



## On probabilistic forecasting of wind power time-series

**Pinson, Pierre**

*Publication date:*  
2010

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Pinson, P. (2010). *On probabilistic forecasting of wind power time-series*. Technical University of Denmark, DTU Informatics, Building 321. IMM-Technical Report-2010-06

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# On probabilistic forecasting of wind power time-series

Pierre Pinson<sup>1,2</sup>

<sup>1</sup> European Centre for Medium-range Weather Forecasts, Reading, United Kingdom

<sup>2</sup> DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark

---

## Abstract

Wind power generation is a nonlinear and bounded variable, partly owing to the power curve that converts wind to electric power, and partly owing to the very stochastic nature of wind itself. Predictive densities of wind power generation should account for that effect. Such densities are clearly not expected to be Gaussian, with higher order moments being directly related to their expectation and potentially to some external signal. It is proposed here to model such predictive densities with discrete-continuous mixtures of generalized logit-normal distributions and of probability masses at the bounds. In this framework, a transformation is employed so that transformed data can be modeled with censored Normal variables. Two types of models are proposed: a simple autoregressive model and a more advanced conditional parametric autoregressive model, for which wind direction is the variable conditioning the wind power dynamics. In both cases, the model parameters are adaptively and recursively estimated, time-adaptativity being the result of exponential forgetting of past observations. The probabilistic forecasting methodology is applied at the Horns Rev wind farm in Denmark, for 10-minute ahead probabilistic forecasting of wind power generation. Probabilistic forecasts generated from the proposed methodology clearly have higher skill than those obtained from a classical Gaussian assumption about wind power predictive densities. Corresponding point forecasts also exhibit significantly lower error criteria.

**Keywords:** bounded time-series, logit transformation, mixture model, probabilistic forecasting, wind power

---

\* *Corresponding author:*

P. Pinson, European Centre for Medium-range Weather Forecasts,  
Shinfield Park, Shinfield, Berkshire, RG2 9AX, United Kingdom.

Tel: +44 118 949 9662, Fax: +44 118 986 9450, email: pierre.pinson@ecmwf.int, webpage: [www.imm.dtu.dk/~pp](http://www.imm.dtu.dk/~pp)

# 1 Introduction

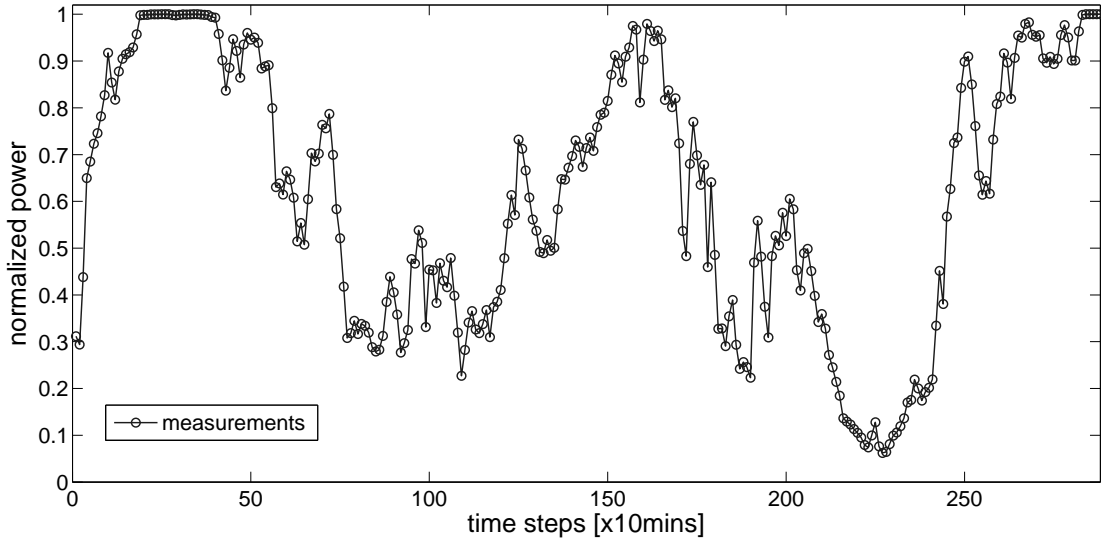
Over the past few decades, one of the major breakthroughs in forecasting meteorological variables for applications such as weather derivatives and renewable energy generation, comes from the transition from point to probabilistic forecasting<sup>1</sup>. This phenomenon is not only observed in the meteorological literature, as probabilistic forecasts are also becoming customary products in economics and finance<sup>2–4</sup> or more generally in management sciences. For the case of continuous variables e.g. wind power, they optimally take the form of predictive densities — equivalently referred to as predictive distributions or density forecasts, giving the full probability density function of the random variable of interest for a set of look-ahead times. It has been demonstrated that the optimal management and trading of wind power calls for probabilistic forecasts<sup>5</sup>. This actually follows from a more general result which is that for a large class of decision-making problems, optimal decisions directly relate to quantiles of conditional predictive distributions, as discussed by Gneiting<sup>6</sup> for instance.

When forecast horizons are further than a few hours ahead, predictive densities of wind power generation may be issued using either dynamic post-processing of ensemble forecasts of meteorological variables<sup>7,8</sup>, or statistical models generating probabilistic forecasts based on point predictions, and conditional to the predicted weather condition<sup>9,10,21</sup>. Forecast horizons in the range of hours to days have until recently been those attracting the most interest, since corresponding to the decision-making needs when participating in European day-ahead electricity markets<sup>12,13</sup>. In parallel today, it is recognized that wind power producers and traders need shorter-term forecasts as the gate closure of electricity markets is shortening or rolling intra-day markets appearing. On the more technical side, the wind farm controllers themselves can be optimal only if relying on model predictive control methodologies, with temporal resolution of a few minutes<sup>14</sup>. And if considering the Transmission System Operator (TSO), the significant penetration of wind generation in the electricity mix makes that short-term forecasts (with horizons of up to 15 minutes) are essential for the optimal operation of reserves employed, in order to achieve an economic and safe operation of the electricity network<sup>15</sup>.

For such short look-ahead times, purely statistical models taking advantage of meteorological and physical expertise on the problem at hand are to be preferred. Examples of recent works taking that direction, focusing on the wind (speed and potentially direction) variable, include the regime-switching space-time method introduced by Gneiting *et al.*<sup>16</sup>, or the multivariate wind vector models of Hering and Genton<sup>17</sup>. Some other works have been focusing on capturing changes in the wind field dynamics using e.g. Markov-switching approaches<sup>18</sup>. In parallel, Pinson and Madsen<sup>19</sup> discuss the benefits of Markov-switching autoregressive models for point and probabilistic forecasting of wind power generation at look-ahead times of 10-minutes ahead.

Of the wind power variable characteristics, certainly the most crucial one is that this variable is nonlinear and bounded. The power curve converting wind to electrical power indeed takes the form of a sigmoid with minimum production zero, and maximum one the nominal capacity (denoted  $P_n$ ) of the wind turbine, wind farm, or wind energy portfolio considered. In the following, wind power measurements and forecasts are normalized by  $P_n$ , and we thus say that wind power generation takes value in the unit interval  $[0, 1]$ . Whatever the temporal resolution of wind power time-series, or the forecast horizon considered, a consequence of the above is that conditional predictive densities cannot be Gaussian, and also that their higher-order moments may be directly related to their mean, and potentially to some external signal. These two points have already been discussed in a

number of publications<sup>20–22</sup>. To our knowledge however, there exists no work in the literature where these aspects have been appropriately taken into consideration. Another important characteristic of the wind power generation process is that similarly to other meteorology-related processes, its dynamics are nonstationary in the sense that they smoothly vary over time, at the time scales of months, seasons and years<sup>16,19,23,24</sup>, with features that cannot be easily accounted for in the models themselves in a parametric framework. This can instead be handled in a nonparametric fashion with e.g. sliding estimation windows or exponential forgetting schemes. For illustrative purposes, Figure 1 depicts an episode with 2 days of successive 10-minute average power output at the Horns Rev offshore wind farm which will be the focus of the application and results part of the paper.



**FIGURE 1:** Episode of 2 days (288 time-steps, with a temporal resolution of 10 minutes) with wind power measurements at the Horns Rev wind farm in Denmark.

In order to account for these various characteristics, we propose here a methodology for the probabilistic forecasting of wind power time-series. Predictive densities take the form of discrete-continuous mixtures consisting of a generalized logit-normal distribution with potential concentration of probability mass at the bounds of the unit interval  $[0, 1]$ . These predictive densities are fully characterized by a location and a scale parameter only. The generalized logit-normal distribution is originally introduced in the present paper, as a generalization of the logit-normal distribution extensively described and discussed by Aitchison and Shen<sup>25</sup> for instance. Such distribution directly relates to a transformation of original wind power data through a generalized logit transform. Both the transformation and the distribution are introduced in Section 2, while their related properties are discussed. We claim that such family of transformations and corresponding distributions should be considered as a generic feature in the modeling and forecasting of wind power generation, whatever the temporal resolution and look-ahead times considered. In the modeling approach introduced in Section 3, the transformed variable can be modeled with censored Gaussian distributions. Both model building and estimation aspects are subsequently dealt with, with focus on simple autoregressive (AR) models or more advanced conditional parametric autoregressive (CP-AR) models. These models are considered for the sake of example, and could be further improved in the future, accounting for e.g. regime-switching aspects or additional seasonalities. The methodology is then applied in Section 4 for the probabilistic forecasting of wind power generation at the

Horns Rev wind farm in Denmark, with look-ahead times of 10-minutes ahead. The paper ends with concluding remarks in Section 5.

