Danmarks Tekniske Universitet

Spatio-temporal correction targeting Nysted Offshore. Probabilistic forecasts

---

Julija Tastu [1], Pierre Pinson, Pierre-Julien Trombe, Henrik Madsen

July 7^{th}, 2011

[1] jvl@imm.dtu.dk

Contents

Summary

This report concerns probabilistic forecasts for Nysted Offshore. Different approaches for issuing predictive densities are studied, discussed in details and compared. The results show that the spatial correction of the first order moments of the predictive densities improves the quality of the corresponding forecasts. The spatial correction of the higher order moments is shown to be unnecessary as does not bring any additional amelioration. The best performing of the studied models is based on the adaptive quantile regression using the spatially corrected point predictions as input. This model is shown to outperform the benchmark approach in termps of the CRPS score (accuracy measure) by 1.5%-8.29% depending on the considered prediction horizon.

Introduction

Previous studies within *DONG Spatio-temporal* concern improvements in the forecasts of the expected power at Nysted Offshore using spatial information. Such predictions are never 100% accurate as they are always associated with some uncertainty due to incomplete knowledge one has about the future events. The uncertainty of the point forecasts is highly variable and depends on the considered look-ahead time, the quality of the related meteorological forecasts and many more factors. As the verity of the point-forecasts is time-varying, they are not to be trusted in the same way at every step. Providing information on how much those forecasts can actually be trusted at a certain moment, plays a very important role in many decision making tasks. Such information is directly related to knowing how much risk is being taken by relying on the prediction. Complementing point predictions with the estimates of the associated uncertainty is a task of probabilistic forecasting.

The main objective of this work is to extend the point forecasts for Nysted Offshore to probabilistic forecasts. The aim is to "dress" the point predictions with the uncertainty estimates in order to obtain full predictive densities of the generated wind power instead of giving only the expected value (which is the case for the point forecasts). There are two ways of getting such predictive densities. One way is to assume that the wind power generation follows a known distribution. Then in order to build a predictive density one needs to estimate the parameters of the assumed distribution using the past observations. This approach is called a parametric, or distribution-based method. Another way is to assume that the exact distribution is unknown. In that case one needs to consider a wind power expectation given by the point prediction and assume that the accuracy of it will be directly related to the previously observed performance of the corresponding point forecast model. The quantiles of the past point forecast errors are then estimated. A 0.95 quantile, for example, gives a value which will not be exceeded by the observation with a probability (nominal proportion) of 0.95. After estimating the quantiles, the point forecasts are being "dressed" with a set of prediction intervals of different nominal coverage rates in order to obtain a full density. This approach is called a non-parametric, or distribution-free approach.

When building probabilistic forecasts for a wind power variable which account for the spatial information, some knowledge could be taken from the earlier works concerning spatial modelling of a wind speed data with a consideration of probability forecasts. Gneiting et al. [8] propose regime switching models which account for two dominant wind directions while predicting wind speed up to two hours ahead. The corresponding probabilistic forecasts for the wind speed are based on a truncated Normal distribution. The truncation is carried out in order to ensure that the final forecasts are within the valid range, i.e. no negative wind speed values occur. Hering and Genton [11] extend the work of Gneiting by considering a circular wind direction and applying both truncated Normal distribution and a skew-T distribution for describing the model related uncertainties. In their work the authors also link the predictive wind speed densities to the corresponding wind power variable using a power curve model. Lau in [13] looked at the spatio-temporal models for a wind power generation while considering multi-step ahead forecasts described by the censored Normal distribution with spatial covariance structures.

In parallel, without considering spatial effects, but rather concentrating on a single site, Pinson [20] compares several types of parametric predictive densities for describing very short-term wind power generation at Horns Rev wind farm. A censored Normal, a Beta and a generalized logit-Normal distributions are considered in the work. The author shows that the generalized logit-Normal distribution is a better candidate for describing the wind power generation than the censored Normal or Beta distributions.

In addition some research was carried out for proposing non-parametric approaches for probabilistic forecasting of the generated wind power. Møller et al. [14] introduced an adaptive estimation scheme for a quantile regression and used the method for issuing probabilistic wind power forecasts. In parallel, in [17] Pinson and Kariniotakis proposed another non-parametric approach called adaptive re-sampling. Similarly to the adaptive quantile regression this approach does not assume any particular distribution for the wind power generation. Instead it builds prediction intervals of different nominal coverage rates based on the past deviations from the expected power. Both mentioned non-parametric approaches have been compares on the same test case - hourly wind power predictions at Klimt wind farm. The results are documented in [19] where it is shown that the two methods perform similarly.

To our knowledge, no studies comparing parametric and non-parametric approaches to probabilistic wind power forecasting exist. As far as the spatial models for the wind power data are concerned, nothing but the censored Normal distribution has been considered (see [13]). The goal of this study is therefore to apply both parametric and non-parametric methods for issuing probabilistic forecasts for Nysted Offshore. The comparison of the forecasts performance will be carried out considering different prediction horizons. The most important questions which are aimed to be answered are:

- which modelling approach gives better results: parametric or non-parametric one? Is it consistent for all the considered horizons?
- does accounting for the spatial information improve the performance of the probabilistic forecasts?

The outline of the report is as follows:

A short description of the data used in this work is given in Section 1. It is followed by the methodological aspects on the parametric probabilistic forecasting (Section 2). The distributions considered in this work are presented in Sections 2.1 and 2.2. The parameter estimation for the considered densities is discussed in Section 2.3 and 2.4. A non parametric approach for building the predictive densities is presented in Section 3. Section 4 discusses the methods for probabilistic forecast assessment. The results are presented in Section 5. The paper ends with the concluding remarks in Section 6.

1 Data and previous work

All analysis within *Dong Spatio-temporal* are primary based on the WPPT power (point) forecasts with a resolution of 15 min and a horizon of 48 hours available at a number of wind farms listed in Table 1. The location of the farms are shown in Figure 1. The WPPT forecasts are based on the power measurements with a resolution of 5 min and meteorological forecasts with a time resolution of 1 hour and a horizon of 48 hours. The meteorological forecasts include wind direction and wind speed at an altitude of 10 meters. Meteorological forecasts are provided by the HIRLAM model operated by the Danish Meteorological Institute (DMI). The WPPT forecasts used in the study are historical runs where the delay of the meteorological forecasts has been set to 4 hours. This implies that the meteorological forecasts originally issued, for instance, at midnight become available as an input to the wind power prediction models only at 4 a.m. This means that the wind power forecasts issued at 4 a.m. and targeting 5 a.m. (one hour ahead predictions) use the meteorological forecast issued at midnight as input. That implies that the one-hour ahead wind power predictions issued at 4 a.m. are actually using five-hours ahead meteorological forecasts as input. The data covers the period from the 1st of May, 2008 to the 1st of February, 2010. All the models considered in the project have been run for the period from the 1st of May, 2008 to the 31st of December, 2009. The actual evaluation of the models' performance is based on the period from the 1st of January, 2009 to the 31st of December, 2009.

Spatio-temporal correction of the WPPT power forecasts at Nysted Offshore was carried out in [3, 2]. Different configurations of the correction models have been tried in [3, 2]. The best performance was observed when the spatial correction models (Conditional Parametric models) for Nysted Offshore have been run using a reduced set of explanatory variables proposed in [2] (the corresponding model is denoted as a CP-model) and additionally to that incorporated data transformations described in [3] (the corresponding model is denoted as a Logit-CP model). Detailed analysis of those two models is given in [3]. Complementary to the initial WPPT forecasts in this work the corrected predictions made by the CP and the Logit-CP models are considered.

2 Parametric approaches to probabilistic forecasting

Parametric approach is based on the assumption that a distribution function of the wind power generation is known. In order to choose a specific type of density, the crucial features of the wind power variable are to be taken into the account. Firstly, when normalized by the nominal capacity, it is double bounded between a minimum production of 0 and a maximum one of 1. It is in parallel a non-linear function of the forecasted wind speed in the shape of a sigmoid, which means that higher variations in the wind power are observed when the predicted wind speed is in its medium range. For the wind speeds close to the cut-in or cut-off values the fluctuations in the observed power are lower. This means that a distribution of the observed power changes with the level of the predicted wind speed and as the latter is directly related to the expected power, one can say that the density of the observed power depends on the level of the expected power. This is further demonstrated in Figure 2. Another point clearly seen in Figure 2 is a non-negligible concentration of probability mass at the bounds. When the expected power is far from the natural generation bounds (see Figure 2(d)), the conditional histogram resembles that of a Gaussian distribution (a characteristic bell-shape around the expected value can be seen). The closer to the bounds, the less dispersed distributions become and the higher the probability concentration at the closest bound can be noted.

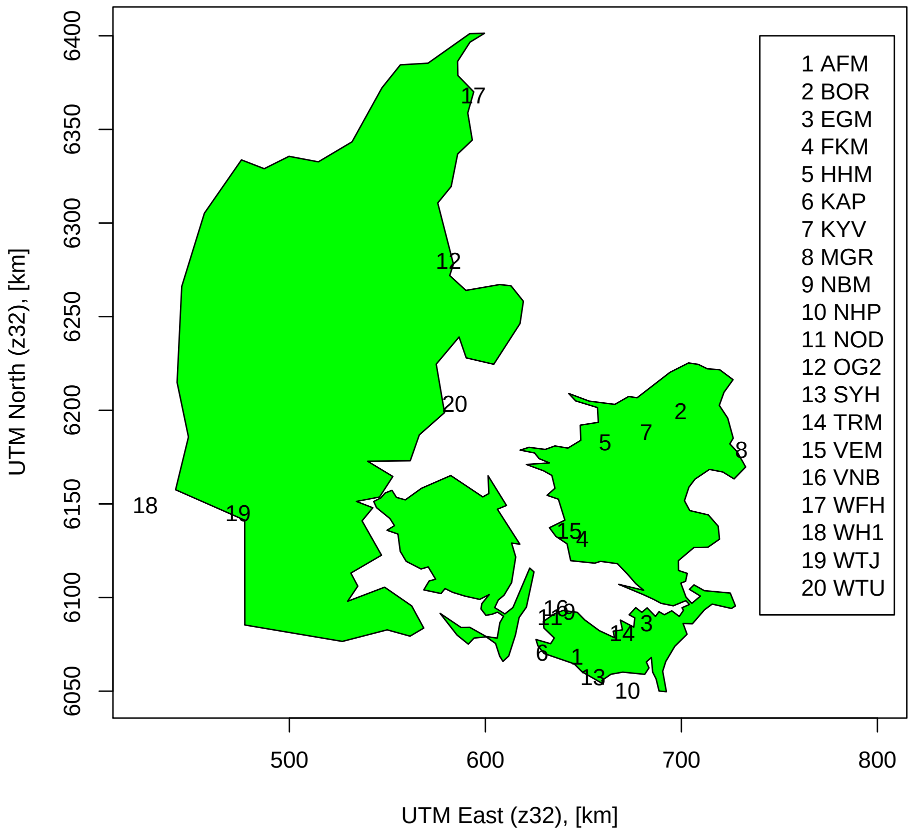
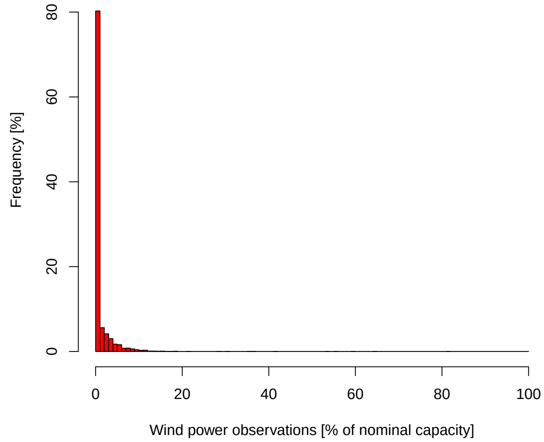


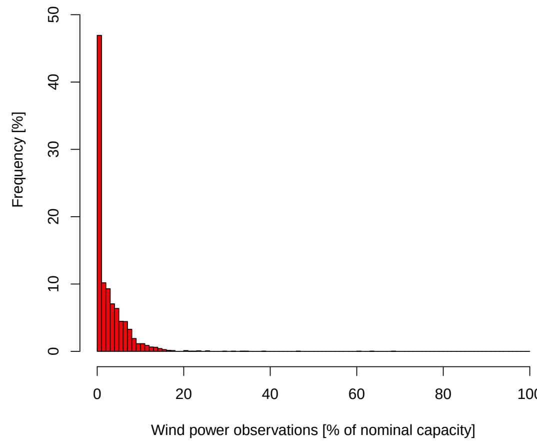
Figure 1: Map of wind farms included in the analysis. Nysted (target wind farm) is marked by number 10

Name	Abbrev.	Cap. (MW)	Number of turbines	UTM E	UTM N	Hub h.
Askoe Faeland	AFM	2.40	4	646777.00	6068525	40.00
Borup (sum)	BOR	3.96	6	699583.00	6199488	44.00
Egelev	EGM	1.80	4	682256.00	6086204	35.00
Frankerup/Faardrup	FKM	2.70	4	649515.50	6131569	40.50
Hagesholm	HHM	6.00	3	661093.00	6182659	68.00
Kappel	KAP	10.20	24	628911.00	6070409	32.00
Kyndby	KYV	3.06	19	682049.00	6188266	30.00
Middelgrunden	MGR	20.00	10	730597.30	6178847	64.00
Nyboelle	NBM	1.00	2	642797.00	6092329	35.00
Nysted Offshore	NHP	165.60	72	672504.00	6050365	68.00
Noejsomheds Odde	NOD	21.00	21	633010.00	6089671	50.00
Overgaard 2	OG2	11.50	5	581199.00	6279823	80.00
Syltholm	SYH	18.75	25	654876.00	6057536	44.00
Taars	TRM	1.00	2	669788.00	6081120	40.00
Vemmelev (sum)	VEM	4.65	8	642791.67	6135634	38.33
Vindeby ("offshore")	VNB	4.95	11	635956.00	6094187	38.00
Frederikshavn	WFH	10.60	4	593839.00	6368078	80.00
Horns Rev 1	WH1	160.00	80	426487.00	6149224	80.00
Tjaereborg Enge	WTJ	2.00	1	473833.00	6144928	57.00
Tunoe Knob	WTU	5.00	10	584391.00	6203478	45.00

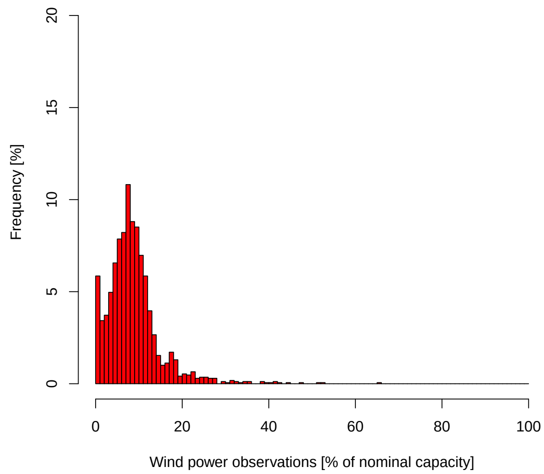
Table 1: Wind farms included in the analysis. UTM coordinates corresponds to zone 32.



