Time-Adaptive Quantile-Copula for Wind Power Probabilistic Forecasting

R.J. Bessa[1*], V. Miranda[1], A. Botterud[2], Z. Zhou[2], J. Wang[2]

[1]*Instituto de Engenharia de Sistemas e Computadores do Porto (INESC Porto), Faculty of Engineering, University of Porto, Campus da FEUP Rua Dr. Roberto Frias, 378, 4200 - 465 Porto Portugal*
[2]*Argonne National Laboratory, 9700 S. Cass Avenue, Bldg. 221, Argonne, IL 60439, USA*

Abstract

This paper presents a novel time-adaptive quantile-copula estimator for kernel density forecast and a discussion of how to select the adequate kernels for modeling the different variables of the problem. Results are presented for different case-studies and compared with splines quantile regression (QR). The datasets used are from NREL's Eastern Wind Integration and Transmission Study, and from a real wind farm located in the Midwest region of the United States. The new probabilistic forecasting model is elegant and simple and yet displays advantages over the traditional QR approach. Especially notable is the quality of the results achieved with the time-adaptive version, namely when evaluated in terms of forecast calibration, which is a characteristic that is advantageous for both system operators and wind power producers.

Keywords: Wind power, forecasting, probabilistic, Kernel, density estimation, copula, time-adaptive.

1. Introduction

The importance of forecasting wind power has grown in parallel with the increase in penetration of wind generation on interconnected power systems. It can be loosely said, for the short term, that the uncertainty in wind power prediction is higher than in load forecasting. This very fact has strong implications for the security and costs associated with decision making in systems with high penetration of wind power, for instance in generator scheduling and determination of operating reserve margins. Furthermore, in regions where there is a transition from a feed-in tariff scheme to direct market participation for wind power, an accurate representation of forecasting uncertainty also has an important function in controlling the trade-off between risk and return when wind energy is scheduled in the electricity market.

This necessity of having uncertainty estimation and characterization in wind power forecasting for the next hours/days has motivated the development of advanced physical and statistical based approaches. State-of-the-art algorithms in wind power uncertainty forecasting are referenced in a recent comprehensive report [1].

Two of the most popular statistical methods are: splines quantile regression [2], which consists of a linear quantile regression with the base functions formulated as cubic B-splines, and adapted resampling [3], which uses a fuzzy inference model and resampling to determine the distributions of forecast errors associated with the power output forecast. As an alternative, physics based models can also be used; two examples are: meteorological ensembles [4] and spatial fields displaying wind forecast information from multiple grid points [5].

Three key features have taken researchers' attention: i) the representation of wind power uncertainty; ii) the chain of models for uncertainty forecasting; iii) time-adaptive (or online) models to cope with evolving data streams.

Wind power uncertainty can take the form of probabilistic forecasts, risk indices [6], or scenarios

*Correspondence to: Ricardo Bessa, INESC Porto, Campus da FEUP Rua Dr. Roberto Frias, 378, 4200 - 465 Porto Portugal.
E-mail: rbessa@inescporto.pt

[5][7] for short-term wind power generation. Probabilistic forecasting consists of expressing the wind power generation or forecast error in "probabilistic terms", such as: a) parametric representation (e.g. Gaussian distribution); b) moments of the distributions (e.g. mean, standard deviation, skewness); c) a set of quantiles; d) probability mass function (*pmf*); and e) 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.

The traditional model chain for wind power uncertainty forecasting, according to Juban *et al.* [8], consists of using as input the forecast errors or point forecasts from a wind power deterministic forecasting model. The uncertainty estimation model is placed after the model that produces deterministic wind power forecasts. A preferred approach consists of either using the Numerical Weather Prediction (NWP) forecast error as input for the uncertainty estimation method or computing the uncertainty directly from the NWP points. This class of algorithms avoids an intermediate step (conversion of wind to power) since probabilistic and deterministic forecasts can be produced directly from the NWP. For instance, the local quantile regression described by Bremnes [9] forecasts the wind power generation quantiles based on information about explanatory variables (e.g. NWP forecasts); a set of quantiles characterize the uncertainty, and the point forecast could be associated to the median (quantile 50%). The Kernel Density Estimation described by Juban *et al.* [10] also provides uncertainty estimation and point forecasts.

Some models available in the literature are trained in an offline mode, meaning that the models are unable to cope with changes in the underlying distributions of the several variables. Examples of offline forecasting algorithms are the quantile regression presented by Bremnes [9] and the model described by Juban *et al.* [10]. On the other hand, the tendency in the state of the art is to develop algorithms capable of adapting to changes in data; one example is the time-adaptive quantile regression model described by Møller *et al.* [11].

Consequently, an algorithm for wind power uncertainty forecasting shall ideally have as requisites: i) a high flexibility to represent wind power uncertainty; ii) time-adaptive characteristics, and iii) avoiding an intermediate step that computes wind power point forecasts, which is particular important when both point and uncertainty forecasts are time adaptive. However, point forecasts from different models could represent additional and useful information for uncertainty forecasting.

In this paper, a kernel density forecast (KDF) method, which respects the three above mentioned requisites, is described. The main differences that this methodology displays, in comparison to the current state-of-the-art in kernel density estimation (e.g. [10]) are: i) a time-adaptive version of the quantile copula estimator [12] is described and tested in a real wind farm; ii) our method is based on selecting adequate kernels for modeling the different variable types of the wind power problem.

This paper is organized as follows: section 2 presents the motivation for representing the wind power uncertainty by density functions; section 3 describes the kernel density forecast methodology: the time-adaptive quantile-copula estimator; section 4 presents the evaluation results for two case-studies with datasets from NREL's Eastern Wind Integration and Transmission Study (EWITS) and from a real U.S. wind farm; section 6 presents the conclusions.

2. Motivation to Represent Wind Power Uncertainty by Probability Density Functions

From an information theory perspective, the probability density function (*pdf*) contains all the information associated with a random variable. For instance, it enables the computing of the moments of the forecasted distribution. Therefore, we may argue that the *pdf* is generic and can be transformed into several uncertainty forms of representation, such as quantiles, standard deviation, or skewness.

The best way to represent uncertainty is determined by the end user's requests and the nature of the decision-making problem being addressed. In general, one cannot talk about better and worse uncertainty representations, only of more or less adequate representations (for a similar discussion regarding point forecasts, see [13]). However, the *pdf* by itself gives the necessary flexibility for several decision-making problems.

The problem of finding the "optimal" wind power bidding strategy for the electricity market can be formulated with different methods when wind power uncertainty is considered. When the objective is to maximize the expected profit (or minimize the expected cost of imbalances) the aim consists of finding the optimal quantile, which for some electricity markets is determined by imbalances in price ratios [9]. It is possible to extract the optimal quantile from the *pdf* for each hour and, consequently, the "optimal" decision under the expected value paradigm. Botterud *et al.* [14] presented an approach based on maximizing the utility; for this approach the *pdf* enables the production of a probability mass function (*pmf*) that can be used to compute the expected utility. Bourry *et al.* [15] described an approach based on portfolio theory where a trade-off between expected income and risk (described by the conditional value-at-risk) is evaluated to find the "optimal" bid. For this problem, the knowledge of the *pdf* allows the computation of any risk measure. For instance, it is possible to evaluate a trade-off between expected income and risk described by the variance and skewness.

The *pdf* representation is also useful for system operators in setting the required operating reserve for the current and next days, using for instance the method presented by Matos and Bessa [16]. The *pdf* representation provides the full probability distribution, which allows for a better characterization of the tails. According to Bessa and Matos [17] the tails in the operating reserve problem are the critical factor, in particular if the system operator prefers a higher security (e.g. loss of load probability around 0.1%).