## 2 The Generalized Logit transformation and the Generalized Logit-Normal distribution

Probably the most employed general class of transformations is that introduced by Box and Cox<sup>26</sup>, commonly known as Box-Cox or power transformations. For an original time-series  $\{x_t\}$ , its transform is given by applying

$$y_t = \begin{cases} (x_t^\nu - 1)/\nu, & \nu \neq 0 \\ \log(x_t), & \nu = 0 \end{cases} \quad (1)$$

to each element  $x_t$  of the original time-series.  $\nu$  is commonly taken equal to 0, corresponding to a logarithmic transformation, or estimated from the data in a maximum-likelihood framework. The aim of employing such transformation is to have  $\{y_t\}$  Gaussian, or at least “more Gaussian” than the original time-series  $\{x_t\}$ . It is indeed the case that one then works in a Gaussian framework for the modeling and forecasting of the transformed time-series  $\{y_t\}$ , even though this assumption may still not be completely correct<sup>27</sup>. It is certainly more correct though than employing a Gaussian framework directly on the original time-series  $\{x_t\}$ . In the following, we write  $\{X_t\}$  and  $\{Y_t\}$  the stochastic processes whose realizations define the time-series  $\{x_t\}$  and  $\{y_t\}$ .

The logistic (also called logit) transformation has originally been suggested by Johnson<sup>28</sup>, and consequently used e.g. for bounded economic variables<sup>29</sup>. Such transformation relates to the idea of employing logit-normal distributions for modeling the original effect. Indeed if the transformed variable is Gaussian, the original variable is then distributed logit-normal. Logit-normal distributions have been introduced and thoroughly examined by Aitchison and Shen<sup>25</sup>. Wallis<sup>29</sup> actually goes further, and argues that logistic transformations should be used as a general transformation even if the original time-series considered are not bounded. Both power and logistic transformations are part of the more general family of transformations described and discussed by Aranda-Ordaz<sup>30</sup>. Similarly, Aitchison and Shen<sup>25</sup> raises the question of the necessity of widening the class of logit-normal distributions. It will be explained below how the wind power modeling and forecasting applications actually call for a generalization of the logit-normal distribution, and thus also a generalization of the logit transform.

Considering the fact that wind power is nonlinear and bounded, the logistic transformation is obviously appealing. It is known however that the sigmoid corresponding to the power curve commonly has a different shape of its inflection for low and high power values. In such a case it appears more appropriate to employ a generalized version of the logistic function allowing for asymmetry, inspired by Richards<sup>31</sup>. We propose the name of *generalized logit* (GL) transformation, and denote that transformation by  $\gamma$ . Similarly to the above, write  $\{x_t\}$  the original time-series, and  $\{y_t\}$  its transform. The GL transformation  $\gamma$  is defined by

$$y_t = \gamma(x_t; \nu) = \ln \left( \frac{x_t^\nu}{1 - x_t^\nu} \right), \quad x_t \in (0, 1) \quad (2)$$

while its inverse transformation, referred to as inverse generalized logit (IGL), is defined as

$$x_t = \gamma^{-1}(y_t; \nu) = \{1 + \exp(y_t)\}^{-1/\nu}, \quad y_t \in \mathbb{R} \quad (3)$$

The GL transformation generalizes the logistic one by just adding one shape parameter  $\nu$ , the interest of which will be detailed below when discussing general properties of the GL transform. Roughly speaking, it permits to account for the fact that the nonlinearities and bound effects may not be symmetric. For  $\nu = 1$ , one retrieves the more classical logit transformation. Note that the GL transformation can be employed whatever the bounds of the closed intervals, as the original time-series could initially be rescaled in order to take values in  $[0, 1]$  instead. Equivalently, one could also use the most general form of the Richard's function<sup>31</sup>, which allows for any bound for the closed interval.

Our motivation for introducing the GL transformation for nonlinear and bounded variables such as wind power generation comes from the results reported in various analyses of statistical distributions of wind power forecast errors. Under the rationale of minimization of a quadratic loss function, a point forecast for  $X_{t+k}$  issued at time  $t$  is defined as the conditional expectation  $\mu_{X,t+k}$  of the random variable  $X_{t+k}$  given the information set at time  $t$ . Several authors, e.g. <sup>20</sup> and <sup>21</sup>, have observed that the standard deviation  $\sigma_{X,t+k}$  of the distributions of forecast errors was directly linked to its conditional expectation  $\mu_{X,t+k}$ , and this whatever the time  $t$ , the forecast horizon  $k$  or the forecast model employed. This hence indicates that in a general manner, appropriate predictive densities of wind power generation should account for and reproduce this observed feature. Recently, some models for this relationship have been proposed<sup>8,22,32</sup>. It is actually that the general form of the model linking  $\mu_{X,t+k}$  and  $\sigma_{X,t+k}$  may be written as

$$\sigma_{X,t+k} = \alpha \mu_{X,t+k}^\nu (1 - \mu_{X,t+k}^\nu) \quad (4)$$

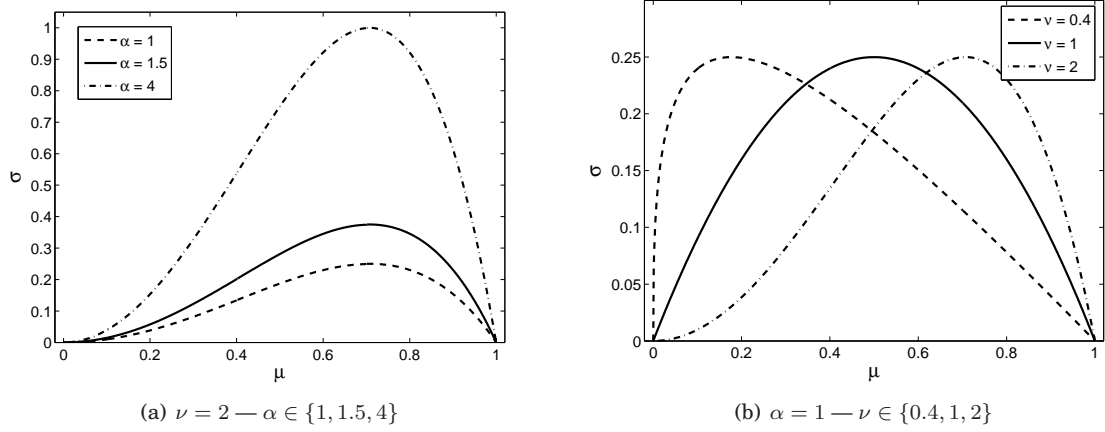
One observes from (4) that we employ a scale and shape parameter denoted by  $\alpha$  and  $\nu$ , respectively, where  $\nu$  already appears in the definition of the GL transformation (2). The reason why we do so will be detailed below. First, let us illustrate in Figure 2 how this model linking  $\mu_{X,t+k}$  and  $\sigma_{X,t+k}$  evolves as one varies the shape and scale parameters  $\nu$  and  $\alpha$ . In Figure 2(a), the shape parameter is set to  $\nu = 2$ , and the scale parameter  $\alpha$  takes values in  $\{1, 1.5, 4\}$ . The model has the same shape and is simply linearly scaled. In contrast in Figure 2(b), it is then the scale parameter that is fixed to  $\alpha = 1$ , and the shape parameter taken in  $\{0.4, 1, 2\}$ . This parameter controls the asymmetry of the model instead, being positively skewed for  $\nu < 1$  and negatively skewed for  $\nu > 1$ .

Given the variance  $\sigma_{X,t+k}^2$  of the random variable  $X_{t+k}$  of interest, that of its transform  $Y_{t+k} = \tilde{\gamma}(X_{t+k})$ , denoted by  $\sigma_{Y,t+k}^2$ , can be approximated by

$$\sigma_{Y,t+k}^2 \simeq \frac{d\tilde{\gamma}}{dx}(\mu_{X,t+k}) \sigma_{X,t+k}^2 = (\mu_{X,t+k}^\nu (1 - \mu_{X,t+k}^\nu))^{-2} \sigma_{X,t+k}^2 \quad (5)$$

as long as  $\sigma_{X,t+k}^2$  is small, which is the case for predictive densities of wind power generation at horizons of few minutes to few-hours ahead.

It follows from (5) that if (4) is the true model that links the standard deviation  $\sigma_{X,t+k}$  and the conditional expectation  $\mu_{X,t+k}$ , then employing the GL transform with the values of  $\alpha$  and  $\nu$  in (4) makes that the corresponding predictive distribution for  $Y_{t+k}$ , issued at time  $t$  and for lead time  $t+k$ , has a variance that can be approximated by  $\alpha^2$ , thus explaining the appellation of scale parameter. This



**FIGURE 2:** The influence of the scale ( $\alpha$ ) and shape ( $\nu$ ) parameters on the relationship between mean and standard deviation of predictive densities.

result is independent of  $t$ ,  $k$ , or of any covariate. This scale parameter  $\alpha$  may, however, be a function of  $t$ ,  $k$  and of some covariate, since it is expected that the characteristics of predictive distributions will change with the look-ahead time and the prevailing weather conditions. It may also be that  $\nu$  is a function of  $t$  if it is thought that the transformation will be nonstationary. This would complicate the problem in a somehow unnecessary manner, and is hence not done here. Consequently if not considering the higher moments of  $Y_{t+k}$ , one may propose a second order representation of  $Y_{t+k}$  as  $Y_{t+k} \sim \mathcal{N}(\mu_{Y,t+k}, \sigma_{Y,t+k}^2)$ , hence being a function of a location parameter  $\mu_{Y,t+k}$  and of a scale parameter  $\sigma_{Y,t+k}$ . Similarly to employing a Box-Cox transformation, it may be that a Gaussian assumption for predictive densities for  $Y_{t+k}$  is not a perfect assumption when applied to real-world data, though it may have less consequences than making an inappropriate Gaussian assumption for predictive densities of the original  $X_{t+k}$ . This will be illustrated and discussed when presenting application results in Section 4.