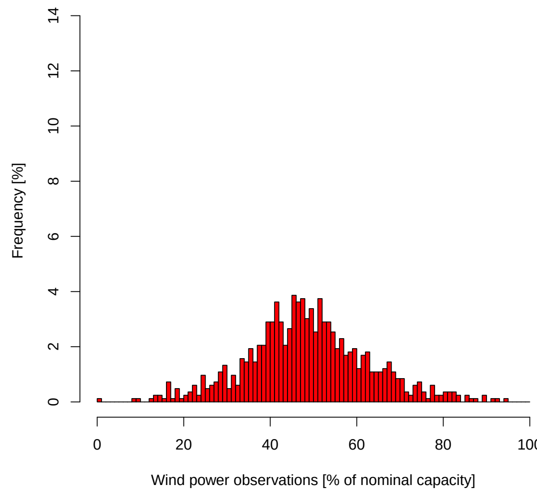
(a) Expected power in 0 - 2 % of nominal capacity



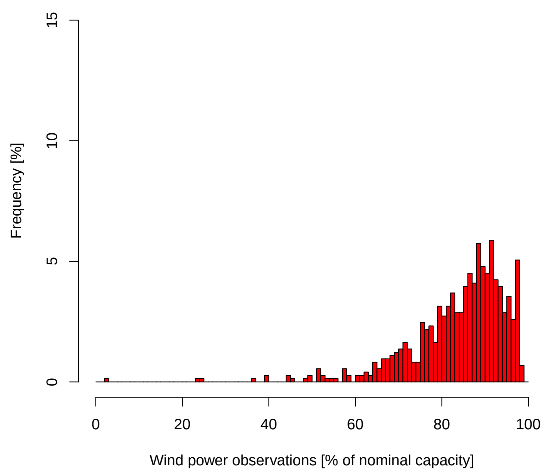
(b) Expected power in 2 - 4 % of nominal capacity



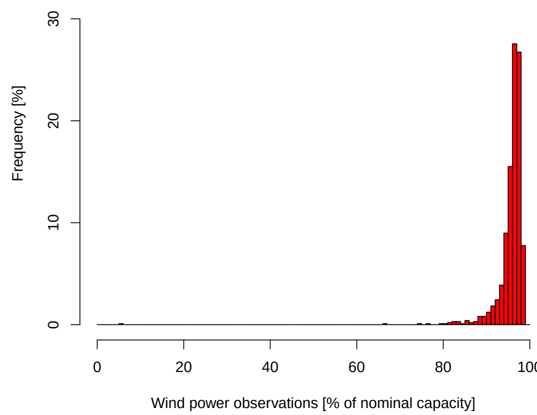
(c) Expected power in 8 - 10 % of nominal capacity



(d) Expected power in 48 - 50 % of nominal capacity



(e) Expected power in 82 - 84 % of nominal capacity



(f) Expected power in 96 - 98 % of nominal capacity

Figure 2: Distribution of the observed power conditional to different levels of the expected power given by the WPPT forecasts. Note that the range of frequencies on the y-axis varies from one plot to the next.

2.1 Censored Normal distribution

For modelling those effects a Censored Normal distribution with varying parameters has been considered in [20, 13]. The censoring is needed due to the fact that a wind power variable is bounded. A Censored Normal wind power variable follows an ordinary Gaussian law within the open unit interval. Since the values outside [0,1] cannot be taken, the corresponding tails of the ordinary Gaussian distribution are cut and deposed to the bounds of the defined interval (to 0 and 1, respectively).

Formally the Censored Normal predictive density for the wind power generation X_{t+k} at time $t+k$ can be written as:

$$X_{t+k} \sim w^0_{t+k}\delta_0(x) + f^{(0,1)}(x; \mu_{t+k}, \sigma^2_{t+k}) + w^1_{t+k}\delta_1(x), x \in [0,1] \tag{1}$$

where δ_0 and δ_1 are Dirac delta functions at 0 and 1, respectively, representing the location of the potential concentration of probability mass:

$$\delta_j(x) = \begin{cases} \infty & \text{if } x = j \\ 0 & \text{otherwise} \end{cases}$$

w^0_{t+k} and w^1_{t+k} are the weights representing the levels of probability mass concentration at the corresponding bounds of the unit interval.

$$w^0_{t+k} = F(0; \mu_{t+k}, \sigma^2_{t+k}) \tag{2}$$

$$w^1_{t+k} = 1 - F(1; \mu_{t+k}, \sigma^2_{t+k}) \tag{3}$$

with $F(x; \mu, \sigma^2)$ being a cumulative Gaussian distribution function with parameters μ and σ^2. $f^{(0,1)}(x; \mu, \sigma^2)$ follows a Gaussian density function (with a location parameter μ and a scale parameter σ^2) within the open unit interval (0,1) and equals 0 outside this interval:

$$f^{(0,1)}(x; \mu, \sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

2.2 Generalized Logit-Normal distribution

The aforementioned characteristics of the wind power variable (bounds, non-constant variance) indicate that some data transformations could be employed in order to stabilize the variance, reduce the influence of the bounds and as a result could make the assumption of a Gaussian distribution more appropriate. For that purpose a generalized logit (GL) transformation (see equation (4)) has been proposed in [20]. In [3] this transformation has been incorporated into the models for spatially correcting the expected wind power generation at Nysted Offshore and resulted in the improved model performance.

For an original time series x_t the generalized logit transform is given by

$$y_t = \gamma(x_t; v) = \ln\left(\frac{x_t^v}{1 - x_t^v}\right), v > 0, |x_t| < 1 \tag{4}$$

while the inverse transformation is defined as

$$x_t = \gamma^{-1}(y_t; v) = \left(1 + \frac{1}{\exp(y_t)}\right)^{-1/v}, v > 0 \tag{5}$$

A shape parameter v has been set to 0.01 (see [3] for details).

Based on (1), but now considering Y_{t+k}, the GL-transform of the corresponding wind power generation X_{t+k}, the form of its predictive density is given by

$$Y_{t+k} \sim w^0_{t+k}\delta_{-\infty}(y) + f^{(0,1)}(y; \tilde{\mu}_{t+k}, \tilde{\sigma}^2_{t+k}) + w^1_{t+k}\delta_{\infty}(y), \ y \in \mathbb{R} \tag{6}$$

In other words instead of assuming that the wind power generation X_{t+k} follows a censored Normal distribution, this distribution is assumed for the GL-transform of the original variable. Otherwise stated - the original wind power generation is assumed to follow a generalized logit Normal distribution.

2.3 Estimation of the location parameter

A parallel between parameter estimation in the censored Normal and the generalized logit-Normal densities. Both the censored Normal and the generalized logit-Normal densities are characterized by the location and the scale parameters. The methodology for estimating the parameters is analogical for both densities. The only difference comes from the input used in the estimation routines. Suppose μ (the location parameter of the censored Normal density) is described as a function f of the available wind power measurements p and the point predictions $\hat{p}$, i.e. $\mu = f(p, \hat{p})$. Then the location parameter of the generalized logit-Normal density $\widetilde{\mu}$ can be found from $f(\gamma(p; v), \gamma(\hat{p; v}))$. The same parity holds for the scale parameters σ^2 and $\widetilde{\sigma^2}$. Thus in the following we focus in details on the estimation of μ and σ^2, keeping in mind that the equivalent techniques with a corresponding input adjustment can be used for finding $\widetilde{\mu}$ and $\widetilde{\sigma^2}$.

From Normal to censored-Normal distribution: approximation used in the parameter estimation routines As has been mentioned before, the censored Normal predictive density can be fully characterised by the location (μ) and the scale (σ^2) parameters. In case of an ordinary (non-Censored) Gaussian distribution the location parameter is described by the mean (expectation) of the distribution and the scale parameter - by its variance. It is not exactly the case for the Censored Normal distribution. However, as previously demonstrated in Figure 2 censoring effect is only significant close to the bounds. But even close to them it is assumed that the location and the scale parameters can be well approximated by the expectation and the variance, respectively. Such approximation introduces a certain bias (lack of accuracy) as the observed power is close to the generation bounds, but nevertheless this estimation method is being used in the state-of-the art research works (see [20] as an example) where it is argued to show a satisfactory performance.

Estimating the location parameter Thus the location parameter of the censored Normal distribution is approximated by the mean (expected value) of the corresponding density. When modelling wind power, the expectation is given by the point forecast. Therefore an estimate of the location parameter which describes the density of the power generation at time $t + k$ (when predicting k-steps ahead) is given by:

$$\widehat{\mu}_{t+k|t} = \widehat{p}_{t+k|t} \tag{7}$$

where $\widehat{\mu}_{t+k|t}$ denotes the estimate (forecasted value) of the location parameter $\mu_{t+k|t}$ and $\hat{p}_{t+k|t}$ stands for the point prediction of the wind power generation issued at time t for time $t + k$.

Recall, the location parameter of the generalized logit-Normal density could be then found from:

$$\widehat{\widetilde{\mu}}_{t+k|t} = \gamma(\widehat{p}_{t+k|t}; v) \tag{8}$$

where $\widehat{\widetilde{\mu}}_{t+k|t}$ denotes an estimate of $\widetilde{\mu}_{t+k|t}$ and $\hat{p}_{t+k|t}$ stands for the expected wind power generation at time $t + k$ forecasted at time t.

In this work several different point predictions ($\widehat{p}_{t+k|t}$) are considered: initial WPPT forecasts and two spatially corrected forecasts which showed the best improvements over the WPPT in the previous studies ([3]). More precisely, the following power expectations are considered:

1. *WPPT forecasts*
2. *CP forecasts.* Conditional parametric models are used for the spatial correction of the WPPT forecasts. The correction model is run using a reduced set of explanatory variables proposed by ENFOR in [2]. The model itself and the resulting predictions are analysed in details in [3].
3. *Logit-CP forecasts.* WPPT forecast correction is performed using the conditional parametric model which is run on the reduced set of explanatory variables and incorporates the GL data transformation (4). A detailed description of the model is given in [1].

2.4 Estimation of the scale parameter

The scale parameter in case of the Censored Normal density is approximated by the variance of the wind power distribution. Variance is estimated by modelling squared residuals (ϵ^2) made by the point forecasting model:

$$\epsilon_{t+k}^2 = \left(p_{t+k} - \hat{p}_{t+k|t}\right)^2 \tag{9}$$

where p_{t+k} denotes wind power observation at time $t+k$ and $\hat{p}_{t+k|t}$ is the corresponding wind power point forecast issued for time $t+k$ at time t (k-step ahead point prediction).

An estimate ($\hat{\sigma}^2$) of the scale parameter σ^2 is given by the expectation (prediction) of the future squared deviation, i.e.:

$$\hat{\sigma}_{t+k|t}^2 = \hat{\epsilon}_{t+k|t}^2 \tag{10}$$

where $\hat{\epsilon}_{t+k|t}^2$ denotes forecasted value of the squared deviation issued at time t for time $t+k$.
As discussed before, the scale parameter of the generalized logit-Normal density is then found from $\widehat{\tilde{\sigma}}^2$ can be obtained by modelling the squared deviations ($\tilde{\epsilon}_t^2$) of the transformed wind power variable i.e. :

$$\widehat{\tilde{\sigma}^2}_{t+k|t} = \widehat{\tilde{\epsilon}^2}_{t+k|t} \tag{11}$$

where

$$\tilde{\epsilon}_{t+k}^2 = \left(\gamma(p_{t+k}; v) - \gamma(\hat{p}_{t+k|t}; v)\right)^2 \tag{12}$$

The same techniques are used for modelling $\tilde{\epsilon}_{t+k}^2$ as for ϵ_{t+k}^2:

1. Exponential smoothing
2. GARCH / GARCHX model
3. Conditional Parametric ARCH model without/ with eXogenous input (CP-ARCH/ CP-ARCHX)
4. Markov switching models

2.4.1 Exponential smoothing

One of the easiest to implement techniques for the adaptive estimation of the scale parameter σ^2 can be performed using an exponential smoothing scheme. This method can be viewed as a simple weighted average between the previous smoothed statistic and the information given by the recently available observation. The smoothing scheme writes as:

$$\hat{\sigma}_{t+k|t}^2 = (1-\alpha)\hat{\sigma}_{t+k-1|t-1}^2 + \alpha\left(p_t - \hat{p}_{t|t-k}\right)^2 \tag{13}$$

where α is a smoothing parameter. It is arbitrary set to 0.9997 in this work. The chosen value assumes a locally constant value of the variance, allowing for the long-term variations only.

2.4.2 GARCH and GARCHX models

In order to suggest a more advanced model for the squared deviations, the crucial features of the ϵ_t^2 series have to be revealed. Figure 3 depicts squared residuals of the 15-min ahead WPPT point forecasts for the period of time spanning the 20th of March, 2009 to the 20th of June, 2009. Volatility of the prediction errors exhibits clustering in time: large errors tend to be followed by large errors and similarly small errors tend to be followed by the small ones.

The autocorrelation function (ACF) and the partial ACF of the squared errors made by the WPPT 15-min ahead forecasts (see Figures 4) indicate that the conditional variance should be modelled as a weighted sum of approximately the last 18 squared errors (suggesting a high order AutoRegressive Conditional Heteroscedasticity (ARCH) model).

However, for the sake of parsimony, an ARCH model of a high order can be substituted by a Generalized ARCH (GARCH) specification [6].