The method described by Pinson *et al.* [7] to represent the uncertainty by scenarios with temporal correlation of forecast errors could also benefit from the *pdf* representation. With a forecasted *pdf* the distribution is fully characterized and there is no need to perform an exponential interpolation.

For multimodal distributions a density forecast allows the computation of the modes instead of just computing the expected value (which in this case is not a good summary of the distribution). In recent practice, it is unlikely to find wind power multimodal density distributions, but the adoption of ensemble approaches may favor this possibility. In any case, the mode or the median is still a better deterministic forecast, because normally the wind power distributions are highly skewed.

3. Kernel Density Forecast Methodology

3.1 Kernel Density Estimation

Kernel Density Estimation (KDE) consists of a nonparametric estimator of a density function, and was introduced by Rosenblatt [18], with several properties derived by Parzen [19]. 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:

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

where *N* is the number of samples, *K* is a kernel function and *h* the bandwidth parameter.

Eq. 1 represents the placing of a kernel at each sample X_i. The corresponding estimated density function is derived from dividing by *N* the sum of the *N* kernels centered on each sample.

Given i.i.d multivariate data $X_1,\ldots,X_d$ from *d* different variables drawn from an unknown multivariate density function *f*, the multivariate KDE [20] is given by:

$$\hat{f}(x_1,\ldots,x_d) = \frac{1}{N \cdot h_1 \cdot \ldots \cdot h_d} \sum_{i=1}^{N} \prod_{j=1}^{d} K_j\left(\frac{x_j - X_{ij}}{h_j}\right) \tag{2}$$

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

3.2 Conditional Kernel Density Estimation

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 on *X*=*x*, i.e. $f(y|X=x)$. The conditional density can be formulated as follows

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

where *f(x,y)* is the multivariate density function of *X* and *Y* (joint distribution function) and *f(x)* is the marginal density of *X*.

It is also possible to have nonparametric conditional density estimation for Eq. 3. The classic approach is the Nadaraya-Watson kernel smoother proposed by Rosenblatt [21]:

$$\hat{f}(y \mid X = x) = \frac{\hat{f}_{XY}(x, y)}{\hat{f}_X(x)} = \sum_{i=1}^{N} K_{h_y}\left(\frac{y - Y_i}{h_y}\right) \cdot w_i(x) \tag{4}$$

where

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

The wind power density forecasting 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, e.g. Numerical Weather Prediction (NWP) forecasts, wind power measured values, hour of the day.

Translating this sentence to an equation, we have:

$$\hat{f}(y_{t+k} \mid X = x_{t+k|t}) = \frac{f_{Y,X}(y_{t+k}, x_{t+k|t})}{f_X(x_{t+k|t})} \tag{6}$$

where y_{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 the approach that is presented in the next section

3.3 Quantile-Copula Estimator

The quantile-copula estimator was introduced by Faugeras [12]. According to the authors, its main advantages over the existing methods are: the methods based on the NW estimator are numerically unstable when the denominator is close to zero; for a problem with several explanatory variables, this method has only one kernel product, instead of two; at a conceptual level, density estimation should only be based on density estimation methods and not on regression approaches.

The main difference from the estimator in Eqs. 4 and 5 is in the joint density function of *Y* and *X*. Almost at the same time, Faugeras [12] and Bouezmarni *et al.* [22] proposed the idea of using a copula for modeling the dependency structure between *Y* and *X*. Regarding copulas, the Sklar theorem [23] says the following for the bivariate case:

Let H be a two-dimensional distribution function with marginal distribution functions F and G. Then there is a copula C such that

$$H(x, y) = C(F(x), G(y)) \tag{7}$$

Conversely, for any univariate distribution functions F and G and any copula C, the function H is a two-dimensional distribution function with marginals F and G. Furthermore, if F and G are continuous, then C is unique.

This theorem means that the multivariate distribution function can be separated into two parts: i) marginal functions that can be estimated separately; ii) dependency structure between the marginal which is modeled by the copula. For more details about copulas see Nelson [24]. A conditional density estimator can be built departing from the definition of Eq. 7. So, we know that

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)) \tag{8}$$

then Eq. 2 (for the bivariate case) can be replaced by

$$f(x,y) = \frac{\partial^2}{\partial u \cdot \partial v} C(u,v) = f_X(x) \cdot f_Y(y) \cdot c(u,v) \quad (9)$$

where *u* and *v* are a quantile transform of the data, $u=F_X(x)$ and $v=F_Y(y)$, and *c* is the copula density function.

Replacing Eq. 9 in Eq. 3, we have the following conditional density estimator:

$$f(y \mid X = x) = f_Y(y) \cdot c(u,v) \quad (10)$$

Now it is necessary to build a nonparametric estimator for Eq. 10. The idea proposed by Bouezmarni *et al.* [22] was a semiparametric approach, where a parametric model is used for the copula, and the marginal distributions are represented by a nonparametric model (empirical distribution function). However, we followed the idea described by Faugeras [12], where the copula density is estimated with KDE.

The estimator for $f_Y(y)$ is the KDE in Eq. 1. The copula density estimator is the estimator in Eq. 2 as follows:

$$\hat{c}(u,v) = \frac{1}{N \cdot h_u \cdot h_v} \sum_{i=1}^{N} K_{h_u}\left(\frac{u - U_i}{h_u}\right) \cdot K_{h_v}\left(\frac{v - V_i}{h_v}\right) \quad (11)$$

where U_i and V_i are the data transformed by the empirical cumulative distribution function, i.e. $U_i=F_X^e(X_i)$ and $V_i=F_Y^e(Y_i)$. An empirical cumulative distribution function (*cdf*) is defined as:

$$F^e(t) = \frac{1}{N} \sum_{i=1}^{N} I(x_i \le t) \quad (12)$$

where *I* is the indicator function of event $x_i \le t$.

Figure 1 depicts the copula density computed with Eq. 11 for the quantile transform of the wind speed (variable *v*) and wind power (variable *u*) from a real wind farm (case-study of section 4). This copula density function represents the probability density associated to each point plotted in the wind speed against wind power scatter plot in Figure 2.

The copula represents the dependence structure between the two variables. Therefore, we see that there is a strong dependence in the two extreme corners, e.g. when there are lower quantiles in wind speed (lower wind speed values) the wind power quantiles also present lower values with a higher probability.

An interesting conclusion is that this copula density seems very similar to a family of parametric copulas, the elliptical copulas [25].

The quantile-copula conditional KDE is written as:

$$\hat{f}(y \mid X = x) = \frac{1}{N \cdot h_y} \cdot \sum_{i=1}^{N} K_{h_y}\left(\frac{y - Y_i}{h_y}\right) \cdot \frac{1}{N \cdot h_u \cdot h_v} \cdot \sum_{i=1}^{N} K_{h_u}\left(\frac{F_X^e(u) - F_X^e(U_i)}{h_u}\right) \cdot K_{h_v}\left(\frac{F_Y^e(v) - F_Y^e(V_i)}{h_v}\right) \quad (13)$$

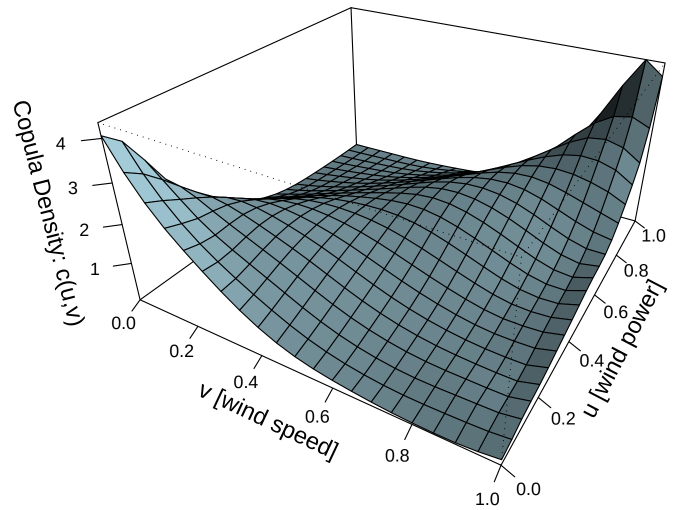

Figure 1: Bivariate copula density function of quantile transforms of forecasted wind speed (v) and measured wind power (u).