In a general manner, even if (4) is not the true model linking mean and standard deviation of wind power predictive densities, if there is additional conditional heteroskedasticity, or an effect of some covariate, the GL transformation introduced here will permit to treat the transformed time-series  $\{y_t\}$  in a simpler framework than if directly working on modeling and forecasting the original  $\{x_t\}$ . We introduce the *generalized logit-normal distribution*, denoted by  $L_\nu(\mu, \sigma^2)$  and referred to as GL-Normal distribution, where  $\mu$  and  $\sigma$  are the location and scale parameter of the corresponding Normal variable after transformation through the GL-transformation with shape parameter  $\nu$ . In the present case, considering that predictive densities for the original effect  $X_{t+k}$  can be modeled with GL-Normal distributions then translates to having predictive densities for  $Y_{t+k}$  modeled with Gaussian distributions. Generalizing the expression of the probability density function of a univariate logit-normal variable as expressed in by Aitchison and Shen<sup>25</sup> or Frederic *et al.*<sup>33</sup>, that for the GL-Normal variable writes

$$f(x) = \frac{1}{\sigma \sqrt{2\pi} x^\nu (1-x^\nu)} \exp \left\{ -\frac{1}{2} \left[ \frac{\gamma(x; \nu) - \mu}{\sigma} \right]^2 \right\}, \quad x \in (0, 1) \quad (6)$$

We do not discuss here the moments of a variable that would be distributed GL-Normal. One could



carry out theoretical developments in the spirit of Aitchison and Shen<sup>25</sup> to derive some properties and perform comparisons with the class of logit-normal and Dirichlet distributions. It could also be envisaged to perform similar numerical developments as in Frederic *et al.*<sup>33</sup> in order to visualize the moments of GL-Normal distributions as a function of their location and scale parameters. Note finally that the question of the bounds has been left aside here, but will be dealt with below instead when introducing the modeling and estimation approaches.

### 3 Modeling and estimation approaches

Our main objective is to show the advantages of modeling and forecasting wind power time-series with GL-Normal variables, since permitting to model complex bounded distributions while allowing to project ourselves in a Gaussian modeling framework. As a consequence, the developments gathered below concentrate on fairly simple dynamic models only. They could however be extended to account for additional effects e.g. conditional heteroskedasticity or regime-switching dynamics, though it would clearly call for modification of the estimation methods described. GL-Normal variables can also be employed for modeling predictive densities for further look-ahead times (i.e. from few hours to few days ahead), for which appropriate statistical approaches will account for nonlinear transformation of the dynamic information provided by meteorological predictions.

To give a brief sketch of our approach, we propose to model predictive densities of wind power generation as discrete-continuous mixtures of GL-Normal variables and of probability masses at the bounds of the unit interval. Autoregressive (AR) and conditional parametric autoregressive (CP-AR) models are proposed for the location parameter of the GL-Normal variables. Their scale parameter is seen as unconditional in the former case, or as a nonlinear function of the covariate conditioning the AR dynamics in the latter one. Only AR types of dynamics are considered, as earlier results on similar offshore data suggest that such dynamics for the mean are appropriate<sup>19</sup>. Since we have insisted on the fact dynamic models for wind power generation should be seen as slowly varying, it is proposed here to adaptively and recursively track model coefficients, where adaptivity is based on exponential forgetting of past observations.

#### 3.1 The general form of predictive densities

For the sake of simplicity, some might argue that the assumption such that wind power generation never takes the value of 0 or 1 is appropriate, so that this variable may be directly seen as being GL-Normal distributed. In practice however, already owing to the finite precision and rounding of measurement devices it is clear that 0 and 1 values will be present in collected data. In addition, since wind power generation is null for wind speed values lower than 3-4 m.s<sup>-1</sup> (the cut-in wind speed) and equal to nominal capacity for wind speed values above 14-18 m.s<sup>-1</sup> (the rated wind speed), it is expected that a fair share of genuine 0 and 1 values will be observed. Note that the values mentioned for cut-in and rated wind speeds are not general values and may vary depending on the turbines and wind power portfolio considered. The phenomenon such that wind power generation has significant observed frequencies of zero and maximum production is general though. This also implies that the  $\nu$ -parameter of the GL transformation will be dependent on the case-study considered, and should hence be selected based on available data.

Based on the point made above, it is fair to write the predictive density for the wind power genera-



tion  $X_{t+k}$  at time  $t+k$  as

$$X_{t+k} \sim \omega_{t+k}^0 \delta_0 + (1 - \omega_{t+k}^0 - \omega_{t+k}^1) L_\nu(\mu_{t+k}, \sigma_{t+k}^2) + \omega_{t+k}^1 \delta_1 \quad (7)$$

with  $\delta_0$  and  $\delta_1$  dirac functions at 0 and 1, respectively, representing the potential concentration of probability mass at the bounds of the unit interval. To be consistent, one has

$$\omega_{t+k}^0, \omega_{t+k}^1 \in [0, 1], \quad \omega_{t+k}^0 + \omega_{t+k}^1 \in [0, 1]$$

Equivalently if considering  $Y_{t+k}$  the GL-transform of  $X_{t+k}$ , the form of its predictive density is given by

$$Y_{t+k} \sim \omega_{t+k}^0 \delta_{-\infty} + (1 - \omega_{t+k}^0 - \omega_{t+k}^1) \mathcal{N}(\mu_{t+k}, \sigma_{t+k}^2) + \omega_{t+k}^1 \delta_{+\infty} \quad (8)$$

Again owing to the finite precision of measurements and to rounding, it is not expected that the time-series  $\{y_t\}$  actually takes all possible values in  $\mathbb{R}$ . It is more the case that one can consider real bounds on the range of potential values for  $y_t$ , i.e.  $y_t \in \mathcal{D}_y = [\gamma(\epsilon; \nu), \gamma(1 - \epsilon; \nu)]$ , where  $\epsilon$  is in the order of measurements precision, say  $\epsilon \leq 10^{-2}$ . One can consequently rewrite (8) as

$$Y_{t+k} \sim \omega_{t+k}^0 \delta_{\gamma(\epsilon; \nu)} + \mathcal{N}(\mu_{t+k}, \sigma_{t+k}^2) \mathbf{1}_{\bar{\mathcal{D}}_y} + \omega_{t+k}^1 \delta_{\gamma(1-\epsilon; \nu)} \quad (9)$$

where  $\bar{\mathcal{D}}_y$  is the open of  $\mathcal{D}_y$  and  $\mathbf{1}_{\bar{\mathcal{D}}_y}$  the indicator function for this open interval, i.e.

$$\mathbf{1}_{\bar{\mathcal{D}}_y} = \begin{cases} 1, & y \in \bar{\mathcal{D}}_y \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

One notices that the  $(1 - \omega_{t+k}^0 - \omega_{t+k}^1)$  weight does not appear in (9). This is due to the fact that the predictive density for  $Y_{t+k}$  takes the form of a censored Normal instead. This allows us to straightforwardly derive the expressions for the weights  $\omega_{t+k}^0$  and  $\omega_{t+k}^1$ , which are indeed given by

$$\omega_{t+k}^0 = \Phi\left(\frac{\gamma(\epsilon; \nu) - \mu_{t+k}}{\sigma_{t+k}}\right), \quad \omega_{t+k}^1 = 1 - \Phi\left(\frac{\gamma(1 - \epsilon; \nu) - \mu_{t+k}}{\sigma_{t+k}}\right) \quad (11)$$

with  $\Phi$  the cumulative distribution function of a standard Normal variable. This formulation of predictive densities of wind power generation as a discrete-continuous mixture is fully characterized by the location and scale parameters  $\mu_{t+k}$  and  $\sigma_{t+k}$  only. Note that the above proposal is somehow similar to the coarsening approach proposed by Lesaffre *et al.* <sup>34</sup> for the modeling of outcome scores in  $[0, 1]$  using classical logit-normal distributions.

### 3.2 Autoregressive dynamics

In the case of AR dynamics, the model for the location parameter  $\mu_{t+k}$  writes

$$\mu_{t+k} = \boldsymbol{\theta}^\top \tilde{\mathbf{y}}_t \quad (12)$$

where

$$\boldsymbol{\theta} = [\theta_0 \ \theta_1 \ \dots \ \theta_l]^\top, \quad \tilde{\mathbf{y}}_t = [1 \ y_t \ \dots \ y_{t-l+1}]^\top$$

while the scale parameter is defined as a constant, i.e.

$$\sigma_{t+k}^2 = \beta \quad (13)$$

The maximum lag  $l$  can be determined after examining the autocorrelation and partial autocorrelation functions of the  $\{y_t\}$  time-series, by minimization of either Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), or finally through a cross-validation exercise.

The  $\{y_t\}$  time-series corresponds to realizations of censored Normal variables. It may therefore be argued that the vector of AR parameters  $\theta$  may be estimated with either Least-Squares (LS) or Maximum-Likelihood (ML) estimation techniques. In the case of censored Normal variables, the LS-estimator for the location parameter is certainly more subject to estimation bias than the ML-estimator. Experiments on wind power generation data has showed us that the LS-estimator of the location parameter was acceptable, while having the advantage of simplicity of implementation if compared with the ML-estimator. This is the reason why we only consider and describe the LS-estimator below.

We explained that the models employed for characterizing the wind power generation dynamics should be seen as slowly varying. This argument is supported by a number of publications e.g. <sup>16,17,19,24</sup>, and we thus do not develop on that aspect. For the case of AR dynamics for wind power generation, the parameters  $\theta$  can be adaptively and recursively estimated with a Recursive Least Squares (RLS) method<sup>35,36</sup>. In parallel, the scale parameter  $\sigma_{t+k}$  can also be easily tracked within this RLS framework. Recursive estimation offers the advantage that only the last available measurements are used at each time step for updating the model parameters. This contrasts with the more computationally intensive idea of employing a sliding window, as done by Gneiting *et al.* <sup>16</sup> or Hering and Genton <sup>17</sup>, for which a full estimation is performed on a bulk of data at each time step.

