}^{N} \delta^{\alpha}_{i,k} \tag{21}$$

We also calculated a skill score from Eq. 22 which gives information about a model's performance (e.g. calibration, sharpness, etc.) in a single measure for a set of m quantiles:

$$S_c\left(\hat{f}_{t+k}, p_{t+k}\right) = \sum_{i=1}^{m}\left(\xi^{\alpha_i} - \alpha_i\right)\left(p_{t+k} - \hat{q}^{\alpha_i}_{t+k}\right) \tag{22}$$

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 of Eq. 17. The higher the scoring rule, the better: the maximum value is 0 for perfect probabilistic forecasts.

For reasons of comparison, the probabilistic forecast is represented through a set of quantiles ranging from 5% to 95% with a 5% increment.

3.3 Evaluation Results: NREL's EWITS Study

3.3.1 Offline Results

The kernel function used in the Nadaraya-Watson (NW) estimators was Chen's beta kernel for both realized and forecasted wind power (i.e. the explanatory variable). The kernel size was 0.001 for both variables (determined experimentally by trial-error).

Fig. 3 shows the average calibration for the whole time horizon (24 hours) for probabilistic forecasts obtained with the linear quantile regression (Linear QR), splines quantile regression (Splines QR) and the NW estimator. Note that what is depicted in the diagram is the difference between forecasted and empirical quantile proportions.

For the quantiles above 55% the NW estimator present a lower deviation than the QR methods. For quantiles below the median the splines QR is competitive with the NW, and for some quantiles it achieves the lowest deviation.

On average, the methods overforecast the quantiles since the forecasted quantile proportions are greater than the empirical ones. The tests performed with different bandwidths showed that by changing the kernel bandwidth the model changes from underestimation to overestimation and vice-versa.

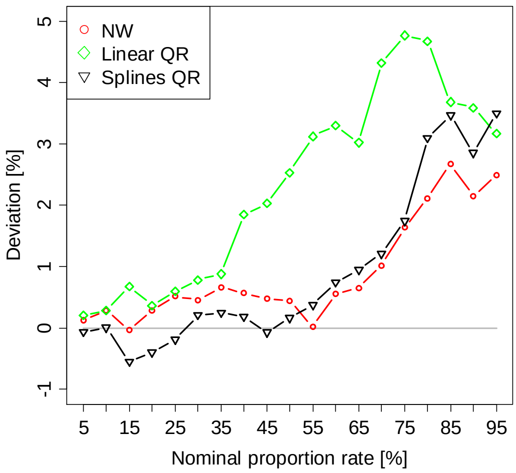

Figure 3: Calibration diagram for the offline test with NREL data.

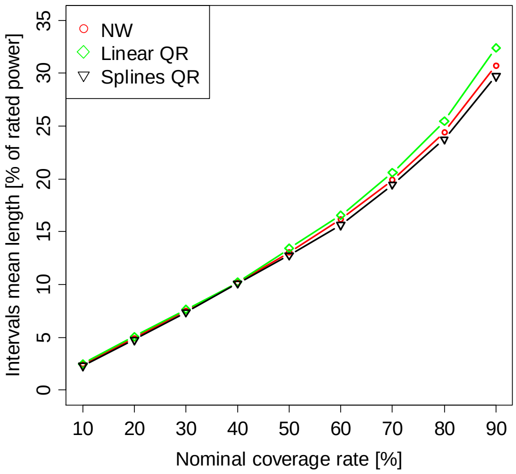

Figure 4: Sharpness diagram for the offline test with NREL data.

Fig. 4 depicts a sharpness diagram where the x-axis is the nominal coverage of the forecast interval (1-α) and the y-axis is the average size of the intervals.

In this case what is desired is to have intervals with smaller size for all coverage rates. In terms of sharpness the forecasted quantiles presented relatively narrow amplitudes for all methods, although splines QR has the lowest sharpness.

There is a trade-off between reliability and sharpness, meaning that improving the reliability will generally degrade the sharpness and vice-versa [14].

3.3.2 *Time-adaptive Algorithm: Proof of Concept*

The aim of this section is to demonstrate the validity of the time-adaptive concept presented in section 2.5. However, in order to introduce a change in the data structure, we "disconnected" two sites (one of 211.6 MW and another of 616.1 MW, out of a 5.19 GW total) during Jan-Sep and "connected" them after Oct.

This change was created artificially; however, it reproduces a situation that could actually happen. For instance: a system operator is receiving forecasts from 13 wind farms (these forecasts are summed up and estimates for the uncertainty associated to the total wind power generation are produced); then, in October two new wind farms are connected to the grid. In this case, the knowledge from past observations is no longer valid. By using a time-adaptive model the system operator is able to adapt to the new situation without requiring an offline training of the model.

Fig. 5 depicts the calibration diagram obtained with the offline and the time-adaptive NW estimators with three different values of λ.