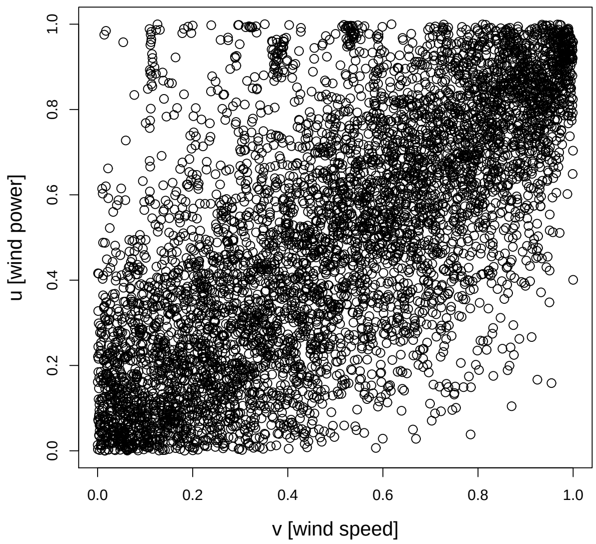

Figure 2: Scatter plot of quantile transform of forecasted wind speed (v) against measured wind power (u).

3.4 Kernel Functions

The choice of the kernel function K depends on the type of variable. In what regards the data type, we have in the wind power problem three different types: i) wind power bounded between 0 (e.g. zero generation) and 1 (e.g. rated power); ii) wind speed bounded between 0 and $+\infty$; iii) circular variables like the hour of the day and the wind direction. For these three types, different kernels should be considered.

We need two different types of kernels for the quantile-copula estimator: i) a kernel for variables bounded between 0 and 1, i.e., after normalizing, wind power and the quantile transforms variables u and v; ii) circular kernel for the wind direction.

For variables bounded between 0 and 1 the following beta kernel proposed by Chen [26] was adopted:

$$\hat{f}_1(x) = \frac{1}{N} \cdot \sum_{i=1}^{N} K_{x/h+1,(1-x)/h+1}(X_i) \tag{14}$$

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 of $K_{p,q}$.

Figure 3 and Figure 4 depicts the beta kernel shape for five different points and the Gaussian kernel for the same points respectively. 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 kernels are non-negative (and consequently the estimator), in contrast to the Gaussian kernel. The Gaussian kernel shape is fixed for any value of x. The Gaussian kernel may lead to inconsistent results at the boundaries.

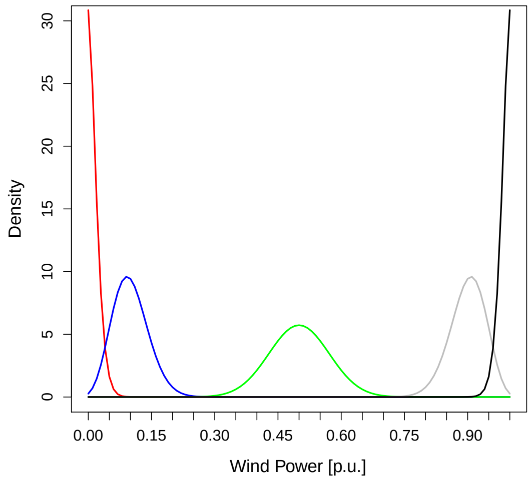

Figure 3: Beta kernels of Eq. 14 for h=0.02 (x=0.01, x=0.1, x=0.5, x=0.9, x=0.99).

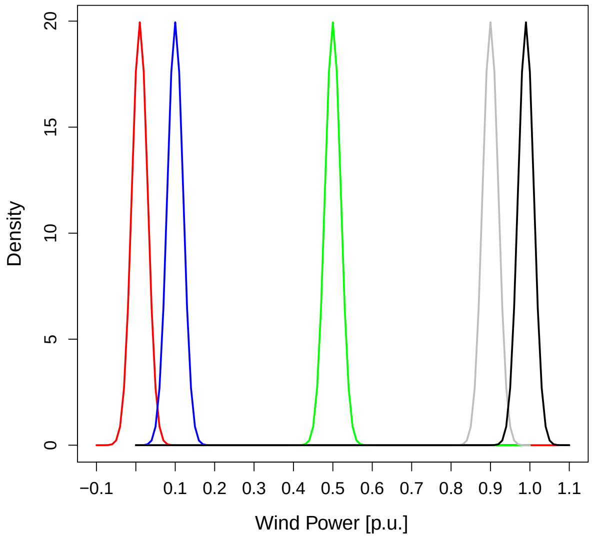

Figure 4: Gaussian kernels for h=0.02 (x=0.01, x=0.1, x=0.5, x=0.9, x=0.99).

As mentioned in [27] the integrals computed from the beta kernels may not converge to their theoretical counterpart. This may lead to distributions that do not have an integral equal to 1. Moreover, the kernel is also inconsistent for distributions that are point mass at 0% and 100%. This is due to lack of normalization, and the idea proposed in [27] consists of a modified beta kernel estimator (named "macro-beta"):

$$\hat{f}'(x) = \frac{\hat{f}(x)}{\int_0^1 \hat{f}(x)dx} \tag{15}$$

Since this is only a change of scale, the normalization for the conditional KDE is employed over the conditional function of Eq. 13.

Circular variables are a particular case for KDE. For instance, the angular difference between a wind direction of 350° and 10° is only 20°, while the linear real valued distance is 340°. The approach is to use circular distributions such as the wrapped normal distribution or the von Mises distribution [28]. In this case, and since it is mathematically more simple and a close approximation to the wrapped normal distribution, we used the von Mises distribution. The von Mises distribution is given by:

$$g(\theta;\mu,\kappa) = \frac{1}{2\pi \cdot I_0(\kappa)} e^{\kappa \cdot \cos(\theta-\mu)} \tag{16}$$

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

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} e^{\kappa\cdot\cos(\theta)} d\theta \tag{17}$$

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 circular kernel is used for the quantile transform of the wind direction. Therefore, it is necessary to perform a change of scale from [0,1] to [0, 6.266] (in radians).

3.5 Time-Adaptive Estimator

Wegman and Davies [29] introduced a recursive estimator of KDE for Eq. 1. The density function can be calculated recursively using the following:

$$\hat{f}_n(x) = \frac{n-1}{n}\cdot \hat{f}_{n-1}(x) + \frac{1}{n\cdot h_i} K_h\left(\frac{x-X_i}{h_i}\right) \tag{18}$$

The extension to the multivariate case (Eq. 2 and 11) is straightforward.

Eq. 18 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 generating structure of the data (non-stationary data), this recursive estimation is incapable of automatically discard older data.

In order to overcome these problems, Wegman and Marchette [30] proposed the KDE estimator with exponential smoothing, and the KDE formulation with adjustable discarding of old data becomes:

$$\hat{f}_n(x) = \lambda\cdot \hat{f}_{n-1}(x) + \frac{(1-\lambda)}{h_i} K_h\left(\frac{x-X_i}{h_i}\right) \tag{19}$$

where λ is called forgetting factor and controls how quickly or slowly the exponential smoothing adapts to the new data (exponential forgetting); λ replaces *(n-1)/n* and $(1-\lambda)$ replaces *1/n*, and its value should be slightly below one.