Imagine being at time  $t$ , and let us introduce the following time-dependent objective function to be minimized at that time,

$$S_t(\theta) = \sum_{i=1}^t \lambda^{t-i} \left( y_i - \theta^\top \tilde{\mathbf{y}}_{i-k} \right)^2 \quad (14)$$

where  $\lambda$  is the forgetting factor,  $\lambda \in (0, 1)$ , allowing for exponential forgetting of past observations, for which the corresponding effective number of observations is defined as  $n_\lambda = (1 - \lambda)^{-1}$ . The value of the forgetting factor is typically slightly below 1. The LS-estimate  $\hat{\theta}_t$  of the AR-dynamics parameter at time  $t$  is then defined as

$$\hat{\theta}_t = \arg \min_{\theta} S_t(\theta) \quad (15)$$

In a recursive framework, the formulae for updating the LS-estimate of  $\theta$  based on the newly available information at time  $t$  summarizes to

$$\hat{\mathbf{R}}_t = \lambda \hat{\mathbf{R}}_{t-1} + \tilde{\mathbf{y}}_{t-k} \tilde{\mathbf{y}}_{t-k}^\top \quad (16)$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \hat{\mathbf{R}}_t^{-1} \tilde{\mathbf{y}}_{t-k} \varepsilon_t \quad (17)$$

where  $\varepsilon_t$  is defined as  $\varepsilon_t = y_t - \theta_{t-1}^\top \tilde{\mathbf{y}}_{t-k}$ . For an extensive description of RLS estimation schemes, we refer to Ljung and Söderström<sup>35</sup> and Madsen<sup>36</sup>.

In parallel, the formula for updating the constant estimate  $\hat{\beta}_t$  that defines the scale parameter at time  $t$  is naturally given by an exponential smoothing scheme

$$\hat{\beta}_t = \lambda \hat{\beta}_{t-1} + (1 - \lambda) \left( y_t - \boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k} \right)^2 \quad (18)$$

The model (4) linking location and scale parameters of predictive densities may not perfectly be the true model for this relationship, especially close to the bounds of the unit interval. In view of (5), one sees that such discrepancy will clearly get more amplified as the location parameter gets closer to these same bounds — i.e. owing to small denominator values. Therefore in order to have a robust estimation of the scale parameter, it is proposed here to introduced a weight function in the exponential smoothing scheme of (18). This weight function is meant to downweight observations as they get closer to the bounds of the unit interval, while respecting the assumed model linking location and scale parameters. This is why the weight function has a shape similar to that of model (4), also illustrated in Figure 2. It however takes a maximum value of 1. The weight  $w_t$  is given by

$$w_t = 4\boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k} \left( 1 - \boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k} \right) \quad (19)$$

while the robust version of the exponential smoothing scheme of (18) writes

$$\hat{\beta}_t = \lambda_t^* \hat{\beta}_{t-1} + (1 - \lambda_t^*) \left( y_t - \boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k} \right)^2 \quad (20)$$

with

$$\lambda_t^* = 1 - (1 - \lambda)w_t \quad (21)$$

It appears appropriate to initialize  $\boldsymbol{\theta}_t$  so that the AR dynamics for the location parameter corresponds to a random walk model, while  $\beta_t$  may be initialized with a small value or alternatively with an expert guess. In parallel, the inverse covariance matrix  $\hat{\mathbf{R}}_t$  can be initialized as an identity matrix times a small constant, yielding a small load on its diagonal. Only the updating formula (16) is used in the recursive estimation procedure as long as  $\hat{\mathbf{R}}_t$  is not invertible. When that point is reached, both (17) and (20) are also employed at each time step for updating the models for the location and scale parameters.

### 3.3 Conditional parametric autoregressive dynamics

If considering CP-AR dynamics instead, the model for the location parameter  $\mu_{t+k}$  is changed to

$$\mu_{t+k} = \boldsymbol{\theta}(\omega)^\top \tilde{\mathbf{y}}_t \quad (22)$$

where

$$\boldsymbol{\theta}(\omega) = [\theta_0(\omega) \ \theta_1(\omega) \ \dots \ \theta_l(\omega)]^\top$$

while  $\tilde{\mathbf{y}}_t$  is defined as in (12). This translates to saying that the vector of constant parameters  $\boldsymbol{\theta}$  in (12) is replaced by coefficient functions of some covariate, these functions being of unknown form.

In parallel the model for the scale parameter then becomes

$$\sigma_{t+k}^2 = \beta(\omega) \quad (23)$$

also implying that this parameter is an unknown function of this same covariate. The maximum lag  $l$  in the dynamic model above can be chosen similarly to that for the more simple AR dynamics. The above model could be extended to the case of several covariates influencing the dynamics of the location parameter, as well as the scale parameter. Hastie and Tibshirani<sup>37</sup> warn us about the fact that the dimensionality of  $u$  should be kept low, say, below 3. Recent developments on conditional parametric and the more general varying-coefficient models have recently been reported by Fan and Zhang<sup>38</sup>.

One may think of a number of meteorological variables that would impact the AR dynamics for the location parameter, as well as the scale parameter itself. The most obvious variable however is wind direction, as it directly induces shadowing effects within the wind parks, thus affecting the dynamic response of a wind park to meteorological conditions. It is also known that for a large number of interesting environments for installing wind parks (offshore, coastal areas or complex terrain topographies for instance), wind direction directly influences the wind speed dynamics themselves<sup>24</sup>. We will denote by  $\omega_t$  the wind direction measurement at time  $t$ ,  $\omega_t \in [0, 2\pi]$ .

It is chosen to estimate the coefficient functions  $\theta(\omega)$  and  $\beta(\omega)$  in a nonparametric framework, i.e. without assuming a shape for these functions. The method for their adaptive and recursive estimation combines local kernel regression and RLS with exponential forgetting. The estimation procedure described below is in the spirit of Nielsen *et al.*<sup>39</sup> and Pinson *et al.*<sup>40</sup>, to which the reader is referred to for more details. Local kernel regression is the simplest version of local polynomial regression<sup>41</sup>. The only assumption on the coefficient functions is that they are sufficiently smooth to be locally approximated with constants. The estimation problem reduces to locally fitting linear models at a number  $m$  of fitting points  $\omega_{(j)}$ . Defining these fitting points is best done by using information on the distribution of  $\omega$ .

Let us focus on a single fitting point  $\omega_{(j)}$  only. The objective function to be minimized at each time  $t$  consists of a modified version of that used for adaptive estimation of the AR-dynamic model considered above. The estimate of the set of local coefficients at  $\omega_{(j)}$  and at time  $t$  is defined as

$$\hat{\theta}_{(j),t} = \arg \min_{\theta_{(j)}} \sum_{i=1}^t \Lambda_{(j),t}(i) c_{(j),i} \left( y_i - \theta_{(j)}^\top \tilde{\mathbf{y}}_{i-k} \right)^2 \quad (24)$$

where the weights  $c_{(j),i}$  to be assigned to past observations are given by a Kernel function of the form

$$c_{(j),i} = T \left( \frac{|\omega_i - \omega_{(j)}|}{\tilde{h}_{(j)}} \right) \quad (25)$$

In the above,  $|\cdot|$  denotes a polar distance, being consistent with  $\omega$  being a circular variable.  $\tilde{h}_{(j)}$  is the bandwidth of the kernel assigned to  $\omega_{(j)}$ .  $T$  may for instance be chosen as a tricube function as in Cleveland and Devlin<sup>41</sup>,

$$T(v) = \begin{cases} (1 - v^3)^3, & v \in [0, 1] \\ 0, & v > 1 \end{cases} \quad (26)$$

In parallel in (24),  $\Lambda_{(j),t}$  is the function that permits exponential forgetting of past observations, i.e.

$$\Lambda_{(j),t}(i) = \begin{cases} \lambda_{(j),t}^{\text{eff}} \Lambda_{(j),t-1}(i), & 1 \leq i \leq t-1 \\ 1, & i = t \end{cases}, \quad (27)$$

$\lambda_{(j),t}^{\text{eff}}$  is the effective forgetting factor for the fitting point  $\omega_{(j)}$ . It follows the definition given by<sup>39</sup>, which links  $\lambda_{(j),t}^{\text{eff}}$  and  $c_{(j),t}$  so that

$$\lambda_{(j),t}^{\text{eff}} = 1 - (1 - \lambda)c_{(j),t} \quad (28)$$

where  $\lambda$  is the classical user-defined forgetting factor,  $\lambda \in (0, 1)$ . This effective forgetting factor ensures that old observations are downweighted only when new information is available.

A little algebra, which is skipped here since extensively covered by e.g. Nielsen *et al.*<sup>39</sup> or Pinson *et al.*<sup>40</sup>, yields the set of recursive formulae for the adaptive estimation of the local coefficients  $\hat{\theta}_{(j),t}$ :

$$\varepsilon_{(j),t} = y_t - \hat{\theta}_{(j),t-1}^\top \tilde{\mathbf{y}}_{t-k} \quad (29)$$

$$\mathbf{R}_{(j),t} = \lambda_{(j),t}^{\text{eff}} \mathbf{R}_{(j),t-1} + c_{(j),t} \tilde{\mathbf{y}}_{t-k} \tilde{\mathbf{y}}_{t-k}^\top \quad (30)$$

$$\hat{\theta}_{(j),t} = \hat{\theta}_{(j),t-1} + c_{(j),t} \mathbf{R}_{(j),t}^{-1} \tilde{\mathbf{y}}_{t-k} \varepsilon_{(j),t} \quad (31)$$

One sees that when the weight  $c_{(j),t}$  equals 0 (thus meaning that the local estimates should not be affected by new information), then one has  $\hat{\theta}_{(j),t} = \hat{\theta}_{(j),t-1}$  and  $\mathbf{R}_{(j),t} = \mathbf{R}_{(j),t-1}$ . This confirms the role of the effective forgetting factor, i.e. to downweight old observations, but only when new information is available.