Due to the increase in the wind power generation with the connection of two wind farms, it is expected that the offline model gives an underestimation of the quantiles for values below the 50% quantile and an overestimation of the quantiles for greater values. As an example, the 95% quantile means that the probability of having a wind generation above its value is only 5%; however, the empirical analysis with the offline approach shows that this probability is 13%.

With the time-adaptive version, under and overestimations are partly corrected. The calibration obtained with λ equal to 0.999 and 0.995 is much better than the offline approach. For instance, for the quantile 95% the empirical proportions obtained with the time-adaptive approach is 92.3% with λ=0.999 and 91.2% with λ=0.995.

3.4 *Evaluation Results: Midwest Wind Farm*

3.4.1 *Offline Results*

The following kernel functions were used: Chen's beta kernel with h=0.008 for the wind power generation; Chen's gamma kernel with h=0.05 for the wind speed forecast; von Mises distribution with κ= 2.5 for the forecasted wind direction; Chen's beta kernel with h=0.1 for the look-ahead time step. The kernel bandwidth values were determined experimentally (trial-error) and using as starting point the values suggested by the R package "hdrcde" [26].

Fig. 6 depicts the calibration obtained with NW and splines QR (with 6 degrees of freedom). The best calibration performance is from the NW estimator. As previously mentioned, due to the trade-off between calibration and sharpness, it is expected from the splines QR a better sharpness performance, as depicted in Fig. 7.

Fig. 8 depicts the skill score computed for each look-ahead time step for both NW and QR estimators. The NW estimator has almost the same performance as QR in terms of skill score. QR is better than KDF for some look-ahead steps, but it is worse in others.

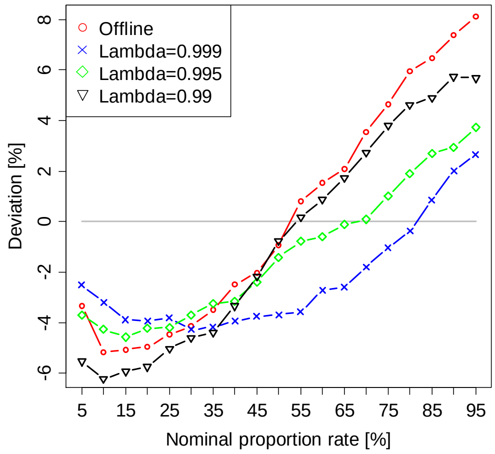

Figure 5: Calibration diagram for the NREL dataset with concept change.

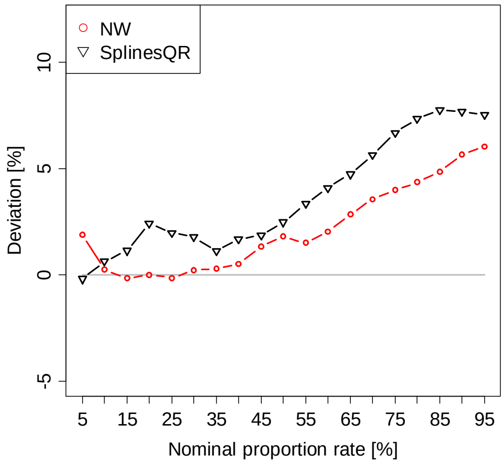

Figure 6: Calibration diagram for the Midwest wind farm with offline NW and QR estimators.

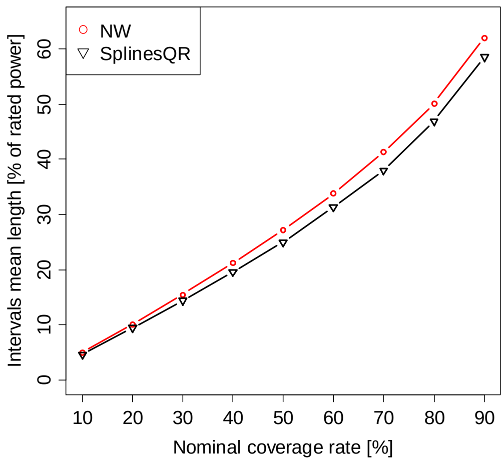

Figure 7: Sharpness diagram for the Midwest wind farm with offline NW and QR estimators.

Note that the skill score does not inform on the contributions from calibration or sharpness. Hence, calibration should be assessed (as a primary requirement), and then the information provided by skill score allows deriving conclusions about the remaining metrics [25].