The quantile-copula estimator described in section 3.4 can be converted to a time-adaptive estimator using Eq. 19. The quantile transform function (the empirical cumulative distribution function of Eq. 12) is transformed to time-adaptive using the following equation:

$$F^e(x)_t = \lambda_e \cdot F^e(x)_{t-1} + (1-\lambda_e)\cdot I(X_i \le x) \tag{20}$$

where λ_e is the forgetting factor of the empirical cumulative distribution.

The estimator becomes

$$\hat{f}(y \mid x = X)_t = \hat{f}_t(y)\cdot \hat{c}_t(u,v) \tag{21}$$

where

$$\hat{f}_t(y) = \lambda\cdot \hat{f}_{t-1}(y) + \frac{(1-\lambda)}{h_y}\cdot K\left(\frac{y-Y_i}{h_y}\right) \tag{22}$$

and

$$\hat{c}_t(u,v) = \lambda\cdot \hat{c}_{t-1}(u,v) + \frac{(1-\lambda)}{h_x\cdot h_y}\cdot\left[K_{h_x}\left(\frac{F_X^e(u)-F_X^e(U_i)}{h_x}\right)\cdot K_{h_y}\left(\frac{F_Y^e(v)-F_Y^e(V_i)}{h_y}\right)\right] \tag{23}$$

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

Note that different values of λ should be defined for Eq. 20 and 22-23, because the quantile transform of the data should change with a lower rate otherwise the model could become unstable.

In the wind power forecast problem, when new values of measured wind generation and NWP data are available, this recent data are used to update the knowledge of the model.

Nevertheless, data pre-processing for filtering erroneous measurements from SCADA systems is crucial for these methods, otherwise the time adaptive training is contaminated and model performance degrades.

4. Case-Studies

4.1 Description

Two different datasets were used as case-studies. The first dataset consists of hourly day-ahead wind power point forecasts and realized wind power generation for 15 hypothetical sites in the state of Illinois, within the MISO footprint for 2006 obtained from NREL's EWITS study [31]. The data were produced by combining a mesoscale NWP model with a composite power curve for a number of potential sites for wind power farms. The day-ahead forecasts were generated based on observed forecast errors from four real wind power plants. The resulting Markov chain forecast models for each of the four sites were randomly assigned to the sites in the data set. The data methodology is explained in [32].

We used the wind power data (forecasts and realized generation) for the period from January to August to train the uncertainty forecast model. The months between September and December were used as a validation dataset. The explanatory variable for this dataset is the point forecast.

The second dataset is from a large-scale wind farm located in a flat terrain in the U.S. Midwest. The complete dataset (SCADA and NWP) correspond to the period between January 2nd 2009 and February 20th 2010. The NWP data was generated by the WRF model [33] at Argonne National Laboratory for one reference point located in the wind farm at 6 AM every day.

The temporal horizon of the predictions used was t+6 up to t+48 hours; the very-short term (up to 6 h) is not addressed in this paper.

The required temporal resolution for wind power forecasts is usually one hour, compatible with market purposes. Both the SCADA and NWP data in the second dataset have temporal resolution of 10 minutes, so a simple average of the 10 minute data was performed to produce hourly data.

The training dataset was selected to have 70% of all examples (30% of examples for testing). Hence, the training set is from 1 January 2009 to 21 November 2009 (12169 points), and the testing set from 22 November 2009 to 20 February 2010 (5203 points). The explanatory variables for these data are the wind speed and direction forecasts from the NWP model, and the look-ahead time step.

4.2 Evaluation Framework

A framework to evaluate wind power probabilistic forecasts is detailed in [34]. The evaluation set consists of a series of quantile forecasts for unique or varying nominal proportions and observations (measured values). The presented classification can be unconditional, but because several variables might influence the quality of probabilistic forecasts, the evaluation can also become conditional in order to reveal the influence of such variables (e.g. by look-ahead time step).

For the evaluation purpose, three metrics were considered: calibration (or reliability), sharpness, and skill score.

A requirement for probabilistic forecasts is that the nominal probabilities (or nominal proportions) of quantile forecasts match the observed probabilities. In other words, in an infinite series of probabilistic forecasts, observed proportions should equal the pre-assigned probability exactly. This property is commonly referred to as reliability or calibration. Here, we will use calibration to avoid confusion with power system reliability. The difference between observed and nominal probabilities is the bias of the probabilistic forecasting method.

Sharpness is the tendency of probability forecasts towards discrete forecasts, measured by the mean size of the forecast intervals (distance between quantiles). Quantiles are gathered by pairs in order to obtain intervals with different nominal coverage rate. This gives an indication on the level of usefulness, where narrow intervals are desired. This measure does not depend on observations.

The objective of scoring rules is to give a global information on model performance in a single measure [34] The skill score is then:

$$S_c\left(\hat{f}_{t+k}, p_{t+k}\right) = \sum_{i=1}^{m}\left(\xi_k^{\alpha_i} - \alpha_i\right)\left(p_{t+k} - \hat{q}_{t+k}^{\alpha_i}\right) \tag{24}$$

where p_{t+k} is the realized wind power, α_i is the quantile proportion, q_{t+k} is the forecasted quantile, m the number of quantiles (i.e. $m=1$ means the evaluation of a single quantile). ξ is an indicator variable which gets 0 or 1 if:

$$\xi_k^{\alpha_i} = \begin{cases} 1\ if\ \ p_{t+k} \leq \hat{q}_{t+k|t}^{\alpha_i} \\ 0\ otherwise \end{cases} \tag{25}$$

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

The skill score can be computed for each look-ahead time step using with the following:

$$S_{c_k} = \frac{1}{N}\sum_{j=1}^{N} S_c\left(\hat{f}_{t+k}, p_{t+k}\right) \tag{26}$$

where N is the number of samples from the test set.

Pinson *et al.* [34] mentioned that using a unique proper skill score allows one to compare the overall skill of competitive approaches, since scoring rules encompass all the aspects of probabilistic forecast evaluation. However, it does not inform on the contributions of calibration or sharpness to the skill score. Hence, these authors suggested that calibration should be assessed in a first analysis (as the primary requirements), and then the information provided by skill score allows to derive conclusions about the remaining metrics.

Although a holistic indicator has undeniable utility, one must bear in mind however that the ultimate usefulness of a model rating system depends on the use and value of the use that will be given to the model. This discussion could follow parallel steps to the one in [13]. Therefore, we suggest that any evaluation of merit cannot be reduced to assessing a single, albeit global, score and distinct criteria should be kept – and, for wind power uncertainty forecasts, calibration seems to be of the utmost importance, over other criteria.

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

The results obtained with the quantile-copula estimator are compared with the splines quantile regression [2], which may be seen as the most widely used for the representation of probabilistic wind power forecasts within the industry. For the circular variables, a periodic cubic spline basis with equidistant knots is used. This is done by the S-PLUS/R functions “pb.bse”, “pb.h” and “bint0” available in [35].

4.3 Evaluation Results: NREL’s EWITS Study

4.3.1 Offline Results

Figure 5 depicts an example of a probabilistic forecast obtained with quantie-copula estimator in the form of a set of interval forecasts.

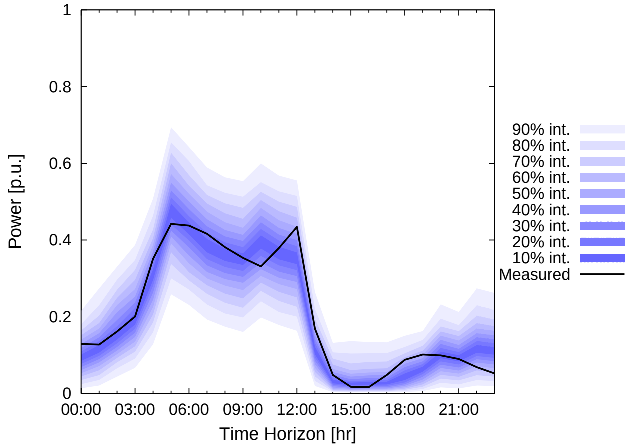

Figure 5: Probabilistic forecast for NREL dataset obtained with the quantile-copula estimator.