In parallel, one has to estimate  $\hat{\beta}_{(j),t}$ , the local constant that defines the scale parameter at time  $t$  and at  $\omega_{(j)}$ . As an extension to (20), it is given by

$$\hat{\beta}_{(j),t} = \lambda_{(j),t}^* \hat{\beta}_{(j),t-1} + (1 - \lambda_{(j),t}^*) \left( y_t - \boldsymbol{\theta}_{(j),t}^\top \tilde{\mathbf{y}}_{t-k} \right)^2 \quad (32)$$

with

$$\lambda_{(j),t}^* = 1 - (1 - \lambda)w_t c_{(j),t} \quad (33)$$

$w_t$  still being defined by (19).

When starting the recursive process, the local inverse covariance matrices  $\mathbf{R}_{(j),t}$ , as well as the local parameter estimates  $\hat{\theta}_{(j),t}$  and  $\hat{\beta}_{(j),t}$ , can be initialized for all fitting points as in Section 3.2 for the case of AR dynamics. The value of the parameter estimates  $\hat{\theta}_t(\omega)$  and  $\hat{\beta}_t(\omega)$  for any  $\omega$  can then be obtained by interpolation through the estimated local parameters at the various fitting points  $\omega_{(j)}$  ( $j = 1, \dots, m$ ).

## 4 Application and results

Focus is given here to the application of the methodology and models described above to the test case of an offshore wind farm, for 10-minute ahead forecasting of wind power production. Such lead time corresponds to one-step ahead forecasts in view of the temporal resolution of the time

series considered. Note though that the methodology has been introduced for general direct  $k$ -step ahead forecasting. Indeed, one could work along the lines of Gneiting *et al.* <sup>16</sup> for instance, and build models for direct 2-step ahead forecasting. It is known that purely statistical models for wind power forecasting are in order for forecast horizons up to 6-8 hours ahead, while for further look-ahead time dynamic information from meteorological forecasts should also be considered as input<sup>12,13</sup>.

The test case and available data are introduced in a first stage, followed by a detailed description of the model configuration and estimation setup employed. We also discuss the way forecasts are evaluated, as well as the benchmarks considered. Application results and related comments are finally gathered.

#### 4.1 Test case and available data

We consider the case of the offshore wind farm located at Horns Rev, off the west coast of Jutland in Denmark. This wind farm has a nominal capacity  $P_n$  of 160 MW. The original wind and power measurement data includes raw one-second measurements for each wind turbine. Focus is given to the total power output, normalized by  $P_n$ , and to the average wind direction for the wind farm. A sampling procedure has been developed in order to obtain time-series of 10-minute wind speed, direction and power averages. This averaging rate makes that the very fast fluctuations related to the turbulent nature of the wind disappear and reveal slower fluctuations at the minute scale. Because there may be some erroneous or suspicious data in the raw measurements, the sampling procedure has a threshold parameter  $\tau_v$ , which corresponds to the minimum percentage of data that need to be considered as valid within a given time interval, so that the related power (or wind direction) average is considered as valid too. The threshold chosen is  $\tau_v = 75\%$ . The available raw data is from 16<sup>th</sup> February 2005 to 25<sup>th</sup> January 2006, consisting of 49536 data points. As a result of the averaging procedure, the wind direction and power generation dataset consists of 41651 valid data points.

#### 4.2 Model configuration, estimation setup and benchmarking

From the available data, two periods are defined, the first one being used for identification (and initial training) of the statistical models, and the second one for evaluating what the performance of these models may be in operational conditions. These two parts of the dataset are referred to as learning and evaluation sets. The first 15120 data points (with 11968 valid ones) are employed as the learning set, exactly covering the months from February to May 2005. The remainder of the dataset, covering a period from June 2005 to January 2006 is used for out-of-sample evaluation of 1-step ahead forecast performance. This evaluation set contains 34536 data points, including 29683 valid ones.

Over the learning period, a part of the data is used for one-fold cross validation (the last 9000 points) in order to select an optimal autoregressive order for the dynamic models, the values for forgetting factors, the shape parameter of the GL-transformation, as well as the fitting points and bandwidths for the case of the conditional parametric models. Actually, instead of considering the forgetting factor itself, it is preferred to use the corresponding effective number of observations  $n_\lambda$ . It allows one to better appraise the size of the equivalent ‘sliding window’ in the adaptive estimation of the dynamic model parameters. In parallel, the censoring parameter  $\epsilon$  is arbitrarily set to  $\epsilon =$

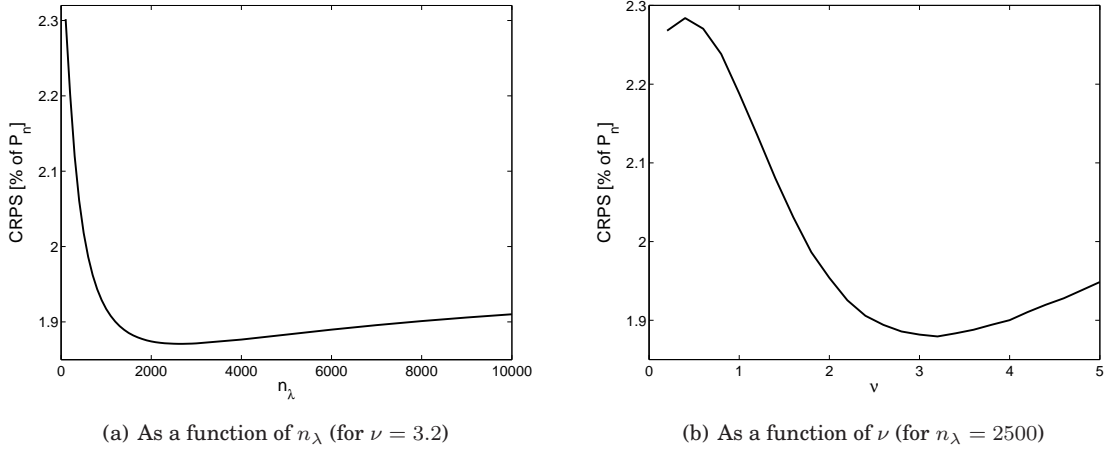
0.001, corresponding to the resolution of the power measurements. Selection of optimal values for the model structure and parameters  $n_\lambda$  and  $\nu$  is done in a trial-and-error manner, by evaluating the results obtained from a set of different setups. For more information on cross validation, we refer to Stone<sup>42</sup>. The criterion to be minimized over the cross-validation set is the Continuous Rank Probability Score (CRPS) of 1-step ahead density forecasts. The CRPS is a common proper skill score used for the evaluation of density forecasts<sup>43</sup>. For the case of the CP-AR dynamics, the bandwidth values are selected similarly.

Focusing on the order of the various models first, results from Pinson and Madsen<sup>19</sup> suggest that a maximum lag  $l = 3$  is sufficient for capturing the dynamics in offshore wind power fluctuations with a 10-minute resolution, at the Horns Rev wind farm. We indeed verified here that for all models considered a significant decrease in 1-step ahead CRPS values can be observed when augmenting  $l$  up to  $l = 3$ , and that no further benefit is obtained by further increasing the value of  $l$ . Applying a parsimonious model principle, we then decide to concentrate on models of order  $l = 3$  only. In all cases, the AR models are initialized as simple random walk models, with very low variance of predictive densities.

In the case of the GL-Normal densities and AR dynamics, one may consider the selection of  $n_\lambda$  and  $\nu$  to be performed jointly or independently. It has been observed for the data of the present case-study that employing both ways yielded similar results. We thus present them below in Figure 3 as if  $n_\lambda$  and  $\nu$  were optimized one after the other, independently. More precisely, Figure 3(a) depicts the variations in 1-step-ahead CRPS for  $\nu$  fixed to its optimal value, and with  $n_\lambda$  varying between 100 and 10.000, with steps of 100 observations. Similarly, Figure 3(b) shows the evaluation of the 1-step-ahead CRPS for  $n_\lambda$  fixed to its optimal value, and with  $\nu$  varying between 0.2 and 5, with increments of 0.2. The overall minimum 1-step-ahead CRPS value on the cross-validation set is obtained for  $n_\lambda = 2500$  and  $\nu = 3.2$ . Figure 3(b) clearly shows the interest of introducing the generalized form of the logit transformation, since a significant decrease in CRPS is seen when considering negatively-skewed mean-variance models, i.e. for  $\nu > 1$  (see Figure 2(b) and related text in Section 2). One notices though that our proposal probabilistic forecasting methodology seems to be fairly robust with respect to the choice of  $n_\lambda$  and  $\nu$  values, as the 1-step-ahead CRPS of predictive densities does not vary much for  $n_\lambda \in [2000, 4000]$  and for  $\nu \in [2.5, 3.5]$ .