Following [6] a noise process ϵ_t is called a GARCH process of order (p,q) if it satisfies:

$$\begin{aligned} &\epsilon_t = \sigma_t v_t, && v_t \overset{i.i.d.}{\sim} D(0,1) \\ &\sigma_t^2 = \alpha_0 + \textstyle\sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^{p} \beta_k \sigma_{t-k}^2 \end{aligned} \tag{14}$$

Where $D(0,1)$ denotes a distribution with a zero mean and a unit variance. For the specific case of the wind power variable analysed in this work (considering various prediction horizons k), the model (15) is applied to data as:

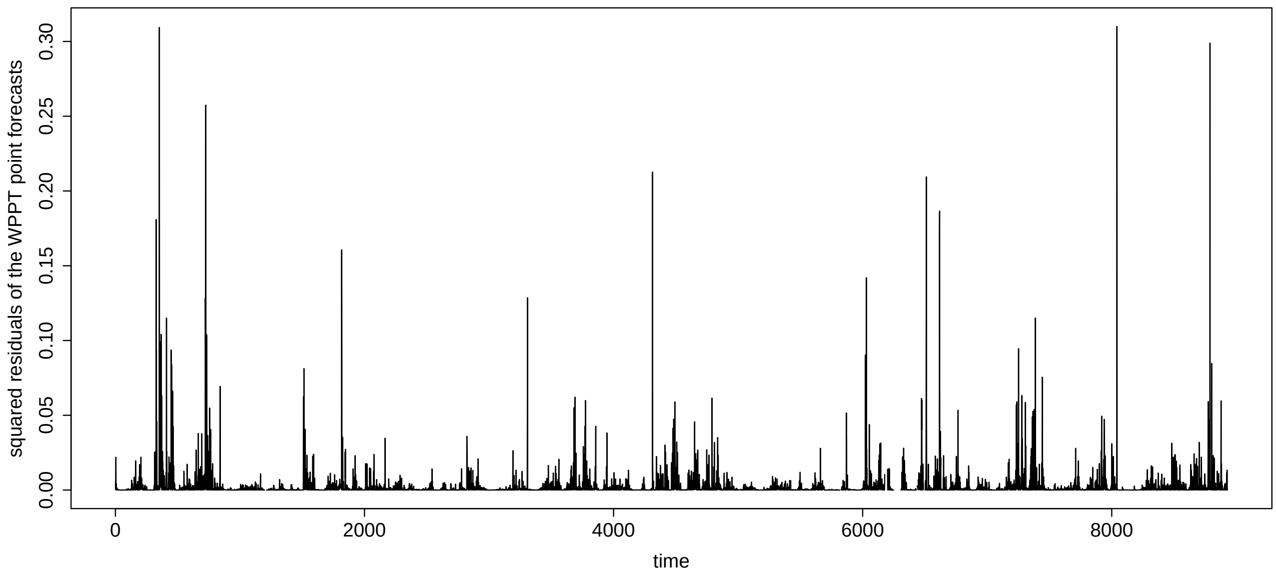


Figure 3: Squared residuals of the WPPT forecasts (prediction horizon is 15 min) in the period of the 20th of March, 2009 to the 20th of June, 2009

$$\begin{array}{ll} \epsilon_{t|t-k} = \sigma_{t|t-k} v_t, & v_t \overset{i.i.d.}{\sim} D(0,1) \\ \sigma^2_{t|t-k} = \alpha_0 + \sum_{j=k}^{k} \alpha_j \epsilon^2_{t-j|t-j-k} + \sum_{l=k}^{k} \beta_l \sigma^2_{t-l|t-l-k} & \end{array} \tag{15}$$

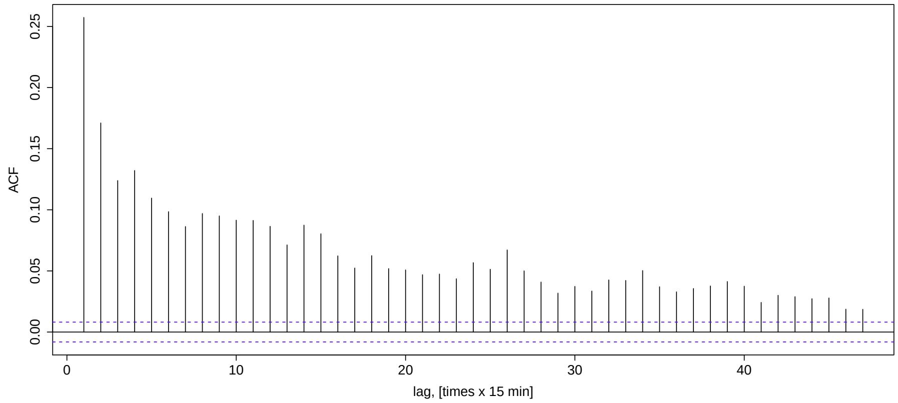
where $\epsilon_{t|t-k}$ stands for the errors made by a wind power (point) forecasting model (see (9)), $\sigma^2_{t|t-k}$ denotes a conditional variance of the point forecast residuals $\epsilon_{t|t-k}$. In other words, when building a GARCH model with targeting a k steps prediction horizon, GARCH(k,k) model is applied on data with some parameters fixed to zero: $\alpha_j = 0$ and $\beta_j = 0$, where $j = 1, ..., k-1$. For the sake of simplicity further on in the report we will skip the notation of the k steps ahead prediction horizon and stay with the simpler notation of the models given by ϵ_t rather than $\epsilon_t | t-k$

It is known that if ϵ_t follows a GARCH(p,q) process, then ϵ_t^2 follows an ARMA(r,q) process where r=max(p,q) [6]. Adaptive estimation of the model parameters can be carried out using a recursive least squares algorithm [15].

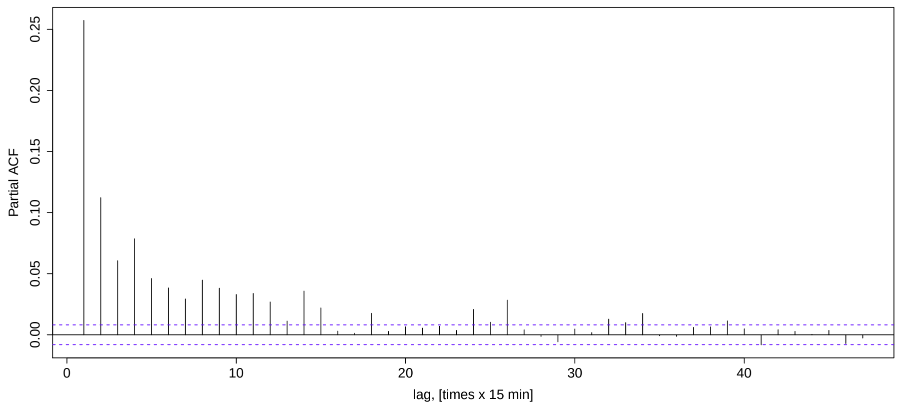
Based on the previous research works documented in [3, 2], the WPPT forecasts for Nysted Offshore can be improved if considering a spatial information in the prediction scheme. As previously highlighted, a point forecast corresponds to the mean (the first order moment) of the wind power variable. When moving to probabilistic framework an important question is whether spatial information can be further used for improving higher order moments, not only the mean.

Censored Gaussian distribution is fully characterised by the location and the scale parameters. They are approximated by the the mean and the variance of the predictive density, respectively. Therefore, the question boils down to wondering whether the spatial information can be incorporated into the models for describing the variance.

Figure 5 shows an empirical cross correlation function between the pre-whitened series of squared deviations at Kappel wind farm and the filtered series of the squared prediction errors observed at Nysted Offshore. The considered wind farms are numbered as 6 and 10 on the map, respectively (see Figure 1). The plot reveals that the strongest correlation is between the squared deviations at time t at Nysted Offshore and the squared WPPT forecast errors at Kappel observed one hour prior. In other words, the situation at Nysted Offshore "now" depends on the situation at Kappel "one hour ago". Such preliminary diagnostics suggests that the GARCH model (15) can be further extended by including an eXogenous input given by the squared errors observed at the distant location. The corresponding model is denoted by GARCHX and is defined as:



(a) Autocorrelation of the squared errors



(b) Partial autocorrelation of the squared errors

Figure 4: Autocorrelation and partial autocorrelation of the squared errors made by the WPPT 15-min ahead forecasts. Estimated on the data spanning the 1st of January, 2009 to the 31st of December, 2009

$$\begin{array}{ll} \epsilon_t = \sigma_t v_t, & v_t \overset{i.i.d.}{\sim} D(0,1) \\ \sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^{p} \beta_k \sigma_{t-k}^2 + \sum_{j=1}^{l} \gamma_j \xi_{t-j}^2 & \end{array} \tag{16}$$

where additionally to the notations of the GARCH model ξ_t^2 denotes squared errors made by WPPT at the remote location. For the prediction horizons up to 1 hour squared residuals at Kappel wind farm are considered. For the larger horizons ξ_t^2 is given by the squared errors at Horns Rev I (farm number 18). The choice is based on the analysis of the cross-correlation functions. The results are in line with the geographical layout of the farms (see Figure 1) and the considered temporal resolution.

If ϵ_t follows a GARCHX(p,q,l) process, then ϵ_t^2 follows and ARMAX(r,q,l) process where r=max(p,q). Adaptive estimation of the model parameters can be carried out using a recursive least squares algorithm [15].

2.4.3 CP-ARCH and CP-ARCHX models

Recall the results demonstrated in Figure 2: conditional densities of the observed power change with the level of the predicted power. This effect can be accounted for if letting the coefficients of the model be dependent on the level of the predicted power. This calls for considering Conditional Parametric ARCH (CP-ARCH) models for estimating σ_t^2.

Analogically to ARCH, but with non-constant parameter values, one can define a CP-ARCH process of order (q) as a noise process ϵ_t satisfying:

$$\begin{array}{ll} \epsilon_t = \sigma_t v_t, & v_t \overset{i.i.d.}{\sim} D(0,1) \\ \sigma_t^2 = \alpha_0(\hat{p}_t) + \sum_{j=1}^{q} \alpha_j(\hat{p}_t) \epsilon_{t-j}^2 & \end{array} \tag{17}$$

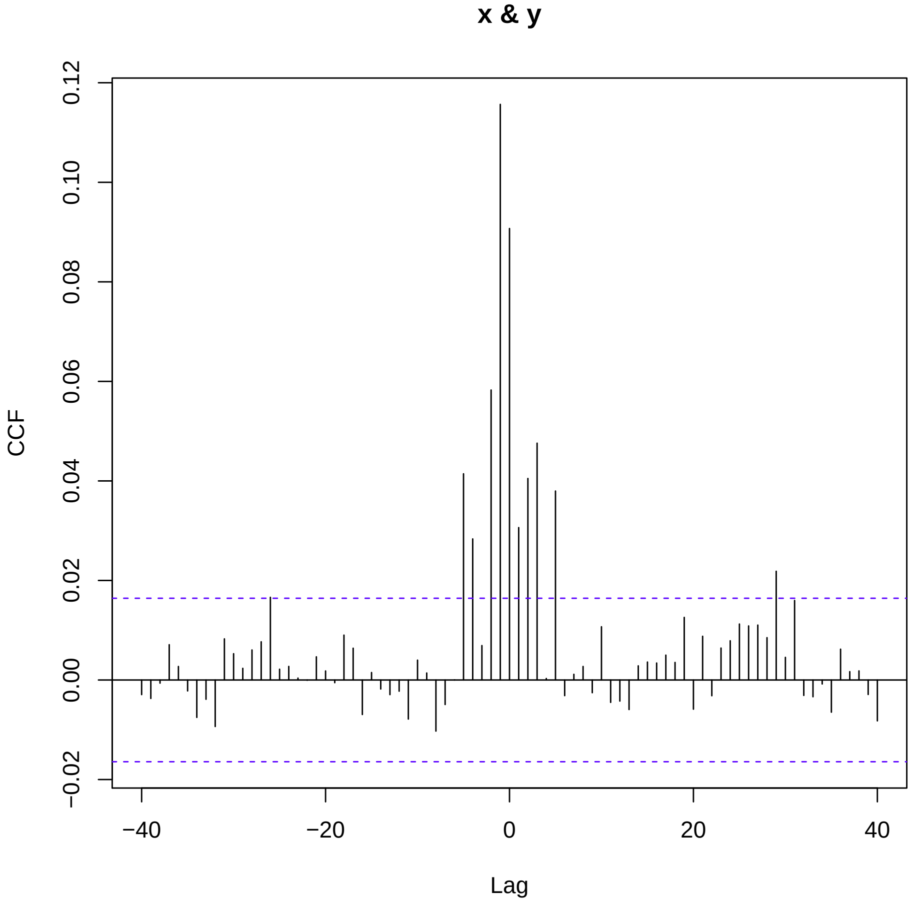


Figure 5: The empirical cross correlation function between the pre-whitened input series (x) of the squared errors made by the WPPT forecast at Kappel wind farm and the filtered output series (y) of the squared errors made by the CP model for Nysted Offshore. Lags are given in hours. 1-hour ahead predictions are considered.

Where $D(0,1)$ denotes a distribution with a zero mean and a unit variance. For the specific case of the wind power variable analysed in this work, ϵ_t stands for the errors made by a wind power (point) forecasting model, σ_t^2 denotes a conditional variance of the residuals ϵ_t. The main difference between the ARCH and the CP-ARCH models comes from the nature of the model parameters. The parameters of the CP-ARCH model α_j, $j = 0, ..., q$ instead of considered constant (as in case of the ARCH model) are assumed to be smooth functions of $\hat{p}_t$. In this particular case $\hat{p}_t$ stands for the wind power (point) prediction at Nysted Offshore issued for time t.

A spatial information (an eXogenous signal) can be included into the CP-ARCH model resulting in the CP-ARCHX specification:

$$\begin{array}{ll} \epsilon_t = \sigma_t v_t, & v_t \overset{i.i.d.}{\sim} D(0,1) \\ \sigma_t^2 = \alpha_0(\hat{p}_t) + \sum_{j=1}^{q} \alpha_j(\hat{p}_t)\epsilon_{t-j}^2 + \sum_{j=1}^{l} \gamma_j(\hat{p}_t)\xi_{t-j}^2 & \end{array} \tag{18}$$

where additionally to the notations of the CP-ARCH model ξ_t^2 denotes squared errors of the WPPT forecast at the remote location. For the prediction horizons up to 1 hour squared deviations at Kappel wind farm are considered. For the larger horizons ξ_t^2 is given by the squared errors at Horns Rev I (farm number 18).

The CP-ARCH process for ϵ_t can be written as a Conditional Parametric AutoRegressive (CP-AR) model for the ϵ^2. Similarly CP-ARCHX translates to a Conditional Parametric AutoRegressive model with an eXogenous input (CP-ARX) for the squared residuals. The parameters can then be estimated using an adaptive recursive algorithm used for the parameter estimation in the conditional parametric models. For the estimation details see [1].