The kernel size was 0.001 for both realized and forecasted wind power. The degrees of freedom for the cubic splines are 6. These values were determined by trial-error, and using as starting point the bandwidth values suggested by the function “cde.bandwidths” from the R package “hdrcde” [36].

Figure 6 depicts the calibration diagram averaged for the whole time horizon (24 hours) for the probabilistic forecasts obtained with the splines quantile regression (Splines QR) and Quantile-copula (QC) estimator. Note that what is depicted in the diagram is the difference from the “perfect calibration” (i.e. perfect match between nominal and observed probabilities).

The two models presented in Figure 6 show a deviation from the “perfect calibration” below 5%, which according to Juban *et al.* [8] is equivalent to what is found in the literature. For the quantiles between 70-95% and 5-25% the QC estimators present a lower deviation than the quantile regression method. For the remaining quantiles, the splines QR achieves the lowest deviation. On average, the methods overestimate (nominal proportions greater than observed) the quantiles.

Figure 7 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 intervals presented relatively narrow amplitudes in the two methods, although splines QR presented a slightly lower sharpness. It is important to note that Juban *et al.* [8] found a trade-off between reliability and sharpness, meaning that improving the reliability will generally degrade the sharpness and vice-versa.

Figure 8 depicts the skill score computed for each look-ahead time step. The performance of both approaches is very similar, in some hours there is a slight advantage for QC and in others is the opposite situation.

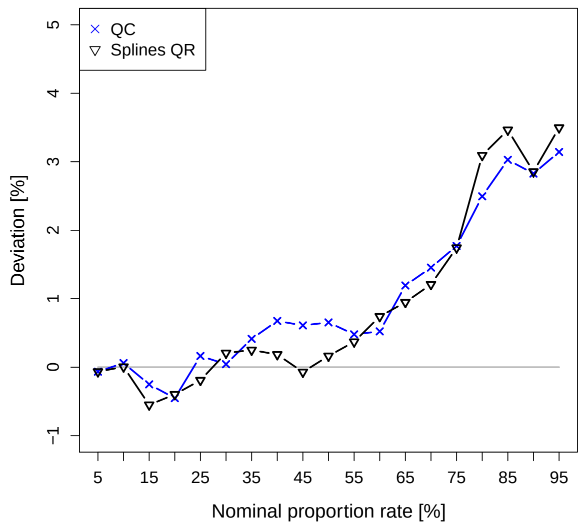

Figure 6: Calibration diagram for the offline test with NREL data.

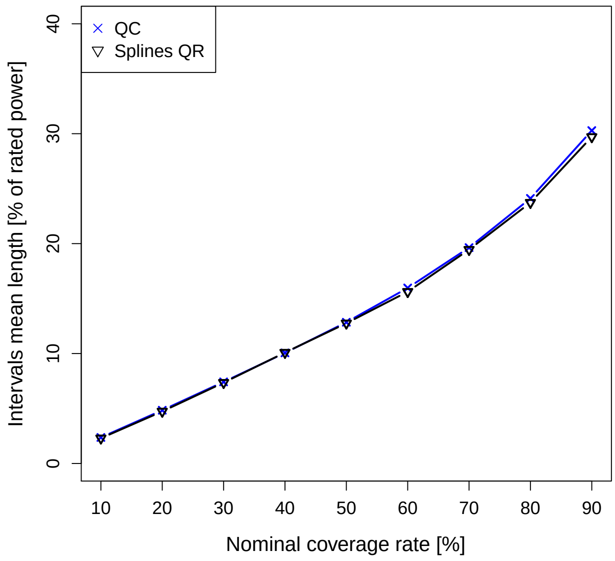

Figure 7: Sharpness diagram for the offline test with NREL data.

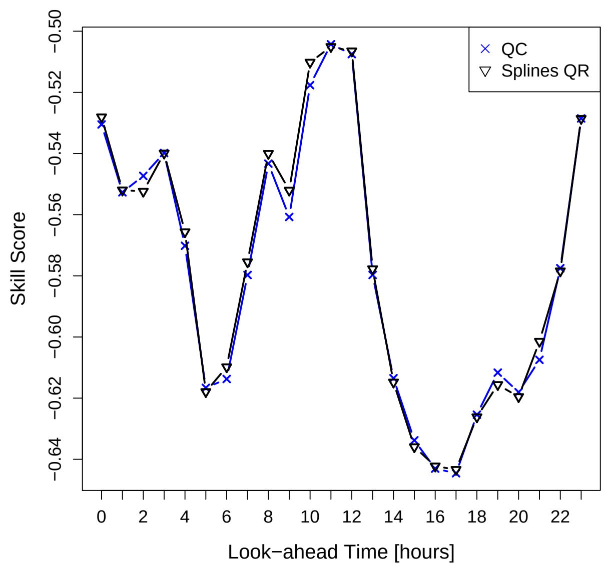

Figure 8: Skill score diagram for the offline test with NREL data.

4.3.2 Time-adaptive Test: Proof of Concept

The aim of this section is to demonstrate that the time-adaptive concept presented in section 3.5 works in conditional KDE and can be applied to the wind power forecast problem. The same EWITS dataset was used for this test. However, in order to introduce a *concept change* (rough, not smooth) 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) between January and October. The same testing period between September and December was used for this simulation.

This situation was created artificially but it simulates a situation that could happen to a system operator. For instance, a system operator is currently receiving forecasts from 13 wind farms, then these forecasts are summed up and estimates the uncertainty associated to the total wind power generation. Then, in October two wind farms are connected to the grid, and in this case the knowledge from past observations has limited validity. 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. Moreover, the system operator would have to wait several months in order to have sufficient data to perform the offline training.

The results will only be analyzed in terms of calibration, since the major impact on the system operator in this situation is in an underestimation or overestimation of the quantiles. Figure 9 depicts the calibration diagram for the offline and time-adaptive QC estimator (the offline splines QR behaves similarly to the offline QC in this case). The preliminary tests showed that the value of λ for the empirical cumulative distribution function should be very low; in this case a value equal to 0.9995 was used.

Due to the increase in the wind power generation because of the connection of two wind farms, it is expected that the offline approach 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 wind generation above its value is only 5%, however, the observed quantiles estimated with the offline approach fix this probability at 13.6%. This means that the probability of having more wind generation in the system is higher than the predicted one. The opposite situation is verified for the quantiles below 50%, e.g. the 10% means that with 90% probability the wind generation is above its value, but the observed proportion for the offline approach reveals that this probability is 84.7%.

The time-adaptive approach can incorporate the recent information and discount the old information (controlled with λ). Therefore, under and overestimations are corrected using this approach.

For instance, for the quantile 95% the observed proportions obtained with the time-adaptive approach is 91.4% with λ=0.999 and 88.5% with λ=0.995. For the 10% quantile the observed proportions are 11.7% (λ=0.999) and 15.7% (λ=0.995). The QC estimator that presents better calibration performance is with λ equal to 0.999. As a first consideration, it is clear that the use of a lower value for λ could be used in order to quickly learn the new data structure. However, as the result for 0.99 shows that a lower value leads to results with larger deviations than the offline approach. The reason is that the QC becomes numerically unstable and it is unable to properly assimilate the recent information. The main conclusion is that a value of λ near 1 should be used and in the case of concept change this value could be reduced, but after a while it should be increased again.