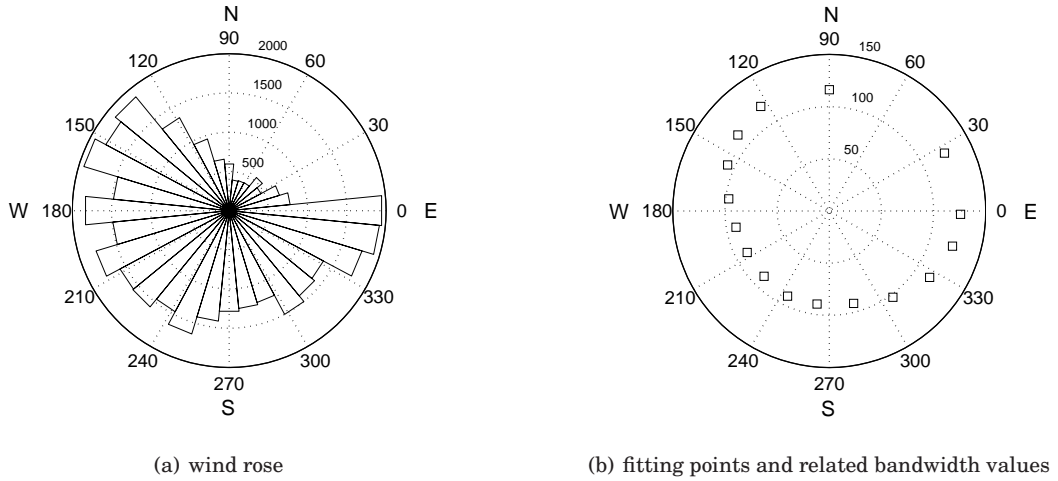
Turning our attention to the GL-Normal predictive densities with CP-AR dynamics, an exercise similar to that performed above led to the same values for  $n_\lambda$  and  $\nu$ , i.e.  $n_\lambda = 2500$  and  $\nu = 3.2$ . In addition here, the location of the fitting points and related bandwidth values for the local fitting of coefficient functions have to be selected. It has been arbitrarily decided to use 16 fitting points, since such number allows a fairly good coverage of potential wind directions values, while limiting the computational power needed. The spreading of these fitting points has been performed in view of the distribution of wind direction values over the learning period, i.e. such that there is the same frequency of data between any two successive fitting points. Regarding the Kernel bandwidth in (25), both fixed and nearest-neighbor bandwidths have been envisaged. The latter concept was preferred, in order to be consistent with the idea introduced above, such that it would be more consistent to respect the distribution of wind direction data when defining the fitting points. Following Nielsen *et al.*<sup>39</sup>, the nearest-neighbor bandwidth is defined by the share of the (wind direction) data that should be covered by the Kernel attached to any single fitting point. This share is chosen here to be 60% in order to obtain sufficiently smooth variations of the coefficient functions in the CP-AR model. Smaller values have been tested, though not improving the forecast performance of the model over the cross-validation set. The resulting setup for the fitting points and related





**FIGURE 3:** Evolution of the 1-step-ahead CRPS of GL-Normal density forecasts of wind power generation based on simple AR-dynamics. These results are obtained on the cross-validation set, as a function of the parameters  $n_\lambda$  and  $\nu$ .

bandwidth values is depicted in Figure 4, along with the corresponding wind rose at Horns Rev. In the polar representation of Figure 4(b), the polar coordinates give the location of the fitting points, while the bandwidths values are given by the ordinate of the various points. It is visually straightforward to recognize that less fitting points and higher bandwidth values are present in areas with less data available.



**FIGURE 4:** Wind rose, fitting points and related bandwidths values based on the distribution of wind direction values over the learning set. All values are in degrees, following the trigonometric convention.

As a basis for the evaluation of the forecast performance of the methodology introduced in the paper, two types of benchmarks are considered. First of all, the persistence forecast method, based on a simple random walk model, is known to be difficult to outperform for such short look-ahead times<sup>12,13,19</sup>. In parallel, in order to demonstrate the interest of proposing GL-Normal predictive densities against the more classical Gaussian framework approach, the other benchmark considered is based on censored Gaussian predictive densities, with AR dynamics for the location parameter, a constant scale parameter, these parameters also being adaptively and recursively esti-

mated. For this benchmark, the optimal value found for the effective number of observations  $n_\lambda$  is  $n_\lambda = 2000$ .

### 4.3 Results and comments

A fully probabilistic approach to forecast evaluation is employed. By that is meant that focus is given to evaluating the quality of predictive densities, including their reliability and overall skill, as well as to assessing the quality of some point forecasts that can be extracted from such predictive densities. Following the point of Gneiting *et al.* <sup>6</sup> among others, one should extract the optimal point forecasts from predictive densities based on the target evaluation score. Indeed, the median of predictive densities should be selected if the target evaluation score is of the Mean Average Error type (abbreviated MAE, or NMAE for its normalized version). Similarly, if the target score is of quadratic nature like the Root Mean Square Error (RMSE, or NRMSE for its normalized version), the optimal point forecast to be extracted is the expectation of predictive densities.

In view of these aspects, let us focus first on the quality of point forecasts related to the median and expectation of predictive densities, since the corresponding NMAE and NRMSE scores appear to be the most employed ones in the wind power forecasting literature <sup>12,13,44</sup>. Owing to the length of our evaluation set (approximately 8 months), it is chosen to study these scores overall and on a monthly basis. The evaluation results for point forecasts based on Persistence, the Gaussian AR benchmark, as well as for the proposed GL-Normal AR and CP-AR predictive densities, are collated in Table 1. Whatever the score considered, the Persistence benchmark indeed seems competitive: the more advanced approaches only propose overall improvements up to 5%. This is due to the inertia in local atmospheric processes at such temporal scales. Even though monthly fluctuations can be observed in the NMAE and NRMSE scores, it appears that the Normal AR benchmark is almost consistently outperforming Persistence, while being outperformed by the point forecasts extracted from GL-Normal predictive densities. A slight gain in point forecast accuracy is also seen if considering that the moments of GL-Normal predictive densities may vary as a function of wind direction, as is the case with the CP-AR dynamic model.

**TABLE 1:** Monthly results related to the evaluation of potential point forecasts extracted from density forecasts: focus on the median and the expectation.

(a) NMAE for the median (in % of the nominal capacity  $P_n$ )

Month	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	All
Persistence	2.40	2.43	2.50	2.50	2.40	2.59	2.61	2.47	2.48
Normal AR	2.34	2.42	2.48	2.43	2.33	2.64	2.63	2.45	2.45
GL-Normal AR	2.32	2.35	2.43	2.38	2.28	2.47	2.54	2.40	2.38
GL-Normal CP-AR	2.31	2.34	2.42	2.37	2.26	2.47	2.53	2.36	2.36

(b) NRMSE for the expectation (in % of the nominal capacity  $P_n$ )

Month	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	All
Persistence	3.98	4.68	4.47	4.66	4.17	6.23	4.76	4.28	4.70
Normal AR	3.80	4.56	4.36	4.46	3.97	6.00	4.62	4.20	4.54
GL-Normal AR	3.81	4.57	4.32	4.43	3.96	5.87	4.55	4.12	4.50
GL-Normal CP-AR	3.79	4.53	4.30	4.42	3.93	5.87	4.56	4.08	4.47

In the second stage, we look at the quality of predictive densities, concentrating first on their overall skill by employing the CRPS criterion, being a proper skill score. Extensive discussion about the evaluation of probabilistic forecasts through the use of proper scoring rules can be found in Gneiting *et al.* <sup>43</sup>. For the case of predictive densities, employing Persistence as a benchmark does not make much sense anymore, since an additional assumption on the variance of the random walk model should be made. This could be performed either by assuming a second-order stationary process and therefore estimating the variance based on the learning set, or alternatively by using an exponential smoothing approach for tracking the variance of the process, as is done here for the various dynamical models. Another benchmark that is often considered for the skill evaluation of probabilistic forecasts of meteorological or weather-related processes is climatology, which is based on always issuing the same unconditional predictive density built from all historical observations available. It is known however that this benchmark is very easy to outperform for short-term horizons, and is hence also not relevant. In contrast here, it is preferred to consider the Normal predictive densities with AR dynamics as the benchmark, since necessarily outperforming Persistence in terms of the CRPS criterion (otherwise, the optimal lag of the model found through the cross-validation exercise would be 0). These Normal predictive densities also are a natural benchmark in view of our argument such that GL-Normal predictive densities should be preferred to Normal ones. Similarly to the above, the CRPS criterion is calculated on a monthly basis as well as globally, and this for the three competing types of predictive densities considered. The corresponding evaluation results are gathered in Table 2.

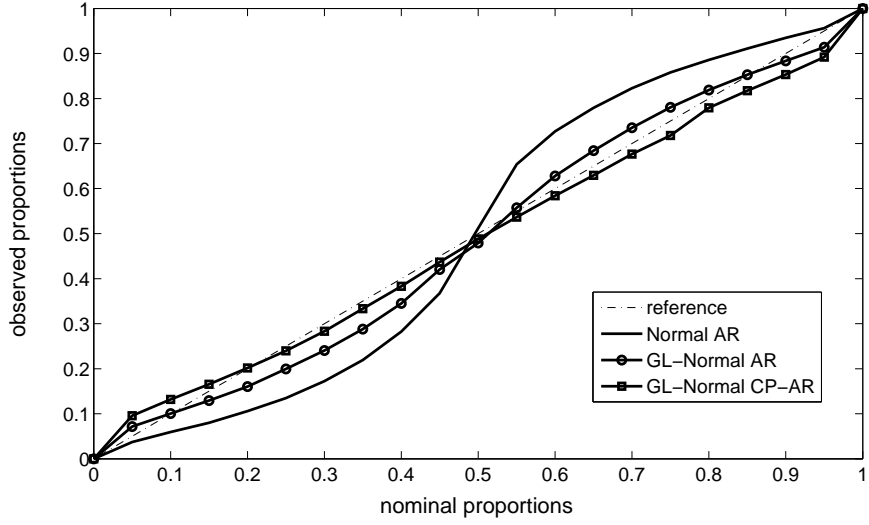
**TABLE 2:** *Monthly results related to the evaluation of density forecasts with a CRPS criterion (in % of the nominal capacity  $P_n$ ).*