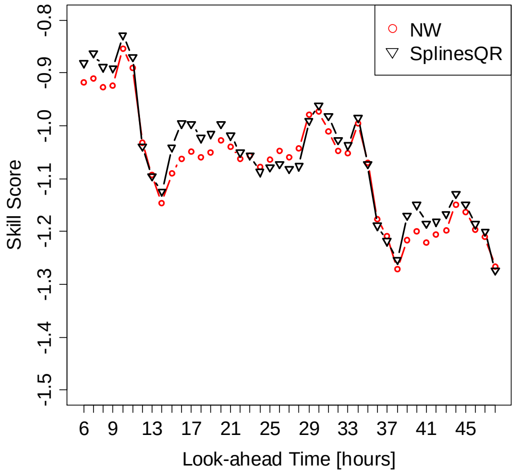

Figure 8: Skill score diagram for the Midwest wind farm with offline NW and QR estimators.

3.4.2 *Time-adaptive Results*

The time-adaptive NW version was compared with the offline version for different values of the forgetting factor (λ). For a better understanding, λ was represented by the corresponding *n* value. 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 kernel and bandwidths as in the offline version was used.

Fig. 9 depicts the calibration results. The time-adaptive version with n=2738 and n=1000 achieved the best performance, while a small number of points in the sliding window leads to a worse performance comparing to the offline results. The version with higher λ does not have a significant impact on the sharpness, as depicted in Fig. 10.

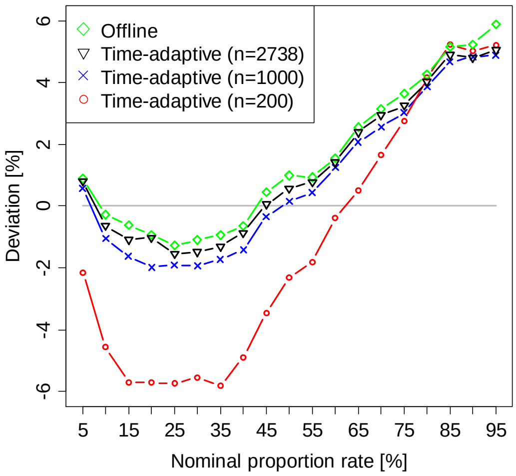

Figure 9: Calibration diagram for the Midwest wind farm with offline and time-adaptive NW.

Fig. 11 depicts the skill score for both versions. The best performance was obtained with 2738 and 1000 points. The difference between offline and time adaptive versions is only noticeable in the first 28 look-ahead time steps.

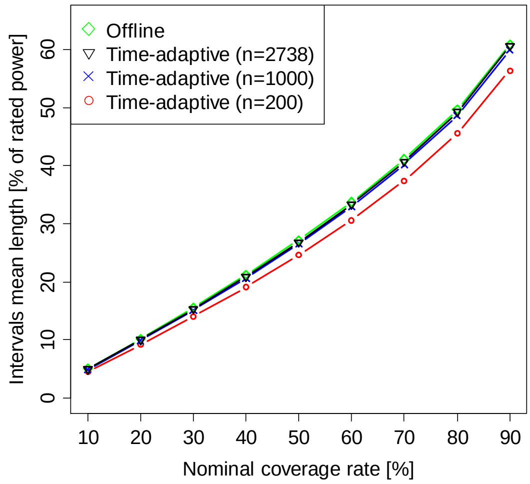

Figure 10: Sharpness diagram for the Midwest wind farm with offline and time-adaptive NW.

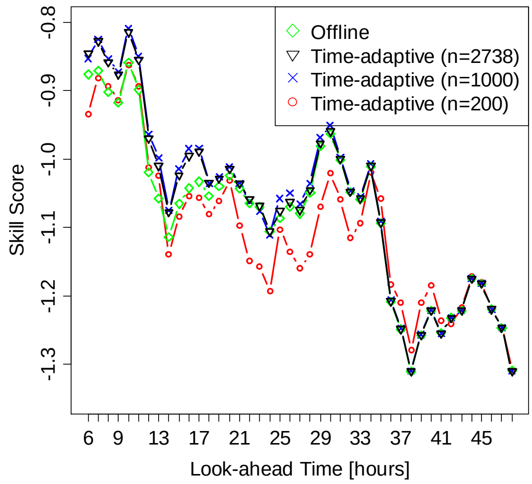

Figure 11: Skill score diagram for the Midwest wind farm with offline and time-adaptive NW.

4 CONCLUSIONS

This paper presents a new approach to estimate the uncertainty in short term wind power forecasts, based on kernel density estimation, including a new time-adaptive model.

Our studies demonstrated that kernel density forecasts with the NW estimator have a tendency to present better performance in terms of calibration, while the QR methods have a tendency to present a better sharpness performance. The skill score of both methods is rather similar. We must underline that the calibration metric is the primary requirement for wind power probabilistic forecasting.

The new time-adaptive version improves the bias of the probabilistic forecasts (calibration), while only slightly changing the sharpness; it improves the skill score when compared with the offline approach.