2.4.4 Markov switching models

Figure 3 indicates that the variance of the wind power forecast errors is not stationary. The succession of periods with volatility of higher and lower magnitudes suggests a regime switching approach. Factors conditioning the changes in the volatility dynamics are not known to the full extent. Motivated by the results shown in Figure 2, the CP-ARCH and CP-ARCHX models assume that the changes in the wind power densities depend on the level of the predicted power. However, this assumption is just a simplification of the underlying complex process. As an alternative Markov Switching models allow a sequence of the regimes to be governed by a hidden, unknown signal, instead of being directly explained by some observable process. The estimation of the transitions between the regimes is carried out in a probabilistic way. Such types of models have been successfully implemented for describing the volatile behaviour of the offshore wind power generation in [22, 18].

Following [24] this can be written as following . Let ϵ_t^2, $t = 1, ..., n$ be the time series of the squared deviations from a wind power point forecast. In parallel, consider c_t a regime sequence taking a finite number of discrete values, $c_t \in \{1, ..., m\}$, $\forall t$. In this work $m = 2$. It is assumed that ϵ_t^2 is an AutoRegressive process (The corresponding Markov Switching model is then denoted by MS-ARCH) governed by the regime sequence in the following way:

$$\epsilon_t^2 = \alpha_0^{(c_t)} + \sum_{j=1}^{p} \alpha_j^{(c_t)} \epsilon_{t-j}^2 + e_t^{(c_t)} \tag{19}$$

or an AutoRegressive process with an eXogenous input (The corresponding Markov Switching model is then denoted by MS-ARCHX) described by:

$$\epsilon_t^2 = \alpha_0^{(c_t)} + \sum_{j=1}^{p} \alpha_j^{(c_t)} \epsilon_{t-j}^2 + \sum_{j=1}^{l} \gamma_j^{(c_t)} \xi_{t-j}^2 + e_t^{(c_t)} \tag{20}$$

where analogically to model (16), ξ_t^2 denotes squared errors of the WPPT forecast at the remote location. For the prediction horizons up to 1 hour squared deviations at Kappel wind farm are considered. For the larger horizons ξ_t^2 is given by the squared errors at Horns Rev I (farm number 18). $e_t^{(c)}$ is a Gaussian white noise process in regime c, i.e. a sequence of independent random variables with the density function η:

$$\eta^{(c)}(e) = \frac{1}{\sigma_t^{(c)}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{e}{\sigma_t^{(c)}}\right)^2\right) \tag{21}$$

In addition it is assumed that the regime sequence c_t follows a first order Markov chain on the finite space $1, ..., m$:

$$p_{ij} = P(c_t = j | c_{t-1} = i, c_{t-2}, ..., c_0) = P(c_t = j | c_{t-1} = i) \tag{22}$$

The probabilities governing the switches between the regimes are gathered is the transition probability matrix Γ:

$$\mathbf{\Gamma} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix} \tag{23}$$

where $p_{ij} >= 0$, $\forall i, j$ and $\sum_{j=1}^{m} p_{ij} = 1$, $\forall i$.
A likelihood L of the Markov Switching model (19) is then given by:

$$L_T = \boldsymbol{\delta} \boldsymbol{P}(e_1) \mathbf{\Gamma} \boldsymbol{P}(e_2) \ldots \mathbf{\Gamma} \boldsymbol{P}(e_T) \mathbf{1}' \tag{24}$$

where $\boldsymbol{P}(e_j) = \begin{pmatrix} \eta^{(1)}(e_j) & & 0 \\ & \ddots & \\ 0 & & \eta^{(m)}(e_j) \end{pmatrix}$

and $\boldsymbol{\delta}$ is an initial distribution which is assumed to correspond to the stationary process distribution found from:

$$\boldsymbol{\delta}(\boldsymbol{I}_m - \mathbf{\Gamma} + \boldsymbol{U}) = \mathbf{1} \tag{25}$$

where $\mathbf{1}$ is a row vector of ones, $\boldsymbol{I}_m$ is an identity matrix of order m and $\boldsymbol{U}$ is an $m \times m$ matrix of ones.

The set of coefficients allowing to fully characterize the Markov switching model (19) is given by $\boldsymbol{\Theta} = \left[\boldsymbol{\Gamma}, \alpha_0^{(1)}, \ldots, \alpha_p^{(1)}, \ldots, \alpha_0^{(m)}, \ldots, \alpha_p^{(m)}, \sigma^{(1)}, \ldots, \sigma^{(m)}\right]$. Estimation of the model parameters is performed by numerically maximizing the likelihood function with respect to $\boldsymbol{\Theta}$. In order to avoid dealing with the constraints in the optimization routine, a re-parametrization of $\boldsymbol{\Gamma}$ and $\sigma^{(1)}, \ldots, \sigma^{(m)}$ is carried out as proposed in [24].

In this work the parameters of the MS-AR model are estimated in a non-adaptive way. We have tried to employ an adaptive estimation scheme (using a sliding window of 5000 observations) and the results were shown not to improve. It was therefore decided to stick to the non-adaptive estimation procedure for the sake of simplicity.

3 Non-parametric approach to probabilistic forecasting

A non-parametric approach for estimating the predictive densities of the wind power generation does not assume a specific, known distribution for the data. Instead, it suggests estimating quantiles of the data directly from the previous observations. Adaptive quantile regression offers a way of estimating quantiles in the slowly varying non-stationary systems using linear regression techniques.

Quantile regression as presented in [12] is based on a linear model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{r} \tag{26}$$

where $\boldsymbol{y}$ denotes a vector of realizations of the random variable Y, $\boldsymbol{X}$ is a design matrix, i.e. a matrix containing explanatory variables, $\boldsymbol{r}$ denotes a vector of the error terms and $\boldsymbol{\beta}$ is a parameter vector to be estimated from the data. Commonly, in case of an ordinary linear regression, a quadratic loss function is applied for obtaining the parameter estimate $\hat{\boldsymbol{\beta}}$. The resulting estimate $\hat{y}_t = \boldsymbol{x}_t\hat{\boldsymbol{\beta}}$ corresponds to the conditional mean, i.e. the expected value of the random variable Y. Instead of being quadratic, a loss function $Q_\tau(r)$ for the quantile regression is based on a weighting of the absolute values of the residuals:

$$Q_\tau(r) = \begin{cases} \tau r & \text{if } r >= 0 \\ (\tau - 1)r & \text{if } r < 0 \end{cases}$$

Given N observations,the best estimate of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\beta}{\text{argmin}} \sum_{t=1}^{N} Q_\tau(r_t) \tag{27}$$

The resulting estimate of y_t given by $\hat{y}_t = \boldsymbol{x_t}\hat{\boldsymbol{\beta}}(\tau)$ is then equal to the conditional τ^{th} quantile of the random variable Y, i.e. the value which is not exceeded by an observation with a nominal proportion (probability) of τ. Solutions for (27) are obtained using linear programming techniques.

A time-adaptive parameter estimation method allows to update the estimate of the parameters as the new observations become available. The method is introduced and described in details in [14]. The main idea behind it is to update the design matrix $\boldsymbol{X}$ by leaving out the oldest observations as a new data becomes available. The parameters are then re-estimated based on the updated version of the design matrix and using the previously estimated parameter value for initialization of the new optimization procedure (27). In this work we follow the routine as described in [14], but skip the idea of using the old parameter value as an initial step in the updating procedure (i.e. we rather follow the exact approach of [23]). This is done for the sake of implementation simplicity and due to the fact that the quality of the results is not affected by omitting the aforementioned step.

Application to the case study The quantile models studied here are given by

$$\epsilon_t = g(\hat{p}_t, \tau) + r_t = \beta_{0,t}(\tau) + f_t(\hat{p}_t, \tau) + r_t = \beta_{0,t}(\tau) + \sum_{j=1}^{K-1} b_j(\hat{p}_t)\beta_{j,t} + r_t \tag{28}$$

where ϵ_t is an error made by the point forecast, $\hat{p}_t$ is the forecasted power, τ is the required quantile, r_t is a noise term. The b_j are natural cubic B-spline basis functions, K is the number of knots used for the spline construction. In other words, the model assumes that the τ^{th} quantile of the forecast error ϵ_t is a non-linear function g of the predicted power. Function g is viewed as a piecewise cubic function of $\hat{p}_t$ and can therefore be expressed as a linear combination of the known basis functions b_j. This representation permits to use the estimation techniques valid for the linear model given by (26).

Some technical details: the model examined in this work has the knots for the spline basis functions placed in steps of 25% quantiles of the forecasted wind power. The design matrix contains 2100 recent observation. It is updated and the model parameters are re-estimated daily at 00:00. Considered quantiles $\tau \in T = \{0.05, 0.10, ..., 0.95\}$

The intuitive explanation of model (28) is the following. Preliminary data analysis given in Figure 2 showed that the density of the wind power generation depends on the level of the predicted power. Model (28) permits to estimate in an optimal way the set of quantiles T of the point forecast prediction error as a non-linear function of the expected wind power. Knowing the quantiles, it is possible to "dress" the point forecasts of the wind power with a set of prediction intervals of different nominal coverage rates and obtain a full predictive density. For instance, knowing 0.05 and 0.95 quantiles of the ϵ_t (denoted as $\epsilon_{0.05,t}$ and $\epsilon_{0.95,t}$, respectively) it is possible to say that an actually observed power generation at time t will with 90% certainty be covered by the interval $[\hat{p}_t + \epsilon_{0.05,t}, \hat{p}_t + \epsilon_{0.95,t}]$.

4 Methods for probabilistic forecasts assessment

Forecast quality relates to the degree of correspondence between forecasts and observations. When evaluating such similarity there are many aspects to be considered. The most common approach is to compute a measure of the overall correspondence between predictions and outcomes. For instance, the accuracy of the forecasts is given by a measure of the average correspondence between individual pairs of forecasts and observations. In this work the accuracy is evaluated using a Conditional Ranked Probability Score (it is defined and discussed below). Computation of measures of such overall correspondence is very helpful and useful when several probabilistic forecasts are to be compared. However, the drawback is that it shuffles different aspects of forecast quality, weights each of them in a certain way and provides a corresponding summarizing value. In order to truly understand how the forecast operates in different situations and how consistent are the issued predictive densities with the observed proportions the overall score has to be decomposed and different aspects have to be assessed separately. In this work additionally to the overall accuracy and skill of the forecasts we will focus on such aspects as calibration, sharpness and conditional evaluation of the predictive densities.

Reliability (calibration) refers to the statistical consistency between the distributional forecasts and the observations: it reflects how close the nominal and the empirical proportions are. Over an evaluation set of a significant size, the observed and the predicted coverage rates should be as close as possible. Calibration is viewed as a very important quality for the probabilistic forecasts. However, alone it does not make for a useful forecast [10]. For example a climatology forecast (forecast based on the overall histogram of the historical data, not based upon the dynamic implications of the current state of the system) constitutes a reliable forecast. This forecast though is constant, not flexible enough and simply unable to satisfactory represent possible outcomes in the current situation. Due to that, besides the wish for the forecasts to be calibrated, they are also desired to provide forecast users with a situation-dependent informative predictions. This is closely related to the quality of sharpness.

Sharpness refers to the degree of concentration of the distribution of the probabilistic forecast. If the density forecast takes the form of a delta function, this would correspond to a maximum possible sharpness and would equate with the idealized concept of the perfect point forecast. In contrast, the climatology forecast is generally not sharp as there is a probability that the future observation can take on any value that has been observed in the past.

Conditional evaluation concerns assessment of the conditional distributions. As has been previously discussed the predictive densities of the wind power generation are not constant, - they change with the level of expected power. Thus is it interesting to check how well the issued probabilistic forecasts perform in the moments when the power generation is high, low or medium.

All three aforementioned characteristics (reliability, sharpness and quality of the conditional densities) are very important for making a good probabilistic forecast and it is a combination of them which plays the most important role. One forecast might be reliable, but lacking sharpness. Another one might be sharp, but lacking calibration. Which one is better? The answer is commonly given by some proper scoring rules.

Scoring rules assess the quality of the predictive density by assigning a numerical score based on the forecast and on the event or value which materializes. They assess different quality aspects simultaneously providing a single numerical value summarizing the quality of the forecast performance. Following [7] let $S(F, x)$ denote a score assigned when a forecaster issues the predictive density F and x is the value which actually materializes. We consider scores as penalties which the forecaster wishes to minimize on average. A scoring rule is proper if the expected value of the penalty $S(F, x)$ for an observation x drawn from G is minimized if $F = G$, i.e. if the forecast is perfect. In other words, the smaller the obtained score is, the better the issued forecast is (the closer it is to the perfect forecast). The score is called strictly proper if the minimum is unique. In estimation problems proper scoring rules encourage a forecaster to do careful assessments and to be honest. Every attempt to speculate on, for instance, sharpness with the price of losing in calibration (or vice versa) is being penalized and is correspondingly reflected in the value of the obtained score.

Conditional Ranked Probability Score (crps) has gained popularity as a means of evaluating wind power probabilistic forecasts. It is defined as:

$$crps(F, x) = -\int_{-\infty}^{\infty} \left(F(y) - \mathbf{1}\{y >= x\}\right)^2 dy \tag{29}$$

where $F(y)$ is a predictive cumulative distribution function, x is a value which materialized (an actual observation), $\mathbf{1}\{y >= x\}$ denotes a function that gives the value 1 if $y >= x$ and 0 otherwise.

For assessing a probabilistic forecast over a data set containing T observations the average of the crps values for each forecast/verification pair is calculated.

$$CRPS = \frac{1}{T}\sum_{t=1}^{T} crps(F_t, x_t) \tag{30}$$

The choice of using CRPS is motivated by the facts, that firstly, it is a proper scoring rule [9]. Secondly, it is a distance sensitive rule, meaning that a credit is given for assigning high probabilities to values near, but not identical to the one materializing. Another useful property of the CRPS score arises from the fact that for point forecasts it reduces to the absolute error. Thus CRPS provides a direct way to compare point and probabilistic forecasts.

Skill score For comparing forecast models it is convenient to introduce a measure of the relative improvement in CRPS with respect to the considered reference forecast model. Such improvements are given by the corresponding skill score δ which is defines as:

$$\delta = \frac{CRPS_{ref} - CRPS_m}{CRPS_{ref}} 100\% \tag{31}$$

where $CRPS_m$ corresponds to the CRPS score of the considered forecast model and $CRPS_{ref}$ is the CRPS score of the reference model.

5 Results

5.1 Notation

Before discussing the results the following notation is introduces: all the considered models are noted as "Expectation Model/Volatility Model" where in front of "/" stands a type of the point predictions used as input and after the "/" the used uncertainty estimation method is introduced. Following that:

Non-parametric models ("/QR"):

- "Model"/QR stands for the non-parametric predictive densities based on the adaptive quantile regression. "Model" specifies the type of point predictions used as input. (WPPT/QR, CP/QR, Logit-CP/QR).