Month	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	All
Normal AR	1.85	2.01	1.99	1.98	1.87	2.31	2.23	1.97	2.01
GL-Normal AR	1.76	1.80	1.86	1.86	1.81	1.93	1.92	2.03	1.86
GL-Normal CP-AR	1.73	1.78	1.82	1.81	1.71	1.94	1.88	1.79	1.80

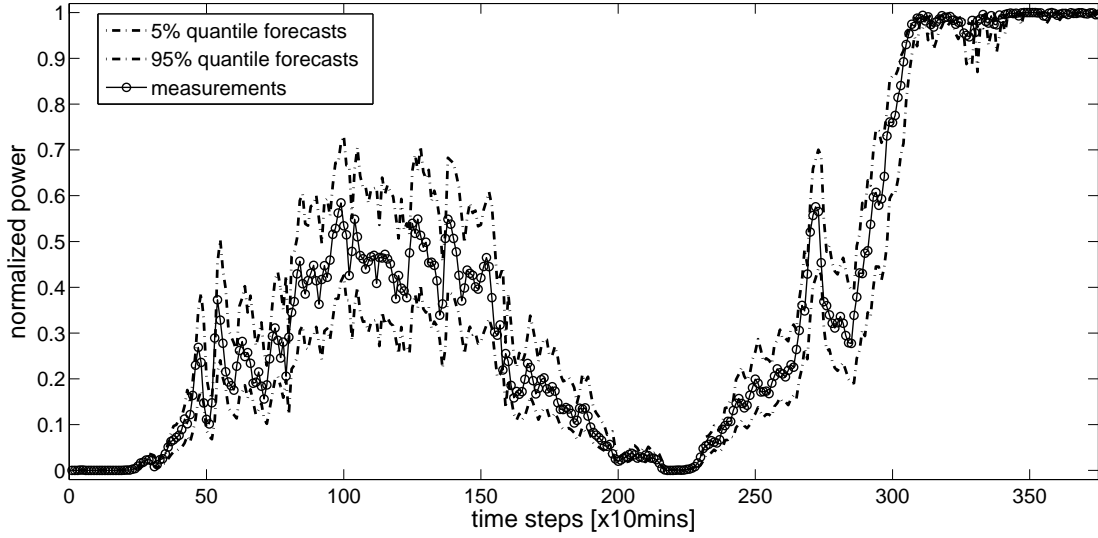
The improvements in terms of the CRPS criterion are significant when going from the Normal predictive densities to GL-Normal predictive densities, in the order of 7.5%, and consistent over the evaluation period. Like for the case of point forecasts in the above, an additional improvement is observed if considering that the parameters of GL-Normal densities should be made a function of wind direction. The overall decrease in the CRPS criterion when going from Normal predictive densities with AR dynamics to GL-Normal predictive densities with CP-AR dynamics is of 10.5%.

We somehow expect that such improvement comes from the fact that GL-Normal predictive densities are more probabilistically reliable (or in other words, calibrated) than the Normal ones. Calibration of predictive densities can be assessed through the use of e.g. Probability Integral Transform (PIT) histograms, as carried out by Gneiting *et al.* <sup>16</sup>, or with reliability diagrams in the form of quantile-quantile plots, as in Figure 5. Such reliability diagrams depict the observed proportions of a set of quantiles composing predictive densities of wind power generation against the nominal ones. Here the set of quantiles is defined with a step of 5% in their nominal proportions. This assessment of the reliability of calibration of the quantiles composing predictive densities directly builds on their very definition: a quantile forecast with nominal proportion  $\alpha$  should cover the actual observation  $\alpha\%$  of the time.

One easily observes from Figure 5 that the benchmark Gaussian predictive densities introduce a systematic and significant probabilistic bias in the probabilistic forecasts they provide. Predictive



**FIGURE 5:** Evaluation of the reliability of predictive densities with reliability diagrams in the form of quantile-quantile plots.



**FIGURE 6:** Episode of 375 time-steps (with a time resolution of 10 minutes), with quantile forecasts of nominal proportions 0.05 and 0.95, along with related normalized wind power measurements. These quantile forecasts are extracted from the predictive densities with CP-AR dynamics.

densities seem to be too wide on average, since observed proportions are larger than the nominal ones for nominal proportions above 0.5, and inversely for nominal proportions below 0.5. A simple explanation for that is that the variance of predictive densities is not conditional to the level of wind generation. Since wind generation is quite often at a medium level, the exponentially smoothed error variance used for the variance of predictive densities is fairly high. The variance is consequently overpredicted when wind generation is at a low and high level. In contrast when employing GL-Normal distributions for probabilistic forecasting, their natural mean-variance model make that

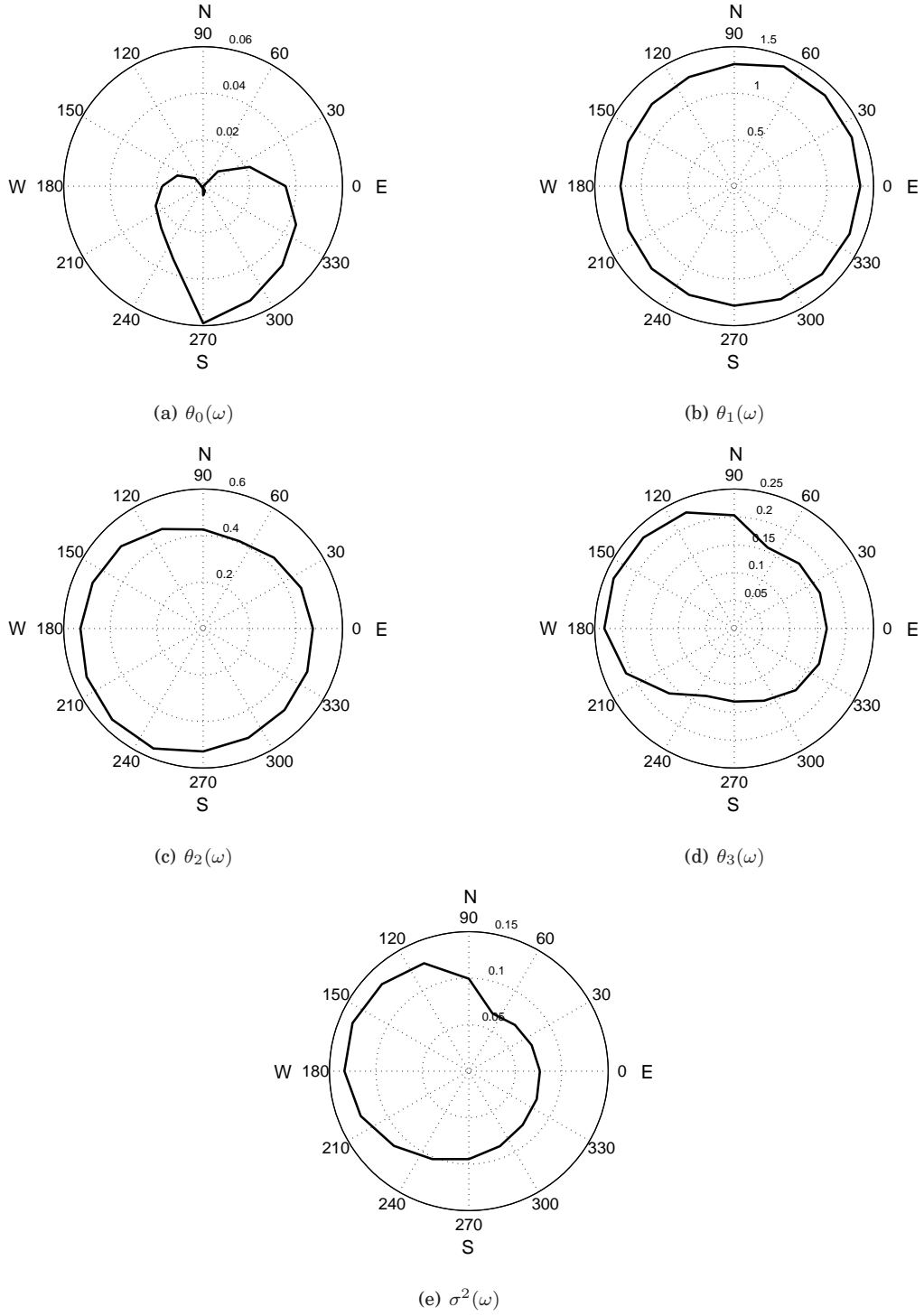
predictive densities do not tend to be systematically over- or under-dispersed. This appears to be valid for both types of dynamics considered for the mean of predictive densities.

For illustrative purposes, Figure 6 depicts an episode with 375 successive time-steps of quantile forecasts (with nominal proportions 0.05 and 0.95) extracted from GL-Normal predictive densities with CP-AR dynamics, along with the corresponding measurements. One can easily observe that the predictive densities are naturally bounded between 0 and 1, and that the variance of GL-Normal densities is larger when the conditional expectation of the process is at a medium level. It is harder though to see that both dynamics and variance of the predictive densities are a function of the wind direction onsite. In that objective, let us also have a look at the parameters of the CP-AR dynamic model, as well as at the variance of the predictive densities, at the end of the evaluation set. These various quantities are represented in polar coordinates in Figure 7. For all parameters, some significant variations as a function of the wind direction can be observed. Most interestingly, the variance of predictive densities significantly vary as a function of  $\omega$ , being higher for Westerly and North-Westerly winds, and lower for the other sectors. This goes in line with the extensive wind data analysis performed by Vincent *et al.*<sup>24</sup> and Vincent *et al.*<sup>45</sup> based on meteorological measurements obtained at Horns Rev. It was indeed observed that wind fluctuations (and by extension wind power fluctuations) tend to have a higher variance for winds coming from the North Sea, thus from Westerly and North-Westerly directions, than if winds come from land.