From a qualitative perspective, density estimation models offer an important advantage over other models. The output is a complete description of the forecasted *pdf*, something that is of importance for several decision problems in the power system domain. Moreover, the full *pdf* is very valuable for forecasting multimodal distributions (i.e. compute the modes instead of just computing the expected value). Hence, the work reported in this paper is seen as a valid contribution to the wind power forecasting activity.

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 some of the wind farm data used in the analysis.

The author Ricardo J. Bessa acknowledges Fundação para a Ciência e a Tecnologia (FCT) for PhD Scholarship SFRH/BD/33738/2009.

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, no. 4, pp. 1657-1666, Nov. 2009.

[2] R.J. Bessa, V. Miranda, A. Botterud and J. Wang, "'Good' or 'Bad' Wind Power Forecasts: A Relative Concept", *Wind Energy*, 2010, *In Press*. DOI: 10.1002/we.444

[3] C. Monteiro, R. 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, Nov. 2009.

[4] 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, pp. 95-108, 2006.

[5] P. Pinson and G. Kariniotakis, "Conditional Prediction Intervals of Wind Power Generation", *IEEE Trans. on Power Sys.*, vol. 25, no. 4, pp. 1845-1856, Nov. 2010.

[6] N.J. Cutler, H.R. Outhred, I.F. MacGill, M.J. Kay and J.D. Kepert, "Characterizing Future Large, Rapid Changes in Aggregated Wind Power Using Numerical Weather Prediction Spatial Fields", *Wind Energy*, vol. 12, pp. 542-555, 2009.

[7] E. Constantinescu, V. Zavala, M. Rocklin, S. Lee, and M. Anitescu, "A Computational Framework for Uncertainty Quantification and Stochastic Optimization in Unit Commitment with Wind Power Genera-

tion", *IEEE Trans. on Power Sys.*, vol. 26, no. 1, pp. 431-441, 2011.

[8] A. Botterud, J. Wang, R.J. Bessa, H. Keko and V. Miranda, "Risk Management and Optimal Bidding for a Wind Power Producer," IEEE PES General Meeting, Minneapolis, USA, 25-29 Jul. 2010.

[9] M.A. Matos and R.J. Bessa, "Setting the Operating Reserve Using Probabilistic Wind Power Forecasts", *IEEE Trans. on Power Sys.*, vol. 26, no. 2, pp. 594-603, May 2011..

[10] J. Wang, A. Botterud, R. Bessa, H. Keko, L. Carvalho, D. Issicaba, J. Sumaili, V. Miranda, "Representing Wind Power Forecasting Uncertainty in Unit Commitment", *Applied Energy*, 2011, *In Press*: DOI: 10.1016/j.apenergy.2011.04.011.

[11] J. Usaola, "Probabilistic Load Flow with Correlated Wind Power Injections", *Elect. Power Sys. Res.*, vol. 80, no. 5, pp. 528-536, May 2010.

[12] 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, no. 1, pp. 51-62, 2009.

[13] J.K. Møller, H.A. Nielsen and H. Madsen, "Time-Adaptive Quantile Regression", *Comp. Stat. & Data Anal.*, vol. 52, no. 3, pp. 1292-1303, Jan. 2008.

[14] J. Juban, N. Siebert and G. Kariniotakis, "Probabilistic Short-term Wind Power Forecasting for the Optimal Management of Wind Generation" IEEE Power Tech Conference, Switzerland, July 2007.

[15] M. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function", *The Annals of Math. Stat.*, vol. 27, no. 3, pp. 832-837, 1956.

[16] M.P. Wand and M.C. Jones, "Multivariate Plug-in Bandwidth Selection", *Comp. Stat.*, vol. 9, pp. 97-116, 1994.

[17] R.J. Hyndman, D.M. Bashtannyk and G.K. Grunwald, "Estimating and Visualizing Conditional Densities", *Journal of Comp. and Graph. Stat.*, vol. 5, no. 4, pp. 315-336, Dec. 1996.

[18] S.X. Chen, "Beta Kernel Estimators for Density Functions", *Comp. Stat. & Data Anal.*, vol. 31, no. 2, pp. 131-145, 1999.

[19] S.X. Chen, "Probability Density Function Estimation using Gamma Kernels", *Annals of the Inst. of Stat. Math.*, vol. 52, no. 3, pp. 471-480, Sept. 2000.

[20] K. V. Mardia and P. E. Jupp, "Directional Statistics", Wiley, Nov. 1999. ISBN: 978-0-471-95333-3

[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] 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, no. 4, pp. 893–914, 2003.

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

[24] W. Skamarock, *et al.*, "A Description of the Advanced Research WRF Version 2", NCAR/TN–468+STR Technical note, Jun. 2005.

[25] 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, no. 6, pp. 497-516, Nov. 2007.

[26] R.J. Hyndman, J. Einbeck and M. Wand, "Highest Density Regions and Conditional Density Estimation. Package hdrcde", Technical Report, Nov. 2010.