Parametric models under the assumption of the censored Normal distribution ("CN:") :

- CN:WPPT/Exp.smooth - A parametric model under the assumption of the censored Normal distribution. The WPPT point predictions are used as input. The scale parameter of the suggested distribution is estimated using the exponential smoothing technique (13).
- CN:CP/Exp.smooth - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested distribution is estimated using the exponential smoothing technique (13).
- CN:CP/GARCH - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the GARCH approach (15).
- CN:CP/GARCHX - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input.The scale parameter of the suggested error distribution is estimated using the GARCH approach (16).
- CN:CP/CP-ARCHX - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the CP-ARCHX approach (18).
- CN:CP/CP-ARCH - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the CP-ARCH approach (17).
- CN:Logit-CP/CP-ARCH - A parametric model under the assumption of the censored Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the CP-ARCH approach (17).
- CN:CP/MS-AR - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the MS-AR approach (19).
- CN:CP/MS-ARX - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the MS-ARX approach (20).

Parametric models under the assumption of the generalized logit-Normal distribution ("GLN:") :

- GLN:Logit-CP/Exp.smooth - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the exponential smoothing technique (13).
- GLN:Logit-CP/GARCHX - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the GARCHX approach (15).
- GLN:Logit-CP/CP-ARCHX - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the CP-ARCHX approach (18).
- GLN:Logit-CP/MS-AR - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the MS-AR approach (19).
- GLN:Logit-CP/MS-ARX - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the MS-ARX approach (20).

5.2 Accuracy and skill assessment

Different parametric and non-parametric models have been used for issuing probabilistic forecasts for Nysted Offshore. The CRPS scores for all the models have been evaluated over a period from the 1st of January, 2009 to the 31st of December, 2009. Only the data points where none of the considered models have a missing value were used in the evaluation set. This resulted in approximately 25 000 ($\approx$ 8.5 months) active data points for every of the considered prediction horizons. The CRPS results for all the considered models run with the different prediction horizons are presented in Table 2.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Non-parametric densities														
WPPT/QR	*2.34*		*5.17*		*7.69*		*7.97*		*8.19*		*8.34*		*8,47*	
CP/QR	2.31	1.5	4.94	4.6	7.43	3.5	7.73	2.9	8.00	2.3	8.13	2.5	8.29	2.1
Logit-CP/QR	2.31	1.5	4.92	4.9	7.42	3.5	7.74	2.8	8.02	2.1	8.15	2.3	8.31	1.9
Censored Normal predictive densities														
CN:WPPT/Exp. smooth.	2.49		5.40		7.89		8.17		8.39		8.54		8.67	
CN:CP/Exp. smooth.	2.46	-5.2	5.14	0.6	7.63	0.8	7.94	0.3	8.21	-0.2	8.37	-0.4	8.53	-0.7
CN:CP/GARCH	2.37	-1.4	5.06	2.2	7.60	1.2	7.92	0.6	8.06	1.6	8.36	-0.2	8.52	-0.5
CN:CP/GARCHX	2.35	-0.5	5.03	2.7	7.60	1.3	7.92	0.6	8.20	-0.1	8.35	-0.1	8.52	-0.5
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0
CN:CP/CP-ARCHX	2.31	1.2	4.95	4.4	7.46	3.0	7.78	2.4	8.07	1.5	8.24	1.3	8.40	0.8
CN:Logit-CP/CP-ARCH	2.31	1.2	4.93	4.7	7.42	3.5	7.75	2.7	8.03	1.9	8.19	1.8	8.36	1.4
CN:CP/MS-AR	2.37	-1.3	5.06	2.2	7.62	0.9	8.31	-4.4	8.21	-0.2	8.41	-0.8	8.57	-1.1
CN:CP/MS-ARX	2.37	-1.1	5.05	2.4	7.61	1.1	7.97	0.0	8.21	-0.3	8.37	-0.4	8.55	-0.9
Generalized logit-Normal predictive densities														
GLN:Logit-CP/Exp. smooth	2.43	-3,6	5.09	1,7	7.64	0,7	7.96	0,1	8.23	-0,5	8.38	-0,5	8.55	-0,9
GLN:Logit-CP/GARCH	2.33	0.6	4.95	4.2	7.59	1.4	7.93	0.4	8.21	-0.2	8.36	-0.3	8.54	-0.7
GLN:Logit-CP/CP-ARCH	2.31	1.5	4.94	4.6	7.54	1.9	7.87	1.3	8.15	0.5	8.32	0.3	8.50	-0.3
GLN:Logit-CP/MS-AR	2.35	-0.2	4.98	3.8	7.65	0.5	8.01	-0.5	8.26	-0.8	8.42	-1.0	8.63	-1.8
GLN:Logit-CP/MS-ARX	2.34	-0.2	4.97	3.9	7.63	0.9	7.99	-0.3	8.23	-0.5	8.39	-0.6	8.58	-1.3

Table 2: Evaluation of the density forecasts with a CRPS criterion. An adaptive quantile regression over the initial WPPT point forecasts is considered a benchmark(given in italic). Other models are compared to the benchmark model and the corresponding relative improvements are given by δ.

As a benchmark it is chosen to use the adaptive quantile regression using the WPPT forecast as input. This is a very strong benchmark as both the WPPT and the adaptive quantile regression are the state-of-the-art approaches. Other models are compared to the benchmark model and the relative improvements δ are calculated. The results as they are presented in Table 2 may look a bit overwhelming - too many of them to be able to note all the details. In order to make it easier to spot the main characteristic, in the following paragraphs several snapshots of the overall results are taken and discussed in more details.

5.2.1 Parametric predictive densities

Estimating the first order moments

Spatial correction of the first order moments One way for describing a random variable in a probabilistic framework is to estimate all order moments (mean, variance, skewness, kurtosis, higher order moments). In case of the parametric approach, i.e. when the data is assumed to follow a known distribution, a finite number of moments will fully characterize the variable. The parametric densities considered in this work can be fully characterized by the first two order moments: the mean and the variance (see Section 2.1 for details).

This implies that using those two moments as input we can fully recreate the predictive density. The mean is given by the point predictions of the wind power generation. The results given in Table 3 show that the quality of the parametric predictive densities is improved if instead of the original WPPT forecasts the spatially corrected point predictions given by the CP model are considered. This indicates that accounting for the spatio-temporal effects while estimating the first order moment of the predictive density improves the quality of the corresponding probabilistic forecasts.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Censored Normal predictive densities														
CN:WPPT/Exp. smooth.	2.49		5.40		7.89		8.17		8.39		8.54		8.67	
CN:CP/Exp. smooth.	2.46	-5.2	5.14	0.6	7.63	0.8	7.94	0.3	8.21	-0.2	8.37	-0.4	8.53	-0.7

Table 3: Evaluation of the density forecasts with a CRPS criterion. Comparison of the Censored Normal predictive densities when using the WPPT and the spatially corrected CP forecasts as input. δ shows the corresponding relative improvements over the WPPT/QR

Incorporating the generalized logit transformation in to the input correction model In [3] it is shown that the quality of the point predictions is further improved if the generalized logit data transformation is incorporated into the spatial correction models (resulting in the Logit-CP point predictions). The results given in Table 4 show that using the Logit-CP predictions instead of the CP ones as input improves the quality of the corresponding predictive densities. In other words, considering the data transformation when estimating the expectation of the Censored Normal distribution brings slight improvements in all the considered horizons.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Non-parametric densities														
WPPT/QR	*2.34*		*5.17*		*7.69*		*7.97*		*8.19*		*8.34*		*8,47*	
Censored Normal predictive densities														
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0
CN:Logit-CP/CP-ARCH	2.31	1.2	4.93	4.7	7.42	3.5	7.75	2.7	8.03	1.9	8.19	1.8	8.36	1.4

Table 4: Evaluation of the density forecasts with a CRPS criterion. Comparison of the Censored Normal predictive densities using the CP and the Logit-CP forecasts as input. δ shows the corresponding relative improvements over the WPPT/QR

Estimating the second order moments

Accounting for the dynamics in variance of the wind power generation As has been discussed in Section 2 the variance of the wind power generation is not constant. Accounting for this improves the results significantly. This is demonstrated in Table 5 where several different approaches for estimating the variance are compared.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Non-parametric densities														
WPPT/QR	*2.34*		*5.17*		*7.69*		*7.97*		*8.19*		*8.34*		*8,47*	
Censored Normal predictive densities														
CN:CP/Exp. smooth.	2.46	-5.2	5.14	0.6	7.63	0.8	7.94	0.3	8.21	-0.2	8.37	-0.4	8.53	-0.7
CN:CP/GARCH	2.37	-1.4	5.06	2.2	7.60	1.2	7.92	0.6	8.06	1.6	8.36	-0.2	8.52	-0.5
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0

Table 5: Evaluation of the density forecasts with a CRPS criterion. Comparison of the Censored Normal predictive densities based on different techniques to model the dynamics of the variance. δ shows the corresponding relative improvements over the WPPT/QR

Exponential smoothing technique assumes a locally constant variance, allowing for the long-term variations, only. The GARCH model accounts for the changes in the dynamics of the variance based on the previous (most recent) observations. This technique accounts for the changing dynamics. The main drawback of the GARCH method is that it is based purely on the past observations, meaning that the predictive performance of the model has a certain delay - i.e. the model is not capable to predict when exactly the change will occur, but once the dynamics has actually changed the model is able to notice this and adapt correspondingly. That is why the model performs best on the data exhibiting the clustering effect. The CP-ARCH model (as discussed in Section 2.4.3) captures the changes in the dynamics driven by variations in the level of the expected power generation. This approach is motivated by the results depicted in Figure 2 where it is shown that the variance of the wind power generation depends on the level of the expected power. Differently from the GARCH approach, the conditional parametric method allows to describe the dynamics based not only on the previously observed fluctuations, but takes into the consideration the power forecast as well. Therefore this model is better adapted to foreseeing the changes in the dynamics of the wind power generation. The results given in Table 5 show that the CP-ARCH model provides the best input for the Censored Normal predictive density. The fact that the exponential smoothing shows the poorest results, proves that the variance of the wind power generation is not constant and accounting for the changing dynamics is important. As the GARCH model is outperformed by the CP-ARCH model, it is possible to conclude that the changes in the dynamics of the variance can be (partly) explained by the level of the expected power.

Modelling the dynamics of the variance: explicit modelling conditional on the level of the expected power versus the implicit Markov switching approach As has been previously discussed, conditioning the density of the wind power generation on any observable input would always be just an approximation of the actual complex meteorological phenomenon. Markov switching models propose an alternative approach by modelling the dynamics driven by the unknown signal. As can be seen from the results shown in Table 6 the explicit approach suggesting that the dynamics is driven by the level of the expected power outperforms the implicit one given by the Markov switching model.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Non-parametric densities														
WPPT/QR	*2.34*		*5.17*		*7.69*		*7.97*		*8.19*		*8.34*		*8,47*	
Censored Normal predictive densities														
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0
CN:CP/MS-AR	2.37	-1.3	5.06	2.2	7.62	0.9	8.31	-4.4	8.21	-0.2	8.41	-0.8	8.57	-1.1
Generalized logit-Normal predictive densities														
GLN:Logit-CP/CP-ARCH	2.31	1.5	4.94	4.6	7.54	1.9	7.87	1.3	8.15	0.5	8.32	0.3	8.50	-0.3
GLN:Logit-CP/MS-AR	2.35	-0.2	4.98	3.8	7.65	0.5	8.01	-0.5	8.26	-0.8	8.42	-1.0	8.63	-1.8

Table 6: Evaluation of the density forecasts with the CRPS criterion. Two methods for capturing density dynamics are compared. The explicit method suggests that it is the level of the power generation which causes a variability in the density dynamics. The implicit method assumes that the driving force is unknown and tries to implicitly capture it by the Markov switching approach. An adaptive quantile regression over the initial WPPT point forecasts is considered as benchmark(given in italic). Other models are compared to the benchmark model and the corresponding relative improvements are given by δ.

This supports the idea that the dynamics of the variance of the wind power generation are indeed strongly affected by the level of the expected power. Therefore accounting for it directly in the modelling procedure is more beneficial than trying to capture the underlying complex phenomenon by applying Markov switching methods. The latter does not seem to capture any additional information than is extracted directly from the changes in the power expectation.

Accounting for the spatial information As has been discussed, the spatial correction of the first order moments of the predictive densities improves the performance of the corresponding probabilistic forecast. As the considered parametric densities are fully described by both the mean and the variance, an interesting question is whether the spatial correction of the variance can bring additional ameliorations. The answer is given by the results shown in Table 7. No significant improvements are achieved when considering the CP/GARCHX (CP/CP-ARCHX) model instead of the CP/GARCH (CP/CP-ARCH). This shows that no additional improvements in the model performance are achieved when including the spatial information into the variance estimation.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Non-parametric densities														
WPPT/QR	*2.34*		*5.17*		*7.69*		*7.97*		*8.19*		*8.34*		*8,47*	
Censored Normal predictive densities														
CN:CP/GARCH	2.37	-1.4	5.06	2.2	7.60	1.2	7.92	0.6	8.06	1.6	8.36	-0.2	8.52	-0.5
CN:CP/GARCHX	2.35	-0.5	5.03	2.7	7.60	1.3	7.92	0.6	8.20	-0.1	8.35	-0.1	8.52	-0.5
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0
CN:CP/CP-ARCHX	2.31	1.2	4.95	4.4	7.46	3.0	7.78	2.4	8.07	1.5	8.24	1.3	8.40	0.8

Table 7: Evaluation of the density forecasts with a CRPS criterion. Comparison of the models performance with and without accounting for the spatial effects when estimating the variance. δ shows the corresponding relative improvements over the WPPT/QR

Comparing two type of parametric densities: Censored Normal versus the Generalized logit-Normal Previously it has been discussed that applying the generalized logit transformation on the data permits to estimate the expectation of the wind power generation in a more accurate way (see [3] for more details). This suggests that the proposed transformation is suitable for stabilizing the variance of the wind power generation (making it less dependent on the power expectation) allowing for more robust point predictions. The following question is whether the transformation also makes the data look more Gaussian, ie whether the assumption that the transformed data is Normally distributed surpasses the suggestion that the non-transformed data is Gaussian. This is analogical to wondering which of the studied parametric densities describes the data better - the Censored Normal distribution or the generalized logit-Normal one. The results given in table 8 indicate that the censored Normal distribution shows better results. This indicates that even though the generalized logit transformation helps to stabilize the variance, makes it less dependent on the bounds, it does not make the assumption of Gaussianity more appropriate.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Censored Normal predictive densities														
CN:Logit-CP/CP-ARCH	2.31	1.2	4.93	4.7	7.42	3.5	7.75	2.7	8.03	1.9	8.19	1.8	8.36	1.4
Generalized logit-Normal predictive densities														
GLN:Logit-CP/CP-ARCH	2.31	1.5	4.94	4.6	7.54	1.9	7.87	1.3	8.15	0.5	8.32	0.3	8.50	-0.3

Table 8: Evaluation of the density forecasts with a CRPS criterion. Comparison of the Censored Normal and the generalized logit-Normal distributions. δ shows the corresponding relative improvements over the WPPT/QR

5.2.2 Non-parametric predictive densities

spatial correction of the input Similarly to the parametric predictive densities, the performance of the non-parametric predictive densities improves if instead of the WPPT forecasts the spatially corrected point predictions are considered as input. This once again proves that the quality of the probabilistic forecasts can be improved if the spatial effects are taken into consideration.