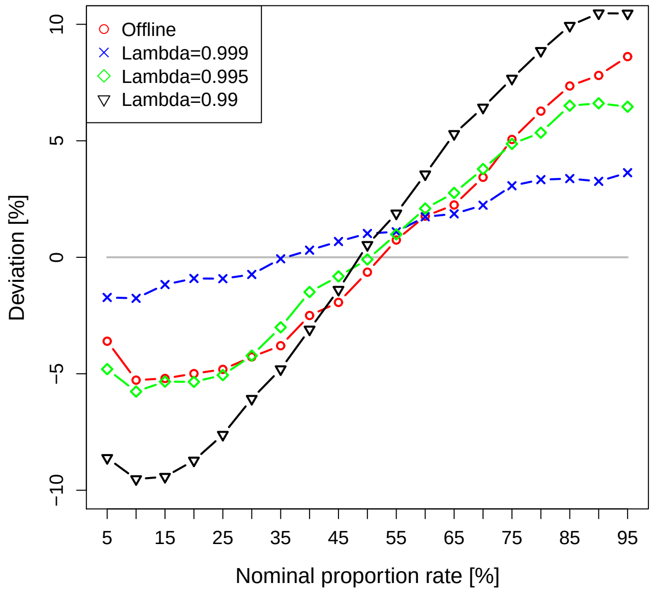

Figure 9: Calibration diagram for the NREL dataset with concept change and QC estimator (offline and time-adaptive).

4.4 Evaluation Results: Midwest Wind Farm

4.4.1 Offline Results

The following kernel functions were used in the Quantile-Copula (QC) estimator:

- wind power generation: Chen's beta kernel from Eq. 14 with a bandwidth equal to 0.008;
- wind speed forecast: Chen's beta kernel from Eq. 14 with a bandwidth equal to 0.008;
- wind direction: von Mises distribution from Eq. 16 with a bandwidth equal to 1.0;
- look-ahead time step: Chen's beta kernel from Eq. 14 with a bandwidth equal to 0.2;

The degrees of freedom for the cubic splines are 8.

Figure 10 depicts an example of a probabilistic forecast obtained with the QC estimator for the Midwest wind farm.

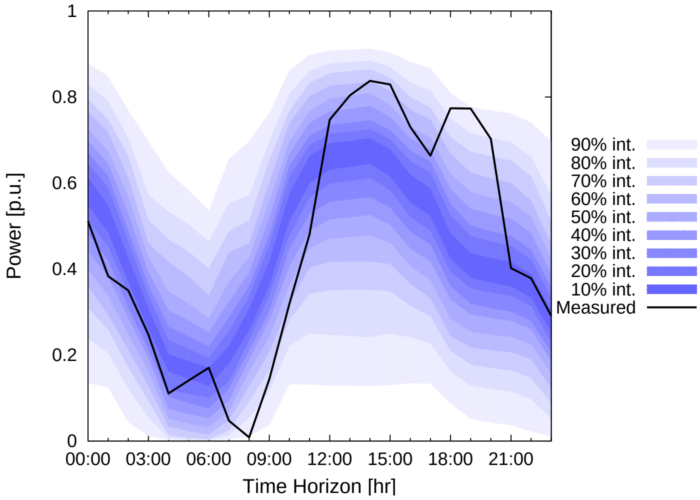

Figure 10: Probabilistic forecast for the Midwest wind farm obtained with the quantile-copula estimator.

Figure 11 depicts the calibration obtained with an offline QC and splines QR. Note that the calibration is presented for quantiles between 1% and 5% in 1% steps, them from 5% to 95% in 5% steps, and finally from 95% to 99% in 1% steps. The QC estimator presents the best calibration performance compared to splines QR, e.g. the maximum deviation of QR is around 8%, while for QC it is less than 6%. Splines QR presents the best calibration for the left tail, while QC presents the best calibration for the right tail.

Figure 12 depict the sharpness obtained for QC and QR. As expected, the method with better calibration presents a worse performance in terms of sharpness. There is slighly better sharpness performance from QR.

Figure 13 presents the skill score computed for each look-ahead time step. The performance of both methods is very similar, with a slight advantage for QC in a majority of the look-ahead time steps.

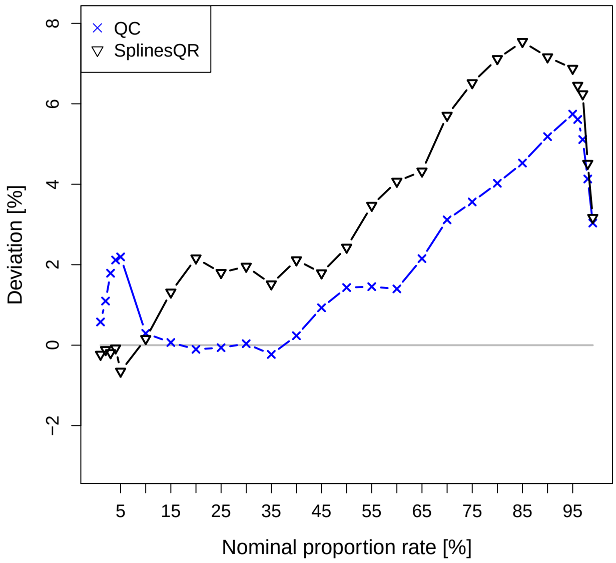

Figure 11: Calibration diagram for the Midwest wind farm offline QC and splines QR.

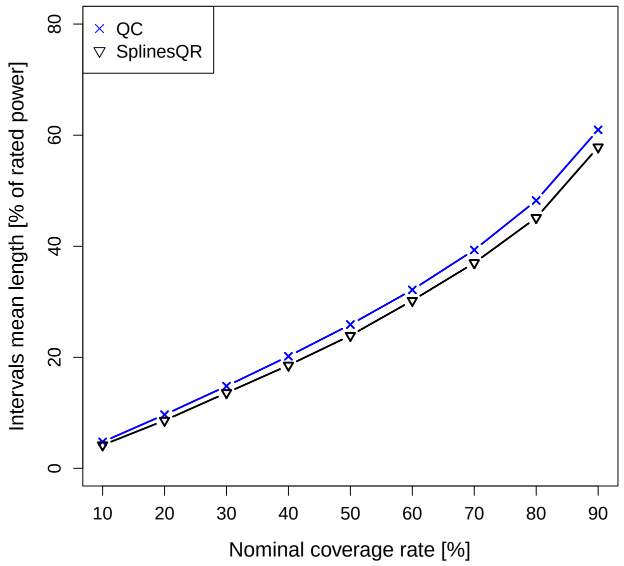

Figure 12: Sharpness diagram for the Midwest wind farm offline QC and splines QR.

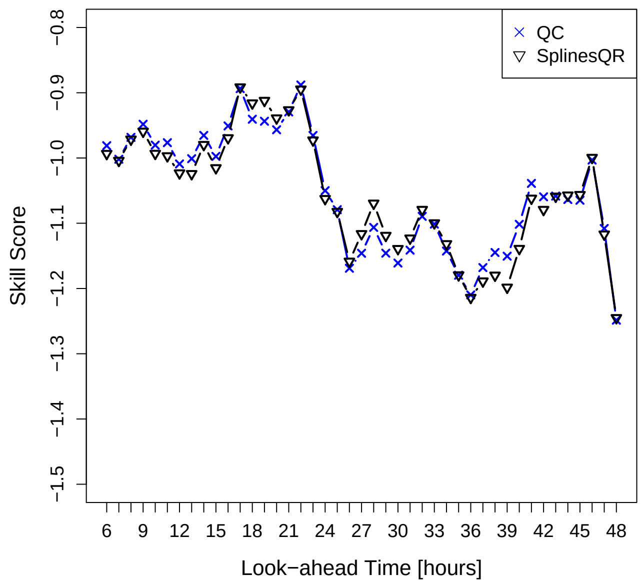

Figure 13: Skill score diagram for the Midwest wind farm for offline QC and splines QR.