## 5 Concluding remarks

A general approach suitable for probabilistic forecasting of wind power time-series has been presented. Predictive distributions take the form of discrete-continuous mixtures consisting of a newly introduced GL-Normal distribution, with potential concentration of probability mass at the bounds of the unit interval  $[0, 1]$ . It has been argued that GL-Normal distributions permit to account for the nonlinear and bounded characteristics of wind power generation, and that they should be seen as a basic feature of probabilistic forecasting methodologies whatever the forecast horizons considered, from a few minutes to few days ahead. Being fully characterized by a location and a scale parameter only (assuming that the shape parameter is a meta-parameter estimated once and for all from the available data), these predictive densities allow for the proposal of fairly simple dynamic models. These models can for instance, as is done here, account for the own autoregressive dynamics of the wind power generation process, or the nonlinear effect of some covariate e.g. wind direction on these dynamics.

These models could be extended in the future to account for additional effects or for additional covariates e.g. meteorological measurements or predictions, depending upon the temporal resolution and look-ahead times of interest. The proposal of new dynamical models for probabilistic time-series forecasting of wind power generation should be supported by some of the recent extensive data analysis studies performed based on wind, wind power and other meteorological data. For the case of this offshore site, the results from Pinson and Madsen<sup>19</sup> indeed suggest that regime-switching may be present in the time-series of wind power generation, while the results presented in Vincent *et al.*<sup>45</sup> appear to confirm such a statement, with potential explanations being related to local thermal effects and/or to the passage of rain fronts in the vicinity or over the site of interest. Note that such regime-switching may affect the dynamics of both the mean and variance of predictive densities. This regime-switching idea may also be relevant for onshore sites, going along the



**FIGURE 7:** *CP-AR model parameters as a function of the wind direction  $\omega$  at the end of the evaluation set.*

lines of the models proposed by <sup>16</sup>, in order to account for channeling and thermal effects.

From a more methodological point of view, the use of the proposed GL-Normal distributions may be envisaged for the modeling of other nonlinear and bounded processes where it is expected or known that the effects of the bounds is not symmetrical. If focusing on the wind power application, the

shape parameter  $\nu$  of GL-Normal distributions will surely be depending upon the site of interest. In the general manner, this parameter will certainly depend on the process considered and should be estimated from data. The cross-validation approach employed here appeared to yield acceptable results. It may however be envisaged to develop a more rigorous framework based on e.g. maximum likelihood estimation for selecting  $\nu$ .

## Acknowledgments

The work presented has been partly supported by the European Commission under the SafeWind project (ENK7-CT2008-213740), by the Danish ForskEL programme through the project ‘Radar@Sea’ (ForskEL 2009-1-0226), and finally by the Danish Research Council for Technology and Production Sciences (FTP-274-08-0573), which are hereby acknowledged. The authors would like to thank Vattenfall Denmark for originally providing the wind power measurements for the Horns Rev wind farm, Pierre-Julien Trombe for the data processing and quality checking, as well as Tilmann Gneiting for inspiring discussions on probabilistic forecasting and bounded time-series.

## References

1. Gneiting T. Editorial: Probabilistic forecasting. *Journal of the Royal Statistical Society, Series A* 2008; **171**: 319–321.
2. Abramson B, Clemen R. Probability forecasting. *International Journal of Forecasting* 1995; **11**: 1–4.
3. Tay AS, Wallis KF. Density forecasting: a survey. *Journal of Forecasting* 2000; **19**: 235–254.
4. Timmermann A. Density forecasting in economics and finance. *Journal of Forecasting* 2000; **19**: 231–234.
5. Pinson P, Chevallier C, Kariniotakis G. Trading wind generation with short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems* 2007; **22**: 1148–1156.
6. Gneiting T. Quantiles as optimal point predictors. *International Journal of Forecasting* 2010; available online.
7. Nielsen HAa, Nielsen TS, Madsen H, Badger J, Giebel G, Landberg L, Sattler K, Voulund L, Tøfting J. From wind ensembles to probabilistic information about future wind power production - Results from an actual application. *Proc. IEEE PMAPS Conference, Probabilistic Methods Applied to Power Systems*, Stockholm, Sweden.
8. Pinson P, Madsen H. Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy* 2009; **12**: 137–155.
9. Bremnes JB. A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy* 2006; **9**: 3–11.
10. Møller JK, Nielsen HAa, Madsen H. Time-adaptive quantile regression. *Computational Statistics and Data Analysis* 2008; **52**: 1292–1303.
11. Pinson P, Kariniotakis G. Conditional prediction intervals of wind power. *IEEE Transactions on Power Systems* 2010; in press.
12. Giebel G, Kariniotakis G, Brownsword R. The state of the art in short-term prediction of wind power - A literature overview. Technical report, ANEMOS EU project, deliverable report D1.1, 2003. [Available online: <http://www.anemos-project.eu>]
13. Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews* 2008; **12**: 1725–1744.
14. Kristoffersen JR, Christiansen P. Horns Rev offshore wind farm: its main controller and remote control system. *Wind Engineering* 2003; **27**: 351–359.
15. Sørensen PE, Cutululis NA, Viguera-Rodríguez A, Jensen LE, Hjerrild J, Donovan MH, Madsen H. Power fluctuations from large wind farms. *IEEE Transactions on Power Systems* 2007; **22**: 958–965.



16. Gneiting T, Larson K, Westrick K, Genton MG, Aldrich E. Calibrated probabilistic forecasting at the state-line wind energy center - The regime-switching space-time method. *Journal of the American Statistical Association* 2006; **101**: 968–979.
17. Hering AS, Genton MG. Powering up with space-time wind forecasting. *Journal of the American Statistical Association* 2010; **105**: 96–104.
18. Ailliot P, Monbet V, Prevosto M. An autoregressive model with time-varying coefficients for wind fields. *Environmetrics* 2006; **17**: 107–117.
19. Pinson P, Madsen H. Adaptive modeling and forecasting of wind power fluctuations with Markov-switching autoregressive models. Technical report no. IMM-2009-11, Technical University of Denmark, DTU Informatics, Kgs. Lyngby, Denmark, 2009.
20. Lange M. On the uncertainty of wind power predictions - Analysis of the forecast accuracy and statistical distribution of errors. *Journal of Solar Energy Engineering - Transactions of the ASME* 2005; **127**: 177–184.
21. Pinson P. Estimation of the uncertainty in wind power forecasting. Ph.D. thesis, Ecole des Mines de Paris, Paris, France, 2006. [Available online: <http://pastel.paristech.org>]
22. Bludszuweit H, Domínguez-Navarro JA, Llombart A. Statistical analysis of wind power forecast errors. *IEEE Transactions on Power Systems* 2008; **23**: 983–991.
23. Sanchez I. Short-term prediction of wind energy production. *International Journal of Forecasting* 2006; **22**: 43–56.
24. Vincent CL, Giebel G, Pinson P, Madsen H. Resolving non-stationary spectral signals in wind speed time-series using the Hilbert-Huang transform. *Journal of Applied Meteorology and Climatology* 2010; **49**: 253–267.
25. Aitchison J, Shen S. Logistic-normal distributions: some properties and uses. *Biometrika* 1980; **67**: 261–272.
26. Box GEP, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 1964; **26**: 211–243.
27. Carroll RJ, Ruppert D. On prediction and the power transformation family. *Biometrika* 1981; **68**: 609–615.
28. Johnson N. Systems of frequency curves generated by methods of translation. *Biometrika* 1949; **36**: 146–176.
29. Wallis KF. Time series analysis and bounded economic variables. *Journal of Time Series Analysis* 1987; **8**: 115–123.
30. Aranda-Ordaz FJ. On two families of transformations to additivity for binary response data. *Biometrika* 1981; **66**: 59–65.
31. Richards FJ. A flexible growth function for empirical use. *Journal of Experimental Botany* 1959; **10**: 290–300.
32. Usaola J. Probabilistic load flow with wind production uncertainty using cumulants and Cornish–Fisher expansion. *International Journal of Electrical Power and Energy Systems* 2009; **31**: 474–481.
33. Frederic P, Lad F. Two moments of the logitnormal distribution. *Communications in Statistics - Simulation and Computation* 2008; **37**: 1263–1269.
34. Lesaffre E, Rizopoulos D, Tsonoka R. The logistic-transform for bounded outcome scores. *Biostatistics* 2007; **8**: 72–95.
35. Ljung L, Söderström T. *Theory and Practice of Recursive Estimation*. MIT Press: Boston, 1983.
36. Madsen H. *Time Series Analysis*. Chapman & Hall/CRC: London, 2007.
37. Hastie T, Tibshirani R. *Generalized Additive Models*. Chapman & Hall/CRC: London, 1990.
38. Fan JQ, Zhang W. Statistical methods with varying-coefficient models. *Statistics and its Interface* 2008; **1**: 179–195.
39. Nielsen HAa, Nielsen TS, Joensen AK, Madsen H, Holst J. Tracking time-varying coefficient functions. *International Journal of Adaptive Control and Signal Processing* 2000; **14**: 813–828.
40. Pinson P, Nielsen HAa, Madsen H, Nielsen TS. Local linear regression with adaptive orthogonal fitting for the wind power application. *Statistics and Computing* 2008; **18**: 59–71.

41. Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 1988; **83**: 596–610.
42. Stone M. Cross-validation and assessment of statistical predictions (with discussion), *Journal of the Royal Statistical Society, Series B* 1974; **36**: 111–147.
43. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society, Series B* 2007; **69**: 243–268.
44. Madsen H, Pinson P, Nielsen TS, Nielsen HAa, Kariniotakis G. Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering* 2005; **29**: 475–489.
45. Vincent CL, Giebel G, Pinson P. Wind fluctuations in the North Sea. *International Journal of Climatology* 2010, submitted.