Considering data transformation for the input correction One can see that in case of the non-parametric predictive densities, the implementation of the generalized logit transformation does not help improving the corresponding probabilistic densities.

5.2.3 Parametric versus the non-parametric densities

In order to compare the parametric and the non-parametric approaches, we consider the CN:Logit-CP/CP-ARCH and the CP/QR which are the best performing parametric and non-parametric models, respectively. The results given in Table 10 show that the non-parametric densities outperform the parametric ones when considering larger prediction horizons. For shorter prediction horizons both approaches show similar results.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Non-parametric densities														
WPPT/QR	*2.34*		*5.17*		*7.69*		*7.97*		*8.19*		*8.34*		*8,47*	
CP/QR	2.31	1.5	4.94	4.6	7.43	3.5	7.73	2.9	8.00	2.3	8.13	2.5	8.29	2.1
Logit-CP/QR	2.31	1.5	4.92	4.9	7.42	3.5	7.74	2.8	8.02	2.1	8.15	2.3	8.31	1.9

Table 9: Evaluation of the non-parametric densities forecasts with a CRPS criterion.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %	CRPS, %	δ, %
Non-parametric densities														
WPPT/QR	*2.34*		*5.17*		*7.69*		*7.97*		*8.19*		*8.34*		*8,47*	
CP/QR	2.31	1.5	4.94	4.6	7.43	3.5	7.73	2.9	8.00	2.3	8.13	2.5	8.29	2.1
Censored Normal predictive densities														
CN:Logit-CP/CP-ARCH	2.31	1.2	4.93	4.7	7.42	3.5	7.75	2.7	8.03	1.9	8.19	1.8	8.36	1.4

Table 10: Evaluation of the density forecasts with a CRPS criterion. Comparison of parametric and non-parametric approaches

5.3 Reliability assessment

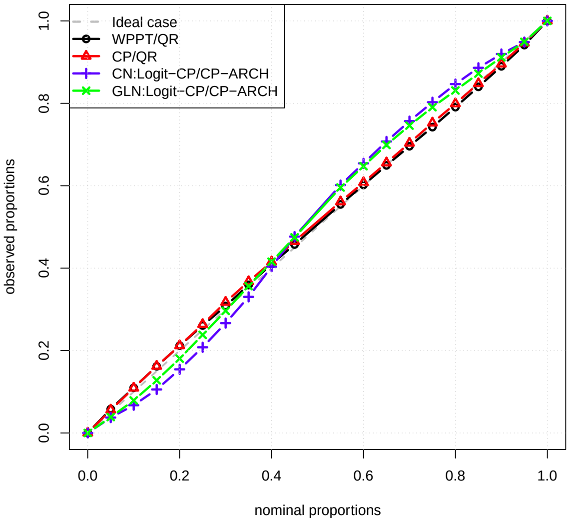
As has been previously discussed in Section 4 one of the most important qualities for the distributional forecast is reliability. In the earlier scientific works reliability is even viewed as a requirement rather than a desired property. The recent development, however, indicates that, as has been argued in Section 4 reliability by itself does not guarantee a useful forecast and should be only evaluated in relation with the sharpness and resolution (the joint performance evaluation is given by the score functions). However in this section the focus in on reliability, due to the particular importance given to it.

In the following analysis it is chosen to focus only on the benchmark model (WPPT/QR), the best performing non-parametric model (CP/QR) and two best performing parametric models (CN:Logit-CP/CP-ARCH) and (GLN:Logit-CP/CP-ARCH).

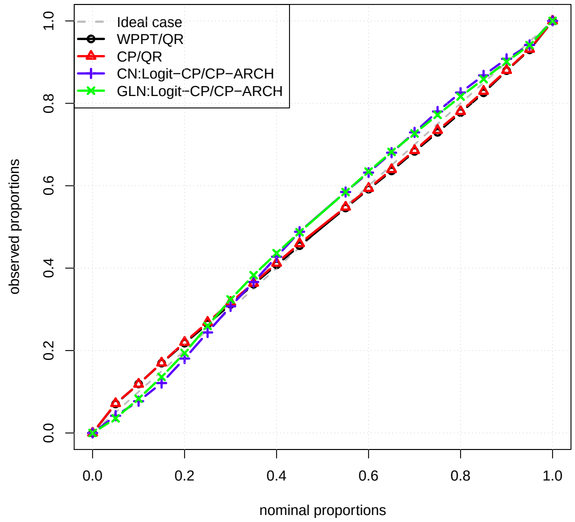
Figure 6 depicts reliability diagrams for the considered models when working on 1 hour, 4 hours and 8 hours ahead predictions. One can see that the non-parametric approach (based on the adaptive quantile regression) provides calibrated forecasts. On another hand, the models based on the parametric assumptions, deviate more from the nominal proportions and therefore one can claim that the parametric approaches result in probabilistically biased forecasts. Why are those forecasts biased? Partly the lack of calibration could be explained by the fact that parametric approaches imposes certain theoretical shapes for the distribution of the wind power generation. An important part of the assumption is based on the suggestion that with a non-zero probability a measured wind power will reach 0 and 100% of the nominal capacity. In fact, in the considered data set there were no observations reaching 100%. What was actually happening when the measured wind power was approaching the nominal capacity is shown in Figure 7. There is a clear indication that some human factor has been involved, i.e. most probably the power was down-regulated at those periods. Since such regulations are not of a constant level through the considered data set, it is difficult to replace the theoretical maximum of the nominal capacity by the adjusted value of the down-regulated power. Such deviations from the theoretical settings, where the power is allowed to reach the nominal capacity create a certain bias which is also reflected in the reliability diagrams of the parametric models. This could be corrected if the corresponding information on the wind farm regulation policy was available.

5.4 Conditional operation

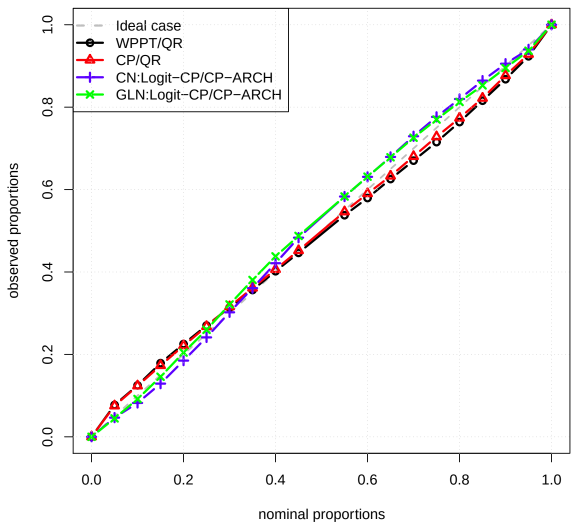
In this section the focus is on the conditional evaluation of the performance of the WPPT/QR, CP/QR, CN:logit-CP/CP-ARCH and GLN:Logit-CP/CP-ARCH models. In Section 5.2 it is shown that when evaluated on all the data available in the validation set, the CP/QR shows the best results in terms of the CRPS score. The goal of this section is to check whether this conclusion holds for various subsets of the validation set.



(a) Horizon=1 hour



(b) Horizon=4 hour



(c) Horizon=8 hour

Figure 6: Reliability diagram

5.4.1 Accuracy as a function of time

. Firstly the validation period is divided into 5 equally populated subsets and the CRPS scores of the considered models are calculated. The results are given in Figure 8.

One can see that the CP/QR model consistently (in all the considered subsets with different prediction horizons) outperforms the benchmark approach (WPPT/QR). The performance of the parametric densities with respect to the WPPT/QR is less steady. When considering 1 hour ahead forecasts, the parametric predictive densities, similarly to CP/QR outperform the WPPT/QR in all the subsets of the validation set. However, as the prediction horizon increases, the parametric models fail to outperform the benchmark approach in major part of the subsets of the validation set.

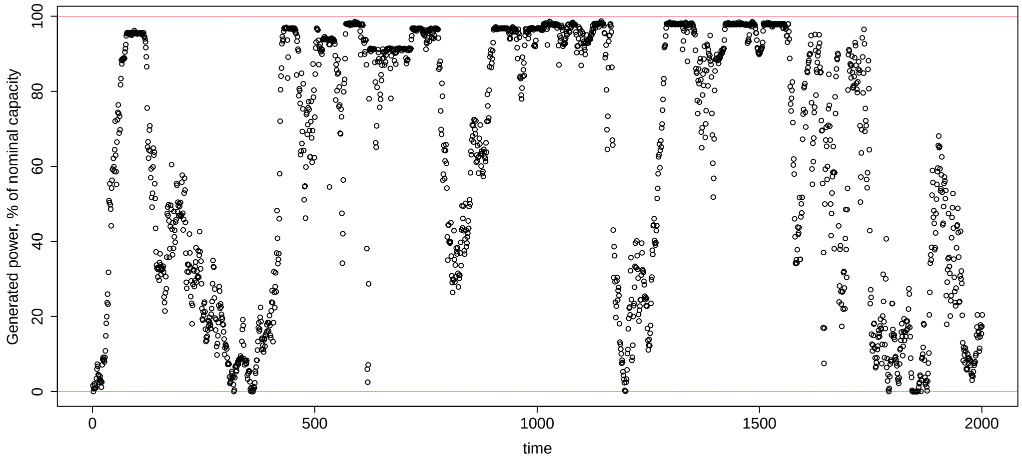


Figure 7: A period of the observed power generation at Nysted Offshore recorded in the period from 19:45 on the 4th of November, 2008 to 15:45 on the 25th of November, 2008

5.4.2 Sharpness as a function of the level of expected power generation

.

Recall, sharpness corresponds to the ability of probabilistic forecasts to concentrate the probabilistic information about future outcomes. In this work, following the approach by Pinson et at. [19] the sharpness is assessed by the mean widths of the central predictive intervals with a nominal coverage rate of 50%, ı.e. if writing

$$\beta_{t,h} = \hat{q}(0.75)_{t|t-h} - \hat{q}(0.25)_{t|t-h} \tag{32}$$

the size of the central interval forecast estimated at time $t-h$ for lead time t. Here $\hat{q}^{0.75}$ and $\hat{q}^{0.25}$ define the corresponding quantiles of the predictive densities. Then a measure of sharpness for horizon h is given by $\overline{\beta}_h$, the mean size of the intervals:

$$\overline{\beta}_h = \frac{1}{N}\sum_{t=1}^{N}\beta_{t,h} \tag{33}$$

The results given in Figure 9 show that with all the considered prediction horizons, the non-parametric predictive densities provide sharper forecasts than the parametric ones. This holds for all the levels of the expected power.

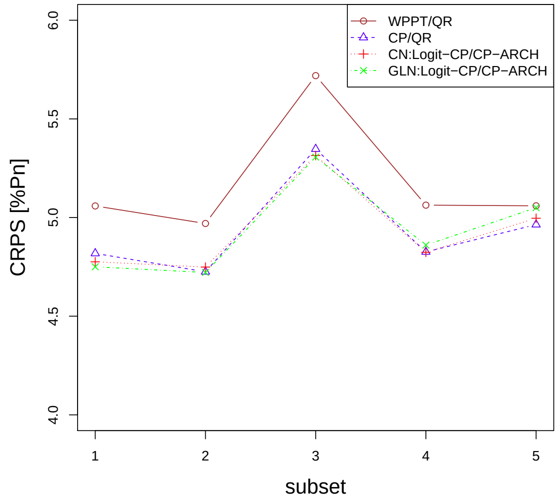
5.4.3 Accuracy and skill as functions of the level of expected power generation

.

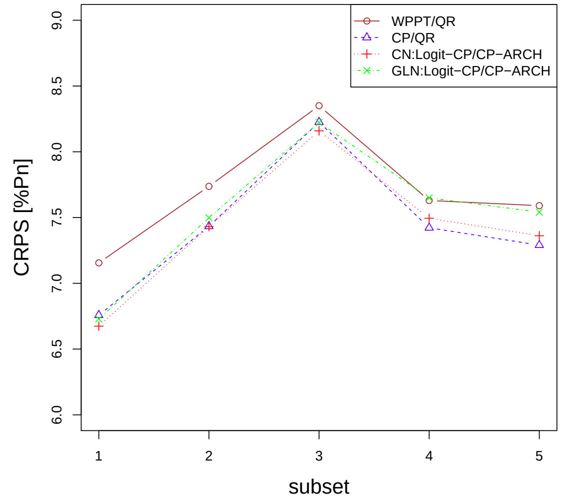
As has been previously discussed in this work, the densities of the wind power generation depend on the level of the expected power. It is thus interesting to see how the models perform conditional to the level of the forecasted power generation. The results in terms of conditional accuracy given by the CRPS scores are depicted in Figure 10 and the corresponding skill scores are shown in Figure 11.

From Figure 10 one can see that the CRPS scores increase when the expected power is not close to the generation bounds. This leads to higher uncertainty associated with the corresponding predictive densities. As the result the increase in the CRPS scores is observed.

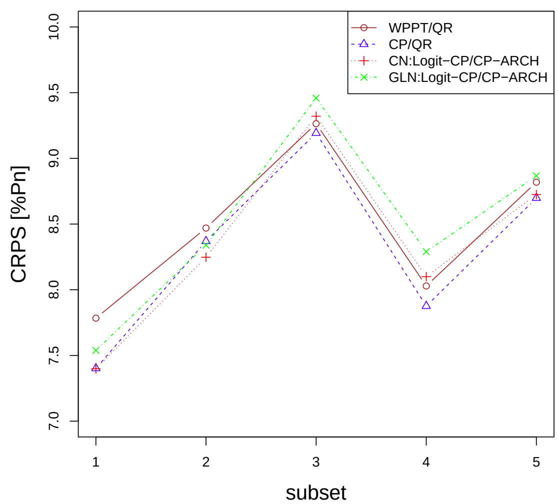
The analysis of the conditional skill score depicted in Figure 11 indicates that the parametric models perform similarly to the non-parametric ones when the expected power is its medium range. However, close to the generation bounds the parametric densities perform worse than the non-parametric ones. This can be explained by the fact that the closer to the bounds, the more significant the censoring effect in the parametric densities become. This leads to higher bias in parameter estimates. The drop in operation quality of the parametric predictive densities is especially evident when the level of the expected power is close to the nominal capacity. This is in line with the discussion given in Section 5.3 on the down-regulation policy.