4.4.2 Time-Adaptive Results

In this section the time-adaptive version was compared with the offline version 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 of the equivalent sliding window (*n*), according to $\lambda = n/(n+1)$.

So, three values for λ (of Eq. 22-23) 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 used in the offline version was also considered for the time-adaptive versions. Note that the time-adaptive version of the empirical cumulative distribution function (λ_e in Eq. 20) has a different value. Since this dataset does not have significant variations in the data structure (in contrast to the dataset used in section 4.3.2) the adopted value was 0.9999. Note that a smaller value would lead to very poor results.

Figure 14 depicts the calibration results. The time-adaptive version with λ=0.99963477 (n=2738 points) and λ=0.999 (n=1000 points) achieved the best performance.

The version with higher λ does not have a significant impact on sharpness (depicted in Figure 15).

Figure 16 depicts the skill score for the offline and time-adaptive versions. The best performance was obtained with the time-adaptive versions with 2738 and 1000 points, while the version with 200 points presents the worst performance.

One conclusion that can be derived from these results is that the time-adaptive approach changes calibration, or in other words, it changes the bias of the probabilistic forecasts. This change in probabilistic bias is performed in a rather uniform fashion for each quantile, i.e. it is almost a linear shift for high λ values.

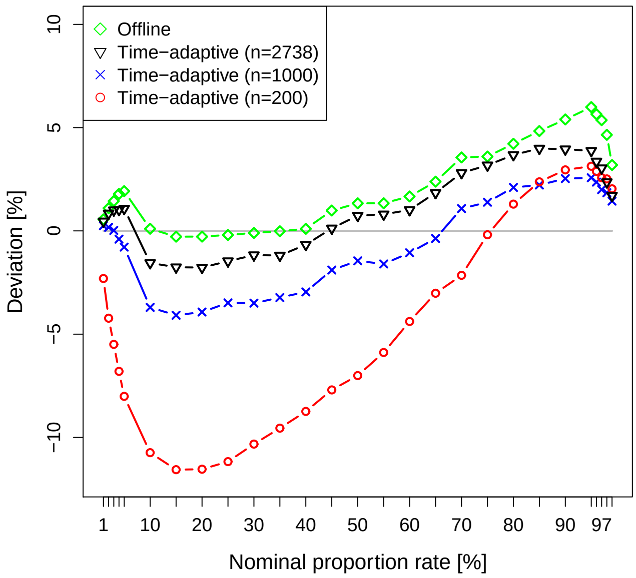

Figure 14: Calibration diagram for the time-adaptive QC with different λ values.

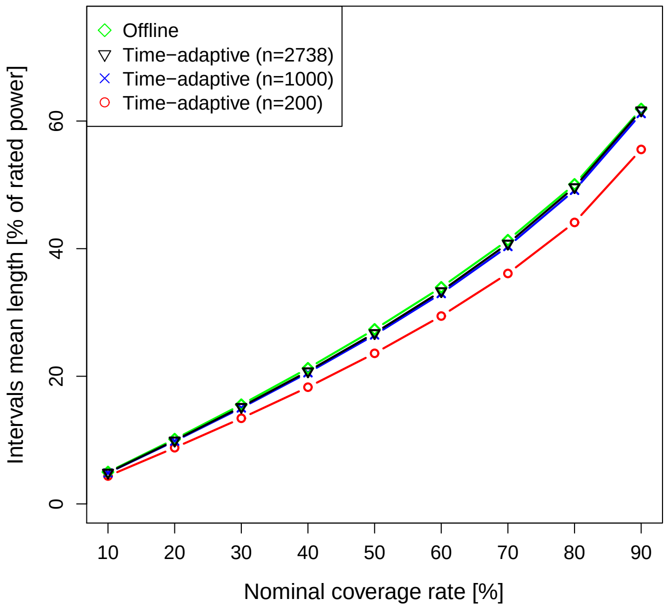

Figure 15: Sharpness diagram for the time-adaptive QC with different λ values.

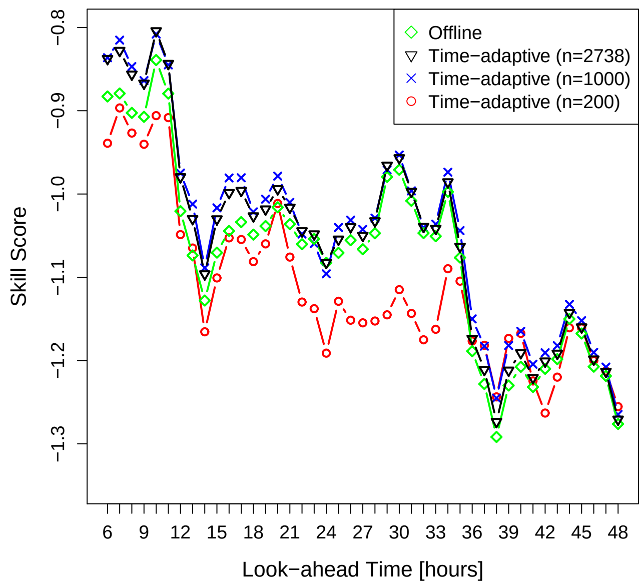

Figure 16: Skill score diagram for the time-adaptive QC with different λ values.

5. Conclusions

This paper presented new contributions to the advancement of wind power probabilistic forecasting beyond the current state-of-the-art. A new time-adaptive quantile-copula method applied to the wind power uncertainty forecast problem was described. The method is based on conditional kernel density estimation and produces as output the wind power *pdf* for each look-ahead time step.

The results obtained for the two case-studies (i.e. NREL's EWITS Study and a Midwest wind farm) have shown that the quantile-copula method led to a better calibration performance, while the splines quantile regression (a method from the state-of-the-art) presented slightly better sharpness performance. The skill score performance is rather similar for both approaches, with a slightly advantage for the new model (quantile-copula).

In this paper, only the quality (i.e. correspondence between forecasts and observations) of the probabilistic forecast was evaluated. No considerations were made about the forecast value, i.e. the incremental benefits (economic/or other) from the new technique when employed by end-users as an input into their decision-making processes. Nevertheless, we defend that model comparison in forecast evaluation, and in probabilistic forecasts in particular, should also be done by an *ex-post* evaluation of the benefits obtained within the decision making problems that use the forecasts as input. We wish to stress that the incremental forecast value will differ from problem to problem, and from end-user to end-user. In the authors opinion, the "ideal" evaluation metric should be a kind of skill score that may account for end-user preferences (and idea of what a "good" forecast is), oriented to a particular decision-making context, and not an "abstract" formula unrelated to the problem, albeit mathematically sound.

Nevertheless, the three metrics used in this paper give indications about the forecast value, even if the interpretation may differ from problem to problem. For instance, the calibration metric is particularly important for the wind power bidding problem. As proven in [37] for convex loss functions, the optimal solution that minimizes the expected loss is a forecast quantile related with an asymmetry parameter that reflects the possibly distinct costs of underforecast and overforecast. This can be extrapolated for electricity markets with asymmetric penalizations for positive and negative deviations between bids and measured values. In this case, the optimal quantile (in the expected value paradigm) differs from the median. Therefore, a calibration deviation may lead to situations of under or overestimation of the optimal quantile, and consequently a deviation from what is optimal in the economic sense.

The same idea can be applied to the problem of setting the operating reserve for a power system. As shown in [38], deviations in calibration may lead to under and overestimation of the loss of load probability, which can increase the operators' stress. A lower sharpness performance may also lead to high amounts of recommend reserve, which means an increase in the cost of providing reserves.