(a) Horizon=1 hour



(b) Horizon=4 hour



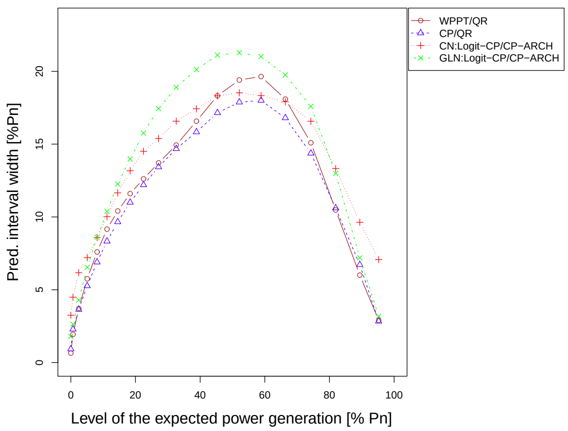
(c) Horizon=8 hour

Figure 8: CRPS evaluated over 5 equally sized subsets of the validation data set. P_n denotes the nominal capacity.

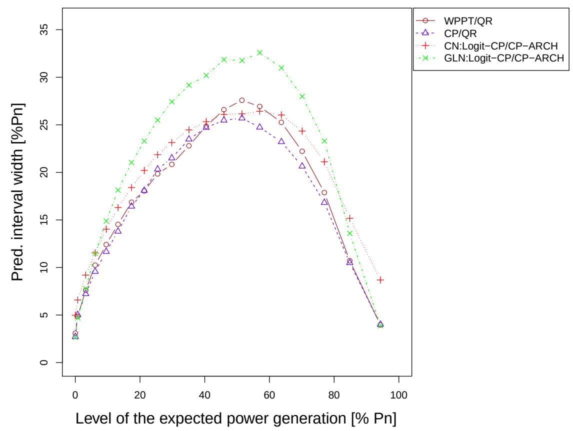
5.4.4 Accuracy and skill as functions of the measured power generation

.

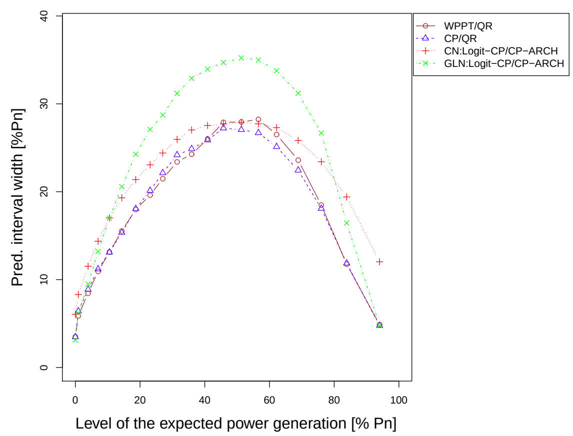
Figures 12 and 13 depict the accuracy and skill scores as functions of the observed power generation. From those figures one can conclude that the parametric densities managed to describe the situations when the observed power was in its medium range better than the non-parametric ones. In the situations when the power measurements fell close to the generation bounds, the non-parametric models performed much better than the parametric densities. This once again indicates that the non-parametric densities outperform the parametric ones mainly in the situations when the observed power approaches 0 or the nominal capacity. This is the region where the parametric densities suffer the most from the approximations used in parameter estimation methods and the power regulation policies.



(a) Horizon=1 hour

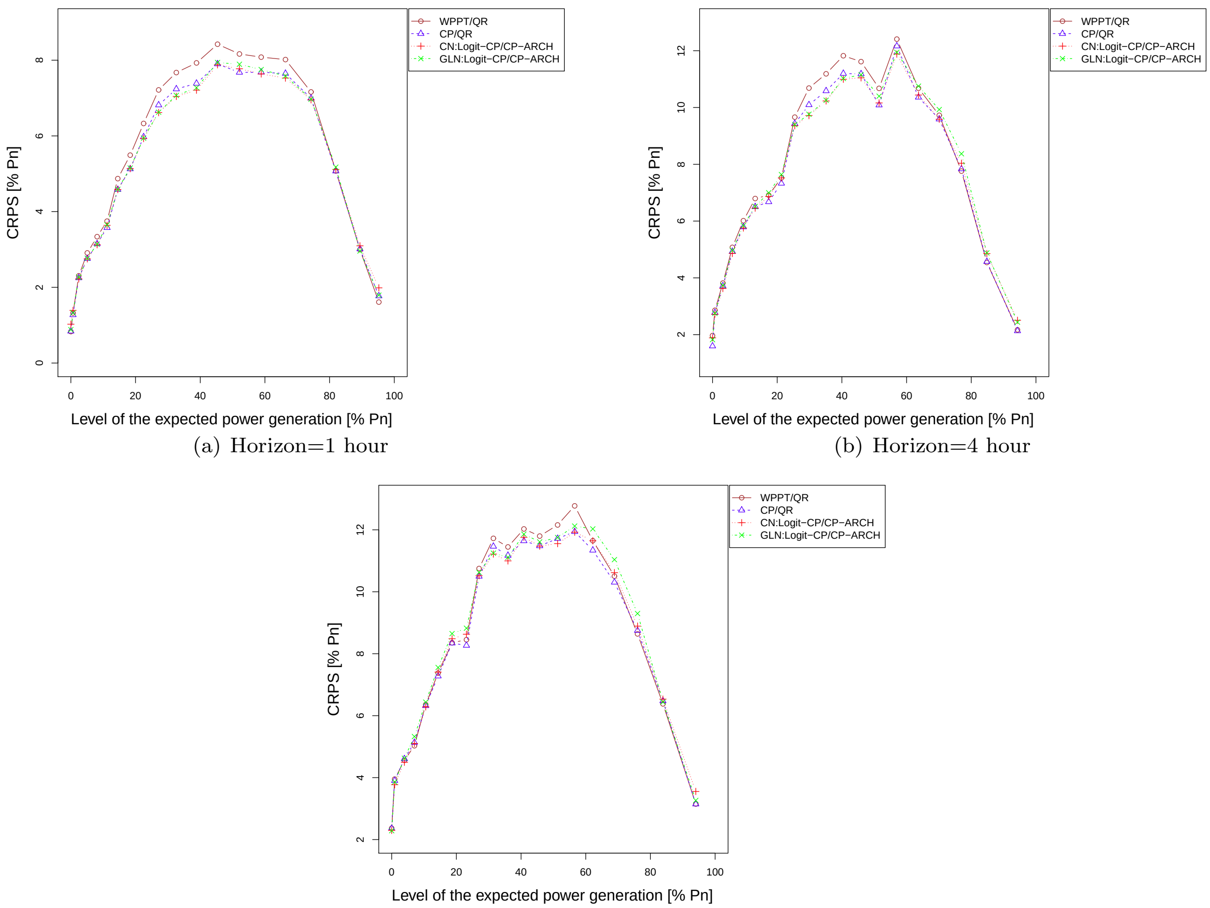


(b) Horizon=4 hour



(c) Horizon=8 hour

Figure 9: Sharpness is given by the width of the prediction intervals with the nominal coverage of 50 %. It is evaluated as a function of the expected power generation level. Levels of the point predictions represent 20 equally populated classes based on the quantiles of the CP point forecasts. P_n denotes the nominal capacity.

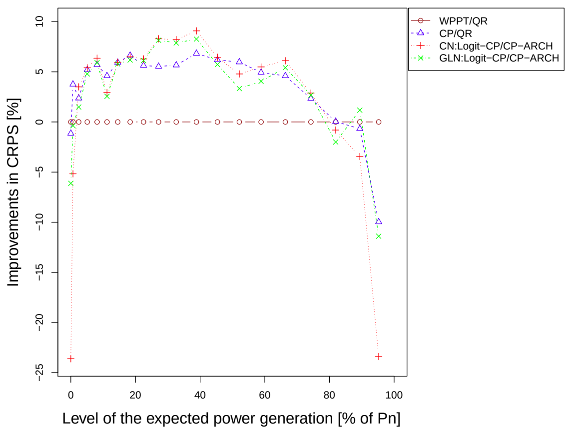


(a) Horizon=1 hour

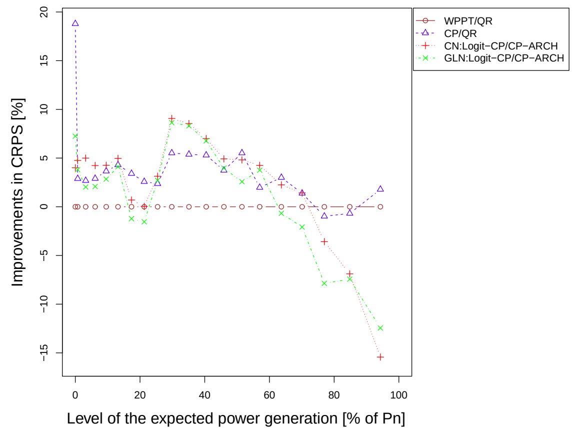
(b) Horizon=4 hour

(c) Horizon=8 hour

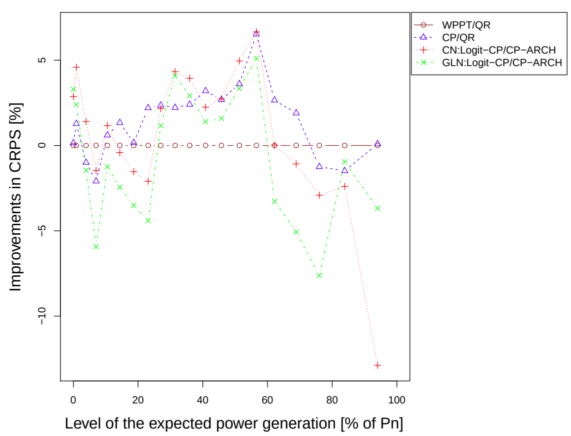
Figure 10: CRPS evaluated as a function of the expected power generation level. Levels of the point predictions represent 20 equally populated classes based on the quantiles of the CP point forecasts. P_n denotes the nominal capacity.



(a) Horizon=1 hour

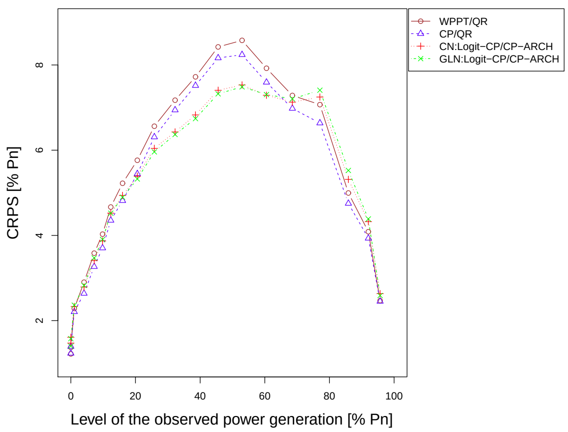


(b) Horizon=4 hour

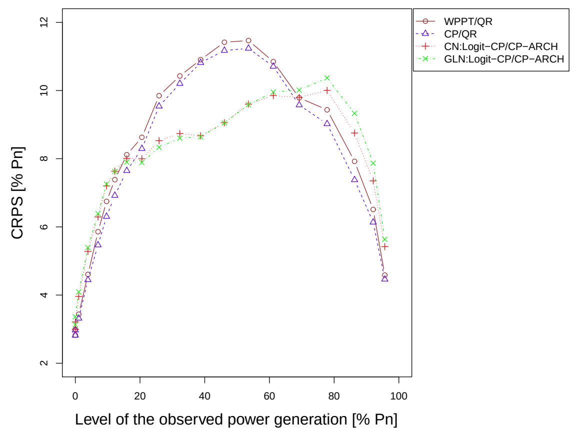


(c) Horizon=8 hour

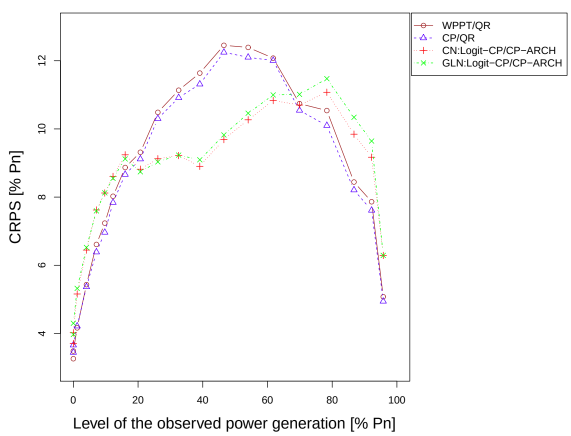
Figure 11: Relative improvements in CRPS when compared to the benchmark model WPPT/QR. Evaluation is conditional to the expected power generation level. Levels of the point predictions represent 20 equally populated classes based on the quantiles of the CP point forecasts. P_n denotes the nominal capacity.