Therefore, a wind power probabilistic forecast method should provide an adequate compromise between calibration and sharpness, keeping in mind that calibration is the major requirement. The quantile-copula method described in this paper shows a good balance between both metrics, which indicates a likely improvement in several decision making problems.

For future work, we foresee three important aspects: i) a time-adaptive forgetting factor; ii) a robust and time-adaptive approach (or rules) for setting the kernels' bandwidth; iii) the possibility of producing multi-period forecast intervals including information about the temporal correlation of forecast errors. These three lines of progress will deal with the ultimate goal of having robust and reliable time-adaptive processes and representing other important characteristics of the wind power series that are not captured by the conditional estimates discussed in this paper. Nevertheless, the Quantile-Copula model presented in this paper offers not only advantages over existing approaches, but also to serve as a stepping stone for further progress.

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, and Emil Constantinescu at Argonne National Laboratory for providing the Numerical Weather Prediction data.

The work of R. J. Bessa was supported by Fundação para a Ciência e Tecnologia (FCT) Ph.D. Scholarship SFRH/BD/33738/2009.

References

[1] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, G. Conzelmann, Wind power forecasting: state-of-the-art 2009, Report ANL/DIS-10-1, Argonne National Laboratory, 2009. Available online at: http://www.dis.anl.gov/projects/windpowerforecasting.html (accessed on Apr. 2011)

[2] H.A. Nielsen, H. Madsen, T. S. Nielsen, Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts, Wind Energy. 9 (2006) 95-108.

[3] P. Pinson, G. Kariniotakis, Conditional prediction intervals of wind power generation, IEEE Transactions on Power Systems. 25 (2010) 1845-1856.

[4] P. Pinson, H. Madsen, Ensemble-based probabilistic forecasting at Horns Rev, Wind Energy. 12 (2009) 137-155.

[5] N.J. Cutler, H.R. Outhred, I.F. MacGill, M.J. Kay, J.D. Kepert, Characterizing future large, rapid changes in aggregated wind power using Numerical Weather Prediction spatial fields, Wind Energy. 12 (2009) 542-555.

[6] P. Pinson, H.Aa. Nielsen, H. Madsen, G. Kariniotakis, Skill forecasting from ensemble predictions of wind power, Applied Energy. 86 (2009) 1326-1334.

[7] P. Pinson, G. Papaefthymiou, B. Klockl, H.Aa. Nielsen, H. Madsen, From probabilistic forecasts to statistical scenarios of short-term wind power production, Wind Energy. 12 (2009) 51-62.

[8] J. Juban, L. Fugon, G. Kariniotakis, Uncertainty estimation of wind power forecasts, in Proceedings of the European Wind Energy Conference EWEC'08, Brussels, Belgium, March 31–April 03, 2008.

[9] John B. Bremnes, Probabilistic wind power forecasts using local quantile regression, Wind Energy. 7 (2004) 47-54.

[10] J. Juban, N. Siebert, G. Kariniotakis, Probabilistic short-term wind power forecasting for the optimal management of wind generation, in Proceedings of the IEEE Power Tech Conference, Lausanne, Switzerland, July 2007.

[11] J. K. Møller, H. A. Nielsen, H. Madsen, Time-adaptive quantile regression, Computational Statistics & Data Analysis. 52 (2008) 1292-1303.

[12] O.P. Faugeras, A quantile-copula approach to conditional density estimation, Journal of Multivariate Analysis. 100 (2009) 2083-2099.

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

[14] A. Botterud, J. Wang, R. Bessa, H. Keko, V. Miranda, Risk management and optimal bidding for a wind power producer, in Proceedings of the IEEE PES General Meeting, Minneapolis, USA, 2010.

[15] F. Bourry, J. Juban, L.M. Costa, G. Kariniotakis, Advanced strategies for wind power trading in short-term electricity markets, in Proceeding of the European Wind Energy Conference & Exhibition EWEC 08, Brussels, Belgium, 31 March - 3 April 2008.

[16] M.A. Matos, R. Bessa, Setting the operating reserve using probabilistic wind power forecasts, IEEE Transactions on Power Systems. 26 (2011).

[17] R. Bessa, M.A. Matos, Comparison of probabilistic and deterministic approaches for setting operating reserve in systems with high penetration of wind power, in Proceedings of the 7th Mediterranean Conference and Exhibition on Power Generation, Transmission, Distribution and Energy Conversion - MedPower2010, Agia Napa, Cyprus, 7-10 Nov. 2010.
[18] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, The Annals of Mathematical Statistics. 27 (1956), 832-837.
[19] E. Parzen, On Estimation of a probability density function and mode, The Annals of Mathematical Statistics. 33 (1962) 1065-1076.
[20] M.P. Wand, M.C. Jones, Multivariate plug-in bandwidth selection, Computational Statistics. 9 (1994) 97-116.
[21] M. Rosenblatt, Conditional probability density and regression estimators, in: Multivariate Analysis II, Academic Press, New York, 1969, pp. 25–31.
[22] T. Bouezmarni, J. V. K. Rombouts, Semiparametric multivariate density estimation for positive data using copulas, Computational Statistics & Data Analysis. 53 (2009) 2040-2054.
[23] M. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de Statistique de L' Université de Paris. 8 (1959) 229-231.
[24] R. Nelson, An Introduction to Copulas. Lecture Notes in Statistics, Springer, New York, 1999.
[25] G. Frahm, M. Junker, Szimayer, Elliptical copulas: applicability and limitations, Statistics & Probability Letters. 63 (2003) 275-286.
[26] S.X. Chen, Beta kernel estimators for density functions, Computational Statistics & Data Analysis. 31 (1999) 131-145.
[27] C. Gourieroux, 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, 2006.
[28] K. V. Mardia, P. E. Jupp, Directional Statistics. Wiley's Series in Probability and Statistics, New York, 1999.
[29] E.J. Wegman, H.I. Davies, Remarks on recursive estimators of a probability density, The Annals of Statistics. 7 (1979) 316-327.
[30] E.J. Wegman, D.J. Marchette, On some techniques for streaming data: a case study of internet packet headers, Journal of Computational and Graphical Statistics. 12 (2003) 893-914.
[31] Eastern Wind Integration and Transmission Study (EWITS), National Renewable Energy Laboratory (NREL). Information at: http://www.nrel.gov/wind/systemsintegration/ewits.html (accessed on Apr. 2011)
[32] M. Brower, Development of eastern regional wind resource and wind plant output datasets, NRE, Subcontract Report NREL/SR-550-46764, 2009.
[33] W. Skamarock, J.B. Klemp, J. Dudhia, D. Gill, D.M. Barker, W. Wang, J.G. Powers, A Description of the advanced research WRF version 2, NCAR/TN–468+STR Technical note, 2005.
[34] P. Pinson, H.A. Nielsen, J.K. Moller, H. Madsen, G. Kariniotakis, Nonparametric probabilistic forecasts of wind power: required properties and evaluation, Wind Energy. 10 (2007) 497-516.
[35] www.imm.dtu.dk/~han/pub (accessed on Apr. 2011)
[36] R. J. Hyndman, J. Einbeck, M. Wand, Package hdrcde, R Manual, 2010. Available online at: http://www.robjhyndman.com/software/hdrcde (accessed on Apr. 2011)
[37] T. Gneiting, Quantiles as optimal point forecasts, International Journal of Forecasting. 27 (2011) 197-207.
[38] R. Bessa, M.A. Matos, Comparison of probabilistic and deterministic approaches for setting operating reserve in systems with high penetration of wind power, in Proceedings of MedPower2010, Agia Napa, Cyprus, 2010.