(a) Horizon=1 hour

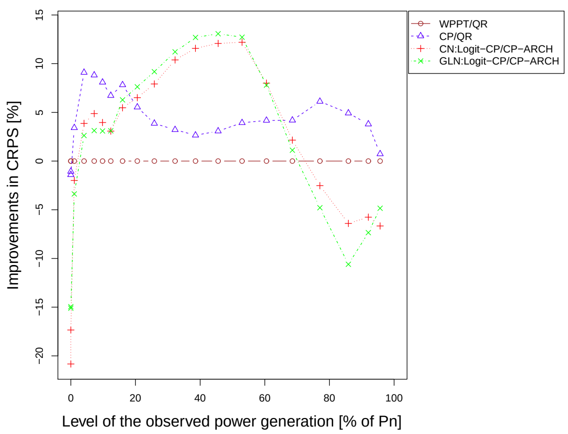


(b) Horizon=4 hour

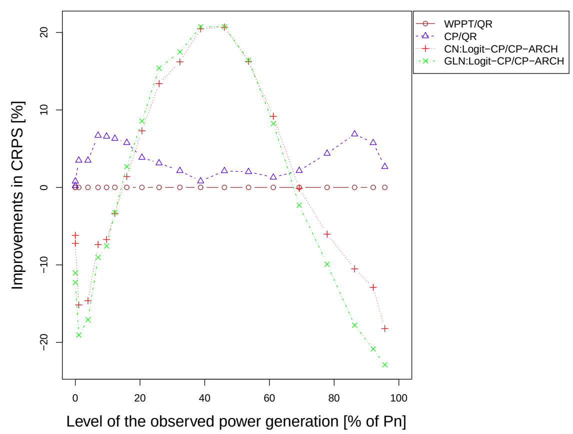


(c) Horizon=8 hour

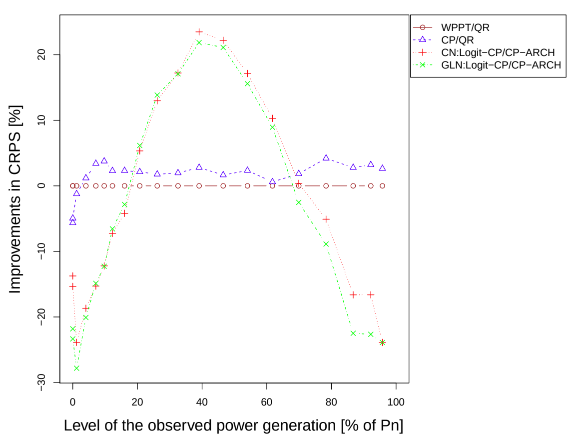
Figure 12: CRPS evaluated as a function of the power measurements. Levels of the power measurements represent 20 equally populated classes based on the quantiles of the CP point forecasts. P_n denotes the nominal capacity.



(a) Horizon=1 hour



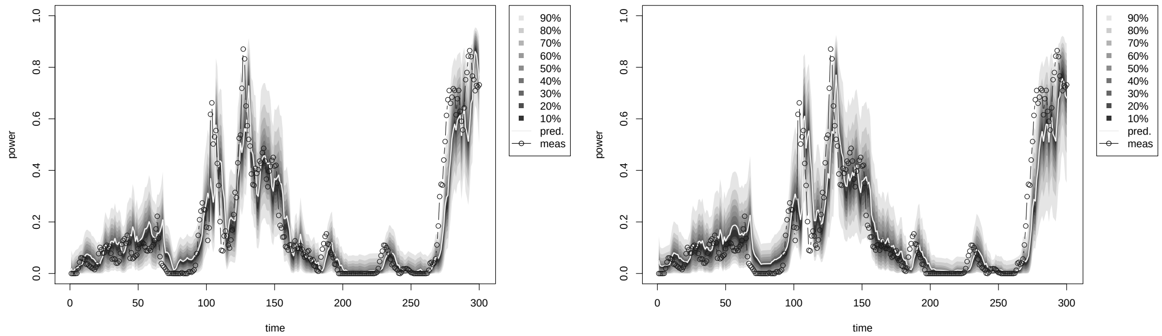
(b) Horizon=4 hour



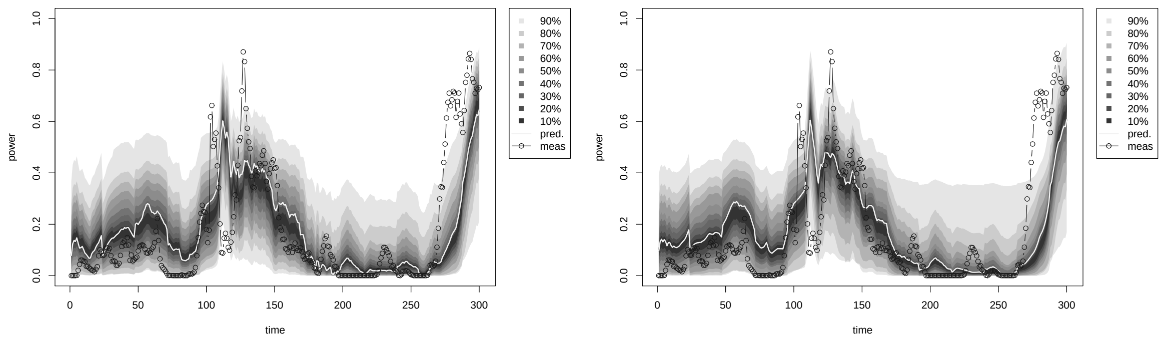
(c) Horizon=8 hour

Figure 13: Relative improvements in CRPS when compared to the benchmark model WPPT/QR. Evaluation is conditional to the expected power generation level. Levels of the power measurements represent 20 equally populated classes based on the quantiles of power observations. P_n denotes the nominal capacity.

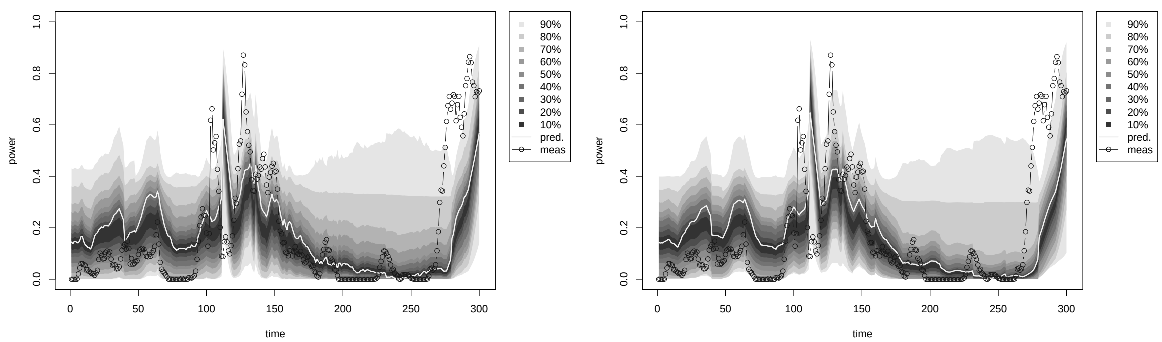
5.5 Demonstration of the operation of the best-performing model



(a) Horizon=1 hour



(b) Horizon=4 hour



(c) Horizon=8 hour

Figure 14: Example of the predictive densities for the wind power generation at Nysted Offshore given by the CP/QR (left) and the WPPT/QR (right) models. Considered period is from 10:15 on the 27th of July, 2009 to 13:45 on the 30th of July, 2009 (corresponding to 300 time steps of 15 min). Graphically the difference in predictive densities seems subtle and mainly related to the difference in the means of the predictive densities.

Figure 14 gives an example of probabilistic predictions given by the CP/QR and the benchmark approach. In general one can see that the difference in the distributions seems rather subtle and mainly related to the expectation (mass center) of the predictive densities. Figure 15 demonstrates an example of how the CP/QR model could be used in practice- at a given time probabilistic predictions could be issued for up to agreed amount of hours ahead. In this work we consider the predictions up to 8 hours ahead.

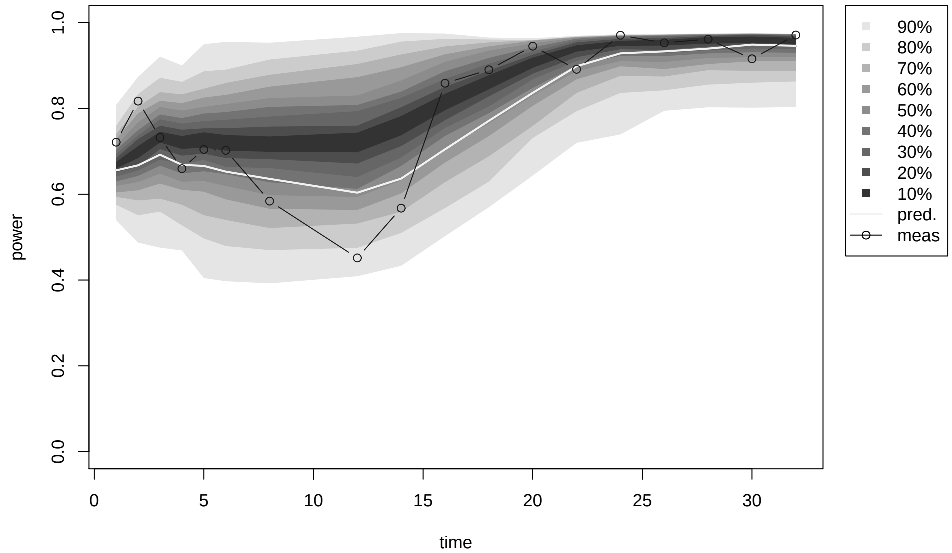


Figure 15: Example of probabilistic forecasts of wind power generation obtained with the CP/QR model. Predictive density is given by the prediction intervals of the different nominal coverage rates. Power values are standardized by the nominal capacity of the wind farm. Predictions for different lead times (from 15 min to 8 hours ahead) are issued at 17:15 on the 24th of November, 2009. A solid line shows the point predictions.

6 Conclusions

Focus has been given to probabilistic forecasts for Nysted Offshore. Both parametric and non-parametric methods for building the predictive densities have been considered. Following the parametric approach two different densities have been proposed for modelling the wind power generation: the Censored Normal and the generalized logit-Normal. The non-parametric approach was based on the time-adaptive quantile regression. The results show that using the spatially corrected point forecasts instead of the original WPPT predictions as input to building the predictive densities improves the performance of the models in both parametric and non-parametric approaches. It translates to saying that the spatial correction of the first order moments improves the quality of the corresponding predictive densities. The spatial correction of the higher order moments was shown not to ameliorate the quality of the predictive densities any further.

It has been shown in the work that the densities of the wind power generation are not constant. Two different methods for modelling the dynamics of the predictive densities have been proposed. Firstly, it has been noted that the distribution of the wind power generation is related to the level of the expected power. Capturing this effect was shown to improve the performance of the corresponding forecasts. In the second place, it has been decided to model the changes in the predictive densities implicitly, ı.e. assume that the dynamics of the system is governed by some unobservable process rather than by any particular explanatory variable. This was implemented using the Markov-switching models. It has been shown that the direct approach with conditioning the density dynamics of the level of the expected power outperforms the Markov switching models. This indicates that the dynamics of the predictive densities can be rather well explained by the changes in the expected level of the power generation. Thus accounting for it directly in the modelling procedure is more beneficial than trying to capture the underlying complex phenomenon in a probabilistic framework.

In this work the two types of the proposed parametric densities have been compared. It is shown that the censored Normal distribution describes the data better than the generalized logit-Normal. This holds for all the considered prediction horizons.

The parametric and the non-parametric probabilistic forecasts have also been compared. It is shown that both approaches perform similarly (in terms of the average accuracy) in the short prediction horizons (up to 5-6 hours ahead). In the longer horizons the difference in performance becomes more significant with the quantile-regression based models taking the leading position. Even in the shorter horizons, even though the two approaches show similar results in terms of accuracy, the corresponding densities are different. The parametric densities are shown to be less sharp than the non-parametric ones. The analysis of the situation-based performance of the two approaches has been carried out. It is shown that the parametric densities outperform the non-parametric ones in the periods when then the power generation is in its medium range. Closer to the bounds, the operation quality of the parametric models is poor compared to the quantile regression based approach.

Summarizing, based on the overall operation quality, the best performing model is the CP/QR. This is a probabilistic forecast based on the adaptive quantile regression using the spatially corrected CP point predictions as input. The model consistently outperforms the benchmark approach in all the considered horizons. The relative improvements in overall quality(compared to the benchmark approach WPPT/QR) are ranging from 1.5% to 8.29% depending on the prediction horizon.

References

[1] DTU Informatics/Report 1. Spatio-temporal analysis and correction of wind power forecasts for DONG wind farms. 2010

[2] ENFOR/04EKS0017A002-A. Spatio-temporal analysis and correction targeting Nysted Offshore. 2011

[3] DTU Informatics/Report 2. Spatio-temporal analysis and correction targeting Nysted Offshore. 2011

[4] ENFOR/04EKS0017A001-A[2]. Spatio-temporal analysis and correction of wind power forecasts for DONG wind farms. 2010

[5] M. R. Berthold and F. Hoppner. On clustering time series using euclidean distance and Pearson correlation. Technical report, University of Konstanz, 2008.

[6] Bollerslev, T. Generalized AutoRegressive Conditional Heteroscedasticity. Journal of Econometrics, 1986, 22, 43-56.

[7] Gneiting T., Balabdaoui F., Raftery A. E. Probabilistic forecasts, calibration and sharpness, J. Royal Stat. Soc. A, 2009

[8] Gneiting, T., Larson, K., Westrick, K, Genton, M. G. and Aldrich, E. Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method. Journal of the American Statistical Association, 2006, 101, 968-979.

[9] Gneiting T., Raftery A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. 2007, Vol. 102, Issue 477, pp. 359

[10] Hamill, T. M. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. Monthly Weather Review, 2001: 129, 550-560.

[11] Hering, A. S., Genton, M. Powering Up with Space-Time Wind Forecasting. Journal of the American Statistical Association, 2010, Vol. 105.

[12] Koenker, R., Basset Jr., G., Regression quantile. Econometrica 1978. 46 (1), 33-50

[13] Lau, A. Probabilistic Wind Power Forecasts: From Aggregated Approach to Spatiotemporal Models. Mathematical Institute University of Oxford. A thesis submitted for the degree of Doctor of Philosophy 2011

[14] Møller, J. K., Nielsen H.,Aa., Madsen, H. Time-adaptive quantile regression. Computational Statistics and Data Analysis, 2008, 51, 1292-1303

[15] Madsen, H. Time Series Analysis. Chapman & Hall 2008

[16] J. P. Marques de Sa. Pattern Recognition: Concepts, Methods ans Applications. Berlin: Springer Verlag, 2001.

[17] Pinson, P. and Kariniotakis, G. Conditional prediction intervals of wind power generation. IEEE Transactions on Power Systems. 2010. 25(4), pp. 1845-1856.

[18] Pinson, P. and Madsen, H. Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models. Journal of Forecasting. 2010.

[19] Pinson, P., Nielsen, H. A., Møller, J. K., Madsen, H. and Kariniotakis, G. N. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. Wind Energy, 2007, 10, 497-516

[20] P. Pinson. Very short-term probabilistic forecasting of wind power with generalised logit-Normal distributions. Journal of the Royal Statistical Society, Series C. available online. 2012.

[21] N. Siebert. Development of methods for regional wind power forecasting. PhD thesis, Mines Paris, 2008

[22] Trombe, P.J., Pinson, P., Madsen, H. A general probabilistic forecasting framework for offshore wind power fluctuations. Submitted to Energies, 2012

[23] Jónsson, T., Pinson P. Density Forecasting of Day-Ahead Electricity Prices. working paper

[24] Zucchini, W., MacDonald, I. L. Hidden Markov Models for Time Series. An Introduction Using R. Monographs on Statistics and Applied Probability 110. Chapman & Hall 2009